主管单位:中华人民共和国工业和信息化部
主办单位:北京航空航天大学

ISSN 1001-5965
CN 11-2625/V

北京航空航天大學学报

JOURNAL OF BEIJING UNIVERSITY OF AERONAUTICS AND ASTRONAUTICS



掠海导弹触水滑跳运动特性数值模拟研究

雷娟棉　吴志翔　解文洋

10 2024
Vol.50

北京航空航天大学学报

第 50 卷　第 10 期　（总第 380 期）　2024 年 10 月

目　　次

掠海导弹触水滑跳运动特性数值模拟研究 …… 雷娟棉，吴志翔，解文洋（2975）
融合双注意力机制和动态图卷积的点云语义分割 …… 杨军，张琛（2984）
液压破碎锤活塞横裂机理分析与工艺优化 …… 童桂英，万怡男，张磊，石岩，张业明（2995）
区间划分改进相似性的发动机剩余寿命预测 …… 赵洪利，白令德（3005）
施工工艺对复合材料蒙皮电气性能影响分析 …… 杨占刚，柯中澍，杨旭伟，包兴旺（3013）
基于可能性理论的红外与可见光视频自适应分层融合算法 …… 吉琳娜，郭小铭，杨风暴（3021）
边防哨所风光耦合制氢系统的配置优化 …… 齐海涛，刘咄，赵东澳，刘旭（3032）
基于 IFC 标准的机场道面结构信息模型存储与传递方法 …… 蔡靖，王瀚雪，戴轩，李岳，马磊（3042）
停留时间与飞行方向对飞行员警觉性的影响 …… 李敬强，胡超，张希凝，张璐，刘安南（3054）
星基增强系统单频服务性能评估方法 …… 金彪，李锐，王盾，刘磊，原晋栩，李子潇（3062）
非均匀热流影响超临界二氧化碳流动换热数值研究 …… 代威，石申舟，付衍琛，左洋（3074）
基于本体的软件安全漏洞模式 …… 胡璇，陈俊名，李海峰（3084）
基于视觉 Transformer 飞行员姿态估计 …… 吴红兰，刘豪，孙有朝（3100）
侧风对直升机沙盲特性的影响 …… 谭剑锋，杨宇霄，张卫国，王世雄，吕琳（3111）
采用温度梯度的干燥反应器 H_2O 穿透深度评估方法 …… 张勇平，卞强，杨润泽，张震，白攀峰（3123）
基于时序知识图谱的电力设备缺陷预测 …… 纪鑫，武同心，杨智伟，余婷，李君婷，何禹德（3131）
基于自适应参数优化 RSSD-CYCBD 的行星齿轮箱复合故障诊断 ……
…… 孙环宇，杨志鹏，王艺玮，郭琦（3139）
面向突防的滑翔制导炮弹弹道规划方法 …… 尹秋霖，陈琦，王中原，王庆海（3151）
基于 SINS/OD/高度计的车载重力异常测量方法 …… 李新宇，张志利，周召发，常振军，赵芝谦（3162）
连续变弯度后缘飞机的滚转机动载荷减缓 …… 雷朝辉，杨超，宋晨，金天燚，吴志刚（3172）
点接触共轭渐开线蜗杆副啮合理论及性能分析 …… 任雯，李杰，王天羽，李海涛（3183）
基于行为特征的雷达辐射源威胁评估 …… 王俊迪，王星，田元荣，陈游（3196）
振镜电机系统高性能数模混合驱动控制方法 …… 赵家康，郭宏，徐金全（3208）
基于 MDk-DPC 的空中目标自动分群方法 …… 马钰棠，孙鹏，张杰勇，闫云飞，赵亮（3219）
机载导航设备辅助的 LEO 卫星定位方法与仿真验证 …… 徐鉴名，黄智刚，李锐（3230）
变弯度柔性后缘机翼动响应减缓试验 …… 张育鸣，戴玉婷，尉濡恺，胡雅婷，杨超（3239）
进口拐角波系对超燃冲压发动机尾喷管的性能影响 …… 陈冲，葛建辉，徐惊雷，苏鹏，俞凯凯（3250）
链上、链下协同的智能合约即服务问题 …… 刘浩宇，马泽祁，叶亚飞，胡凯（3262）
相变储能拓扑翅片的性能研究 …… 殷健宝，邢玉明，王仕淞，叶梦岩，王子贤，侯煦（3274）
一种基于多链协同的联盟链改进模型 …… 赵鲜鲜，谭海波，赵赫，周桐，程昊天，李金泽（3283）

期刊基本参数：CN 11-2625/V * 1956 * m * A4 * 322 * zh * P * ￥50.00 * 400 * 30 * 2024-10

（**编　辑**　张　嵘　孙　芳　卞欢欢　王　茜　李亚泰）

JOURNAL OF BEIJING UNIVERSITY OF AERONAUTICS AND ASTRONAUTICS

Vol. 50 No. 10 (Sum 380) October 2024

CONTENTS

Numerical simulation investigation on water surface skipping motion characteristics of sea-skimming projectile
…… LEI Juanmian, WU Zhixiang, XIE Wenyang (2975)
Semantic segmentation of point clouds by fusing dual attention mechanism and dynamic graph convolution
…… YANG Jun, ZHANG Chen (2984)
Mechanism analysis and process optimization of transverse cracking of hydraulic crushing hammer piston
…… TONG Guiying, WAN Yinan, ZHANG Lei, SHI Yan, ZHANG Yeming (2995)
Remaining life prediction of engine by improved similarity with interval partition
…… ZHAO Hongli, BAI Lingde (3005)
Analysis of effect of construction process on electrical properties of composite skins
…… YANG Zhangang, KE Zhongshu, YANG Xuwei, BAO Xingwang (3013)
Adaptive layered fusion algorithm for infrared and visible video based on possibility theory
…… JI Linna, GUO Xiaoming, YANG Fengbao (3021)
Configuration optimization of wind/photovoltaic hydrogen production system at frontier sentries
…… QI Haitao, LIU Duo, ZHAO Dong'ao, LIU Xu (3032)
Storage and transfer method of airport pavement structure information model based on IFC standard
…… CAI Jing, WANG Hanxue, DAI Xuan, LI Yue, MA Lei (3042)
Influence of layover length and flight direction on pilots' alertness
…… LI Jingqiang, HU Chao, ZHANG Xining, ZHANG Lu, LIU Annan (3054)
Service performance assessment method of single frequency SBAS
…… JIN Biao, LI Rui, WANG Dun, LIU Lei, YUAN Jinxu, LI Zixiao (3062)
Numerical study on flow and heat transfer of supercritical carbon dioxide under non-uniform heat flux influences
…… DAI Wei, SHI Shenzhou, FU Yanchen, ZUO Yang (3074)
Software security vulnerability patterns based on ontology
…… HU Xuan, CHEN Junming, LI Haifeng (3084)
Vision Transformer-based pilot pose estimation
…… WU Honglan, LIU Hao, SUN Youchao (3100)
Influence of crosswind on helicopter brownout
…… TAN Jianfeng, YANG Yuxiao, ZHANG Weiguo, WANG Shixiong, LYU Lin (3111)
Evaluation method of H_2O penetration depth of drying reactor based on temperature gradient
…… ZHANG Yongping, BIAN Qiang, YANG Runze, ZHANG Zhen, BAI Panfeng (3123)
Power equipment defect prediction based on temporary knowledge graph
…… JI Xin, WU Tongxin, YANG Zhiwei, YU Ting, LI Junting, HE Yude (3131)
Compound fault diagnosis of planetary gearbox based on RSSD-CYCBD by adaptive parameter optimization
…… SUN Huanyu, YANG Zhipeng, WANG Yiwei, GUO Qi (3139)
Trajectory programming method of gliding-guided projectiles for penetration
…… YIN Qiulin, CHEN Qi, WANG Zhongyuan, WANG Qinghai (3151)
Vehicle gravity anomaly measurement methods based on SINS/OD/altimeter
…… LI Xinyu, ZHANG Zhili, ZHOU Zhaofa, CHANG Zhenjun, ZHAO Zhiqian (3162)
Rolling maneuver load alleviation of aircraft with continuously variable-camber trailing edge
…… LEI Chaohui, YANG Chao, SONG Chen, JIN Tianyi, WU Zhigang (3172)
Meshing theory and performance analysis of point-contact conjugate involute worm gear pair
…… REN Wen, LI Jie, WANG Tianyu, LI Haitao (3183)
Threat assessment of radar radiation sources based on behavioral characteristics
…… WANG Jundi, WANG Xing, TIAN Yuanrong, CHEN You (3196)
High performance control method for galvanometer laser scanner system
…… ZHAO Jiakang, GUO Hong, XU Jinquan (3208)
Aerial target automatic grouping method based on MDk-DPC
…… MA Yutang, SUN Peng, ZHANG Jieyong, YAN Yunfei, ZHAO Liang (3219)
LEO satellite positioning method and simulation verification aided by airborne navigation equipment
…… XU Jianming, HUANG Zhigang, LI Rui (3230)
Experiment on dynamic response alleviation of a wing with variable-camber flexible trailing edge
…… ZHANG Yuming, DAI Yuting, WEI Rukai, HU Yating, YANG Chao (3239)
Influence of inlet corner wave system on performance of scramjet nozzle
…… CHEN Chong, GE Jianhui, XU Jinglei, SU Peng, YU Kaikai (3250)
Smart contracts-as-a-service with on-chain and off-chain collaboration
…… LIU Haoyu, MA Zeqi, YE Yafei, HU Kai (3262)
Study of performance of topological fin for phase change energy storage
…… YIN Jianbao, XING Yuming, WANG Shisong, YE Mengyan, WANG Zixian, HOU Xu (3274)
A consortium chain improvement model based on multi-chain collaboration
…… ZHAO Xianxian, TAN Haibo, ZHAO He, ZHOU Tong, CHENG Haotian, LI Jinze (3283)

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2022.0813

掠海导弹触水滑跳运动特性数值模拟研究

雷娟棉[1,*]，吴志翔[1]，解文洋[2]

(1. 北京理工大学 宇航学院，北京 100081；2. 北京机电工程研究所，北京 100074)

摘 要：水面滑跳运动是一种飞行体在近水面运动时底部反复触水并被弹起的运动过程，该运动方式可以提高远程对海打击导弹在末端近水面无动力运动时的航程和机动能力。为研究导弹的水面滑跳运动特性，通过求解非定常 Navier-Stokes 方程和 Realizable k-ε 湍流模型，结合重叠网格技术，采用 VOF 方法捕捉空气和水之间的自由液面变化，通过六自由度方程模拟刚体运动，对典型导弹外形水面滑跳运动过程进行模拟研究，给出并分析了导弹一次水面滑跳过程的运动参数和表面压力变化。结果表明：根据水面滑跳过程中导弹姿态和相对水面位置的变化特点，可将导弹水面滑跳运动过程分为入水前空中飞行段、第 1 次触水弹起段、近水面滑行段、第 2 次触水滑行段、滑跳出水飞行段 5 个阶段；第 1 次触水过程中，船尾段受到水面冲击被迅速弹离水面，同时使弹体低头，导弹整体仍向下朝水面运动；第 2 次触水过程中，导弹下表面大面积拍击水面，不仅产生较大的水面冲击力，且产生使弹体抬头的力矩，在向上冲击力和抬头力矩作用下，导弹被抬升滑跳出水面。船尾收缩段产生的抬头力矩及导弹前端触水处产生的冲击载荷共同作用使得导弹完成水面滑跳过程。

关 键 词：跨介质飞行器；水面滑跳；多相流；刚体动力学；数值模拟
中图分类号：O359^{+}.1; V19
文献标志码：A **文章编号**：1001-5965(2024)10-2975-09

随着军事侦察技术和防御手段的增强，传统单一介质飞行器的突防能力越来越受到制约，在此背景下，跨介质飞行器开始受到了广泛关注。跨介质飞行器既能在空中巡航，又能在水下潜航，还可根据来袭拦截系统自主跨越水空界面躲避攻击，实现隐蔽突防。但是，由于水和空气介质的密度、黏度和可压缩性等物理性质的巨大差异，2 种介质下飞行器的设计存在很多矛盾。目前，该类飞行器仅停留在理论研究和概念样机设计阶段，还不具备实际工程应用的条件。

受到扁平石块水面弹跳这项娱乐运动的启发，有学者提出一类利用“打水漂”现象在水面滑跳的飞行器。“打水漂”这种运动方式已经应用在航天器返回和高超声速飞行器再入大气层的过程中，通过飞行器底部产生的激波使得飞行器在大气层内实现多次跳跃，达到高速突防和增程的效果。在海面，远程对海打击导弹在末端通过底部小角度冲击水面获得从水面弹起的升力，通过多次弹跳，不仅可以在低能耗状态下实现增程，还可以利用海面杂波实现隐蔽机动突防，同时也避免了多次出入水运动带来的动力和制导技术难题。

国外对“打水漂”现象的研究开始较早。2003 年，Bocquet[1] 对圆形和方形石块打水漂运动过程进行了研究，通过给出圆盘水漂运动过程中的动力学方程，解释了旋转对打水漂运动的重要性，并根据能量耗散规律给出了圆盘水漂运动弹跳所需最低速度和最大弹跳次数。Rosellini[2] 和 Clanet[3] 等对旋转圆盘在水面的单次弹跳和多次弹跳运动规律进

收稿日期：2022-09-29；**录用日期**：2022-12-16；**网络出版时间**：2023-02-08 10:43
网络出版地址：link.cnki.net/urlid/11.2625.V.20230208.0900.001
基金项目：国家部委基金 (6141B012871)
* **通信作者**. E-mail：leijm@bit.edu.cn

引用格式：雷娟棉，吴志翔，解文洋. 掠海导弹触水滑跳运动特性数值模拟研究 [J]. 北京航空航天大学学报，2024，50（10）：2975-2983.
LEI J M，WU Z X，XIE W Y. Numerical simulation investigation on water surface skipping motion characteristics of sea-skimming projectile [J]. Journal of Beijing University of Aeronautics and Astronautics，2024，50（10）：2975-2983 （in Chinese）.

行了实验研究，结果表明，旋转的主要作用是在冲击过程中通过陀螺效应稳定圆盘，圆盘的能量耗散取决于圆盘运动过程中受到的水动力。Nagahiro和Hayakawa[4]利用光滑粒子流体力学(smoothed particle hydrodynamics, SPH)方法按Rosellini等[2]的实验条件对旋转圆盘在水面的弹跳运动过程进行了数值模拟，得出了20°为旋转圆盘弹跳运动的最佳入水角度。侯东伯[5]对平板触水滑跳运动过程分别建立了入水力学模型和滑行力学模型，并进行了仿真计算与试验结果对比，结果表明，滑行力学模型更适合滑跳运动，同时复杂运动体底部平缓构型有利于滑跳运动。付晓琴等[6]对二维平板在波浪条件下的水漂运动过程进行了数值模拟，证明了俯仰角为20°可以实现最小投掷速度完成水漂运动，在波峰位置入水时平板能够获得最大水面接触面积，在波谷处入水能够获得最长的触水时间。解文洋等[7]利用计算流体力学(computational fluid dynamics, CFD)数值方法对三维圆盘水面滑跳过程进行了模拟，分析了圆盘运动过程中产生侧向偏移和转动的原因。Li等[8]基于大涡模拟(large eddy simulation, LES)数值方法对三维不同旋转速度下的圆盘打水漂运动过程进行了数值模拟，分析了圆盘撞击水面过程中产生的水域空腔非对称演变、水动力作用和侧向运动，指出了小攻角低转速和大攻角高转速时，空腔和流场更对称，圆盘壁面法向受力在圆盘侧向运动中起到重要作用。唐介等[9]对圆盘水漂过程建立了理论和实验模型，揭示了适当攻角和速度是产生足够水动力以满足弹跳条件的关键因素，以及圆盘偏转是由于陀螺效应和马格努斯效应共同作用的结果。孙士明等[10]设计了一种适宜滑跳的航行体构型，并进行了开放水面自由滑跳实验，同时对模型重心位置的变化规律进行了数值模拟计算，根据实验结果证明了将滑跳运动应用在飞行器近海面飞行的可行性，而且运动过程中模型整体应保持光滑过渡。田北晨等[11]设计了一种带滑板的跨介质飞行器，并对其近水面滑跳运动进行了数值模拟，研究了入水速度和俯仰角对滑跳运动的影响，初步确定了入水参数的合理范围。

另外，飞机水面迫降过程中也可能会发生类似“打水漂”的滑跳运动。一些学者[12-13]在研究飞机水上迫降时发现，底部具有大曲率变化的飞机会出现剧烈俯仰，可能会发生在水面的弹跳运动。其中，机体撞击水面时的剧烈冲击力会影响飞机迫降的平稳性和结构的完整性。Bensch等[14]对运输机在平静水面的迫降过程进行了数值模拟，快速估算出了飞机在水上迫降时受到的载荷，并根据计算结果分析了机翼和机身的结构变形。Streckwall等[15]利用Comet数值模型研究了NACA TN2929飞机模型中的3种机身外形在不同工况下的入水姿态变化规律。Delsart等[16]使用欧拉和拉格朗日耦合方法对直升机在水上迫降的过程进行了数值模拟。Qu等[17]利用整体动网格方法对NACA TN2929飞机的水上迫降能力进行了研究，发现在迫降过程中飞机不仅受到水动力，机翼和襟翼上还同时受到气动力的作用。吴宗成等[18]利用滑移动网格数值方法对NACA TN2929飞机在迎浪、顺浪和平行于波浪条件下的迫降过程进行了研究，并对比分析了3种条件下飞机的运动规律和受力情况。杨晓彬和许国冬[19]基于重叠网格技术和DFBI模型对NACA TN2929飞机的水上迫降过程进行了数值模拟研究，结果表明，飞机受到的载荷峰值与飞机入水速度的平方呈线性相关，随着飞机初始攻角的增加，受到的砰击载荷峰值会逐渐减小。实际上，飞机水面迫降过程中也会有滑跳的趋势，伴随有机体的剧烈俯仰运动和反复砰击水面，不利于人员和飞机安全；已有文献采用多种数值模型和网格方法对飞机水上迫降过程进行了相关研究，重点关注机体冲击水面时的水、气、固力学特性，目的是保证飞机在迫降过程中所受的冲击载荷不要过大，机体结构完整，降落过程平稳过渡至水面漂浮。飞机水面迫降与水面滑跳过程在水动力作用机理和运动方式有相似之处，对于设计飞行器水面滑跳有指导意义。

目前，对水面滑跳运动过程的实验和数值模拟研究，主要集中于对旋转圆盘这一简单外形的运动学和动力学研究，且圆盘本身由于高速自旋的陀螺效应而较为稳定；针对复杂外形近水面滑跳过程的现象研究较少，飞机水面迫降虽涉及水面冲击、滑翔和漂浮等问题，但因关注的是飞机迫降过程中的安全问题，研究的侧重点是避免飞机触水后再次弹起离开水面。

本文以远程对海攻击导弹为对象，其末端掠海突防过程中导弹在近水面的滑跳运动过程为应用背景，采用基于雷诺平均Navier-Stokes方程的多相流非定常数值方法对典型导弹外形近水面滑跳运动过程进行模拟，研究一次完整触水滑跳过程及其受力等，分析了导弹船尾段的作用。本文研究工作对设计和发展水面滑跳飞行器具有重要参考意义。

1 数值计算方法

本文主要研究低速条件下导弹在水面的滑跳运动过程，计算时，水和空气均认为是不可压缩流体且运动过程无热量交换，考虑重力加速度，忽略

表面张力。数值计算流动控制方程为非定常雷诺平均 Navier-Stokes 方程, 湍流模型为采用标准壁面函数的 Realizable k-ε 模型, 多相流场采用 VOF 多相流模型, 弹体运动过程采用六自由度运动方程进行求解。压力场和速度场采用 Coupled 耦合算法, 时间离散采用隐式格式, 空间离散均采用二阶迎风格式。

1.1　流体控制方程

本文采用 VOF 多相流模型捕捉水气界面, 该模型利用网格单元不同相的体积分数ε_i加权平均描述混合介质, 通过求解连续性方程和动量方程来捕捉交界面相分布和形状。对于无相间质量输运的多相流动, 混合介质的连续性方程和动量方程分别为

$$\frac{\partial \rho_{\mathrm{m}}}{\partial t}+\frac{\partial\left(\rho_{\mathrm{m}} u_i\right)}{\partial x_i}=0 \tag{1}$$

$$\frac{\partial\left(\rho_{\mathrm{m}} u_i\right)}{\partial t}+\frac{\partial\left(\rho_{\mathrm{m}} u_i u_j\right)}{\partial x_j}=-\frac{\partial p}{\partial x_i}+\frac{\partial}{\partial x_i}\left[\left(\mu_{\mathrm{m}}+\mu_{\mathrm{t}}\right)\left(\frac{\partial u_i}{\partial x_j}+\frac{\partial u_j}{\partial x_i}\right)\right] \tag{2}$$

式中: i、j 为坐标方向; u_i、u_j为速度分量; ρ_{m}和μ_{m}分别为混合介质的密度和动力黏性系数; p 为流场压力; μ_{t}为湍流黏性系数。

$$\rho_{\mathrm{m}}=\sum_{i=1}^{n}\varepsilon_i\rho_i \tag{3}$$

$$\mu_{\mathrm{m}}=\sum_{i=1}^{n}\varepsilon_i\mu_i \tag{4}$$

$$\mu_{\mathrm{t}}=\rho_{\mathrm{m}}C_\mu k^2/\varepsilon \tag{5}$$

式中: C_μ为经验常数; k为湍流动能; ε为湍流动能耗散率。

1.2　动力学方程

本文对弹体水面滑跳运动过程模拟时, 弹体的六自由度运动分别通过在固定的惯性坐标系和随物体运动的体坐标系求解平动和转动动力学方程得到。平动和转动方程分别为

$$m\dot{\boldsymbol{V}}_{\mathrm{G}}=\sum\boldsymbol{F}_{\mathrm{G}} \tag{6}$$

$$\dot{\boldsymbol{\omega}}=\boldsymbol{I}^{-1}\left(\sum\boldsymbol{M}_{\mathrm{B}}-\boldsymbol{\omega}_{\mathrm{B}}\times\boldsymbol{I}\boldsymbol{\omega}_{\mathrm{B}}\right) \tag{7}$$

式中: m 为刚体质量; $\dot{\boldsymbol{V}}_{\mathrm{G}}$为刚体的平动加速度; $\boldsymbol{F}_{\mathrm{G}}$为刚体所受合力; $\boldsymbol{\omega}_{\mathrm{B}}$为刚体转动速度; $\dot{\boldsymbol{\omega}}$为角加速度矢量; $\boldsymbol{I}$ 为惯性张量; $\boldsymbol{M}_{\mathrm{B}}$为刚体所受力矩。

力矩由惯性坐标系转换到体坐标系方程为

$$\boldsymbol{M}_{\mathrm{B}}=\boldsymbol{R}_{\mathrm{r}}\boldsymbol{M}_{\mathrm{G}} \tag{8}$$

式中: $\boldsymbol{M}_{\mathrm{G}}$为刚体在惯性坐标系下所受力矩; $\boldsymbol{R}_{\mathrm{r}}$为转动矩阵, 表示为

$$\boldsymbol{R}_x=\begin{bmatrix}1 & 0 & 0\\ 0 & \cos\phi & \sin\phi\\ 0 & -\sin\phi & \cos\phi\end{bmatrix} \tag{9}$$

$$\boldsymbol{R}_y=\begin{bmatrix}\cos\theta & 0 & -\sin\theta\\ 0 & 1 & 0\\ \sin\theta & 0 & \cos\phi\end{bmatrix} \tag{10}$$

$$\boldsymbol{R}_z=\begin{bmatrix}\cos\psi & \sin\psi & 0\\ -\sin\psi & \cos\psi & 0\\ 0 & 0 & 1\end{bmatrix} \tag{11}$$

$$\boldsymbol{R}_{\mathrm{r}}=\boldsymbol{R}_x\boldsymbol{R}_y\boldsymbol{R}_z \tag{12}$$

其中: φ、θ、ψ 为刚体运动的 3 个欧拉角, 即滚转角、俯仰角和偏航角。

根据式 (6)、式 (7) 计算出平移加速度和角加速度后, 通过数值积分得出速度, 并使用角速度和平移速度给出弹体的位置和姿态参数, 同时对计算和网格进行更新。

2　数值计算外形及条件

飞行器在水面的滑跳运动过程是一个涉及气、液、固多相耦合的非定常过程, 但本文将导弹视为刚体, 仅考虑空气和水气液两相的流动。因远程对海打击导弹在弹道末端以滑跳运动方式飞行时, 主要是弹体表面与海面相互作用产生相应的弹跳升力, 且滑跳运动时速度较低(<30 m/s), 舵面较小, 受到的水、气动力可忽略, 因此, 将模型简化为不带翼舵的导弹外形进行水面滑跳运动过程研究。

本文导弹水面滑跳运动, 数值模拟基本外形参考美国标准 2.Block 反舰导弹外形, 其主要由弹头、圆柱段和船尾 3 部分组成, 导弹几何外形参数和初始计算条件如图 1 所示。

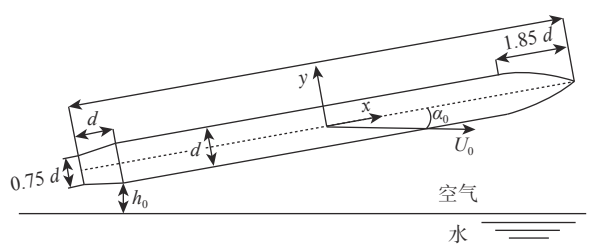


图 1　导弹几何外形参数及计算条件示意图

Fig. 1　Schematic diagram of geometry parameters and calculation conditions of projectile

导弹主要几何外形参数为: 直径 d=95 mm, 长度 L=1 219 mm, 质量 M=1.33 kg, 导弹绕弹体 x 轴的转动惯量 I_x=0.005 6 kg·m^2, 绕 y 轴与 z 轴的转动惯量 I_y=I_z=0.14 kg·m^2。导弹在水面滑跳运动过程中, 可根据导弹最低点与水面距离 h 判断导弹相对水

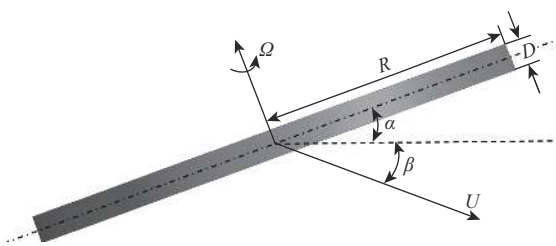
面的位置。当 h>0 时，导弹未入水；当 h=0 时，导弹刚触水；当 h<0 时，导弹有一部分没入水中。进行导弹水面滑跳运动过程数值模拟计算的条件为：初始攻角 α_0=10°，初始速度沿水平方向 U_0=25 m/s，初始时刻导弹最低点距离水面高度 h_0=0.05 m。

3 数值计算方法及网格验证

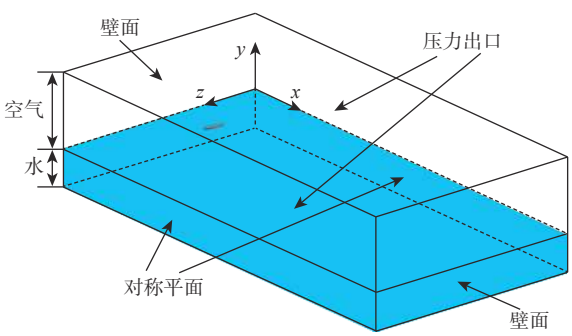
3.1 数值计算方法验证

为验证数值计算方法的可靠性，先按文献 [2] 中的实验条件对圆盘打水漂过程进行数值模拟。图 2(a) 给出了圆盘外形及运动参数示意图，实验中采用铝制圆盘，半径 R=25 mm，厚度 D=2.75 mm，圆盘质量 M_d=0.014 kg。图 2(b) 给出了数值模拟三维计算域及边界条件示意图，其中，x 轴方向沿圆盘初始水平运动方向，y 轴沿竖直方向向上。计算域 x 方向长度为 1.2 m，y 方向高度为 0.3 m，其中，水域高度为 0.1 m，z 方向宽度为 0.6 m。竖直 y 方向顶部和底部为压力出口，顶部空气域压力值与大气压相同，底部水域压力值对应同深度下静水压力；两侧为对称边界，水平方向前后两端边界条件为壁面。圆盘表面网格尺寸为 0.6 mm，总网格量约为 400 万，时间步长取 0.1 ms。

计算条件为：圆盘入水速度 U=3.5 m/s，圆盘倾角 α_d=20°，圆盘速度矢量与水平面夹角 β=20°，旋转转速 Ω=65 r/s。由本文计算条件得到圆盘触水时间 τ=32 ms，文献 [2] 中实验结果为 31 ms。图 3 给出了本文数值模拟计算得到的圆盘水平速度随时间变化曲线，并与文献 [8] 中采用 LES 数值方法得到的结果进行比对。可以看出，本文数值模拟结果与实验结果和 LES 模拟结果有较好的一致性，证明了数值方法可靠性，后续计算仍采用此表面网格尺寸。



(a) 圆盘外形和运动参数示意图



(b) 计算域示意图

图 2 圆盘运动参数和计算域示意图

Fig. 2 Schematic diagram of disk motion parameters and calculation domain

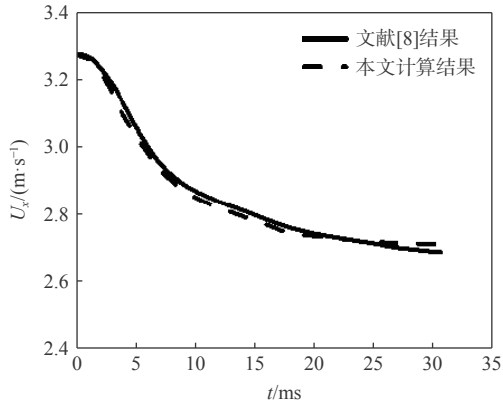


图 3 圆盘水平速度随时间变化

Fig. 3 Variation of horizontal velocity of disk with time

3.2 数值模拟计算网格及时间步长验证

图 4 给出了进行导弹水面滑跳运动过程数值模拟时的三维计算域及边界条件示意图。计算域沿水平方向 (x 轴方向) 的长度为 20 m，沿竖直方向 (y 轴方向) 的高度为 3 m(其中，水域高度为 1 m)，沿侧向 (z 轴方向) 的宽度为 1.2 m。在竖直方向，计算域的顶部和底部为压力出口，顶部空气域压力值与大气压相同，底部水域压力值对应同深度下静水压力；两侧为对称边界，水平方向前后两端边界为壁面。图 5 为导弹周围网格示意图。导弹表面附近网格尺度与 3.1 节圆盘表面网格尺度相同，同时对水气交界面处网格适当加密以提高自由液面处计算精度，整体网格量约为 500 万，其中，导弹周围重叠网格量约为 50 万。

选取 Δt =2,1,0.5 ms 这 3 个时间步长进行时间步无关性验证。图 6 给出了分别取 3 种不同时间步长时通过数值模拟计算得到的导弹质心高度 H 和攻角 α 随时间的变化曲线。可以看出，时间步

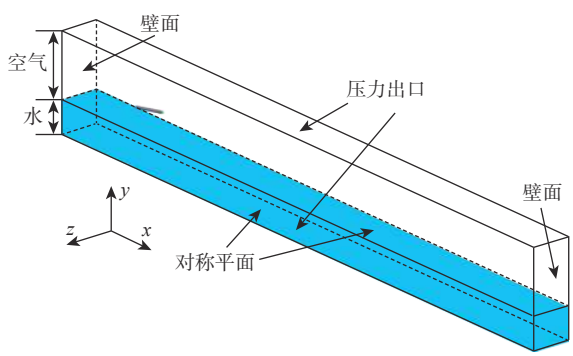


图 4 计算域边界条件示意图

Fig. 4 Schematic diagram of boundary conditions of calculation domain

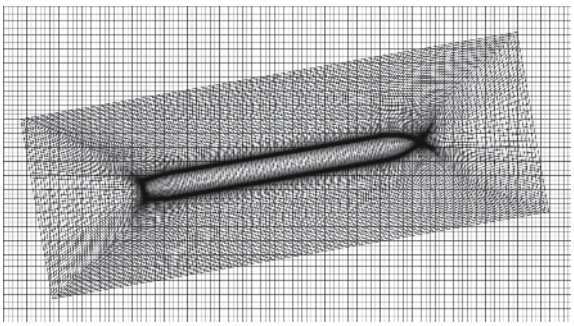


图 5　导弹周围网格示意图

Fig. 5　Schematic diagram of grid around projectile

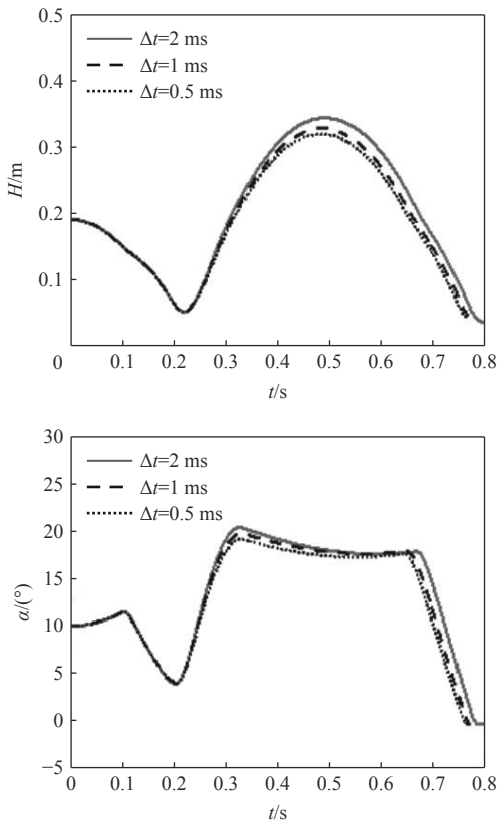


图 6　导弹质心距水面高度和攻角随时间变化

Fig. 6　Variation of centroid height above water surface and angle of attack of projectile with time

长取 0.5 ms 和 1 ms 的计算结果比较接近, 而时间步长取 2 ms 时的结果相比于另外 2 个时间步长的计算结果差异较大。综合考虑计算结果的准确性和计算成本, 后续导弹滑跳运动过程数值模拟计算时取时间步长为 Δt=1 ms。

4　导弹滑跳运动特性数值模拟

对图 1 所示的美国标准 2.Block 反舰导弹在水面的滑跳运动过程进行数值模拟, 得到了导弹在水面滑跳运动过程中的轨迹、姿态变化规律及与水的互作用力, 下面给出数值模拟结果并进行运动特性分析。

图 7 给出了通过非定常数值模拟计算得到的一次滑跳过程中不同时刻导弹相对水面的位置。图 8 给出了导弹在水面滑跳运动过程中最低点与水面距离 h 和垂向速度 U_y 随时间的变化曲线。可以看出, 导弹在水面的一次滑跳运动过程中出现了二次触水的现象, 第 1 次触水时入水较浅, 很快跃起离开水面; 跃起后基本紧贴水面向前运动, 在贴水面向前运动很短距离后再第 2 次触水, 第 2 次触水后弹体入水较深, 在触水滑行一段距离后再第 2 次跃起离开水面。

图 9 给出了一次滑跳运动过程中导弹攻角 α 和俯仰角速度 ω 随时间的变化曲线。图 10 给出了一次滑跳运动过程中导弹垂直方向和水平方向所受外力 F_y 和 F_x (不包括重力) 随时间的变化曲线。

根据图 7～图 10 中导弹在水面的一次滑跳运动过程中与水面的相对位置、导弹的姿态变化及受力特点, 可将一次近水面滑跳运动过程分为 5 个阶段: ①入水前空中飞行段, 0～0.1 s; ②第 1 次触水弹起段, 0.1～0.13 s; ③近水面滑行段, 0.13～0.175 s; ④第 2 次触水滑行段, 0.175～0.34 s; ⑤滑跳出水飞行段, 0.34 s 以后。下面分别对滑跳运动过程中各

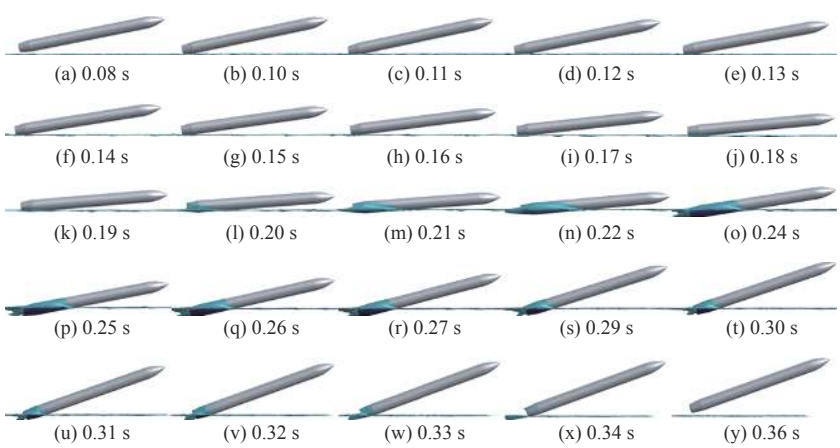


图 7　一次滑跳过程中导弹相对水面位置变化

Fig. 7　Position of projectile relative to water surface during a skipping motion

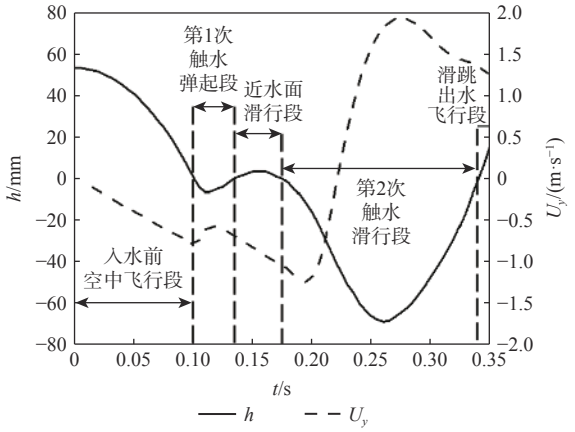


图 8 导弹最低点距水面高度和垂向速度随时间变化

Fig. 8 Variation of height from the lowest point of projectile to water surface and vertical velocity with time

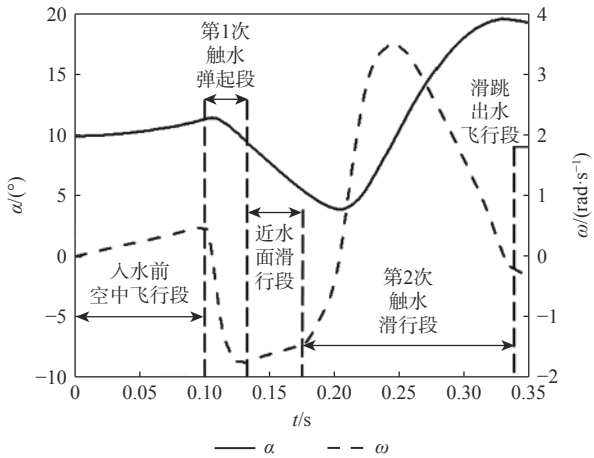


图 9 攻角和俯仰角速度随时间变化

Fig. 9 Variation of angle of attack and pitch angular velocity with time

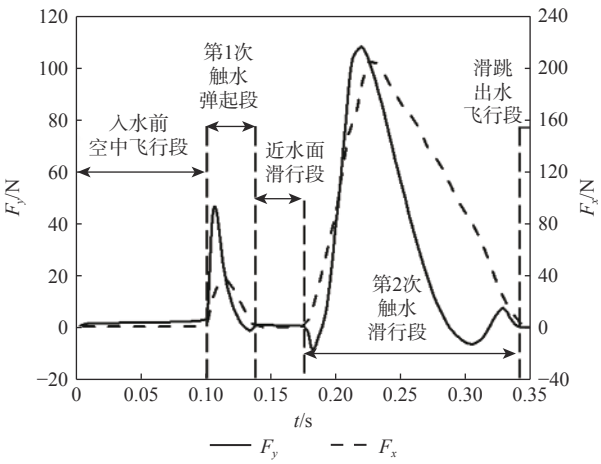


图 10 导弹垂直和水平方向受力随时间变化

Fig. 10 Variation of force on projectile in vertical and horizontal directions with time

阶段的运动及受力特点进行分析。

1）入水前空中飞行段。由图 7～图 10 可见，在 0～0.1 s 的入水前空中飞行段，导弹的垂向速度增大，与水面之间的距离减小，导弹攻角和俯仰角速度增大，除重力之外所受的水平方向和垂直方向的外力 F_x 和 F_y 非常小。

这是因为：在入水前空中飞行段，导弹所受的外力主要是气动力，而此时导弹速度较低，也没有翼面提供升力，所受气动力很小。导弹在重力作用下，高度降低，与水面之间的距离减小，沿负 y 轴的垂向速度分量增大；由于导弹没有气动稳定面，是静不稳定的，在入水前的空中飞行过程，不稳定气动力矩作用下导弹攻角和俯仰角速度一直在增大。

2）第 1 次触水弹起段。在 0.1～0.13 s 的运动过程中，导弹船尾部先接触水面，导弹在受到水面冲击后被迅速弹离水面且低头，但导弹整体仍向下朝水面运动。在此过程中，弹体尾部触水受到的冲击力短时间内急剧增大，增大到最大值后迅速减小，导弹垂向速度快速减小一定值后又逐渐增大；俯仰角速度从空中飞行时的正值迅速减小变为负值且一直在增大，使得该过程中导弹具有一定的低头角速度，攻角迅速减小，导弹尾部被抬起脱离水面。

图 11 给出了第 1 次触水弹起段导弹底部压力云图。导弹船尾在第 1 次接触水面时，短时间内受到较为剧烈的冲击载荷作用而出现局部高压区，并且因为船尾距离质心较远，产生了一个很大的低头力矩，所以导弹急剧低头；而后迅速弹起，与水接触面积减小，竖直方向受力也急剧减小。

因为此阶段导弹触水面积小，触水时间较短，水面对导弹冲击力的累积作用并未完全改变导弹运动方向，所以导弹整体向下运动速度受冲击影响虽有小幅度减小，但仍保持向下朝水面运动。

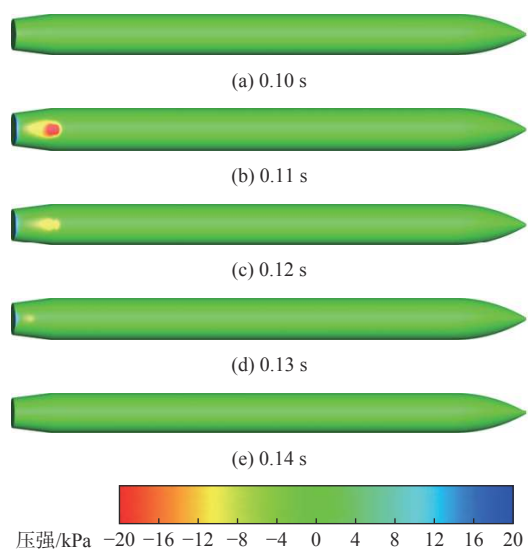


图 11 第 1 次触水弹起段导弹底部压力云图

Fig. 11 Pressure contour at bottom of projectile under bounce stage of the first contact with water

3）近水面滑行段。导弹在第 1 次触水弹离水面后的 0.13～0.175 s 间进入近水面滑行段，在该阶段，导弹与水面之间的最小距离变化很小，导弹与水面保持很小的距离向前滑行。由于上一阶段的船尾段触水受到剧烈冲击，使导弹具有一定的低头角速度，其攻角还在不断减小；相较于第 1 阶段的空中飞行过程，由于此时导弹攻角更小，所受到的气动力也更小，在重力的作用下导弹质心高度不断降低。

4）第 2 次触水滑行段。在近水面滑行一小段距离后，导弹受重力作用再次靠近水面，在 0.175～0.34 s 间为第 2 次触水滑行段。与第 1 次触水不同，在 0.175 s，导弹第 2 次触水后，船尾段受到冲击并未马上弹离水面，导弹先是有较大面积沉入水中，而后逐渐弹离水面，沿竖直向上的速度不断增加；导弹竖直方向受力经历由负到正再到负的变化过程，在 0.22 s 时垂向作用力达到最大，并且这一阶段竖直力和水平力作用峰值和作用时间均比第 1 次触水时大；在这一阶段，导弹低头角速度先逐渐减小至零，在 0.205 s 导弹攻角达到最小值，之后弹体开始抬头，抬头角速度先增大再减小。下面对这一阶段受力和导弹姿态变化特点进行分析。

在第 2 次触水初期，由图 10 可知，导弹此时受到竖直向下的力，与第 1 次触水时受到竖直向上的力完全相反。图 12 给出了第 2 次触水滑行段导弹底部压力云图。由图 12 观察到这段时间导弹底部触水区域出现了面积很小的低压区和高压区，这可以解释为导弹第 1 次触水撞击时攻角较大，船尾段迎面撞击水面受到冲击载荷，整个导弹受到竖直向上的力；而第 2 次触水时攻角较小，导致导弹直接冲击水面驻点区域前移至圆柱段与船尾段交界处，此处水流流速最低，压力最大。而船尾处存在收缩，前端驻点处水流受冲击后经过收缩处时流速加快，根据伯努利原理，此处的压力较低。由于触水初期，导弹水平速度较大，船尾段流速与作用面积更大，且导弹接触水面面积较小，高压区作用面积较小，低压区产生的吸力占主导作用，导弹的垂向力表现为负值。

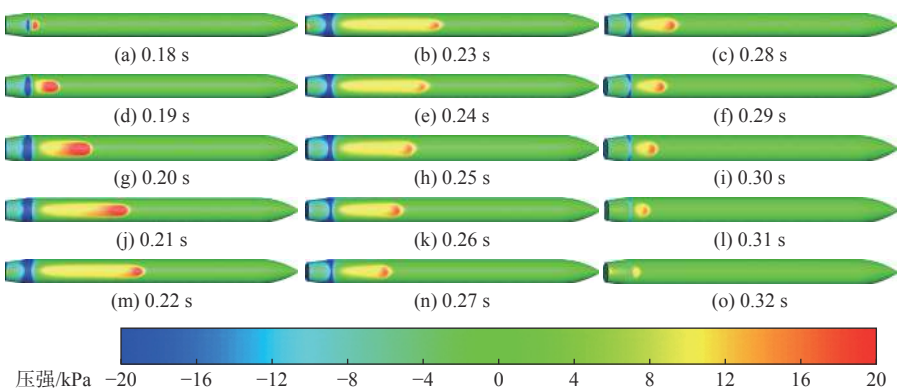


图 12　第 2 次触水滑行段导弹底部压力云图

Fig. 12　Pressure contour at bottom of projectile under bounce stage of the second contact with water

随着导弹继续浸入水面，浸水面积增大，触水高压区前移且作用面积增大，冲击力开始占主导作用，所受垂向力转为正值并开始不断增大，直至与重力相等，导弹的纵向速度达到负极值，垂直向下速度此后开始减小。而在触水后期，可以观察到图 10 中垂向力改变了方向，根据图 12 可知，在这个时间段，高压区作用面积和幅值均在减小，导弹所受的抬升力已经非常小，此时船尾所受的吸力占主导。之后，导弹的船尾部分开始离开水面，吸力逐渐消失，抬升力又重新主导，垂向力变为正值。

根据图 12 导弹底部高压低压区域作用变化特点，在第 2 次触水前期（0.175～0.24 s），随着导弹质心高度不断下降，船尾的低压区产生的吸力和触水驻点高压区产生的冲击力均随着导弹浸入水中的深度增加而增大，但由于高压区在不断向质心前移，低压区相对于高压区离质心更远，综合来看，作用在导弹上的力产生了抬头力矩，低头角速度逐渐减小并转为正的抬头角速度，导弹有抬头的趋势。

在 0.24 s 之后，导弹抬头角速度开始减小，从图 9 和图 11 看出，这是由于此时竖直方向力开始减小，并且高压区逐渐后移产生低头力矩，与船尾底部的低压区吸力产生的抬头力矩相互抵消，使得导弹受到的抬头力矩减小。

这一阶段中，导弹撞击水面产生的冲击力更为剧烈，弹体表面高压区既要起抬升导弹脱离水面的作用，又要与船尾低压区相互作用产生抬头力矩，控制出水时的攻角。相较于上一次船尾触水，导弹整体与水面接触面积更大，表面高压区与低压区压力幅值与位置变化也更剧烈，共同作用下导致导弹在这一阶段的运动和姿态参数（如垂向速度和攻

角),变化更为剧烈,也决定了导弹是否能够完成滑跳。

5)滑跳出水段。在0.34 s导弹第2次跳出水面,之后以约20°攻角在水面滑行,完成了一次完整的滑跳运动。

综上所述,导弹在水面一次滑跳过程中先后经历2次触水弹起过程,受到水动力载荷作用,运动趋势从向着水面运动到最后滑跳弹离水面。根据这2次触水阶段的受力和姿态变化特点,导弹船尾在2次触水过程中都起到了关键性的作用:第1次触水过程中,导弹只有船尾受到水面冲击作用,被从水中向上弹起,并使导弹有小幅度低头,避免了尾部抬起角度过高使攻角变为负值,头部直接扎进水中;导弹第2次触水时,船尾部分与水面相互作用形成的低压区产生吸力,一方面使导弹质心的高度进一步降低,另一方面使其产生抬头力矩,保证了导弹能够弹出水面,完成滑跳运动。根据低压区表现出的特点可以得出,低压区产生的吸力大小与导弹收缩段浸入水中的面积和导弹的水平速度呈正相关。

5 结 论

本文基于非定常CFD数值模拟方法,结合重叠网格技术和六自由度模型,对导弹一次水面滑跳全过程进行了研究,给出了导弹一次触水滑跳过程中运动参数随时间的变化,并结合水动力载荷作用区域及幅值特点,分析了导弹触水和滑跳出水过程中的运动特性,指出了导弹船尾在整个滑跳过程中的作用。根据水面滑跳过程中导弹姿态和相对水面位置的变化特点,将导弹水面滑跳运动过程分成5个阶段:入水前空中飞行段、第1次触水弹起段、近水面滑行段、第2次触水滑行段和滑跳出水飞行段。

1)导弹先在空中飞行,受重力作用竖直向下速度逐渐增加,攻角有小幅度增加,并逐渐靠近水面。

2)第1次触水撞击,导弹入水很浅,船尾段受到剧烈冲击力很快弹离水面,攻角受冲击力产生的低头力矩开始减小。

3)弹离水面后,导弹与水面保持很小的距离,整体保持滑行姿态并受重力作用继续靠近水面。

4)第2次触水撞击时,导弹攻角更小,且撞击前的垂向速度值更大,因此,导弹的沾湿面积和浸水体积更大,冲击载荷峰值更大,持续作用时间也更长,水流经船尾收缩段时加速从而形成了低压区,其在第2次触水初期和后期占主导作用,船尾段吸力作用产生的抬头力矩及导弹前端触水驻点处较大的冲击载荷产生的抬升作用,使得导弹得以离开水面,完成一次水面滑跳过程。

通过研究发现,外部构型对导弹触水后的运动和受力特性有显著影响,而本文模型为旋成体弹身,今后还需要开展复杂构型运动体及真实海浪条件下触水滑跳的数值研究和试验验证。

参考文献(References)

[1] BOCQUET L. The physics of stone skipping[J]. American Journal of Physics, 2003, 71(2): 150-155.

[2] ROSELLINI L, HERSEN F, CLANET C, et al. Skipping stones[J]. Journal of Fluid Mechanics, 2005, 543: 137.

[3] CLANET C, HERSEN F, BOCQUET L. Secrets of successful stone-skipping[J]. Nature, 2004, 427: 29.

[4] NAGAHIRO S I, HAYAKAWA Y. Theoretical and numerical approach to "magic angle" of stone skipping[J]. Physical Review Letters, 2005, 94(17): 174501.

[5] 侯东伯. 运动体触水滑跳过程运动特性研究[D]. 哈尔滨: 哈尔滨工业大学, 2020.
HOU D B. Research on kinematic characteristics of moving object during water-skipping[D]. Harbin: Harbin Institute of Technology, 2020(in Chinese).

[6] 付晓琴, 李阳辉, 卢昱锦, 等. 二维平板水漂运动数值模拟[J]. 航空学报, 2021, 42(6): 124351.
FU X Q, LI Y H, LU Y J, et al. Numerical simulation of two-dimensional plate skipping[J]. Acta Aeronautica et Astronautica Sinica, 2021, 42(6): 124351(in Chinese).

[7] 雷娟棉, 解文洋, 于勇, 等. 圆盘水漂运动特性数值模拟研究[J]. 北京理工大学学报, 2021, 41(7): 687-695.
LEI J M, XIE W Y, YU Y, et al. Study on numerical simulation of skipping stone movement characteristics[J]. Transactions of Beijing Institute of Technology, 2021, 41(7): 687-695(in Chinese).

[8] LI C H, WANG C, WEI Y J, et al. Numerical investigation on the cavity dynamics and deviation characteristics of skipping stones[J]. Journal of Fluids and Structures, 2021, 104: 103301.

[9] TANG J, ZHAO K, CHEN H T, et al. Trajectory and attitude study of a skipping stone[J]. Physics of Fluids, 2021, 33(4): 043316.

[10] 孙士明, 陈玮琪, 王宝寿, 等. 航行体近水面滑跳运动试验研究[J]. 数字海洋与水下攻防, 2019, 2(2): 79-83.
SUN S M, CHEN W Q, WANG B S, et al. Experimental research on near-water-surface skipping motion of vehicle[J]. Digital Ocean & Underwater Warfare, 2019, 2(2): 79-83(in Chinese).

[11] 田北晨, 刘涛涛, 吴钦, 等. 跨介质飞行器触水滑跳运动特性数值模拟[J]. 兵工学报, 2022, 43(3): 586-598.
TIAN B C, LIU T T, WU Q, et al. Numerical simulation on kinematic characteristics of trans-media aircraft during water-skipping [J]. Acta Armamentarii, 2022, 43(3): 586-598(in Chinese).

[12] 郭保东, 屈秋林, 刘沛清, 等. 混合翼身布局客机SAX-40水上迫降力学性能数值研究[J]. 航空学报, 2013, 34(11): 2443-2451.
GUO B D, QU Q L, LIU P Q, et al. Ditching performance of silent aircraft SAX-40 in hybrid wing-body configuration[J]. Acta Aeronautica et Astronautica Sinica, 2013, 34(11): 2443-2451(in Chinese).

[13] 赵芸可, 屈秋林, 刘沛清. 水上飞机水面降落全过程力学特性数值研究[J]. 北京航空航天大学学报, 2020,46(4):830-838.
ZHAO Y K, QU Q L, LIU P Q. Numerical study on mechanical properties of seaplane in whole water surface landing process[J]. Journal of Beijing University of Aeronautics and Astronautics, 2020, 46(4): 830-838(in Chinese).

[14] BENSCH L, SHIGUNOV V, BEUCK G, et al. Planned ditching simulation of a transport airplane[C]//KRASH Users' Seminar. Phoenix: [s. n.], 2001.

[15] STRECKWALL H, LINDENAU O, BENSCH L. Aircraft ditching: A free surface/free motion problem[J]. Archives of Civil and Mechanical Engineering, 2007, 7(3): 177-190.

[16] DELSART D, LANGRAND B, VAGNOT A. Evaluation of a Euler/Lagrange coupling method for the ditching simulation of helicopter structures[C]//Proceedings of the 5th International Conference on Fluid Structure Interaction. Southampton: WIT Press, 2009, 105: 259-268.

[17] QU Q L, HU M X, GUO H, et al. Study of ditching characteristics of transport aircraft by global moving mesh method[J]. Journal of Aircraft, 2015, 52(5): 1550-1558.

[18] 吴宗成, 黄波恩, 吴亚聪. 滑移动网格在波浪水面迫降数值模拟中的应用[J]. 哈尔滨工业大学学报, 2019, 51(1): 80-86.
WU Z C, HUANG B E, WU Y C. Application of sliding dynamic grid to wavy water ditching simulation[J]. Journal of Harbin Institute of Technology, 2019, 51(1): 80-86(in Chinese).

[19] 杨晓彬, 许国冬. 基于重叠网格法的飞机水上降落水动力砰击载荷研究[J]. 振动与冲击, 2020, 39(2): 57-63.
YANG X B, XU G D. Identification of hydrodynamic impact loads during the airplane ditching based on overset grid method[J]. Journal of Vibration and Shock, 2020, 39(2): 57-63(in Chinese).

Numerical simulation investigation on water surface skipping motion characteristics of sea-skimming projectile

LEI Juanmian[1,*], WU Zhixiang[1], XIE Wenyang[2]

(1. School of Aerospace Engineering, Beijing Institute of Technology, Beijing 100081, China;
2. Beijing Electro-Mechanical Engineering Institute, Beijing 100074, China)

Abstract: Water surface skipping is a kind of motion in which the bottom of the aircraft touches the water surface and bounces up repeatedly when it moves near the water surface. This motion can improve the unpowered range and maneuverability of the long-range projectile near the water surface. This paper uses the VOF method to capture the change in the water surface, the overset grid, and the equation of six degrees of freedom to simulate the rigid body motion in order to study the characteristics of projectile water surface skipping. This is done by solving the unsteady Navier-Stokes equation and the Realizable k-ε model. According to the change of projectile attitude and relative water surface position, the water surface skipping motion process of the projectile is divided into five stages. The process of the projectile sliding on the water surface is simulated. The findings indicate that there are five stages in the process of projectile water surface sliding: the air flight prior to contact with the water, the first water contact and bounce, the near-water surface gliding, the second water contact and sliding out of the water. These stages are based on changes in the projectile's attitude and relative position to the water surface. The first time, the stern of the projectile touches the water and quickly bounces up from the water surface by the impact of the water surface, making the projectile pitch down. In this process, the projectile still moves toward the water surface. During the second water contact, the lower surface of the projectile slaps the water surface in a large area. The projectile is not only subjected to a large impact force on the water surface but also subjected to a nose-up moment. Under the lift impact force and the nose-up moment, the projectile glides out of the water surface. The combined effect of the lift torque generated by the stern of the projectile and the impact load generated at the water contact point of the projectile enables the projectile to complete the water surface skipping process.

Keywords: trans-media aircraft; water surface skipping; multiphase flow; rigid body dynamics; numerical simulation

Received: 2022-09-29; **Accepted**: 2022-12-16; **Published Online**: 2023-02-08 10:43
URL: link.cnki.net/urlid/11.2625.V.20230208.0900.001
Foundation item: National Ministries and Commissions Foundation of China (6141B012871)
* **Corresponding author**. E-mail: leijm@bit.edu.cn

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2022.0775

融合双注意力机制和动态图卷积的点云语义分割

杨军[1, 2, *]，张琛[1]

(1. 兰州交通大学 电子与信息工程学院，兰州 730070；2. 兰州交通大学 测绘与地理信息学院，兰州 730070)

摘 要：针对现有基于深度学习的三维点云语义分割算法通常在提取局部特征时忽略邻域点间深层次语义信息，聚合局部邻域特征时忽略其他邻域特征中有用信息的问题，提出融合双注意力机制和动态图卷积神经网络(DGCNN)的三维点云语义分割算法。通过动态图卷积操作构造边缘特征，并将中心点与邻域点的相对距离输入到核点卷积操作得到增强后的边缘特征，进一步加强中心点与邻域点之间的联系；引入空间注意力模块以建立邻域点之间的依赖关系，将特征相似点相互关联，从而在局部邻域内提取到更深层次的上下文信息，丰富邻域点的几何特征；在聚合局部邻域特征时引入通道注意力模块，通过给不同通道赋予不同的权值以达到增强有用通道同时抑制无用通道的目的，从而提高语义分割的准确率。在S3DIS数据集和SemanticKITTI数据集上的实验结果表明：所提算法的语义分割精度分别达到了66.0%和59.4%，与其他经典的网络模型相比，取得了更好的点云分割效果。

关 键 词：三维点云；语义分割；深度学习；注意力机制；动态图卷积
中图分类号：TP391
文献标志码：A **文章编号**：1001-5965(2024)10-2984-11

近年来，随着深度传感器技术(如激光雷达设备、RGB-D相机)的不断发展，三维点云被广泛应用于室内导航[1]、机器人[2]、自动驾驶[3]、三维重建[4]等领域。作为计算机视觉长期研究的课题之一，语义分割旨在利用计算机对场景进行逐点分类，并将场景分割成若干具有特定语义类别的区域，是三维场景理解和分析的基础[5]。然而，由于点云数据的不规则、无序性和稀疏性等特点，使自动、精确的点云语义分割成为一项具有挑战性的任务，设计针对大规模、复杂场景三维点云的语义分割网络模型成为当前的研究重点。早期学者们借鉴二维图像处理的方法，将三维点云数据转化为二维数据，进而使用二维图像处理中较为成熟的卷积神经网络(convolutional neural network, CNN)、全连接神经网络(fully connected neural network, FCN)算法，则基于投影和基于体素化的方法应运而生。基于投影的方法先获取三维点云形状在不同视角下的二维图像，再对每个视图进行卷积处理，通过池化层和全连接层将每个视角得到的特征进行聚合，得到最终的语义分割结果，但投影过程不可避免地会产生信息损失，导致分割结果并不理想。基于体素化的方法将点云规则化为网格结构，很大程度上保留了物体的几何信息，但该结构仍然无法细分物体边界的几何信息，并且存在大量无效、冗余的信息，大大增加了计算量，空间复杂度高。PointNet[6]提出直接对点云数据进行处理，但其注重独立提取点云全局特征而忽略了点之间的几何关联，从而丢失了许多局部特征。为解决该问题，文献[7]在动态图卷积神经网络(dynamic graph convolutional neural network, DGCNN)中提出了边缘卷积，但在实验过程中发

收稿日期：2022-09-14；**录用日期**：2023-01-02；**网络出版时间**：2023-01-10 13:51
网络出版地址：link.cnki.net/urlid/11.2625.V.20230109.1713.007
基金项目：国家自然科学基金(42261067,61862039)；兰州市人才创新创业项目(2020-RC-22)；兰州交通大学天佑创新团队(TY202002)
* **通信作者**. E-mail：yangj@mail.lzjtu.cn

引用格式：杨军，张琛. 融合双注意力机制和动态图卷积的点云语义分割[J]. 北京航空航天大学学报，2024，50(10)：2984-2994.
YANG J，ZHANG C. Semantic segmentation of point clouds by fusing dual attention mechanism and dynamic graph convolution[J]. Journal of Beijing University of Aeronautics and Astronautics，2024，50(10)：2984-2994 (in Chinese).

现, 该方法仍会丢失局部深层次语义信息。

针对上述问题, 本文提出一种融合双注意力机制和 DGCNN 的架构。主要创新点如下:

1) 构建空间注意力模块, 用来建立局部邻域内的邻域点间的依赖关系, 提取更深层次的语义信息。

2) 构建通道注意力模块, 用来增强有用的通道, 并尽可能保留局部邻域内的重要信息, 从而提高网络的训练效果。

1　相关工作

目前, 基于深度学习的三维点云语义分割算法主要分为三大类: 基于投影的方法、基于体素化的方法和基于点云的方法。

1.1　基于投影的方法

基于投影的方法通常将三维点云投影到二维图像中, 利用二维卷积方法提取三维模型的表面信息。Tatarchenko 等[8] 提出了 TangentConv 网络, 将点云投影到曲面上, 利用切线卷积进行点云分割, 但该方法会丢失点云的几何和结构信息。Wu 等[9] 提出并设计了一种利用球面投影完成三维点云语义分割的 SqueezeSeg 网络, 该方法先将三维点云投影到球面中, 每个球面都是一张二维图像, 再利用 SqueezeNet 网络进行特征提取, 最终通过条件随机场(conditional random field, CRF)判别模型的优化, 实现对输入的三维点云进行每个点的语义信息预测。之后, 为优化点云语义信息预测的结果, 利用无监督的管道方法提出了 SqueezeSegV2 网络[10], 并提升了点云语义分割的精度。Milioto 等[11] 提出了 RangeNet++网络, 将二维范围图像的语义标签传输到三维点云, 使用基于 GPU 加速的后处理方法来减轻离散化错误和推理输出模糊的问题。球形投影虽然保留了更多信息, 并且适合于激光雷达 LiDAR 点云的标记, 但这种通过投影的中间表示仍会不可避免地损失一些几何和结构信息。

1.2　基于体素化的方法

体素化方法主要是将不规则的点云转化为规则的体素网格, 利用标准的三维卷积分析和处理。VoxNet 模型[12] 最早将点云采样成体积网格后输入到 Volumetric CNN 网络中实现点云分割, 相较于二维图像数据, 其计算开销更大, 效率也更低。针对体素网格低分辨率的限制, SEGCloud 网络[13] 引入了三线性插值(trilinear interpolation, TI)和全连接条件随机场(fully connected conditional random field, FCCRF)来实现细粒度和全局一致性。为减少不必要的计算和内存消耗, Riegler 等[14] 采用八叉树构建网络 OctNet, 用于将存储器分配和计算集中到相关的密集区域, 在不影响分辨率的情况下实现更深层网络; Zeng 和 Gevers[15] 构建了基于 KD-tree 的 KD-Net, 用于提高点云计算和存储效率, 但此类树结构方法依赖体素边界, 并未充分利用局部几何结构。

1.3　基于点云的方法

基于点云的方法直接对不规则点云进行处理。PointNet++[16] 是在原有 PointNet 的基础上对点的特征的构建提出了更加有效的方法, 其利用点的邻域空间, 通过下采样的方式同时构建每个采样点的局部空间, 对每个局部空间利用 PointNet 的基础结构进行空间的特征提取, 并进行对应的上采样特征插值, 最终得到了每个点的特征, 但忽略了单个点与其邻域点的几何关系。针对这一问题, 文献 [17] 提出了一种多特征融合方法, 通过构建注意力融合层学习全局单点特征和局部几何特征的隐含关系来充分挖掘更能表征模型类别的细粒度几何特征, 并将全局单点特征和细粒度几何特征进一步融合, 达到优势互补、增强特征丰富性的目的。然而, 点云数据中, 并不是所有点之间都有关联, 有些点与点之间的关系强度很低。为此, Zhao 等[18] 设计了一种基于点对联系的点云分割网络 PointWeb, 其主要模块为自适应特征调整(adaptive feature adjustment, AFA), 该算法完整描述了在点集构成的局部空间中, 每个点与其余点之间的强度联系, 可以根据这种联系使每个点都能从所有其他点收集特征, 从而增强特征描述局部邻域的能力, 但对于点云边缘部分的分割效果并不理想。为解决这一问题, Wang 等[19] 提出一种图注意力卷积网络 GACNet, 主要通过建立每个点与周围点的图结构, 并通过引入注意力机制计算中心点与每一个邻接点的边缘权重, 根据动态学习到的特征有选择地关注它们中最相关的部分, 捕获点云的结构化特征以进行细粒度分割, 但该算法在一定程度上忽略了邻域点间的深层次语义信息。

综上所述, 目前大多数基于深度学习的点云语义分割算法过分关注中心点与邻域点的特征提取, 而忽略了邻域点间相互结构关系的学习, 并且在聚合局部邻域特征时, 通常选择最大值特征来代表局部邻域特征, 导致丢失一些具有鉴别性的特征。因此, 如何在局部邻域内将中心点的局部特征与邻域点间的深层次语义特征相结合, 并在聚合局部邻域特征时保留其他通道的重要特征是三维点云语义分割研究的一个亟待解决的问题。

2　融合双注意力机制和动态图卷积神经网络

本文构建的融合双注意力机制和 DGCNN 的架

构主要由改进的边缘卷积(EdgeConv++)模块、空间注意力机制(spatial attention mechanism, SAM)模块和通道注意力机制(channel attention mechanism, CAM)模块组成,网络结构如图1所示。在提取局部邻域特征时,首先,通过边缘卷积操作EdgeConv++构造边缘特征,并将中心点与邻域点的相对距离 $\boldsymbol{y}_{ij}$ 输入到核点卷积(kernel point convolution, KPConv模块)[20],得到增强后的局部特征 $\boldsymbol{c}_{ij}$,通过多层感知机(multilayer perceptron, MLP)将局部特征和全局特征连接起来得到边缘特征 $\boldsymbol{h}_{ij}$;然后,通过SAM模块学习局部邻域内的邻域点深层语义特征,并对边缘特征进行增强,通过CAM模块对边缘特征各通道赋予不同的权重;最后,使用最大池化操作得到聚合后的局部邻域特征作为中心点的新特征。

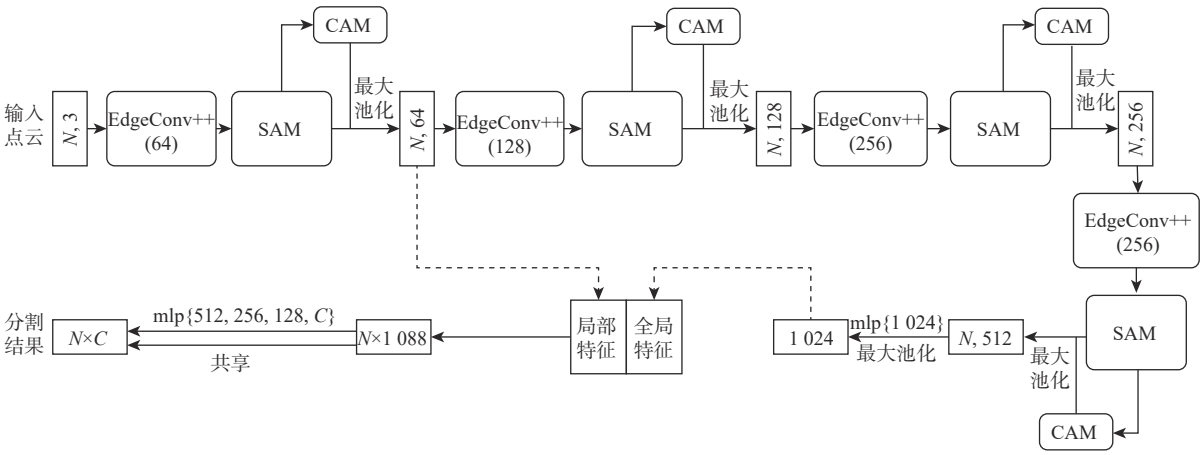


图1 融合双注意力机制和DGCNN的模型结构

Fig. 1 Structure of model fusing dual attention mechanism and DGCNN

2.1 EdgeConv++模块

对EdgeConv模块进行改进,与DGCNN中定义的边缘特征 $\boldsymbol{e}_{ij}$ 相似,既考虑中心点 $\boldsymbol{x}_i$ 的全局信息,也考虑中心点与邻域点的局部信息,如下:

$$\boldsymbol{e}_{ij}=l_\theta\left(\boldsymbol{x}_i,\boldsymbol{x}_{ij}-\boldsymbol{x}_i\right) \tag{1}$$

式中:$l_\theta:\mathbf{R}^F\times\mathbf{R}^F\rightarrow\mathbf{R}^{F'}$ 为具有一组可学习参数 θ 组成的非线性函数。

但本文的局部信息不只是通过计算中心点与邻域点的相对距离 $\boldsymbol{y}_{ij}=\boldsymbol{x}_{ij}-\boldsymbol{x}_i$,而是受KPConv的启发。将 $\boldsymbol{y}_{ij}$ 作为输入,经过一个KPConv后得到增强后的局部特征 $\boldsymbol{c}_{ij}$,KPConv的计算过程如图2所示。将全局特征与局部特征拼接形成新的边缘特征 $\boldsymbol{h}_{ij}$,如下:

$$\boldsymbol{h}_{ij}=l_\theta\left(\boldsymbol{x}_i,\boldsymbol{c}_{ij}\right)=l_\theta\left(\boldsymbol{x}_i,\mathrm{KPConv}\left(\boldsymbol{y}_{ij}\right)\right) \tag{2}$$

EdgeConv++模块结构如图3所示。

具体做法为:首先,对于点云中的任意一个点 $\boldsymbol{x}_i$,通过K最近邻(K-nearest neighbor, KNN)算法构建一个以点 $\boldsymbol{x}_i$ 为中心点的局部邻域图,得到 $\boldsymbol{x}_i$ 的 k 个邻居点 $N(\boldsymbol{x}_i)=\{\boldsymbol{x}_{i1},\boldsymbol{x}_{i2},\cdots,\boldsymbol{x}_{ik}\}$。然后,以 $\boldsymbol{x}_i$ 为球心,以 $\boldsymbol{x}_i$ 到 $N(\boldsymbol{x}_i)$ 邻居点的最远距离为半径 r 构建一个球。在球体内,找 d 个点作为核心点 $\tilde{\boldsymbol{x}}_d$,核心点并不是点云中的点,而是通过球心与各核心点之间的引力和斥力得到的。对于每个核心点 $\tilde{\boldsymbol{x}}_d$,分别有一个权重矩阵 $\boldsymbol{W}_d$ 与之对应。对于任意的 $\boldsymbol{y}_{ij}$,定义核函数 $\boldsymbol{g}$ 为

$$\boldsymbol{g}\left(\boldsymbol{y}_{ij}\right)=\sum_{d<D}h\left(\boldsymbol{y}_{ij},\tilde{\boldsymbol{x}}_d\right)\boldsymbol{W}_d \tag{3}$$

式中:D 为核心点总数。

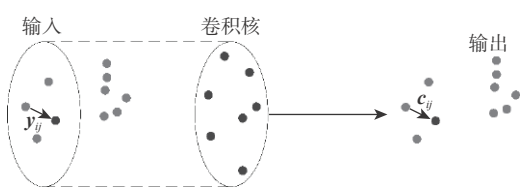


图2 KPConv模块结构示意图

Fig. 2 Structure of KPConv module

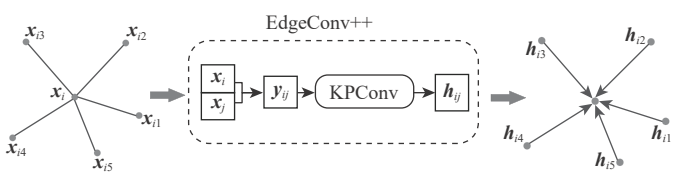


图3 EdgeConv++模块结构示意图

Fig. 3 Structure of EdgeConv++ module

可以看出, $\boldsymbol{g}(\boldsymbol{y}_{ij})$ 相当于对 d 个权重矩阵加权求和。权重系数 $h(\boldsymbol{y}_{ij},\tilde{\boldsymbol{x}}_d)$ 定义为

$$h\left(\boldsymbol{y}_{ij},\tilde{\boldsymbol{x}}_d\right)=\max\left(0,\ 1-\frac{\left\|\boldsymbol{y}_{ij}-\tilde{\boldsymbol{x}}_d\right\|}{\beta}\right) \tag{4}$$

式中: β 为核心点的影响距离, 用来确保每个核心点的影响区域之间有重叠, 其大小由点云的输入密度确定。

每个权重矩阵 $\boldsymbol{W}_d$ 的系数由 $\boldsymbol{W}_d$ 对应的核心点 $\tilde{\boldsymbol{x}}_d$ 到 $\boldsymbol{y}_{ij}$ 的相对距离来确定。相对距离越小, 权重越大, 最大值为 1; 相对距离越大, 权重越小, 最小值为 0。即距离越近, 相关性越大; 反之亦然。$\boldsymbol{g}(\boldsymbol{y}_{ij})$ 可以看作专门针对每个 x_{ij} 计算出来一个权重矩阵, 利用该权重矩阵对局部特征 $\boldsymbol{y}_{ij}$ 进行特征变换得到 $\boldsymbol{c}_{ij}$, 此时得到的 $\boldsymbol{c}_{ij}$ 既包含了中心点与邻域点的局部几何信息, 也包含了邻域点与核心点之间的位置信息。

$$\boldsymbol{c}_{ij}=\boldsymbol{g}\left(\boldsymbol{y}_{ij}\right)\boldsymbol{y}_{ij} \tag{5}$$

$\boldsymbol{g}(\boldsymbol{y}_{ij})$ 可通过邻域点与核心点之间的相对位置关系计算得到, 因此, 在 KPConv 中直接将 $\boldsymbol{c}_{ij}$ 当作边缘特征而没有考虑中心点的全局特征, 核心点虽然由中心点生成, 可在一定程度上保留原始中心点的全局信息, 但还是会不可避免地导致中心点的描述能力下降, 进而丢失部分全局特征。

2.2　SAM 模块

现有网络中的注意力机制[17,19]侧重于采样点与邻域内的局部信息交互, 普遍忽略了邻域点相互结构关系的学习, 而空间注意力的引入可以建立邻域点之间的依赖关系, 无论距离远近, 只要特征相似就会相互关联, 从而在局部邻域内提取到更深层次的邻域上下文信息。SAM 模块结构如图 4 所示。

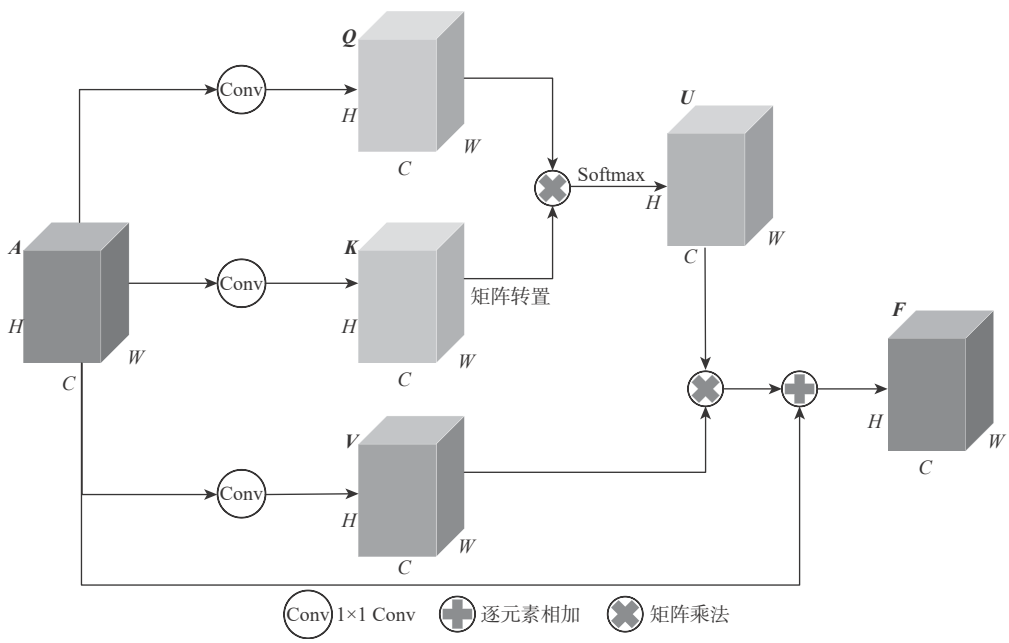


图 4　SAM 模块结构示意图

Fig. 4　Structure of SAM module

受 DANet[21] 启发, 对于给定的特征向量 $\boldsymbol{A}\in\mathbf{R}^{C\times H\times W}$, 分别使用 2 个独立的卷积操作生成 2 个新的特征向量 $\boldsymbol{Q}\in\mathbf{R}^{C\times H\times W}$ 和 $\boldsymbol{K}\in\mathbf{R}^{C\times H\times W}$。对 $\boldsymbol{Q}$ 和 $\boldsymbol{K}$ 进行矩阵相乘, 并应用 Softmax 函数得到归一化的空间注意力权重矩阵 $\boldsymbol{U}_{ij}\in\mathbf{R}^{N\times N}$:

$$\boldsymbol{U}_{ij}=\mathrm{Softmax}\left(\boldsymbol{Q}_i\boldsymbol{K}_j\right)=\frac{\exp\left(\boldsymbol{Q}_i\boldsymbol{K}_j^{\mathrm{T}}\right)}{\sum_{i=1}^{N}\sum_{j=1}^{N}\exp\left(\boldsymbol{Q}_i\boldsymbol{K}_j^{\mathrm{T}}\right)} \tag{6}$$

式中: N 为特征图大小, 即 $N=H\times W$; $\boldsymbol{Q}_i$ 和 $\boldsymbol{K}_j$ 分别为点 i 和点 j 在 $\boldsymbol{Q}$ 和 $\boldsymbol{K}$ 中的特征表示; $\boldsymbol{U}_{ij}$ 衡量第 i 个点与第 j 个点之间的相似性, 相似性越高, 值越大, 在一定程度上反映了点之间的相关性。对特征向量 $\boldsymbol{A}$ 应用卷积操作得到新的特征向量 $\boldsymbol{V}\in\mathbf{R}^{C\times H\times W}$, 并将其与注意力权重 $\boldsymbol{U}_{ij}$ 进行矩阵相乘操作, 最终与尺度参数 α 相乘, 并与输入特征 $\boldsymbol{A}$ 进行逐元素求和操作, 即可得到增强的输出特征 $\boldsymbol{F}$。

$$\boldsymbol{F}=\alpha\sum_{j=1}^{N}\left(\boldsymbol{U}_{ij}\boldsymbol{V}_j\right)+\boldsymbol{A}_i \tag{7}$$

式中: α 为初始值为 0 的可学习的尺度参数。

2.3　CAM 模块

目前, 大多数方法[22-23]在提取到点云的局部邻域特征 $\boldsymbol{f}_i$ 时, 通常选择局部邻域特征中的最大值特

征来代表该局部邻域特征，定义为$\boldsymbol{F}_i=\max(\boldsymbol{f}_i)$。但是，简单地对局部邻域进行最大值池化操作，难免会丢失局部邻域内的其他特征。因此，在聚合特征时，有必要选择具有鉴别性和有效性的特征。最近的工作[24-25]在二维图像分割领域取得了不错的效果，受此启发，本文在对局部邻域特征的聚合过程中设计了一个针对三维点云数据的CAM模块，通过给不同通道赋予不同的权值以达到强调有用通道同时抑制无用通道的目的，从而提高网络训练的效果。CAM模块结构如图5所示。

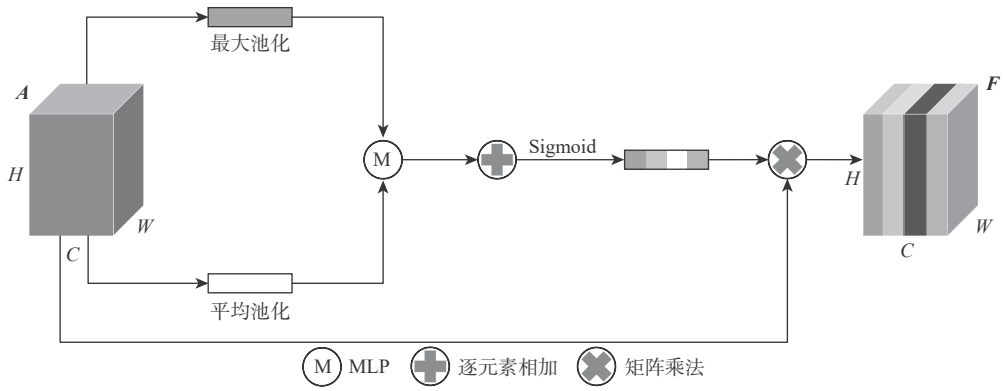


图5 CAM模块结构示意图

Fig. 5 Structure of CAM

首先，对输入特征向量$\boldsymbol{A}$在空间维度分别进行最大池化和平均池化，使其只保留通道的维度，其他维度为1，得到一个一维的矢量。然后，分别经过一个共享权重的MLP，将两者的输出做逐元素相加，再经过Sigmoid激活函数，得到通道注意力权重。最后，将该权重和输入特征向量$\boldsymbol{A}$做逐元素乘法，得到经过通道注意力机制后的特征向量$\boldsymbol{F}$：

$$\boldsymbol{F}=\sigma\left(\mathrm{MLP}(\mathrm{MaxPool}(\boldsymbol{A}))+\mathrm{MLP}(\mathrm{AvgPool}(\boldsymbol{A}))\right)\boldsymbol{A} \tag{8}$$

式中：σ表示Sigmoid激活函数。

3 实 验

3.1 评估指标及参数设置

为评估本文所构建的融合双注意力机制和DGCNN架构的有效性，选用流行的大规模室内场景数据集S3DIS[26]和大规模室外场景数据集SemanticKITTI[27]进行实验。评估指标采用总体准确率(overall accuracy, OA)和平均交并比(mean intersection-over-union, mIoU)，分别表示为

$$A=\frac{T_{\mathrm{P}}+T_{\mathrm{N}}}{T_{\mathrm{P}}+T_{\mathrm{N}}+F_{\mathrm{P}}+F_{\mathrm{N}}} \tag{9}$$

$$M=\frac{1}{c+1}\sum_{i=0}^{c}\frac{T_{\mathrm{P}}}{F_{\mathrm{N}}+F_{\mathrm{P}}+T_{\mathrm{P}}} \tag{10}$$

式中：T_{P}表示被模型预测为正类的正样本；T_{N}表示被模型预测为负类的负样本；F_{P}表示被模型预测为正类的负样本；F_{N}表示被模型预测为负类的正样本；c为数据集中的类别数。

本文的实验环境配置为：硬件为Intel(R) Core(TM) i9-10900K CPU@3.70GHz和NVIDIA RTX3090 GPU(24 GB)，软件为Linux Ubuntu 18.04和深度学习框架PyTorch1.11.0。实验采用了动量因子为0.99的基于动量的随机梯度下降(stochastic gradient descent, SGD)优化算法，使用Adam算法更新SGD的步长。

3.2 室内场景语义分割

S3DIS数据集是从3个不同建筑物的6个区域所包含的271个房间中采样生成的RGB-D点云数据。将房间内的点云数据划分为1 m×1 m的块，再对每个块预测其中每个点的语义标签。每个点由一个9维向量表示，分别为：$\boldsymbol{X}$、$\boldsymbol{Y}$、$\boldsymbol{Z}$、$\boldsymbol{R}$、$\boldsymbol{G}$、$\boldsymbol{B}$、$\boldsymbol{X}'$、$\boldsymbol{Y}'$、$\boldsymbol{Z}'$。其中，$\boldsymbol{X}$、$\boldsymbol{Y}$、$\boldsymbol{Z}$表示位置坐标，$\boldsymbol{R}$、$\boldsymbol{G}$、$\boldsymbol{B}$表示颜色信息，$\boldsymbol{X}'$、$\boldsymbol{Y}'$、$\boldsymbol{Z}'$为标准化后的每个块相对于所在房间的位置坐标(值为0～1)。

为验证本文算法的有效性，与经典算法在S3DIS数据集上进行了分割对比实验，结果如表1所示。可以看出，本文算法获得了66.0%的mIoU和92.8%的OA，在OA方面表现最优，虽然mIoU略逊于KPConv的可变形核，但在3个类别(ceiling、wall、table)上实现了最佳性能。由于本文算法使用的是KPConv的刚性核而非可变形核，避免了学习局部偏移量而修改核点位置的操作，这也导致了KPConv的可变形核对于大小各异的物体(如chair、bookcase)的分割效果优于本文算法。与BAAF-Net相比，本文算法的mIoU和OA这2个分割精度评价指标均高于该算法，主要由于BAAF-Net在提取局部邻域特征时，并未考虑邻域点间的深层次语义信息。与常规的6折交叉验证方法不

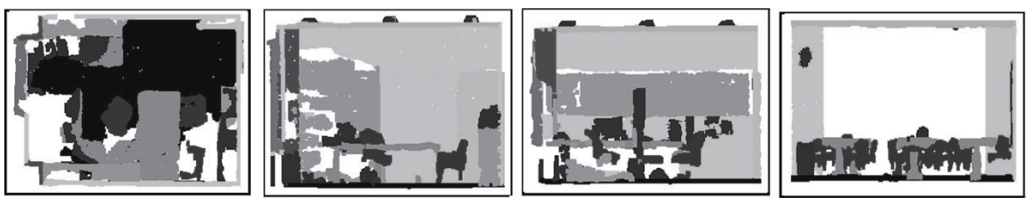
表 1　不同算法在 S3DIS 数据集上的分割精度对比（Area 5 作为测试）

Table 1　Comparison of segmentation accuracy of different algorithms on S3DIS dataset (Area 5 as a test)　%

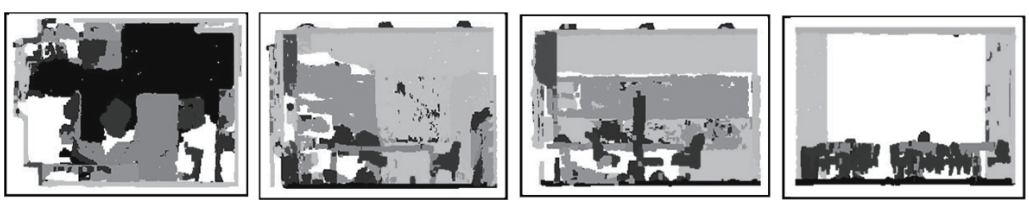
算法	OA	mIoU	IoU												
			ceiling	floor	wall	beam	column	window	door	table	chair	sofa	bookcase	board	clutter
PointNet[6]	79.3	41.1	88.8	97.3	69.8	0.1	3.9	46.3	10.8	59.0	52.6	5.9	40.3	26.4	33.2
TangentConv[8]	82.5	52.6	90.5	97.7	74.0	0	20.7	39.0	31.3	77.5	69.4	57.3	38.5	48.8	39.8
PointCNN[28]	85.9	57.3	92.3	98.2	79.4	0	17.6	22.8	62.1	74.4	80.6	31.7	66.7	62.1	56.7
SPG[29]	86.4	58.0	89.4	96.9	78.1	0	42.8	48.9	61.6	84.7	75.4	69.8	52.6	2.1	52.2
PCCN[30]		58.3	92.3	96.2	75.9	0.3	6.0	69.5	63.5	66.9	65.6	47.3	68.9	59.1	46.2
PointWeb[18]	87.0	60.3	92.0	98.5	79.4	0	21.1	59.7	34.8	76.3	88.3	46.9	69.3	64.9	52.5
HPEIN[31]	87.2	61.9	91.5	98.2	81.4	0	23.3	65.3	40.0	75.5	87.7	58.5	67.8	65.6	49.4
RandLA-Net[32]	87.2	62.4	91.1	95.6	80.2	0	24.7	62.3	47.7	76.2	83.7	60.2	71.1	65.7	53.8
GACNet[19]	87.8	62.8	92.3	98.3	81.9	0	20.3	59.1	40.8	78.5	85.8	61.7	70.7	74.7	52.8
BAAF-Net[33]	88.9	65.4	92.9	97.9	82.3	0	23.1	65.5	64.9	78.5	87.5	61.4	70.7	68.7	57.2
KPConv[20]		67. 1	92.8	97.3	82.4	0	23.9	58.0	69.0	81.5	91.0	75.4	75.3	66.7	58.9
本文	92.8	66.0	92.9	98.3	82.8	0	21.2	56.6	68.5	91.3	81.3	74.4	64.2	68.9	57.8

同, 本文将 S3DIS 数据集的 Area 5 区域单独作为测试区域, 而将 Area 1～Area 4 和 Area 6 区域用于训练。由于 Area 5 和其他区域中的房间有所不同, 包含的物体存在差异, 对网络模型的要求更高。通过引入 SAM 模块, 本文算法提取到了局部邻域内任意点对之间更深层次的特征, 同时引入 CAM 模块, 在聚合局部邻域特征时能够根据不同通道的重要程度产生更具代表性的特征, 从而提高点云语义分割的精度。

此外, 为定性分析语义分割结果, 对 Area 5 的测试结果进行了可视化, 如图 6 所示。可以看出, 本文算法可以高效、准确地提取到点云特征, 在某些场景下得到的分割结果已经很接近真实标签。



(a) 真实标签



(b) 分割结果

图 6　S3DIS 数据集分割结果的可视化

Fig. 6　Visualization of segmentation results on S3DIS dataset

3.3　室外场景语义分割

SemanticKITTI 数据集包含了德国卡尔斯鲁厄附近的市内交通、居民区、高速公路和乡村道路场景, 是目前最大的汽车激光雷达点云数据集。该数据集由 22 个点云序列组成, 00～10 序列作为训练集(08 序列用作验证集并生成可视化结果), 共包含 23 201 帧三维点云场景; 11～21 序列作为测试集, 共包含 20 351 帧三维点云场景。训练集中每点都有自己的标签, 人为进行了语义标注, 共有 19 种类别标签。表 2 为本文算法在 SemanticKITTI 数据集上与其他经典算法的场景语义分割定量结果对比。可以看出, 与目前主流算法相比, 本文算法的 mIoU 为 59.4%, 仅次于 BAAF-Net, 并且在其中 6 个类别上也取得了最好的分割结果。而 BAAF-Net 的输入与目前主流算法不同, 不仅包含点云的位置信息, 还包含一些额外的特征(如法线、颜色等), 因此, BAAF-Net 对一些小物体 (如 other-vehicle 和 motorcyclist) 进行分割时, 能够利用这些额外的特

表2 不同算法在SemanticKITTTI数据集上的分割精度对比

Table 2 Comparison of segmentation accuracy of different algorithms on SemanticKITTI dataset %

算法	mIoU	IoU								
		road	sidewalk	parking	other-ground	building	car	truck	bicycle	motorcycle
PointNet[6]	14.6	61.6	35.7	15.8	1.4	41.4	46.3	0.1	1.3	0.3
SPG[29]	17.4	45.0	28.5	1.6	0.6	64.3	49.3	0.1	0.2	0.2
PointNet++[16]	20.1	72.0	41.8	18.7	5.6	62.3	53.7	0.9	1.9	0.2
TangentConv[8]	40.9	83.9	63.9	33.4	15.4	83.4	90.8	15.2	2.7	16.5
SpSequenceNet[34]	43.1	90.1	73.9	57.6	27.1	91.2	88.5	29.2	24.0	0
PointASNL[35]	46.8	87.4	74.3	24.3	1.8	83.1	87.9	39.0	0	25.1
HPGCNN[36]	50.5	89.5	73.6	58.8	34.6	91.2	93.1	21.0	6.5	17.6
RangeNet++[11]	52.2	91.8	75.2	65.0	27.8	87.4	91.4	25.7	25.7	34.4
RandLA-Net[32]	53.9	90.7	73.7	60.3	20.4	86.9	94.2	40.1	26.0	25.8
PolarNet[37]	54.3	90.8	74.4	61.7	21.7	90.0	93.8	22.9	40.3	30.1
3D-MiniNet[38]	55.8	91.6	74.5	64.2	25.4	89.4	90.5	28.5	42.3	42.1
SAFFGCNN[39]	56.6	89.9	73.9	63.5	35.1	91.5	95	38.3	33.2	35.1
KPConv[20]	58.8	88.8	72.7	61.3	31.6	90.5	96.0	33.4	30.2	42.5
BAAF-Net[33]	59.9	90.9	74.4	62.2	23.6	89.8	95.4	48.7	31.8	35.5
本文	59.4	92.0	77.4	70.1	35.2	89.7	94.4	38.4	43.7	40.3

算法	IoU									
	other-vehicle	vegetation	trunk	terrain	person	bicyclist	motorcyclist	fense	pole	traffic-sign
PointNet[6]	0.8	31.0	4.6	17.6	0.2	0.2	0	12.9	2.4	3.7
SPG[29]	0.8	48.9	27.2	24.6	0.3	2.7	0.1	20.8	15.9	0.8
PointNet++[16]	0.2	46.5	13.8	30.0	0.9	1.0	0	16.9	6.0	8.9
TangentConv[8]	12.1	79.5	49.3	58.1	23.0	28.4	8.1	49.0	35.8	28.5
SpSequenceNet[34]	22.7	84.0	66.0	65.7	6.3	0	0	67.7	50.8	48.7
PointASNL[35]	29.2	84.1	52.2	70.6	34.2	57.6	0	43.9	57.8	36.9
HPGCNN[36]	23.3	84.4	65.9	70.0	32.1	30.0	14.7	65.5	45.5	41.5
RangeNet++[11]	23.0	80.5	55.1	64.6	38.3	38.8	4.8	58.6	47.9	55.9
RandLA-Net[32]	38.9	81.4	61.3	66.8	49.2	48.2	7.2	56.3	49.2	47.7
PolarNet[37]	28.5	84.0	65.5	67.8	43.2	40.2	5.6	61.3	51.8	57.5
3D-MiniNet[38]	29.4	82.8	60.8	66.7	47.8	44.1	14.5	60.8	48.0	56.6
SAFFGCNN[39]	28.7	84.4	67.1	69.5	45.3	43.5	7.3	66.1	54.3	53.7
KPConv[20]	31.6	84.8	69.2	69.1	61.5	61.6	11.8	64.2	56.4	48.4
BAAF-Net[33]	46.7	82.7	63.4	67.9	49.5	55.7	53.0	60.8	53.7	52.0
本文	30.3	84.3	64.9	70.0	60.1	47.4	7.6	66.9	53.1	62.4

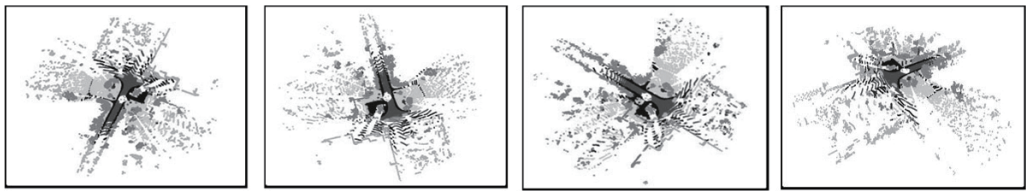
征，从而取得较好的结果。本文在提取局部特征时使用了KPConv的核点方法，但由于加入了SAM模块与CAM模块，得到的大部分场景mIoU均优于KPConv算法。图7为本文算法在SemanticKITTI数据集上的语义分割可视化结果。可以看出，由于本文算法能够提取到更深层次的局部邻域信息，即使在处理大规模复杂室外点云数据时，仍然能够产生较好的分割效果。

3.4 消融实验

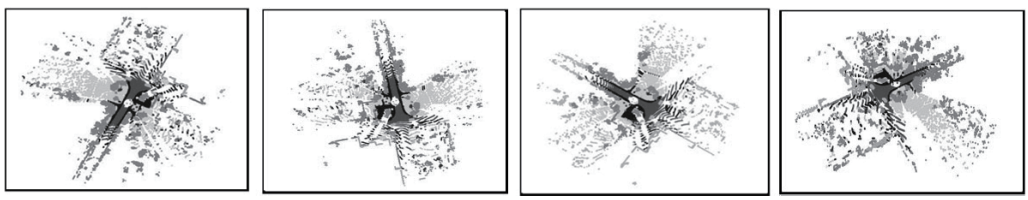
3.4.1 双注意力机制模块

为进一步说明本文引入的2种注意力模块的有效性，在S3DIS数据集上分别对这2个模块进行消融实验，结果如表3所示。其中，网络模型Net Model-3为完整的融合双注意力机制和动态图卷积神经网络，而网络模型Net Model-1是在所构建网络的基础上移除了SAM模块，网络模型Net Model-2则是移除了CAM模块。

从表3中结果可以看出，网络模型Net Model-3在S3DIS数据集上的mIoU比网络模型Net Model-1高出3.2%，验证了空间注意力的有效性。原因在于：SAM模块能够在局部邻域内学习到任意2个特征相似的点的深层次语义信息，从而使每个点的特征都得到一定程度的增强。而网络模型Net Model-3在S3DIS数据集上的mIoU比网络模型Net Model-2



(a) 真实标签



(b) 分割结果

图 7　SemanticKITTI 数据集分割结果的可视化

Fig. 7　Visualization of segmentation results on SemanticKITTI dataset

表 3　不同模块在 S3DIS 数据集上的消融实验

Table 3　Ablation experiments of different modules on S3DIS dataset

网络模型	模块名	mIoU/%
Net Model-1	CAM	62.8
Net Model-2	SAM	65.2
Net Model-3	SAM+CAM	66.0

只高出 0.8%, 因为在加入 CAM 模块后, 虽然会增强有用通道的权重, 但还是会通过最大池化操作来聚合局部邻域特征, 所以, 对于点云的语义分割效果提升相比于 SAM 模块并不是很大。

3.4.2　改进的动态图卷积模块

为进一步说明本文改进的动态图卷积模块的有效性, 在 S3DIS 数据集上分别对 EdgeConv++和 KPConv 这 2 个模块进行消融实验, 结果如表 4 所示。其中, 网络模型 Net Model-6 为完整的融合双注意力机制和 DGCNN, 而网络模型 Net Model-4 是在所构建网络的基础上移除了刚性核的核点卷积(KPConv rigid)模块, 网络模型 Net Model-5 则是移除了动态图卷积模块, 网络模型 Net Model-7 为使用 KPConv 原网络模型的刚性核。

从表 4 中结果可以看出, 虽然在单独使用动态图卷积或刚性核的核点卷积进行边缘特征提取时, 本文的结果不如 KPConv 原网络模型的刚性核, 但将二者相结合(即改进的动态图卷积)得到的结果优于 KPConv 原网络模型的刚性核, 这也证明了本文所提出的改进的动态图卷积模块的有效性。

表 4　改进的动态图卷积模块有效性验证

Table 4　Effectiveness verification of improved dynamic graph convolution module

网络模型	模块名	mIoU/%
Net Model-4	EdgeConv++	64.2
Net Model-5	KPConv	64.7
Net Model-6	KPConv+EdgeConv++	66.0
Net Model-7	KPConv rigid	65.4

3.4.3　核点卷积模块中的核心点数量

核心点的数量会影响网络提取到的局部几何特征的优劣, 较少的核心点数量 d 使网络无法充分学习到有效的局部特征, 导致分割精度较差; 而 d 的数量过大反而会造成更多冗余的局部特征, 增加网络计算量。从表 5 可以看出, 当核心点数量 d<15 时, 随着 d 的增大, 分割结果也在逐步提升; 而当 d>15 时, 由于核心点数量过多, 难免会学习到冗余的特征, 反而影响了分割的结果。因此, 本文在实验中的参数设置为 d=15。

表 5　核心点数量对分割结果的影响

Table 5　Influence of number of kernel points on segmentation results

d	mIoU/%
11	64.8
13	65.3
15	66.0
17	65.7

4　结　论

本文针对三维点云语义分割算法在提取局部特征时忽略邻域点间深层次语义信息, 聚合局部邻域特征时忽略其他邻域特征中有用信息的问题, 提出融合双注意力机制和 DGCNN 的三维点云语义分割算法。

1) 采用动态图采样边缘特征, 以及使用通道、空间注意力有效提升语义分割的精度, 在 S3DIS 和 SemanticKITTI 数据集上, 语义分割精度分别达到了 66.0% 和 59.4%。

2）由于网络采用KNN算法构建局部邻域图，会导致网络鲁棒性较差，从而影响点云的分割结果。因此，如何处理稀疏数据集中具有相似几何结构特征的物体仍是需要研究的重点。

参考文献（References）

[1] ZHU Y K, MOTTAGHI R, KOLVE E, et al. Target-driven visual navigation in indoor scenes using deep reinforcement learning[C]// Proceedings of the IEEE International Conference on Robotics and Automation. Piscataway: IEEE Press, 2017: 3357-3364.

[2] ZHANG K G, XIONG C H, ZHANG W, et al. Environmental features recognition for lower limb prostheses toward predictive walking[J]. IEEE Transactions on Neural Systems and Rehabilitation Engineering, 2019, 27(3): 465-476.

[3] QI C R, LIU W, WU C X, et al. Frustum PointNets for 3D object detection from RGB-D data[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2018: 918-927.

[4] ZHANG G C, YU J, WANG Z H, et al. Visual 3D reconstruction system based on RGBD camera[C]//Proceedings of the IEEE 4th Information Technology, Networking, Electronic and Automation Control Conference. Piscataway: IEEE Press, 2020: 908-911.

[5] 姜枫, 顾庆, 郝慧珍, 等. 基于内容的图像分割方法综述[J]. 软件学报, 2017, 28(1): 160-183.
JIANG F, GU Q, HAO H Z, et al. Survey on content-based image segmentation methods[J]. Journal of Software, 2017, 28(1): 160-183 (in Chinese).

[6] CHARLES R Q, HAO S, MO K C, et al. PointNet: Deep learning on point sets for 3D classification and segmentation[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2017: 77-85.

[7] WANG Y, SUN Y B, LIU Z W, et al. Dynamic graph CNN for learning on point clouds[J]. ACM Transactions on Graphics, 2019, 38(5): 1-12.

[8] TATARCHENKO M, PARK J, KOLTUN V, et al. Tangent convolutions for dense prediction in 3D[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2018: 3887-3896.

[9] WU B C, WAN A, YUE X Y, et al. SqueezeSeg: Convolutional neural nets with recurrent CRF for real-time road-object segmentation from 3D LiDAR point cloud[C]//Proceedings of the IEEE International Conference on Robotics and Automation. Piscataway: IEEE Press, 2018: 1887-1893.

[10] WU B C, ZHOU X Y, ZHAO S C, et al. SqueezeSegV2: Improved model structure and unsupervised domain adaptation for road-object segmentation from a LiDAR point cloud[C]//Proceedings of the International Conference on Robotics and Automation. Piscataway: IEEE Press, 2019: 4376-4382.

[11] MILIOTO A, VIZZO I, BEHLEY J, et al. RangeNet++: Fast and accurate LiDAR semantic segmentation[C]//Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems. Piscataway: IEEE Press, 2019: 4213-4220.

[12] MATURANA D, SCHERER S. VoxNet: A 3D convolutional neural network for real-time object recognition[C]//Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems. Piscataway: IEEE Press, 2015: 922-928.

[13] TCHAPMI L, CHOY C, ARMENI I, et al. SEGCloud: Semantic segmentation of 3D point clouds[C]//Proceedings of the International Conference on 3D Vision. Piscataway: IEEE Press, 2017: 537-547.

[14] RIEGLER G, ULUSOY A O, GEIGER A. OctNet: Learning deep 3D representations at high resolutions[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2017: 6620-6629.

[15] ZENG W, GEVERS T. 3DContextNet: k-d tree guided hierarchical learning of point clouds using local and global contextual cues[C]// Proceedings of the European Conference on Computer Vision. Berlin: Springer, 2018: 314-330.

[16] QI C R, YI L, SU H, et al. PointNet++: Deep hierarchical feature learning on point sets in a metric space[C]//Proceedings of the 31st International Conference on Neural Information Processing Systems. Cambridge: MIT Press, 2017: 5099-5108.

[17] 党吉圣, 杨军. 多特征融合的三维模型识别与分割[J]. 西安电子科技大学学报, 2020, 47(4): 149-157.
DANG J S, YANG J. 3D model recognition and segmentation based on multi-feature fusion[J]. Journal of Xidian University, 2020, 47(4): 149-157(in Chinese).

[18] ZHAO H S, JIANG L, FU C W, et al. PointWeb: Enhancing local neighborhood features for point cloud processing[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2019: 5560-5568.

[19] WANG L, HUANG Y C, HOU Y L, et al. Graph attention convolution for point cloud semantic segmentation[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2019: 10288-10297.

[20] THOMAS H, QI C R, DESCHAUD J E, et al. KPConv: Flexible and deformable convolution for point clouds[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. Piscataway: IEEE Press, 2019: 6410-6419.

[21] FU J, LIU J, TIAN H J, et al. Dual attention network for scene segmentation[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2019: 3141-3149.

[22] LE T, DUAN Y. PointGrid: A deep network for 3D shape understanding[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2018: 9204-9214.

[23] ZHANG Z Y, HUA B S, YEUNG S K. ShellNet: Efficient point cloud convolutional neural networks using concentric shells statistics [C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. Piscataway: IEEE Press, 2019: 1607-1616.

[24] HU J, SHEN L, ALBANIE S, et al. Squeeze-and-excitation networks[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2020, 42(8): 2011-2023.

[25] WOO S, PARK J C, LEE J Y, et al. CBAM: Convolutional block attention module[C]//Proceedings of the European Conference on Computer Vision. Berlin: Springer, 2018: 3-19.

[26] ARMENI I, SENER O, ZAMIR A R, et al. 3D semantic parsing of large-scale indoor spaces[C]//Proceedings of the IEEE Conference

on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2016: 1534-1543.

[27] BEHLEY J, GARBADE M, MILIOTO A, et al. SemanticKITTI: A dataset for semantic scene understanding of LiDAR sequences[C]// Proceedings of the IEEE/CVF International Conference on Computer Vision. Piscataway: IEEE Press, 2019: 9296-9306.

[28] LI Y Y, BU R, SUN M C, et al. PointCNN: Convolution on X-transformed points[C]//Proceedings of the Advances in Neural Information Processing Systems. Cambridge: MIT Press, 2018, 31: 828-838.

[29] LANDRIEU L, SIMONOVSKY M. Large-scale point cloud semantic segmentation with superpoint graphs[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2018: 4558-4567.

[30] WANG S L, SUO S, MA W C, et al. Deep parametric continuous convolutional neural networks[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2018: 2589-2597.

[31] JIANG L, ZHAO H S, LIU S, et al. Hierarchical point-edge interaction network for point cloud semantic segmentation[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. Piscataway: IEEE Press, 2019: 10432-10440.

[32] HU Q Y, YANG B, XIE L H, et al. RandLA-Net: Efficient semantic segmentation of large-scale point clouds[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2020: 11105-11114.

[33] QIU S, ANWAR S, BARNES N. Semantic segmentation for real point cloud scenes via bilateral augmentation and adaptive fusion[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2021: 1757-1767.

[34] SHI H Y, LIN G S, WANG H, et al. SpSequenceNet: Semantic segmentation network on 4D point clouds[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2020: 4573-4582.

[35] YAN X, ZHENG C D, LI Z, et al. PointASNL: Robust point clouds processing using nonlocal neural networks with adaptive sampling[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2020: 5588-5597.

[36] DANG J S, YANG J. HPGCNN: Hierarchical parallel group convolutional neural networks for point clouds processing[C]//Proceedings of the Asian Conference on Computer Vision. Berlin: Springer, 2020: 20-37.

[37] ZHANG Y, ZHOU Z X, DAVID P, et al. PolarNet: An improved grid representation for online LiDAR point clouds semantic segmentation[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2020: 9598-9607.

[38] ALONSO I, RIAZUELO L, MONTESANO L, et al. 3D-MiniNet: Learning a 2D representation from point clouds for fast and efficient 3D LiDAR semantic segmentation[J]. IEEE Robotics and Automation Letters, 2020, 5(4): 5432-5439.

[39] 杨军, 李博赞. 基于自注意力特征融合组卷积神经网络的三维点云语义分割[J]. 光学 精密工程, 2022, 30(7): 840-853.
YANG J, LI B Z. Semantic segmentation of 3D point cloud based on self-attention feature fusion group convolutional neural network[J]. Optics and Precision Engineering, 2022, 30(7): 840-853(in Chinese).

Semantic segmentation of point clouds by fusing dual attention mechanism and dynamic graph convolution

YANG Jun[1, 2, *], ZHANG Chen[1]

(1. School of Electronic and Information Engineering, Lanzhou Jiaotong University, Lanzhou 730070, China;
2. Faculty of Geomatics, Lanzhou Jiaotong University, Lanzhou 730070, China)

Abstract: The existing semantic segmentation methods of 3D point clouds based on deep learning usually ignore the profound semantic information between neighboring points when extracting local features and fail to consider the useful information in other neighboring features when aggregating local neighboring features. To solve these problems, a semantic segmentation algorithm of 3D point clouds fusing dual attention mechanism and dynamic graph convolution neural network (DGCNN) was proposed. Firstly, edge features were constructed by dynamic graph convolution operation, and the relative distance between the center point and the neighboring points was input to the kernel point convolution operation to obtain enhanced edge features, further strengthening the relationship between the center point and the neighboring points. Secondly, the spatial attention module was introduced to establish the dependence between neighboring points, and the similar feature points were intercorrelated, so as to extract profound context information in the local neighborhood and enrich the geometric features of neighboring points. Finally, the channel attention module was introduced when local neighboring features were aggregated. By giving different weights to different channels, the purpose of enhancing useful channels and suppressing useless channels was achieved, so as to improve the accuracy of semantic segmentation. The experimental results on the S3DIS dataset and SemanticKITTI dataset show that the semantic segmentation accuracy of this algorithm has reached 66.0% and 59.4%, respectively. Compared with other classical network models, this algorithm has achieved a better point cloud segmentation effect.

Keywords: 3D point cloud; semantic segmentation; deep learning; attention mechanism; dynamic graph convolution

Received: 2022-09-14; **Accepted**: 2023-01-02; **Published Online**: 2023-01-10 13:51

URL: link.cnki.net/urlid/11.2625.V.20230109.1713.007

Foundation items: National Natural Science Foundation of China (42261067,61862039); Talent Innovation and Entrepreneurship Project of Lanzhou City (2020-RC-22); Tianyou Innovation Team of Lanzhou Jiaotong University (TY202002)

* **Corresponding author**. E-mail: yangj@mail.lzjtu.cn

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2024.0130

液压破碎锤活塞横裂机理分析与工艺优化

童桂英[1]，万怡男[2,*]，张磊[3]，石岩[4]，张业明[5]

(1. 烟台大学 机电汽车工程学院，烟台 264005；2. 烟台理工学院 机电工程学院，烟台 264005；
3. 烟台艾迪精密机械股份有限公司，烟台 264000；4. 北京航空航天大学 自动化科学与电气工程学院，北京 100191；
5. 河南理工大学 机械与动力工程学院，焦作 454000)

摘 要：针对破碎锤中的活塞横裂现象，分析了活塞产生横裂的机理，并提出了一种新的活塞热处理制造工艺。采用硬度计和金相组织显微镜对活塞材料进行分析，得出发生活塞横裂故障部位的材料成分和金相组织符合设计要求，而硬度和硬化层深度低于设计要求。利用ANSYS求解活塞“卡死”情况下的最大应力为1 229.8 MPa，超出材料的屈服极限850 MPa，最大应力的位置在活塞横裂处；利用Fluent求解活塞所受的径向不平衡力，在活塞的回油槽中受到的最大径向力为3 408 N，最小径向力为10 N，在活塞的进油槽中，最大径向力为15 675 N，最小径向力为73 N。结果表明，活塞在径向不平衡力作用下形成的“卡死”现象是活塞发生横裂的主要原因。提出延长活塞的渗碳时间、增加活塞在热处理过程中的校直次数等热处理新工艺，经耐久实验验证，所提工艺可以有效解决活塞横裂问题。

关 键 词：液压破碎锤；活塞；横裂；卡死；径向不平衡力
中图分类号：TH137.9; TP391.9
文献标志码：A **文章编号**：1001-5965(2024)10-2995-10

液压破碎锤属于一种新型高效的液压冲击振动机具，是挖掘机或装载机的配套工作部件，工作时借助工作母机输出的高压油实现破碎器活塞在缸体中的高频往复运动，从而打击钎杆，对外做功，主要应用于矿山开采、冶金工业、市政工程、建筑施工、公路铁路[1-2]等诸多领域。

根据工作原理的不同，破碎锤分为全液压式、气液联合式和氮爆式3类[3-4]。目前，市场上应用最多的是气液联合式破碎锤。活塞是破碎锤的关键零件[5]，在工作中，活塞相对缸体做往复连续的高速运动，运动往复间存在剧烈的摩擦，下端面又要频繁撞击钎杆，承受冲击应力。因此，活塞性能的好坏直接决定破碎锤的可靠性和寿命[6]。经市场使用统计发现，活塞故障产生最多的是横裂，最终“卡死”在缸筒内无法运动，造成了重大经济损失。

针对活塞开裂问题，国内外学者进行了相关的研究。Sokolski等[7]采用有限元方法分析了不同冲击速度下钎杆的冲击区应力大小；Bakhshandi等[8]重点分析了2种不同钢种的圆柱形冲击活塞，表明2种活塞的失效形式都始于冲击表面的退化，产生的径向和环向裂纹最终造成活塞断裂；Andersson等[9]采用一维流体与三维结构模型联合仿真分析的方法，模拟了活塞与衬套之间的油膜形成机制，表明活塞在运动过程中受到径向不平衡力；王加林等[10]采用ABAQUS软件对活塞冲击过程中的应力状态进行了研究；连萌[1]分析了不同冲击角度下钎杆所受到的应力；张涛[11]采用金相检测、电镜检测等方法对柴油发动机活塞的开裂现象进行了分析，

收稿日期：2024-03-07；录用日期：2024-05-11；网络出版时间：2024-06-06 10:56
网络出版地址：link.cnki.net/urlid/11.2625.V.20240605.1822.002
基金项目：山东省重点研发计划(2018GGX104007)；流体动力与机电系统国家重点实验室开放基金课题(GZKF-202016)；河南省科技攻关计划(202102210081)
*通信作者. E-mail：wanyinan@163.com

引用格式：童桂英，万怡男，张磊，等. 液压破碎锤活塞横裂机理分析与工艺优化[J]. 北京航空航天大学学报，2024，50(10)：2995-3004.
TONG G Y，WAN Y N，ZHANG L，et al. Mechanism analysis and process optimization of transverse cracking of hydraulic crushing hammer piston[J]. Journal of Beijing University of Aeronautics and Astronautics，2024，50(10)：2995-3004 (in Chinese).

并改进了一种有效的热处理方案；王中伟[12]采用Workbench平台分析了破碎锤5种极限工况下的应力分布。可见，目前针对破碎锤钎杆和活塞的受力研究较多，尚未对破碎锤活塞开裂的原因作出细致的研究。

本文首先采用硬度计和金相组织显微镜对活塞材料的成分、热处理硬度和金相组织进行了分析，可知，发生活塞横裂故障部位的材料成分和金相组织符合设计要求，而硬度和硬化层深度低于设计要求，从而造成其耐磨性下降，强度降低。然后，通过有限元分析软件ANSYS对活塞正常工作情况下和受工作中径向不平衡力的作用造成“卡死”情况下的最大应力的大小和位置进行了分析，可知，活塞在正常工作情况下最大应力为355.11 MPa，不会产生横裂，活塞在“卡死”状态受到的最大应力为1 229.8 MPa，远远超过了材料本身的屈服极限，活塞发生横裂。最后，利用Fluent求解了活塞所受的径向不平衡力，结果表明，工作中活塞在径向不平衡力作用下，造成活塞单侧接触、磨损，若此时表面硬度和硬化层深度低，就会造成磨损加剧，金属黏连，形成活塞“卡死”现象。

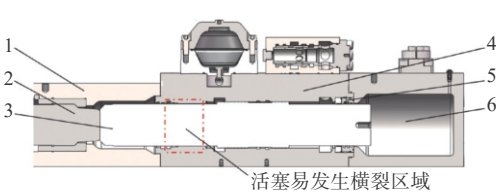
基于破碎锤活塞“卡死”进而造成活塞横裂的原因，本文提出了一种新的活塞热处理制造工艺，用于提高活塞杆耐磨性和圆柱度，防止金属黏连，从而有效避免活塞“卡死”，从根本上解决了活塞发生横裂的问题。

1 横裂活塞材料硬度和组织分析

1.1 活塞缸体间的横裂现象分析

本文破碎锤正常工作压力为28 MPa，后壳中氮气压力为3 MPa，活塞质量为284 kg，活塞的材质为SNCM26V合金钢。热处理工艺为：渗碳+高温回火+淬火+低温回火+深冷+回火+消除应力处理。热处理后要求表面硬度为58～64 HRC，磨削前硬化层厚度为1.6～1.9 mm，心部硬度为40～44 HRC。

液压活塞与缸体组件的三维装配图如图1所示。活塞与缸体配合为间隙配合，活塞与缸体的表面粗糙度和同轴度均有一定的要求。



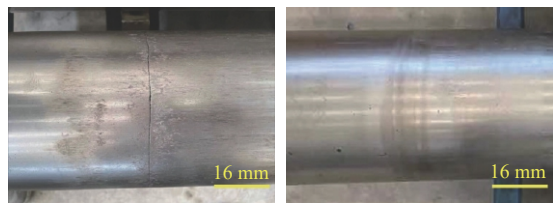
1. 前壳；2. 钎杆；3. 活塞；4. 缸体；5. 油封固定器；6. 后壳。

图1 活塞与缸体组件的装配三维剖视图

Fig. 1 3D sectional view of assembly of piston and block assembly

图2和图3分别为工作时长小于200 h的破碎锤发生活塞横裂图和与活塞配合的缸体内腔图。可见，横裂活塞的裂纹沿圆周径向延展，约占整个圆周的四分之一，裂口较为均匀，无肉眼可见的塑性变形[13-14]，符合脆性[15-16]开裂的特征，活塞表面与配合的缸体内腔表面均有被磨损、拉伤的痕迹[17]。如图4所示，在拆卸返回的缸体时，发现大部分活塞抱死在缸体中，活塞的拆卸比较困难，密封圈出现拉丝、变形，已经失效。

经统计，故障活塞发生横裂的位置大部分距离打击头部300 mm左右。根据断口情况，先从材料和热处理方面进行原因分析。



(a) 横裂圆周面 (b) 横裂圆周背面

图2 横裂活塞图

Fig. 2 Transverse cracking piston diagram

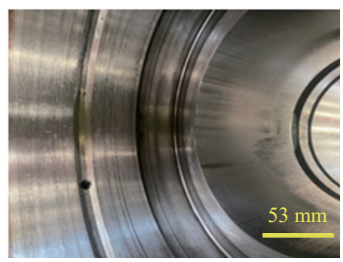


图3 拉伤缸体内腔图

Fig. 3 Inner cavity diagram of the strained cylinder block

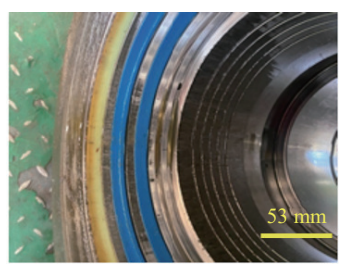


图4 密封圈失效图

Fig. 4 Seal ring failure diagram

1.2 活塞材料成分分析

为确定活塞发生横裂是否由材料成分不当造成强度不够引起，使用金属材质光谱仪分别对活塞正常区域和横裂区域进行金属材质光谱分析[18-20]，如表1～表3所示。结果表明，2个样本区材料中的微量元素均在要求的范围内，且材料中的杂质元素也未超标。由此可见，活塞的横裂不是由材料成

表 1　活塞合格金属材质元素质量分数范围

Table 1　Piston qualified metal material element mass fraction range

元素	质量分数
Ni	0.028～0.033
Si	0.001 5～0.003
Mn	0.008～0.012
P	≤0.000 15
S	≤0.000 1
Cr	0.014～0.018
Mo	0.004～0.006

表 2　活塞正常区域金属材质光谱分析

Table 2　Spectral Analysis of metal material spectrum in the normal area of piston

元素	质量分数
Ni	0.032 4
Si	0.001 9
Mn	0.011 6
P	0.000 07
S	0.000 06
Cr	0.017 3
Mo	0.005 87

表 3　活塞横裂区域金属材质光谱分析

Table 3　Piston cross-crack area metal material spectrum analysis

元素	质量分数
Ni	0.031 8
Si	0.002 1
Mn	0.010 8
P	0.000 09
S	0.000 08
Cr	0.017 5
Mo	0.005 84

分引起的。

1.3　硬度梯度分析

检测金属材料硬度时，检测点从表面到心部分多层检测，测得硬度值成递减的趋势，该方法即硬度梯度法。如图 5 所示，本文分别对活塞的正常部分与发生横裂部分的径向端面进行洛氏硬度检测[16]，沿试样圆周均匀取 8 个测试位置进行硬度数据采集，测量从活塞表面到心部的硬度，结果取平均值，如图 6 所示。可以看出，从活塞表面到心部，材料硬度值总体趋势是递减的，材料的表层硬度幅

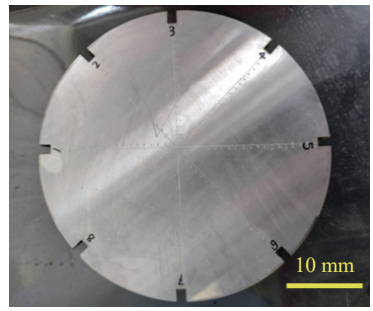


图 5　活塞硬度梯度取样位置示意图

Fig. 5　Schematic diagram of the sampling location of the piston hardness gradient

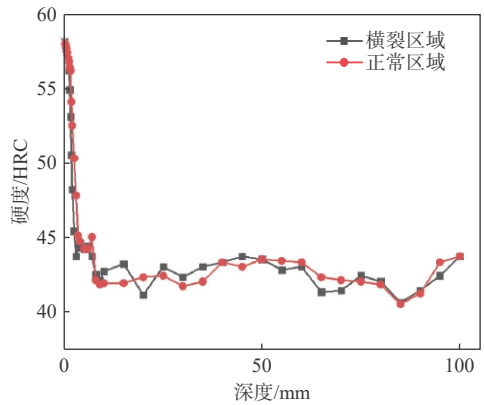


图 6　活塞硬度梯度折线图

Fig. 6　Line diagram of piston hardness gradient

度下降很大，此后材料硬度整体趋于稳定。

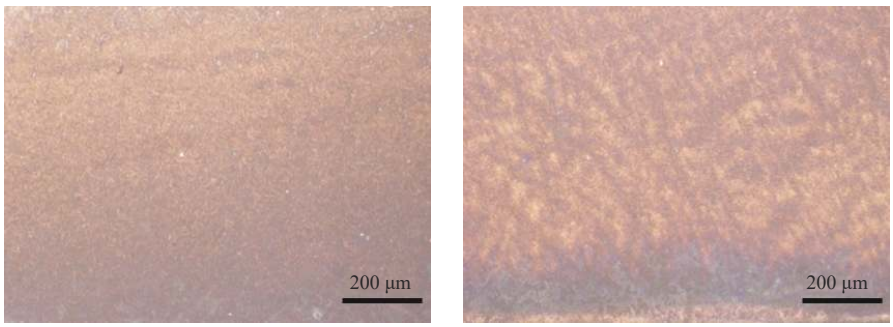
结果显示，横裂区域的活塞表面硬度较正常区域要低，正常区域的硬化层深度在 1.54～1.72 mm 之间，横裂区域的硬化层深度在 0.87～1.62 mm 之间，远远小于图纸要求的硬化层深度，且硬化层不均匀。心部硬度 43.5 HRC 符合要求。

1.4　微观结构分析

对正常活塞和横裂活塞进行金相组织试样分析，将活塞切片腐蚀后，在金相组织显微镜下放大 50 倍后观察组织的形貌[16-17]。如图 7 所示，横裂位置处的活塞金相组织外围有明显的白亮层，这是活塞拉伤导致的组织异常。硬化层的厚度相对较薄，裂口为明显的脆性断裂特征。

将试样在金相组织显微镜下放大 500 倍后观察组织形貌，如图 8 所示。横裂活塞内部组织与正常活塞相同，马氏体与残余奥氏体等级均为 1 级；横裂活塞的碳化物等级为 2 级，正常活塞碳化物等级为 1 级，符合设计要求。

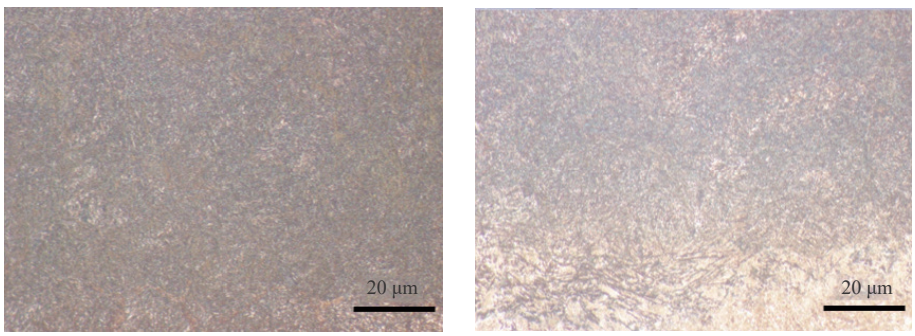
由以上分析可知，发生横裂活塞故障部位的材料成分和金相组织符合设计要求，而硬度和硬化层深度低于设计要求，从而造成其耐磨性下降，强度降低。



(a) 正常活塞50倍放大后金相组织 (b) 横裂活塞50倍放大后金相组织

图 7 50 倍放大后活塞金相组织图

Fig. 7 Piston metallographic structure diagram of 50 times piston



(a) 正常活塞500倍放大后金相组织形貌 (b) 横裂活塞500倍放大后金相组织形貌

图 8 500 倍放大后活塞金相组织形貌图

Fig. 8 Metallographic structure and morphology of 500 times piston

2 破碎锤活塞力学分析

零件上的受力大小是造成零件破坏的重要因素。本节从研究破碎锤的工作机理出发，对活塞正常工作情况下的受力情况进行分析，计算活塞在工作中的最大冲击力，并将该冲击力输入 ANSYS 软件进行静力学模拟求解，研究不同工作状态下活塞应力集中的位置，进一步确定活塞发生横裂的原因。

2.1 活塞受力分析

破碎锤工作分为 4 个阶段[21]：回程阶段、换向阶段、冲程阶段和打击停顿阶段。其中，换向阶段的油液压力处于回程阶段与冲程阶段之间，与活塞横裂关系不大，因此，本文不做分析。

如图 9(a)所示，回程阶段活塞受到的合力为

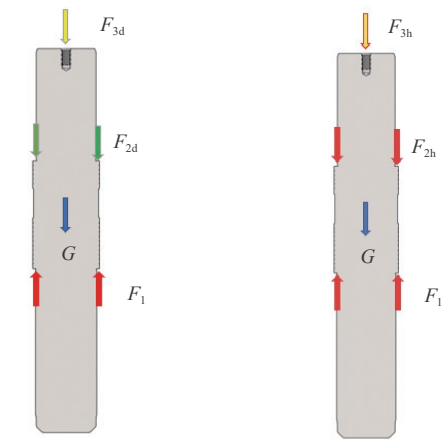
$$F_{\mathrm{d}} = F_1 + F_{2\mathrm{d}} + F_{3\mathrm{d}} + G \tag{1}$$

式中：F_{d}为活塞回程阶段受到的合力，方向向上，N；F_1为活塞前腔受到的高压油作用力，N；$F_{2\mathrm{d}}$为活塞后腔受到的低压油作用力，N；$F_{3\mathrm{d}}$为活塞氮气室受到的低气压作用力，N；G为活塞自身重力，N。

如图 9(b)所示，冲程阶段活塞受到的合力为

$$F_{\mathrm{h}} = F_1 + F_{2\mathrm{h}} + F_{3\mathrm{h}} + G \tag{2}$$

式中：F_{h}为活塞冲程阶段受到的合力，方向向下，N；$F_{2\mathrm{h}}$为活塞后腔受到的高压油作用力，N；$F_{3\mathrm{h}}$为活塞氮气室受到的高气压作用力，N。



(a) 活塞回程受力 (b) 活塞冲程受力

图 9 活塞工作受力示意图

Fig. 9 Schematic diagram of piston working force

打击停顿阶段，活塞处于最下端位置，换向阀还没有完全换向，此时活塞受到后壳氮气的低气压作用力$F_{3\mathrm{d}}$，活塞前腔一直受高压油作用力F_1，活塞后腔受高压油作用力$F_{2\mathrm{h}}$：

$$F_1 = pA_3 \tag{3}$$

式中：p为系统的工作压力，Pa；A_3为活塞前腔的作用面积，m^2。

$$F_{2\mathrm{h}} = pA_4 \tag{4}$$

式中：A_4为活塞后腔的作用面积, m^2。

$$F_{3d} = p_1 A_5 \tag{5}$$

式中：p_1为后壳氮气压力, Pa; A_5为活塞后端面积, m^2。

另外, 活塞打击钎杆时受到巨大的瞬时冲击反力, 由此引起的破碎锤的最大打击力F_m为

$$F_m = \sigma A_1 \tag{6}$$

式中：F_m为破碎锤打击钎杆时的最大打击力, N; σ为打击钎杆时活塞受到的正应力, MPa; A_1为活塞打击头的面积, m^2。

打击钎杆时活塞受到的正应力为

$$\sigma = \frac{E}{1 + A_1/A_2} \cdot \frac{v_m}{c} \tag{7}$$

式中：E为活塞材料的弹性模量, N/m^2; v_m为活塞冲程时最大的打击速度, m/s; A_2为钎杆后尾的作用面积, m^2; c为在活塞材料中的纵波速度, m/s。

活塞材料中的纵波速度为

$$c = \sqrt{E/\rho} \tag{8}$$

式中：ρ为活塞材料的密度, kg/m^3。

对氮气而言, 绝热指数 k=1.4, 破碎锤的结构参数与性能参数如表 4 所示, 表中直径尺寸如图 10 所示。

表 4　破碎锤参数

Table 4　Crushing hammer parameters

参数	数值
活塞上端直径D_2/mm	185
活塞中部直径D/mm	208
活塞下端直径D_1/mm	190
钎杆尾部直径D_3/mm	125
破碎锤工作压力P_1/MPa	28
氮气室的初始容积V/L	24
氮气室的初始压力P/MPa	3.0
活塞质量M_p/kg	284
活塞材料弹性模量E/(N·m^2)	2.12×10^{11}
活塞材料密度ρ/(kg·m^{-3})	7 850
活塞打击头直径D_4/mm	146

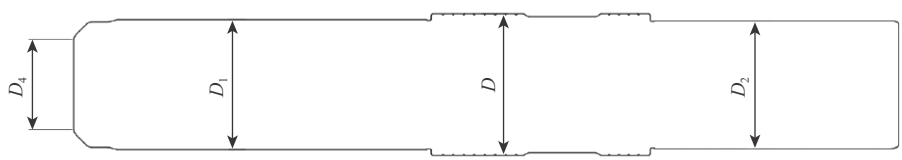


图 10　活塞直径尺寸

Fig. 10　Dimensional of piston diameter

经计算, 活塞各腔受力如下。

1) 前腔受力为

$$F_1 = 28\pi\left(\frac{D^2}{4} - \frac{D_1^2}{4}\right) = 28\pi\left(104^2 - 95^2\right) = 157\,544 \tag{9}$$

2) 后腔受力为

$$F_{2h} = 28\pi\left(\frac{D^2}{4} - \frac{D_2^2}{4}\right) = 28\pi\left(104^2 - 92.5^2\right) = 198\,777 \tag{10}$$

3) 后壳氮气力为

$$F_{3d} = \frac{3\pi D_2^2}{4} = 3\times92.5^2\pi = 80\,640 \tag{11}$$

根据现场高速摄影机计算得到活塞的末端速度为 13 m/s。

4) 活塞受到的正应力为

$$\sigma = \frac{2.12\times10^{11}}{1 + \dfrac{D_4^2}{4}\Big/\dfrac{D_3^2}{4}} \cdot \frac{13}{\sqrt{\dfrac{2.12\times10^{11}}{7.85\times10^3}}} = 220 \tag{12}$$

5) 活塞的瞬时冲击力为

$$F_m = 220\pi\times73^2 = 3\,680\,000 \tag{13}$$

2.2　活塞 ANSYS 静力学模拟

为确定活塞打击钎杆时受到最大应力的位置, 利用有限元软件 ANSYS 进行静力学模拟求解, 采用 SPAR。设置仿真边界条件, 活塞打击钎杆时处于瞬时停顿状态, 对活塞的后端进行固定约束; 按照图 9 所示活塞受力情况对活塞施加载荷约束, 分别添加外力F_1、F_{2h}、F_{3d}、自身重力 G 和活塞的瞬时冲击力F_m。利用网格结构划分出活塞的具体结构, 在局部处, 如退刀槽、倒圆(倒角)进行了网格加密, 再进行网格无关性验证。如表 5 所示, 随着网格节点数增加到 4 600 000 个之后, 所得计算结果已相对稳定。

表 5　网格无关性验证结果

Table 5　Grid independence verification results

网格单元数量	应力值/MPa
2 893 229	457.62
3 170 064	514.68
3 674 387	510.92
4 460 354	356.35
4 699 218	364.33
4 976 105	355.11

因此，本文将在网格单元数为4 976 105个、节点数5 547 446个的模型下进行计算分析，如图11所示。应力分析结果如图12所示，图中显示活塞最大应力位置在活塞退刀槽内，最大应力为355.11 MPa，活塞横裂位置的应力约130 MPa，均小于活塞材料的屈服极限。实际工作时，活塞在缸体内处于自由状态，可以沿轴向运动，因此，活塞受到的应力会更小。在正常工作下，活塞不会发生过载开裂情况。

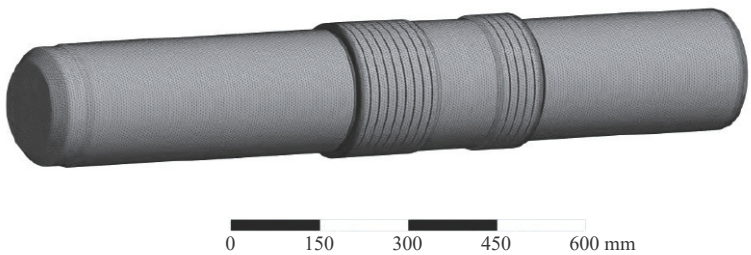


图11 活塞有限元模型

Fig. 11 Finite element model of piston

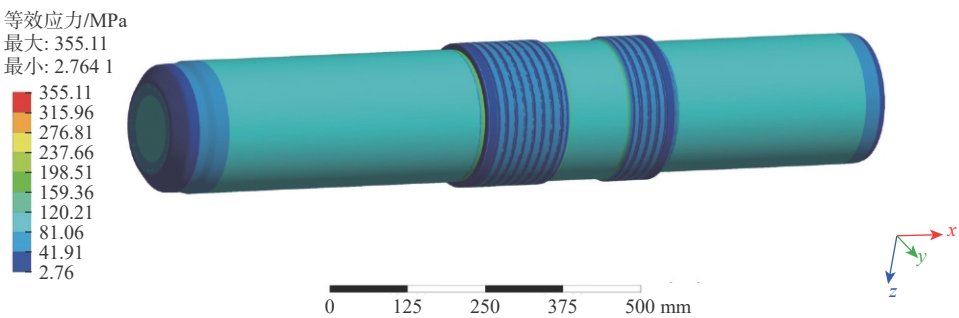


图12 活塞在正常受力下应力云图

Fig. 12 Stress cloud map of piston under normal stress

2.3 Fluent径向力模拟

由图1和图2可知，发生横裂的活塞与缸体内腔表面发生了不对称的局部磨损、拉伤痕迹。破碎锤在工作过程中，进油口与主阀相通，有2个进油孔位于缸体前方单侧，进油压力较大，回油孔位于缸体两侧，回油压力较小。此时，活塞在缸体内腔中受到的油液压力可能存在分布不均匀的问题，导致活塞在高速往复运动中受到径向不平衡力[2,22]作用，增加了活塞单方向磨损、拉伤的可能性。因此，对拉伤位置附近的流场进行Fluent流体仿真[2,23]，求解流场内径向力的大小。

图13为缸体与活塞前端的剖面图。因为活塞前端的横裂率高，所以对这部分的流场进行分析。图14为回油槽与进油槽的局部放大图。

利用Fluent软件进行流体仿真，首先，对回油槽和进油槽进行油道抽取，如图15所示。回油槽有2个回油孔，进油槽有2个进油孔。然后，对流道模型用六面体网格进行划分，划分后的有限元模型如图16所示。最后，将有限元模型导入Fluent软件，采用k-epsilon湍流模型进行求解。

活塞打击钎杆的瞬时状态是冲击力最大的，因此，对打击的瞬时状态进行分析。活塞打击钎杆时，进油槽中进油口压力为50 MPa，前腔压力为30 MPa；回油槽中进油口压力为30 MPa，回油背压为3 MPa，工作介质为46#抗磨液压油，密度为

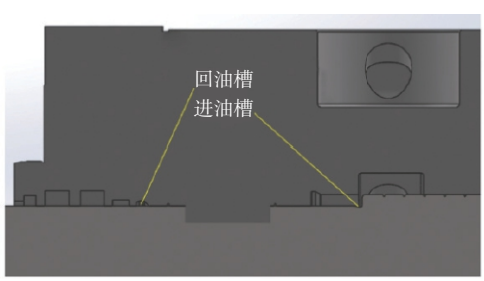
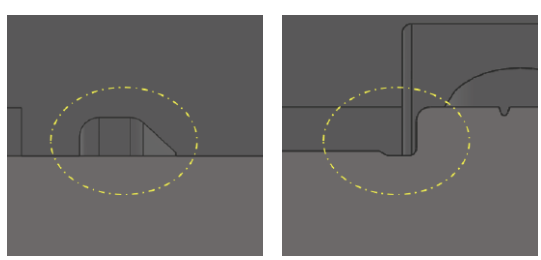


图13 缸体与活塞前端的剖面图

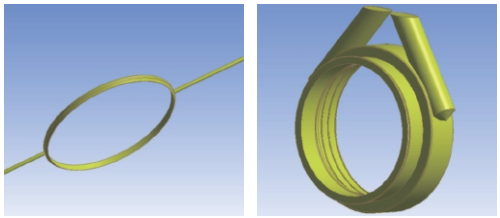
Fig. 13 Front sectional view of cylinder block and piston



(a) 回油槽局部放大图 (b) 进油槽局部放大图

图14 回油槽和进油槽局部放大图

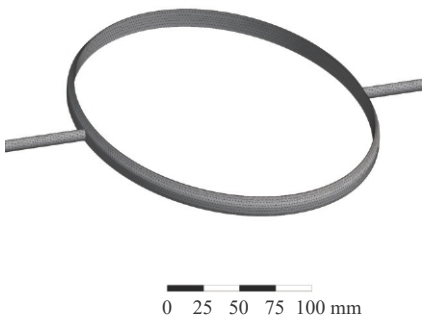
Fig. 14 Partial enlarged view of oil return groove and oil inlet groove



(a) 回油槽油道抽取模型　(b) 进油槽油道抽取模型

图 15　油道抽取模型

Fig. 15　Oil passage extraction model



(a) 回油槽油道有限元模型



(b) 进油槽油道有限元模型

图 16　油道有限元模型

Fig. 16　Finite element model of oil passage

850 kg/m^3, 运动黏度为 46 mm^2/s。

通过仿真求解可知, 活塞受到了液压不平衡压力的作用, 结果如图 17 所示。回油槽中活塞受到的最大压力为 2.9 MPa, 靠近回油孔的位置压力最小, 为 1.42 MPa; 进油槽中活塞受到的最大压力位置在进油口的对面, 最大压力为 46 MPa, 最小压力为 36.3 MPa。

采用 Fluent 软件中的 ForceReport 命令计算油道壁面所受的力, 回油槽中受到的最大径向力为 3 408 N, 最小径向力为 10 N, 进油槽中受到的最大径向力为 15 675 N, 最小径向力为 73 N。

通过以上仿真分析可知, 活塞在缸体内腔中受到了径向不平衡油液压力和力的作用, 造成活塞与缸体配合表面单面机械接触, 发生磨损、拉伤, 一旦磨损部位存在硬度低、硬化层小的情况, 磨损就会加重, 有可能造成 2 个表面发生金属黏连, 使活塞“卡死”在缸腔内。

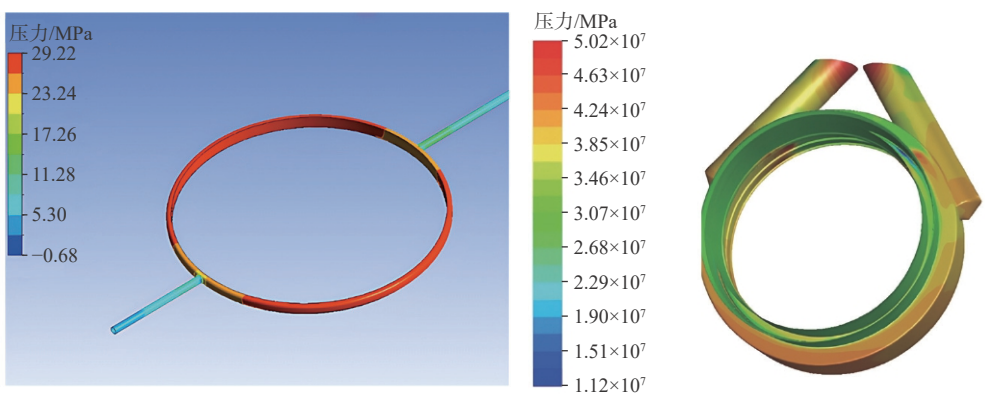
2.4　活塞“卡死”状态下的受力分析

分析得知, 活塞在工作中, 因受到径向不平衡力作用, 有可能造成活塞“卡死”在缸体内。本节对“卡死”后的活塞进行静力学分析。

对“卡死”位置设置固定约束, 活塞“卡死”时轴心线偏斜, 瞬时冲击力不再垂直[1], 根据现场实际情况设置反冲力作用方向偏斜 2°, 其余有限元分析模型和加载条件与正常工作情况一致。

应力分析结果如图 18 所示, 应力最大位置在活塞“卡死”位置, 最大应力为 1 229.8 MPa, 远远超过了材料本身的屈服极限 850 MPa, “卡死”位置另一侧最大应力为 480 MPa, 小于材料的屈服极限。结果表明, 在偏载作用下, 活塞“卡死”位置会发生横裂, 不会造成整个圆周完全横裂, 与图 1 观察现象一致。

从以上分析可知, 活塞不“卡死”在缸体内是不会造成横裂的, 而“卡死”的原因是活塞局部硬度低, 抗磨性下降, 在不平衡的径向力作用下造成了活塞外圆和缸体孔壁表面的黏连。因此, 解决横裂的措施是优化活塞现有工艺, 保证其不“卡死”在



(a) 回油槽压力云图　(b) 进油槽压力云图

图 17　回油槽和进油槽压力云图

Fig. 17　Pressure cloud chart of oil return groove and oil inlet groove

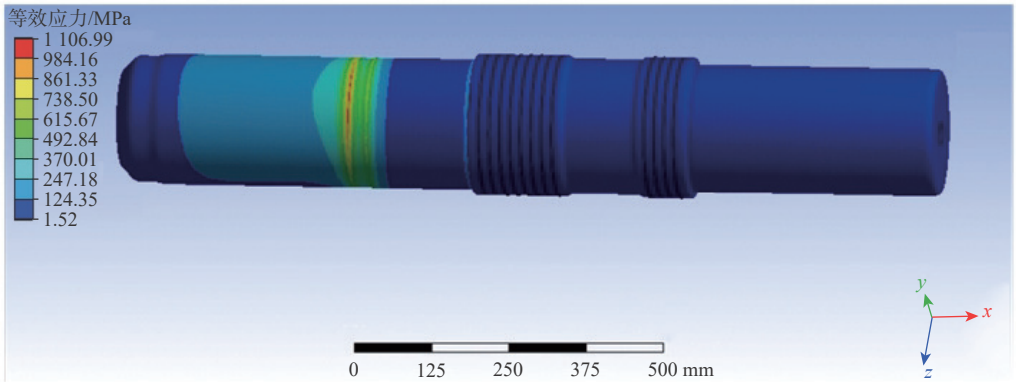


图 18 活塞卡死下的应力云图

Fig. 18 Stress cloud map under piston jamming

缸体内。

2.5 工艺优化

活塞外圆表面与缸体内孔表面属于小间隙配合，在径向不平衡力作用下，两者容易产生单边机械摩擦，进而“卡死”。解决活塞“卡死”在缸体内目前有 2 个措施：①提高活塞表面的耐磨性，提高表面圆柱度；②增大两者配合间隙。增大配合间隙直接影响泄漏量，进而影响活塞的打击力，因此不能太大。为此，提出将渗碳时间延长 2 h，提高活塞表面的硬化层深度。同时，增加活塞在热处理过程中的校直次数，由原来的 3 次增加到 5 次。一方面，可以保证活塞精磨后硬化层深度的均匀性；另一方面，可以提高磨后活塞外圆的圆柱度和线性度。

3 实验验证

对 3 种不同机型的活塞分别取 5 支做工艺优化，再进行现场打击耐久测试。因为活塞发生横裂现象绝大部分工作时长小于 200 h，甚至小于 100 h，所以，本次耐久实验采用每天不间断打击 10 h，连续工作 300 h，分别对这些活塞做测试记录。

经过 300 h 的打击测试，均未发现“卡锤”现象，破碎锤均可正常工作，随机对其中 3 支破碎锤进行拆卸，拆下的活塞如图 19 所示，易横裂的区域沿圆周有磨损痕迹，但未发现拉伤和裂纹源。实验结果与仿真结果一致。

对 3 支活塞做洛氏硬度实验，实验方法与图 5 的实验一致，结果如图 20 所示。结果显示，3 支活塞的硬化层厚度在 1.58～1.83 mm 之间，且硬化层较为均匀，圆周硬度差不超过 1 HRC，硬度梯度符合使用要求，磨损位置硬度有轻微降低，在使用要求之内。结果证明，工艺优化可以有效解决活塞的“卡死”现象，大大提高了活塞的使用寿命。

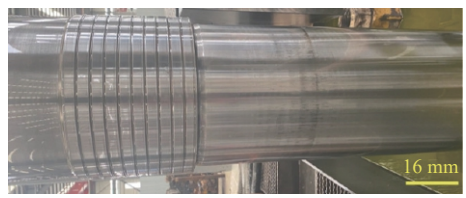


图 19 热处理优化测试活塞

Fig. 19 Heat treatment optimization test piston

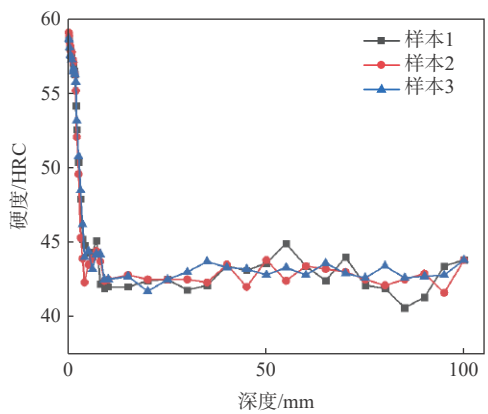


图 20 测试活塞硬度梯度

Fig. 20 Test piston hardness gradient

4 结 论

1）利用硬度计和金相组织显微镜对活塞材料的成分、热处理硬度和金相组织进行分析，得出发生横裂的活塞的材料成分与金相组织均合格，该因素不是造成活塞横裂的原因。

2）运用有限元分析软件 ANSYS 对活塞最大应力的大小和位置进行分析，得出在正常工作下，活塞的最大应力（应力为 355.11 MPa）在退刀槽处，并不是活塞横裂处（应力为 130 MPa 左右），活塞受到的力小于材料的屈服极限，不会发生横裂现象。

3）通过 Fluent 软件分析了活塞在缸体内腔受到的径向不平衡力大小，得出若活塞“卡死”在缸体内，其受到的最大应力为 1 229.8 MPa，超过了材

料的屈服极限, 会产生活塞横裂。

4) 经数据分析得出, 造成活塞“卡死”的主要原因是其在工作中受到不平衡液压力的作用, 且局部材料硬度低, 热处理硬化层不均匀, 磨损加大产生金属黏连所致。

5) 经耐久实验验证优化活塞工艺, 提高渗碳时间, 增加活塞热处理过程中的校直次数, 能够提高活塞表面的耐磨性、硬化层深度均匀性和圆柱度, 同时提高活塞的使用寿命。

参考文献（References）

[1] 连萌. 液压挖掘机破碎锤钎杆的冲击特性分析[J]. 黄河水利职业技术学院学报, 2017, 29(1): 46-50.
LIAN M. The impact characteristics analysis of hydraulic excavator broken hammer drill rod[J]. Journal of Yellow River Conservancy Technical Institute, 2017, 29(1): 46-50(in Chinese).

[2] 周广星. 液压破碎锤回油孔处结构对卡紧力影响研究[J]. 液压气动与密封, 2021, 41(5): 68-70.
ZHOU G X. Study on the effect of hydraulic crushing hammer on clamping force[J]. Hydraulics Pneumatics & Seals, 2021, 41(5): 68-70 (in Chinese).

[3] 曹傲. 液压凿岩机的参数优化研究[D]. 西安: 长安大学, 2020.
CAO A. Research on parameter optimization of hydraulic rock drill[D]. Xi'an: Changan University, 2020(in Chinese).

[4] 罗亮光. 内回转气动凿岩机活塞运动设计计算的新理念与实践[J]. 凿岩机械气动工具, 2021, 47(1): 6-15.
LUO L G. New concepts and practices on design computation for movements of pistons in interanlly rotary pneumatic rock drills[J]. Rock Drilling Machinery & Pneumatic Tools, 2021, 47(1): 6-15(in Chinese).

[5] 高波, 徐雪锋, 夏剑辉. 液压凿岩机活塞应力分析[J]. 凿岩机械气动工具, 2020, 46(4): 15-20.
GAO B, XU X F, XIA J H. Stress analysis of piston of hydraulic rock drill[J]. Rock Drilling Machinery & Pneumatic Tools, 2020, 46(4): 15-20(in Chinese).

[6] SOKOLSKI M, SOKOLSKI P. Problems of strength estimation of the vulnerable zones in the tools of hydraulic hammers for mining[C]//Proceedings of the International Conference on Computer Aided Engineering. Berlin: Springer, 2019: 696-704.

[7] SOKOLSKI M, SOKOLSKI P. Strength estimation of the impact zone–A critical area of the tools of the hydraulic hammers[J]. Archives of Civil and Mechanical Engineering, 2016, 16(4): 767-776.

[8] BAKHSHANDI R K, TKACHUK A, SADEK M, et al. Failure analysis of two cylindrical impact pistons subjected to high velocity impacts in drilling applications[J]. Engineering Failure Analysis, 2022, 140: 106623.

[9] ANDERSSON H, HOLMBERG L J, SIMONSSON K, et al. Simulation of leakage flow through dynamic sealing gaps in hydraulic percussion units using a co-simulation approach[J]. Simulation Modelling Practice and Theory, 2021, 111: 102351.

[10] WANG J L, HAN B, WANG C, et al. Failure analysis of the piston used in a pneumatic down the hole impactor[J]. Engineering Failure Analysis, 2021, 127: 105561.

[11] 张涛. 某高功率柴油机活塞失效分析及解决方案[J]. 内燃机与配件, 2021(1): 23-25.
ZHANG T. Analysis and solution on fracture failure of the piston used in a high power diesel engine[J]. Internal Combustion Engine & Parts, 2021(1): 23-25(in Chinese).

[12] 王中伟. 基于 Workbench 的液压碎石车工作装置有限元分析[J]. 凿岩机械气动工具, 2015(2): 18-24.
WANG Z W. Finite element analysis of hydraulic rock breaker working device based on workbench[J]. Rock Drilling Machinery & Pneumatic Tools, 2015(2): 18-24(in Chinese).

[13] HAO S R, SHEN Y X, CHENG J B. Phase field formulation for the fracture of a metal under impact with a fluid formulation[J]. Engineering Fracture Mechanics, 2022, 261: 108142.

[14] LANG L, ZHU Z M, DENG S, et al. Study on the arresting mechanism of two arrest-holes on moving crack in brittle material under impacts[J]. Engineering Fracture Mechanics, 2020, 229: 106936.

[15] HÜTTER G, ZYBELL L, KUNA M. Micromechanisms of fracture in nodular cast iron: From experimental findings towards modeling strategies–A review[J]. Engineering Fracture Mechanics, 2015, 144: 118-141.

[16] 陈富强, 陈晓伟. 球墨铸铁曲轴断裂失效分析[J]. 金属热处理, 2011, 36(12): 124-126.
CHEN F Q, CHEN X W. Fracture failure investigation on nodular cast iron crankshaft[J]. Heat Treatment of Metals, 2011, 36(12): 124-126(in Chinese).

[17] 闫佳兵. 船用柴油机活塞损坏原因分析[J]. 内燃机与配件, 2020(7): 163-164.
YAN J B. Cause analysis of piston damage of marine diesel engine [J]. Internal Combustion Engine & Parts, 2020(7): 163-164(in Chinese).

[18] 王宏伟, 李君, 张峥, 等. 压缩机活塞杆断裂原因分析[J]. 材料工程, 2009, 37(12): 5-9.
WANG H W, LI J, ZHANG Z, et al. Failure analysis on fracture of compressor piston rod[J]. Journal of Materials Engineering, 2009, 37(12): 5-9(in Chinese).

[19] XU L Q, YU X K, HOU Z G. Analysis of the causes of driving gear shaft fractures in gear pumps[J]. Journal of Failure Analysis and Prevention, 2020, 20(1): 242-248.

[20] 中华人民共和国国家质量监督检验检疫总局, 中国国家标准化管理委员会. 钢件渗碳淬火回火金相检验: GB/T 25744—2010[S]. 北京: 中国标准出版社, 2011.
General Administration of Quality Supervision, Inspection and Quarantine of the People's Republic of China, Standardization Administration of the People's Republic of China. Metallographic examination for carburizing quenching and tempering of steel parts: GB/T 25744—2010[S]. Beijing: Standards Press of China, 2011(in Chinese).

[21] 张明松, 苏志成, 王宏. 液压破碎锤动态特性的研究[J]. 南方农机, 2021, 52(3): 106-107.
ZHANG M S, SU Z C, WANG H. Research on dynamic characteristics of hydraulic breaker[J]. China Southern Agricultural Machinery, 2021, 52(3): 106-107(in Chinese).

[22] 杨文军, 陈新元, 邓江洪. 间隙密封液压缸的活塞卡紧力分析[J]. 液压与气动, 2012(9): 107-109.
YANG W J, CHEN X Y, DENG J H. Clamping force analysis of

hydraulic cylinder with piston sealing[J]. Chinese Hydraulics & Pneumatics, 2012(9): 107-109(in Chinese).

[23] 朱德, 刘杰, 肖欣谕, 等. 基于 Fluent 的液压密封流体膜压力精确化研究[J]. 机床与液压, 2021, 49(10): 122-129.
ZHU D, LIU J, XIAO X Y, et al. Precise study on fluid film pressure of hydraulic seal based on fluent[J]. Machine Tool & Hydraulics, 2021, 49(10): 122-129(in Chinese).

Mechanism analysis and process optimization of transverse cracking of hydraulic crushing hammer piston

TONG Guiying[1], WAN Yinan[2,*], ZHANG Lei[3], SHI Yan[4], ZHANG Yeming[5]

(1. School of Electromechanical and Automotive Engineering, Yantai University, Yantai 264005, China;
2. School of Mechatronic Engineering, Yantai Institute of Technology, Yantai 264005, China;
3. Yantai Aidi Precision Machinery Co., Ltd., Yantai 264000, China;
4. School of Automation Science and Electrical Engineering, Beihang University, Beijing 100191, China;
5. School of Mechanical and Power Engineering, Henan Polytechnic University, Jiaozuo 454000, China)

Abstract: In view of the phenomenon of piston transverse crack in crushing hammer, the paper analyzes the transverse crack of piston, and puts forward a new heat treatment process of piston. The piston material was analyzed by using a hardness tester and a metallographic microscope. It is found that the material composition and metallographic structure at the location of the piston transverse crack failure are in line with the design requirements, while the hardness and hardened layer depth are lower than the design. Using Ansys, the maximum stress of the piston is 1 229.8 MPa, which exceeds the yield limit of the material by 850 MPa, and the maximum stress is located at the transverse crack of the piston. Fluent was used to solve the radial unbalance force on the piston. The maximum radial force and the minimum radial force in the oil return groove of the piston were 3 408 N and 10 N. In the oil inlet groove of the piston, the maximum radial force is 15 675 N and the minimum radial force is 73 N. The results show that the "stuck" phenomenon formed by the radial unbalanced force is the main reason for the transverse crack of the piston. A new heat treatment process such as extending the carburizing time of piston and increasing the number of times of piston straightening during heat treatment is proposed. The durability test shows that the proposed process can effectively solve the problem of piston transverse cracking.

Keywords: hydraulic crushing hammer; pistons; transverse cracking; jamming; radial unbalance force

Received: 2024-03-07; **Accepted**: 2024-05-11; **Published Online**: 2024-06-06 10:56

URL: link.cnki.net/urlid/11.2625.V.20240605.1822.002

Foundation items: Key research and development program of Shandong Province (2018GGX104007); Open Fund of State Key Laboratory of Fluid Power and Mechatronic Systems (GZKF-202016); Henan Province Science and Technology Research Program (202102210081)

* **Corresponding author**. E-mail: wanyinan@163.com

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2022.0762

区间划分改进相似性的发动机剩余寿命预测

赵洪利*，白令德
(中国民航大学 航空工程学院，天津 300300)

摘 要：针对传统相似性匹配方法易引入伪相似发动机导致预测精度不高的问题，提出一种结合发动机性能退化特点与相似性匹配特点的区间划分改进相似性剩余寿命(RUL)预测方法。利用堆栈自编码器基于选择的参数构建健康指数；根据测试发动机的已知运行循环进行区间划分，利用传统相似性匹配方法进行初步筛选；对于不同区间下的测试发动机，分别使用不确定性修正和退化一致性检验去除初选参考发动机中的异常发动机，得到最终的剩余寿命预测结果。利用NASA的C-MAPSS数据集进行验证，结果表明：所提方法的预测精度比当前相似性匹配方法提升了34%，证明了方法的有效性。

关 键 词：航空发动机；寿命预测；区间划分改进相似性；退化一致性检验；不确定性修正
中图分类号：V233
文献标志码：A **文章编号**：1001-5965(2024)10-3005-08

航空发动机内部结构高度复杂且外部工作环境恶劣，作为飞机的“心脏”，一旦出现意外，将会导致巨大的事故。《中国民用航空发展第十三个五年规划》中，将安全性水平放在首位，降低事故发生率成为首要目标[1]。在发动机维修领域，由于预测与健康管理(prognostics and health management, PHM)能够基于发动机的实时监测数据提高运行的可靠性和安全性，受到了广泛关注，剩余寿命(remaining useful life, RUL)预测是PHM中最重要的部分，利用发动机的各项信息提前预知可能失效的时间，并制订相应的维护计划，提供维修决策支持，能够提高发动机的安全水平，降低相应的维修成本[2-3]。

航空发动机的RUL预测大致可分为物理模型、数据驱动和二者混合的方法[4-5]。其中，数据驱动方法具有泛用性好、预测结果准确的优点，受到了广大研究人员的青睐。Chen等[6]利用Lasso回归选择变量并进行窗口化处理，将每个窗口中的均值和趋势特征输入随机森林模型预测发动机的RUL，得到了较高的预测精度。车畅畅等[7]将多状态参数与性能退化量相结合，代入一维卷积神经网络中进行训练以提取退化特征，并使用双向长短期记忆网络进行RUL预测。郭庆等[8]基于发动机的性能退化特点，提出指数形式的Wiener过程预测航空发动机的RUL。Zhang等[9]建立多层长短期记忆网络，逐层提取时间特征，并结合注意力机制预测旋转机械的RUL，解决了长时间序列的信息丢失问题。

另外，相似性匹配方法由于操作简单、结果准确，被广泛用于RUL预测，该方法分析并计算测试样本与训练样本之间的相似度，从样本库中筛选出退化趋势最相近的样本以进行预测[10]。张妍等[11]提取出数据的低维正交特征进行时延多模型相似性匹配，并使用核密度函数估计RUL的概率分布，按照密度加权计算RUL。Cai等[12]在相似性匹配过程中引入核双样本检验，结合威布尔分布得到了RUL的概率分布，对预测的不确定性有较强的解释能力。Yu等[13]提出了零中心化的相似性匹配规则，解决了发动机初始健康状态不同及相似样本较

收稿日期：2022-09-07；**录用日期**：2022-12-02；**网络出版时间**：2022-12-15 11:41
网络出版地址：link.cnki.net/urlid/11.2625.V.20221214.1142.003
基金项目：中央高校基本科研业务费专项资金(3122021049)；中国民航大学实验技术创新基金(2021CXJJ90)
* **通信作者**. E-mail：henleytrent@163.com

引用格式：赵洪利，白令德. 区间划分改进相似性的发动机剩余寿命预测[J]. 北京航空航天大学学报，2024，50(10)：3005-3012.
ZHAO H L，BAI L D. Remaining life prediction of engine by improved similarity with interval partition[J]. Journal of Beijing University of Aeronautics and Astronautics，2024，50(10)：3005-3012 (in Chinese).

少的问题。Xue 等[14]在传统相似性匹配方法的基础上引入退化一致性检验，筛选出符合整体与局部退化趋势的发动机，得到了较好的预测结果。

上述研究中，构建发动机的健康指数时没有充分考虑不同参数对健康指数的影响差异，融合时可能会添加部分无关数据信息，另外，使用传统相似性匹配方法对发动机进行 RUL 预测时，容易引入相似程度较高但预测误差较大的伪相似发动机。本文针对以上问题，首先，基于发动机的工作原理筛选与性能退化关系密切的传感器参数，并使用堆栈自编码器 (stacked autoencoders, SAE) 进行特征提取，构建发动机的健康指数；其次，使用区间划分改进相似性方法进行针对性预测，识别并去除初选参考发动机中的伪相似发动机，预测发动机的 RUL；最后，利用 NASA 的 C-MAPSS 数据集验证本文方法的有效性。

1 健康指数的构建

为评估发动机运行周期中的健康状态，需要构建一个综合的指标来表征发动机运行过程中状态的变化规律[15]。近年来，大部分学者使用深度学习方法构建健康指数，但缺少与航空发动机退化特点之间的联系。因此，本文基于发动机的工作原理筛选出与性能退化联系密切的传感器参数，由于不同参数之间存在非线性关系，采用非线性特征提取能力较强的 SAE 提取退化特征，输入实时数据得到最终的健康指数。

1.1 传感器参数的筛选

航空发动机结构复杂，性能退化时涉及多个组件，以高压压气机 (high pressure compressor, HPC) 为例，根据发动机的工作原理，当 HPC 发生性能退化时，其流通能力和效率会降低，为保证发动机的压比不变，要提高燃油的供给，这会导致燃烧室后各站位的温度增加，低压转子转速小幅度增加，也会影响涡轮冷却空气的引气量。将退化趋势不明显的参数去除后，最终筛选的传感器参数包括 HPC 出口温度、低压涡轮出口温度、HPC 出口总压、燃油流量同 HPC 出口静压的比值。

1.2 堆栈自编码器

由于筛选出的多传感器参数之间呈非线性关系，使用 SAE 进行特征提取。SAE 是一种由多个自编码器 (autoencoders, AE) 栈式堆叠构建的深层网络结构[16]。AE 的网络结构如图 1 所示，由输入层、隐藏层、输出层构成。

AE 的训练过程分为编码和解码过程。编码过程将高维输入样本映射到低维抽象表示，解码过程将抽象表示转换为期望输出。训练的目标为降低输入层与输出层之间的重构误差，实现样本的复现。

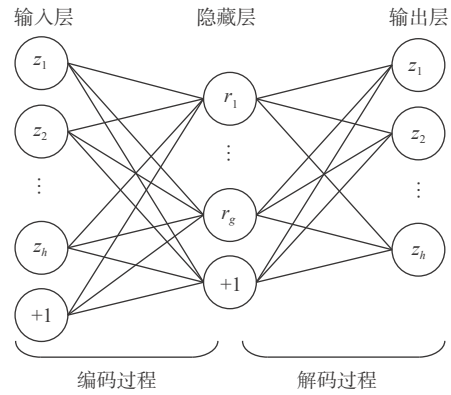


图 1 自编码器结构

Fig. 1 Autoencoder structure

给定输入 $\boldsymbol{Z}=[z_1,z_2,\cdots,z_h]$，$h$为输入的维度，对其进行线性映射并叠加非线性激活函数后，得到隐藏状态 $\boldsymbol{R}$，完成编码过程，如下：

$$\boldsymbol{R}=\boldsymbol{f}_{\mathrm{e}}(\boldsymbol{W}\boldsymbol{Z}+\boldsymbol{b}) \tag{1}$$

式中：$\boldsymbol{f}_{\mathrm{e}}$为编码器的激活函数；$\boldsymbol{W}$和$\boldsymbol{b}$分别为输入层与隐藏层之间的权重和偏置；$\boldsymbol{R}$为隐藏层向量。

解码过程与编码过程类似，将隐藏状态 $\boldsymbol{R}$重构为输出 $\boldsymbol{Z}'$，如下：

$$\boldsymbol{Z}'=\boldsymbol{f}_{\mathrm{d}}(\boldsymbol{W}'\boldsymbol{R}+\boldsymbol{b}') \tag{2}$$

式中：$\boldsymbol{f}_{\mathrm{d}}$为解码器的激活函数；$\boldsymbol{W}'$和$\boldsymbol{b}'$分别为隐藏层与输出层之间的权重和偏置。

AE 使用梯度下降算法反向传播误差，逐步调整网络参数，使损失函数达到最小。当损失函数足够小时，认为隐藏层的向量包含了原始数据的全部信息。损失函数使用均方误差损失函数，表示如下：

$$L_{\mathrm{SAE}}=\frac{1}{h}\sum_{i=1}^{h}(x_i-x_i')^2 \tag{3}$$

SAE 可看作由多个 AE 串联而成，上一级隐藏层的抽象特征作为下一级输入层向量，通过深层次的网络结构，逐层提取退化特征，能够得到更具有代表性的表达特征[17]。图 2 展示了具有 3 个隐藏层结构的 SAE 网络。当原始数据输入输入层后，先由隐藏层 1 进行初步特征提取，并将该特征作为隐藏层 2 的输入向量进行深度特征提取，最终利用隐藏层 3 和输出层进行数据重构。

1.3 样本库的构建

模型训练完成后，输入预处理后发动机的传感器参数即可得到对应的健康指数。以第m台训练发动机和第n台测试发动机为例，其健康指数集合 X^m和Y^n分别定义为

$$X^m=\left\{x_1^m,x_2^m,\cdots,x_{C_r^m}^m\right\} \tag{4}$$

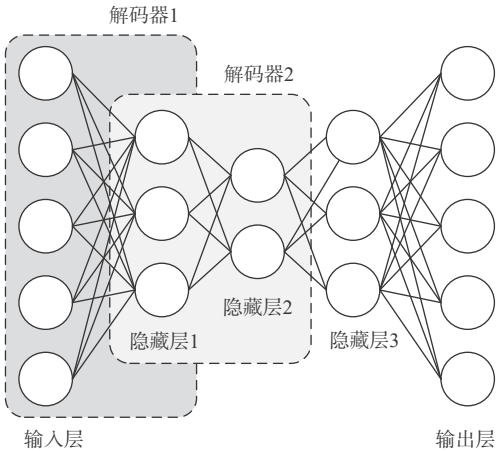


图 2　堆栈自编码器结构

Fig. 2　Stacked autoencoder structure

$$Y^n=\left\{y_1^n,y_2^n,\cdots,y_{C_t^n}^n\right\} \tag{5}$$

式中: x_1^m为第m台训练发动机在第 1 个循环时的健康指数; C_r^m为第m台训练发动机的循环结束点; y_1^n为第n台测试发动机在第 1 个循环时的健康指数; C_t^n为第n台测试发动机的当前循环点。

所有训练发动机的健康指数构成了相似性匹配的样本库, 为进一步分析发动机的退化趋势, 对样本库中的发动机进行指数拟合, 得到拟合样本库。其中, 第m台训练发动机拟合后的健康指数集合X'^m为

$$X'^m=\left\{x_1'^m,x_2'^m,\cdots,x_{C_r^m}'^m\right\} \tag{6}$$

式中: $x_1'^m$为第m台训练发动机在第 1 个循环时的健康指数。

2　区间划分改进相似性方法

在相似性匹配过程中, 样本的时间序列越长, 则相似样本的数量越少, 但相似性越高的样本, 预测结果越精确, 产生的误差就越小; 反之, 样本的时间序列越短, 相似样本的数量越多, 但预测的结果具有较大的不确定性, 误差较大。基于该特点, 设计区间划分改进相似性方法以解决测试发动机易引入伪相似发动机的问题。对已知运行循环较短、中等和较长 3 种区间下的测试发动机进行针对性处理, 识别、去除初选参考发动机中的异常发动机以得到最终的预测结果。该方法容易集成在传统相似性匹配方法上, 具有简单、方便的优点。

具体来讲, 针对已知运行循环较短的测试发动机在预测时具有较大的不确定性问题, 提出一种修正方法以获取更加保守的结果; 针对已知运行循环为中等长度的发动机存在易出现伪相似发动机的问题, 采用退化一致性检验进行筛选, 以寻找更加值得信任的发动机; 对已知运行循环较长的测试发动机进行预测时, 发动机的性能退化趋势明显, 因此, 使用传统的相似性匹配方法进行预测, 在保证精度的前提下, 相较于退化一致性检验方法, 有更加快速的运算效率。

综上所述, 区间划分改进相似性方法预测流程如图 3 所示。

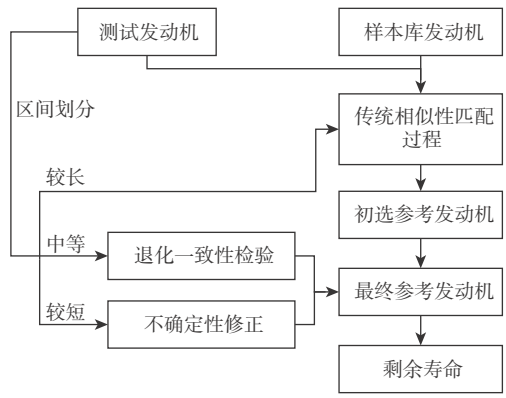


图 3　区间划分改进相似性方法流程

Fig. 3　Flow of improved similarity method with interval partition

2.1　传统相似性匹配方法

传统相似性匹配过程通过选取一定的距离尺度, 计算测试发动机与每台训练发动机之间的距离来反映相似度, 并基于此筛选出最相似的多台训练发动机进行预测。常用的距离尺度包括欧氏距离和曼哈顿距离[18], 本文选择平均欧氏距离的平方作为距离尺度, 逐点计算测试发动机与训练发动机之间的距离, 定义为

$$d_t^{m,n}=\frac{1}{C_t^n}\sum_{i=1}^{C_t^n}\left(x_{t+i-1}^m-y_i^n\right)^2 \tag{7}$$

式中: $d_t^{m,n}$为第m台训练发动机与第n台测试发动机在时滞t下的距离; x_{t+i-1}^m为训练发动机在时滞t下第i个循环时的健康指数。

另外, 为解决不同发动机初始误差不同的问题, 增加时滞t进行相似性匹配, 从而获得测试发动机在不同时滞下与训练发动机之间的相似度, 如下:

$$S_t^{m,n}=\exp(-d_t^{m,n}/\lambda) \tag{8}$$

式中: $S_t^{m,n}$为第m台训练发动机与第n台测试发动机在时滞t下的相似度; λ为距离放缩因子, 通常取 0.002。

当发动机的历史退化数据较少时, 进行相似性匹配易产生较大误差, 为获取更多相似的样本, 使用零中心化规则进行匹配[13]。零中心化规则允许发动机的健康指数曲线在纵向进行移动, 直至找到最佳的匹配位置, 如图 4 所示。最佳匹配位置t_0对应的相似度为

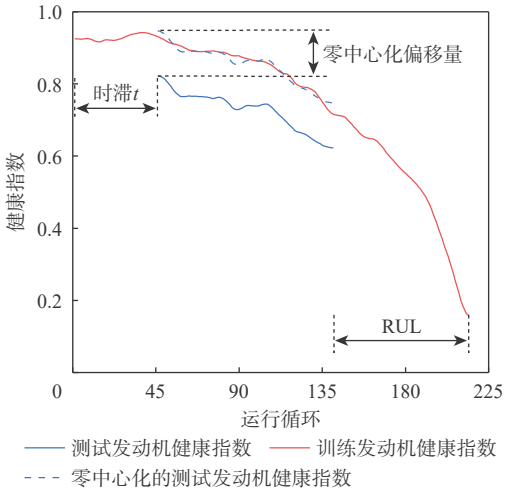


图4 传统相似性匹配过程

Fig. 4 Traditional similarity matching process

$$S_{t_o}^{m,n}=\max_{1<t<C_r^m-C_t^n}S_t^{m,n} \tag{9}$$

式中：$S_{t_o}^{m,n}$为第m台训练发动机与第n台测试发动机在最佳匹配位置时的相似度。

相似度阈值由相似滤波系数γ和参考发动机的相似度计算得到，超过阈值的发动机作为参考发动机，如下：

$$S_{t_o}^{m,n}>\gamma\max(S_{t_o}^{m,n}) \tag{10}$$

最终，将参考发动机的预测结果按照式(11)进行融合，得到测试发动机的最终预测结果。

$$R^n=\frac{\sum S_{t_o}^{m,n}R^{m,n}}{\sum S_{t_o}^{m,n}} \tag{11}$$

式中：R^n为第n台测试发动机的最终预测结果；$R^{m,n}=C_r^m-t_o-C_t^n$为第n台测试发动机与第m台训练发动机的预测结果。

2.2 退化一致性检验

退化一致性检验通过计算发动机之间的伪相似度来衡量整体趋势对局部趋势的影响。首先，将测试发动机健康指数窗口化后进行传统相似性匹配过程，相似度最高的窗口数据称为最优匹配段；然后，分析最优匹配段上游和下游(统称为邻段)的退化趋势，并计算伪相似度；最后，将通过一致性检验的发动机作为最终参考发动机。具体过程如图5所示。

退化一致性检验方法的核心思想是：最优匹配段与其相邻段之间的退化趋势偏差和训练发动机健康指数与其拟合轨迹之间的偏差相关，即最优匹配段与邻段的偏差越大，拟合曲线与健康曲线偏差越大，反之亦然。因此，定义时间窗口W^n的上游和下游分别为

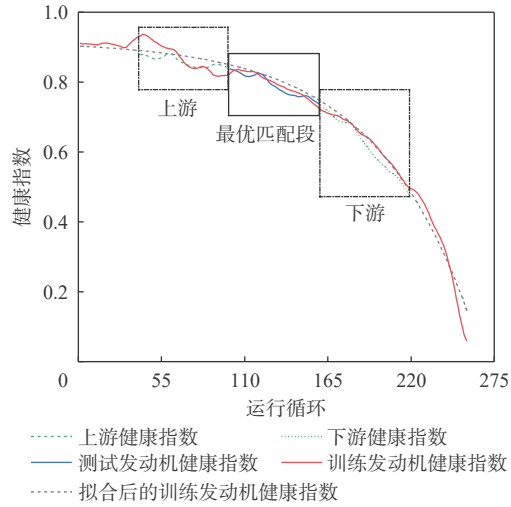


图5 退化一致性检验过程

Fig. 5 Degradation consistency test process

$$W_{-,o}^{m,n}=\{y_{-,1}^{m,n},y_{-,2}^{m,n},\cdots,y_{-,\Delta t^n}^{m,n}\} \tag{12}$$

$$W_{+,o}^{m,n}=\{y_{+,1}^{m,n},y_{+,2}^{m,n},\cdots,y_{+,\Delta t^n}^{m,n}\} \tag{13}$$

式中：$W_{-,o}^{m,n}$和$W_{+,o}^{m,n}$分别为最优匹配段上游和下游的健康指数集合；$y_{-,1}^{m,n}$和$y_{+,1}^{m,n}$分别为上游和下游在第1个循环时的健康指数；Δt^n为窗口长度。

根据核心思想，时间窗口的平移距离和健康指数可以定义为

$$\Delta t_{-,o}^{m,n}=\min(\beta\Delta t,t_o-\beta\Delta t) \tag{14}$$

$$\Delta t_{+,o}^{m,n}=\min(\beta\Delta t,C_r^m-t_o-\Delta t) \tag{15}$$

$$y_{-,i}^{m,n}=y_i^n+x'^m_{t_o-\Delta t_{-,o}^{m,n}+i-1}-x'^m_{t_o+i-1} \tag{16}$$

$$y_{+,i}^{m,n}=y_i^n-x'^m_{t_o-\Delta t_{+,o}^{m,n}+i-1}+x'^m_{t_o+i-1} \tag{17}$$

式中：$\Delta t_{-,o}^{m,n}$和$\Delta t_{+,o}^{m,n}$分别为最优匹配段向上和向下的平移距离；β为时间窗口平移系数。

按式(18)～式(22)计算最优匹配段及其上下游和训练发动机的相似度以得到伪相似度，上下游的相似度越低，则伪相似度越大，表示最优匹配段附近的一致性越差；反之亦然。

$$d_{-,t_o}^{m,n}=\frac{1}{\Delta t^n}\sum_{t=1}^{\Delta t^n}\left(y_{-,i}^{m,n}-x^m_{t_o-\Delta t_{-,o}^{m,n}+i-1}\right)^2 \tag{18}$$

$$d_{+,t_o}^{m,n}=\frac{1}{\Delta t^n}\sum_{t=1}^{\Delta t^n}\left(y_{+,i}^{m,n}-x^m_{t_o-\Delta t_{+,o}^{m,n}+i-1}\right)^2 \tag{19}$$

$$S_{-,t_o}^{m,n}=\exp(-d_{-,t_o}^{m,n}/\lambda) \tag{20}$$

$$S_{+,t_o}^{m,n}=\exp(-d_{+,t_o}^{m,n}/\lambda) \tag{21}$$

$$\alpha_{t_o}^{m,n}=2S_{t_o}^{m,n}/\left(S_{-,t_o}^{m,n}+S_{+,t_o}^{m,n}\right) \tag{22}$$

式中：$\alpha_{t_o}^{m,n}$为第m台训练发动机与第n台测试发动机

最优匹配段的伪相似度。

得到测试发动机与每台初选参考发动机的伪相似度后, 按照伪相似度百分比θ去除不满足整体退化趋势的发动机, 按式(11)进行融合, 得到已知运行循环为中等长度的发动机的最终 RUL 预测值。

2.3　不确定性修正方法

发动机的性能退化大致可分为平稳期和衰退期, 处于衰退期的发动机退化趋势明显, 而平稳期则较为稳定[19]。因此, 对已知运行循环较短的发动机进行预测时, 结果往往存在较大的不确定性[20], 一方面, 导致多模型融合后的结果产生较大的误差, 另一方面, 无法在实际中指导维修决策。因此, 为稳定预测结果的区间, 本文结合多模型相似性匹配提出了一种预测不确定性的修正方法。

测试发动机的已知运行循环较短时, 其参考发动机的数量较多, 因此, 先利用不使用零中心化规则的传统相似性匹配过程得到与测试发动机相似的多台参考发动机, 再对比 RUL 预测值的大小, 去除较大的异常值, 保证预测值尽可能不超过实际失效值, 以更好地指导维修决策。不同的异常值阈值对预测精度有一定影响。将测试发动机的 RUL 预测值从小到大进行排序, 选取小于异常值阈值δ的发动机, 如下:

$$R^{m,n} < \delta \max(R^{m,n}) \tag{23}$$

将满足异常值阈值的发动机按照式(11)进行融合, 得到已知运行循环较短的测试发动机 RUL。使用均方根误差 (root mean square error, RMSE) 评价不同异常值阈值的预测结果, 选取最优的异常值阈值预测发动机的 RUL。

3　方法验证

3.1　数 据 集

采用 NASA 的 C-MAPSS 涡扇发动机数据集中的 FD001 数据集进行方法验证, 该数据集被广泛用于发动机的 RUL 预测[21]。数据集包含 100 台全寿命周期的训练发动机数据和 100 台不完整寿命的测试发动机数据及其对应的 RUL 值。发动机数据是在飞行高度为 0、油门角度为 100°、马赫数为 0.84 的运行条件下, 由 HPC 性能退化所导致的 21 个传感器参数变化值。传感器参数符号及具体含义如下: T_2 为风扇进口温度, T_{24} 为低压压气机出口温度, T_{30} 为 HPC 出口温度, T_{50} 为低压涡轮出口温度, P_2 为风扇入口压力, P_{15} 为外涵道总压力, P_{30} 为 HPC 出口总压, N_f 为风扇转速, N_c 为核心机转速, e_{pr} 为发动机压比, P_{s30} 为 HPC 出口静压, P_{hi} 为燃油流量和 P_{s30} 比值, N_{Rf} 为风扇修正转速, N_{Rc} 为核心机修正转速, B_{PR} 为涵道比, B_{far} 为燃烧室油气比, h_{tBleed} 为引气焓, N_{f_dmd} 为风扇转速命令值, P_{CNfR_dmd} 为风扇修正转速命令值, W_{31} 为高压涡轮冷却空气流量, W_{32} 为低压涡轮冷却空气流量。

3.2　评价指标

采用均方根误差和评分函数来评价发动机的 RUL 预测结果, 定义为

$$E_{\mathrm{RMSE}} = \sqrt{\frac{1}{N}\sum_{n=1}^{N}\left(R_{\mathrm{a}}^{n} - R^{n}\right)^{2}} \tag{24}$$

$$E_{\mathrm{Score}} = \begin{cases} \displaystyle\sum_{n=1}^{N}\left[\mathrm{e}^{-\left(R_{\mathrm{a}}^{n}-R^{n}\right)/13}-1\right] & R_{\mathrm{a}}^{n} < R^{n} \\ \displaystyle\sum_{n=1}^{N}\left[\mathrm{e}^{\left(R_{\mathrm{a}}^{n}-R^{n}\right)/10}-1\right] & R_{\mathrm{a}}^{n} \geqslant R^{n} \end{cases} \tag{25}$$

式中: N为测试发动机总台数; R_{a}^{n}为第n台发动机的 RUL 真实值。

区别于均方根误差, 评分函数对预测误差的惩罚具有非对称性, 认为延迟预测比超前预测具有更差的性能, 如图 6 所示。均方根误差和评分函数越小, 则预测的误差越小, 结果越精确。

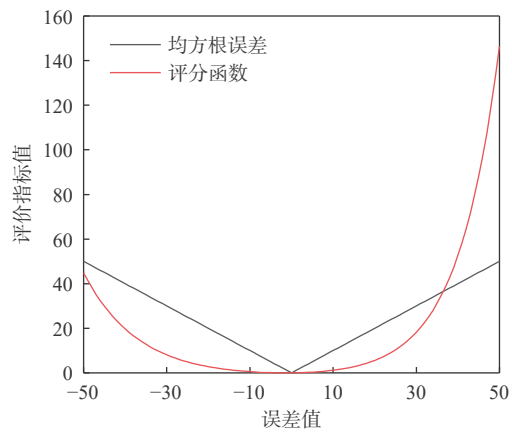


图 6　不同误差值的评价指标比较

Fig. 6　Evaluation metrics comparison of different errors

3.3　构建健康指数

使用卡尔曼滤波方法对筛选出的发动机传感器参数进行滤波、平滑处理, 对其进行去量纲化后得到预处理后的发动机数据, 代入 SAE 中进行训练。SAE 的网络结构如表 1 所示。

表 1　堆栈自编码器网络参数

Table 1　Stacked autoencoder network parameters

层数	节点数	激活函数
1	4	Sigmoid
2	2	Sigmoid
3	1	Sigmoid

模型训练完成后，将发动机预处理后的数据输入SAE中，得到某台发动机的健康指数，如图7所示。

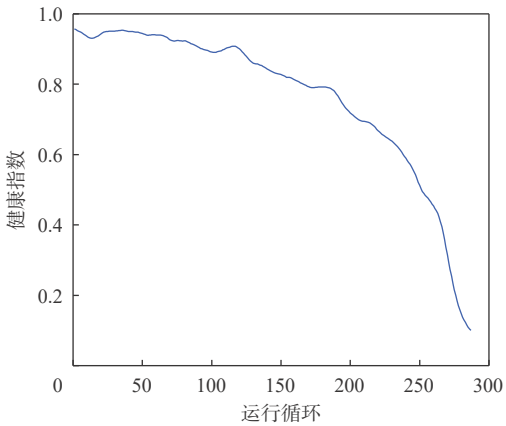


图7 某台发动机健康指数

Fig. 7 Health index of an engine

3.4 参数的确定

衡量发动机的整体退化趋势时，忽略系统初始和最终阶段的退化趋势[14]，基于此理念对训练发动机进行统计分析。在确定运行区间分界点时，将每台发动机的数据代入隐马尔可夫模型进行健康状态识别，并定义发动机的4个退化状态为健康、磨损、严重磨损和失效。计算发动机第1个和第3个健康状态分界点的均值，并基于此进行不同区间的划分，得到较短和中等运行循环区间的分界点为60，中等和较长运行循环区间的分界点为162。

对训练发动机进行交叉验证，并采用随机搜索方法进行寻优，最终确定区间划分改进相似性匹配方法中的参数，如表2所示。

3.5 结果评价

确定已知运行循环较短的测试发动机后，对其进行预测不确定性的修正，得到均方根误差随异常值阈值变化的曲线，如图8所示。

随着异常值阈值的下降，均方根误差先下降后上升，最优预测精度时的异常值阈值为0.8，这是因为已知运行循环较短的发动机预测时的不确定性较大，去除该比例下的异常值后能够有效提高预测精度。而过小的异常值阈值会去除正确的参考发动机，影响RUL的预测结果。

表2 区间划分改进相似性参数

Table 2 Improved similarity parameters with interval partition

参数	数值
时间窗口长度W	60
距离放缩因子λ	0.002
相似滤波系数γ	0.9
多模型个数N	20
时间窗口平移系数β	1
伪相似度百分比θ	80

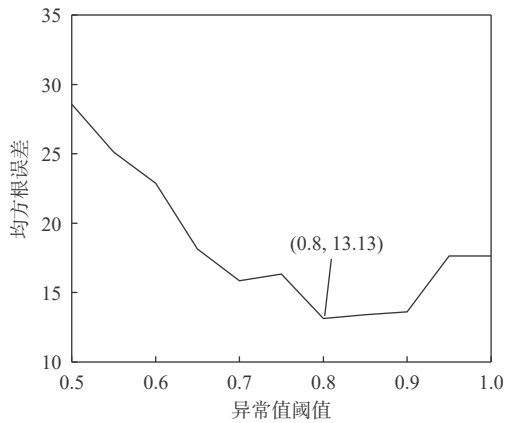


图8 不同异常值阈值的均方根误差比较

Fig. 8 RMSE comparison of different abnormal thresholds

为说明退化一致性检验方法对预测精度的影响，以第89台测试发动机为例进行进一步的解释。表3为第89台测试发动机及对应的伪相似度实例。测试发动机的相似度阈值为0.906，将超过阈值的发动机作为初选参考发动机，进行加权融合

表3 第89台测试发动机及对应的伪相似度

Table 3 The 89th test engine and its pseudo-similarity

第89台测试发动机RUL真实值	初选参考发动机序号	初选参考发动机RUL预测值	传统相似性匹配方法				退化一致性检验			
			相似度	相似度阈值	RUL预测值	误差	伪相似度	伪相似度阈值	RUL预测值	误差
	20	27	0.985				1.018			
	54	20	0.988				1.100			
	59	26	0.997				1.217			
	80	62	0.991				1.389			
28	91	62	0.992	0.906	42	14	1.388	1.234	32	4
	52	26	0.992				1.202			
	49	29	0.987				1.485			
	95	61	0.994				1.476			
	⋮	⋮	⋮				⋮			

得到的 RUL 预测结果为 42, 实际寿命与预测寿命的误差值为 14。以其中 8 台相似程度较高的训练发动机为例, 进行退化一致性检验后确定伪相似度阈值为 1.234, 将超过阈值的发动机去除, 得到测试发动机的预测结果为 32, 预测误差为 4, 与真实值相近。但该方法在实际应用中存在错误识别正确发动机的情况, 如第 49 台训练发动机。总体来讲, 作为传统相似性匹配方法的二次筛选过程, 退化一致性检验通过计算发动机之间的伪相似度识别并去除有突变段或退化趋势不同的发动机, 能够提高已知运行循环为中等长度的发动机预测准确性。

为进一步验证本文方法的有效性, 将本文的预测结果与当前的研究成果进行比较, 包括多层感知机 (multilayer perceptron, MLP)[22]、支持向量回归 (support vector regression, SVR)[23]、深度长短期记忆 (deep long short-term memory, DLSTM) 网络[24] 和基于退化特征的相似性匹配 (similarity matching based on degraded features, SMD)[11], 如表 4 所示。可知, 本文方法的均方根误差相对较好, 且评分函数相较于 SMD 方法降低了 34%, 能够有效、及时地预测发动机的 RUL, 证明了方法的有效性。

表 4　不同预测方法的预测精度对比

Table 4　Comparison of prediction accuracy of different prediction methods

方法	均方根误差	评分函数
MLP[22]	25.23	1 205
SVR[23]	20.96	1 381
DLSTM[24]	18.33	655
SMD[11]		682
本文方法	16.26	450

4　结　论

1) 基于发动机的工作原理确定了性能退化过程中的关键参数, 结合 SAE 构建的健康指数能够更好地反映发动机的性能衰退。

2) 针对相似性匹配方法的时间序列特点, 提出了区间划分改进相似性方法以识别伪相似发动机。对于已知运行循环较短的发动机, 提出了不确定性问题的修正方法; 对于已知运行循环为中等长度的发动机, 使用退化一致性检验筛选更加值得信任的发动机, 用于提高预测的精度。

3) 利用 C-MAPSS 数据集验证表明, 相较于当前的相似性匹配方法, 本文方法评分函数降低了 34%, 这是因为本文将发动机的退化特点与相似性匹配的特点相结合对发动机进行区间划分改进相似性匹配, 提高了传统相似性匹配方法的性能。

但在实际应用过程中, 没有考虑环境及多种故障模式带来的影响, 后续将会对这方面进行深入研究。

参考文献 (References)

[1] 中国民用航空局. 中国民用航空发展第十三个五年规划[EB/OL]. (2017-02-16)[2022-08-18]. http://www.gov.cn/xinwen/2017-02/16/content_5168506.htm.
Civil Aviation Administration of China. The Thirteenth Five-Year Plan for the development of China's civil aviation[EB/OL]. (2017-02-16)[2022-08-18]. http://www.gov.cn/xinwen/201702/16/content_5168506.htm(in Chinese).

[2] 尉询楷, 杨立, 刘芳, 等. 航空发动机预测与健康管理[M]. 北京: 国防工业出版社, 2014: 132-133.
WEI X K, YANG L, LIU F, et al. Aero-engine prediction and health management[M]. Beijing: National Defense Industry Press, 2014: 132-133(in Chinese).

[3] ELSHEIKH A, YACOUT S, OUALI M S. Bidirectional handshaking LSTM for remaining useful life prediction[J]. Neurocomputing, 2019, 323: 148-156.

[4] 周俊. 数据驱动的航空发动机剩余使用寿命预测方法研究[D]. 南京: 南京航空航天大学, 2017.
ZHOU J. Research on data-driven prediction methods for remaining useful life of aero-engine[D]. Nanjing: Nanjing University of Aeronautics and Astronautics, 2017(in Chinese).

[5] PECHT M. Prognostics and health management of electronics[M]. New York: Wiley, 2008.

[6] CHEN X, JIN G, QIU S Q, et al. Direct remaining useful life estimation based on random forest regression[C]//Proceedings of the Global Reliability and Prognostics and Health Management. Piscataway: IEEE Press, 2020: 1-7.

[7] 车畅畅, 王华伟, 倪晓梅, 等. 基于 1D-CNN 和 Bi-LSTM 的航空发动机剩余寿命预测[J]. 机械工程学报, 2021, 57(14): 304-312.
CHE C C, WANG H W, NI X M, et al. Residual life prediction of aeroengine based on 1D-CNN and Bi-LSTM[J]. Journal of Mechanical Engineering, 2021, 57(14): 304-312(in Chinese).

[8] 郭庆, 李印龙, 郑天翔. 基于非线性 Wiener 过程航空发动机性能退化预测[J]. 推进技术, 2021, 42(9): 1956-1963.
GUO Q, LI Y L, ZHENG T X. Performance degradation prediction of aero-engine based on nonlinear Wiener process[J]. Journal of Propulsion Technology, 2021, 42(9): 1956-1963(in Chinese).

[9] ZHANG H, ZHANG Q, SHAO S Y, et al. Attention-based LSTM network for rotatory machine remaining useful life prediction[J]. IEEE Access, 2020, 8: 132188-132199.

[10] WANT T. Trajectory similarity based prediction for remaining useful life estimation [D]. Cincinnati: University of Cincinnati, 2010.

[11] 张妍, 王村松, 陆宁云, 等. 基于退化特征相似性的航空发动机寿命预测[J]. 系统工程与电子技术, 2019, 41(6): 1414-1421.
ZHANG Y, WANG C S, LU N Y, et al. Remaining useful life prediction for aero-engine based on the similarity of degradation characteristics[J]. Systems Engineering and Electronics, 2019, 41(6): 1414-1421(in Chinese).

[12] CAI H S, JIA X D, FENG J S, et al. A similarity based methodology for machine prognostics by using kernel two sample test[J]. ISA Transactions, 2020, 103: 112-121.

[13] YU W N, KIM I Y, MECHEFSKE C. An improved similarity-based prognostic algorithm for RUL estimation using an RNN autoencoder scheme[J]. Reliability Engineering & System Safety, 2020, 199: 106926.

[14] XUE B, XU F M, HUANG X, et al. Improved similarity based prognostics method for turbine engine degradation with degradation consistency test[J]. Applied Intelligence, 2022, 52(9): 10181-10201.

[15] TKACZ E, KOZANECKA D, KOZAN-ECKI Z. Investigations of oil-free support systems to improve the reliability of ORC hermetic high-speed turbomachinery[J]. Mechanics and Mechanical Engineering, 2011, 15(3): 355-365.

[16] 王昆, 郭迎清, 赵万里, 等. 基于 SSAE 和相似性匹配的航空发动机剩余寿命预测[J]. 北京航空航天大学学报, 2023, 49(10): 2817-2825.
WANG K, GUO Y Q, ZHAO W L, et al. Remaining useful life prediction of aeroengine based on SSAE and similarity matching[J]. Journal of Beijing University of Aeronautics and Astronautics, 2023, 49(10): 2817-2825(in Chinese).

[17] 李浩, 王卓健, 李哲, 等. 基于堆栈自编码器和 DeepAR 的航空发动机剩余寿命预测[J]. 推进技术, 2022, 43(11): 72-80.
LI H, WANG Z J, LI Z, et al. Prediction of remaining useful life of aero-engine based on stacked autoencoder and DeepAR[J]. Journal of Propulsion Technology, 2022, 43(11): 72-80(in Chinese).

[18] KHELIF R, MALINOWSKI S, CHEBEL-MORELLO B, et al. RUL prediction based on a new similarity-instance based approach[C]// Proceedings of the IEEE 23rd International Symposium on Industrial Electronics. Piscataway: IEEE Press, 2014: 2463-2468.

[19] RAMASSO E. Investigating computational geometry for failure prognostics in presence of imprecise health indicator: results and comparisons on C-MAPSS datasets[C]//Proceedings of the 2nd European Conference of the Prognostics and Health Management Society. Piscataway: IEEE Press, 2014: 1-13.

[20] MALHOTRA P, TV V, RAMAKRISHNAN A, et al. Multi-sensor prognostics using an unsupervised health index based on LSTM encoder-decoder[EB/OL]. (2016-08-22)[2022-08-18]. http://arxiv.org/abs/1608.06154.

[21] SAXENA A, GOEBEL K, SIMON D, et al. Damage propagation modeling for aircraft engine run-to-failure simulation[C]//Proceedings of the International Conference on Prognostics and Health Management. Piscataway: IEEE Press, 2008: 1-9.

[22] 李杰, 贾渊杰, 张志新, 等. 基于融合神经网络的航空发动机剩余寿命预测[J]. 推进技术, 2021, 42(8): 1725-1734.
LI J, JIA Y J, ZHANG Z X, et al. Remaining useful life prediction of aeroengine based on fusion neural network[J]. Journal of Propulsion Technology, 2021, 42(8): 1725-1734(in Chinese).

[23] SUN W J, ZHAO R, YAN R Q, et al. Convolutional discriminative feature learning for induction motor fault diagnosis[J]. IEEE Transactions on Industrial Informatics, 2017, 13(3): 1350-1359.

[24] WU J, HU K, CHENG Y W, et al. Data-driven remaining useful life prediction via multiple sensor signals and deep long short-term memory neural network[J]. ISA Transactions, 2020, 97: 241-250.

Remaining life prediction of engine by improved similarity with interval partition

ZHAO Hongli*, BAI Lingde

(School of Aeronautical Engineering, Civil Aviation University of China, Tianjin 300300, China)

Abstract: The traditional similarity matching method is liable to introduce pseudo-similar engines, which leads to low prediction accuracy. To address this issue, a remaining useful life (RUL) prediction method of improved similarity with interval partition was proposed, which combined the engine performance degradation characteristics and similarity matching characteristics. Firstly, the health index was constructed based on the selected parameters by using the stacked autoencoder. Then, based on the known running cycles of the test engine, the interval was divided, and the traditional similarity matching method was used for preliminary screening. For the test engines in different intervals, the uncertainty correction and degradation consistency test were performed to remove the abnormal engine from the preliminarily selected reference engines, and the final remaining life prediction results were obtained. The C-MAPSS dataset of NASA was used for verification, and the results show that the prediction accuracy is 34% higher than the current similarity matching method, which proves the effectiveness of the proposed method.

Keywords: aero-engine; life prediction; improved similarity with interval partition; degradation consistency test; uncertainty correction

Received: 2022-09-07; **Accepted**: 2022-12-02; **Published Online**: 2022-12-15 11:41

URL: link.cnki.net/urlid/11.2625.V.20221214.1142.003

Foundation items: The Fundamental Research Funds for the Central Universities (3122021049); Civil Aviation University of China Experimental Technology Innovation Fund (2021CXJJ90)

* **Corresponding author**. E-mail: henleytrent@163.com

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2022.0763

施工工艺对复合材料蒙皮电气性能影响分析

杨占刚*，柯中澍，杨旭伟，包兴旺
(中国民航大学 电子信息与自动化学院，天津 300300)

摘 要：新型商用飞机蒙皮中复合材料的占比越来越高，然而较差的导电性使其需要额外的导电结构实现电流在蒙皮内的流通。为提高蒙皮导电能力，降低复合材料与导电结构的搭接对电流通路的影响，基于各组成结构的电气特性对蒙皮结构进行建模，并利用解析方法分析搭接时可能产生影响的施工工艺。针对蒙皮不同部位各自处理工艺带来的接触面间的电流与整体结构阻抗的变化，分析各结构在装配过程中的施工工艺对飞机蒙皮搭接结构电气特性的影响情况。理论分析与仿真计算结果表明：正确的施工工艺处理方法，尤其是对蒙皮表面树脂与金属连接件氧化膜恰当的处理程度，能够有效减小整体结构阻抗对蒙皮电气性能的影响，并能够为面向现场的施工提供一定的操作建议。

关 键 词：飞机蒙皮；复合材料；搭接结构；施工工艺；电气性能
中图分类号：V261.97
文献标志码：A **文章编号**：1001-5965(2024)10-3013-08

由于复合材料所具有的质量轻、强度高、抗腐蚀性强等特点，现代大型客机蒙皮选材从最初的金属材料，逐渐转变为以碳纤维增强聚合物(carbon fiber reinforced plastic, CFRP)为主的先进复合材料，该材料在整机各区域的使用比例也逐代上升[1]。传统金属机身蒙皮一般为电的良导体，可视为理想的接地等电位体，在蒙皮内的电流导通过程中起着低阻抗通路的作用[2-4]；而CFRP材料电阻各向异性，其最小方向的阻值仍超铝材料的60倍[5]，直接用于电气接地时会引入远大于标准值的电阻，造成飞机蒙皮接地回路阻抗偏大，影响电流的正常流通。基于此，复合材料飞机蒙皮一般通过嵌入金属铜箔和金属连接件作为主要导电结构，再通过紧固件的连接形成一条完整稳定的蒙皮内电气通路，从而实现辅助接地点电流及故障电流的导通，降低接地网络带来的额外电压降对机载设备的影响，最终与飞机接地回流网络共同组成电流导通回路[6]。

在各结构件的搭接装配过程中，作为主要支撑结构的CFRP与导电结构接触，产生的搭接阻抗会对蒙皮整体结构的电学特性产生影响[7]。为降低该影响，一般通过使用电学性能更优异的导电结构或在装配时提供更适宜的工作环境以提高复合材料蒙皮整体的导电性能[8-9]。同时，各搭接组成部分制造品质的优良(如接触面上接触点数量及形貌)会影响接触材料的微观结构和物理性能，并最终对蒙皮的阻抗特性产生影响[10-11]。文献[12-13]研究了不同装配温度下，复合材料的正温度及负温度系数效应对搭接结构导电性能的影响。文献[14]通过减少生产过程中微粒污染膜和无机污染膜的产生，使触点的接触电阻控制在可接受的范围内，并长期保持稳定状态。文献[15-16]基于机电阻抗的研究方法，分析了搭接结构胶接的质量对接触面键合性能的影响，给出了评估各结构间键合品质的方法。然而，当复合材料飞机蒙皮所使用的构件材质和工作环境得到确定后，搭接过程中采取的施工工艺就成为影响蒙皮结构电气特性的决定性因素之一。表面结构的打

收稿日期：2022-09-07；**录用日期**：2022-10-07；**网络出版时间**：2023-01-10 10:03
网络出版地址：link.cnki.net/urlid/11.2625.V.20230109.1005.001
基金项目：航空科学基金(20182667010)；国家商用飞机制造工程技术研究中心创新基金(COMAC-SFGS-2021-609)
* **通信作者**. E-mail：yangcauc@163.com

引用格式：*杨占刚，柯中澍，杨旭伟，等. 施工工艺对复合材料蒙皮电气性能影响分析[J]. 北京航空航天大学学报，2024，50(10)：3013-3020. YANG Z G，KE Z S，YANG X W，et al. Analysis of effect of construction process on electrical properties of composite skins[J]. Journal of Beijing University of Aeronautics and Astronautics，2024，50(10)：3013-3020 (in Chinese).*

磨程度[17]，以及复合材料与导电连接件接触面的处理方式[18]，都会对搭接结构阻抗特性和接触区域的电流密度产生影响。

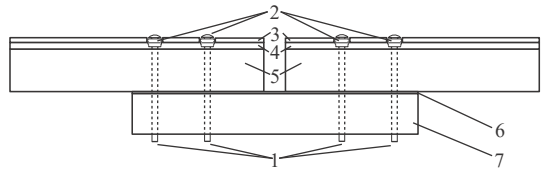
目前，针对搭接结构电气性能的计算，主要集中于各独立结构的电气分布及故障电流的泄放，而较少关注整体结构的阻抗及2种材质连接处的接触面电流问题。例如，文献[19-20]使用基于体积的部分元等效电路算法，分析了结构件的阻抗对接地回流网内压降的影响，忽略了构件之间接触阻抗的产生及其受施工工艺影响的变化情况。文献[5]借助建模工具InCa3D对复合材料飞机接地回流网络建模，通过机载电缆的电流值和压降变化分析最佳电流返回路径，但未考虑结构件间的电阻影响。文献[21-22]提出基于恒流场理论的接地网参数计算方法，在已知搭接结构中任一点的电位满足Laplace方程的前提下，将导体模型按照不同结构进行分段处理，通过计算各分段的自电阻和互电阻，求取整体结构的电流分布。文献[23]分析了不同结构组成的接地回流网络性能，但仅局限于规则结构的搭接分析，且未关注接触面间结构变化对整体阻抗产生的影响。为探究施工工艺对复合材料飞机蒙皮的搭接完整性及电气性能的影响情况，本文基于可能对蒙皮结构产生电学性能影响的施工操作进行理论分析，结合蒙皮典型搭接结构仿真计算不同施工工艺带来的接触面电流密度分布及整体阻抗变化情况，给出提高复合材料飞机蒙皮搭接性能的施工工艺设计优化方案。

1 复合材料飞机蒙皮结构电学性能分析

复合材料飞机蒙皮一般由CFRP、金属连接件、延展金属箔或金属网、树脂、螺栓及凹形垫圈组成。其中，CFRP作为主体部分，是支撑整块蒙皮的结构基础；金属连接件除了在结构上连接分离的板材外，还是电流在2块蒙皮间导通的主要载体；在CFRP外表面铺设一块平整的金属箔，一般为铜材质；金属箔的外侧再覆盖一层树脂，保护金属铜箔不被划损破坏。树脂、延展金属箔、CFRP、金属连接件等结构通过紧固件实现结构上的连接，同时也确保了电流在整块复合材料飞机蒙皮内的导通。各结构的搭接装配如图1所示。

1.1 打磨金属件氧化铝的电气特性分析

由于CFRP中一般存在少量导电颗粒，使得部分电流通过相应接触面进入CFRP中，目前，关于CFRP内部导电原理较为普遍的说法是通道导电理论[24]。该理论认为，由于CFRP中的导电填料较少，



1. 螺栓；2. 凹型垫圈；3. 树脂；4. 延展金属箔；5. CFRP；6. 金属氧化膜；7. 金属连接件。

图1 蒙皮搭接结构示意图

Fig. 1 Schematic of skin bonding structure

导电微粒一般均匀分布其中，彼此之间相互隔绝，使得CFRP呈现绝缘的特性，因此，可将其类比于电容。

进一步对CFRP与金属连接件的接触面情况进行探究，考虑到金属连接件为铝材质，而铝极易在空气中与氧气反应生成一层致密的氧化膜，在2个连接结构间阻碍电流导通。一般，需在装配前对该层氧化膜进行打磨，刷上一层特制的导电溶液，以提高结构连接处的导电性能，同时预防铝制连接件在空气中再次被氧化。基于螺杆附近区域的CFRP与金属连接件的接触情况，进行结构界面间及金属氧化物的等效电路分析，该部分模型如图2所示。

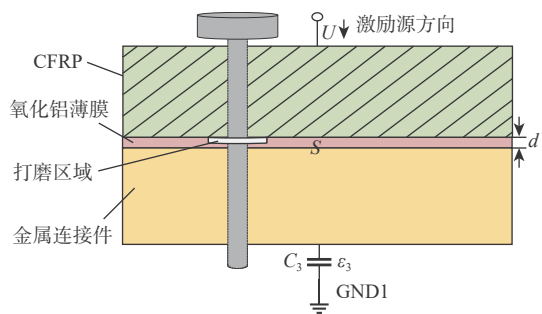


图2 考虑氧化膜的接触界面示意图

Fig. 2 Schematic of contact surface with oxide film

对搭接结构接触区域进行等效分析，设施加在CFRP顶端的电压为U，在CFRP、氧化铝薄膜部分的电压分别为U_1、U_2，对应介电常数分别为ε_1、ε_2。将金属连接件等效为介电常数为ε_3的电容与一个电阻串联，只对电容的影响因素进行探究，其电容值为C_3，分得的电压为U_3，且有$U = U_1 + U_2 + U_3$。当整体结构稳定后，根据串联电路的电流处处相等原理，即各结构的电荷量相等：

$$Q_1 = Q_2 = Q_3 \tag{1}$$

式中：Q_1、Q_2、Q_3分别为CFRP、氧化铝薄膜、金属连接件的电荷量。

设C_1、C_2分别为CFRP、氧化铝薄膜的电容，根据电容的定义式$C = Q/U$及式(1)，有

$$\begin{cases} U_1 = \dfrac{U}{1+C_1/C_2+C_1/C_3} \\ U_2 = \dfrac{U}{1+C_2/C_1+C_2/C_3} \\ U_3 = \dfrac{U}{1+C_3/C_1+C_3/C_2} \end{cases} \tag{2}$$

由电容的决定式$C=\varepsilon S/d$, 当氧化铝的材质确定后, 即介电常数保持恒定, 氧化铝薄膜的电容与其电压垂直方向上的表面积S和自身厚度d相关, 如下:

$$\begin{cases} U_1 = \dfrac{U}{1+C_1\Big/\left(\varepsilon_2\dfrac{S}{d}\right)+C_1C_3} \\ U_2 = \dfrac{U}{1+\varepsilon_2\dfrac{S}{d}\Big/C_1+\varepsilon_2\dfrac{S}{d}\Big/C_3} \\ U_3 = \dfrac{U}{1+C_3/C_1+C_3\Big/\left(\varepsilon_2\dfrac{S}{d}\right)} \end{cases} \tag{3}$$

由式 (3) 可知, 各结构分得电压与其相连结构的物理量有关。对金属氧化铝薄膜进行打磨时, 会改变其表面积S和厚度d, 打磨程度不同, 则其电容大小不同。当打磨区域的面积增大时, 被打磨变薄的极板正对面积增多, 氧化铝薄膜的电容逐渐增大。同理, 当氧化铝薄膜的厚度d逐渐减小时, CFRP、金属连接件上的电压U_1、U_3逐渐增大, 氧化铝薄膜上的电容C_2逐渐增大, d取极限时, 有

$$\begin{cases} U_1 = \dfrac{U}{1+C_1/C_3} \\ U_2 = 0 \\ U_3 = \dfrac{U}{1+C_3/C_1} \end{cases} \tag{4}$$

由电场及其与电流密度间关系, 有

$$U_m = \boldsymbol{E}_m \boldsymbol{l} \tag{5}$$

$$\boldsymbol{J}_m = \sigma_m \boldsymbol{E}_m \tag{6}$$

式中: U_m、$\boldsymbol{E}_m$分别为结构件上下两表面间电压、电场; $\boldsymbol{l}$为电场在结构件上降落方向的距离; σ_m为结构件的整体等效电导率; $\boldsymbol{J}_m$为结构件表面电流。

由式 (5) 与式 (6) 可知, 当仅对氧化铝薄膜打磨而不刷涂导电溶液时, 会使 CFRP 与金属连接件上的电流值变大, 而长时间的大电流易对结构产生破坏, 因此, 需要进一步施工处理。

1.2　刷涂导电溶液后接触面电气特性分析

当对金属氧化铝薄膜打磨完全且刷涂导电溶液后, 可将螺杆附近等效电路简化, 即图 2 中的氧化铝薄膜完全替代为导电溶液层。考虑到导电涂层较薄, 可进一步化简为具有表面电导率的电气结构, 如图 3 所示。图中, E_1、E_3分别为导电层上、下 2 个表面的电场大小, 根据结构件内电流与电场的

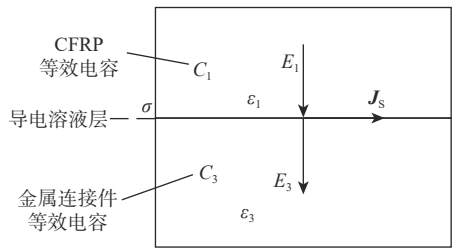


图 3　导电溶液处理后接触面示意图

Fig. 3　Schematic of contact surface after using conductive solution

关系, 分别指向与背离导电层; J_S为导电层上的电流大小; σ为刷涂上导电溶液后的该薄层的电导率。区域中, 任意一点m处的电势φ_m可由加在整体结构上的电势φ与导电溶液层边界Γ上的电场$\boldsymbol{E}$的竖直分量E_n表示:

$$\frac{1}{2}\varphi_m = \int_\Gamma E_n\varphi^*\mathrm{d}\Gamma + \int_\Gamma \varphi(\partial\varphi^*/\partial n)\mathrm{d}\Gamma \tag{7}$$

式中: φ^*为泊松方程$\nabla^2\varphi^* = -\Delta^P$的基础解, Δ^P为狄拉克δ函数, 除P点外, 其余位置该函数均为 0。

根据表面电荷密度ρ_S特性, 有

$$\varepsilon_1 E_1 - \varepsilon_3 E_3 = -\rho_S \tag{8}$$

$$\nabla\cdot\boldsymbol{J}_S = -\frac{\mathrm{d}\rho_S}{\mathrm{d}t} \tag{9}$$

此时, 可认为导电层上下表面电势相等, 即$\varphi_1=\varphi_3$, 其中, φ_1、φ_3分别为导电层上、下表面的电势。综合式 (8)、式 (9), 有

$$\mathrm{j}\omega\varepsilon_1 E_1 - \mathrm{j}\omega\varepsilon_3 E_3 = \nabla\cdot\boldsymbol{J}_S \tag{10}$$

式中: ω为角频率。

将 CFRP 与金属连接件中的电介质均视为理想电介质, 即无损电介质, 中间界面仍具有电导率。又根据$\boldsymbol{J}=\sigma\boldsymbol{E}$、$\boldsymbol{E}=-\nabla U$, 可将式 (10) 右侧化简为

$$\mathrm{j}\omega\varepsilon_1 E_1 - \mathrm{j}\omega\varepsilon_3 E_3 = -\sigma\nabla\cdot\nabla U' \tag{11}$$

式中: U'为 CFRP 上表面与金属连接件下表面间的电压差。

为提高计算精度, 将带有任意加权函数w的加权残差法应用于式 (11), 得

$$\int_{\Gamma_S} w(\mathrm{j}\omega\varepsilon_1 E_1 - \mathrm{j}\omega\varepsilon_3 E_3)\mathrm{d}\Gamma = -\sigma\int_{\Gamma_S} w\nabla\cdot\nabla\varphi\mathrm{d}\Gamma \tag{12}$$

式中: Γ_S表示电导率为σ的导电溶液层接触面。

根据格林定理, 对于闭合边界线长为L_Γ的任意表面Γ, 有

$$\int_\Gamma A\nabla\cdot\boldsymbol{B}\mathrm{d}\Gamma = -\int_\Gamma \nabla A\cdot\boldsymbol{B}\mathrm{d}\Gamma + \oint_{L_\Gamma} A\boldsymbol{B}\cdot\boldsymbol{a}_n\mathrm{d}L \tag{13}$$

式中: A、$\boldsymbol{B}$为积分中的任意函数(x,y,z); $\boldsymbol{a}_n$为表面Γ的法向量。

联立式 (12)、式 (13), 可得

$$\int_{\Gamma_S} w(\mathrm{j}\omega\varepsilon_1 E_1 - \mathrm{j}\omega\varepsilon_3 E_3)\mathrm{d}\Gamma = \sigma\int_{\Gamma_S}\nabla w\cdot\nabla\varphi\mathrm{d}\Gamma - \sigma\oint_{L_{\Gamma_S}} w\nabla\varphi\cdot\boldsymbol{a}_n\mathrm{d}L \tag{14}$$

式(14)表明，在材料特性及电学激励确定后，电导率为σ的导电界面，其接触面上的电场强度受导电界面边界条件的影响程度，即受打磨趋势的影响情况。

2 典型蒙皮搭接结构的建模

基于图1所示的蒙皮搭接结构对复合材料飞机蒙皮进行建模，树脂与延展金属箔的高度均为0.5 mm，CFRP、金属连接件及其氧化膜的高度分别为10.5 mm、5.8 mm及0.2 mm，上述规则结构件的长、宽均分别为500 mm、300 mm，螺栓及凹形垫圈的尺寸规格如表1所示。

复合材料飞机蒙皮各结构件所使用材料材质及相对应的电磁参数如表2所示。

表1 螺栓与凹形垫圈的尺寸

Table 1 Dimensions of bolts and concave washers mm

结构	参数
沉头螺栓总长度	18
沉头螺栓螺杆直径	6.4
沉头螺栓头直径	9
沉头螺栓头厚度	0.5
凹形垫圈高度	2.2
凹形垫圈内径	6.45
凹形垫圈外径	19
凹形垫圈厚度	0.4

表2 各结构件材质及电磁参数

Table 2 Material and electromagnetic parameters of each component

部件名称	材质	电导率/(S·m^{-1})
延展金属箔	铜	5.998×10^7
CFRP	碳纤维复合材料	1.5×10^4
金属氧化膜	氧化铝	0.01
金属连接件	铝	3.03×10^7
沉头螺栓	钢	4.032×10^6
凹形垫圈	钛	2.38×10^6

本文的研究对象是施工工艺带来的影响，除了各结构搭接装配外，连接结构的表面预处理是影响仿真计算结果的主要因素，一般在模型搭建时通过结构间接触面区域的模型调整实现。紧固件除了连接各结构外，还是电流导通的主要载体；为增大螺栓的导电能力，需预先对螺栓四周的表面树脂进行打磨，再通过凹形垫圈扩大螺栓头与金属铜箔的接触区域，使铜箔中的电流更易通过紧固件导通。此外，在空气中，铝制金属连接件的表面极易形成一层致密的氧化膜，一般是电的不良导体，需要在装配前对该层氧化膜进行打磨；打磨后，还需要刷上一层特制的导电溶液，以提高结构连接处的导电性能，同时预防金属连接件在空气中再次被氧化。

为探究复合材料飞机蒙皮搭接结构内电流的流通路径，在2块延展金属箔的两侧施加1 A的直流电流源激励，在金属连接件的下底面添加接地条件，观测电流在整体结构内的电流流向，结果如图4所示。可知，施加的电流通过延展金属铜箔进入蒙皮，由于CFRP及树脂具有较低的导电率，箭头所示的内部电流较为集中地从金属铜箔通过螺栓进入金属连接件中。

当电流流经紧固件螺栓杆后，通过CFRP与金属连接件的接触面，进入金属连接件内部，由于金属连接件较为优良的导电性能，电流会在该结构与金属螺栓杆中共同导通，最终在金属连接件的下表面接地端流出。

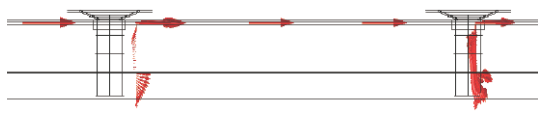


图4 螺栓附近电流流向

Fig. 4 Current flow direction near bolts

3 施工工艺对搭接结构电气性能影响分析

根据图1所示的复合材料飞机蒙皮典型搭接结构，考虑对蒙皮搭接结构产生较大电学性能影响的措施，选取在构建蒙皮结构时可以体现其中2个部位各自处理工艺的操作，探究施工工艺对复合材料飞机蒙皮阻抗性能和接触面电流变化趋势的影响情况。

3.1 紧固件锪窝处理程度对结构的电气影响

复合材料飞机蒙皮各部分结构在进行装配时，需预先在结构件上开孔，通过紧固件的旋拧实现整体结构的机械连接，紧固件由带螺纹的螺栓及凹形垫圈组成。凹形垫圈除了避免螺栓头对金属铜箔造成结构性磨损破坏外，还可以增大螺栓头与搭接结构的接触面积，提高该区域的电流传导性能。在对搭接结构进行装配时，需预先处理树脂等结构。对树脂表面的处理程度，会影响紧固件与主导电通路的接触面积大小，进而对搭接结构的电气性能产生影响。

如图5所示，对搭接结构的螺栓头附近区域进行打磨处理，使凹形垫圈恰可置于其中，从而既可以增大电流通路的导通面积，也不会在蒙皮结构的

上表面产生凸起影响。

树脂锪窝处理半径会影响凹形垫圈与金属铜箔的接触面积。考虑到锪窝打磨工具恒定, 即对锪窝的处理深度始终不变, 与凹形垫圈的高度一致。针对树脂打磨区域设置不同半径的布尔操作差集处理, 仿真计算树脂与延展金属铜箔接触面的电流密度分布情况, 选取锪窝处理半径为 9.5 mm 时情况, 结果如图 6 所示。

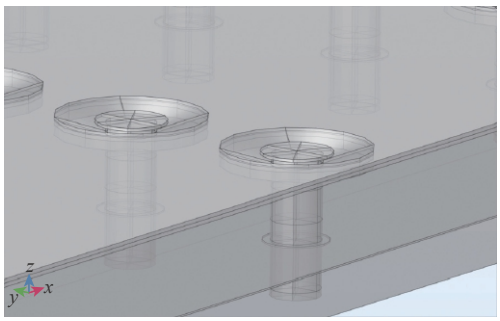


图 5　树脂表面的锪窝处理

Fig. 5　Countersink treatment of resin surface

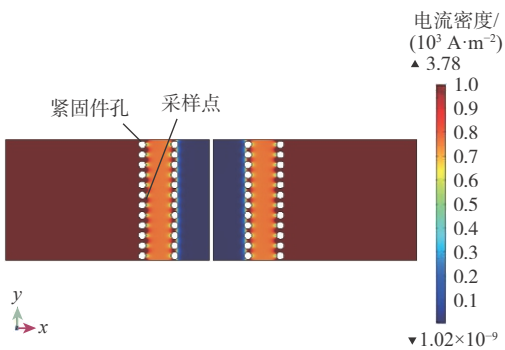


图 6　树脂-延展金属铜箔接触面电流密度分布

Fig. 6　Current density distribution on resin-extended meta copper foil contact surface

图 6 中, 4 列小孔代表紧固件穿过的区域。可知, 电流在流经紧固件前, 主要由金属铜箔传输, 此时电流密度较大; 在穿过外侧紧固件后, 大量电流通过紧固件流入金属连接件中, 在 2 排紧固件间, 电流密度迅速降低; 在经过第 2 排紧固件后, 已基本无电流通过金属铜箔流入 CFRP 中, 故表面电流密度较小。如图 6 所示, 在接触面取一采样点, 观察接触面上电流密度最大值及采样点处电流密度在不同处理半径下的变化情况, 结果如图 7 所示。

当锪窝处理半径为 9.5 mm 时, 凹形垫圈与金属铜箔接触面积最大, 电流更容易通过垫圈进入紧固件, 结合图 7 可知, 此时铜箔表面的电流密度极大值最小, 且采样点处电流密度模值也最小, 对铜箔的电流冲击损坏最弱。针对不同锪窝处理半径对搭接结构的整体阻抗影响结果进行仿真分析, 如图 8 所示。

当锪窝处理半径小于 9.5 mm 时, 由于凹形垫

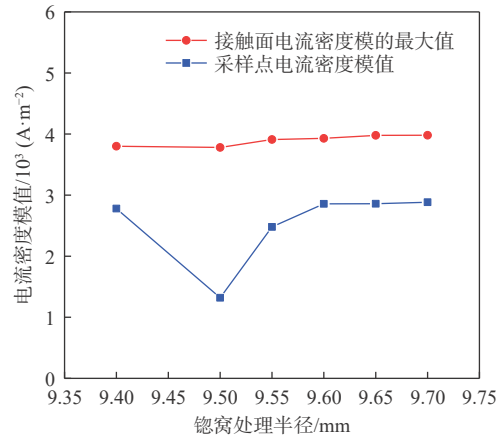


图 7　延展金属铜箔上表面电流密度变化趋势

Fig. 7　Variation trend of current density on upper surface of extended metal copper foil

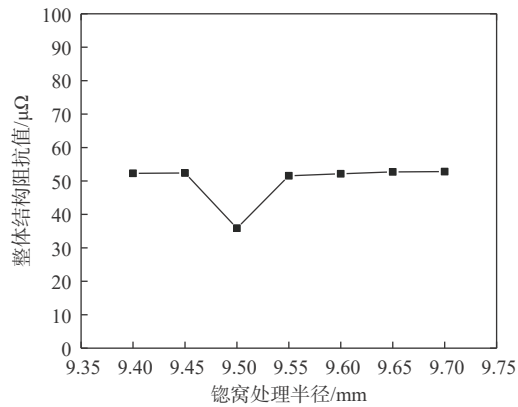


图 8　锪窝处理半径对结构阻抗的影响

Fig. 8　Influence of countersink radius on structural impedance

圈外直径为 19 mm, 垫圈无法放置于锪窝内, 使其架于树脂之上。此时, 凹形垫圈外侧面无法与延展金属铜箔接触, 阻抗值较大, 且突出的凹形垫圈易使树脂表面磨损, 破坏其完整性。随着锪窝处理半径的增大, 当凹形垫圈恰可放入锪窝中时, 此时垫圈与铜箔接触面积最大, 整体结构阻抗值最小, 且符合实际应用意义。随着锪窝处理半径的继续增大, 与凹形垫圈直接接触的金属铜箔也同时被打磨, 两者刚分离时, 初始阶段的打磨使接触面积减少最多, 其后接触面积减小不明显, 搭接结构阻抗值的变化趋势也验证了这一原理。

3.2　金属表面氧化膜处理对结构的电气特性影响

铝及其合金具有比重小、易加工、导电性能好等特点, 在复合材料飞机蒙皮搭接过程中, 一般充当金属连接件, 起到连接 2 块 CFRP 板材, 使两侧独立的导电通路连接为一个整体的作用。同时, 铝制连接件还通过接线端子实现与机载设备的连接, 承担机载设备短路电流的泄放功能。然而, 铝制品表面极易生成一层导电性能较差的氧化膜, 对金属连接件与 CFRP、紧固件等结构的接触面产生影响, 进

而影响蒙皮的导电性能。因此，在施工过程中，对该层氧化膜进行打磨处理，同时在打磨区域刷涂导电溶液，是提高蒙皮结构在该搭接区域电学性能的有效措施。

对金属连接件表面氧化膜的处理程度，包括处理的深度与处理范围，是后续导电溶液刷涂区域的基础。在现场施工中，一般通过碱性溶液的刷洗来去除金属连接件表面的氧化铝，处理溶液的浓度和刷涂范围会对处理结果产生影响。考虑到整体结构中金属连接件阻抗值较小，分得电压较多，对恒流源而言，其电压变化对整体结构的阻抗值影响更为明显，因此，通过整体结构阻抗值变化反映金属连接件的电压变化情况。实验仿真中，通过布尔操作减去金属连接件表面围绕螺杆四周的氧化铝，以减去区域的范围大小和深度作为变量，观察整体结构阻抗值的变化情况，分析金属连接件表面处理程度带来的影响，结果如图 9 所示。

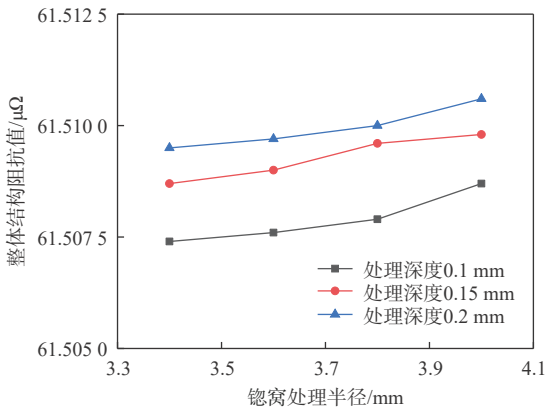


图 9 碱洗氧化膜对结构阻抗的影响

Fig. 9 Influence of oxide film after alkali wash on structural impedance

由图 9 可知，仅对金属连接件进行碱洗无法降低搭接结构的阻抗值。这是因为仅对氧化铝进行打磨而后续不继续刷涂导电溶液，只去除了导电性不良的介质，没有使螺杆与金属连接件间的易导电接触通道增多，电流仍主要通过螺杆进入金属连接件中。同时，导电性较差的介质虽导电性能不好，但仍可起到分担部分电流的作用，若仅打磨则会减小此部分电流导通通道，使整体结构阻抗增加。

对金属连接件的表面氧化膜碱洗后，还需在打磨区域刷涂一层导电溶液，该溶液主要由磷酸根和六价铬组成，一般还添加有氟离子，整体溶液的导电率为 4.98×10^5 S/m，刷涂该溶液的操作称为阿洛丁处理。对打磨后的金属表面进行不同深度及处理半径的阿洛丁处理后，整体搭接结构阻抗变化如图 10 所示。可知，碱洗区域进行阿洛丁处理后，整体结构的阻抗值显著降低。当处理半径为 3.2 mm 时，此时与螺杆半径一致，不同的处理深度对结构的阻抗变化并不明显；随着处理半径的增加，处理深度越大，阻抗值降低的趋势越明显。当导电溶液的刷涂深度恒定，随着处理半径的增大，整体搭接结构的阻抗值降低明显，但当处理半径大于 4 mm 后，阻抗值降低的速度随处理半径的增加逐渐降低。

在对金属连接件的氧化铝薄膜进行不同程度的碱洗及阿洛丁处理后，可基于 CFRP 与金属连接件氧化膜的接触面上电流密度变化趋势，探究施工工艺对搭接结构电学性能的影响情况。选取处理半径为 4 mm、处理深度为 0.2 mm 的打磨状态，观察接触面的电流分布情况，如图 11 所示。

图 11 中，经过阿洛丁处理的靠近紧固件穿孔区域电流密度模值较大。在碱洗和阿洛丁处理共同作用下，对不同深度和不同半径的阿洛丁处理区域电流密度模的最大值变化情况进行计算仿真，结果如图 12 所示。

由图 12 可知，随着碱洗及阿洛丁处理半径的增大，CFRP 与金属连接件的接触面上电流密度模的极大值逐渐升高，当处理半径达到 4 mm 后，模的最大值不再增加。结合图 10 可知，随着处理半径的增大，结构阻抗值显著降低，螺杆四周的阿洛丁处理表面相比于原有的氧化铝表面，电流受到阻碍降低，大量电流通过此处流入金属连接件中。同时，处理范围越大，电流导通通道越多，电流密度的模值越大。结合图 4，随着处理半径的继续增加，接

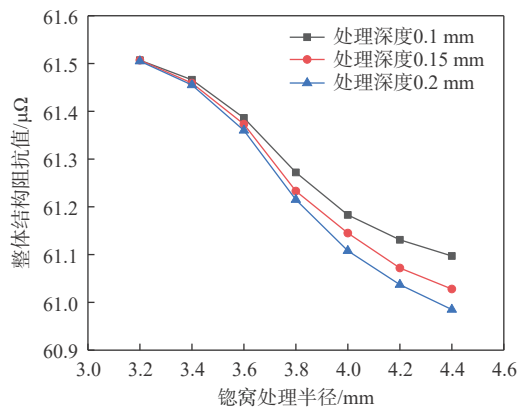


图 10 阿洛丁处理程度对结构阻抗的影响

Fig. 10 Influence of alodine treatment on structural impedance

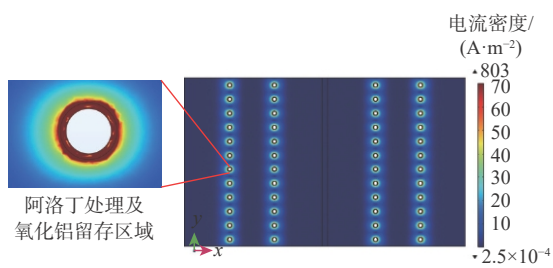


图 11 氧化膜接触面电流密度分布

Fig. 11 Current density distribution on contact surface of metal oxide film

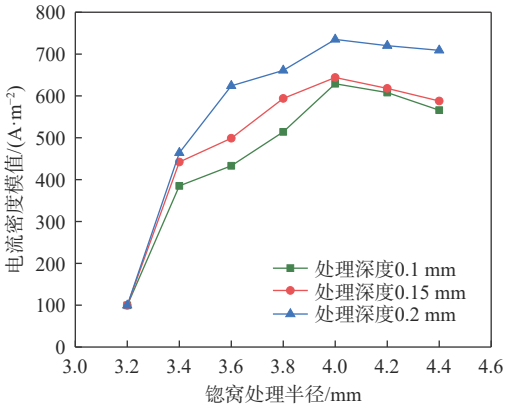


图 12　阿洛丁处理对氧化膜表面电流密度的影响

Fig. 12　Influence of alodine treatment on surface current density of oxide film

触面上大部分电流已从螺杆周围流过, 此时再增大阿洛丁处理区域半径, 对接触面上的电流影响较小。当处理深度不完全时, 阿洛丁溶液下方残留金属氧化铝会阻碍电流通过已处理表面进入金属连接件, 且处理深度越小, 残余的氧化铝越多, 对电流的阻碍作用越大, 接触面上电流密度模值也就越小。结合整体搭接结构的阻抗变化趋势, 对金属表面的处理深度应超过金属氧化膜的厚度, 处理范围应大于或等于 4 mm。

4　结　论

通过对复合材料飞机蒙皮搭接结构 2 个部位分别进行相对应的工艺处理, 分析其整体结构阻抗值、接触面电流密度模值情况, 得出以下结论:

1) 通过恰当的施工工艺处理, 能够有效降低结构间接触部分的阻抗, 减小接地阻抗对机载设备回流通路的影响。对树脂表面进行打磨能有效增大凹形垫圈与金属铜箔的接触面积, 当锪窝处理范围与凹形垫圈大小一致时, 对蒙皮结构的电气性能提升效果最好。

2) 仅对金属连接件的氧化膜进行碱洗操作, 蒙皮结构的电学性能受影响较小, 后续还需进行阿洛丁处理。阿洛丁处理越充分, 氧化膜去除越彻底, 蒙皮搭接结构的电学性能表现越优异; 阿洛丁处理区域半径增大, 可以有效提高复合材料飞机蒙皮的电气性能; 当处理半径达到一定程度 (4 mm) 后, 蒙皮搭接结构电气性能提升的程度有限。

参考文献 (References)

[1] 刘国春, 郭荣辉, 秦文峰. 民用飞机复合材料结构制造与维修[M]. 北京: 清华大学出版社, 2020: 10-11.
LIU G C, GUO R H, QIN W F. Manufacture and maintenance of composite structure of civil aircraft[M]. Beijing: Tsinghua University Press, 2020: 10-11(in Chinese).

[2] GOLEANU A, DUNAND M, FREYCON A, et al. Search for quasi-optimal current return paths in a composite environment[C]//Proceedings of the 10th International Symposium on Electromagnetic Compatibility. Piscataway: IEEE Press, 2011: 254-259.

[3] JONES C E, SZTYKIEL M, PEÑA-ALZOLA R, et al. Grounding topologies for resilient, integrated composite electrical power systems for future aircraft applications[C]//Proceedings of the AIAA/IEEE Electric Aircraft Technologies Symposium. Piscataway: IEEE Press, 2019: 1-15.

[4] US Air Force. Military specification, bonding, electrical and lightning protection for aerospace systems: MIL-STD-464A[S]. Washington, D. C. : US Department of Defense, 2002.

[5] GOLEANU A L, DUNAND M, GUICHON J M, et al. Towards the conception and optimisation of the current return path in a composite aircraft[C]//Proceedings of the IEEE International Systems Conference. Piscataway: IEEE Press, 2010: 466-471.

[6] 张宇. 复合材料飞机的电流回路接地技术研究[J]. 航空科学技术, 2011, 22(6): 27-30.
ZHANG Y. Research of current return path of composite aircraft[J]. Aeronautical Science & Technology, 2011, 22(6): 27-30(in Chinese).

[7] JONES C E, NORMAN P J, GALLOWAY S J, et al. Electrical model of carbon fibre reinforced polymers for the development of electrical protection systems for more-electric aircraft[C]//Proceedings of the 18th European Conference on Power Electronics and Applications. Piscataway: IEEE Press, 2016: 1-10.

[8] 易健宏, 杨平, 沈韬. 碳纳米管增强金属基复合材料电学性能研究进展[J]. 复合材料学报, 2016, 33(4): 689-703.
YI J H, YANG P, SHEN T. Research progress of electrical properties for carbon nanotubes reinforced metal matrix composites[J]. Acta Materiae Compositae Sinica, 2016, 33(4): 689-703(in Chinese).

[9] 郭永超, 高成, 宋双, 等. 不同激励源作用下的碳纤维复合材料的导电性能研究[C]//第 22 届全国电磁兼容学术会议. 北京: 人民邮电出版社, 2012: 62-66.
GUO Y C, GAO C, SONG S, et al. Study of electric properties of carbon fiber composite materials using different excitation[C]//Proceedings of the 22nd National Electromagnetic Compatibility Academic Conference. Beijing: Posts & Telecom Press, 2012: 62-66(in Chinese).

[10] LIN Z, SUN X, LIU S, et al. Effect of SnO_2 particle size on properties of Ag-SnO_2 electrical contact materials prepared by the reductive precipitation method[J]. Advanced Materials Research, 2014, 936: 459-463.

[11] WATTS P C P, HSU W K, BARNES A, et al. High permittivity from defective multiwalled carbon nanotubes in the X-band[J]. Advanced Materials, 2003, 15(7-8): 600-603.

[12] CHEN P W, CHUNG D D L. Concrete as a new strain/stress sensor[J]. Composites Part B: Engineering, 1996, 27(1): 11-23.

[13] CLINGERMAN M L, WEBER E H, KING J A, et al. Development of an additive equation for predicting the electrical conductivity of carbon-filled composites[J]. Journal of Applied Polymer Science, 2003, 88(9): 2280-2299.

[14] 张强, 蔡旭龙, 王建宏. 触点制造过程表面膜电阻对接触电阻的影响[J]. 机电元件, 2014, 34(2): 43-47.
ZHANG Q, CAI X L, WANG J H. Influence of surface film resist-

ance on contact resistance in contact manufacturing process[J]. Electromechanical Components, 2014, 34(2): 43-47 (in Chinese).

[15] MALINOWSKI P H, WANDOWSKI T, OSTACHOWICZ W M. Characterisation of CFRP adhesive bonds by electromechanical impedance[C]//Proceedings of the Health Monitoring of Structural and Biological Systems. Bellingham: SPIE, 2014: 906415.

[16] MALINOWSKI P H, OSTACHOWICZ W M, BRUNE K, et al. Study of electromechanical impedance changes caused by modifications of CFRP adhesive bonds[J]. Fatigue & Fracture of Engineering Materials & Structures, 2017, 40(10): 1592-1600.

[17] 周猛. CFRP 锪窝制孔尺寸及误差对连接结构疲劳性能的影响[D]. 天津: 天津大学, 2018: 4-5.
ZHOU M. CFRP countersink hole size and error influence of fatigue performance on the connection structure[D]. Tianjin: Tianjin University, 2018: 4-5(in Chinese).

[18] 王力强, 吕玲, 胡彦卿, 等. 阿洛丁 1200S 处理工艺参数对铝合金耐盐雾腐蚀性能的影响[J]. 腐蚀与防护, 2019, 40(12): 912-915.
WANG L Q, LÜ L, HU Y Q, et al. Effects of processing parameters in alodine 1200S treatment on salt spray corrosion resistance of Al alloy[J]. Corrosion & Protection, 2019, 40(12): 912-915(in Chinese).

[19] 刘建英, 隋政, 张起浩, 等. 复合材料飞机接地回流网络建模与阻抗分析[J]. 北京航空航天大学学报, 2021, 47(5): 885-893.
LIU J Y, SUI Z, ZHANG Q H, et al. Modeling and impedance analysis of composite material aircraft grounded return network[J]. Journal of Beijing University of Aeronautics and Astronautics, 2021, 47(5): 885-893(in Chinese).

[20] 杨占刚, 隋政, 张起浩, 等. 复合材料飞机接地回流网络网内压降分析[J]. 航空学报, 2022, 43(1): 324859.
YANG Z G, SUI Z, ZHANG Q H, et al. Voltage drop in composite aircraft grounding and current return network[J]. Acta Aeronautica et Astronautica Sinica, 2022, 43(1): 324859(in Chinese).

[21] HARRINGTON R F. Field computation by moment methods[M]. New York: Macmillan, 1968.

[22] 周琳, 黄江涛, 高正红. 基于离散伴随方程的三维雷达散射截面几何敏感度计算[J]. 航空学报, 2020, 41(5): 623361.
ZHOU L, HUANG J T, GAO Z H. Three dimensional radar cross section geometric sensitivity calculation based on discrete adjoint equation[J]. Acta Aeronautica et Astronautica Sinica, 2020, 41(5): 623361(in Chinese).

[23] BANDINELLI M, MORI A, GALGANI G, et al. A surface PEEC formulation for high-fidelity analysis of the current return networks in composite aircrafts[J]. IEEE Transactions on Electromagnetic Compatibility, 2015, 57(5): 1027-1036.

[24] 裴春雨. 碳纳米纤维/环氧树脂基复合材料自感知性能[D]. 大连: 大连理工大学, 2014: 12-14.
PEI C Y. Self-sensing properties of carbon nanofiber/epoxy resin matrix composites materials[D]. Dalian: Dalian University of Technology, 2014: 12-14(in Chinese).

Analysis of effect of construction process on electrical properties of composite skins

YANG Zhangang*, KE Zhongshu, YANG Xuwei, BAO Xingwang

(School of Electronic Information and Automation, Civil Aviation University of China, Tianjin 300300, China)

Abstract: Composite materials are increasingly used in modern commercial aircraft skin. However, the poor conductivity makes it necessary to have additional conductive structures to ensure the current flow in the skin. In order to improve the conductivity of the skin and reduce the influence of the bonding between the composite material and the conductive structure on the current path, the skin structure was modeled based on the electrical characteristics of each component structure, and the construction process that may affect the bonding was analyzed by using the analytical method. In view of the changes in the current in the contact surfaces and the overall structural impedance caused by different construction processes applied at different parts of the skin, the influence of the construction process of each structure during assembly on the electrical characteristics of the aircraft skin bonding structure was analyzed. Theoretical analysis and simulation results show that the appropriate construction process, especially the correct treatment of the resin and oxide film of metal connection on the skin surface, can effectively reduce the influence of overall structural impedance on the electrical performance of the skin and provide some operational suggestions in site-oriented construction.

Keywords: aircraft skin; composite material; bonding structure; construction process; electrical performance

Received: 2022-09-07; **Accepted**: 2022-10-07; **Published Online**: 2023-01-10 10:03

URL: link.cnki.net/urlid/11.2625.V.20230109.1005.001

Foundation items: Aeronautical Science Foundation of China (20182667010); Fund of National Engineering and Research Center for Commercial Aircraft Manufacturing (COMAC-SFGS-2021-609)

* **Corresponding author**. E-mail: yangcauc@163.com

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2022.0765

基于可能性理论的红外与可见光视频自适应分层融合算法

吉琳娜*，郭小铭，杨风暴
(中北大学 信息与通信工程学院，太原 030051)

摘 要：针对目前红外与可见光视频融合模型无法根据视频间差异特征动态调整融合策略，造成融合效果差甚至失效等问题，提出了一种基于可能性理论的红外与可见光视频的自适应分层融合算法。计算视频序列每帧中感兴趣区域的各类差异特征的幅值大小，得到每帧对应的主要差异特征；搭建分层融合框架，并确定各层变元，基于余弦相似性计算不同变元对各差异特征的融合有效度，利用可能性理论构造相应的融合有效度分布；逐层分析不同变元对各类差异特征的融合效果，选择出每层的最优变元；通过变元间的优化组合实现红外与可见光视频的自适应分层融合。实验结果表明：所提算法在保留红外典型目标和可见光结构细节等方面取得了显著的融合效果，并且在定量分析和定性评价上明显优于单一融合算法。

关 键 词：图像处理；视频融合；自适应分层融合；可能性理论；红外视频
中图分类号：TP391
文献标志码：A **文章编号**：1001-5965(2024)10-3021-11

红外与可见光视频传感器成像技术利用不同辐射物理特性、成像方式对场景目标本质特征及其差异进行探测，2类视频具有很强的互补性。将双路视频进行融合能够充分发挥不同成像性能优势，利于探测视频的特征综合和准确解译，是退化视觉环境的态势感知、出入口区域危险事件监测与预警、高速移动目标精准定位与智能避障等应用中对感兴趣目标进行识别的必要前提[1-2]。

目前，红外和可见光视频融合的相关算法主要分为2类：①静态单帧融合算法，一般基于多尺度变换[3]、神经网络[4-5]、稀疏表示[6]和混合式[7]等手段进行融合，实时性要求难以保证；②基于多帧时空信息的融合算法，利用时空能量匹配、时空结构张量、高阶奇异值分解等三维非分离变换[8-9]来实现，考虑连续帧间的空间几何和时间维度信息，可保持融合视频的时间稳定性和一致性。当前，对于2类视频融合，在融合策略选取上多依赖于视频的“先验知识”，然而在真实的目标探测中，由于探测场景的动态性、成像环境的多变性，尤其是视频序列帧内差异特征类型、幅值、频次等属性及其变化更为复杂，事先确定的融合模型难以满足“动态”变化差异特征的自适应融合需求，传统方法已成为制约2类视频融合效果提升的瓶颈。

研究发现，视频间差异特征的幅值及不同算法下各特征的融合有效程度具有非概率性和区间变化性[10-11]，通常是根据已有的、有限且相近的场景的融合策略进行预测和估计，因而也具有可能预测性和小样本性。而可能性分布可结合非样本空间的先验知识，对各类小样本不确定信息进行量化和表征，为融合有效度度量提供理论工具。因此，本

收稿日期：2022-09-07；**录用日期**：2022-11-19；**网络出版时间**：2022-12-14 17:01
网络出版地址：link.cnki.net/urlid/11.2625.V.20221213.1220.005
基金项目：国家自然科学基金 (61702465)；山西省基础研究计划 (201901D211238,202203021221104)；山西省高等学校科技创新项目 (2020L0264)；中北大学研究生科技立项 (2022180501)
* **通信作者**. E-mail：jlnnuc@163.com

引用格式：吉琳娜，郭小铭，杨风暴. 基于可能性理论的红外与可见光视频自适应分层融合算法[J]. 北京航空航天大学学报，2024，50(10)：3021-3031. JI L N，GUO X M，YANG F B. Adaptive layered fusion algorithm for infrared and visible video based on possibility theory[J]. Journal of Beijing University of Aeronautics and Astronautics，2024，50(10)：3021-3031 (in Chinese).

文针对现有融合模型无法根据视频间差异动态调整融合策略，提出一种基于可能性理论的红外和可见光视频自适应分层融合算法，根据可能性理论构造不同分层变元下差异特征的融合有效度分布，建立差异特征、融合策略、融合效果间的复杂不确定关系，从而有效发挥2类成像的互补优势，为提高红外与可见光视频的自适应融合性能提供一条新思路。

1 基于可能性理论的红外与可见光视频自适应分层融合

本文提出的基于可能性理论的红外与可见光视频自适应分层融合算法如图1所示，包括帧内主要差异特征的表征、融合有效度分布的构造和分层融合的实现等。

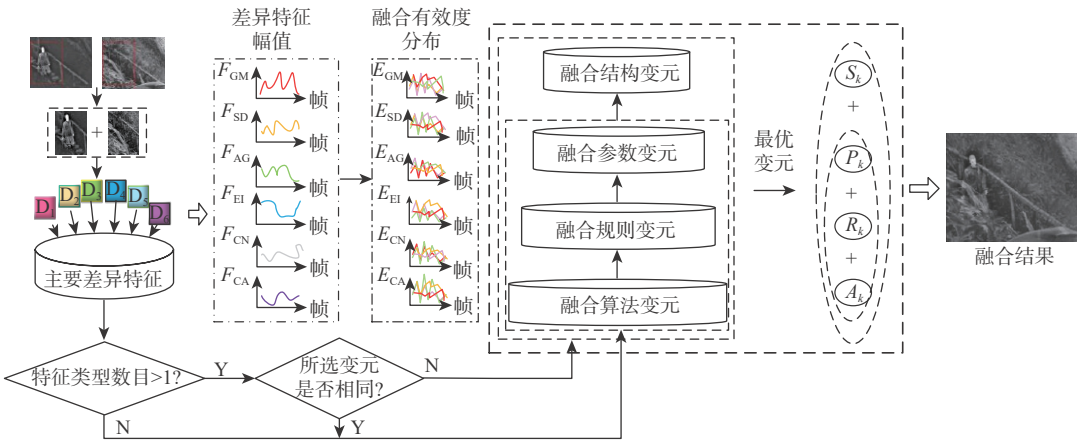


图1 自适应分层融合示意图

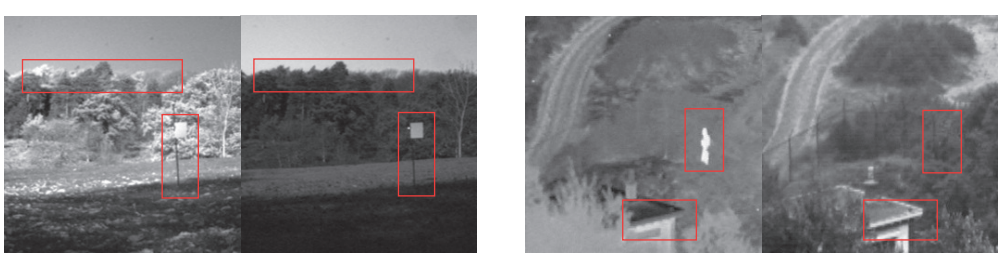
Fig. 1 Schematic of adaptive layered fusion

首先，根据融合需求及运动目标的检测识别任务，对红外与可见光视频序列中每帧的感兴趣区域进行大致划分，计算每帧中感兴趣区域的各类差异特征的幅值大小，得到每帧对应的主要差异特征；其次，搭建分层融合框架并确定各层变元，利用余弦相似性计算不同变元对各差异特征的融合有效度，进而构造相应的融合有效度分布，建立差异特征、融合策略、融合效果间的复杂不确定关系；然后，逐层分析不同变元对各类差异特征的融合效果，选择出各层的最优变元；最后，通过变元间的优化组合构成最佳融合策略，实现红外与可见光视频的自适应分层融合，从而解决现有模型无法根据视频间差异特征动态调整融合策略，造成融合效果差的问题。

1.1 帧内主要差异特征的表征与选取

结合红外与可见光传感器的成像机理及图像特点（见图2），红外视频一般包含较多的热目标信息，所呈现的轮廓较为明显，可见光视频可准确提供所在场景较丰富的细节信息，因此，2类视频的主要差异大多集中在亮度、边缘、纹理细节等3方面。为有效量化这3类差异，选择灰度均值（gray mean, GM）、边缘强度（edge intensity, EI）、标准差（standard deviation, SD）、平均梯度（average gradient, AG）、粗糙度（coarseness, CA）、对比度（contrast, CN）等6类特征来度量。其中，灰度均值量化表征亮度信息；标准差、平均梯度、边缘强度分别表示边缘信息中的反差度信息、边缘的清晰度、边缘幅值强度信息；粗糙度和对比度分别表征纹理信息的粗糙性和像素强度对比的整体布局[12]。

差异特征[13]表示2类视频对应帧内特征的绝对差异度，如下：



(a) 第1组 (b) 第2组

图2 红外与可见光序列图像

Fig. 2 Infrared and visible sequence images

$$F_{i,r}^{\mathrm{Diff}}=\left|F_{i,r}^{\mathrm{I}}-F_{i,r}^{\mathrm{V}}\right| \tag{1}$$

式中: $F_{i,r}^{\mathrm{I}}$、$F_{i,r}^{\mathrm{V}}$和$F_{i,r}^{\mathrm{Diff}}$分别为第 i 帧红外、可见光视频序列特征 r 和差异特征的幅值大小。

这 6 类差异特征的幅值具有不同的量纲, 为增强特征之间的可比性, 并便于后续视频各帧主要差异特征的选取, 本文对 2 类视频所提取的各类差异特征进行归一化处理(见式(2)), 即取值范围均为[0,1]。

$$F_{i,r}^{\mathrm{Diff}}=\frac{F_{i,r}^{\mathrm{Diff}}-\min\left(F_{1,r}^{\mathrm{Diff}},F_{2,r}^{\mathrm{Diff}},\cdots,F_{n,r}^{\mathrm{Diff}}\right)}{\max\left(F_{1,r}^{\mathrm{Diff}},F_{2,r}^{\mathrm{Diff}},\cdots,F_{n,r}^{\mathrm{Diff}}\right)-\min\left(F_{1,r}^{\mathrm{Diff}},F_{2,r}^{\mathrm{Diff}},\cdots,F_{n,r}^{\mathrm{Diff}}\right)} \tag{2}$$

1.2　基于可能性理论的融合有效度分布构造

融合有效度分布具体构造步骤如下:

步骤 1　利用 16×16 的平滑窗口对每帧红外与可见光视频同时进行不重叠分块处理, 得到每帧视频的一系列图像块$\{\mathrm{Patch}\}_{i=1:n}^{\mathrm{I}}$、$\{\mathrm{Patch}\}_{i=1:n}^{\mathrm{V}}$。

步骤 2　对于红外与可见光视频的每一图像块, 分别提取 6 类特征的幅值$F_{i,r=1:6}^{\mathrm{I}}$和$F_{i,r=1:6}^{\mathrm{V}}$。

步骤 3　利用余弦相似性及加权思想构造单一特征融合有效程度的评价函数, 以此来衡量分层融合中各类变元对某一特征的融合效果, 如下:

$$E_{i,r}^{k}=\frac{1}{N}\sum_{\xi=1}^{N}E_{i,r,\xi}^{k}=\frac{1}{N}\sum_{\xi=1}^{N}\left(w_{i,r,\xi}^{\mathrm{I}}\cdot\mathrm{SIM}\left(F_{i,r,\xi}^{k},F_{i,r,\xi}^{\mathrm{I}}\right)+w_{i,r,\xi}^{\mathrm{V}}\cdot\mathrm{SIM}\left(F_{i,r,\xi}^{k},F_{i,r,\xi}^{\mathrm{V}}\right)\right) \tag{3}$$

式中: $E_{i,r}^{k}$为融合有效度; $w_{i,r,\xi}^{\mathrm{I}}$和$w_{i,r,\xi}^{\mathrm{V}}$分别为红外与可见光视频序列帧的权重因子。

$$\begin{cases}\mathrm{SIM}\left(F_{i,r,\xi}^{k},F_{i,r,\xi}^{\mathrm{I}}\right)=\dfrac{\left\langle F_{i,r,\xi}^{k},F_{i,r,\xi}^{\mathrm{I}}\right\rangle}{\sqrt{\left(F_{i,r,\xi}^{k}\right)^{2}+\left(F_{i,r,\xi}^{\mathrm{I}}\right)^{2}}}\\ \mathrm{SIM}\left(F_{i,r,\xi}^{k},F_{i,r,\xi}^{\mathrm{V}}\right)=\dfrac{\left\langle F_{i,r,\xi}^{k},F_{i,r,\xi}^{\mathrm{V}}\right\rangle}{\sqrt{\left(F_{i,r,\xi}^{k}\right)^{2}+\left(F_{i,r,\xi}^{\mathrm{V}}\right)^{2}}}\end{cases} \tag{4}$$

$$\begin{cases}\left\langle F_{i,r,\xi}^{k},F_{i,r,\xi}^{\mathrm{I}}\right\rangle=\sum_{x=1}^{l}F_{i,r,\xi}^{k}(x)F_{i,r,\xi}^{\mathrm{I}}(x)\\ \left\langle F_{i,r,\xi}^{k},F_{i,r,\xi}^{\mathrm{V}}\right\rangle=\sum_{x=1}^{l}F_{i,r,\xi}^{k}(x)F_{i,r,\xi}^{\mathrm{V}}(x)\end{cases} \tag{5}$$

$$\begin{cases}w_{i,r,\xi}^{\mathrm{I}}=\dfrac{F_{i,r,\xi}^{\mathrm{I}}}{F_{i,r,\xi}^{\mathrm{I}}+F_{i,r,\xi}^{\mathrm{V}}}\\ w_{i,r,\xi}^{\mathrm{V}}=\dfrac{F_{i,r,\xi}^{\mathrm{V}}}{F_{i,r,\xi}^{\mathrm{I}}+F_{i,r,\xi}^{\mathrm{V}}}\end{cases} \tag{6}$$

评价函数值越大, 融合有效度越高。$F_{i,r,\xi}^{k}$表示变元 A_k 得到的视频序列第 i 帧融合结果对应像素块 ξ 的差异特征 r 的幅值, 进而可计算出不同变元下 6 类特征的融合有效度。

步骤 4　统计每帧视频各类差异特征的幅值区间$\left[0,\max(F_{i,r}^{\mathrm{Diff}})\right]$, 将其等分成 L 份, 统计每个幅值区间内包含的融合有效度散点数目$n_{i,r}^{k}$, 计算其总占比, 所得的值即为各区间的概率分布$P_{i,r}^{k}$, 利用比较转化法将概率分布转换成可能性分布[14], 如式(7)和式(8)所示, 即可得到不同变元对差异特征的融合有效度分布。

$$P_{i,r}^{k}=\frac{n_{i,r}^{k}}{N_{\mathrm{TO}}} \tag{7}$$

$$\pi_{i,r}^{k}=\frac{P_{i,r}^{k}}{\max(P_{i,r}^{k})} \tag{8}$$

式中: N_{TO}为幅值区间内总散点数目; $\pi_{i,r}^{k}$为融合有效度分布。

1.3　自适应分层融合

分层融合模型自底向上依次为融合算法变元集、融合规则变元集、融合参数变元集和融合结构变元集, 模型中各层变元既相互独立又互相关联, 各层集合中元素动态可变, 其属性及类型取决于当前融合需求。

自适应分层融合指在分析不同变元对主要差异特征融合效果的基础上, 即基于有效度函数值依次对融合算法、融合规则、融合参数和融合结构等进行筛选, 根据每一显著帧的主要差异特征自适应选择变元, 得到每层中的最优变元, 最终通过 4 类变元的变换组合得到融合结果的过程。具体流程如图 3 所示。

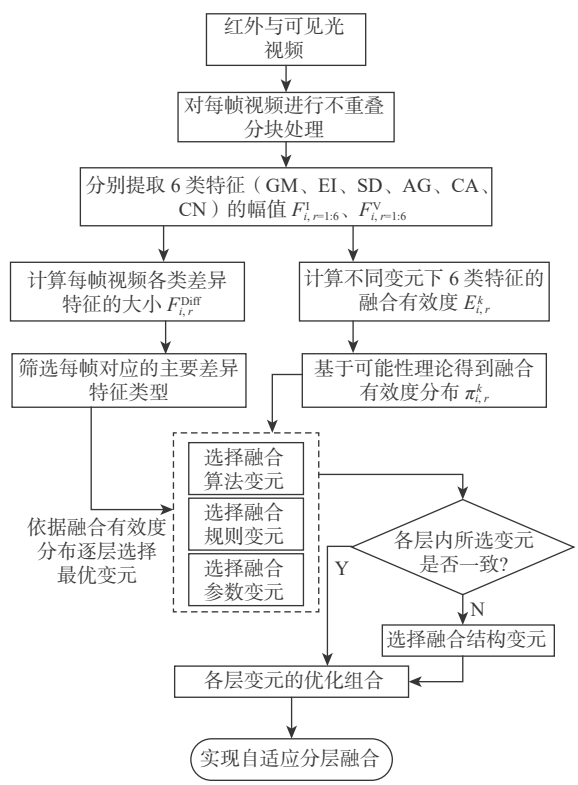


图 3　自适应分层融合算法流程

Fig. 3　Flowchart of adaptive layered fusion algorithm

下面对本文分层融合的变元进行介绍：

1）融合算法变元集。共选取多尺度融合框架中的7种算法，具体包括曲波变换（curvelet transform, CVT）、非下采样剪切波变换（non-subsampling shearlet transform, NSST）、非下采样轮廓波变换（non-subsampling contourlet transform, NSCT）、小波包变换（wavelet packet transform, WPT）、静态小波变换（static wavelet transform, SWT）、拉普拉斯金字塔（Laplace pyramid, LP）变换和双树复小波变换（dual tree complex wavelet transform, DTCWT）。变元中的元素可根据融合需求进行增删或调整。

2）融合规则变元集。多尺度融合框架下的融合规则主要分为高频规则和低频规则。其中，低频规则主要包括简单平均加权（simple average weighting, SAW）、系数取最大（maximum coefficient, MC）、基于窗口能量（window energy, WE）和窗口加权平均（window weighted average, WWA）；高频规则主要包括系数绝对值取最大（absolute maximum coefficient, AMC）、系数取最大、基于窗口能量、基于窗口梯度（window gradient, WG）、基于窗口标准差（window standard deviation, WSD）和改进的拉普拉斯能量和取大（maximum of improved Laplace total energy, MILTE）。高低频融合规则可以两两任意组合。

3）融合参数变元集。本文涉及的多尺度融合框架下的融合参数变元主要指多尺度分解层数及算法内部滤波器类型等，分解层数和滤波器均根据算法本身来设置。例如，NSCT和NSST采用非下采样金字塔和多方向分解，其中，金字塔滤波器选用pyrexc、pyr、maxflat、7-9，方向滤波器选用7-9、ideal、dmaxflat6等，分解层数可设置为{4}、{4,8}、{4,8,8}等；DTCWT第1层使用的滤波器和其他层并不相同，具体为：第1层使用的滤波器为antomini、legall、near_sym_a、near_sym_b，其他层使用的滤波器为qshift_06、qshift_a、qshift_b、qshift_c、qshift_d。

4）融合结构变元集。本文融合结构变元包括串联、并联、内嵌3种。①串联型：前一级的变元输出作为后一级变元的输入。②并联型：不同变元具有相同的输入，同时对图像进行融合，将不同融合结果组合获得最终融合结果。③内嵌型：不同变元层层嵌套获得最终融合结果。

2 实验仿真

2.1 源视频数据集

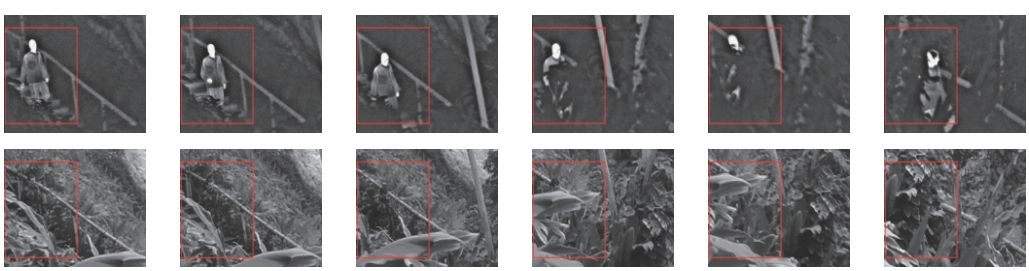
本文在2个公共数据集上分别进行了实验，以证明算法的优势和鲁棒性。①BEPMS数据集，涉及参与者观看一个穿着迷彩服的男人穿过茂密的树叶的视频，同时执行第2个按钮按下任务，由于篇幅受限，本文仅选择前200个对准的视频帧作为研究对象，图像大小为576像素×480像素。②OTCBVS数据集，包含17 089个不同场景的图像对，其中，200个图像对来自OSU Color-Thermal change视频。

2.2 实验案例

以BEPMS数据集所选红外与可见光视频为例，先对视频序列每帧中的感兴趣区域进行大致划分，并尽可能框出，如图4所示，展示了第1帧、第12帧、第50帧、第105帧、第185帧、第200帧；再利用式（1）分别计算每组视频帧中感兴趣区域的各差异特征的幅值，通过式（2）得到每帧对应的主要差异特征的归一化结果（见表1）；每帧视频的GM、SD、AG、EI、CN和CA等幅值最大值对应的差异特征即代表每帧主要差异特征（见表1中加粗数字），如第1帧的EI，第12帧的EI，第185帧的SD、AG和EI。

根据式（3）～式（8）分别构造不同变元（融合算法、融合规则、融合参数和融合结构）下各帧视频内不同差异特征对应的融合有效度分布函数，进而利用融合有效度函数值依次对融合算法、融合规则、融合参数、融合结构4类变元的融合性能进行排序，再根据表1中每帧对应的主要差异特征选择各类变元，通过不同变元的变换组合得到最终融合策略。

图5为不同融合算法变元对各帧视频6类差异



(a) 第1帧 (b) 第12帧 (c) 第50帧 (d) 第105帧 (e) 第185帧 (f) 第200帧

图4 BEPMS数据集中的红外与可见光视频的部分帧

Fig. 4 Some frames of infrared and visible video in BEPMS dataset

表 1　部分帧差异特征的归一化结果

Table 1　Normalized difference features of someframes

差异特征	视频序列					
	第1帧	第12帧	第50帧	第105帧	第185帧	第200帧
GM	0.580 1	0.242 6	**0.484 5**	**1.000 0**	0.319 7	0
SD	0.323 1	0.194 5	0.230 4	0.006 6	**1.000 0**	0.443 7
AG	0.209 7	0.040 3	0.134 4	0	**1.000 0**	0.442 9
EI	**0.595 0**	**0.356 3**	0.288 2	0.266 3	**1.000 0**	**0.597 3**
CN	0.516 5	0.243 1	0.373 5	0	0.568 3	0.448 1
CA	0.385 1	0	0.470 1	0.562 9	0.664 1	0.206 2

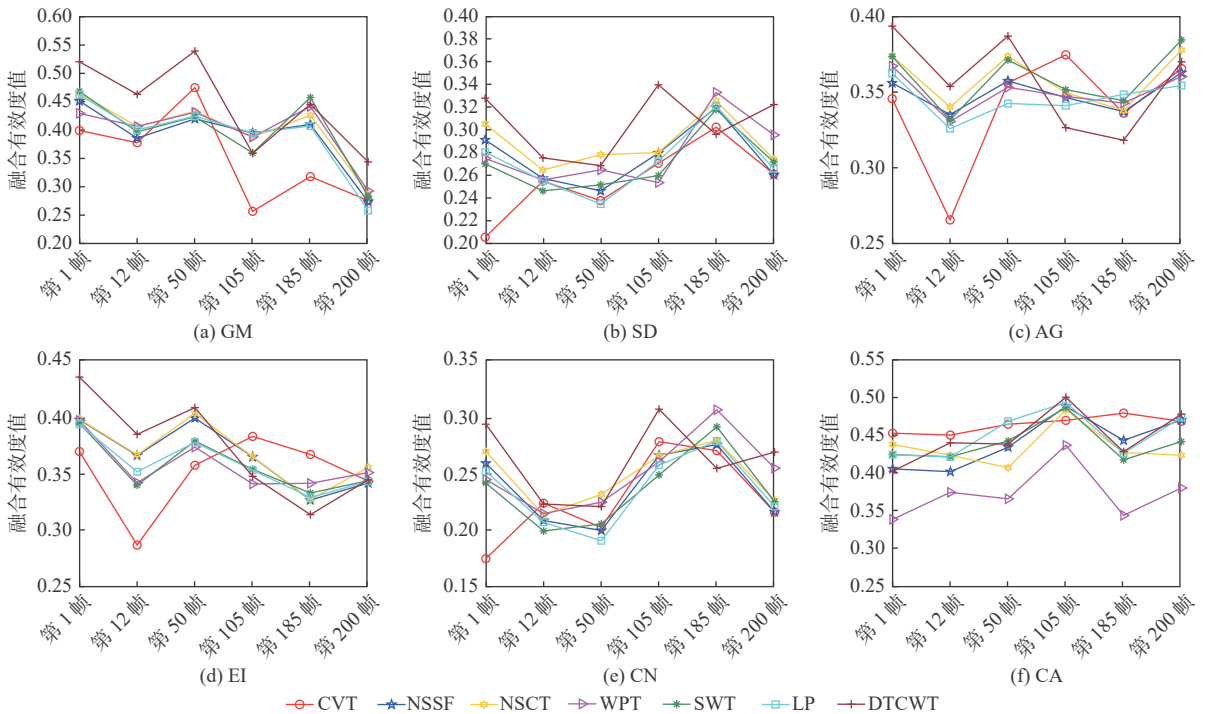


图 5　不同融合算法变元对差异特征的融合有效度分布

Fig. 5　Distribution of fusion effectiveness of different fusion algorithm variables for difference features

特征的融合有效度分布。例如, 第 1 帧对应的主要差异特征为 EI(见图 5(d)), 对应最优融合算法变元为 A_7, 同理可得第 12 帧、第 50 帧对应的最优算法变元均为 A_7, 第 105 帧对应 A_5, 第 200 帧对应 A_3 等, 具体结果如表 2 所示。

在各帧选出最优融合算法变元的基础上, 以不同低频融合规则、高频融合规则变元为变量, 根据融合有效度函数依次比较低频和高频融合规则变元对各差异特征的融合效果, 分析过程同上。由此可得出各帧对应的最优融合规则变元, 结果如表 3

表 2　各帧对应的最优融合算法变元

Table 2　Optimal fusion algorithm variables corresponding to each frame

视频序列	融合算法
第1帧	DTCWT
第12帧	DTCWT
第50帧	DTCWT
第105帧	SWT
第185帧	WPT/LP/CVT
第200帧	NSCT

表 3　各帧对应的最优融合规则变元

Table 3　Optimal fusion rule variables corresponding to each frame

视频序列	低频	高频
第1帧	SWA	WE
第12帧	SWA	AMC
第50帧	SWA	WE
第105帧	SWA	MC
第185帧	SWA	WG
第200帧	MC	AMC

所示。第1帧、第12帧、第50帧虽然对应的最优融合算法变元均为DTCWT,但最优融合规则变元却各有差异,其中,第1帧为SWA-WE,第12帧为SWA-AMC,第50帧为SWA-WE。

由于前3帧最优融合算法变元均为DTCWT,本文以该算法为例分析融合参数变元对各类差异特征的融合性能影响。对于DTCWT,其对应的融合参数是多尺度分解层数、第1层和其他层滤波器的类型。本文分解层数设置为1～7,滤波器类型具体见第2节。根据融合有效度函数值分析每帧的各类差异特征在不同分解层数、不同滤波器的融合结果,得到每帧对应的最优融合参数变元(见表4)。

表4 各帧对应的最优融合参数变元

Table 4 Optimal fusion parameter variables corresponding to each frame

视频序列	滤波器		分解层数
	第1层	其他层	
第1帧	near_sym_a	qshift_06	2
第12帧	near_sym_a	qshift_06	3
第50帧	antomini	qshift_a	4
第105帧	legall	qshift_a	1
第185帧	near_sym_a	qshift_06	3
第200帧	near_sym_a	qshift_06	1

串联、并联和内嵌等组合方式会形成不同的融合结构。由于视频序列某一帧中主要差异特征不止一种,且每种差异特征对应选择的最优融合变元不尽相同,就需要考虑几种变元的协同组合。例如,以上视频序列中,第185帧明显有3种主要差异特征,包括SD、AG和EI,对应选择的最优融合算法变元有WPT、LP和CVT,考虑这3种变元间的组合融合效果。3种融合结构变元实验如下:①串联型。用LP对2帧视频进行融合,其融合结果与红外视频帧采用CVT算法进行融合,融合结果与可见光视频帧采用WPT进行融合。②并联型。采用LP、CVT和WPT分别对2帧视频进行融合,将3种结果合成得到最终融合结果。③内嵌型。用LP对2帧视频进行分解,低频系数用CVT继续分解,高频系数用WPT进行分解,逐层重建得到最终融合结果。

图6为第185帧在串联、并联和内嵌3类融合结构变元下对不同差异特征的融合有效程度,由SD、AG和EI这3类主要差异特征的融合有效度值选出最优融合结构变元为串联型。

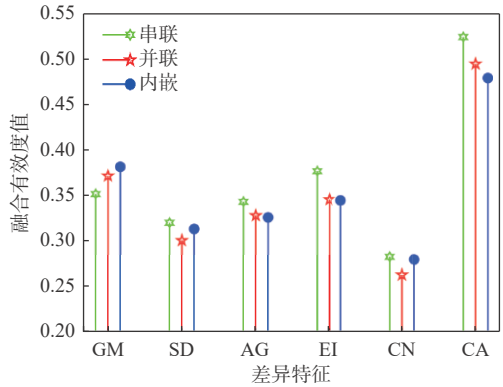


图6 不同融合结构变元的融合性能比较

Fig. 6 Comparison of fusion performance of different fusion structure variables

3 实验结果分析

为验证本文算法的合理性和有效性,将其与第2节所选的7种融合算法的融合结果相比较,部分融合算法的参数设置按照文献[15]所设,融合规则采用高频绝对值取最大,低频窗口加权平均,本文所有实验均在Windows 10的Intel (R) Core (TM) i5-5200U操作系统下运行。同时,由于主观评价易受评价者个人心理因素、精神状态等各方面的差异影响,存在一定的主观能动性,本文选择$Q^{AB/F}$、Q_0、Q_w、Q_e、互信息MI、VIFF、空间频率RCF这7个客观评价指标[16-17]对各算法融合结果进行评价。

对BEPMS和OTCBVS数据集进行定性和定量实验,从展示的几帧融合结果(见图7和图8)来看,典型的单一融合算法并不能在所有帧上均保持良好的融合效果,随着帧内、帧间差异信息的改变,行人的亮度、楼梯的边缘及树木的轮廓都会不可避免地丢失信息,本文算法则可以很好地保留行人亮度及场景细节信息,与其他算法相比,其具有更清晰的视觉效果,定性分析结果证明了其在保留典型热目标和丰富可见细节方面的优势。表5显示了BEPMS数据集的定量分析结果。本文算法在$Q^{AB/F}$、Q_0、Q_e、MI和VIFF这5个评价指标达到最佳值,在Q_w上位于第三,次于NSST和LP,在RCF上仅次于NSST,表明本文算法在融合图像中保留了丰富的有用特征信息,比其他算法具有更好的融合性能,这与上述定性分析一致。表6给出了OTCBVS数据集的定量实验结果。基于本文算法的融合结果在Q_w、Q_e、MI、VIFF和RCF等评价指标具有最大值,$Q^{AB/F}$和Q_0仅次于NSST和CVT。

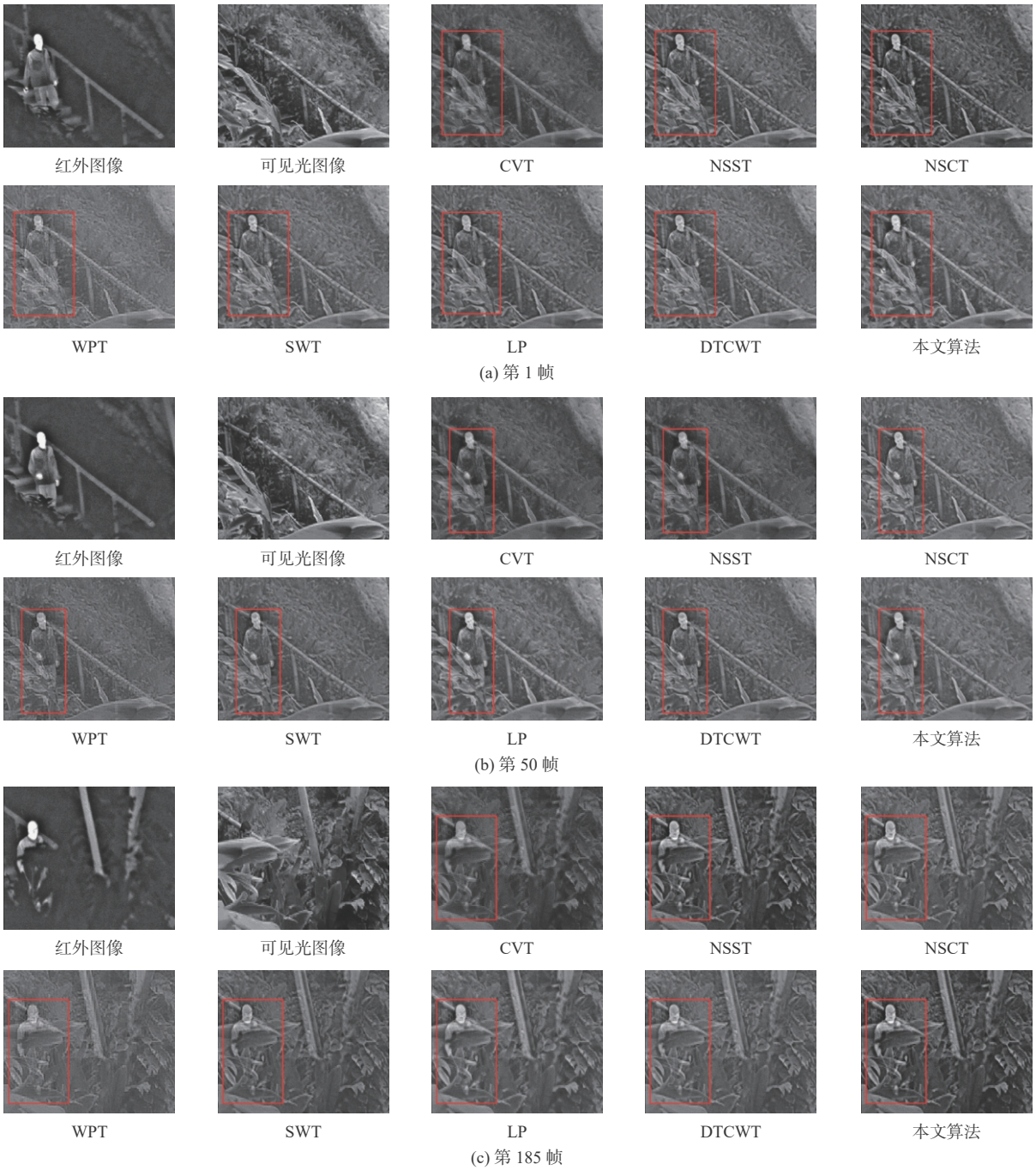


图 7　BEPMS 数据集中部分帧的融合结果

Fig. 7　Fusion results of some frames in BEPMS dataset

从图 9 明显看出, 本文算法在时间效率上也符合实时融合需求。

对整个视频序列来说, 视频相邻帧间图像通常具有很高的相似性, 因此, 可通过帧间图像内容的相似程度来度量相邻帧间的连续性。平均结构相似度 MSSIM 指 2 个相邻视频帧经过不重叠分块处理后, 计算所得的相应位置图像块的结构相似度的平均值, 该指标可有效反映 2 帧视频间画面内容的结构、对比度及亮度等信息的变化情况。因此, 本文利用 MSSIM 曲线的分布特点来判断融合后视频帧间是否连续。SSIM 将 2 幅图像的结构信息、对比度信息、亮度信息 3 个分量联合起来计算图像的相似性。计算过程如下: 首先, 利用 16×16 的平滑窗口对每帧红外与可见光视频同时进行不重叠分块处理, x 和 y 分别表示相邻帧对应的图像块; 然后, 分别计算 x 和 y 对应块之间的结构相似度; 最后, 将这些块之间的结构相似度取平均值, 即得到 2 帧之间的相似度 MSSIM。

x 和 y 对应的 SSIM 值表示为

$$\mathrm{SSIM}(x,y)=[l(x,y)]^{\alpha}[c(x,y)]^{\beta}[s(x,y)]^{\gamma} \tag{9}$$

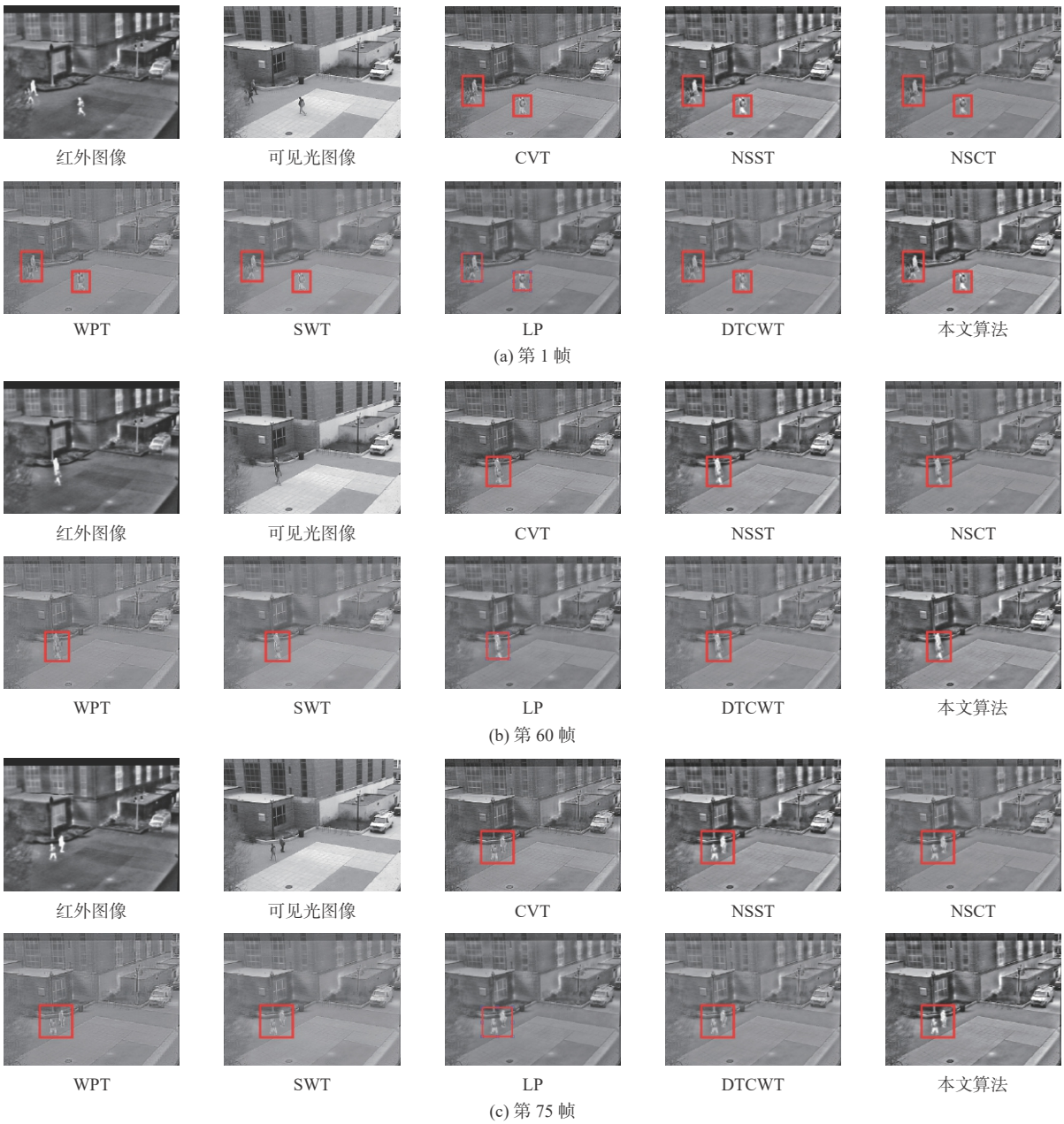


图 8 OTCBVS 数据集中部分帧的融合结果

Fig. 8 Fusion results of some frames in OTCBVS dataset

表 5 BEPMS 数据集各评价指标结果

Table 5 Results of each evaluation index in BEPMS dataset

融合算法	$Q^{AB/F}$	Q_0	Q_w	Q_e	MI	VIFF	RCF
CVT	0.343 5	0.547 4	0.705 0	0.152 7	1.428 2	0.346 2	14.433 3
NSST	0.440 5	0.561 5	0.810 0	0.225 7	1.423 1	0.434 9	14.535 0
NSCT	0.451 0	0.549 8	0.801 6	0.217 9	1.390 7	0.511 2	13.950 4
WPT	0.302 4	0.456 0	0.639 7	0.107 8	1.156 5	0.257 6	14.076 4
SWT	0.433 3	0.547 3	0.785 3	0.217 8	1.334 8	0.385 5	13.546 2
LP	0.447 2	0.552 1	0.803 5	0.206 4	1.468 0	0.477 2	14.222 4
DTCWT	0.411 3	0.571 2	0.784 7	0.205 2	1.324 4	0.387 8	13.664 7
本文算法	0.452 4	0.573 7	0.803 2	0.225 9	1.477 3	0.522 0	14.473 3

式中：α、β、γ一般取值为 1；$l(x,y)$、$c(x,y)$和$s(x,y)$分别为亮度比较函数、对比度比较函数和结构比较函数。

设μ_x、μ_y为图像的均值，σ_x^2、σ_y^2为图像的方差，

表 6　OTCBVS 数据集各评价指标结果

Table 6　Results of each evaluation index in OTCBVS dataset

融合算法	$Q^{AB/F}$	Q_0	Q_w	Q_e	MI	VIFF	RCF
CVT	0.438 1	0.493 1	0.665 8	0.230 7	2.013 9	0.270 6	27.424 3
NSST	0.476 2	0.487 9	0.723 0	0.266 2	2.222 6	0.342 2	27.407 3
NSCT	0.400 4	0.457 1	0.639 9	0.239 0	2.230 4	0.194 2	18.513 0
WPT	0.291 1	0.396 0	0.502 4	0.151 8	1.899 4	0.131 9	18.855 3
SWT	0.304 3	0.457 8	0.556 9	0.198 8	2.197 9	0.187 6	19.440 2
LP	0.368 9	0.460 1	0.623 5	0.235 0	2.032 9	0.200 9	18.443 4
DTCWT	0.336 2	0.428 3	0.596 8	0.223 1	1.885 0	0.155 8	27.009 6
本文算法	0.474 3	0.489 3	0.724 0	0.273 0	2.231 8	0.349 4	27.592 5

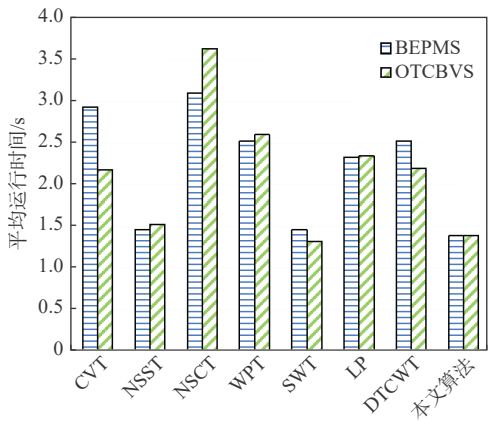


图 9　不同融合算法的运行时间

Fig. 9　Running time of different fusion algorithms

$\sigma_x\sigma_y$表示图像的协方差, C_1、C_2、C_3为值很小的常数, 则有

$$l(x,y)=\frac{2\mu_x\mu_y+C_1}{\mu_x^2+\mu_y^2+C_1} \tag{10}$$

$$c(x,y)=\frac{2\sigma_x\sigma_y+C_2}{\sigma_x^2+\sigma_y^2+C_2} \tag{11}$$

$$s(x,y)=\frac{\sigma_{xy}+C_3}{\sigma_x\sigma_y+C_3} \tag{12}$$

由式(10)～式(12)可以推出

$$\mathrm{SSIM}(x,y)=\frac{(2\mu_x\mu_y+C_1)(2\sigma_{xy}+C_2)}{(\mu_x^2+\mu_y^2+C_1)(\sigma_x^2+\sigma_y^2+C_2)} \tag{13}$$

则可计算出 2 帧之间的相似度 MSSIM, 如下:

$$\mathrm{MSSIM}(x,y)=\frac{1}{N}\sum_{i=1}^{N}\mathrm{SSIM}(x_i,y_i) \tag{14}$$

式中: N 为每帧视频划分的图像块个数。

从图 10 中看出, 2 个数据集融合后的视频序列的结构相似度值较高, 并在很小范围内上下波动, 没有出现异常点或突变点, 整体曲线较为平稳, 由此可以判断融合视频帧间图像是连续的。

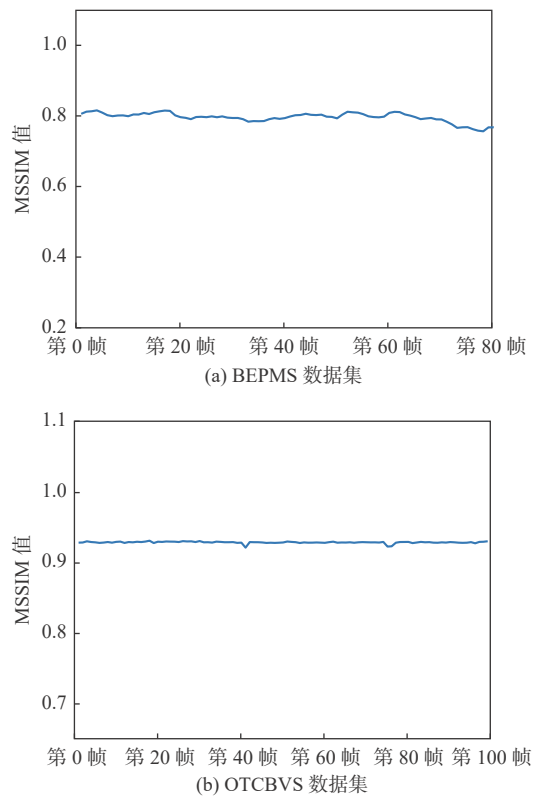


图 10　视频相似性检测结果

Fig. 10　Video similarity detection results

4　结　论

本文提出了一种基于可能性理论的红外与可见光视频自适应分层融合算法。

1) 可根据探测场景中差异信息属性的不同动态选择各层变元, 通过变元间的优化组合构成最佳融合策略, 实现模型的自适应分层融合。

2) 在 BEPMS 和 OTCBVS 数据集上的实验结果分析表明, 本文算法在视频整体上保留了红外典型目标和可见光结构细节, 融合性能明显优于其他单一融合算法。

3) MSSIM 曲线表明, 经本文算法融合后的视频帧结构相似度值较高, 波动范围很小, 整体曲线

较为平稳，即反映出融合视频间图像是连续的。

在未来的工作中，需要进一步研究模型的智能感知机制，以提高差异特征感知的准确性和敏感性，以便更好地为模型动态选择融合算法提供客观依据。

参考文献（References）

[1] 赵志刚, 王鑫, 彭廷海, 等. 国外中长波双波段红外成像技术的发展及应用[J]. 红外技术, 2020, 42(4): 312-319.
ZHAO Z G, WANG X, PENG T H, et al. Status quo and application of middle and long wave dual-band infrared imaging technologies in occident[J]. Infrared Technology, 2020, 42(4): 312-319(in Chinese).

[2] MA J Y, MA Y, LI C. Infrared and visible image fusion methods and applications: A survey[J]. Information Fusion, 2019, 45: 153-178.

[3] DUJ, LI W S, XIAO B, et al. Union Laplacian pyramid with multiple features for medical image fusion[J]. Neurocomputing, 2016, 194: 326-339.

[4] 唐超影, 浦世亮, 叶鹏钊, 等. 基于卷积神经网络的低照度可见光与近红外图像融合[J]. 光学学报, 2020, 40(16): 1610001.
TANG C Y, PU S L, YE P Z, et al. Fusion of low-illuminance visible and near-infrared images based on convolutional neural networks[J]. Acta Optica Sinica, 2020, 40(16): 1610001(in Chinese).

[5] LI H F, WANG Y T, YANG Z, et al. Discriminative dictionary learning-based multiple component decomposition for detail-preserving noisy image fusion[J]. IEEE Transactions on Instrumentation and Measurement, 2020, 69(4): 1082-1102.

[6] 王文卿, 马笑, 刘涵. 基于联合低秩稀疏分解的红外与可见光图像融合[J]. 信号处理, 2021, 37(9): 1770-1780.
WANG W Q, MA X, LIU H. Infrared and visible image fusion via joint low-rank and sparse decomposition[J]. Journal of Signal Processing, 2021, 37(9): 1770-1780(in Chinese).

[7] HU H M, WU J W, LI B, et al. An adaptive fusion algorithm for visible and infrared videos based on entropy and the cumulative distribution of gray levels[J]. IEEE Transactions on Multimedia, 2017, 19(12): 2706-2719.

[8] ZHANG Q, WANG Y B, LEVINE M D, et al. Multisensor video fusion based on higher order singular value decomposition[J]. Information Fusion, 2015, 24: 54-71.

[9] ZHANG Q, WANG L, MA Z K, et al. A novel video fusion framework using surfacelet transform[J]. Optics Communications, 2012, 285(13-14): 3032-3041.

[10] 郭小铭, 吉琳娜, 杨风暴. 基于可能性信息质量合成的双模态红外图像融合算法选取[J]. 光子学报, 2021, 50(3): 0310003.
GUO X M, JI L N, YANG F B. Dual-mode infrared image fusion algorithm selection based on possibility information quality synthesis[J]. Acta Photonica Sinica, 2021, 50(3): 0310003(in Chinese).

[11] JI L N, YANG F B, GUO X M. Set-valued mapping cloud model and its application for fusion algorithm selection of dual mode infrared images[J]. IEEE Access, 2021, 9: 54338-54349.

[12] 杨风暴. 红外偏振与光强图像的拟态融合原理和模型研究[J]. 中北大学学报 (自然科学版), 2017, 38(1): 1-8.
YANG F B. Research on theory and model of mimic fusion between infrared polarization and intensity images[J]. Journal of North University of China (Natural Science Edition), 2017, 38(1): 1-8(in Chinese).

[13] 张雷, 杨风暴, 吉琳娜. 差异特征指数测度的红外偏振与光强图像多算法融合[J]. 火力与指挥控制, 2018, 43(2): 49-54.
ZHANG L, YANG F B, JI L N. Multi algorithm fusion of infrared polarization and intensity image based on difference feature exponential measure[J]. Fire Control & Command Control, 2018, 43(2): 49-54(in Chinese).

[14] JI L N, YANG F B, GUO X M. Image fusion algorithm selection based on fusion validity distribution combination of difference features[J]. Electronics, 2021, 10(15): 1752.

[15] 杨风暴, 吉琳娜, 王肖霞. 可能性理论及应用[M]. 北京: 科学出版社, 2019: 41-45.
YANG F B, JI L N, WANG X X. Possibility theory and application[M]. Beijing: Science Press, 2019: 41-45(in Chinese).

[16] LI S T, YANG B, HU J W. Performance comparison of different multi-resolution transforms for image fusion[J]. Information Fusion, 2011, 12(2): 74-84.

[17] HAN Y, CAI Y Z, CAO Y, et al. A new image fusion performance metric based on visual information fidelity[J]. Information Fusion, 2013, 14(2): 127-135.

Adaptive layered fusion algorithm for infrared and visible video based on possibility theory

JI Linna*, GUO Xiaoming, YANG Fengbao

(School of Information and Communication Engineering, North University of China, Taiyuan 030051, China)

Abstract: The current infrared and visible video fusion model cannot dynamically adjust the fusion strategy according to the difference between videos, resulting in poor fusion effect or even failure. To address this issue,an adaptive layered fusion algorithm for infrared and visible video based on possibility theory was proposed. First, the magnitudes of various difference features of the region of interest in each frame of the video sequence were calculated, and the main difference features corresponding to each frame were obtained. Secondly, a layered fusion framework was built to determine the variables of each layer. The fusion effectiveness of different variables for each difference feature was calculated based on cosine similarity, and the possibility theory was used to construct the corresponding fusion effectiveness distribution.Then, the fusion effect of different variables for various difference features was analyzed layer by layer, and the optimal variable of each layer was selected. Finally, the adaptive layered fusion of infrared and visible video was realized through the optimal combination of variables. The experimental results show that the method in this paper has achieved remarkable fusion results in preserving typical infrared targets and visible structural details, and it is superior to other single fusion methods in quantitative analysis and qualitative evaluation.

Keywords: image processing; video fusion; adaptive layered fusion; possibility theory; infrared video

Received: 2022-09-07; **Accepted**: 2022-11-19; **Published Online**: 2022-12-14 17:01

URL: link.cnki.net/urlid/11.2625.V.20221213.1220.005

Foundation items: National Natural Science Foundation of China (61702465); Fundamental Research Program of Shanxi Province (201901D211238, 202203021221104); Scientific and Technological Innovation Programs of Higher Education Institutions in Shanxi (2020L0264); Postgraduate Science and Technology Project of the North University of China (2022180501)

* **Corresponding author**. E-mail: jlnnuc@163.com

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2022.0770

边防哨所风光耦合制氢系统的配置优化

齐海涛[1,2,*]，刘咄[1]，赵东澳[1]，刘旭[1]

(1. 北京航空航天大学 机械工程及自动化学院，北京 100191；2. 北京航空航天大学 工程训练中心，北京 100191)

摘 要：为解决边防哨所地处偏远、环境复杂、对能源需求大且上网成本高的问题，利用风光资源的互补性及氢能的能量密度高、绿色清洁等特点，提出了建立以风能和太阳能作为主要发电能源的风光耦合制氢系统，并选择氢能单元作为系统的储能单元，以满足边防哨所对能源多样性的需求，提高哨所从周边环境获取能量的能力。建立风光耦合制氢系统各子系统数学模型，以减小系统成本、提高可再生能源利用率、增大氢能产量为目标进行系统建设配置优化，选择系统各组件的额定功率作为设计变量，具体地区气象数据作为输入，利用第三代非支配排序遗传算法(NSGA-Ⅲ)进行求解，对优化得到的帕累托解集进行决策。通过层次图的绘制实现多维帕累托前沿可视化，得到了最优配置方案，并与未优化配置方案等进行对比，保障供电可靠性的同时可大幅度提高能源利用率、降低配置成本。

关 键 词：风力发电；光伏发电；电解制氢；混合能源系统；配置优化
中图分类号：V221$^+$.3; TB553
文献标志码：A **文章编号**：1001-5965(2024)10-3032-10

随着可再生能源技术的快速发展，风、光资源因具有分布广、开发潜力大和对环境影响小等特点得到广泛应用，且风光耦合能源系统被验证比单一可再生能源系统发电效率更高，稳定性更好。同时，又因风、光能源的波动性和间歇性，混合能源系统需要配备储能装置减小其出力波动。氢能作为一种新型绿色清洁能源，能量密度高且便于储存和运输，是风光耦合系统储能载体的优选方案之一。搭建风光耦合制氢系统可有效结合边防地区的周边环境，通过不同类型能源间互联互通、高效转化，解决边防哨所供电供能问题，有效减小能源前送的压力[1]。

目前，国外已建立较多风光氢混合发电系统示范工程，如美国的Wind2H2示范工程和英国的Hydrogen Power Park项目等。国内也有相关的示范工程，如沽源风电制氢项目和张北风电制氢及燃料电池发电系统等[2]。

在风光耦合制氢系统的设计规划阶段，系统的容量配置是构建系统的关键，不仅提供了系统选型方案，也是系统能够高质量跟踪负荷需求、提高可再生能源的利用率、获得更好经济效益的基础。文献[3]总结了离网型风光互补发电系统的几种配置方法，提出将迭代和智能算法与配置问题相结合具有更高的精度；文献[4]基于周平均气象数据，以电解水制氢为目的搭建风光混合能源系统，并以最大氢产量为优化目标，利用帝国竞争优化算法进行配置优化；文献[5]采用ε-constraint方法研究分析了不同优化准则组合下风光氢系统配置优化结果，认为多准则更能实现降低电能损耗的目的；文献[6]结合实时电价，以系统预期成本和预期收益为优化目标对风光氢供电系统进行了配置优化；文献[7]除考虑系统安装成本和负荷缺电率，还以系统输出功率波动率为优化目标，采用第二代非支配排序遗传算法(nondominated sorting genetic algorithm-Ⅱ，

收稿日期：2022-09-14；录用日期：2023-02-17；网络出版时间：2023-03-24 16:59
网络出版地址：link.cnki.net/urlid/11.2625.V.20230323.1636.009
*通信作者. E-mail：qihaitao@buaa.edu.cn

引用格式：齐海涛，刘咄，赵东澳，等. 边防哨所风光耦合制氢系统的配置优化[J]. 北京航空航天大学学报，2024，50(10)：3032-3041.
QI H T，LIU D，ZHAO D A，et al. Configuration optimization of wind/photovoltaic hydrogen production system at frontier sentries[J]. Journal of Beijing University of Aeronautics and Astronautics，2024，50(10)：3032-3041 (in Chinese).

NSGA-Ⅱ) 对风氢系统进行配置优化; 文献 [8] 对比了风光氢储、风光储、光氢储和光储 4 种系统方案下的配置优化结果。

以上文献大多先选定风电、光伏和电解槽等各组件的额定功率, 再以各组件的个数为设计变量, 并选择蓄电池作为优先的储能单元, 多应用于大型的风-光-电解氢混合能源系统。本文针对边防哨所上网成本高和能源需求多样的问题, 综合考虑系统建设经济性、供电可靠性和氢能产出, 将系统年均成本、弃风弃光率和制氢量作为优化目标, 系统负荷缺额率作为保证系统可靠性的主要指标, 选择氢能为主要储能载体, 电解槽作为可控负荷, 结合燃料电池协调控制, 以提高风电、光伏发电品质, 并结合某边防地区实际气象小时平均数据进行配置优化, 求解出最优的耦合系统配置组合。

1　风光耦合制氢系统结构

1.1　系统结构

可再生能源系统多分为并网型和离网型。并网型系统与电网相连, 可借助电网能量平抑风光出力波动; 离网型系统虽缺少电网支撑, 但更适合上网成本高的偏远地区, 如海岛、高原基站等。

据此, 本文所研究的针对边防哨所的风光耦合制氢系统是基于直流微网的离网型多能源综合系统, 包括风电机组、光伏阵列、电解水制氢装置、燃料电池、蓄电池和功率变换器等组件。风电机组和光伏阵列作为系统的能源部分, 制氢装置、燃料电池和储氢罐构成用电—制氢—耗氢—发电的闭环储能单元, 蓄电池作为系统的短时储能设备, 在燃料电池开启等时刻短时放电维持电压和功率平衡, 系统结构如图 1 所示[9]。

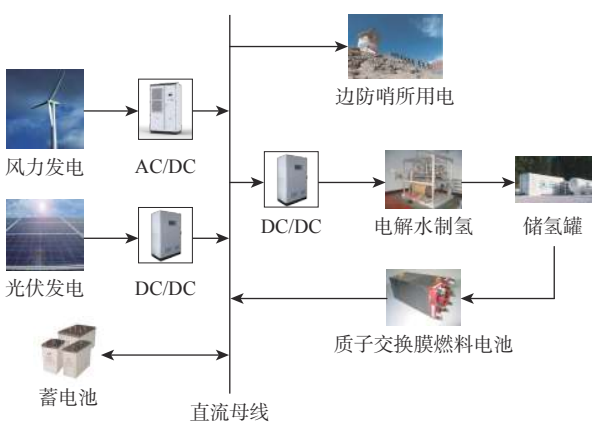


图 1　风光耦合制氢系统结构[9]

Fig. 1　Structure of wind/photovoltaic hydrogen production system[9]

1.2　各子系统数学模型

实际工作时, 风电机组的输出功率取决于当前风速[10-12], 可表示为

$$P_{\mathrm{wt}}(v)=\begin{cases}0 & v>v_{\mathrm{o}},v<v_{\mathrm{i}}\\ P_{\mathrm{wt}}^{\mathrm{N}}\dfrac{v-v_{\mathrm{i}}}{v_{\mathrm{R}}-v_{\mathrm{i}}} & v_{\mathrm{i}}\leqslant v\leqslant v_{\mathrm{R}}\\ P_{\mathrm{wt}}^{\mathrm{N}} & v_{\mathrm{R}}<v\leqslant v_{\mathrm{o}}\end{cases}\tag{1}$$

式中: $P_{\mathrm{wt}}^{\mathrm{N}}$为风力发电机的额定功率; v 为实际风速; v_{i} 为风力机的切入风速, 一般为 2.5～3.5 m/s; v_{o} 为风力机的切出风速, 一般为 20～25 m/s; v_{R} 为额定风速。

光伏电池组件的输出功率与光照强度和温度有关[10,13-14], 可表示为

$$P_{\mathrm{pv}}=P_{\mathrm{pv}}^{\mathrm{N}}f\left(\frac{G_{\mathrm{r}}}{G_{\mathrm{ref}}}\right)\left[1+\alpha_{\mathrm{p}}(T_{\mathrm{c}}-T_{\mathrm{ref}})\right]\tag{2}$$

$$T_{\mathrm{c}}=T_{\mathrm{ae}}+\lambda G_{\mathrm{r}}\tag{3}$$

式中: $P_{\mathrm{pv}}^{\mathrm{N}}$为光伏电池额定输出功率; G_{r} 为当前的光照强度; G_{ref} 为 25℃ 时的参考光照强度; f 为系统发电效率; α_{p} 为功率温度系数; T_{c} 为光伏面板温度; T_{ref} 为标准环境温度; T_{ae} 为实际环境温度; λ 为辐射温度系数。

质子交换膜(proton exchange membrane, PEM)电解槽工作电流密度高, 质量轻且结构紧凑, 响应速度快, 功率调节范围大, 更适配具有波动性和随机性的可再生能源发电系统。目前, PEM 电解槽电耗量为 3.8～5.0 kW·h/Nm^3。电解槽通过电解水制取氢气, 其制取氢气的输出功率可表示为

$$P_{\mathrm{ele,H_2}}=P_{\mathrm{ele}}\eta_{\mathrm{ele}}\tag{4}$$

式中: P_{ele} 为电解槽输入电功率; η_{ele} 为电解槽制氢速率。

燃料电池实际上是一个发电系统, 对于 PEM 燃料电池, 需要有燃料供应系统、氧化剂系统、发电系统、水管理系统、热管理系统、电力系统及控制系统等。燃料电池的输出电功率可表示为

$$P_{\mathrm{fc}}=P_{\mathrm{fc,H_2}}\eta_{\mathrm{fc}}\tag{5}$$

式中: $P_{\mathrm{fc,H_2}}$ 为储氢罐到燃料电池的氢气功率; η_{fc} 为燃料电池的转换效率。

电解槽能吸纳的最大电功率和燃料电池能输出的最大电功率都与当前储氢罐的容量有关。因此, 电解槽能吸纳的最大电功率计算为

$$P_{\mathrm{ele}}^{\max}=\min\left\{P_{\mathrm{ele}}^{\mathrm{N}},\frac{E_{\mathrm{tank}}^{\max}-E_{\mathrm{tank}}(t)}{\Delta t\cdot\eta_{\mathrm{ele}}}\right\}\tag{6}$$

燃料电池能输出的最大电功率为

$$P_{\mathrm{fc}}^{\max}=\min\left\{P_{\mathrm{fc}}^{\mathrm{N}},\frac{E_{\mathrm{tank}}(t)-E_{\mathrm{tank}}^{\min}}{\Delta t}\eta_{\mathrm{fc}}\right\}\tag{7}$$

式中: $P_{\mathrm{ele}}^{\mathrm{N}}$和$P_{\mathrm{fc}}^{\mathrm{N}}$分别为电解槽和燃料电池的额定功

率；$E_{\text{tank}}(t)$ 为 t 时刻储氢罐的储氢能量；$E_{\text{tank}}^{\max}$、$E_{\text{tank}}^{\min}$ 分别为储氢罐的最大、最小储氢能量；Δt 为采样步长。储氢罐的储存体积可以通过电解槽的电耗量来计算。

现有储氢技术包括压缩气体储存、低温压缩液态储存、金属氧化物储存和地下储存等，因为技术问题和系统经济性考量，选择高压气罐作为储氢方案[15]。某一时刻，储氢罐的净进气率为

$$\dot{n}_{\text{sto}}(t)=\dot{n}_{\text{in,H}_2}(t)-\dot{n}_{\text{out,H}_2}(t) \tag{8}$$

式中：$\dot{n}_{\text{in,H}_2}(t)$、$\dot{n}_{\text{out,H}_2}(t)$分别为该时刻储氢罐进、出气速率。

由式 (8) 可得该时刻储氢量为

$$n_{\text{sto}}(t_0+\Delta t)=\int_{t_0}^{t_0+\Delta t}\dot{n}_{\text{sto}}(t)\mathrm{d}t+n_{\text{sto}}(t_0) \tag{9}$$

储氢罐消耗功率为

$$P_{\text{tank}}=C_{\text{p,H}_2}\frac{T_{\text{tank}}}{\eta_{\text{cp}}}\left[\left(\frac{P_2}{P_1}\right)\frac{k-1}{k}-1\right]W_{\text{cp}} \tag{10}$$

式中：$C_{\text{p,H}_2}$为氢气比热容；T_{tank} 为储氢罐进气温度；η_{cp} 为压缩机效率；P_1、P_2 分别为进、出口压强；k 为标况下氢气比热比；W_{cp} 为压缩机出口流速。

蓄电池的荷电状态指当前蓄电池的电量和充满未使用时电量的比值。在蓄电池充放电过程中，其荷电状态会在限定区间内上升和下降，有如下蓄电池充放电输出功率模型：

$$\begin{cases}C_{\text{SOC}}^{1}(t)=(1-\eta_{\text{b}})C_{\text{SOC}}(t-1)+\dfrac{P_{\text{c}}(t)\cdot\Delta t\cdot\eta_{\text{c}}}{E_{\text{b}}}\\ C_{\text{SOC}}^{2}(t)=(1-\eta_{\text{b}})C_{\text{SOC}}(t-1)-\dfrac{P_{\text{d}}(t)\cdot\Delta t}{\eta_{\text{d}}E_{\text{b}}}\end{cases} \tag{11}$$

式中：η_{b}、η_{c} 和 η_{d} 分别为蓄电池的自放电率、充电效率和放电效率；$P_{\text{c}}(t)$ 和 $P_{\text{d}}(t)$ 分别为蓄电池在 t 时刻的充电功率和放电功率；E_{b} 为蓄电池额定容量。蓄电池的充放电次数与充放电功率是影响蓄电池寿命的主要因素，为保证系统长期稳定运行，需对蓄电池充放电过程进行一定约束。

蓄电池在 t 时刻允许充放电功率为

$$\begin{cases}0\leqslant P_{\text{c}}(t)\leqslant\min\left\{P_{\text{c}}^{\max},\dfrac{E_{\text{b}}\left(C_{\text{SOC}}^{\max}-(1-\eta_{\text{b}})C_{\text{SOC}}(t-1)\right)}{\eta_{\text{c}}\cdot\Delta t}\right\}\\ 0\leqslant P_{\text{d}}(t)\leqslant\min\left\{P_{\text{d}}^{\max},\dfrac{E_{\text{b}}\left[(1-\eta_{\text{b}})C_{\text{SOC}}(t-1)-C_{\text{SOC}}^{\min}\right]\eta_{\text{d}}}{\Delta t}\right\}\end{cases} \tag{12}$$

式中：$P_{\text{c}}^{\max}$、$P_{\text{d}}^{\max}$分别为蓄电池允许最大持续充、放电功率。

1.3 系统运行策略

由于风能和太阳能具有波动性和随机性，其工作状态受气象环境影响很大，需先制定系统的能量控制策略，以确定当前状态系统可提供电能和系统所需电能。风力发电和光伏阵列发出的电能供给负载，多余的能量优先供给电解槽制氢，若超出电解槽额定功率且蓄电池荷电状态未达到最大值，则蓄电池充电，否则产生弃风弃光；如果风力发电机和光伏阵列发出的电能不能满足负载功率，优先选择燃料电池发电补充功率缺额，其次考虑蓄电池放电。

根据图 2 流程调度将系统运行模式分为如下 5 种：

1）风光-制氢-储氢模式，即风光发电功率可以满足负荷需求，储氢量未达上限，剩余功率能够完全被电解槽吸纳制氢，或净功率超出电解槽最大工作功率，蓄电池荷电状态已达上限，产生弃风弃光。

2）风光-制氢-储氢-蓄电池模式，即当风光发电功率可以满足负荷需求，储氢量未达上限，电解槽以最大功率运行，且蓄电池荷电状态未达到上限充电，若可调度充电功率超出蓄电池允许最大持续充电功率，则产生弃风弃光。

3）风光-燃料电池模式，即当风光发电无法保

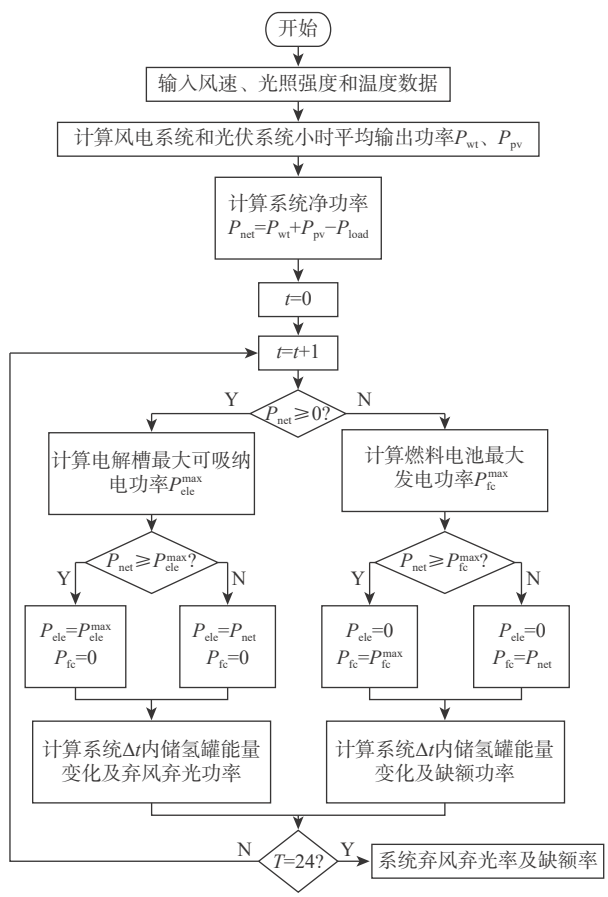


图 2 能量调度流程

Fig. 2 Energy scheduling process

障负荷供电需求时, 储氢量满足燃料电池放电需求, 燃料电池可放电以维持供电平衡, 且燃料电池允许输出功率满足缺额功率, 无须蓄电池参与放电。

4) 风光-燃料电池-蓄电池模式, 即当风光发电无法满足负荷需求时, 燃料电池可放电但最大放电功率不能满足缺额功率时, 调度蓄电池短时放电。

5) 风光-蓄电池模式, 当电解槽和燃料电池因为储氢罐容量限制无法参与维持供电平衡, 由蓄电池进行充放电保证供电可靠性。

2　优化配置模型及求解算法

2.1　优化目标

本文所搭建风光耦合制氢系统要先满足负荷需求, 并通过配置优化储能单元容量提高系统可再生能源利用率和供电可靠性, 在此基础上, 减少系统建设、运行维护等成本。在系统的建设规划过程中, 配置的制氢、储氢、燃料电池和蓄电池的容量越大, 风能和太阳能的利用率越高, 同时系统总成本也会增加。因此, 本文考虑基于 3 kW 负荷进行优化配置, 通过最小化年值成本来表征系统的经济性, 最小化系统弃风弃光率以提高可再生能源利用率, 最大化氢气产量以提高氢能产出, 并通过对系统缺额功率的约束保证系统供电稳定性, 利用优化算法求解实现系统多个指标综合最优。优化目标模型如下:

1) 最小化系统成本。本文中, 风光耦合制氢系统的经济性指标通过系统的总成本来评估, 主要考虑系统投资建设成本、运行维护成本和置换成本, 即

$$\begin{cases}\min F_1=\min C_{\text{total}}\\ C_{\text{total}}=C_{\text{in}}+C_{\text{o\&m}}+C_{\text{re}}\end{cases}\tag{13}$$

式中: C_{total} 为系统总年均成本; C_{in} 为年值投资建设成本, 表征系统初始建设、采购、安装及调试所需的一次性花费; $C_{\text{o\&m}}$ 为运行维护成本, 表征系统在该配置下每年运行维护、故障修理所需花费; C_{re} 为年值置换费用, 表征该配置下各组件到达使用年限后进行置换产生的费用。

系统成本的经济性分析在现值成本的基础上结合寿命周期, 通过资本回收系数来计算等额年值成本[11,16], 系统总年值投资成本为

$$C_{\text{in}}=C_{\text{in},i}P_i^{\text{N}}\frac{i(1+i)^N}{(1+i)^N-1}\tag{14}$$

式中: $C_{\text{in},i}$ 为各组件单位容量投资成本; P_i^{N} 为各组件配置容量; i 为年利率, 可由名义利率和年度通货膨胀率计算得到; N 为系统运行年限。

系统总年均运行维护成本为

$$C_{\text{o\&m}}=\sum C_{\text{o\&m},i}\tag{15}$$

式中: $C_{\text{o\&m},i}$ 为各组件每年的运行维护成本。

系统总年值置换成本为

$$C_{\text{re}}=\sum\left(C_{\text{re},i}P_i^{\text{N}}\sum_{n=1}^{N/y_i}\frac{1}{(1+i)^{ny_i}}\right)\tag{16}$$

式中: $C_{\text{re},i}$ 为各组件单位容量置换成本; y_i 为各组件使用年限。

2) 最小化弃风弃光率。由于风光能源具有随机性和间歇性, 输出功率波动较大, 导致很多情况下被迫产生大量弃风弃光电量, 为提高系统可再生能源利用率, 减少弃风弃光现象, 将系统弃风弃光率作为优化目标之一, 具体表示为

$$\begin{aligned}&\min F_2=\min\eta_{\text{loss}}\\ &\eta_{\text{loss}}=\sum_{t=1}^{T}P_{\text{loss}}(t)/\left(P_{\text{wt}}(t)+P_{\text{pv}}(t)\right)\end{aligned}\tag{17}$$

式中: η_{loss} 为弃风弃光率; $P_{\text{loss}}(t)$ 为 Δt 时间内系统的弃风弃光功率; $P_{\text{wt}}(t)$ 为 Δt 时间内系统风力发电产生功率; $P_{\text{pv}}(t)$ 为 Δt 时间内系统光伏发电功率; T 为系统总运行时间。

3) 最大化系统氢能产量。由于选择制氢系统作为主要的储能单元, 在系统生产周期内, 风光能源充足时制氢量越大, 风光发电失效时越能通过氢燃料电池更有效地保障供电需求。因此, 选择系统制氢量作为优化目标, 即

$$\begin{cases}\max F_3=\max E_{\text{H}_2}\\ E_{\text{H}_2}=E_{\text{ele,H}_2}-E_{\text{fc,H}_2}\end{cases}\tag{18}$$

式中: E_{H_2} 为系统氢能产量; $E_{\text{ele,H}_2}$ 和 $E_{\text{fc,H}_2}$ 分别为电解槽制氢量和燃料电池耗氢量。

2.2　约束条件

根据目标函数, 选择风力发电、光伏发电、电解槽和燃料电池的额定功率, 储氢罐和蓄电池的额定容量作为优化变量, 即 $\boldsymbol{X}=[P_{\text{wt}}^{\text{N}}, P_{\text{pv}}^{\text{N}}, P_{\text{ele}}^{\text{N}}, P_{\text{fc}}^{\text{N}}, Q_{\text{tank}}, E_{\text{b}}]$, 选取约束如下:

1) 系统负荷缺额率。在风光耦合制氢系统中, 当风力发电和光伏发电输出功率不足以满足用电需求, 且燃料电池和蓄电池均不能放电时, 负荷所需电能无法得到满足, 即产生系统缺额功率。在给定时间内, 系统产生缺额功率的概率为缺额率, 是衡量系统可靠性的重要指标, 具体表示为

$$\eta_{\text{lack}}=\sum_{t=1}^{T}P_{\text{lack}}(t)/P_{\text{load}}(t)\tag{19}$$

$$\eta_{\text{lack}}\leqslant 0.5\%\tag{20}$$

式中：$P_{\text{lack}}(t)$为Δt时间内系统的缺额功率；$P_{\text{load}}(t)$为Δt时间内系统需要提供的负荷功率。

2）蓄电池充放电约束。为提高系统可靠性，使用蓄电池作为燃料电池开启时段和紧急情况下的短时供电电源，考虑到蓄电池的寿命因素，对其充放电进行约束，蓄电池的荷电状态SOC也需满足相应约束：

$$C_{\text{SOC}}^{\min} \leqslant C_{\text{SOC}}(t) \leqslant C_{\text{SOC}}^{\max} \tag{21}$$

式中：$C_{\text{SOC}}^{\max}$、$C_{\text{SOC}}^{\min}$分别为蓄电池荷电状态的上、下限制。

3）电解槽和燃料电池功率约束。为保证系统的安全性，以及提高电解槽和燃料电池的使用寿命，对t时刻的输入输出功率进行约束。

$$\begin{cases} P_{\text{ele}}^{\min} \leqslant P_{\text{ele}}(t) \leqslant P_{\text{ele}}^{\max} \\ P_{\text{fc}}^{\min} \leqslant P_{\text{fc}}(t) \leqslant P_{\text{fc}}^{\max} \end{cases} \tag{22}$$

式中：$P_{\text{ele}}^{\max}$、$P_{\text{ele}}^{\min}$分别为电解槽的功率上、下限值；$P_{\text{fc}}^{\max}$、$P_{\text{fc}}^{\min}$分别为燃料电池的功率上、下限值。

4）储氢罐容量约束。在系统运行过程中，电解槽制氢和燃电池耗氢使得储氢量动态变化，储氢罐的容量约束如下：

$$0 \leqslant Q_{\text{H}_2}(t) \leqslant Q_{\text{tank}} \tag{23}$$

式中：$Q_{\text{H}_2}(t)$为t时刻储氢罐中储氢体积；Q_{tank}为储氢罐额定容量。

2.3 优化算法

上述系统配置优化问题是一个多目标非线性带约束优化问题，采用NSGA-Ⅲ遗传算法对问题进行求解。NSGA-Ⅲ算法是Deb和Jain[17]在遗传算法的基础上，针对高维多目标问题提出的第三代改进算法，该算法在框架上与NSGA-Ⅱ十分相似，同样利用快速非支配排序将种群个体分类进入不同的非支配前沿。不同的是，对于在临界层中的环境选择，NSGA-Ⅱ利用拥挤比较操作来选择从而保持多样性，NSGA-Ⅲ最大的变化就是利用良好分布的参考点来保持种群的多样性[18]。算法的流程如下：

步骤1 读入气象数据，初始化产生一个规模为N的种群$\boldsymbol{P}_0$，迭代次数gen = 1。

步骤2 根据控制策略，计算当前种群$\boldsymbol{P}_t$的目标函数值。

步骤3 对当前种群进行选择、重组、变异操作得到子代种群$\boldsymbol{Q}_t$。

步骤4 合并父代种群$\boldsymbol{P}_t$和子代种群$\boldsymbol{Q}_t$，得到规模为$2N$的种群$\boldsymbol{R}_t$。

步骤5 对$\boldsymbol{R}_t$进行非支配排序，将非支配层级为1、2、3等的个体依次加入种群$\boldsymbol{P}_{t+1}$，直至$\boldsymbol{P}_{t+1}$的规模大于N。

步骤6 标准化目标空间，根据参考点Z计算参考线，并将临界层中个体与参考线进行关联操作，统计每条参考线关联个体。

步骤7 根据参考线关联个体数从临界支配层中选择$k(k = N - |\boldsymbol{P}_{t+1}|)$个种群个体进入新一代父代$\boldsymbol{P}_{t+1}$。迭代次数gen = gen + 1。

步骤8 重复操作步骤2～步骤7直至gen > $\text{gen}_{\max}$。其中，$\text{gen}_{\max}$为最大迭代次数。

3 算例结果分析

选取某边防地区的气象数据作为配置优化输入，该地区某年的逐月气象变化趋势如图3所示。在配置优化计算中，采样该地区一年的风、光和温度数据，采样时间间隔为1 h，数据来源于ERA5 hourly data on single levels from 1979 to present数据集。风速和光照强度是衡量风能和太阳能的2个重要指标，可以看出，2个指标在图3中呈现一定的季节互补性。

表1给出成本函数计算所使用数据。

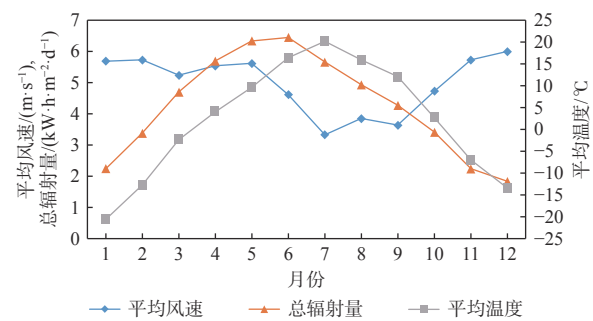


图3 某边防地区逐月气象变化趋势

Fig. 3 Monthly meteorological trend of border area

3.1 优化算法求解结果分析

针对以系统成本、弃风弃光率和系统氢能产量为优化目标的三维多目标优化问题，选择NSGA-Ⅲ算法对配置问题进行求解，综合考虑优化解集的多样性和优化运算量，设置种群规模为500，迭代次数为1 000次，得到优化解集如图4所示。

整个解集在目标空间中分布的广泛程度反映了解集的广泛性，是衡量解集多样性的重要指标，希望所得解集在帕累托前沿上分布尽可能广。图4中，红色点集是三维的优化解集曲面，每个点代表一个优化配置方案，另外3种点集则为解集在3个平面上的投影，可以得到如表2所示的优化配置方案示例。

从图4和表2中可看出，年度氢能产量与系统成本、年度氢能产量与弃风弃光率之间优化趋势相同，即系统成本越高，得到年度氢能产量越大，弃风弃光率越大，年度氢能产量也会越大；系统成本与

表 1　不同组件的单位容量成本数据

Table 1　Cost per unit capacity for different components

风力发电机		光伏阵列		电解槽		
投资建设成本/(元·kW^{-1})	运行维护成本/(元·(kW·年)$^{-1}$)	投资建设成本/(元·kW^{-1})	运行维护成本/(元·(kW·年)$^{-1}$)	投资建设成本/(元·kW^{-1})	运行维护成本/(元·(kW·年)$^{-1}$)	置换成本/(元·kW^{-1})
7 250	145	9 200	184	4 465.5	128	4 654
燃料电池			**储氢罐**		**蓄电池**	
投资建设成本/(元·kW^{-1})	运行维护成本/(元·(kW·年)$^{-1}$)	置换成本/(元·kW^{-1})	投资建设成本/(元·kg^{-1})	运行维护成本/(元·(kW·h·年)$^{-1}$)	投资建设成本/(元·(kW·h)$^{-1}$)	运行维护成本/(元·(kW·h·年)$^{-1}$)	置换成本/(元·(kW·h)$^{-1}$)
3 000	60	2 570	42 000	170	1 190	24	992

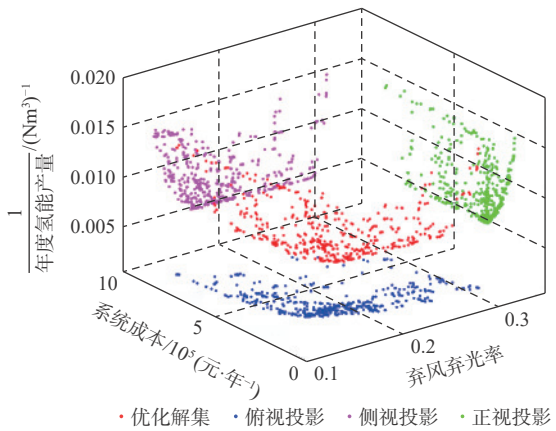


图 4　NSGA-Ⅲ优化结果

Fig. 4　NSGA-Ⅲ optimization results

表 2　优化配置方案示例

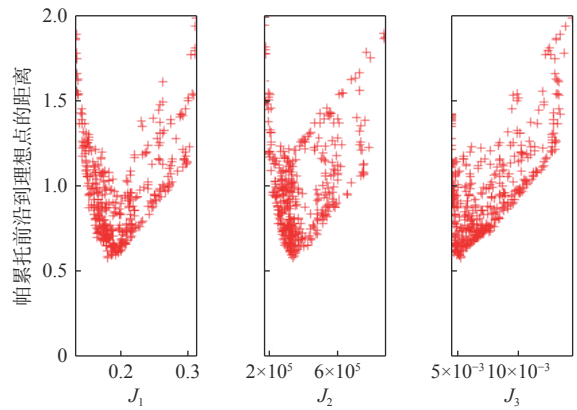
Table 2　Example of optimal configuration scheme

编号	弃风弃光率	系统成本/元	年度氢能产量/Nm^3
1	0.134	795 453	76.923
2	0.181	346 965	200
3	0.277	238 842	142.857
4	0.258	339 334	250
⋮			

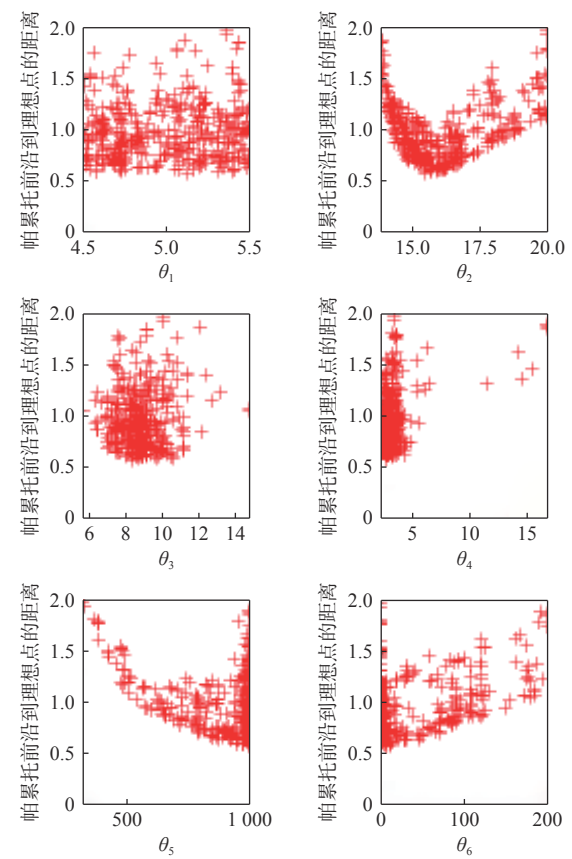
弃风弃光率之间则呈现相悖的优化趋势, 即系统成本越高, 弃风弃光率越小; 且系统成本、弃风弃光率和年度氢能产量 3 个优化目标两两之间呈现非线性趋势, 可通过多目标优化问题的求解和决策得到不同需求下的配置方案。

利用层次图对帕累托解集进行可视化解集决策, 层次图通过计算近似帕累托前沿到理想点的距离来对解集进行分类, 将优化目标和设计变量的每个维度转化为二维图中 x 轴的坐标, y 轴则为通过范数计算的到理想点的距离, 且: ①位于较低层次的点对应于靠近理想点的帕累托前沿区域; ②由于目标函数和设计变量的值对应到 x 轴, 图中空白地带即为目标函数和设计变量不可取值范围[19]。

绘制层次图如图 5 所示。图中, J_1、J_2 和 J_3 分别为弃风弃光率、系统成本和年度氢能产量的倒数; θ_1、θ_2、θ_3、θ_4、θ_5、θ_6 为配置优化问题的优化变量。较低层次的帕累托前沿相对最接近理想点, 通



(a) 帕累托前沿层次图



(b) 帕累托解集层次图

图 5　帕累托前沿及解集层次图

Fig. 5　Level diagrams for Pareto front and Pareto solution set

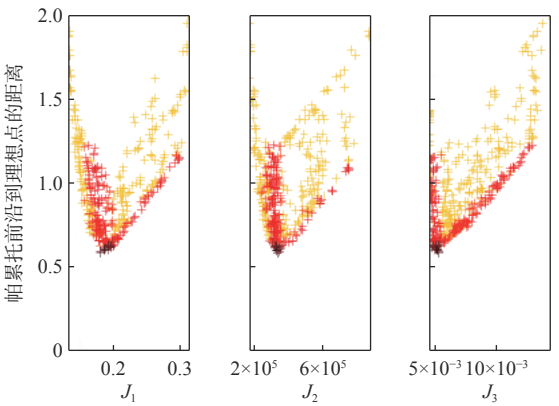
过设定如表 3 所示的理想目标函数取值范围对帕累托解集进行评分, 其中, 颜色越深, 评分越低, 理

想程度越高。$J_i \leqslant J_i^{\mathrm{HD}}$，是相当理想的目标值；$J_i^{\mathrm{HD}} < J_i \leqslant J_i^{\mathrm{D}}$，是相对可取的目标值；$J_i^{\mathrm{D}} < J_i \leqslant J_i^{\mathrm{T}}$，是可以接受的目标值；$J_i^{\mathrm{T}} < J_i \leqslant J_i^{\mathrm{U}}$，是不受欢迎的目标值；$J_i^{\mathrm{U}} < J_i \leqslant J_i^{\mathrm{HU}}$，是较为不理想的目标值；$J_i > J_i^{\mathrm{HU}}$，是不可取的目标值。基于此，可得到如图 6 所示的层次图。最终得到最优配置方案，如表 4 所示。

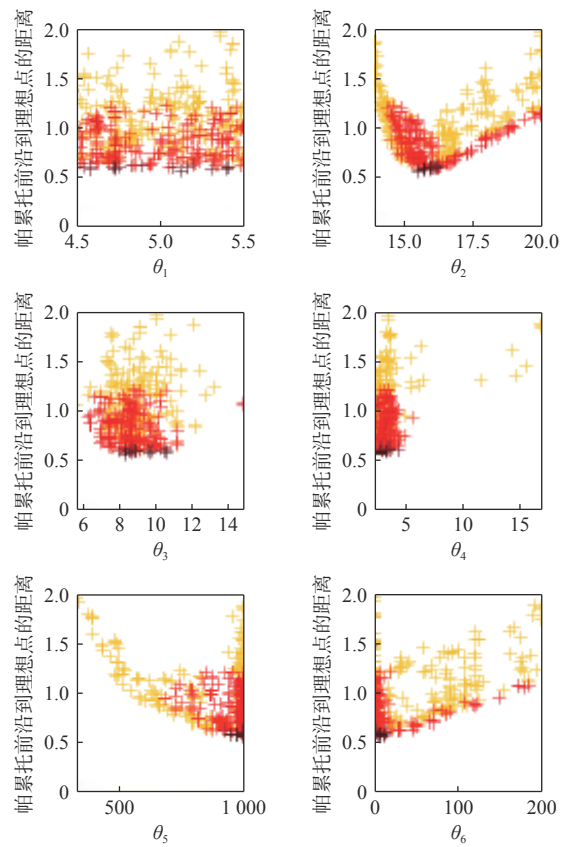
表 3　理想目标函数取值范围

Table 3　Ideal value range of objective function

优化目标	J_i^{HD}	J_i^{D}	J_i^{T}	J_i^{U}	J_i^{HD}	J_i^{HU}
J_1	0.15	0.16	0.17	0.18	0.19	0.2
J_2	220 000	260 000	300 000	320 000	340 000	360 000
J_3	0.005 0	0.005 1	0.005 2	0.005 3	0.005 4	0.005 5



(a) 评分后帕累托前沿层次图



(b) 评分后帕累托解集层次图

图 6　评分后帕累托前沿及解集层次图

Fig. 6　Level diagrams for Pareto front and Pareto solution set after scoring

表 4　最优配置方案

Table 4　Optimal configuration scheme

风力发电机额定功率/kW	光伏阵列额定功率/kW	电解槽额定功率/kW	燃料电池额定功率/kW	储氢罐额定容量/$\mathrm{Nm^3}$	蓄电池额定容量/(kW·h)
10	15.52	8.38	2.86	223	8.38

3.2　优化配置方案分析

传统配置计算方法通常基于地区的统计日平均峰值日照时长和年均达到有效风速时长来计算，为保证系统在风、光模块均不能正常工作的天气下也能运行供电，提前设计相应的储能备份。将优化配置方案与未配置储能单元方案及未优化配置计算方案进行对比，得到如表 5 所示结果。其中，风光互补特性可计算为风力与光伏发电功率之和相对于负荷功率的波动率，即[20]

$$D_{\mathrm{L}} = \frac{1}{\overline{P}_{\mathrm{load}}}\sqrt{\frac{1}{T}\sum_{t=1}^{T}\left(P_{\mathrm{wt}}(t)+P_{\mathrm{pv}}(t)-P_{\mathrm{load}}(t)\right)^2} \tag{24}$$

式中：$\overline{P}_{\mathrm{load}}$为平均负荷功率。波动率 D_{L} 越小，风光输出功率与所需负荷功率越接近，认为所需储能单元容量越小，即风光互补特性越好[21]。

表 5　优化配置方案对比

Table 5　Comparison of optimal configuration schemes

方案	负荷缺额率/%	弃风弃光率	系统成本/(元·年$^{-1}$)	年度氢能产量/$\mathrm{Nm^3}$	风光互补特性
仅风光发电单元	42.9	0.170			
耦合氢能下未优化配置方案	1.1	0.737	439 674	167.4	5.113
优化配置方案	4.3	0.184	338 140	167.6	1.486

如表 5 所示，未配置储能单元下的风光耦合制氢系统，由于可再生能源的波动性和随机性，使得系统难以保障边防哨所的用电需求，负荷缺额率达 42.9%。耦合氢能后，未优化方案及优化配置方案的负荷缺额率大幅度下降，边防哨所供电可靠性得到进一步保障。

优化后的配置方案与优化前相比负荷缺额率达到了指标要求，弃风弃光率降低了约 75%，系统成本降低了约 23%，年度氢能产量相差不多，即优化配置方案在满足负荷需求的同时减少了可再生能源的浪费，在较低的投资成本下获得了氢能储备保障，进一步提高了系统的经济性。图 7 为春、夏、秋、冬 4 个季节典型周风光出力及电解槽和燃料电池工作情况堆叠柱状图。

如图 7 所示，风电和光伏的发电功率峰值呈现一定的互补性，风电在春、冬季节典型周发电量分

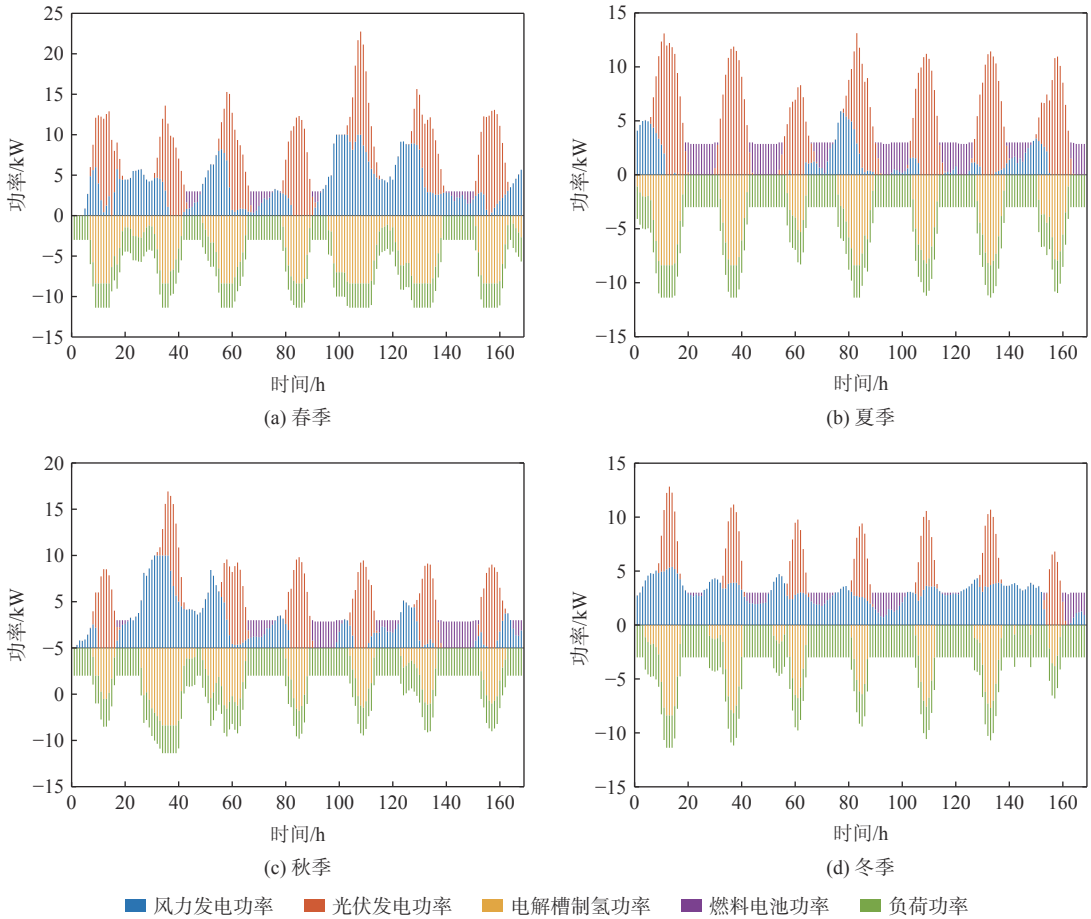


图 7　四季典型周系统各单元出力情况

Fig. 7　Power output of each unit of typical weekly system in four seasons

别约为 615.19 kW·h 和 487.04 kW·h, 较夏、秋季节 156.03 kW·h 和 384.76 kW·h 大, 而光伏则是在夏季发电量最大达 625.22 kW·h; 另外, 电解槽在风光发电量充足时启动, 制取的氢气支撑燃料电池在风光发电量不足时启动, 其中, 夏季典型周发电量最大达 166.73 kW·h, 有效维持了系统功率平衡和满足负荷需求, 可在边防哨所独立运行。

4　结　论

1) 在针对于边防哨所供电和能源需求问题搭建的风光耦合制氢系统基础上, 设计了以氢能为主要储能载体的能量调度策略, 搭建了以系统成本、可再生能源利用率和氢能产量为优化目标的优化配置问题模型, 并利用了 NSGA-Ⅲ算法求解了该多目标优化问题。

2) 对 NSGA-Ⅲ算法得到的优化结果进行了分析和决策, 并对优化配置方案进行了对比分析, 验证了优化配置方法可以有效提高风光耦合制氢系统建设过程中的经济性, 保证供电可靠性及提高可再生能源利用率。

3) 通过春、夏、秋、冬 4 季典型周验证了氢能单元作为风光耦合制氢系统储能单元可平抑风光出力波动, 满足边防哨所的负荷需求及能源多样性需求, 验证了优化配置方案的有效性, 由于风光能源的季节性, 燃料电池发电量在夏季达到最大。

配置优化和能量管理是混合能源系统的两大主要研究方向, 本文针对风光耦合制氢系统的配置优化进行了分析, 后续将进一步对该配置下的能量优化管理方法进行研究, 以进一步实现提高可再生能源利用率和减小输出功率波动等目标。

参考文献（References）

[1] 孔令国. 风光氢综合能源系统优化配置与协调控制策略研究[D]. 北京: 华北电力大学, 2017: 1-12.
KONG L G. Research on optimal sizing and coordinated control strategy of integrated energy system of wind photovoltaic and hydrogen[D]. Beijing: North China Electric Power University, 2017: 1-12(in Chinese).

[2] 曹蕾, 郭婷婷, 殷爱鸣, 等. 风光氢混合发电系统设计与能量管理策略研究进展[J]. 分布式能源, 2021, 6(4): 1-14.
CAO F, GUO T T, YIN A M, et al. Research progress on optimal sizing and energy management strategy of wind-solar-hydrogen hybrid energy systems[J]. Distributed Energy, 2021, 6(4): 1-14(in Chinese).

[3] AL BUSAIDI A S, KAZEM H A, AL-BADI A H, et al. A review of optimum sizing of hybrid PV-wind renewable energy systems in Oman[J]. Renewable and Sustainable Energy Reviews, 2016, 53: 185-193.

[4] KHALILNEJAD A, RIAHY G H. A hybrid wind-PV system performance investigation for the purpose of maximum hydrogen production and storage using advanced alkaline electrolyzer[J]. Energy Conversion and Management, 2014, 80: 398-406.

[5] BAKHTIARI H, ALI NAGHIZADEH R. Multi-criteria optimal sizing of hybrid renewable energy systems including wind, photovoltaic, battery, and hydrogen storage with ε-constraint method[J]. IET Renewable Power Generation, 2018, 12(8): 883-892.

[6] GONZÁLEZ A, RIBA J R, RIUS A, et al. Optimal sizing of a hybrid grid-connected photovoltaic and wind power system[J]. Applied Energy, 2015, 154: 752-762.

[7] 马榕谷, 陈洁, 赵军超, 等. 非并网风氢互补系统的容量多目标优化[J]. 太阳能学报, 2019, 40(2): 422-429.
MA R G, CHEN J, ZHAO J C, et al. Multi-objective optimization for capacity of non-grid-connected wind/hydrogen hybrid power system[J]. Acta Energiae Solaris Sinica, 2019, 40(2): 422-429(in Chinese).

[8] 王侃宏, 赵政通, 罗景辉. 风光氢储系统的两阶优化匹配分析[J]. 科学技术与工程, 2020, 20(26): 10790-10794.
WANG K H, ZHAO Z T, LUO J H. Two-order optimization matching analysis of scenery hydrogen storage system[J]. Science Technology and Engineering, 2020, 20(26): 10790-10794(in Chinese).

[9] KHIAREDDINE A, BEN SALAH C, REKIOUA D, et al. Sizing methodology for hybrid photovoltaic/wind/hydrogen/battery integrated to energy management strategy for pumping system[J]. Energy, 2018, 153: 743-762.

[10] 董伟强. 风光氢蓄混合发电系统的配置及其电池管理研究[D]. 杭州: 浙江大学, 2017: 10-19.
DONG W Q. Optimal sizing and battery management of a stand-alone hybrid power system based on battery/hydrogen[D]. Hangzhou: Zhejiang University, 2017: 10-19(in Chinese).

[11] HADIDIAN MOGHADDAM M J, KALAM A, NOWDEH S A, et al. Optimal sizing and energy management of stand-alone hybrid photovoltaic/wind system based on hydrogen storage considering LOEE and LOLE reliability indices using flower pollination algorithm[J]. Renewable Energy, 2019, 135: 1412-1434.

[12] XU L, RUAN X B, MAO C X, et al. An improved optimal sizing method for wind-solar-battery hybrid power system[J]. IEEE Transactions on Sustainable Energy, 2013, 4(3): 774-785.

[13] SKOPLAKI E, PALYVOS J A. Operating temperature of photovoltaic modules: A survey of pertinent correlations[J]. Renewable Energy, 2009, 34(1): 23-29.

[14] MORIN D, STEVENIN Y, GROLLEAU C, et al. Evaluation of performance improvement by model predictive control in a renewable energy system with hydrogen storage[J]. International Journal of Hydrogen Energy, 2018, 43(45): 21017-21029.

[15] GÖTZ M, LEFEBVRE J, MÖRS F, et al. Renewable power-to-gas: A technological and economic review[J]. Renewable Energy, 2016, 85: 1371-1390.

[16] SHAW S, PETEVES E. Exploiting synergies in European wind and hydrogen sectors: A cost-benefit assessment[J]. International Journal of Hydrogen Energy, 2008, 33(13): 3249-3263.

[17] DEB K, JAIN H. An evolutionary many-objective optimization algorithm using reference-point-based nondominated sorting approach, Part I: Solving problems with box constraints[J]. IEEE Transactions on Evolutionary Computation, 2014, 18(4): 577-601.

[18] 耿焕同, 戴中斌, 王天雷, 等. 基于参考点选择策略的改进型NSGA-Ⅲ算法[J]. 模式识别与人工智能, 2020, 33(3): 191-201.
GENG H T, DAI Z B, WANG T L, et al. Improved NSGA-Ⅲ algorithm based on reference point selection strategy[J]. Pattern Recognition and Artificial Intelligence, 2020, 33(3): 191-201(in Chinese).

[19] BLASCO X, HERRERO J M, SANCHIS J, et al. A new graphical visualization of n-dimensional Pareto front for decision-making in multiobjective optimization[J]. Information Sciences, 2008, 178(20): 3908-3924.

[20] 徐林, 阮新波, 张步涵, 等. 风光蓄互补发电系统容量的改进优化配置方法[J]. 中国电机工程学报, 2012, 32(25): 88-98.
XU L, RUAN X B, ZHANG B H, et al. An improved optimal sizing method for wind-solar-battery hybrid power system[J]. Proceedings of the CSEE, 2012, 32(25): 88-98(in Chinese).

[21] 万家豪, 苏浩, 冯冬涵, 等. 计及源荷匹配的风光互补特性分析与评价[J]. 电网技术, 2020, 44(9): 3219-3226.
WAN J H, SU H, FENG D H, et al. Analysis and evaluation of the complementarity characteristics of wind and photovoltaic considering source-load matching[J]. Power System Technology, 2020, 44(9): 3219-3226(in Chinese).

Configuration optimization of wind/photovoltaic hydrogen production system at frontier sentries

QI Haitao[1, 2, *], LIU Duo[1], ZHAO Dong'ao[1], LIU Xu[1]

(1. School of Mechanical Engineering & Automation, Beihang University, Beijing 100191, China;
2. Engineering Training Center, Beihang University, Beijing 100191, China)

Abstract: In order to solve the problems of the remote location of frontier sentries, complex environment, large energy demand, and high cost of access to the electricity network, the complementary nature of wind/photovoltaic resources, as well as the high energy density and green and clean characteristics of hydrogen energy were used to establish a wind/photovoltaic hydrogen production system with wind and solar energy as the main power generation energy. In addition, hydrogen energy was selected as the energy storage unit of the system to meet the demand for energy diversity of frontier sentries and improve the sentries eability to obtain energy from the surrounding environment. Firstly, the mathematical model of each subsystem of the wind/photovoltaic hydrogen production system was established, and the system construction configuration was optimized to reduce the system cost, improve the utilization rate of renewable energy, and increase hydrogen energy production. The rated power of each component of the system was chosen as the design variable, and the meteorological data of specific regions was used as input. The non-dominated sorting genetic algorithm-Ⅲ (NSGA-Ⅲ) was applied to solve the system, and decision-making was performed on the Pareto solution set obtained by optimization.The visualization of the multi-dimensional Pareto front was realized through the drawing of level diagrams, and the optimal configuration scheme was compared with the unoptimized configuration scheme, the energy utilization rate and configuration cost can be greatly improved while ensuring the reliability of power supply.

Keywords: wind power generation; photovoltaic power generation; hydrogen production by electrolysis; hybrid energy system; configuration optimization

Received: 2022-09-14; **Accepted**: 2023-02-17; **Published Online**: 2023-03-24 16:59
URL: link.cnki.net/urlid/11.2625.V.20230323.1636.009

* **Corresponding author**. E-mail: qihaitao@buaa.edu.cn

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2022.0815

基于IFC标准的机场道面结构信息模型存储与传递方法

蔡靖[1]，王瀚雪[1]，戴轩[1,*]，李岳[1]，马磊[2]

(1. 中国民航大学 交通科学与工程学院，天津 300300；2. 广东省机场管理集团有限公司，广州 510406)

摘 要：为解决机场跑道系统在全生命周期建筑信息模型(BIM)应用中的信息传递和数据交互问题，将工业基础类(IFC)标准在机场跑道系统进行扩展，建立机场道面结构的IFC实体架构，在此基础上创建面向运维应用的刚性道面板实体，并进行几何信息与属性信息的扩展；提出基于IFC的属性信息添加与传递方法，对信息传递过程中的属性信息丢失率进行分析。研究表明：通过建立IFC标准模型，可以实现机场跑道系统的信息传递和数据交互；采用基础模型族扩展法进行机场道面模型传递的属性信息丢失率高达84.21%；采用IFC解析扩展法首次传递可保证信息不丢失，但二次传递信息存在丢失问题；族模型IFC解析扩展法可实现信息首次传递不丢失，二次传递属性信息丢失率低至16.7%，并且采用不同IFC标准版本对模型传递过程中的属性信息丢失率产生较大影响。应用IFC标准进行机场领域扩展具有较高的可行性，所提的信息传递方法可实现机场道面全生命周期信息传递。研究成果可为机场飞行区新IFC实体的扩展提供支持，为机场工程数字化提供技术参考。

关 键 词：机场道面结构；工业基础类；通用数据环境；信息模型；信息传递

中图分类号：V351.11

文献标志码：A **文章编号**：1001-5965(2024)10-3042-12

随着建筑工业化、数字化、智能化水平的不断提升，建筑信息模型(building information modeling, BIM)技术在国内外工程领域已广泛应用。由于BIM相关软件种类繁多，文件格式多样，工程应用过程中的数据信息无损传递与共享是BIM技术成功应用的关键。buildingSMART为此提出了一种中立、开放的数据交换标准——工业基础类(industry foundation classes, IFC)[1]，IFC标准可应用于建设工程全生命周期的各个阶段，提高了信息传递的质量和效率[2]。

在IFC标准整体构架方面，赖华辉等[3-4]分析了其发展特点，并且提出了数据交换的新途径；针对信息传递过程中的映射问题，郭红领等[5]提出了一种自动映射的方法。此外，在不同领域中，均针对IFC标准的信息传递与实体标准开展了研究。Yabuki和Li[6]对桥梁建设中的实体进行了扩展，并建立了基于IFC的数字化模型；Lee和Kim[7]对道路结构进行了IFC描述；Borrmann[8]、钟宇[9]等对隧道工程的空间和物理结构从不同方面进行了IFC扩展，相关研究成果已在IFC4.3标准[1]中有所体现；王松怀[10]、冯山群[11]等提出了铁路工程IFC数据传递方法与流程，并解决了信息传递过程中的缺失问题；胡振中等[12]提出了基于IFC的传感器信息描述和关联机制与信息传递方法。

机场工程涉及专业多、工程体量大、信息交互广、智慧运行对信息传递要求高，基于BIM技术可

收稿日期：2022-09-29；**录用日期**：2023-04-14；**网络出版时间**：2023-05-05 16:36
网络出版地址：link.cnki.net/urlid/11.2625.V.20230505.1432.007
基金项目：天津市自然科学基金(21JCQNJC00850)；中央高校基本科研业务费专项资金(3122022043)
*通信作者. E-mail：tianjindaixuan@126.com

引用格式：蔡靖，王瀚雪，戴轩，等. 基于IFC标准的机场道面结构信息模型存储与传递方法[J]. 北京航空航天大学学报，2024，50(10)：3042-3053. CAI J，WANG H X，DAI X，et al. Storage and transfer method of airport pavement structure information model based on IFC standard[J]. Journal of Beijing University of Aeronautics and Astronautics，2024，50(10)：3042-3053 (in Chinese).

实现机场工程的全生命周期应用[13-14], 中国北京大兴国际机场、湖北鄂州花湖国际机场[15]、南京禄口国际机场[16]、上海浦东国际机场[17]、广州白云国际机场[18] 等在工程建设中均应用了 BIM 技术, 然而其应用主要局限于建设阶段, 针对机场专有基础设施的全生命周期过程缺乏信息传递标准与方法。

机场工程 BIM 信息在传递、共享中的丢失问题是制约其发展的瓶颈, 本质在于传递方法与标准的缺失。在机场领域, 特有 IFC 实体的扩展研究也少见报道。为此, 本文首先基于 IFC 标准对机场跑道系统进行实体构架梳理, 然后, 在此基础上建立道面结构新实体, 对其相关属性进行扩展, 形成机场跑道系统数据模型, 并实现数据信息高效传递, 最后, 提出一种降低属性信息丢失率的机场道面结构信息传递方法。

1　基于 IFC 的机场跑道系统扩展

机场跑道系统的 IFC 实体扩展在国际标准 ISO 16739-1-2024[1] 基础上进行, 所扩展的机场跑道系统实体结构逻辑层次利用 EXPRSS-G 图表示, 如图 1 所示, 其中, 新实体用灰色图框标出。图中, 所有实体类型均由核心层实体 IfcRoot 继承而来; 同时, 根类 IfcRoot 还派生出属性定义实体 IfcPropertyDefinition、关系实体 IfcRelationship、对象实体 IfcObjectDefinition 3 个基本抽象类, 每个基本抽象类又分别派生出相应的子类。对象实体 IfcObjectDefinition 可以数字化描述有形或抽象的实体, 其子类 IfcObject 又可包含 6 个子类, 在机场跑道系统扩展的过程中, 仅用到 IfcObject 的子类 IfcProduct, 该类是所有项目中的基本类, 可以表达空间、物理、结构等多方面的元素。结合机场的实际使用情况, 分别整理添加与机场相关的物理实体(隶属于 IfcCivilElement)和空间实体(隶属于 IfcSpatialStructuralElement)。物理实体根据功能划分为飞行区域结构实体和跑道附属设施实体; 空间实体则按照道面结构的空间组成划分, 以对机场道面结构所占用的空间进行描述。

1.1　机场道面结构物理实体

对于机场道面物理实体, 在工程类物理实体中增加抽象子类, 如图 2 所示。IfcCivilElement 下新增飞行区域结构实体(IfcAirfieldAreaStructure Element)和跑道附属设施实体(IfcRunwaySubsidiaryFacility)。IfcAirfieldAreaStructureElement 是机场道面结构的抽象类, 主要包含跑道系统 (IfcRunwaySystem)、滑行道系统 (IfcTaxiwaySystem)、净空区域 (IfcClearanceArea) 与停机坪 (IfcApron) 四大类; 跑道系统实体则由跑道结构道面(IfcRunwayStructurePavement)、道肩(IfcShoulder)、升降带(IfcRunwayStrip)、防吹坪(IfcRunwayBlastPad)、跑道端安全区(IfcRunwayEndSafetyArea)、净空道(IfcClearway)、停止道(IfcStopway)等构成, 跑道结构道面则由道面块体(IfcRunwayBlock)、传力杆 (IfcDowelBars)、拉杆 (IfcTieBars)、钢筋网 (IfcSteelNet) 共同组成。道面块体指具有相同属性且连续跑道空间内的最小实体单元, 其几何边界由运维过程中自定义所需评价道面板数量分割形成, 也可根据实际道面板接缝进行划分, 同时作为道面结构的板类型, 用于表示不同材料特征和多层形状, 并定义其性质。跑道附属设施实体由标志物(IfcMarker)、标志(IfcMarking)、标记牌(IfcSign)、助航灯光(IfcNavigationalLightingAid)等派生而成。

1.2　机场道面结构空间实体

现有 IFC 标准中的 IfcBuilding、IfcBuildingStorey、IfcSite、IfcSpace 等空间描述类无法描述机场道面结构, 因此, 在空间实体的抽象类 IfcSpatialStructuralElement 下增加抽象子类跑道模型 (IfcRunway)、跑道系统分部 (IfcRunwaySystemPart) 来描述对跑道模型中的各类空间概念实体, 如图 3 所示。跑道模型是对机场道面结构本身的特征数据描述, 包括跑道方向(IfcRunwayDirection)、等级、位置等空间信息。跑道系统分部是从地质空间中由面层 (IfcSufaceCourseSpace)、基层 (IfcBaceCourseSpace)、底基层 (IfcSubgradeSpace)、垫层 (IfcCushronSpace) 共同组成。飞行区道面结构的空间位置特征则由 PartType 属性表达, 其中, PartType 的类型为枚举类型 IfcRunwayTypeEnum, 定义如下: TYPE IfcRunwayTypeEnum = ENUMERATION OF(RUNWAY-TAXIWAY,CLEARWAY,APRON,USERDEFINED,NOTDEFINED); END_TYPE。

1.3　机场道面结构实体间关系

实体间关系包括空间实体之间及空间实体和物理实体之间 2 种。项目 (IfcProject) 下包含场地 (IfcSite), 而跑道模型 (IfcRunway) 在场地内, 跑道模型可由多个跑道系统分部 (IfcRunwaySystemPart) 聚合而成, 跑道系统分部由多结构层组成 (面层、基层、底基层、垫层等), 同时也可包含单个分层空间的元素类型。空间实体之间的关系如图 4 所示, 跑道模型与跑道系统分部之间的关系通过 IfcRelAggregates 进行描述。

在空间实体和物理实体的关系方面, 跑道系统分部与道面块体 (IfcRunwayBlock) 的关系通过 IfcRelContainedInSpatialStructure 进行表达, 一个跑

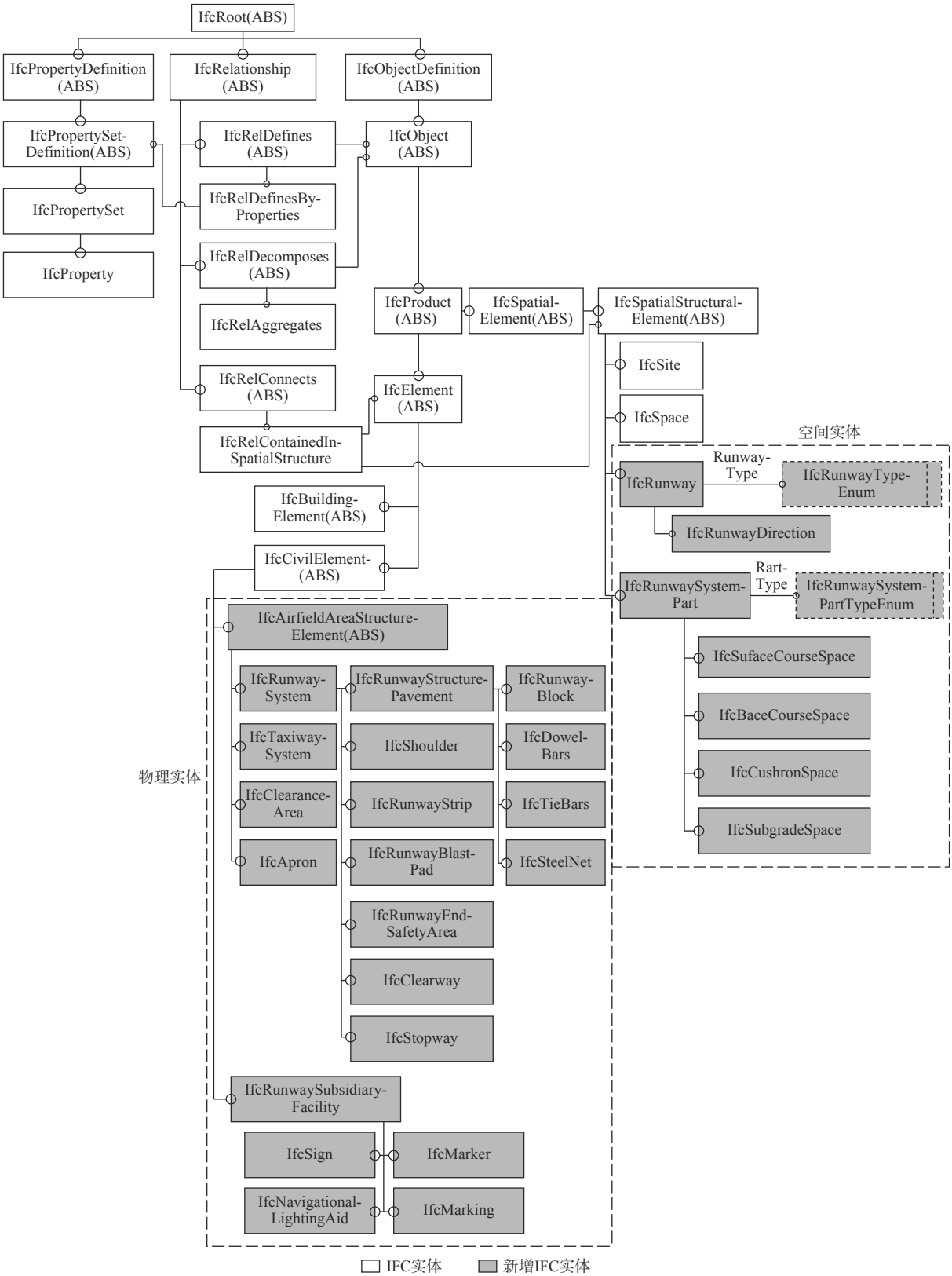


图 1 EXPRESS-G 图表示机场跑道系统模型

Fig. 1 EXPRESS-G diagram of airport runway system model

道系统分部包含一个或多个道面块体，如图 5 所示。

1.4 机场道面信息的存储和表达

IFC 数据模型存储和表达道面板的相关信息都集成于跑道系统 (IfcRunwaySystem) 中，跑道信息主要包括道面的几何信息及属性信息 2 类。道面几何信息包括各结构层的平面尺寸、厚度等，道面板的属性信息包括材料信息、运维所需的道面损坏类型、损坏类型所对应的数量、损坏密度、损坏程

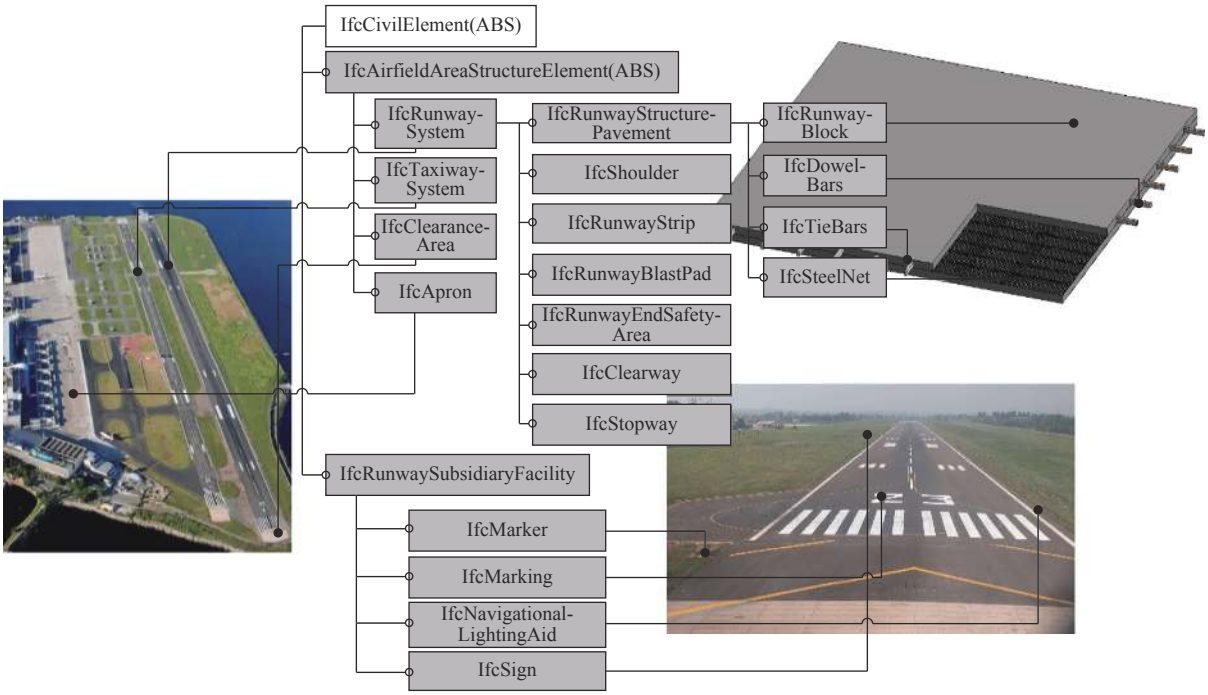


图 2　基于 IFC 标准的物理结构层次关系

Fig. 2　Hierarchical relationship of physical structure based on IFC standard

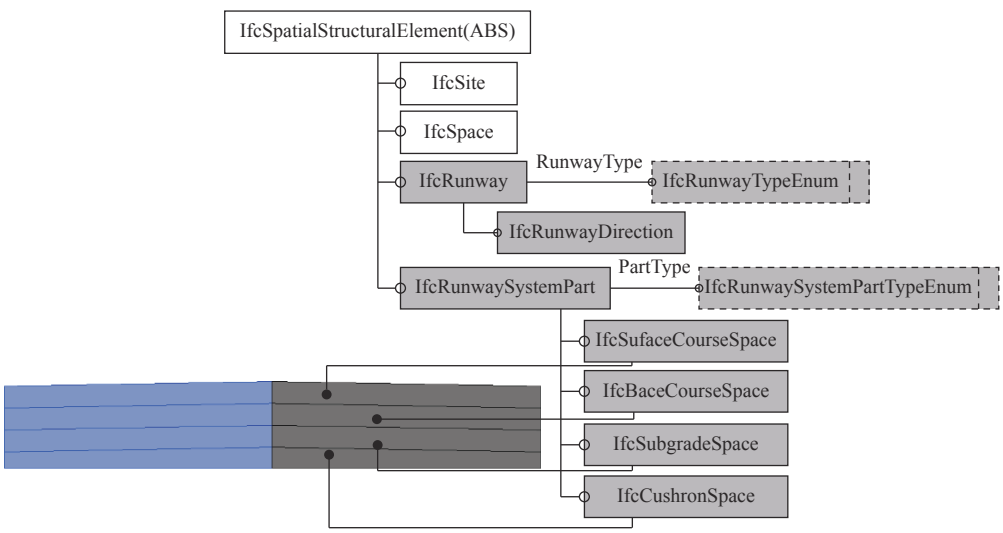


图 3　基于 IFC 标准的空间结构层次关系

Fig. 3　Hierarchical relationship of spatial structure based on IFC standard

度、损坏坐标、弯拉强度、道面状况指数 PCI、道面结构状况指数 SCI、道面等级号 PCN、平整度 IRI 等关键指标。

针对机场道面结构，需扩展的属性如表 1 所示。相关属性数据信息来源于机场全生命周期中产生的数据信息，如设计阶段产生的工程设计数据、勘察数据，施工阶段产生的工程建设数据，运维阶段产生的运行维护数据，包括道面监测、检测、感知数据等。因此，将道面块体属性信息划分为道面基础参数集、道面板病害参数集、道面板评价参数集和道面板运维数据参数集，相应定义了 Pset_RunwayBlockCommon、Pset_PavementDisease-Common、Pset_PavementElementEvaluation、Pset_PavementDate 这 4 种属性集信息，这 4 类信息也是机场道面结构运维所需要的，由于不存在道面的相关定义，其信息通过扩展属性集的方式集成至 IfcRunwayBlock 实体中。

1）道面基础参数集。适用于机场道面工程的基础参数集 (Pset_RunwayBlockCommon 属性) 如表 2 所示，各属性类型均为 IfcPropertySingleValue，该属性集主要用于表示道面的基本信息和性能参数。基础参数包括道面的面积、体积、坡度、铺设时间、结构层及各层对应的厚度和材料。

2）道面板病害参数集。道面板病害参数集

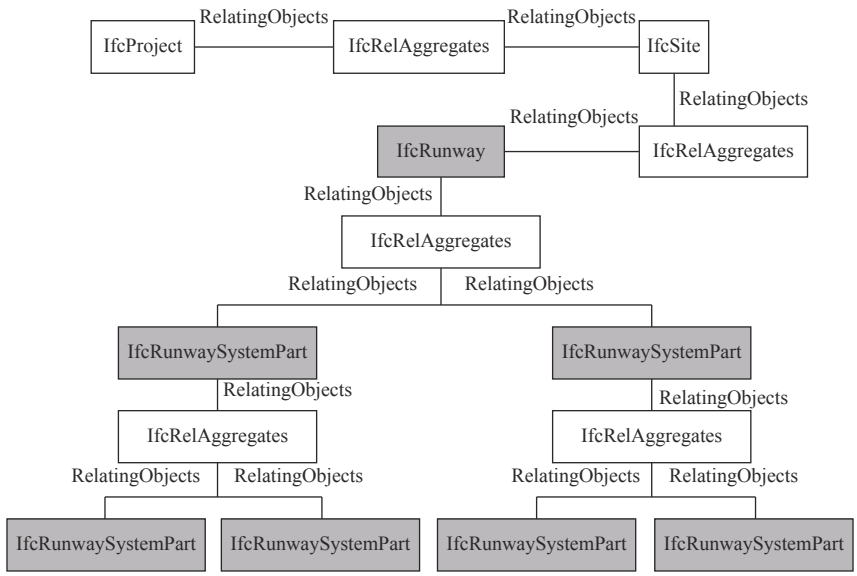


图 4 跑道模型空间结构层次关系

Fig. 4 Hierarchical relationship of spatial structure of runway model

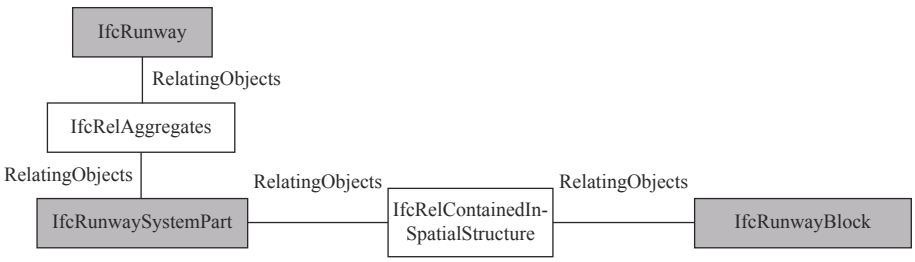


图 5 跑道模型物理实体和空间实体关系示意图

Fig. 5 Schematic diagram of relationship between physical entities and spatial entities of runway model

表 1 道面板新增属性集定义

Table 1 Added property set definition to pavement

属性集名称	使用的实体	类型值	说明
Pset_RunwayBlockCommon	IfcRunwayBlock	UserDefined	描述道面组成层特性和尺寸
Pset_PavementDiseaseCommon	IfcRunwayBlock	UserDefined	道面的常见病害类型和相关数据信息
Pset_PavementElementEvaluation	IfcRunwayBlock	UserDefined	道面评价属性(PCI、SCI、PCN等)
Pset_PavementDate	IfcRunwayBlock	UserDefined	道面板运维监测数据信息

表 2 Pset_RunwayBlockCommon 属性定义

Table 2 Pset_RunwayBlockCommon property definition

属性名称	数据类型	说明
PavementArea	IfcAreaMeasure	道面板面积
PavementVolume	IfcVolumeMeasure	道面板体积
VerticalSlope	IfcPlaneAngleMeasure	纵坡坡度
RunwayLayer	IfcCountMeasure	跑道层数
SufaceCourseSlope	IfcPlaneAngleMeasure	面层坡度
SufaceCourseThickness	IfcPositiveLengthMeasure	面层厚度
SufaceCourseMaterialType	IfcMaterialSelect	面层材料
BaceCourseThickness	IfcPositiveLengthMeasure	基层厚度
BaceCourseMaterialType	IfcMaterialSelect	基层材料
CushronThickness	IfcPositiveLengthMeasure	垫层厚度
CushronMaterialType	IfcMaterialSelect	垫层材料
SubgradeThickness	IfcPositiveLengthMeasure	道基厚度
SubgradeMaterialType	IfcMaterialSelect	道基材料
ConstructionDate	IfcDateTime	铺设时间

(Pset_PavementDiseaseCommon 属性) 如表 3 所示, 该属性集主要用于记录道面运维所需的道面板类型、总数、道面板损坏的种类、程度及对应的数量、检测时间、人员等信息。以刚性道面为例, 将病害种类划分为 15 类, 根据标准的已有分类按“1 纵向、横向和斜向裂缝, 2 角隅断裂, 3 破碎板或交叉裂缝, 4 沉陷或错台, 5 胀裂, 6 填缝料损坏, 7 接缝破碎, 8 唧泥和板底脱空, 9 耐久性裂缝, 10 收缩裂缝, 11 坑洞, 12 起皮、龟裂和细微裂纹, 13 板角剥落, 14 小补丁, 15 大补丁或开挖补丁”的顺序进行 1～15 的编号, 实现文字信息与数据的数字化映射, 同时根据 15 种病害类型的重度、中度、轻度 3 等程度进行属性信息的添加与描述。

3) 道面板评价参数集。道面板评价参数集 (Pset_PavementElementEvaluation 属性) 如表 4 所示, 该属性用于存储道面板相关评价指标数据, 在运维平台中将道面板损坏的相关数据信息进行计算, 并把计算后的指标与结果进行属性整合与管理, 其属性定义包括道面板的 PCI、SCI、PCN、IRI 等数据信息, 实际应用时还可根据不同的评价指标进行扩展。

4) 道面板运维数据参数集。道面板运维信息属性集 (Pset_PavementDate 属性) 如表 5 所示, 该属性用于存储道面板在运维使用过程中的维修数据信息, 包括维修板块编号 ID、维修时间、维修人员、维修材料、处理方式等。

考虑运维过程中数据信息的动态变化, 所采集的道面板损坏种类、损坏程度、损坏数量、损坏坐标及在维修维护过程中产生的相关数据信息均需

表 3　Pset_PavementDiseaseCommon 属性定义

Table 3　Pset_PavementDiseaseCommon property definition

属性名称	属性类型	数据类型	说明
MaterialType	IfcPropertySingleValue	IfcMaterialSelect	跑道类型
DamageQuantity	IfcPropertyTableValue	IfcCountMeasure	单元板块损坏数量
PlateNumber	IfcPropertyTableValue	IfcCountMeasure	单元板块总数
DamageDensity	IfcPropertyTableValue	IfcCountMeasure	损坏密度
DamageType	IfcPropertyTableValue	IfcLabel	损坏类型
DamageCoordinates	IfcPropertyTableValue	IfcLabel	损坏道面板坐标
1Cracks:Longitudinal,Transversa and Diagonal	IfcPropertySingleValue	IfcLabel	1纵向、横向和斜向裂缝
2Corner Break	IfcPropertySingleValue	IfcLabel	2角隅断裂
3Shattered Slab/Intersecting Cracks	IfcPropertySingleValue	IfcLabel	3破碎板或交叉裂缝
4Settlement or Faulting	IfcPropertySingleValue	IfcLabel	4沉陷或错台
5Blowup	IfcPropertySingleValue	IfcLabel	5胀裂
6Alkali Silica Reaction(ASR)	IfcPropertySingleValue	IfcLabel	6填缝料损坏
7Joint Seal Damage	IfcPropertySingleValue	IfcLabel	7接缝破碎
8Pumping and bottom	IfcPropertySingleValue	IfcLabel	8唧泥和板底脱空
9Durability(“D”) Cracking	IfcPropertySingleValue	IfcLabel	9耐久性裂缝
10Shrinkage Cracking	IfcPropertySingleValue	IfcLabel	10收缩裂缝
11Pot hole	IfcPropertySingleValue	IfcLabel	11坑洞
12Scaling	IfcPropertySingleValue	IfcLabel	12起皮、龟裂和细微裂缝
13Spalling(Corner)	IfcPropertySingleValue	IfcLabel	13板角剥落
14Patching, Small	IfcPropertySingleValue	IfcLabel	14小补丁
15Patching, Large and Utility Cuts	IfcPropertySingleValue	IfcLabel	15大补丁或开挖补丁
SurveyDate	IfcPropertyTableValue	IfcDateTime	检测采集时间
SurveyStaff	IfcPropertyTableValue	IfcLabel	检测采集人员

表 4　Pset_PavementElementEvaluation 属性定义

Table 4　Pset_PavementElementEvaluation property definition

属性名称	属性类型	数据类型	说明
PavementConditionIndex	IfcPropertyTableValue	IfcCountMeasure	PCI
SubgradeConditionIndex	IfcPropertyTableValue	IfcCountMeasure	SCI
PavementClassficationNumber	IfcPropertyTableValue	IfcCountMeasure	PCN
InternationalRoughnessIndex	IfcPropertyTableValue	IfcCountMeasure	IRI

表5 Pset_PavementDate属性定义

Table 5 Pset_PavementDate property definition

属性名称	属性类型	数据类型	说明
TargetID	IfcPropertyTableValue	IfcLabel	维修板块编号ID
MaintainDate	IfcPropertyTableValue	IfcDateTime	维修时间
MaintainStaff	IfcPropertyTableValue	IfcLabel	维修人员
MaintainMaterialType	IfcPropertyTableValue	IfcCountMeasure	维修材料
ProcessMode	IfcPropertyTableValue	IfcLabel	处理方式
Remarks	IfcPropertyTableValue	IfcPositiveLengthMeasure	备注

与模型信息相对应，以便实现运维动态管理，为数字孪生跑道提供基础数据。因此，属性类型采用IfcPropertyTableValue，该类型具有唯一性，同时，IfcPropertyTableValue实体的数据为DefiningValues和DefinedValues，分别可以存储运维时间和不同时间对应的运维情况。

2 机场道面结构IFC实体信息传递方法及验证

在明确IFC信息扩展架构后，需对IFC实体属性进行扩展，同时对信息的传递方法进行研究与验证，以避免机场道面信息在不同软件、不同环境传递时丢失，发挥IFC标准在openBIM中的价值。

2.1 信息扩展与传递方法

IFC属性集的扩展及信息集成管理有3种途径。第1种为基础模型族扩展法，即通过BIM相关建模软件建立族模型，通过添加分类属性信息，实现族模型与属性集的映射。第2种为IFC解析扩展法，即通过对IFC中性文件的解析，寻找对应的属性关系，并添加定义新的属性集。2种方法和途径都可以实现信息的扩展，但在操作难度和兼容性方面都存在不同的问题[19]。

基础模型族扩展法可采用商用BIM建模软件（如Revit）实现新属性的添加，操作起来较为容易，但由于IFC内包含信息的领域较广，同类型模型在不同领域描述方式及作用不同，导致BIM在机场建模过程中导出的IFC中性文件存在数据遗失情况。而IFC解析扩展法借助IFC中性文件进行属性的添加，使用者需具备较高的编程能力，易用性差。前2种方法信息传递的过程依靠几何信息与3D物体的相关转换，但IFC中性文件对几何的描述与BIM建模对物体模型的描述存在差异，导致信息传递时存在不同程度的属性遗失问题。于是，通过前2种方法的结合提出第3种“族模型IFC解析扩展法”，即通过BIM建模软件在建立新的族模型时添加扩展新的属性，并对IFC文件进行解析，将丢失信息以属性集的形式进行二次添加，所提出的新属性传递流程如图6所示，具体步骤如下：

1）利用BIM建模软件建立满足适用于机场跑道系统设计要求的道面板族模型（可采用.rfa格式文件），将所需属性信息（如机场道面结构运维所需的结构层信息、病害信息）进行整理和集合，按属性集进行分类，在软件环境下（如Revit）进行部分属性集的一次添加。

2）将新建道面板族BIM模型（.rvt文件）导出为IFC中性文件，通过读取文本可以对比验证出相关属性是否得到扩展，如图7所示，选中部分为新增的“PCN=95、SCI=80”属性。

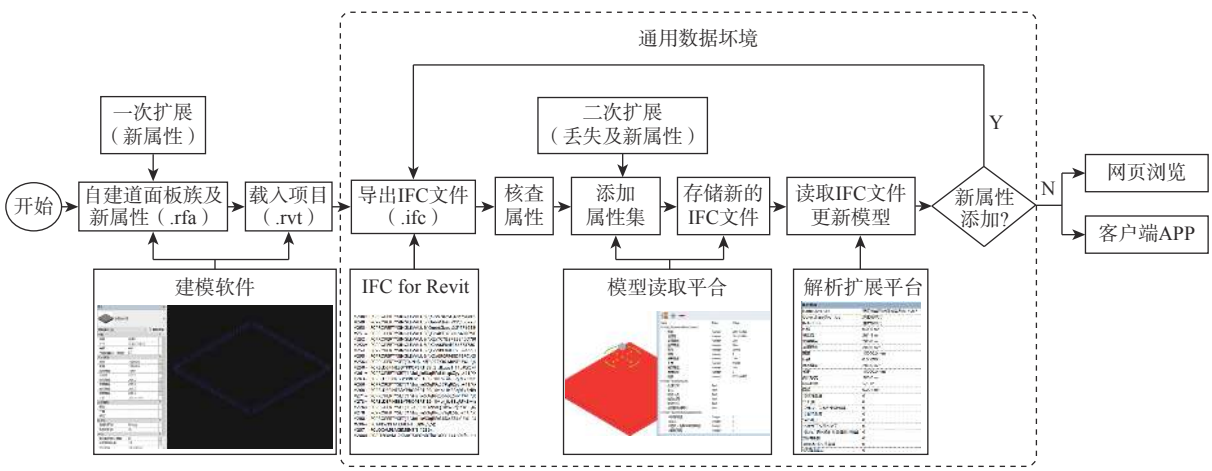


图6 族模型IFC解析扩展法总流程

Fig. 6 General flow chart of family model IFC analytical expansion method

```
#237=IFCPROPERTYSINGLEVALUE('\X2\65CF\X0\', $, IFCLABEL('\X2\90539762677
#238=IFCPROPERTYSINGLEVALUE('\X2\65CF4E0E7C7B578B\X0\', $, IFCLABEL('\X2
#239=IFCPROPERTYSINGLEVALUE('\X2\7C7B522B\X0\', $, IFCLABEL('\X2\5E3889C
#240=IFCPROPERTYSINGLEVALUE('\X2\7C7B578B\X0\',$, IFCLABEL('\X2\9053976
#241=IFCPROPERTYSINGLEVALUE('\X2\7C7B578B\X0\ID', $, IFCLABEL('\X2\90539
#242=IFCPROPERTYSINGLEVALUE('PCN', $, IFCINTEGER(95), $);
#243=IFCPROPERTYSINGLEVALUE('SCI', $, IFCREAL(80.), $);
#244=IFCPROPERTYSINGLEVALUE('\X2\57AB5C4267508D281\X0\', $, IFCTEXT(''), $);
#245= IFCPROPERTYSINGLEVALUE('\X2\57FA5C4267508D28\X0\', $, IFCTEXT(''), $);
#246= IFCPROPERTYSINGLEVALUE('\X2\635F574F7A0B5EA6\X0\', $, IFCTEXT(''), $);
#247= IFCPROPERTYSINGLEVALUE('\X2\635F574F7C7B578B\X0\', $, IFCTEXT(''), $);
#248= IFCPROPERTYSINGLEVALUE('\X2\675065997C7B578B\XO\', $, IFCTEXT(''), $);
```

图 7　新增属性道面板 IFC 中性文件

Fig. 7　Part files of added property pavement IFC model

3) 对 IFC 文件进行解析, 读取 IFC 文件中的几何与属性信息。对解析后 IFC 中性文件中所缺失的新属性进行二次添加, 同时可对原有属性和新添加的属性进行整合并提取合成电子表单, 将新扩展和整理属性导出形成新的 IFC 中性文件, 进行后续运维管理所需的计算与分析。

4) 使用运维软件读取道面板新的 IFC 中性文件, 并利用传递的属性信息进行运维管理[20-22], 如进行道面状态的计算、风险的预测分析[23], 实现运维管理。当需要动态更新信息时, 在步骤 3)中持续对 IFC 中性文件进行新实体和新属性的更新, 并与运维时间映射, 形成道面信息动态数据库。

5) 基于该流程所搭建的通用数据环境, 最终可根据运维管理的需求呈现在 Web 端或 APP 内。

2.2　传递方法的信息丢失验证

对属性信息扩展的基础模型族扩展法、IFC 解析扩展法及族模型 IFC 解析扩展法信息传递过程中的属性信息丢失率进行分析。定义属性信息丢失率 η 为

$$\eta = 1 - \left(\frac{n_1}{n_0}\right) \times 100\% \tag{1}$$

式中: n_0 为传递前模型的属性信息个数; n_1 为传递后模型所保留的属性信息个数。

属性信息验证时, 采用支持 BIM 系列软件(如 Revit、Navisworks)或 BIM 解析平台, 进行 IFC 文件的输出和读取, 对读取后模型中属性信息的丢失情况进行检验。分析时, 分别对信息进行二次传递验证, 一次传递指基础建模软件或解析软件对建模后导出的 IFC 中性文件的第 1 次读取, 二次传递指对一次传递的 IFC 文件进行编辑、添加属性等处理后, 继续传递至基础建模软件或解析软件的过程。各方法的属性信息丢失率结果如表 6 所示, 表中“已有族”指 IFC 文件中已具备实体标准模型信息传递情况, “机场道面族”为本文新建机场结构信息模型的信息传递情况。

可以看出, 对于已有族模型采用 IFC2×3 版本进行传递时, 属性信息丢失率都较低。基础模型族扩展法一次传递和二次传递过程中属性信息丢失率都为 0%; IFC 解析扩展法的 2 次传递属性信息丢失率也都在 5% 以内; 结合两者扩展方法的族模型 IFC 解析扩展法可以保证一次传递的属性信息丢失率为 0%, 二次传递的属性信息丢失率降低维持在 5%。但是, 采用 IFC4 版本进行传递时, 属性信息丢失率较高, 使用基础模型族扩展法可以保证信息传递的高效性, 但利用 IFC 解析扩展法和族模型 IFC 解析扩展法传递时属性信息丢失率为 100%, 原因在于目前 IFC2×3 仍为 IFC 标准最为稳定与成熟的版本。

对于新建的道面结构族采用 IFC2×3 版本进行传递时, 基础模型族扩展法一次传递属性信息丢失率高达 84.21%, 在二次传递时属性信息丢失率达到 100%, IFC 解析扩展法一次传递时属性未丢失, 但在二次传递的过程中, 新建的相关属性信息丢失率达到 100%。上述丢失的信息主要为本文新建的机场道面族的新增属性, 如跑道各层的材料类型、道面板的损坏类型、损坏程度、损坏数量、损坏密度等, 严重制约了跑道信息模型在不同阶段、不同

表 6　不同信息传递方法属性信息丢失率比较

Table 6　Comparison of information loss rate for attributes of different information transmission methods

方法	传递次数	传递版本	η/%	
			已有族(未新增属性)	机场道面族(新增属性)
基础模型族扩展法	一次传递	IFC2×3	0	84.21
		IFC4	0	100
	二次传递	IFC2×3	0	100
		IFC4	0	100
IFC解析扩展法	一次传递	IFC2×3	5	0
		IFC4	100	100*
	二次传递	IFC2×3	5	100*
		IFC4	100	100*
族模型IFC解析扩展法	一次传递	IFC2×3	0	0
		IFC4	100	100*
	二次传递	IFC2×3	5	16.7
		IFC4	100	100*

注: “*” 表示新增属性完全丢失而原有属性全部传递。

使用方间的信息传递。既有方法导致信息丢失的根源在于缺乏应用于机场领域的IFC标准扩展，导致信息模型与标准格式间缺乏映射关系与映射方法；此外，不同模型环境中缺乏机场领域的统一数据名称标准与分类（数据字典），导致模型信息与IFC标准映射错误，从而造成信息传递时的丢失。

从表6可知，族模型IFC解析扩展法一次传递属性未丢失，二次传递属性信息丢失率能降低为16.7%，保留了大部分道面运维所需信息，可见该方法属性信息丢失率较低，在保证正常传递的过程中，将相关属性信息进行最大限度的保留。3种方法在使用IFC4版本进行传递时，基础模型族扩展法属性信息丢失率增高至100%，对于IFC解析扩展法和族模型IFC解析扩展法仅丢失新增属性，保留传递了族模型原有的所有属性。

根据上述分析可知，对于IFC标准，既有实体属性信息丢失率比新增实体属性信息丢失率低，同时IFC版本的差异也将导致属性传递结果，IFC2×3版本可在传递过程中保留更多数据信息。影响信息传递过程中属性信息丢失的因素主要有实体类型、传递方法、传递次数、数据标准完备性、IFC版本等。

2.3 实例应用

为进一步验证本文提出的符合机场跑道系统的属性扩展方法，采用西部某4E机场跑道进行基于IFC标准的实体扩展与病害信息管理。该机场拥有1条跑道，2条滑行道。跑道全长3 600 m，宽45 m，跑道进行日常维护与修理，同时进行了2次完整盖被。在该机场跑道运行管理过程中，运维信息仍采用纸质文件记录，历史数据难以获取，道面病害管理也以人工记录、文字报告形式存储，信息化程度低，场务管理人员对于跑道病害信息化管理与高效维护的需求迫切。针对上述问题，采用本文提出的自定义道面板族单元进行完整跑道的模型建立，应用BIM技术完成机场跑道病害信息的集成管理。

首先，利用BIM建模软件建立道面板族模型，并添加相关属性；然后，基于图6中的族模型IFC解析扩展法，利用IFC解析平台对导出的IFC中性文件进行添加和修改；最后，利用扩展后的新IFC中性文件进行读取，建立完整跑道数字化模型，在模型中增加运维所需新属性，利用IFC文件进行信息传递，通过管理系统对运维信息进行管理和方法验证。

图8为增加运维属性后对道面病害情况读取情况。可以看出，采用族模型IFC解析扩展法扩展的相关新属性（Pset_RunwayBlockCommon、Pset_PavementDiseaseCommon、Pset_PavementElementEvaluation、Pset_PavementDate）在模型中可以传递，同时利用BIM技术可将不同位置道面板的病害信息与模型进行

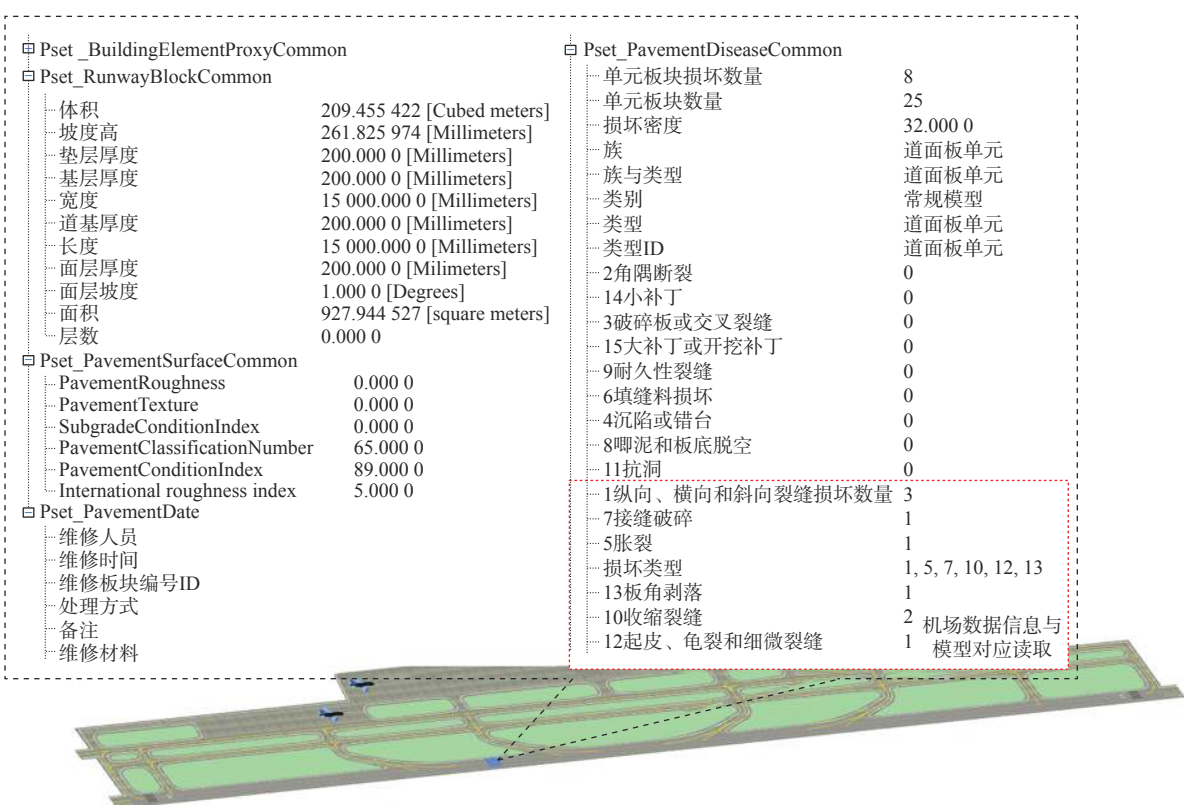


图8 基于IFC标准的某西部机场道面运行状况评价管理

Fig. 8 Pavement condition evaluation and maintenance of a western airport based on IFC standard

对应并存储。图 8 中展示的是不同道面评价单位的病害信息集成情况, 可以看出, 所选道面板单元块所存在病害标号为“1、5、7、10、12、13”6 种病害, 同时评价单元中 6 种病害所对应的道面板数量为“3、1、1、2、1、1”; 相关评价指标(如 PCI、PCN 等)也依次有所体现, 实现了道面结构运维病害信息的可视化高效管理, 进一步结合历史数据, 可实现机场道面服役趋势的分析与状态预警, 为运行安全服务。

3　结　论

本文基于 IFC 标准对机场道面结构信息模型进行建立, 并展开了相关研究, 主要结论如下:

1) 提出了机场跑道系统的 IFC 扩展方法与流程, 验证了该方法的适用性, 对机场道面结构进行 IFC 扩展, 对于实现 BIM 技术在机场全生命周期的应用较为必要。

2) 基于 BIM 技术形成了参数化的机场道面结构信息模型, 完善了面向运维的属性信息框架; 提出了一种适用于道面运维的属性信息扩展方法——族模型 IFC 解析扩展法, 实现建设阶段机场道面模型向运维阶段的信息传递与信息复用。

3) 对基础模型族扩展法、IFC 解析扩展法、族模型 IFC 解析扩展法的属性信息丢失率进行了对比分析, 使用基础 BIM 建模软件扩展的属性信息丢失率高达 84.21%; 利用 IFC 中性文件扩展属性的方法在首次读取属性时可保证数据的属性信息丢失率为 0%, 但二次传递新增属性信息全部丢失; 利用本文方法可将数据信息首次传递的属性信息丢失率降低到 0%, 二次传递的属性信息丢失率降低为 16.7%。影响机机场道面模型传递属性信息丢失率的因素包括实体类型、传递方法、传递次数、数据标准完备性、IFC 版本等。

BIM 应用较为广泛, IFC 标准的建立也日益完善, 但机场工程仍有大量专业实体没有纳入 IFC 标准, 同时其全生命周期中产生的数据也较为复杂, 亟须在扩展新实体的同时对属性信息进行完善。此外, 如何依托模型与数据标准进行原始数据的高效获取、道面实时状况的合理评价是进一步的研究方向。

参考文献（References）

[1] ISO. Industry foundation classes (IFC) for data sharing in the construction and facility management industries-Part 1: Data schema: ISO 16739-1-2024[S]. Geneva: ISO, 2024.

[2] 赖华辉, 侯铁, 钟祖良, 等. BIM 数据标准 IFC 发展分析[J]. 土木工程与管理学报, 2020, 37(1): 126-133.
LAI H H, HOU T, ZHONG Z L, et al. Analysis on development of BIM data standard IFC[J]. Journal of Civil Engineering and Management, 2020, 37(1): 126-133(in Chinese).

[3] 赖华辉, 侯铁, 钟祖良, 等. 基于知识图谱的国内 IFC 研究综述[J]. 土木工程与管理学报, 2020, 37(3): 96-102.
LAI H H, HOU T, ZHONG Z L, et al. Review of domestic IFC research based on mapping knowledge domains[J]. Journal of Civil Engineering and Management, 2020, 37(3): 96-102(in Chinese).

[4] 赖华辉, 邓雪原, 刘西拉. 基于 IFC 标准的 BIM 数据共享与交换[J]. 土木工程学报, 2018, 51(4): 121-128.
LAI H H, DENG X Y, LIU X L. IFC-based BIM data sharing and exchange[J]. China Civil Engineering Journal, 2018, 51(4): 121-128 (in Chinese).

[5] 郭红领, 周颖, 叶啸天, 等. IFC 数据模型至关系型数据库模型的自动映射[J]. 清华大学学报 (自然科学版), 2021, 61(2): 152-160.
GUO H L, ZHOU Y, YE X T, et al. Automated mapping from an IFC data model to a relational database model[J]. Journal of Tsinghua University (Science and Technology), 2021, 61(2): 152-160 (in Chinese).

[6] YABUKI N, LI Z T. Development of new IFC-BRIDGE data model and a concrete bridge design system using multi-agents[C]//Proceedings of the Intelligent Data Engineering and Automated Learning. Berlin: Springer, 2006: 1259-1266.

[7] LEE S H, KIM B G. IFC extension for road structures and digital modeling[J]. Procedia Engineering, 2011, 14: 1037-1042.

[8] BORRMANN A, FLURL M, JUBIERRE J R, et al. Synchronous collaborative tunnel design based on consistency-preserving multi-scale models[J]. Advanced Engineering Informatics, 2014, 28(4): 499-517.

[9] 钟宇, 陈健, 陈国良, 等. 基于建筑信息模型技术的盾构隧道结构信息模型建模方法[J]. 岩土力学, 2018, 39(5): 1867-1876.
ZHONG Y, CHEN J, CHEN G L, et al. Shield tunnel structure information modelling method based on building information modeling technology[J]. Rock and Soil Mechanics, 2018, 39(5): 1867-1876 (in Chinese).

[10] 王怀松, 王学林, 邹少文. 铁路信号 IFC 国际标准的研究与制定[J]. 铁道标准设计, 2019, 63(8): 162-168.
WANG H S, WANG X L, ZOU S W. Research and development of IFC international standards for railway signaling[J]. Railway Standard Design, 2019, 63(8): 162-168(in Chinese).

[11] 冯山群, 杨绪坤, 马永昌, 等. 基于 IFC 扩展的铁路隧道 BIM 数据存储标准研究[J]. 铁路技术创新, 2015(6): 24-27.
FENG S Q, YANG X K, MA Y C, et al. Research on BIM data storage standard of railway tunnel based on IFC extension[J]. Railway Technical Innovation, 2015(6): 24-27(in Chinese).

[12] 胡振中, 田佩龙, 李久林. 基于 IFC 的传感器信息存储与应用研究[J]. 图学学报, 2018, 39(3): 522-529.
HU Z Z, TIAN P L, LI J L. Research on IFC-based storage and application of sensor information[J]. Journal of Graphics, 2018, 39(3): 522-529(in Chinese).

[13] 董强, 刘英林, 李志辉. BIM 技术在机场航站区建设施工中的应用[J]. 建筑技术, 2015, 46(12): 1118-1121.
DONG Q, LIU Y L, LI Z H. Application of BIM technology in airport terminal area construction[J]. Architecture Technology, 2015, 46(12): 1118-1121(in Chinese).

[14] 戴轩, 王瀚雪, 蔡靖, 等. 我国机场 BIM 技术应用现状及运维阶段信息传递方法初探[C]// 2022 世界交通运输大会. 北京: 人民交通出版社股份有限公司, 2022: 6.
DAI X, WANG H X, CAI J, et al. A preliminary study on the application status and information transmission methods of BIM technology in airports in China[C]//Proceedings of the 2022 World Transport Congress. Beijing: People's Communications Publishing House Co. , Ltd., 2022: 6(in Chinese).

[15] 周俊波, 黄亚栋, 谭慧, 等. 鄂州民用机场配套工程中 Civil 3d 软件的应用[J]. 中华建设, 2020(增刊 1): 111-113.
ZHOU J B, HUANG Y D, TAN H, et al. Application of Civil 3d software in Ezhou civil airport supporting project[J]. China Construction, 2020(Sup 1): 111-113(in Chinese).

[16] 赵健, 房健康, 宋扬, 等. BIM 技术在南京禄口国际机场运维阶段的深入应用[J]. 土木建筑工程信息技术, 2015, 7(2): 50-55.
ZHAO J, FANG J K, SONG Y, et al. Application of BIM technology in the operation of Nanjing lukou international airport[J]. Journal of Information Technology in Civil Engineering and Architecture, 2015, 7(2): 50-55(in Chinese).

[17] 董政民. BIM 技术在航站楼运维管理中的应用: 以浦东国际机场 T1 航站楼为例[J]. 建筑经济, 2013, 34(9): 91-93.
DONG Z M. BIM application in the airport terminal operation and maintenance management: Pudong International Airport terminal T1 as an example[J]. Construction Economy, 2013, 34(9): 91-93(in Chinese).

[18] 杨冬, 刘勇, 邹家撇, 等. BIM 技术在广州白云国际机场综合交通枢纽项目中的应用[J]. 施工技术, 2018, 47(6): 133-136.
YANG D, LIU Y, ZOU J P, et al. Application of BIM technology in the Guangzhou Baiyun International Airport integrated transportation hub project[J]. Construction Technology, 2018, 47(6): 133-136(in Chinese).

[19] AI S D, MAO N, HE X. Research on rail BIM data storage standards based on IFC4 extension[J]. Railway Technology Innovation, 2017(4): 48-54.

[20] TANG F L, MA T, ZHANG J H, et al. Integrating three-dimensional road design and pavement structure analysis based on BIM[J]. Automation in Construction, 2020, 113: 103152.

[21] 吴生海, 刘陕南, 刘永晓, 等. 基于 Dynamo 可视化编程建模的 BIM 技术应用与分析[J]. 工业建筑, 2018, 48(2): 35-38.
WU S H, LIU S N, LIU Y X, et al. Application and analysis of BIM technology based on Dynamo visual programming modeling[J]. Industrial Construction, 2018, 48(2): 35-38(in Chinese).

[22] 王茹, 权超超. 公路立交 BIM 参数化快速精确建模方法研究[J]. 图学学报, 2019, 40(4): 766-770.
WANG R, QUAN C C. Research on fast and accurate modeling method of BIM parameterization for highway interchange[J]. Journal of Graphics, 2019, 40(4): 766-770(in Chinese).

[23] 李岳, 胡宇祺, 蔡靖, 等. 基于变权重-正态云模型的飞机轮胎滑水风险[J]. 北京航空航天大学学报, 2023, 49(9): 2299-2305.
LI Y, HU Y Q, CAI J, et al. Hydroplaning risk of aircraft tire based on variable weight theory-normal cloud model[J]. Journal of Beijing University of Aeronautics and Astronautics, 2023, 49(9): 2299-2305(in Chinese).

Storage and transfer method of airport pavement structure information model based on IFC standard

CAI Jing[1], WANG Hanxue[1], DAI Xuan[1,*], LI Yue[1], MA Lei[2]

(1. School of Transportation Science and Engineering, Civil Aviation University of China, Tianjin 300300, China;
2. Guangdong Airport Authority, Guangzhou 510406, China)

Abstract: To address the issues of information transfer and data sharing in the full-life-cycle building information modeling (BIM) application of the airport runway system, the industry foundation classes (IFC) standard was extended to the airport pavement structure. The IFC entity framework of the airport runway system was summarized. Besides, the rigid runway pavement entity considering the operation and maintenance stage was proposed. The graphical information and non-graphical information of the proposed entity were expanded. In addition, an analysis was conducted on the rate of attribute information loss during information model transfer, as well as the most effective way to add and transmit attribute information using the IFC standard. The study shows that the model based on the uniform IFC standard can transfer and share information about airport runway systems effectively. The information loss rate of the family parameter extension method during runway pavement model transfer is as high as 84.21%. The IFC file extension method can ensure the model information integrity during the first transfer process. However, the model information is lost during repeated transfer. A proposed method, combining both the family parameter extension and IFC file parsing, can achieve the model information integrity during the first transfer. In addition, the information loss rate is less than 16.7% during the second transfer. The information loss rate is also influenced by different versions of the IFC standard. It is necessary to extend the IFC entity in the airport domain. As a whole, the proposed information transfer method is appropriate for airport runway lifecycle information management. The research can provide support for the establishment of a new IFC entity in the airfield area, and make a technical reference for digital airport construction.

Keywords: airport pavement structure; industry foundation classes; common data environment; information model; information transfer

Received: 2022-09-29; **Accepted**: 2023-04-14; **Published Online**: 2023-05-05 16:36

URL: link.cnki.net/urlid/11.2625.V.20230505.1432.007

Foundation items: Tianjin Natural Science Foundation (21JCQNJC00850); The Fundamental Research Funds for the Central Universities (3122022043)

* **Corresponding author**. E-mail: tianjindaixuan@126.com

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2022.0777

停留时间与飞行方向对飞行员警觉性的影响

李敬强*，胡超，张希凝，张璐，刘安南
(中国民航大学 安全科学与工程学院，天津 300300)

摘 要：基于疲劳风险管理系统的应用正在全球航空业中迅速展开，在防疫豁免期间，收集实际运行模式下64名国际豁免航班飞行员在不同中转停留时间下的主观自评、客观测试和睡眠信息等数据。运用SPSS27.0统计学软件，利用相关性分析和重复测量方差分析的方法，深入地分析不同中转停留时间及飞行方向对国际豁免航班飞行员警觉性的影响，并用Origin软件绘制相关性系数图和半小提琴图。结果表明：停留时间为1～2 h时，国际豁免航班飞行员的警觉性水平明显较低，向东飞行的飞行员要比向西飞行的警觉性水平更低。这些发现有助于航空公司在防疫豁免运行环境中监测和分析国际豁免航班飞行员的警觉性，为加强中转航班过站保障措施提供借鉴。

关 键 词：风险管理；国际豁免航班；警觉性；飞行员；停留时间；飞行方向
中图分类号：X949
文献标志码：A **文章编号**：1001-5965(2024)10-3054-08

中转停留时间是指飞行员从一个任务周期结束到为下一个任务开始做准备的时间，为飞行员提供了一个从高负荷值勤时间中恢复和准备的机会，是影响飞行员警觉性的一个重要因素[1]。中国民用航空局发布的《疫情期间豁免机组成员值勤期、飞行时间限制的实施办法》(以下简称“豁免办法”)[2]于2020年12月2日正式颁布实施，要求国际豁免航班在延长机组飞行时间限制的同时，在原有的机组基础上再增加1～2套机组，并且在目的地中转停留后直接飞回出发地，不在当地过夜。在该运行模式下，飞行员的中转停留时间明显减少。由于飞行方向不同导致飞行员自身调节生理节律能力的差异性，再加上各个国家防疫的相关政策多变，飞行员在短时间内需要及时应对，精神压力不免随之增加。这些在一定程度上都影响到了飞行员中转期间的休息及回程时的警觉性。鉴于此，鉴别不同停留时间及飞行方向对飞行员警觉性的影响是航空公司建设疲劳风险管理系统(fatigue risk management system, FRMS)过程中识别疲劳危险源的必要环节，同时也为其加强中转过站的疲劳缓解措施的制定提供借鉴。

近年来，很多学者对飞行员的警觉性开展了一系列研究。Reis等[3]利用自我反应疲劳严重程度量表，对比调查了葡萄牙航空公司长途和中短程飞行员的警觉性，分析得出，相较于长途飞行的飞行员，中短程飞行员的总体警觉性程度更低，更易疲劳。van Drongelen等[4]依托对502名飞行员问卷调查的数据，使用单变量和多变量逻辑回归分析确定了与飞行员警觉性相关联的危险因素。于洮等[5]通过分析陆军航空兵某旅142名飞行员的问卷调查数据，得出了睡眠时间、睡眠效率等与飞行员的脑力疲劳和警觉性水平有着密切的关系。曾绍洁等[6]基于多个航空兵部队飞行员的调研数据，利用Logistic逐步回归分析探索了睡眠和疲劳之间的危险因素，并发现全封闭式管理前后的空军飞行员睡眠质量和警觉性状态具有显著差异。刘艳等[7]依托自主研发的飞行员睡眠-疲劳-脑力负荷测量APP对112名飞行员进行了研究，结果发现脑力负

收稿日期：2022-09-14；**录用日期**：2022-12-04；**网络出版时间**：2022-12-27 17:31
网络出版地址：link.cnki.net/urlid/11.2625.V.20221226.1511.003
基金项目：国家自然科学基金与民航局联合基金(U1933122)；中央高校基本科研业务费专项资金(3122019186)
* **通信作者**. E-mail：jqli@cauc.edu.cn

引用格式：李敬强，胡超，张希凝，等. 停留时间与飞行方向对飞行员警觉性的影响[J]. 北京航空航天大学学报，2024，50(10)：3054-3061. LI J Q，HU C，ZHANG X N，et al. Influence of layover length and flight direction on pilots' alertness[J]. Journal of Beijing University of Aeronautics and Astronautics，2024，50(10)：3054-3061 (in Chinese).

荷、睡眠质量和警觉性状态都会对飞行员作训期间的身心健康产生影响。根据以上文献可知, 对于飞行员警觉性的研究, 目前国内外学者侧重于采用问卷或量表的形式研究飞行员的睡眠质量, 很少涉及到中转停留时间的影响。由于商业航空受到监管机构的监督、航空公司排班及飞行员自身条件的限制等条件影响, 从而使收集飞行员在实际运行环境中的警觉性数据变得比较困难。特别地《豁免办法》[2]是近两年在疫情防控新常态下才颁布实施的, 因而针对停留时间对实际运行环境中国际豁免航班飞行员警觉性的影响程度尚不清楚。

本文深入地分析 64 名国际豁免航班飞行员在实际运行环境中的警觉性数据, 探究不同中转停留时间及飞行方向对飞行员警觉性的影响, 旨在为航空公司科学制订排班计划、优化中转过站的管理程序及加强服务保障等提供借鉴。

1　警觉性测试

1.1　试验方法

1.1.1　试验对象

某航空公司 64 名 B787 飞行员自愿参加了试验, 年龄范围为 26 ～37 岁, 平均年龄为 34.38±5.89 岁。飞行员总飞行小时数范围为 9 600 ～23 000 h, 平均总飞行小时数为 13 365.43 h。飞行员健康状态良好, 32 名飞行员执飞北京–洛杉矶–北京的国际豁免航班, 其他的 32 名飞行员执飞北京–阿姆斯特丹–北京的国际豁免航班。

所有参与研究的机组人员都是在获得其书面知情同意后作为志愿者自愿参加试验的。并且都得到了各自监管机构、工会和航空公司管理层的相应支持。

1.1.2　试验工具

1) 卡罗林斯卡嗜睡量表(Karolinsaka sleepiness scale, KSS)。KSS 采用 9 级制评分, 从 1～9 分中选出符合当下嗜睡程度的得分[8]。

2) Samn-Perelli 机组人员疲劳状态检查表 (Samn-Perelli fatigue checklist, SP)。SP 采用 7 级制评分。从 1～7 分中选出符合当下疲劳程度的得分[9]。

3) 安装有精神运动警觉性测试任务的终端平板。用来测试受试者的客观警觉性数据[10]。

4) ActiGraph 体动记录仪。用来监测受试者的睡眠数据[11]。

1.1.3　试验流程

飞行员被要求在报到后、航班起飞前、航班降落后 1 h 内对其自身的警觉性状态及疲劳感受进行评估。评估方式主要包括主观自评、客观测试及睡眠监测。

试验中所使用的飞行员警觉性评估的方式遵循国际民用航空组织 (International Civil Aviation Organization, ICAO) 所给出的疲劳风险管理指导建议[12]。使用 KSS 和 SP 来评估飞行员的主观警觉性水平: KSS 侧重评估受试者的嗜睡程度, 更多地反映受试者当下的困意,而 SP 是专门为机组人员设计的, 以工作/休息概况为基础评估机组人员的疲劳程度, 侧重于检测飞行和值班时间是否会导致机组人员疲劳, 更多地反映受试者当下的警觉性水平。将 KSS 和 SP 结合起来可以更加全面地评估飞行员的警觉性水平, 而且 KSS 和 SP 结合也是被 ICAO 极为推崇的, 并已经在航空业中得到了相应的验证[13]。

飞行员客观警觉性的测试主要是依托安装在机组人员手持终端平板上的 APP 进行测试, 内容主要是 10 min 的精神运动警觉性任务(psychomotor vigilance task, PVT)。本文选取最快 10% 反应时(fastest 10% of responses)和最慢 10% 反应时(slowest 10% of responses)来综合反映 PVT 测试的结果, 进而评估飞行员警觉性水平的变化[14]。

试验期间, 通过佩戴在每位参与者手腕上的 ActiGraph 体动记录仪来收集飞行员睡眠的客观数据, 监测其睡眠状态。同时机组人员还要完成一份睡眠日记, 其记录了机组人员的睡眠信息, 包括小睡和机上休息。使用如图 1 所示的 ActiLife 软件并结合睡眠日记以确定飞行员的 24 h 总睡眠时间(在目的地中转停留后值勤开始前的 24 h 睡眠时间)。

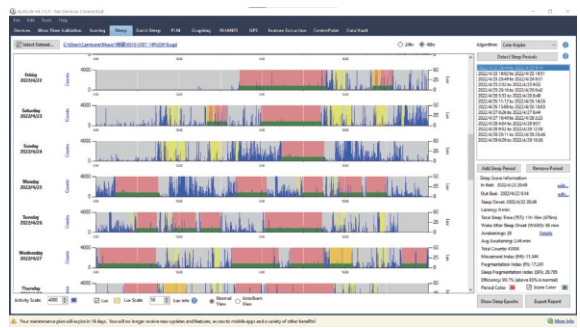


图 1　ActiLife 软件界面

Fig. 1　Interface of ActiLife software

1.2　数据收集

通过查询航空公司的排班信息表及机组人员的值班时间表等来收集国际豁免航班的相关信息。由于本次收集到的航班中转停留时间是 1～5 h(所有航班时间均转换为居住地时间), 故将中转停留的时间划分 1～2 h、2～3 h、3～4 h 和 4～5 h 这 4 个时间段。详细的国际豁免航班信息汇总如表 1 所示。

表1 航班信息汇总

Table 1 Summary of flight informations

航班	停留时间/h	飞行员人次	飞行方向	飞行时间/h	值勤时间/h	跨越时区/h
洛杉矶–北京	1～2	6	西飞	12.81±1.77	13.53±1.61	−16
洛杉矶–北京	2～3	12	西飞	13.60±1.55	13.82±1.59	−16
洛杉矶–北京	3～4	9	西飞	13.17±1.90	13.81±2.19	−16
洛杉矶–北京	4～5	5	西飞	13.04±1.57	14.32±2.27	−16
阿姆斯特丹–北京	1～2	6	东飞	11.54±0.61	11.62±0.68	+7
阿姆斯特丹–北京	2～3	9	东飞	11.46±0.65	12.08±0.69	+7
阿姆斯特丹–北京	3～4	10	东飞	11.40±0.43	11.94±0.53	+7
阿姆斯特丹–北京	4～5	7	东飞	12.15±0.16	13.01±0.35	+7

1.3 安全性能指标

在ICAO建议的疲劳风险管理体系中，将反映飞行员警觉性水平的指标称为安全性能指标[15]。本文收集到的中转停留后飞行员的安全性能指标信息如下：①起飞前KSS评分；②起飞前SP评分；③起飞前PVT性能(fastest 10% of responses和slowest 10% of responses)；④24 h总睡眠时间(total sleep time in 24 h periods, TST in 24 h)。

TST in 24 h反映了飞行员值勤前的24 h睡眠状态；KSS评分反映了飞行员的嗜睡程度；SP评分反映了飞行员的疲劳程度；fastest 10% of responses是行为活动中的最优表现，反映了飞行员的警觉水平的变化；slowest 10% of responses主要包括忽视范围内的响应时间，反映个体注意水平的变化[14]。如果飞行员在过去24 h内睡眠时间较短和清醒时间较长会使飞行员KSS评分较高，更易感到困倦，使SP评分较高，更易感到疲劳。进而导致飞行员fastest 10% of responses较长，警觉水平低下，导致slowest 10% of responses较长，注意水平低下。

2 数据处理方法

2.1 相关性分析

对收集的安全性能指标数据做相关性检验，以判断其之间是否具有一致性。基于KSS与SP均为有序变量，而客观测试和睡眠监测的数据是连续变量，因此，选用Spearman相关系数来计算安全性能指标数据之间的相关性。具体计算过程如下。

假设原始的数据x_i和y_i已经按从大到小的顺序排列，记$R_{(x)}$和$R_{(y)}$，其分别为x_i和y_i在排列后数据所在的位次。Spearman相关系数为

$$\rho=\frac{\frac{1}{n}\sum_{i=1}^{n}\left[\left(R_{(x_i)}-\overline{R}_{(x)}\right)\left(R_{(y_i)}-\overline{R}_{(y)}\right)\right]}{\sqrt{\left[\frac{1}{n}\sum_{i=1}^{n}\left(R_{(x_i)}-\overline{R}_{(x)}\right)^2\right]\left[\frac{1}{n}\sum_{i=1}^{n}\left(R_{(y_i)}-\overline{R}_{(y)}\right)^2\right]}} \tag{1}$$

式中：$\overline{R}_{(x)}$为x_i数据的平均位次，$\overline{R}_{(y)}$为y_i数据的平均位次。

更简单易行的计算式如下：

$$\rho=1-\frac{6\sum d_i^2}{n(n^2-1)} \tag{2}$$

式中：d_i为第i个数据对的位次值之差；n为总的观测样本数。

利用式(2)，运用统计学软件SPSS27.0，计算KSS与SP评分之间的相关性如表2所示，其中，KSS与SP的样本数均为64。

表2 KSS与SP相关性分析

Table 2 Correlation analysis of KSS and SP

指标	相关系数R		显著性p	
	KSS	SP	KSS	SP
KSS	1.000	0.874	0.000	0.000
SP	0.874	1.000	0.000	0.000

由表2可知，KSS与SP评分之间呈现出较强的相关性(R=0.874, p=0.000)，2种主观自评量表的关系紧密，飞行员的主观嗜睡程度在各个阶段与疲劳程度呈现出相反的趋势。说明当飞行员感受到困倦时，他当下更易感到疲劳。因此，嗜睡程度和疲劳程度所反映出的警觉性状态具有较高的一致性，可以将KSS和SP结合起来全面地评估飞行员的主观警觉性水平。

计算主观自评指标与客观测试指标之间的关系，从而判断主、客观数据是否具有一致性，其之间的相关性如表3所示。

表3 主观自评与客观测试指标间的相关性分析

Table 3 Correlation analysis of subjective rating and objective test indicators

指标类型	相关系数R		显著性p	
	KSS	SP	KSS	SP
fastest 10% of responses	0.262	0.253	0.013	0.016
slowest 10% of responses	0.227	0.240	0.032	0.023
TST in 24 h	−0.240	−0.326	0.023	0.002

2.2　差异性分析

双因素重复测量方差试验的数据模型 $\boldsymbol{X}$ 是一个结构为 $r\times s\times t$ 的三维矩阵, 其中, X_{ijk} 取自于总体指标 $N(\mu_{ij}, \sigma^2)$, i=1, 2, ⋯, r, j=1, 2, ⋯, s, k=1, 2, ⋯, t。

根据式 (3)～ 式 (12) 计算样本总均值$\overline{X}$。因素 A 对应的样本均值总偏差平方和为 S_{T}, 误差平方和为 S_{E}, 因素 A 和因素 B 的效应平方和分别为 S_{A} 和 S_{B}, 以及因素 A 与因素 B 交互效应平方和为 $S_{\mathrm{A\times B}}$, 样本总均值$\overline{X}$为

$$\overline{X}=\frac{1}{rst}\sum_{i=1}^{r}\sum_{j=1}^{s}\sum_{k=1}^{t}X_{ijk} \tag{3}$$

式中: r 为 A 因素的水平数; s 为因素 B 的水平数; t 为试验次数; X_{ijk} 为样本。因素 A 与因素 B 组合因素样本均值$\overline{X}_{ij}$为

$$\overline{X}_{ij}=\frac{1}{t}\sum_{k=1}^{t}X_{ijk} \tag{4}$$

因素 A 对应的样本均值$\overline{X_{i\cdot}}$为

$$\overline{X}_{i\cdot}=\frac{1}{st}\sum_{j=1}^{s}\sum_{k=1}^{t}X_{ijk} \tag{5}$$

因素 B 对应的样本均值$\overline{X_{\cdot j}}$为

$$\overline{X}_{\cdot j}=\frac{1}{rt}\sum_{i=1}^{r}\sum_{k=1}^{t}X_{ijk} \tag{6}$$

样本总偏差平方和 S_{T} 为

$$S_{\mathrm{T}}=\sum_{i=1}^{r}\sum_{j=1}^{s}\sum_{k=1}^{t}\left(X_{ijk}-\overline{X}\right)^2 \tag{7}$$

样本误差平方和 S_{E} 为

$$S_{\mathrm{E}}=\sum_{i=1}^{r}\sum_{j=1}^{s}\sum_{k=1}^{t}\left(X_{ijk}-\overline{X}_{ij}\right)^2 \tag{8}$$

因素 A 的效应平方和 S_{A} 为

$$S_{\mathrm{A}}=st\sum_{i=1}^{r}\left(X_{i\cdot}-\overline{X}\right)^2 \tag{9}$$

因素 B 的效应平方和 S_{B} 为

$$S_{\mathrm{B}}=rt\sum_{j=1}^{s}\left(X_{\cdot j}-\overline{X}\right)^2 \tag{10}$$

因素 A 与因素 B 的交互效应平方和 $S_{\mathrm{A\times B}}$ 为

$$S_{\mathrm{A\times B}}=t\sum_{i=1}^{r}\sum_{j=1}^{s}\left(X_{ij}-X_{i\cdot}-X_{\cdot j}+\overline{X}\right)^2 \tag{11}$$

对总离差平方和 S_{T} 进行分解, S_{T} 分解为 3 部分: S_{A}、S_{B} 和 S_{E}(见式(12)), 以分别反映因素 A 的组间差异、因素 B 的组间差异和随机误差。

$$S_{\mathrm{T}}=S_{\mathrm{A}}+S_{\mathrm{B}}+S_{\mathrm{E}} \tag{12}$$

利用式 (1)～式 (12) 算出的 r、s、t、S_{A}、S_{B}、$S_{\mathrm{A\times B}}$ 和 S_{E} 来计算假设:

$$\begin{cases}H_{01}:\mu_{i\cdot}-\mu=0 & i=1,2,\cdots,r\\ H_{02}:\mu_{\cdot j}-\mu=0 & j=1,2,\cdots,s\\ H_{03}:\mu_{ij}-\mu_{i\cdot}-\mu_{\cdot j}+\mu=0 & \\ & i=1,2,\cdots r, j=1,2,\cdots,s\end{cases}$$

其中:

$$\mu=\frac{1}{rs}\sum_{i=1}^{r}\sum_{j=1}^{s}\mu_{ij} \tag{13}$$

$$\mu_{i\cdot}=\frac{1}{s}\sum_{j=1}^{s}\mu_{ij} \tag{14}$$

$$\mu_{\cdot j}=\frac{1}{r}\sum_{i=1}^{r}\mu_{ij} \tag{15}$$

接受 H_{01} 意味着因素 A 对试验没有产生显著影响; 接受 H_{02} 意味着因素 B 对试验没有产生显著影响; 接受 H_{03} 意味因素 A 和因素 B 对试验结果没有交互影响。

通过将上述结果纳入方差分析表, 从而得出如表 4 所示的结果。

本节在双因素方差分析表中引入效应量 η^2, 效应量一种度量效应大小的指标, 表示变量之间的紧密或差异程度。效应量不受样本数量的影响并且与测量单位无关[16]。一般对于方差分析中的效应量, 当效应量大于 0.059 时, 则可以说明具有较大效应。当统计学意义显著, 效应量较大时, 认为分析

表 4　双因素方差分析表

Table 4　Two-factor analysis of variance

方差来源	平方和	自由度	均方	方差	效应量
因素A	S_{A}	$r-1$	$\overline{S}_{\mathrm{A}}=\frac{S_{\mathrm{A}}}{r-1}$	$F_{\mathrm{A}}=\frac{\overline{S}_{\mathrm{A}}}{S_{\mathrm{E}}}$	$\eta_A^2=\frac{S_{\mathrm{A}}}{S_{\mathrm{T}}-S_{\mathrm{B}}-S_{\mathrm{A\times B}}}$
因素B	S_{B}	$s-1$	$\overline{S}_{\mathrm{B}}=\frac{S_{\mathrm{B}}}{s-1}$	$F_{B}=\frac{\overline{S}_{\mathrm{B}}}{S_{\mathrm{E}}}$	$\eta_B^2=\frac{S_{\mathrm{B}}}{S_{\mathrm{T}}-S_{\mathrm{A}}-S_{\mathrm{A\times B}}}$
交互作用A×B	$S_{\mathrm{A\times B}}$	$(r-1)(s-1)$	$\overline{S}_{\mathrm{A\times B}}=\frac{S_{\mathrm{A\times B}}}{(r-1)(s-1)}$	$F_{\mathrm{A\times B}}=\frac{\overline{S}_{\mathrm{A\times B}}}{S_{\mathrm{E}}}$	$\eta_{\mathrm{A\times B}}^2=\frac{S_{\mathrm{A\times B}}}{S_{\mathrm{T}}-S_{\mathrm{A}}-S_{\mathrm{B}}}$
误差	S_{E}	$rs(t-1)$	$\overline{S}_{\mathrm{E}}=\frac{S_{\mathrm{E}}}{rs(t-1)}$		
总和	S_{T}	$rst-1$			

结果具有较大的差异，统计学意义不显著，效应量较小时，可以认为分析结果没有差异。在统计分析结果中报告效应量越来越受到国内外专家学者的推崇[17]。

利用统计学软件SPSS27.0软件，运用双因素重复测量方差分析的方法对飞行方向、中转停留时间与安全性能指标进行方差分析。模型的结构如下：受试者ID(随机因子)；飞行方向(固定因子)和停留时间(固定因子)。表5为方差分析的结果，其中，F为方差。

表5 方差分析结果

Table 5 Results from analysis of variance

安全性能指标	F			p			η^2		
	飞行方向	停留时间	飞行方向与停留时间	飞行方向	停留时间	飞行方向与停留时间	飞行方向	停留时间	飞行方向与停留时间
KSS	19.656	7.221	0.872	0.000	0.000	0.403	0.169	0.126	0.012
SP	2.640	5.794	0.714	0.015	0.001	0.545	0.067	0.092	0.008
fastest 10% of responses	11.302	2.751	0.522	0.001	0.043	0.667	0.141	0.063	0.006
slowest 10% of responses	7.602	6.798	0.487	0.006	0.003	0.692	0.118	0.109	0.005
TST in 24 h	6.224	2.524	0.679	0.013	0.045	0.613	0.105	0.062	0.007

3 结果与讨论

3.1 相关性分析结果

由表3的结果表明，fastest 10% of responses与KSS的相关性显著(R=0.262, p =0.013)，fastest 10% of responses与SP的相关性显著(R=0.253, p=0.016)；slowest 10% of responses与KSS的相关性显著(R=0.227, p=0.032)，slowest 10% of responses与SP的相关性显著(R=0.240, p=0.023)；TST in 24 h与KSS之间呈现出负相关(R=−0.240, p=0.023)，TST in 24 h与SP之间呈现出负相关(R=−0.326, p=0.002)。相关性系数图如图2所示。

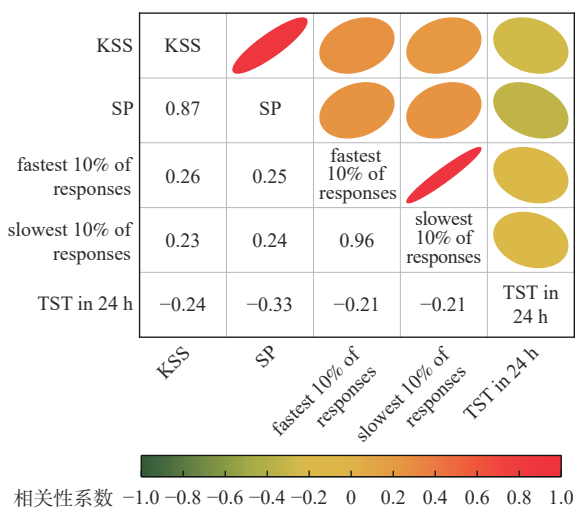


图2 安全性能指标之间的相关系数

Fig. 2 Correlation coefficient of safety performance indicators

由图2可知，主观自评指标与客观测试指标之间呈现出一定的相关性，即当所获得的睡眠时间更短时，KSS和SP主观评估得分更高，相应的PVT客观测试的指标fastest 10% of responses与slowest 10% of responses更长，警觉性水平更低。

3.2 差异性分析结果

3.2.1 主观自评差异性分析

由表5可知，不同飞行方向下的KSS评分统计学意义显著(F=19.656, p=0.000)，且偏η^2=0.169，具有显著差异。不同停留时间下的KSS评分统计学意义显著(F=7.221, p=0.000)，且偏η^2=0.126，具有显著差异。详细的信息利用Origin软件绘制成了如图3所示的半小提琴图。半小提琴图融合了小提琴图、蜜蜂群图和箱线图的优点，不仅可以呈现出数据的概率密度和数据的整体分布情况，还可以反映出数据点的个数及其分布的位置。

由图3可知，与其他时间段相比，飞行员在停留4～5 h后执行任务的KSS主观自评得分明显更低，警觉性水平更高。与此同时，向东飞行的比向西飞行飞行员的KSS主观自评得分更高，更容易感到困倦。

通过表5的方差分析可知，不同飞行方向下的SP评分统计学意义显著(F=2.640, p=0.015)，且偏η^2=0.067，具有显著差异。不同停留时间下的SP评分

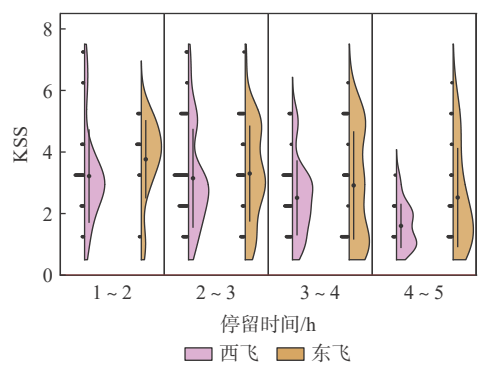


图3 不同停留时间的KSS自评评分

Fig. 3 KSS ratings under different layover lengths

统计学意义显著 (F=5.794, p=0.001), 且偏 η^2=0.092, 具有显著差异。

由图 4 可知, 飞行员在停留 1～2 h 后执行任务, 其 SP 主观自评得分最高, 停留时间越长, SP 的评分也越来越低。向东飞行的飞行员比向西飞行的 SP 主观自评得分更高, 感觉到更疲劳。

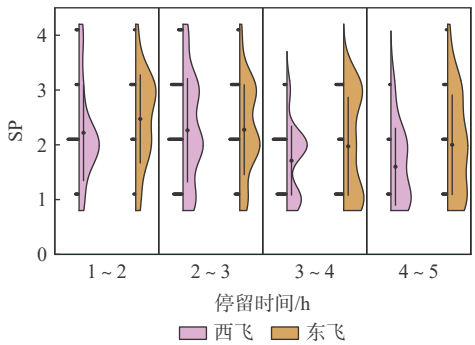


图 4　不同停留时间的 SP 自评评分

Fig. 4　SP ratings under different layover lengths

3.2.2　客观测试差异性分析

根据表 5 的方差分析可知, 不同飞行方向下的 fastest 10% of responses 统计学意义显著 (F=11.302, p=0.001), 且偏 η^2=0.141, 具有显著差异。不同停留时间下的 fastest 10% of responses 的统计学意义显著 (F=2.751, p=0.043), 且偏 η^2=0.063, 具有显著差异。

由图 5 可知, 飞行员在停留 1～2 h 后执行任务的 fastest 10% of responses 要明显长于其他时间段。与向西飞行的飞行员相比, 向东飞行的飞行员的 fastest 10% of responses 更长, 警觉水平更低。

根据表 5 的方差分析可知, 不同飞行方向下的 slowest10% of responses 统计学意义显著 (F=7.602, p=0.006), 且偏 η^2=0.118, 具有显著差异。不同停留时间下的 slowest10% of responses 统计学意义显著 (F=6.798, p=0.003), 且偏 η^2=0.109, 具有显著差异。

由图 6 可知, 飞行员在停留 1～2 h 后执行任务的 slowest 10% of responses 要显著长于其他时间段。与西飞飞行员相比, 向东飞行的飞行员的 slowest 10% of responses 更长, 注意水平更低。

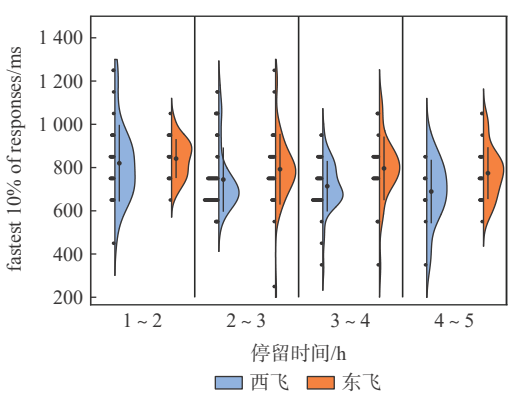


图 5　不同停留时间的 PVT fastest 10% of responses

Fig. 5　Fastest 10% of responses of PVT under different layover lengths

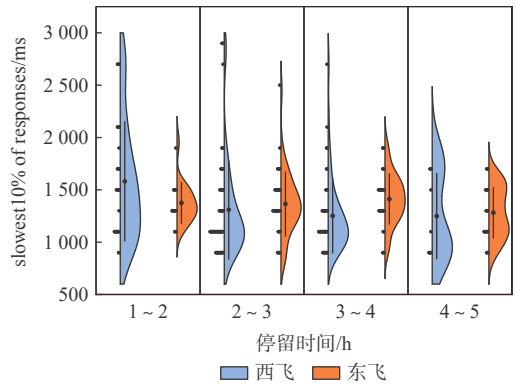


图 6　不同停留时间的 PVT slowest 10% of responses

Fig. 6　Slowest 10% of responses of PVT under different layover lengths

3.2.3　睡眠数据差异性分析

根据表 5 的方差分析可知, 不同飞行方向的 TST in 24 h 的统计学意义显著 (F=6.224, p=0.013), 且偏 η^2=0.105, 具有显著差异。不同停留时间与 TST in 24 h 的统计学意义显著 (F=2.524, p=0.045), 且偏 η^2=0.062, 具有显著差异。

如图 7 所示, 飞行员在停留 1～2 h 后执行任务的 TST in 24 h 时间最少, 随着停留时间的增加, 飞行员所获得的 TST in 24 h 也越来越多。飞行员向东飞行时的 TST in 24 h 时间要少于向西飞行的。

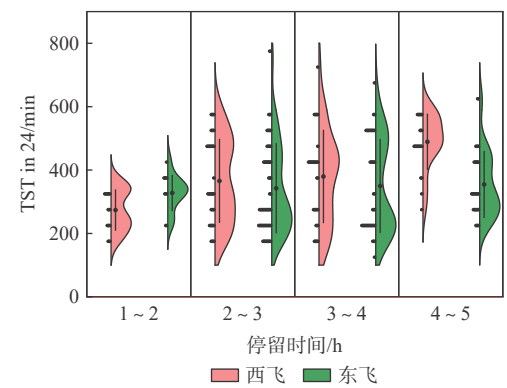


图 7　不同停留时间的 TST in 24 h

Fig. 7　TST in 24 h under different layover lengths

由以上的分析可知, 飞行员在过去 24 h 内睡眠时间较短和清醒时间较长会导致主观困倦和疲劳评分较低, PVT 表现更好。这与 Kosmadopoulos 等[18]在实验室研究得出的观点相一致。而中转停留时间和飞行方向会影响飞行员生理节律的调节及对环境的适应能力, 这反过来又进一步影响了机组人员能够获得的入睡时间, 以及他们的主观困倦、疲劳和起飞前的 PVT 表现。将这 3 类指标结合起来,

在国际豁免航班实际运行环境中监测飞行员的警觉性状态是可行的，符合ICAO给的参考建议[19]。

3.3 讨 论

结果表明，不同的飞行方向和中转停留时间的长短对飞行员的警觉性水平具有显著影响。与西飞的国际豁免航班飞行员相比，向东飞行的飞行员警觉性水平更低。这种现象可能是因为向西飞行时，飞行员的昼夜节律通常适应的更快，因而飞行员在中转停留期间可能会获得更长的睡眠时间。这与学者Signal等[20]的研究相吻合，不同的飞行方向下飞行员适应生理节律的情况也不一致。对于西飞航班来说，每次时区改变一次的时间仅约为1 d，而与之相比，东行航班时区每次改变所需的恢复时间大约为1.5 d。

当停留时间较短时，飞行员主观警觉自评、客观测试和睡眠状态都表现不佳；而当停留时间较长时，飞行员的警觉性水平明显提高。上述警觉性的差异可以解释为，当中转停留时间较短时，飞行员可能还没从前序飞行任务中积累的睡眠债和疲劳中恢复过来，因中转过站的必要程序（例如进出海关）及回程前的飞行准备等任务，持续的觉醒时长和工作负荷给飞行员的警觉性带来了不利影响。此外，在当前疫情新常态的情况下，各个国家和地区的防疫政策不尽相同、中转过站的安全保障措施并非运行良好、飞行机组担心感染疫情风险的忧虑等因素，致使在中转停留结束后飞行员执行任务时的警觉性水平更低，更容易感到疲劳。

4 结 论

1）中转停留时间对于国际豁免航班飞行员警觉性水平的影响是显著的。停留时间为1～2 h时，飞行员的警觉性水平明显较低。该结论可以帮助航空公司在防疫豁免期间优化排班时刻表，适当地延长飞行员中转停留的时间，并且尽可能地保证飞行员在中转停留期间的睡眠质量，以期使飞行员在中转停留后执行任务时维持较高的警觉性状态。

2）向东飞行的国际豁免航班飞行员要比向西飞行的警觉性水平更低。因此，航空公司可以根据机组人员的飞行方向而提前制定休息策略。例如当飞行员向东飞行时，应该让其提前睡觉休息，尽快调整生理节律，提升飞行员的警觉性水平，保障航班的安全运行。

本文参与者的数量因受到时间和成本限制尚需增加，其次飞行员警觉性状态与规定飞机操纵参数偏差等数据之间的联系目前尚不清楚。从科学的角度来看，对于实际航班运行过程中的警觉性研究以及疲劳风险管理系统的建设，需要对更大、更多样的数据集进行分析，朝着标准化数据库结构的方向继续优化。

参考文献（References）

[1] ROACH G D, PETRILLI R M A, DAWSON D, et al. Impact of layover length on sleep, subjective fatigue levels, and sustained attention of long-haul airline pilots[J]. Chronobiology International, 2012, 29(5): 580-586.

[2] 中国民用航空局. 疫情期间豁免机组成员值勤期、飞行时间限制的实施办法[EB/OL]. (2020-12-02) [2022-09-14]. http://www.caac.gov.cn/XXGK/XXGK/ZCFBJD/202012/t20201223_205708.html.
Civil Aviation Administration of China. Implementation measures for exempting crew members from duty period and flight time limitations during the epidemic period.[EB/OL]. (2020-12-02) [2022-09-14].http://www.caac.gov.cn/XXGK/XXGK/ZCFBJD/202012/t20201223_205708.html.

[3] REIS C, MESTRE C, CANHÃO H. Prevalence of fatigue in a group of airline pilots[J]. Aviation, Space, and Environmental Medicine, 2013, 84(8): 828-833.

[4] VAN DRONGELEN A, BOOT C R L, HLOBIL H, et al. Risk factors for fatigue among airline pilots[J]. International Archives of Occupational and Environmental Health, 2017, 90(1): 39-47.

[5] 于洮, 刘德宝, 孙梦婷, 等. 陆军航空兵飞行员睡眠状况与疲劳程度的相关性分析[J]. 华南国防医学杂志, 2018, 32(9): 639-641.
YU T, LIU D B, SUN M T, et al. Correlation analysis between sleep condition and fatigue degree of army aviation pilots[J]. Military Medical Journal of South China, 2018, 32(9): 639-641(in Chinese).

[6] 曾绍洁, 李红霞, 刘微卫, 等. 全封闭式管理对飞行员睡眠质量及疲劳状态的影响[J]. 空军医学杂志, 2022, 38(2): 102-106.
ZENG S J, LI H X, LIU W W, et al. Effects of fully enclosed restrictive management on sleep quality and fatigue of pilots[J]. Medical Journal of Air Force, 2022, 38(2): 102-106 (in Chinese).

[7] 刘艳, 程宝权, 于鑫, 等. 直升机飞行员脑力负荷、睡眠和疲劳的典型相关分析[J]. 陆军军医大学学报, 2022, 44(11): 1087-1093.
LIU Y, CHENG B Q, YU X, et al. Mental workload, sleep and fatigue in helicopter pilots: A canonical correlation analysis[J]. Journal of Army Medical University, 2022, 44(11): 1087-1093 (in Chinese).

[8] 汪磊, 郑竹寒, 张之洋. 基于警觉性的疲劳程度自测系统研究[J]. 航天医学与医学工程, 2021, 34(5): 361-366.
WANG L, ZHENG Z H, ZHANG Z Y. Research on self-test system of fatigue degree based on vigilance[J]. Space Medicine & Medical Engineering, 2021, 34(5): 361-366 (in Chinese).

[9] PAN T, WANG H B, SI H Q, et al. Identification of pilots' fatigue status based on electrocardiogram signals[J]. Sensors, 2021, 21(9): 3003.

[10] HONN K A, SATTERFIELD B C, MCCAULEY P, et al. Fatiguing effect of multiple take-offs and landings in regional airline operations[J]. Accident; Analysis and Prevention, 2016, 86: 199-208.

[11] GASPERETTI C E, DONG L, HARVEY A G. The effect of the transdiagnostic sleep and circadian intervention (TranS-C) on actigraphic estimates of sleep and rest-activity rhythms in adolescents with an evening circadian preference[J]. Sleep Health, 2022, 8(2):

191-194.

[12] International Civil Aviation Organization. Fatigue risk man a gement systems manual for regulators[EB/OL]. (2011-07-24) [2022-09-16]. https://www.icao.int/safety/fatiguemanagement/FRMS2011/Documents/Reference%20Documents/Doc%209966%20-%20FRMS%20Manual%20for%20Regulators.pdf#:~:text=Chapter%201.%20Introduction%20to%20fatigue%20risk%20management%20systems%20(FRMS)%201.1.

[13] GOFFENG E M, WAGSTAFF A, NORDBY K C, et al. Risk of fatigue among airline crew during 4 consecutive days of flight duty[J]. Aerospace Medicine and Human Performance, 2019, 90(5): 466-474.

[14] 张英, 陈晨, 刘振华, 等. 精神运动警觉性任务在不同知觉负荷下的疲劳状态监测[J]. 心理学探新, 2018, 38(4): 377-384.
ZHANG Y, CHEN C, LIU Z H, et al. Psychomotor vigilance monitors different fatigue states[J]. Psychological Exploration, 2018, 38(4): 377-384 (in Chinese).

[15] GANDER P H, MULRINE H M, VAN DEN BERG M J, et al. Effects of sleep/wake history and circadian phase on proposed pilot fatigue safety performance indicators[J]. Journal of Sleep Research, 2015, 24(1): 110-119.

[16] 温忠麟, 范息涛, 叶宝娟, 等. 从效应量应有的性质看中介效应量的合理性[J]. 心理学报, 2016, 48(4): 435-443.
WEN Z L, FAN X T, YE B J, et al. Characteristics of an effect size and appropriateness of mediation effect size measures revisited[J]. Acta Psychologica Sinica, 2016, 48(4): 435-443 (in Chinese).

[17] 沈光辉, 范涌峰, 陈婷. 教育研究中的 P 值使用: 问题及对策: 兼谈效应量的使用[J]. 数学教育学报, 2019, 28(4): 92-98.
SHEN G H, FAN Y F, CHEN T. Using P value in educational research: problems and countermeasures—concurrently discussing about the use of effect quantity[J]. Journal of Mathematics Education, 2019, 28(4): 92-98 (in Chinese).

[18] KOSMADOPOULOS A, SARGENT C, ZHOU X, et al. The efficacy of objective and subjective predictors of driving performance during sleep restriction and circadian misalignment[J]. Accident; Analysis and Prevention, 2017, 99: 445-451.

[19] GANDER P H, SIGNAL T L, VAN DEN BERG M J, et al. In-flight sleep, pilot fatigue and psychomotor vigilance task performance on ultra-long range versus long range flights[J]. Journal of Sleep Research, 2013, 22(6): 697-706.

[20] SIGNAL T L, MULRINE H M, VAN DEN BERG M J, et al. Mitigating and monitoring flight crew fatigue on a westward ultra-long-range flight[J]. Aviation, Space, and Environmental Medicine, 2014, 85(12): 1199-1208.

Influence of layover length and flight direction on pilots' alertness

LI Jingqiang*, HU Chao, ZHANG Xining, ZHANG Lu, LIU Annan

(College of Safety Science and Engineering, Civil Aviation University of China, Tianjin 300300, China)

Abstract: Survey based on the fatigue risk management system has been conducted at an unprecedented speed in the global aviation industry. During the quarantine exemption period, data such as subjective ratings, objective tests, and sleeping information were collected from 64 pilots of international flights exempted from quarantine under different layover lengths in actual operation mode. The SPSS27.0 software, correlation analysis, and analysis of variance of repeated measures were used to deeply analyze the influence of different layover lengths and flight directions on the alertness of pilots of international flights exempted from quarantine. Correlation coefficient diagrams and half violin plots were painted by Origin software. The results show that when the layover length is 1–2 h, the alertness level of pilots of international flights exempted from quarantine is significantly lower. The alertness level of pilots of eastward flights is lower than that of westward flights. These findings can help airlines monitor and analyze the alertness of pilots of international flights exempted from quarantine during the quarantine exemption period and provide references for strengthening the protection measures for layover flights.

Keywords: risk management; international flights exempted from quarantine; alertness; pilot; layover length; flight direction

Received: 2022-09-14; **Accepted**: 2022-12-04; **Published Online**: 2022-12-27 17:31

URL: link.cnki.net/urlid/11.2625.V.20221226.1511.003

Foundation items: The Civil Aviation Administration Joint Funds of the National Natural Science Foundation of China (U1933122); The Fundamental Research Funds for the Central Universities (3122019186)

* **Corresponding author**. E-mail: jqli@cauc.edu.cn

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2022.0785

星基增强系统单频服务性能评估方法

金彪[1]，李锐[2]，王盾[1,*]，刘磊[3]，原晋栩[1]，李子潇[1]

(1. 航天恒星科技有限公司，北京 100095；2. 北京航空航天大学 电子信息工程学院，北京 100191；
3. 空装驻北京地区第七军事代表室，北京 100076)

摘 要：针对星基增强系统(SBAS)用户差分距离误差(UDRE)完好性参数评估时双频无电离层组合放大观测噪声的问题，提出伪距域完好性评估方法，该方法反映了SBAS电离层改正数和星历时钟改正数的综合影响。利用240天共100个国际全球导航卫星系统(GNSS)服务组织(IGS)测站的数据对美国广域增强系统(WAAS)和欧洲地球静止轨道导航重叠服务(EGNOS)的性能进行评估。结果表明：WAAS和EGNOS格网电离层垂直误差(GIVE)参数的完好性分别为99.96%和99.91%，电离层延迟改正精度达到0.56 m。WAAS和EGNOS伪距域完好性优于99.999%，经改正后的GPS用户测距误差(URE)精度达到0.34 m。WAAS 95%的水平和垂直定位误差为1.19 m和1.72 m，EGNOS的则为1.18 m和1.68 m。评估时段内均未发生完好性风险事件，验证了测地型接收机用于SBAS服务性能评估的可行性。WAAS在一类垂直引导进近(APV-I)阶段的定位精度平均可用性为99.9%，二类垂直引导进近(APV-II)阶段在服务中心区域的定位精度可用性为80%～95%；EGNOS APV-I阶段平均定位精度可用性为99.9%，APV-II阶段平均定位精度可用性为94.2%，可用性较高，主要是由于EGNOS评估测站多处于服务区中心。SBAS定位精度可用性在服务区边缘部分下降较明显。

关 键 词：星基增强系统；格网电离层垂直误差；伪距域；完好性；可用性
中图分类号：P228
文献标志码：A **文章编号**：1001-5965(2024)10-3062-12

为满足民航用户对全球导航卫星系统(global navigation satellite system, GNSS)导航定位精度、完好性、连续性及可用性的需求，世界各主要国家相继建立了星基增强系统(satellite based augmentation system, SBAS)，包括美国广域增强系统(wide area augmentation system, WAAS)[1]、欧洲地球静止轨道导航重叠服务(European geostationary navigation overlay service, EGNOS)[2]、印度GPS辅助地球静止轨道卫星增强导航(GPS aided geostationary Earth orbit augmented navigation, GAGAN)[3]和日本多功能卫星增强系统(multi-functional satellite augmentation system, MSAS)[4]等。中国的北斗星基导航增强系统(BeiDou satellite based augmentation system, BDSBAS)[5-6]、俄罗斯的差分改正和监测系统(system for differential corrections and monitoring, SDCM)、澳大利亚新西兰SBAS(Australia-New Zealand SBAS, AUS-NZ SBAS)[7]及韩国增强卫星系统(Korea augmentation satellite system, KASS)[8]也在建设或测试之中。SBAS利用区域监测站数据解算导航卫星星历和电离层改正数及完好性参数[9-11]。用户终端利用SBAS信息进行定位解算并生成保护级以判断系统的可用性。当前，SBAS的

收稿日期：2022-09-15；录用日期：2022-11-04；网络出版时间：2023-02-01 14:55
网络出版地址：link.cnki.net/urlid/11.2625.V.20230131.1834.002
基金项目：国家自然科学基金(41974041)
*通信作者. E-mail：dawun04@163.com

引用格式：金彪，李锐，王盾，等. 星基增强系统单频服务性能评估方法[J]. 北京航空航天大学学报，2024，50(10)：3062-3073.
JIN B，LI R，WANG D，et al. Service performance assessment method of single frequency SBAS[J]. Journal of Beijing University of Aeronautics and Astronautics，2024，50(10)：3062-3073 (in Chinese).

研究主要集中在改正数解算[6,12-13]、用户差分距离误差(user differential range error, UDRE)和双频距离误差(dual frequency range error, DFRE)参数生成[10-11,14-15]、电离层完好性建模[9,16-18]等方面。

在 SBAS 服务性能评估方面, Walter 等[19]于 1994 年先期开展 WAAS 飞行试验, 利用特高频(ultra high frequency, UHF)电台播发增强信息, 机载接收机实时定位解算, 结果表明 95% 的垂直定位误差优于 3.0 m, 极大提升了 GPS 的定位精度。Perrin 等[20]在瑞士阿尔卑斯山卢加诺机场对 EGNOS 进行了飞行测试, 验证其在复杂山区环境下的服务性能, 结果表明, EGNOS 水平和垂直定位精度达到亚米级, 一类垂直引导进近(approach with vertical guidance I, APV-I)和二类垂直引导进近(approach with vertical guidance II, APV-II)阶段的定位精度可用性均为 100%, 未发生完好性风险事件。Specht 等[21]研究了 EGNOS 在海事应用的定位精度, 结果显示, 靠近陆地区域, 其定位精度达到 0.5～2.0 m。Heβelbarth 和 Wanninger[22]对 SBAS 轨道和时钟改正数用于精密单点定位(precise point positioning, PPP)的性能进行了评估, 结果表明, WAAS 改正数定位精度较高, 水平和垂直定位精度优于 30 cm。楼益栋等[23]利用中国东部区域观测数据对准天顶卫星系统(quasi-zenith satellite system, QZSS) L1 频段亚米级完好性增强(submeter class augmentation with integrity function, SAIF)信号的性能进行评估, 发现其用户测距误差(user range error, URE)为 0.6 m, 可实现与 WAAS 相当的定位精度。Hirschberger 等[24]对不同飞行状态下的 WAAS 电文丢包率进行评估, 并指出为满足完好性要求, 电文丢包率不应超过 10^{-3}。陈姗姗等[25]对当前已通过民航认证并正式提供服务的 SBAS 播发的电文进行了分析, 包括电文类型、更新周期、电文内容等, 其中, WAAS 播发的卫星数和电离层格网点(ionospheric grid point, IGP)数量最多、服务区域最大。常志巧等[26]对 WAAS 电离层改正数的性能进行了评估。金彪等[27]研究了 SBAS 地球同步轨道(geostationary Earth orbit, GEO)卫星的测距性能及其对定位结果的影响, 结果表明, 在 SBAS 定位中引入 GEO 卫星测距值可提升系统的可用性。蔡洪亮等[28]利用国内 3 个测站的数据对 BDSBAS 的服务性能进行评估, 结果显示, 其水平定位精度优于 1.5 m, 高程定位精度优于 3.0 m(95%)。

当前对 SBAS 服务性能的研究多基于国际 GNSS 服务组织(international GNSS service, IGS)、国际 GNSS 监测评估系统(international GNSS monitoring and assessment system, iGMAS)等机构测站短时间的观测数据, 缺乏足够的样本对服务性能尤其是完好性进行评估。同时国际民航组织(International Civil Aviation Organization, ICAO)标准与建议措施(standards and recommended practices, SARPs)等标准对 SBAS 接收机的设计提出了明确要求, 包括前端带宽、带内群延迟稳定性、鉴相器类型及鉴相间隔等[29-30]。测地型接收机用于 SBAS 服务性能评估的可行性需要进一步验证。针对上述问题, 本文利用 100 个 IGS 测站 240 天的实测数据对 WAAS 和 EGNOS 的定位精度、完好性和可用性进行评估, 分析测地型接收机用于 SBAS 评估的可行性。本文可用性指定位精度的可用性。针对 UDRE 完好性参数评估时双频无电离层组合放大观测噪声的问题, 本文提出伪距域完好性评估方法, 为后续 BDSBAS 民航应用验证提供了参考。

1　数学模型和数据源

1.1　电离层完好性

GPS L1 伪距电离层延迟 I_1 计算式为[9,31]

$$\begin{cases} I_1 = \dfrac{P_1 - P_2 + \tau_{\mathrm{r}} - \tau^{\mathrm{s}}}{1-\gamma} \\ \gamma = \dfrac{f_1^2}{f_2^2} \end{cases} \tag{1}$$

式中: P_1 和 P_2 为 GPS 载波 L1 和 L2 上的伪距观测值; τ_{r} 和 τ^{s} 分别为接收机和卫星的硬件延迟; γ 为比例系数; f_1 和 f_2 分别为 L1 和 L2 载波的频率。卫星端硬件延迟采用欧洲定轨中心(Center for Orbit Determination in Europe, CODE)发布的差分伪距偏差(differential code bias, DCB)产品进行改正。利用 IGS 全球测站和本文选用的测站数据进行全球电离层延迟建模, 同时估计接收机硬件延迟[32], 考虑到接收机硬件延迟的稳定性, 每周解算一次。

SBAS 播发格网点上电离层垂直延迟, 用户电离层穿刺点处倾斜延迟及其方差分别为

$$\begin{cases} I_{\mathrm{s}} = F_{\mathrm{p}} \sum_{i=1}^{n} W_i I_{\mathrm{v},i} \\ \sigma^2_{\mathrm{UIRE}} = F_{\mathrm{p}}^2 \sum_{i=1}^{n} W_i \sigma^2_{\mathrm{GIVE},i} \end{cases} \tag{2}$$

式中: I_{s} 为穿刺点处电离层斜延迟; n 为穿刺点处可用的电离层格网点数量; $I_{\mathrm{v},i}$ 为 SBAS 播发的格网点 i 的电离层垂直延迟; W_i 为加权因子; F_{p} 为映射函数, 用于将电离层垂直延迟转换为斜延迟。加权因子和映射函数的计算方法可参考文献 [29]。σ^2_{UIRE} 和 $\sigma^2_{\mathrm{GIVE},i}$ 分别为用户穿刺点和格网点处电离层延迟方差。将观测数据计算的电离层延迟 I_1 和利用 SBAS 改正数计算的电离层延迟 I_{s} 作差可得到电离

层残差，当残差超过$3.29\sigma_{\mathrm{UIRE}}$时认为发生电离层完好性风险事件。

1.2 伪距域完好性

UDRE参数用于描述星历和时钟改正数的精度。由于轨道和钟差的相关性，利用IGS精密轨道和钟差产品对SBAS星历改正数进行评估会存在系统性偏差[11]。另外，利用观测值残差评估改正数精度时，通常采用双频无电离层组合观测值，放大了观测噪声，同时观测值残差还受到对流层延迟残余误差的影响。为此，本文提出伪距域完好性方法，用于综合评估SBAS UDRE和电离层垂直误差(grid ionospheric vertical error, GIVE)参数的完好性。经SBAS改正后的GPS伪距残差为

$$\begin{aligned}\Delta\rho_{\mathrm{r}}^{i}=&P_{\mathrm{r}}^{i}-c\Delta t_{\mathrm{r}}+c\Delta t^{i}+\Delta a_{0,i}+\\&\Delta a_{1,i}(t-t_{\mathrm{LT},i})+\Delta r_i+\Delta\dot{r}_i(t-t_{\mathrm{FC},i})-I_{\mathrm{s}}-d_{\mathrm{tro},i}-\rho_{\mathrm{r}}^{i}\end{aligned}\tag{3}$$

式中：P_{r}^{i}为GPS L1 C/A码伪距观测值，r表示接收机，i为卫星；c为光速；Δt_{r}和Δt^{i}分别为接收机和卫星的钟差；$\Delta a_{0,i}$为SBAS播发的卫星i的慢变钟差改正数；$\Delta a_{1,i}$为钟漂改正数；t为当前时刻；$t_{\mathrm{LT},i}$为慢变改正数参考时刻；Δr_i为快变改正数；$\Delta\dot{r}_i$为快变改正数变化率；$t_{\mathrm{FC},i}$为快变改正数的参考时刻；$d_{\mathrm{tro},i}$为对流层延迟，利用文献[29]中的模型进行改正；ρ_{r}^{i}为卫星和测站间的几何距离。式(3)中得到的观测值残差受星历和时钟改正误差、电离层和对流层残余误差及测量误差的影响，其方差为

$$\sigma_i^2=\sigma_{\mathrm{flt},i}^2+\sigma_{\mathrm{UIRE},i}^2+\sigma_{\mathrm{air},i}^2+\sigma_{\mathrm{tro},i}^2\tag{4}$$

其中：σ_i^2为观测值误差方差；$\sigma_{\mathrm{flt},i}^2$为卫星星历和时钟改正数方差，与UDRE、电文类型27(message type27, MT27)或MT28及用户位置相关；$\sigma_{\mathrm{air},i}^2$为接收机测量误差的方差；$\sigma_{\mathrm{tro},i}^2$为对流延迟残余误差方差，各参数的具体计算可参考文献[29]。伪距域完好性评估时将伪距残差$\Delta\rho_{\mathrm{r}}^{i}$与门限值$3.29\sigma_i$相比，当残差小于门限时，则认为该伪距观测值完好。

计算几何距离ρ_{r}^{i}时，卫星位置为

$$\begin{bmatrix}x_{\mathrm{cor},i}\\y_{\mathrm{cor},i}\\z_{\mathrm{cor},i}\end{bmatrix}=\begin{bmatrix}x_i\\y_i\\z_i\end{bmatrix}+\begin{bmatrix}\Delta x_i\\\Delta y_i\\\Delta z_i\end{bmatrix}+\begin{bmatrix}\Delta\dot{x}_i\\\Delta\dot{y}_i\\\Delta\dot{z}_i\end{bmatrix}(t-t_{\mathrm{LT},i})\tag{5}$$

式中：$x_{\mathrm{cor},i}$、$y_{\mathrm{cor},i}$、$z_{\mathrm{cor},i}$为改正后的卫星位置；x_i、y_i、z_i为利用广播星历计算的卫星位置；Δx_i、Δy_i、Δz_i为SBAS播发的卫星位置慢变改正数；$\Delta\dot{x}_i$、$\Delta\dot{y}_i$、$\Delta\dot{z}_i$为慢变改正数的变化率。式(3)中接收机钟差Δt_{r}为未知数，利用多颗卫星的伪距残差加权计算：

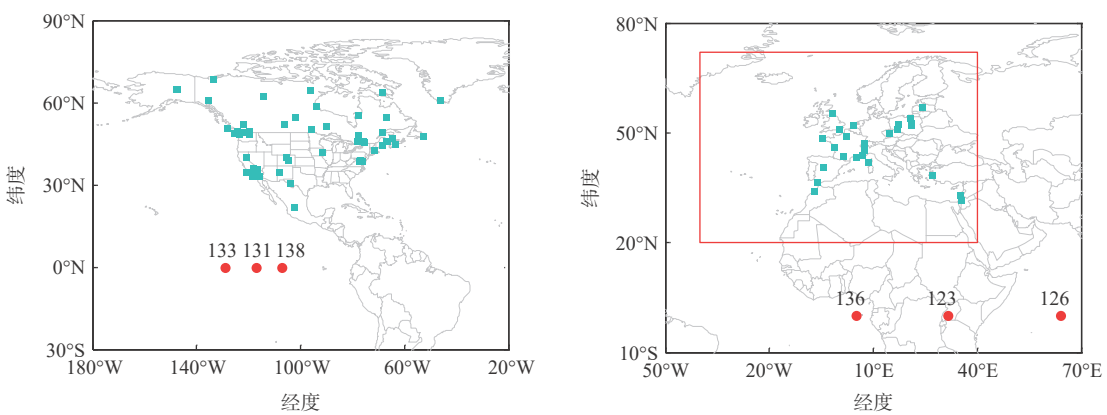
$$\Delta t_{\mathrm{r}}=\sum_{i=1}^{N}\frac{1}{\sigma_i}\Delta\rho_{\mathrm{r}}^{i}\Big/\sum_{i=1}^{N}\frac{1}{\sigma_i}\tag{6}$$

式中：N为使用的卫星数。

1.3 数据源选取

SBAS服务性能评估时需重点考虑：①极小的完好性风险概率对观测数据样本量的要求；②不同地理位置和不同太阳活动情况下电离层的时空变化；③服务区域内不同位置的服务性能存在差异。根据ICAO SARPs的要求，SBAS定位完好性风险概率须小于2×10^{-7}/150 s。假设各测站观测数据相互独立，当采样间隔为30 s、测站数为50时，需174 d的观测数据才可能出现一次完好性风险事件。对于已经通过民航测试认证的SBAS，在验证其性能时可据此确定数据时长。对于尚未经过民航认证的SBAS，则需要更大样本的观测数据对系统性能进行评价。

本文旨在对WAAS和EGNOS的服务性能进行评估，验证本文评估方法的可行性和正确性。因此，共选择100个IGS测站，其中，WAAS评估测站66个，EGNOS评估测站34个，如图1中的正方形点所示。WAAS评估测站对GPS卫星的平均覆盖重数为43重，EGNOS为24重，可较为全面地评估



(a) WAAS卫星和测站　　(b) EGNOS卫星和测站

图1 SBAS GEO卫星位置和选用的IGS测站位置

Fig. 1 Positions of SBAS GEO satellites and selected IGS stations

可视卫星的服务性能。所选测站覆盖了中、高纬度区域, 分布较广, 可对服务区域中心和边缘的服务性能进行评估。考虑到电离层的季节变化特性[33]及 WAAS 和 EGNOS 的最新状态, 选择评估时段为 2021 年 11 月 1 日至 2022 年 6 月 30 日, 共 240 d, 观测数据采样间隔 30 s。所选 IGS 测站均采用测地型接收机。

WAAS 和 EGNOS 当前均使用 3 颗 GEO 卫星提供服务, 其中, WAAS 卫星伪随机噪声(pseudo random noise, PRN)码为 131、133 和 138, 如图 1(a)中圆点所示, EGNOS PRN 码为 123、126 和 136, 如图 1(b)中的圆点所示。根据 EGNOS 生命安全服务定义文件[34], 其利用 MT27 电文播发系统的服务区域, 如图 1(b)中的矩形框所示。WAAS 和 EGNOS 播发的增强电文使用法国国家空间研究中心(Centre National D'Etudes Spatiales, CNES)提供的数据文件。

2　结果分析

2.1　电文统计

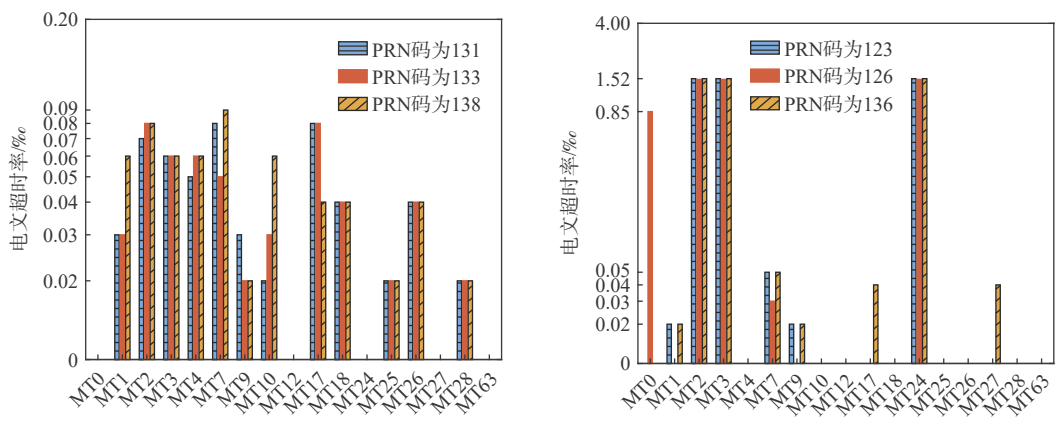
先对 WAAS 和 EGNOS 的电文进行分析, 包括丢包率、超时率、同步头、循环冗余校验(cyclic redundancy check, CRC)和供应商标识正确性。评估时段内 WAAS 的 3 颗卫星及 EGNOS 的 PRN 码为 123 和 136 卫星为服务状态, EGNOS 的 PRN 码为 126 卫星为测试状态, 播发 MT0 电文。WAAS 利用 MT2～MT4 电文播发快变改正数, 而 EGNOS 利用 MT2、MT3 电文播发前 26 颗 GPS 卫星的快变改正数, 利用 MT24 电文播发 GPS 27～GPS 32 卫星的快变和慢变改正数, 以提升电文的使用效率。

WAAS 的 3 颗卫星电文平均丢包率为 4×10^{-4}, EGNOS 平均丢包率为 8×10^{-6}, 均满足小于 10^{-3} 的丢包率要求[24]。图 2 为 WAAS 和 EGNOS 电文的超时率。WAAS PRN 码为 131、133 和 138 的卫星在图中分别以标有横线、无横线及标有斜线的柱状图表示, 同样, EGNOS PRN 码为 123、126 和 136 的卫星也分别以标有横线、无横线及标有斜线的柱状图表示。WAAS 电文超时率小于 9×10^{-5}, 其中 MT2～MT4、MT7 和 MT17 电文超时率相对较高。EGNOS 电文超时率差异较大, MT2、MT3 和 MT24 电文超时率较高, 为 1.52×10^{-3}。在电文丢包和超时情况下, 终端将利用之前播发的仍在有效期内的改正数, 并结合改正数变化率、降效参数进行定位, 会导致定位精度下降和保护级增大, 影响系统的可用性。

图 3 为各 SBAS 电文播发的平均时间间隔。值的注意的是, WAAS MT25 电文播发间隔为 8 s, 极大提升了慢变改正数的播发频率, 降低了电文延迟对定位精度的影响。同时 WAAS 利用 MT63 电文调节电文时序。EGNOS PRN 码为 126 的卫星由于播发 MT0 电文导致其更新周期较长。EGNOS MT25 电文播发周期超过 1 000 s, 表明大部分慢变改正数已通过 MT24 电文播发, 其播发周期为 4 s, 同样提升了慢变改正数的播发频率。统计时段内, WAAS 和 EGNOS 电文同步头、CRC 校验及供应商标识均正常。

2.2　电离层改正性能

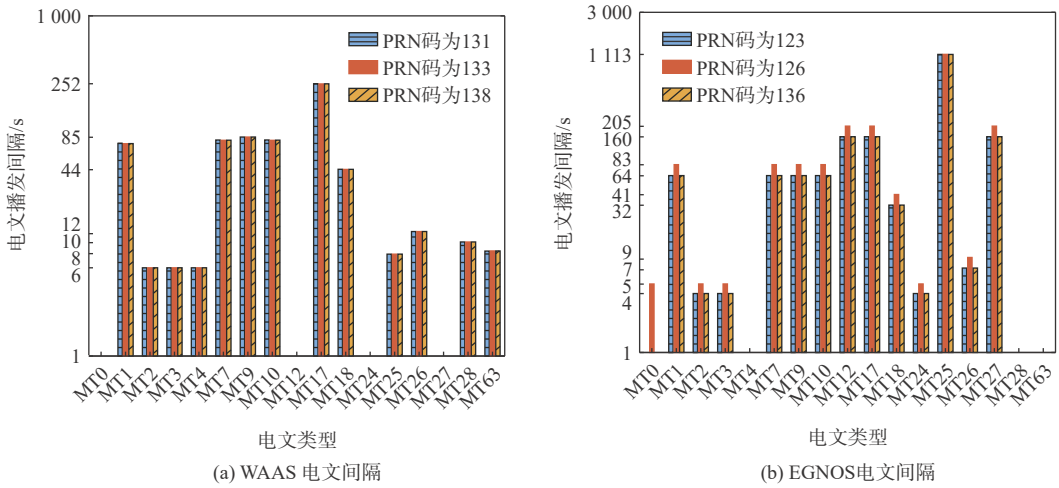
利用第 1.1 节介绍的方法对电离层 GIVE 参数的完好性进行分析, 卫星截止高度角为 5°。分析过程中, WAAS 使用 PRN 码为 133 的卫星电文, EGNOS 使用 PRN 码为 123 的卫星电文。结果如图 4 所示, 图中矩形框表示 EGNOS 的服务区域。WAAS GIVE 参数平均完好性为 99.96%, EGNOS GIVE 参数平均完好性为 99.91%, 均满足 99.9% 的误差包络要求。部分测站 GIVE 完好性低于 99.9%, 如 EGNOS 最东北部测站。



(a) WAAS电文超时率　　(b) EGNOS电文超时率

图 2　SBAS 电文超时率

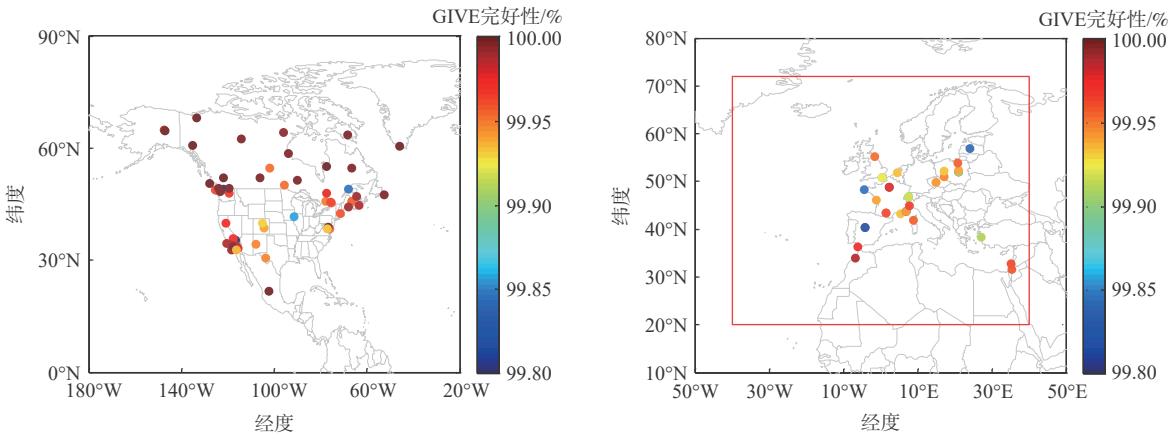
Fig. 2　Timeout rate of SBAS message



(a) WAAS 电文间隔　　(b) EGNOS电文间隔

图 3　SBAS 电文平均播发间隔

Fig. 3　SBAS average message broadcast interval



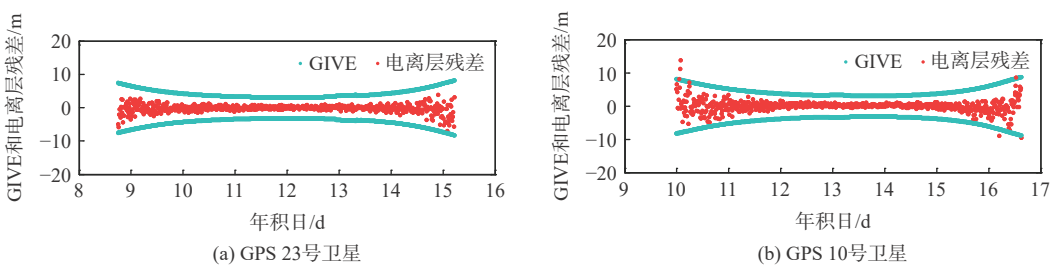
(a) WAAS GIVE完好性　　(b) EGNOS GIVE完好性

图 4　各测站 SBAS GIVE 参数完好性

Fig. 4　SBAS GIVE integrity of each station

为分析电离层完好性风险事件发生的原因，图 5 为 WAAS 服务区内 NLIB 测站 2022 年第 2 天 GPS PRN 码为 23 卫星和 GPS PRN 码为 10 卫星所对应的 GIVE 和电离层残差。图中，GPS PRN 码为 23 的卫星对应的 GIVE 能够完全包络电离层残差，而 GPS PRN 码为 10 的卫星在观测弧段开始和结束处则出现欠包络情况，主要是由于观测开始和结束时伪距观测值噪声较大，使用双频伪距观测值计算的电离层延迟精度较低。GIVE 参数主要用于描述 SBAS 电离层改正数的完好性，其综合考虑了观测数据精度、电离层欠采样等因素[9]。较低的高角度可充分利用观测数据，增加对服务区域上空电离层的采样点。当测站密度较大时，电离层可被充分采样，可提升卫星高度角以减小观测噪声的影响。

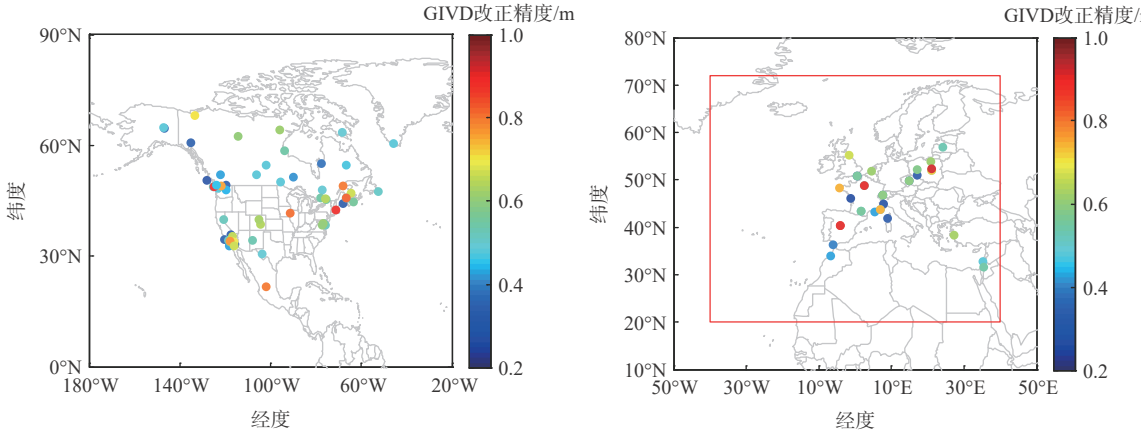
对 WAAS 和 EGNOS 的格网电离层垂直延迟（grid ionospheric vertical delay, GIVD）改正精度进行分析。图 6 为 WAAS 和 EGNOS 电离层延迟改正精度，其中，WAAS 的平均改正精度为 0.53 m，EGNOS 的平均改正精度为 0.56 m，与文献 [35] 中 WAAS 和 EGNOS 的平均改正精度 0.5 m 相当。图中矩形



(a) GPS 23号卫星　　(b) GPS 10号卫星

图 5　NLIB 测站 2022 年第 2 天 GIVE 和电离层残差

Fig. 5　GIVE and ionospheric residuals of NLIB station on day 2 in 2022



(a) WAAS GIVD改正精度　　(b) EGNOS GIVD改正精度

图 6　各测站 SBAS 电离层延迟改正精度

Fig. 6　SBAS ionospheric delay correction accuracy of each station

框表示 EGNOS 的服务区域。

2.3　伪距域性能

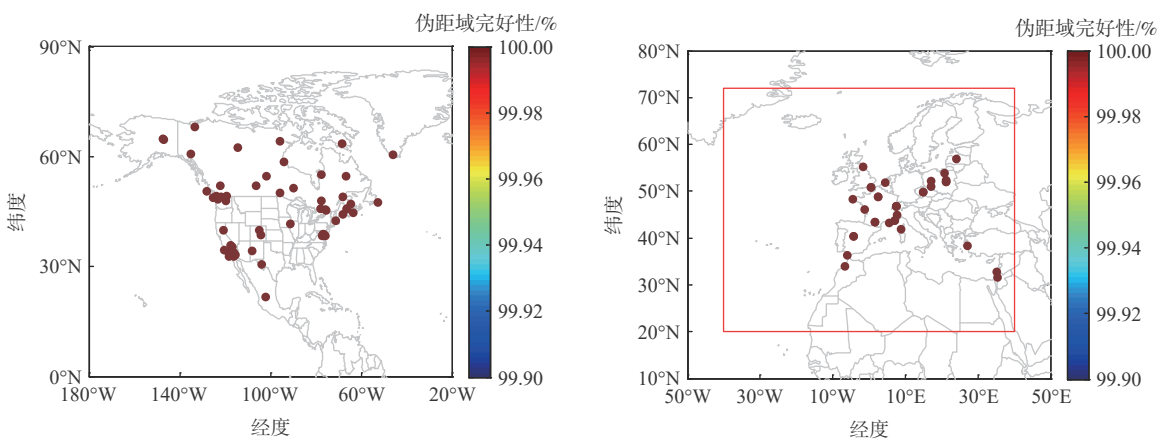
利用第 1.2 节的方法对测站伪距域完好性进行分析, 结果如图 7 所示。图 7(a)为 WAAS 评估测站的伪距域完好性, 平均完好性为 99.999 95%; 图 7(b)为 EGNOS 评估测站的伪距域完好性, 完好性均值为 99.999 7%, 二者均优于 99.9% 的完好性包络要求。

综合考虑卫星星历和时钟改正数、电离层改正数、对流层残余误差和接收机观测噪声的影响, 对改正后的 GPS 卫星 URE 进行分析, 结果如图 8 所示。图 8(a)为经 WAAS 改正后 GPS URE 的均方根(root mean square, RMS)值, 均值为 0.34 m, 服务区边缘部分精度稍低; 图 8(b)为 EGNOS 对应的 GPS URE 结果, 均值为 0.33 m。与图 6 的 GIVD 精度相比, URE 精度较高, 主要是由于 GIVD 评估时使用双频伪距观测值计算电离层延迟, 放大了观测噪声。

2.4　定位性能

对 WAAS 和 EGNOS 的定位性能进行分析, 包括定位精度、完好性及定位精度可用性。由于本文数据采样间隔为 30 s, 因此, 未对服务连续性进行统计。图 9 为 WAAS 和 EGNOS 服务区域内各测站的垂直定位精度, WAAS 和 EGNOS 各测站垂直定位精度为 0.8 m 左右。在服务区边缘, 受可视卫星数和电离层格网点可用性影响, 定位精度较服务区中心低, 在 WAAS 服务区北纬 60°以北及 EGNOS 服务区东南部表现较为明显。

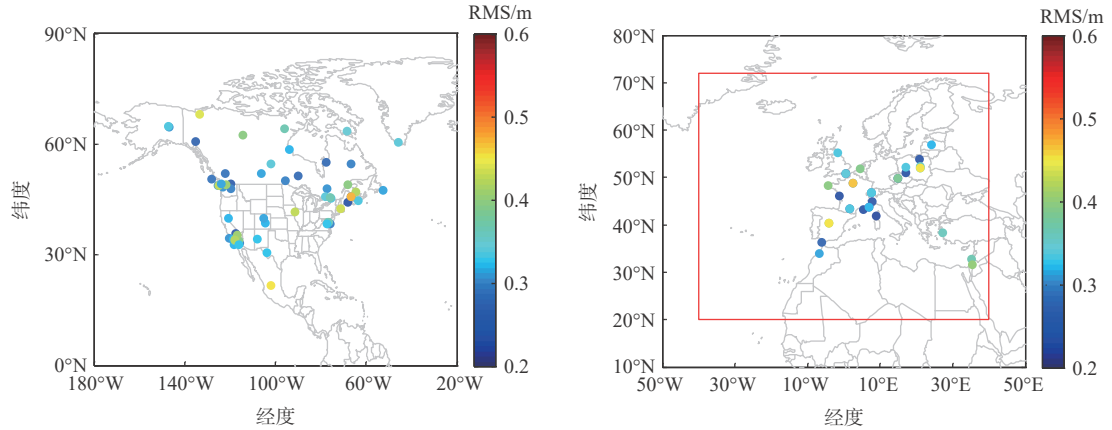
对所有测站的水平和垂直定位精度进行统计, 结果如图 10 所示。图 10(a)为 WAAS 东西、南北和高程方向的定位误差分布。WAAS 68% 分位数的水平和垂直定位误差为 0.60 m 和 0.82 m, 95% 分位数的定位误差为 1.19 m 和 1.72 m。由于定位精度与测站位置相关, 68% 和 95% 分位数的定位误差统计值与正态分布存在差异。与美国联邦航空管理局(Federal Aviation Administration, FAA)发布的 WAAS 性能评估报告相比[36], 本文定位精度稍低, 主要是因为本文对所有测站的定位精度进行统计, 处于服务区边缘的测站定位精度较低, 所以影响了



(a) WAAS伪距域完好性　　(b) EGNOS伪距域完好性

图 7　各测站 SBAS 伪距域完好性

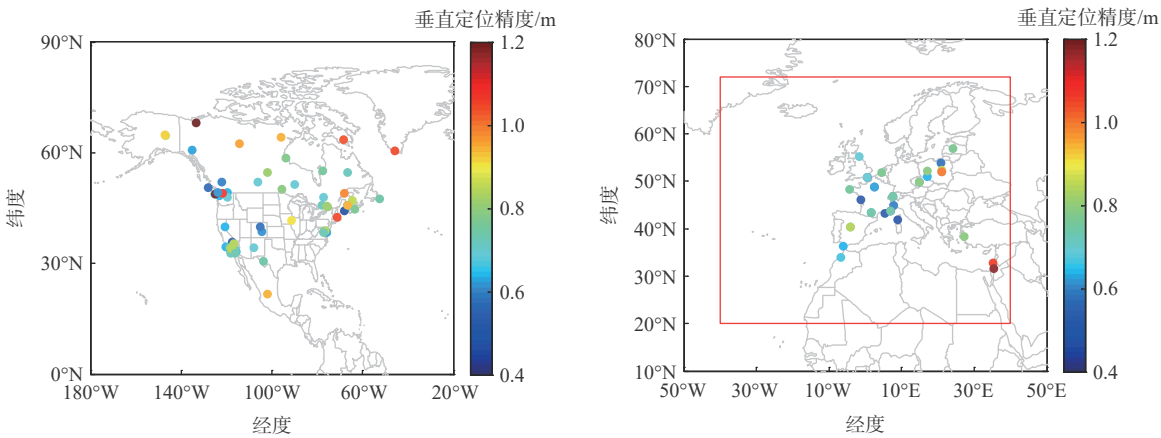
Fig. 7　SBAS pseudo range domain integrity of each station



(a) WAAS URE精度　　(b) EGNOS URE精度

图 8　经 SBAS 改正后各测站 GPS 卫星 URE 精度

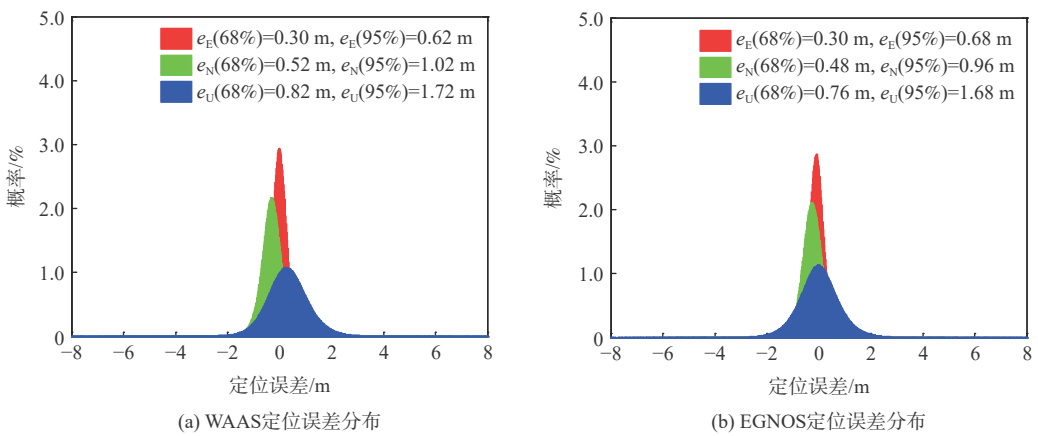
Fig. 8　GPS satellites URE accuracy of each station after using the SBAS corrections



(a) WAAS垂直定位精度　　(b) EGNOS垂直定位精度

图 9　各测站 SBAS 垂直定位精度

Fig. 9　SBAS vertical positioning accuracy of each station



(a) WAAS定位误差分布　　(b) EGNOS定位误差分布

图 10　SBAS 定位误差分布

Fig. 10　SBAS positioning error distribution

统计结果。图 10(b)为 EGNOS 的定位误差结果，68% 分位数的水平和垂直定位误差为 0.57 m 和 0.76 m，95% 分位数对应 1.18 m 和 1.68 m，与欧洲卫星服务提供商(European satellite services provider, ESSP)发布的 EGNOS 统计结果[37] 较一致。

为分析电文更新周期对定位精度的影响，利用 EGNOS PRN 码为 126 卫星播发的电文对测站进行定位。图 11 为 AJAC 测站 2022 年第 1～第 10 天分别利用 PRN 码为 126 和 123 的卫星播发的电文的定位结果，上、中、下 3 幅子图分别表示东西、南北和高程方向的误差，图中“×”号为使用 PRN 码为 123 的卫星电文的结果，“+”号为使用 PRN 码为

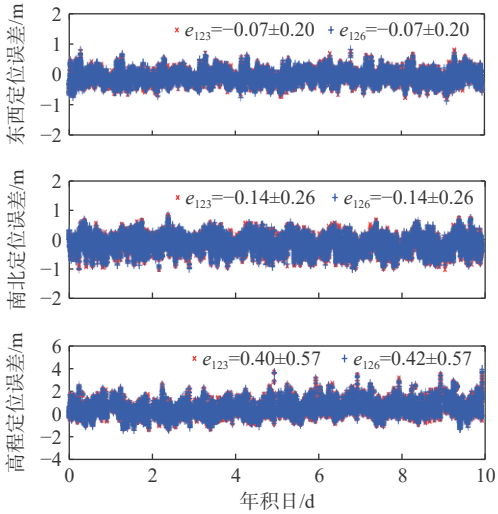


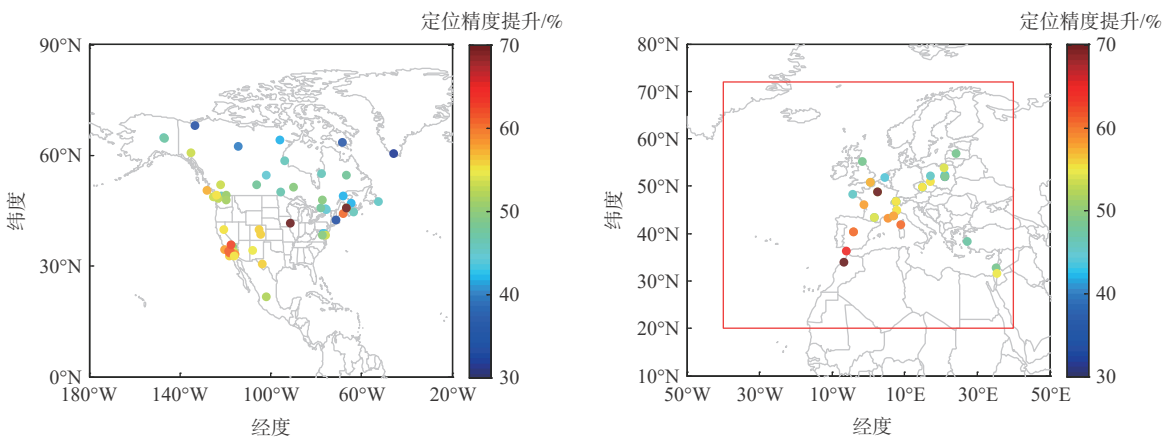
图 11　AJAC 测站利用 EGNOS 126 和 123 号卫星电文定位的误差

Fig. 11　Positioning errors of station AJAC based on the SBAS messages from EGNOS 126 and 123 satellites

126 的卫星电文的结果。2 种电文情况下, 测站东西和南北方向定位结果一致, 只在高程方向出现了 2 cm 的均值偏差。图中高程方向 0.42±0.57 cm 表示误差均值为 0.42 cm, 标准差为 0.57 cm。其他测站结果类似。因此, 尽管 PRN 码为 126 的卫星由于播发 MT0 电文导致更新周期变长, 但其对定位结果的影响较小, 主要是因为卫星轨道、钟差等改正数在短时间内变化较小。

图 12 为 SBAS 相对于 GPS 三维定位精度的提升。WAAS 和 EGNOS 二者定位精度相对于 GPS 的精度提升均超过 50%, 在低纬度地区提升更为明显, 达到 50%～70%, 主要是低纬度地区电离层较为活跃, GPS 广播电离层模型精度有限, 而 WAAS 和 EGNOS 利用测站实时数据生成的电离层改正数精度较高。

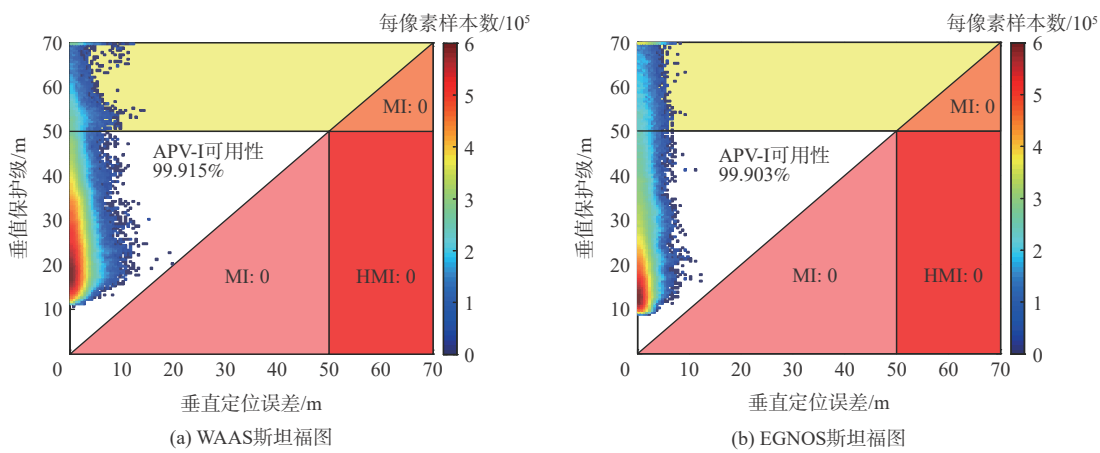
对测站定位结果的完好性进行分析, 图 13 为 WAAS 和 EGNOS 垂直定位结果的斯坦福图。图中对角线上方区域表示定位保护级大于定位误差, 未发生完好性风险事件, 当定位结果处于白色区域时, 定位保护级小于告警门限, 认为定位结果可用; 对角线下方区域则表示定位误差超过保护级, 此时



(a) WAAS定位精度提升　　(b) EGNOS定位精度提升

图 12　SBAS 定位精度相对于 GPS 的提升

Fig. 12　SBAS positioning accuracy improvement compared with GPS



(a) WAAS斯坦福图　　(b) EGNOS斯坦福图

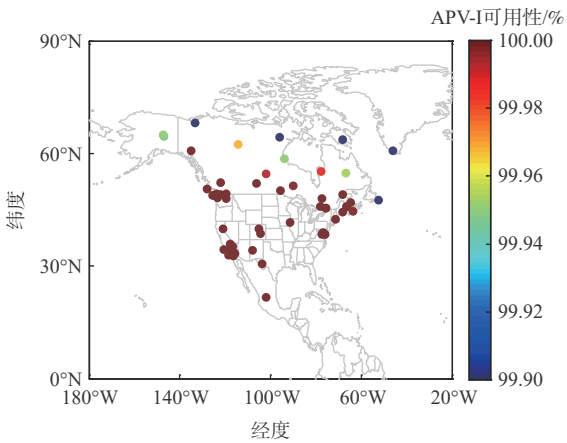
图 13　SBAS 垂直定位结果斯坦福图

Fig. 13　Stanford diagram of SBAS vertical positioning results

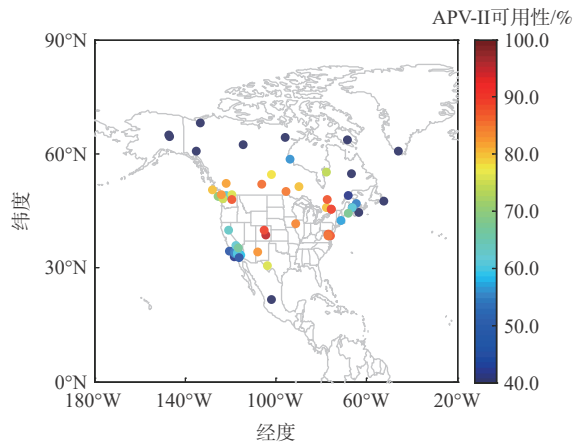
系统产生误导信息(misleading information, MI),其中当定位结果处于右下方区域时,表示定位误差大于告警门限,而保护级仍小于告警门限,系统并不会发出告警信息,此时将产生危险误导信息(hazardous misleading information, HMI),影响民航飞行安全。可以看出,在评估时段内二者均未发生完好性风险事件。由于部分评估测站位于服务区边缘,定位保护级较大,超过了告警门限,导致定位结果不可用。考虑到所选取测站均采用测地型接收机,验证了其较高的数据可靠性及用于SBAS评估的可行性。

对定位结果在APV-I和APV-II阶段的定位精度可用性进行分析,如图14所示,其中,APV-I阶段对应水平和垂直告警门限为40 m和50 m,APV-II阶段的水平和垂直告警门限为40 m和20 m[29]。图14(a)为WAAS在APV-I阶段的定位精度可用性,所有测站平均可用性为99.9%,中心区域可用性较高,边缘部分可用性降低较为明显。图14(b)为APV-II阶段定位精度的可用性,相较于APV-I阶段,其定位精度可用性降低明显,中心区域可用性为80%～95%,边缘区域则为20%～40%。

图15为EGNOS的定位精度可用性,APV-I阶段平均定位精度可用性优于99.9%,APV-II阶段平均定位精度可用性优于94.2%,主要是EGNOS评估测站集中在服务区中心,定位保护级较小,矩形框所示区域为服务区域。与WAAS类似,EGNOS中心区域定位精度可用性较高,边缘区域定位精度可用性较低。



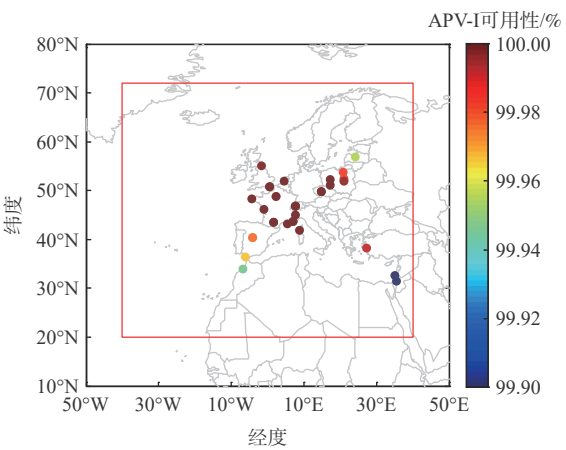
(a) APV-I阶段可用性



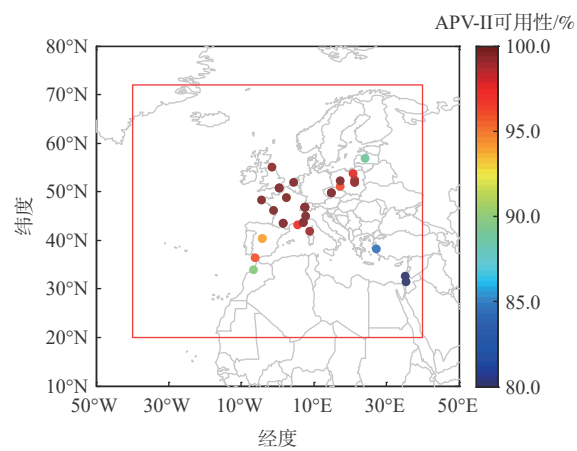
(b) APV-II阶段可用性

图14 WAAS定位精度的可用性

Fig. 14 WAAS availability of positioning accuracy



(a) APV-I阶段可用性



(b) APV-II阶段可用性

图15 EGNOS定位精度的可用性

Fig. 15 EGNOS availability of positioning accuracy

3 结 论

1) 本文分析了SBAS服务性能评估时应重点考虑的因素,给出了SBAS评估对测站分布、测站数量和数据时长的要求。在此基础上,利用100个IGS站240天的观测数据对WAAS和EGNOS的服

务性能进行研究, 包括电离层改正精度、GIVE 参数完好性、伪距域改正精度、伪距域完好性、定位精度、完好性和可用性等。针对 UDRE 参数评估时双频无电离层组合放大观测噪声的问题, 提出伪距域完好性评估方法。

2）WAAS 和 EGNOS 均满足小于 10^{-3} 的电文丢包率要求, 电离层 GIVE 参数完好性均优于 99.9%, 伪距域完好性分别为 99.999 95% 和 99.999 7%, 经改正后的 GPS URE 精度达到 0.3 m。WAAS 和 EGNOS 定位精度相对于 GPS 提升超过 50%, APV-I 阶段定位精度可用性达到 99.9%。评估时段内, 均未发生完好性风险事件, 体现了系统较好的服务性能, 同时验证了测地型接收机用于评估 SBAS 服务性能的可行性。

本文选取的时段电离层较为稳定, 后续 BDSBAS 民航验证评估时, 需考虑太阳活动高峰年电离层异常情况下的服务性能以及更大样本量的数据。

参考文献（References）

[1] WALTER T, SHALLBERG K, ALTSHULER E, et al. WAAS at 15[J]. Navigation, 2018, 65(4): 581-600.

[2] TABTI L, KAHLOUCHE S, BENADDA B, et al. Improvement of single-frequency GPS positioning performance based on EGNOS corrections in Algeria[J]. Journal of Navigation, 2020, 73(4): 846-860.

[3] DAMMALAGE T, DE SILVA D N, SATIRAPOD C. Performance analysis of GPS aided geo augmented navigation (GAGAN) over Sri Lanka[J]. Engineering Journal, 2017, 21(5): 305-314.

[4] SAITO S. MSAS system development [EB/OL]. (2019-06-05) [2022-08-25]. https://www.icao.int/APAC/APAC-RSO/GBASS-BAS%20Implementation%20Workshop/1-6_MSAS%20System%20Development_Rev2%20(S%20Saito).pdf.

[5] CHEN S S, JIN B, LI D J, et al. Study on the prediction method of single and dual frequency service area for BDSBAS[C]//Proceedings of the China Satellite Navigation Conference. Berlin: Springer, 2019: 228-237.

[6] ZHAO L Q, HU X G, TANG C P, et al. Generation of DFMC SBAS corrections for BDS-3 satellites and improved positioning performances[J]. Advances in Space Research, 2020, 66(3): 702-714.

[7] WU J Z, WANG K, EL-MOWAFY A. Preliminary performance analysis of a prototype DFMC SBAS service over Australia and Asia-Pacific[J]. Advances in Space Research, 2020, 66(6): 1329-1341.

[8] LEE E. System development-KASS[EB/OL]. (2019-06-05)[2022-08-25]. https://www.icao.int/APAC/APAC-RSO/GBASSBAS%20Implementation%20Workshop/1-8_KASS_Development_Status_final%20(E%20Lee).pdf.

[9] JIN B, CHEN S S, LI D J, et al. Ionospheric correlation analysis and spatial threat model for SBAS in China region[J]. Advances in Space Research, 2020, 66(12): 2873-2887.

[10] 金彪, 魏巍, 陈姗姗, 等. SBAS 星历改正数及 UDRE 参数生成算法分析[J]. 武汉大学学报 (信息科学版), 2021, 46(1): 111-117.
JIN B, WEI W, CHEN S S, et al. Analysis of SBAS ephemeris correction and UDRE generation algorithm[J]. Geomatics and Information Science of Wuhan University, 2021, 46(1): 111-117 (in Chinese).

[11] JIN B, CHEN S S, LI D J, et al. Performance analysis of SBAS ephemeris corrections and integrity algorithms in China region[J]. Satellite Navigation, 2021, 2(1): 15.

[12] WU J T, PECK S. An analysis of satellite integrity monitoring improvement for WAAS[C]//Proceedings of the 15th International Technical Meeting of the Satellite Division of the Institute of Navigation. Washington, D. C. : Institute of Navigation, 2002: 756-765.

[13] BLANCH J, WALTER T, ENGE P. A Clock and ephemeris algorithm for dual frequency SBAS[C]//Proceedings of the 24th International Technical Meeting of the Satellite Division of The Institute of Navigation. Washington, D. C. : Institute of Navigation, 2011: 2513-2519.

[14] BLANCH J, WALTER T, ENGE P. Evaluation of a covariance-based clock and ephemeris error bounding algorithm for SBAS[C]//Proceedings of the 27th International Meeting of the Satellite Division of The Institute of Navigation. Washington, D. C.: Institute of Navigation, 2014: 3270-3276.

[15] SHAO B, DING Q, WU X B. Estimation method of SBAS dual-frequency range error integrity parameter[J]. Satellite Navigation, 2020, 1(1): 1-8.

[16] WALTER T, HANSEN A, BLANCH J, et al. Robust detection of ionospheric irregularities[J]. Navigation, 2001, 48(2): 89-100.

[17] SPARKS L, BLANCH J, PANDYA N. Estimating ionospheric delay using Kriging: 1. Methodology[J]. Radio Science, 2011, 46(06): 1-13.

[18] SPARKS L, BLANCH J, PANDYA N. Estimating ionospheric delay using Kriging: 2. Impact on satellite-based augmentation system availability[J]. Radio Science, 2011, 46(6): 1-10.

[19] WALTER T, KEE C, CHAO Y, et al. Flight trials of the wide area augmentation system (WAAS)[C]//Proceedings of the 7th International Technical Meeting of the Satellite Division of The Institute of Navigation. Washington, D. C.: Institute of Navigation, 1994: 1537-1546.

[20] PERRIN O, SCARAMUZZA M, BUCHANAN T, et al. Flying EGNOS approaches in the Swiss Alps[J]. Journal of Navigation, 2006, 59(2): 177-185.

[21] SPECHT C, PAWELSKI J, SMOLAREK L, et al. Assessment of the positioning accuracy of DGPS and EGNOS systems in the bay of gdansk using maritime dynamic measurements[J]. Journal of Navigation, 2019, 72(3): 575-587.

[22] HEßELBARTH A, WANNINGER L. SBAS orbit and satellite clock corrections for precise point positioning[J]. GPS Solutions, 2013, 17(4): 465-473.

[23] 楼益栋, 郑福, 龚晓鹏, 等. QZSS 系统在中国区域增强服务性能评估与分析[J]. 武汉大学学报 (信息科学版), 2016, 41(3): 298-303.
LOU Y D, ZHENG F, GONG X P, et al. Evaluation of QZSS system augmentation service performance in China region[J]. Geomatics and Information Science of Wuhan University, 2016, 41(3): 298-

303 (in Chinese).

[24] HIRSCHBERGER M J, LO S, WALTER T. Reevaluating the message loss rate of the wide area augmentation system (WAAS) in flight[C]//Proceedings of the 2021 International Technical Meeting of The Institute of Navigation. Washington, D. C. : Institute of Navigation, 2021: 218-228.

[25] 陈姗姗, 金彪, 赵立谦, 等. SBAS 电文时序动态编排算法[J]. 北京航空航天大学学报, 2021, 47(10): 1996-2005.
CHEN S S, JIN B, ZHAO L Q, et al. Dynamic SBAS message scheduler algorithm[J]. Journal of Beijing University of Aeronautics and Astronautics, 2021, 47(10): 1996-2005 (in Chinese).

[26] 常志巧, 陈金平, 刘利, 等. WAAS 电离层格网播发特性及其性能评估[J]. 天文学进展, 2021, 39(2): 265-275.
CHANG Z Q, CHEN J P, LIU L, et al. Broadcast characteristics and performance evaluation of WAAS ionospheric grid[J]. Progress in Astronomy, 2021, 39(2): 265-275 (in Chinese).

[27] 金彪, 陈姗姗, 李祝莲, 等. SBAS GEO 卫星 URE 精度及定位增强研究[J/OL]. 武汉大学学报 (信息科学版), (2021-09-16)[2022-08-27]. https://kns.cnki.net/kcms/detail/42.1676.TN.20210915.1217.007.html.
JIN B, CHEN S S, LI Z L, et al. SBAS GEO satellite user range error and position augmentation research[J/OL]. Geomatics and Information Science of Wuhan University, (2021-09-16)[2022-08-27]. https://kns.cnki.net/kcms/detail/42.1676.TN.20210915.1217.007.html (in Chinese).

[28] 蔡洪亮, 孟轶男, 耿长江, 等. 北斗三号全球导航卫星系统服务性能评估: 定位导航授时、星基增强、精密单点定位、短报文通信与国际搜救[J]. 测绘学报, 2021, 50(4): 427-435.
CAI H L, MENG Y N, GENG C J, et al. BDS-3 performance assessment: PNT, SBAS, PPP, SMC and SAR[J]. Acta Geodaetica et Cartographica Sinica, 2021, 50(4): 427-435 (in Chinese).

[29] International Civil Aviation Organization. International standards and recommended practices annex 10 Vol. I. radio navigation aids: ISBN 978-92-9258-504-4[S]. Quebec: ICAO, 2018.

[30] RTCA. Minimum operational performance standards for global positioning system/wide area augmentation system airborne equipment: RTCA-DO-229D[S]. Washington, D. C. : RTCA, 2006.

[31] REN X D, CHEN J, LI X X, et al. Ionospheric total electron content estimation using GNSS carrier phase observations based on zero-difference integer ambiguity: methodology and assessment[J]. IEEE Transactions on Geoscience and Remote Sensing, 2021, 59(1): 817-830.

[32] WANG N B, LI Z S, DUAN B B, et al. GPS and GLONASS observable-specific code bias estimation: comparison of solutions from the IGS and MGEX networks[J]. Journal of Geodesy, 2020, 94(8): 74.

[33] LIU J B, HERNANDEZ-PAJARES M, LIANG X L, et al. Temporal and spatial variations of global ionospheric total electron content under various solar conditions[J]. Journal of Geodesy, 2017, 91(5): 485-502.

[34] European GNSS Agency. EGNOS safety of life (SoL) service definition document: 9789292060510[S]. Luxembourg: European GNSS Agency, 2021: 53-57.

[35] NIE Z X, ZHOU P Y, LIU F, et al. Evaluation of orbit, clock and ionospheric corrections from five currently available SBAS L1 services: Methodology and analysis[J]. Remote Sensing, 2019, 11(4): 411.

[36] FAA. Wide area augmentation system performance analysis report: NJ08405 [R]. Washton, D. C.: Federal Aviation Administration, 2022.

[37] European Satellite Services Provider. Monthly performance report: ESSP-DRD-24152 [R]. Toulouse: ESSP SAS, 2019.

Service performance assessment method of single frequency SBAS

JIN Biao[1]，LI Rui[2]，WANG Dun[1,*]，LIU Lei[3]，YUAN Jinxu[1]，LI Zixiao[1]

(1. Space Star Technology Co.，Ltd.，Beijing 100095，China；
2. Electronic Information Engineering，Beihang University，Beijing 100191，China；
3. Seventh Military Representive Room of Air Loading Beijing，Beijing 100076，China)

Abstract: Dual frequency ionosphere-free observations are usually applied to evaluate the satellite based augmentation system (SBAS) user differential range error (UDRE) integrity with the drawback of noise amplification. Instead, a range domain integrity method is proposed to assess the combined effect of SBAS ionospheric corrections, ephemeris and clock corrections. 240 days data of from 100 international global navigation satellite system（GNSS）service (IGS) stations are employed to evaluate the service performance of wide area augmentation system (WAAS) and European geostationary navigation overlay service (EGNOS). Results show that: The grid ionospheric vertical error (GIVE) integrity of WAAS and EGNOS achieves 99.96% and 99.91% respectively and the accuracy of ionospheric correction is 0.56 m. The range domain integrity of WAAS and EGNOS is better than 99.999% and the GPS user range error (URE) after applying the corrections achieves 0.34 m. For WAAS, the 95% horizontal and vertical positioning error is 1.19 m and 1.72 m, whereas for EGNOS, it is 1.18 m and 1.68 m. During the test, there were no integrity risk events. It is confirmed that it is feasible to evaluate the SBAS performance—particularly the integrity—using a geodetic receiver. The average availability of positioning accuracy for WAAS during approach with vertical guidance I (APV-I) is 99.9%, while for approach with vertical guidance II (APV-II), it ranges from 80% to 95% within the center of the service area. The average availability of positioning accuracy for APV-I of EGNOS is 99.9%, and for APV-II, it is 94.2%, indicating a higher level of availability, largely because most of the EGNOS assessment stations are located in the center of the service area. The SBAS availability noticeably decreases at the edge of the service area.

Keywords: satellite based augmentation system; grid ionospheric vertical error; range domain; integrity; availability

Received：2022-09-15；**Accepted**：2022-11-04；**Published Online**：2023-02-01 14：55
URL：link.cnki.net/urlid/11.2625.V.20230131.1834.002
Foundation item：National Natural Science Foundation of China (41974041)
* **Corresponding author**. E-mail: dawun04@163.com

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2022.0769

非均匀热流影响超临界二氧化碳流动换热数值研究

代威[1]，石申舟[2]，付衍琛[2,*]，左洋[2]
(1. 北京宇航系统工程研究所，北京 100076；2. 北京航空航天大学 航空发动机研究院，北京 100191)

摘 要：针对部分极端工况下表面热流密度高、分布不均匀的特点，采用数值计算分析方法研究非均匀热流对超临界二氧化碳流动与换热特性的影响。针对内径为4 mm水平圆管内超临界压力二氧化碳的流动传热，依次在相同总热流不同变化斜率、不同总热流相同变化斜率的2组线性非均匀热流工况下进行数值模拟，得到线性非均匀热流工况下热物性分布、浮升力影响对流换热特性规律。仿真结果表明：相较于均匀热流，非均匀热流工况下加热段入口处附近上壁面传热恶化更严重，线性非均匀热流分布下该处管壁温度可达均匀热流工况下的1.85倍；对给定的实际表面热流边界条件下的超临界二氧化碳流动换热进行数值仿真，加热段入口处上壁面最大温度达出口温度的3.41倍，改变流动方向后这一比值降低到1.50，据此对采取超临界二氧化碳为工质的相关冷却方案给出了建议。

关 键 词：非均匀热流；超临界流体；二氧化碳；对流换热；浮升力
中图分类号：V231.1
文献标志码：A **文章编号**：1001-5965(2024)10-3074-10

超临界流体在临界及拟临界区域热物性参数随温度变化剧烈，表现出优异的流动传热特性，在微小管道内通入超临界压力二氧化碳作为冷却工质可实现高效、高质的主动冷却。此外，超临界压力二氧化碳又有临界参数低、性质无毒、不易燃、存量丰富和易于获得等优点，在工程领域得到广泛应用[1-3]。国内外学者高度关注其流动换热特性，从实验、数值仿真、理论分析[4-6]三大方面分别展开了大量研究。

Adebiyi和Hall[7]对22.14 mm内径的水平圆管内超临界二氧化碳流动传热展开了实验研究，结果表明，浮升力使管径周向壁面温度出现差异，管道底部换热得到强化而顶部换热恶化。这一结论在Pidaparti等[8]与Chu和Laurien[9]的研究中得到验证。

淮秀兰和Koyama[10]对31 mm内径的水平管中超临界二氧化碳临界区域的热物性与传热能力变化展开了研究。研究发现，超临界二氧化碳的流体温度和压力达到拟临界区域时，温度升高，流体压降急剧增大，且随质量流量的变大而减小。超临界二氧化碳的热物性在临界点附近剧烈变化，定压比热和热导率迅速上升，而密度、黏度迅速减小，流体传热能力增强，随着温度上升对流换热系数不断升高。流体温度达到亚临界时换热系数最大，此后开始逐渐减小。

目前学界关于超临界压力下传热强化与恶化发生的原因及具体机理提出了2种理论：变物性单相传热理论和拟沸腾传热理论。研究表明，拟沸腾现象在超临界流体传热过程中确实存在，但拟沸腾传热理论适用范围有限，对于部分实际传热恶化和强化的现象无法给出合理解释。当前学术界主流更倾向于变物性单相传热理论。变物性单相传热理论将导致超临界流体传热强化、传热恶化发生的主要因素为热加速效应、浮升力效应和物性非均匀性的有机耦合，并在实验、数值研究中得到了不同

收稿日期：2022-09-14；**录用日期**：2022-12-18；**网络出版时间**：2023-01-10 15:40
网络出版地址：link.cnki.net/urlid/11.2625.V.20230109.1825.009
基金项目：国家自然科学基金(51906009)
*通信作者. E-mail：yanchenfu@buaa.edu.cn

引用格式：代威，石申舟，付衍琛，等. 非均匀热流影响超临界二氧化碳流动换热数值研究[J]. 北京航空航天大学学报，2024，50(10)：3074-3083. DAI W，SHI S Z，FU Y C，et al. Numerical study on flow and heat transfer of supercritical carbon dioxide under non-uniform heat flux influences[J]. Journal of Beijing University of Aeronautics and Astronautics，2024，50(10)：3074-3083 (in Chinese).

维度、不同角度的验证。

针对浮升力影响超临界流体传热, 学者展开了相关研究。Palko 和 Anglart [11] 研究发现, 质量流速 G 相对小、热流密度 q 相对大时浮升力效应更显著, 传热恶化更严重。类似地, 在 G 相对小、q 相对大的工况下, 王军辉等[12] 开展了水平圆管超临界二氧化碳对流传热实验, 采用浮升力数 Bu 对实验数据进行分析, 结果表明, 此时传热的恶化主要受浮升力效应的影响, 而流动加速效应可以忽略不计。吴超等[13] 针对水平管采用粒子图像测速 (particle image velocity, PIV) 实验方法分析了浮升力导致的二次流对超临界流体流动传热的影响。Kao 等[14] 采用数值模拟法研究发现, 浮升力会抑制湍动能和对流强度, 进而导致传热恶化。

针对热物性变化影响超临界流体传热, Jackson[15] 通过大量超临界流体实验进行理论分析, 认为在变物性单相传热理论所涉及的三大主要影响因素中, 流体热物性的非均匀性最为关键, q 较大时, 壁面附近温度差异大, 超临界流体密度差异导致流动加速效应和浮升力效应更加显著, 进而影响流场分布和湍流结构, 导致局部湍流流场层流化, 传热能力下降。目前 Kim 等[16-18] 许多学者对此观点表示支持。

Jiang 等[19-23] 针对不同管径的竖直管内超临界二氧化碳流动换热特性进行了深入研究。研究表明, 内径为 2 mm 的圆管进口雷诺数 Re_{in} =9 000 时, 仅在向上流动的工况时会出现局部换热恶化, 这主要由热物性非均匀分布和浮升力效应导致; 1 mm 内径圆管中, Re_{in}>6 000 且热流密度高时, 局部换热恶化现象仍仅出现在向上流动工况中, 此时下浮升力影响仍十分显著, 而热加速效应可忽略不计; 内径为 0.27 mm 的圆管内, Re_{in}<2 700 且当地热流密度较高(113 kW/m^2)时, 向上、向下流动的工况均会使内壁面温度非线性变化, 此时热加速效应剧烈而浮升力效应不显著。其研究结果在一定程度验证、支持了变物性单相传热理论。而在浮升力影响湍流度的机理研究上, Zhao 等[24] 认为入口雷诺数从 1 900 增加到 9 000 时, 浮升力通过剪切力作用对湍流产生的贡献比密度及速度波动的直接影响更为显著。

通过以上的总结发现, 目前研究超临界二氧化碳的流动换热大都集中在均匀热流的边界条件。在实际工程应用中, 热流往往呈非均匀分布, 特定工作环境下热流分布具备高度非均匀性, 局部热流密度极大而其他热流密度相对较小, 若按照均匀热流来进行主动冷却与热防护方案设计容易产生过大设计裕度与换热能力浪费。尤其在换热面积、装置质量受限的要求下需要进行针对性、最优化设计, 需要对于非均匀热流分布下流体的流动换热表现有更加具体的认识。本文通过 ANSYS Fluent 数值仿真方法研究非均匀热流对超临界二氧化碳流动与换热特性的影响, 探究超临界二氧化碳管内热物性分布、流动与换热特性规律, 最终为采用二氧化碳为工质以满足主动对流冷却需求的相关方案提供理论参考。

1　物理模型与计算方法

1.1　物理模型及边界条件

本文以 Zhang 等[25] 对超临界压力下水平管内二氧化碳流动换热的实验的研究数据为计算依据。如图 1 所示, 圆管内径 d 为 4 mm, 水平方向布置, 实验段总长度为 2 500 mm, 为保证流体充分发展在加热段左右两端分别设置 400 mm 和 100 mm 的绝热段, 加热段总长度为 2 000 mm, 施加均匀热流密度。

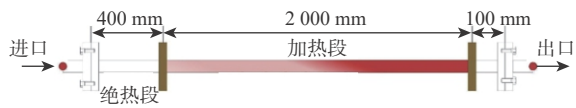


图 1　内径为 4 mm 的水平圆管几何模型示意图[25]

Fig. 1　Schematic diagram of geometric model of horizontal circular tube with a diameter of 4 mm [25]

内径 d 为 4 mm 的水平圆管物理模型的边界条件为: 入口处流动充分发展, 温度 t_{in}=25 ℃, 压力 p_{in}=8 MPa, 质量流速 G=283 kg/(m^2·s); 壁面处速度无滑移, 加热段均匀热流密度 q=32.5 kW/m^2。

超临界压力二氧化碳的宽温度范围物性数据调用 NIST 数据库[26] 获取, 通过 JOURNAL 文件使用内置的 piecewise-linear 方法导入算例中。如图 2 所示, 所用物性数据: 压力为 8 MPa, 温度为 200～1 600 K。该区间内定压比热最大值 $c_{p,max}$= 33.336 kJ/(kg·K),密度最大值 ρ_{max}=1 181.7 kg/m^3, 导热系数最大值 λ_{max}=182.1 mW/(m·K), 黏度最大值 μ_{max}= 256.45 μPa·s, 其中, a 为无量纲化热物性参数, T 和 T_{pc} 分别为温度和拟临界温度。

1.2　湍流模型验证

采用 ICEM CFD 划分三维网格, 网格截面如图 3 所示。根据 Cheng 等[27]、Yang 等[28] 和 Du 等[29] 的研究, 无量纲壁面坐标 Y^+≤1 时, 边界层网格对于计算的准确度影响较小, 因此, 设置第 1 层网格高度为 0.003 mm, 同时保持增长比不高于 1.1, 网格质量不低于 0.85。

陈玮玮[30] 于 2016 年对采用不同湍流模型的超

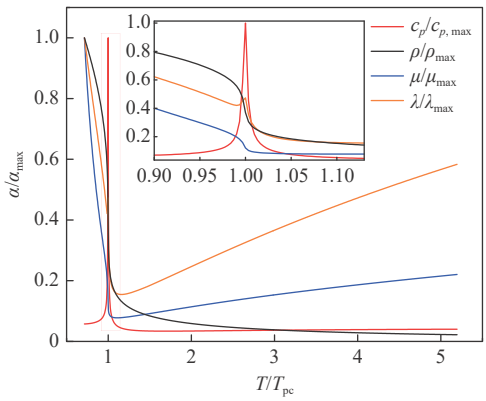


图 2 压力为 8 MPa 时热物性随温度分布

Fig. 2 Variation of thermophysical properties with temperature under pressure is 8 MPa

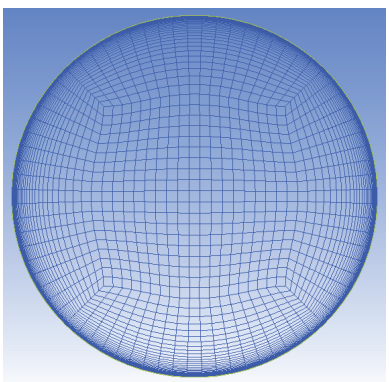


图 3 网格截面划分

Fig. 3 Meshing in cross-section

临界流体管内传热数值模拟研究进行了统计，包括标准 k–ε、Realizable k–ε、重整化群（renormalization group, RNG）k–ε、Modified k–ε、剪切应力传输（shear stress transfer, SST）k-ω、雷诺应力模型（Reynolds stress model, RSM）、代数 Boussinesq 模型（algebraic Boussinesq model, AB）、层流 Boussinesq 模型（laminar Boussinesq model, LB）、长度尺度模型（length scale model, LS）、杨-斯密模型（Yang-Shih model, YS）、Abe-Kondoh-Nagano 模型（AKN）、修正 k 模型（modified k model, MK）、垂直速度波动-平流湍流模型（velocity-vertical fluctuation model, V2F）、複合混合长宽比湍流模型（composite hybrid cubic model, CHC）等。分析其统计结果发现，在被统计的 29 篇论文中，9 篇论文认为 SST k-ω 模型仿真效果最理想，8 篇论文研究认为 RNG k-ε 模型仿真结果误差最小。He 等[31]的研究表明，带简单壁面修正的 k-ε 模型在对流体物性不均匀和浮升力所引起的二次流模上并不适用。SST k-ω 或 RNG k-ε 湍流模型都不能很好地预测存在较大浮升力影响的超临界流体换热。对于水平流动，浮升力对于换热的影响比较小，因此，采用上述湍流模型可以较好地对换热进行预测。

本文计算数据处理中，壁面温度 $T_{\rm wi}$ 为监测截面在周向上、下、左、右各 4 个点温度的平均值；流体温度 $T_{\rm b}$ 为监测截面流体温度的质量流量加权平均；对流换热系数定义为[32]

$$h=\frac{q}{T_{\rm wi}-T_{\rm b}} \tag{1}$$

针对水平管内浮升力对超临界流体换热影响的判别，Petukhov 等[33]提出当临界格拉晓夫数大于实际格拉晓夫数时，浮升力影响可以忽略；Jackson[34]提出忽略浮升力效应的准则关联式，并认为式中浮升力忽略阈值为 10；Kim 等[35]在 Jackson 之前的研究基础[34]上推导出了针对竖直圆管的浮升力准则关联式。

吕海财等[36]针对上述研究做了水平管实验验证，最终引入并优化了 Kim 等[35]的浮升力判据，发展了适合小通道的实验关联式，其中，浮升力数 Bu 定义为

$$Bu=\frac{Gr_{\rm b}}{Re_{\rm b}^{3.425}Pr_{\rm b}^{0.8}} \tag{2}$$

格拉晓夫数 $Gr_{\rm b}$ 由当地热流密度及物性计算得出

$$Gr_{\rm b}=\frac{\beta_{\rm b}gq_{\rm w}d^4}{\lambda_{\rm b}\nu_{\rm b}^2} \tag{3}$$

式中：下标 b 表示主流。

管径 d、壁面热流密度 $q_{\rm w}$ 由几何条件与边界条件获取，主流热物性 $\beta_{\rm b}$、$\lambda_{\rm b}$、$\nu_{\rm b}$ 根据流体温度从 NIST[26]数据库获取，重力加速度 g 取 9.81 m/s²。

使用同一网格，分别使用 SST k-ω 与 RNG k-ε 模型对照文献 [25] 内的实验工况进行数值模拟。控制除湍流模型外的其他所有设置相同，仿真结果与实验值对比如图 4 和图 5 所示，其中，$T_{\rm w}$ 为壁温，$T_{\rm b}$ 为主流温度，h 为对流换热系数。壁温仿真结果表明，2 种模型都能对流动传热总体变化趋势有良好把握：RNG k-ε 模型误差略大一些，平均偏差在 5 K 以内，整体误差为 1.6%；SST k-ω 的壁温结果对比实验值平均误差更小，在 0.5% 以内。上述结果

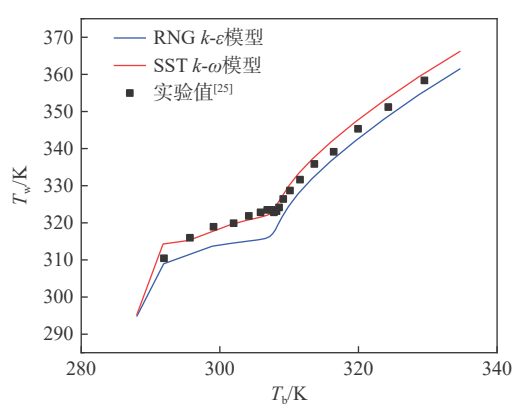


图 4 壁温仿真结果对比

Fig. 4 Comparison of simulation results of tube wall temperature

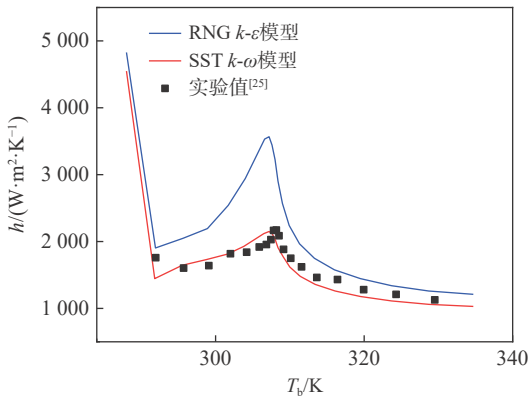


图 5　对流换热系数结果对比

Fig. 5　Comparison of convective heat transfer coefficients

与 Wang 等[37] 的研究结论相符。因此, 本文计算将采用 SST k-ω 模型。

1.3　网格无关性验证

进一步划分不同数量(70, 138, 300 万)的三维网格进行计算, 其他设置相同, 使用 SST k-ω 模型进行验证。壁温与对流换热系数的计算结果分别如图 6 和图 7 所示, 138 万网格与 300 万网格仿真平均结果相差不足 0.5%, 满足无关性要求。

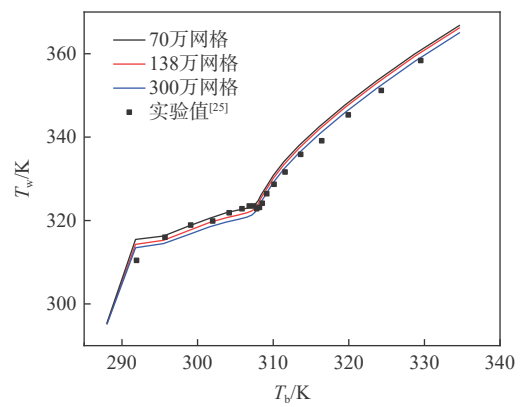


图 6　验证网格无关性的加热段壁温分布

Fig. 6　Tube wall temperature distribution in heating section for grid independence verification

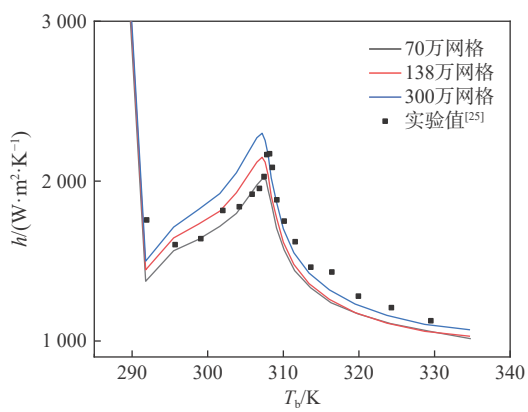


图 7　验证网格无关性的对流换热系数

Fig. 7　Convective heat transfer coefficient verification for grid independence verification

2　结果与讨论

2.1　热流非均匀度影响

为研究热流分布非均匀度对于超临界压力二氧化碳流动换热的影响, 设置如表 1 所示的 5 种工况, 其中, x 为轴向位置。控制温度 t_{in}=25 ℃, 压力 p_{in}=8 MPa, 质量流速 G=283 kg/(m^2·s), 仅使热流以不同沿程斜率分布, 加热段热流密度分布如图 8 所示。

表 1　非均匀度研究工况设置

Table 1　Conditions setting on non-uniformity

工况编号	q沿程分布/(kW·m^{-2})	q平均值/(kW·m^{-2})
1	q=32.5	32.5
2	q=−8.125x+40.625	32.5
3	q=−16.25x+48.75	32.5
4	q=−24.375x+56.875	32.5
5	q=−32.5x+65	32.5

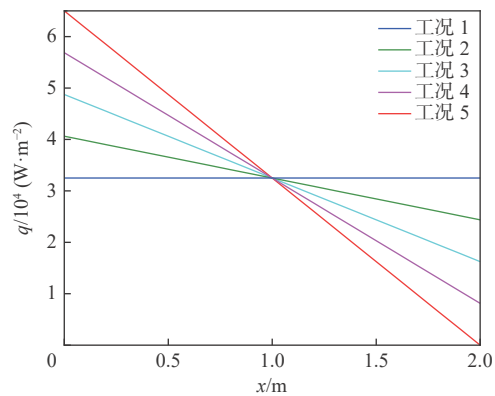


图 8　工况 1～工况 5 加热段热流密度分布

Fig. 8　Heat flux density distribution in heating section in condition 1～condition 5

工况 1～工况 5 温度场沿流向分布如图 9 所示。工况 1 热流均匀分布, 整体温度场未出现明显高温区。对比工况 1～工况 5, 随着热流变化斜率绝对值逐渐变大, 热流分布越不均匀, 加热段入口处出现越来越明显的局部高温区。在加热段入口处, 温度相对低的流体开始受到加热, 热边界层发展尚不完全, 导热能力差。热量堆积在管壁附近流体上, 使其温度迅速上升, 流体密度减小并向上聚集, 导热系数迅速减小, 导热能力进一步减弱, 最终形成了一个位于上壁面的传热恶化区域。入口处更大的局部热流导致热边界层更薄及上壁面附近流体温度更高、导热能力更差, 加剧传热恶化。

由图 9～图 11 纵向对比工况 1～工况 5 可知, 在入口处当地热流密度越大, 该处的主流温度变化不大而壁面温度明显更高, 同时对流换热系数更小, 传热恶化愈发严重。工况 1 中入口处上壁面最

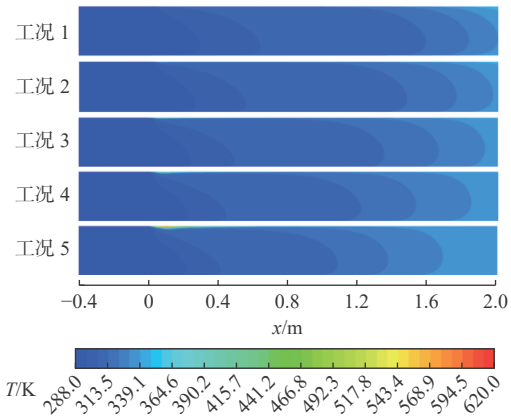


图 9 工况 1～工况 5 流向温度分布

Fig. 9 Temperature distribution along flow direction in condition 1～condition 5

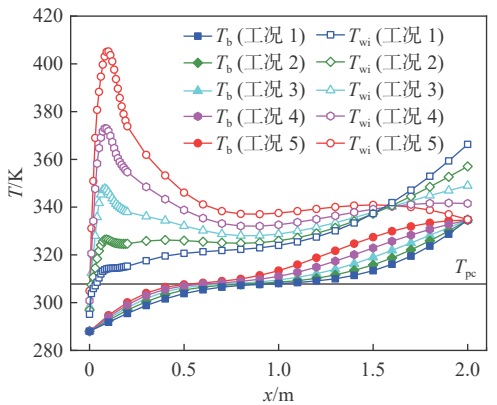


图 10 工况 1～工况 5 温度沿程分布

Fig. 10 Axial temperature distribution in condition 1～condition 5

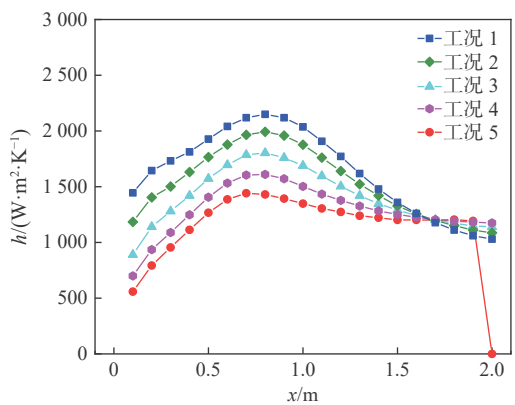


图 11 工况 1～工况 5 对流换热系数沿程分布

Fig. 11 Axial convective heat transfer coefficient distribution in condition 1～condition 5

大温度为 335.9 K，而在工况 5 中这一温度升高到 620.7 K，为原来的 1.85 倍，其中，p=8 MPa，G=283 kg/(m^2·s)。

流经加热段入口之后，流体内部温度也逐渐升高。随着流体主流温度达到临界点，定压比热迅速升高到原来的数十倍，吸热能力大幅升高，此时主流温度增长速度减缓，如图 10 所示；流体密度急剧减小，流体径向截面上发生较强的二次流动；黏性系数迅速减小，黏滞作用削弱，流动向旺盛湍流发展，具有更强掺混作用。对比图 10 和图 11，各个工况对流换热系数峰值均出现于主流温度T_b达到拟临界温度附近时，说明在临界区域换热得到强化。

如图 11 所示，随着热流分布非均匀度增加，从工况 1 到工况 5，平均对流换热系数依次下降，工况 5 的平均对流换热系数为工况 1 的 0.71 倍。工况 5 相较于工况 1 而言，前半段热流密度大，相应的浮升力效应更强，对传热的抑制也更强；后半段热流密度小，浮升力效应相对小，对传热的抑制减弱。

进一步分析热流非均匀分布对于换热的影响，提取对流换热系数、浮升力数Bu随主流温度的变化关系，并以Bu表征浮升力效应。

如图 12 和图 13 所示，以工况 1 为例，前半段对应的Bu在 5 种工况中最小，对应的对流换热系数则最大，后半段对应的Bu在 5 种工况中最大，对流换热系数则最小。5 种工况的Bu交叉区域与对流换热系数交叉区域高度对应。其中，p=8 MPa，G=283 kg/(m^2·s)。

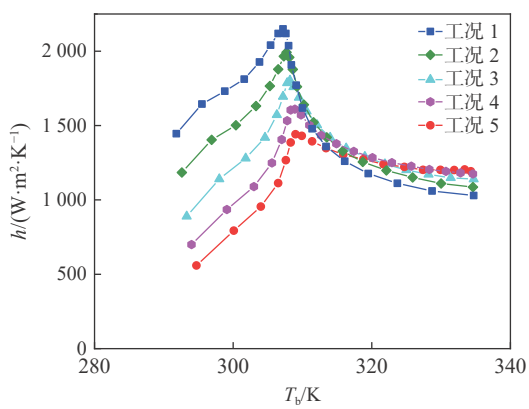


图 12 对流换热系数随主流温度变化趋势

Fig. 12 Variation of convective heat transfer coefficient with main stream temperature

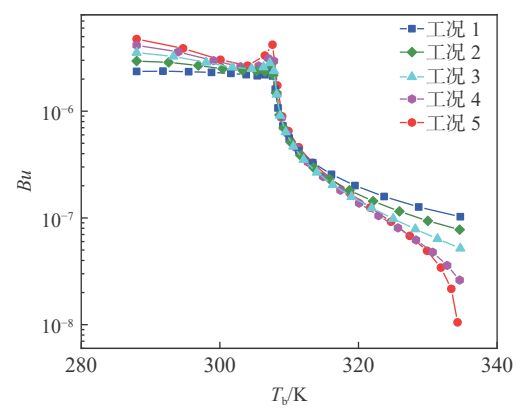


图 13 Bu随主流温度变化趋势

Fig. 13 Variation of Bu with main stream temperature

2.2　热流密度平均值影响

为研究热流密度平均值对于超临界压力二氧化碳流动换热的影响, 在工况 3 基础上拓展为如表 2 所示工况。仅改变热流密度平均值, 其余设置与工况 3 相同, 热流密度沿程分布如图 14 所示。

表 2　热流密度平均值工况设置

Table 2　Condition setting on average value of heat flux density

工况编号	q沿程分布/(kW·m^{-2})	q平均值/(kW·m^{-2})
3	q=−16.25x+48.75	32.5
3-1	q=−16.25x+56.875	40.625
3-2	q=−16.25x+65	48.75
3-3	q=−16.25x+73.125	56.875
3-4	q=−16.25x+81.25	65

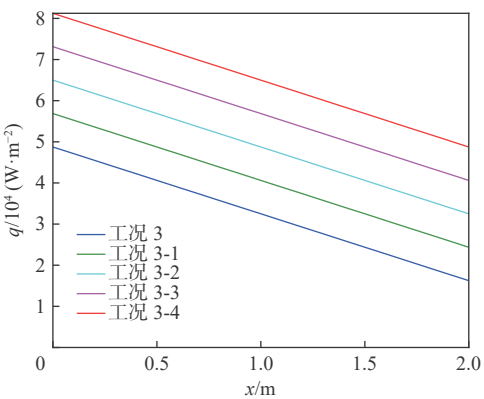


图 14　工况 3 及拓展工况热流沿程分布

Fig. 14　Axial heat flux distribution in condition 3 and its expanded conditions

流向温度分布如图 15 所示, 沿程温度曲线如图 16 所示。同样地, 由于加热段入口热流密度高, 换热性能差的流体向上聚集, 上壁面出现局部高温区; 工况 3 到工况 3-4 随热流密度增大, 加热段入口局部高温区愈发明显。

对比工况 3 与其拓展工况, 随着总体平均热流

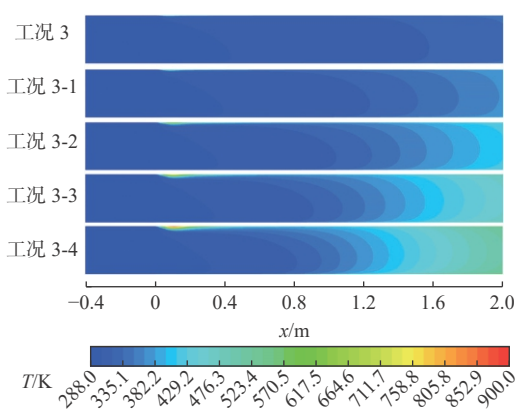


图 15　工况 3 及拓展工况流向温度分布

Fig. 15　Temperature distribution in condition 3 and its expanded conditions

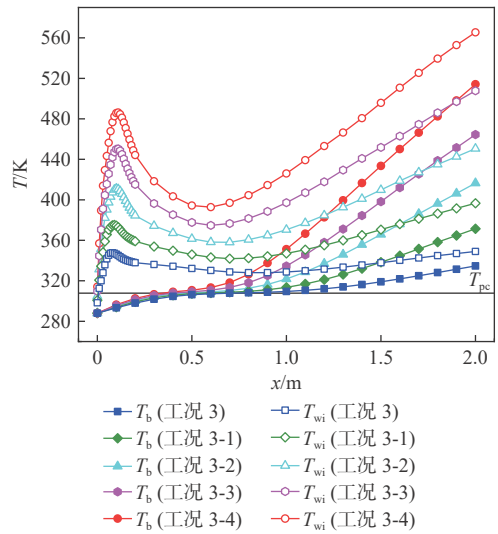


图 16　工况 3 及拓展工况温度

Fig. 16　Temperature of condition 3 and its expanded conditions

密度逐渐增大, 对流换热系数整体逐步减小, 且热流密度增大带来的影响也随之衰减, 如图 17 所示, 热流密度下降幅度在依次递减, 平均降幅从 20.03% 依次递减到 13.30%、8.13%; 从工况 3-3 到工况 3-4, 对流换热系数平均降幅仅有 4.96%, 对流换热系数曲线在部分区域甚至出现相当程度重合。

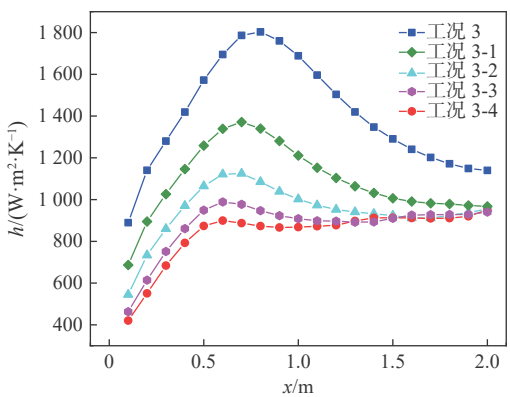


图 17　工况 3 及拓展工况对流换热系数

Fig. 17　Convective heat transfer coefficient of condition 3 and its expanded conditions

对热物性进行分析, 工况 3 到工况 3-1 不同工况间热流整体而言不断增大, 以工况 3-1 到 3-4 为例, 后半段主流温度远高于拟临界温度, 密度减小、黏度增大, 虽然此时速度仍在增加, 但Re整体上反而出现下降趋势, 如图 18 所示, 出现左拐的趋势, 流动的湍流度下降, 内部掺混作用削弱。另外, 后半段导热系数不断增大, 对流换热系数不断下降或基本不变, 因此, 努塞尔数Nu呈减小趋势, 图中出现下拐趋势。

对比工况 3 及其拓展工况, 随着热流密度平均值增大, 流场整体温度将远离拟临界点; 热流越大, 浮升力效应如图 19 所示更严重, 后期Re、Nu减小

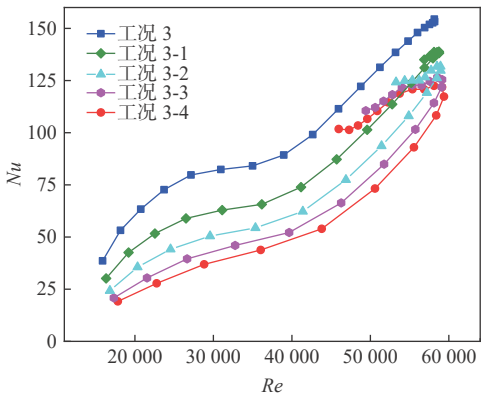


图 18 工况 3 及拓展工况*Nu*随*Re*变化趋势

Fig. 18 Variation of *Nu* with *Re* in condition 3 and its expanded conditions

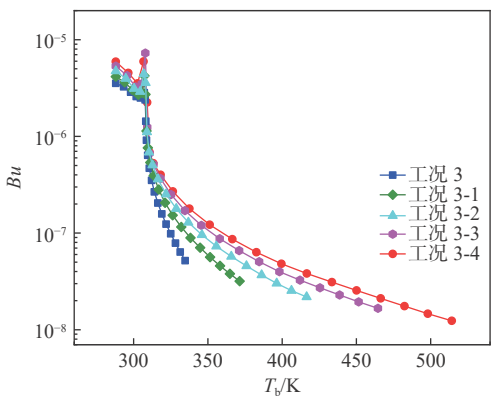


图 19 工况 3 及拓展工况*Bu*分布

Fig. 19 Distribution of *Bu* in condition 3 and its expanded conditions

越快，换热效果越差。

2.3 实际表面热流仿真

考虑到实际材料的承热能力，调节入口边界质量流速G=780 kg/(m^2·s)，其余设置不变。使用文献[38]中的实际热流进行拟合，并以软件内置拟合得到如图 20 所示热流分布。

实际非均匀热流的温度场分布如图 21 所示。仿真结果与 2.2 节中线性非均匀热流仿真结果相似，加热段入口处附近浮升力效应显著，导温能力差，上壁面最大温度为 1 600 K，高达流体出口温度的 3.41 倍。

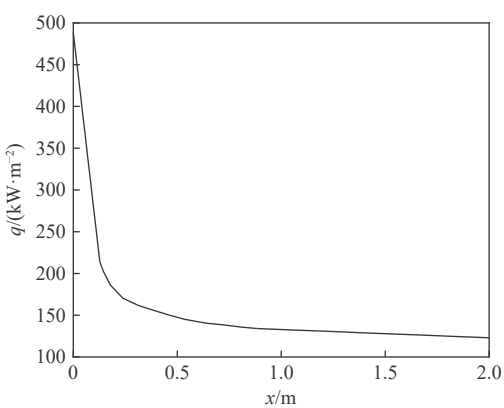


图 20 实际非均匀热流密度分布

Fig. 20 Distribution of actual non-uniform heat flux density

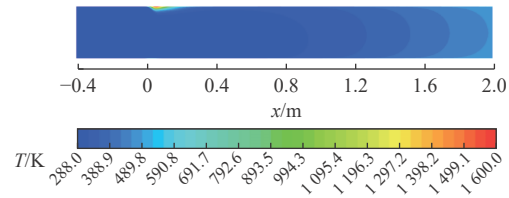


图 21 实际非均匀热流下温度场分布

Fig. 21 Temperature distribution under actual non-uniform heat flux

如图 22 和图 23 所示，对流换热系数的极大值仍然出现在主流温度接近临界点时，而高热流密度使得流场后半段主流温度远高于临界点，*Re*、*Nu*均先增后降，如图 24 所示，*Bu*峰值出现在临界区域，如图 25 所示。

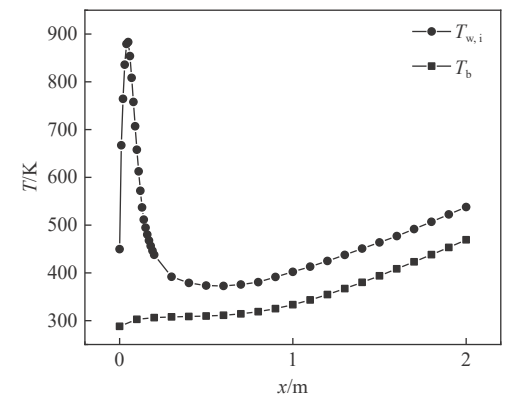


图 22 实际非均匀热流温度曲线

Fig. 22 Temperature curve of actual non-uniform heat flux

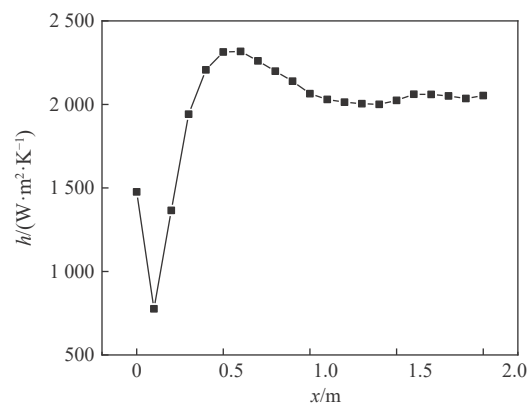


图 23 实际非均匀热流对流换热系数曲线

Fig. 23 Convective heat transfer coefficient curve of actual non-uniform heat flux

为探究实际非均匀热流分布与工质流动方向的相对关系，在原有模型基础上对调进出口后的流向温度场如图 26 所示。

改变流动方向后发现相较于之前总体传热效果得到了有效改善，局部最高温度降低到 706 K，从

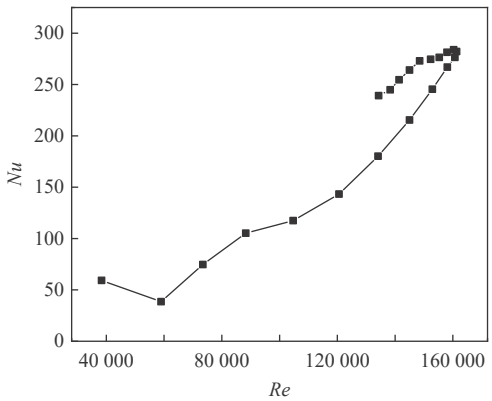


图 24　实际非均匀热流*Nu*随*Re*变化趋势

Fig. 24　Variation of *Nu* with *Re* under actual non-uniform heat flux

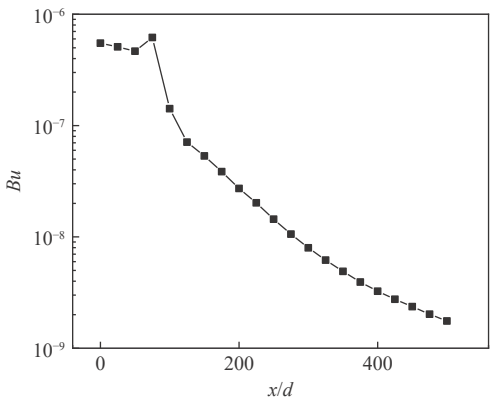


图 25　实际非均匀热流边界沿程*Bu*分布

Fig. 25　Distribution of *Bu* under actual non-uniform heat flux boundary

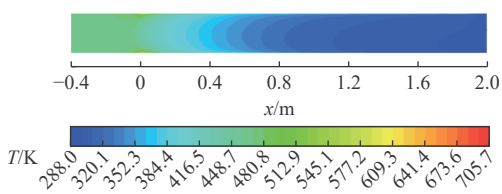


图 26　反向热流仿真温度云图

Fig. 26　Simulated temperature cloud image of reverse heat flux

原结果中的 3.41 倍出口温度降低到 1.50 倍, 整体传热效果显著增强。对流换热系数对比如图 27 所示, 改变流动方向后监测面的对流换热系数平均提升了 42.6%, 整体传热效果显著增强。

因此, 对于采用通入超临界压力二氧化碳作为冷却工质以满足主动对流冷却需求的相关方案提出以下建议:

1) 加热段入口处附近上壁面处传热恶化严重, 此处应为防护重点之一。

2) 入口附近热流密度大, 热边界层发展不完全, 浮升力效应显著, 导致传热恶化的出现。增大进口流量可以有效减弱浮升力效应, 缓解上壁面传热恶化。

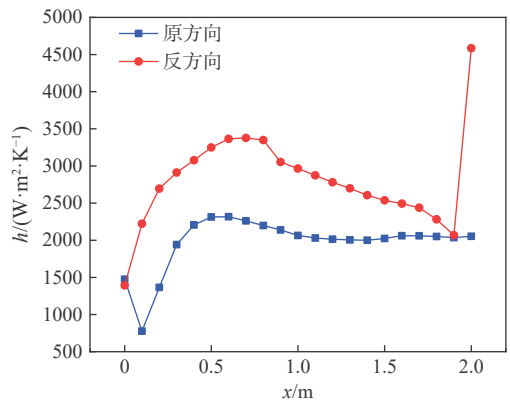


图 27　对流换热系数对比

Fig. 27　Comparison of convective heat transfer coefficients

3) 二氧化碳流动方向与热流密度降低方向相反时相比同向具有更好的传热效果, 建议采用该流动方向以缓解原方向工况的传热恶化。

3　结　论

本文利用数值模拟方法针对非均匀热流工况下水平管内超临界压力二氧化碳的流动传热特性及热物性分布规律进行研究。本文定义了 2 种不同非均匀热流的变化趋势并进行了数值模拟, 得到了线性非均匀热流下超临界压力二氧化碳流动传热特性。在此基础上, 对实际非均匀热流工况下管内二氧化碳流动传热进行了数值模拟, 得到了其流动传热特性及热物性分布规律, 并对实际非均匀分布热流工况下的热防护与主动冷却方案设计给出了建议。具体结论如下:

1) 对于热流密度平均值相同的线性热流分布, 热流非均匀度越高, 入口处热流越大, 上壁面温度越高, 该处传热恶化越严重; 随着热流非均匀度升高, 管壁最大温度可达均匀热流工况的 1.85 倍, 平均对流换热系数可降低到原来的 0.71 倍。

2) 对于相同变化斜率下的线性热流分布, 在流场出口温度达到二氧化碳物性拐点(340 K 左右)后, 热流密度越大, 整体浮升力效应增强, 传热效果越差, 但热流密度对于对流换热系数的影响会逐渐衰减。仿真结果表明, 不同工况间平均热流密度依次升高, 因而对流换热系数下降, 下降平均幅度由 20.03% 依次减小到 13.30%、8.13%, 最终减小到 4.96%。

参考文献 (References)

[1] ESKANDARI MANJILI F, CHERAGHI M. Performance of a new two-stage transcritical CO_2 refrigeration cycle with two ejectors[J]. Applied Thermal Engineering, 2019, 156: 402-409.

[2] TSAI W C, WANG Y D. Progress of supercritical fluid technology

in polymerization and its applications in biomedical engineering[J]. Progress in Polymer Science, 2019, 98: 101161.

[3] LIU Y H, LI Q L, DUAN X L, et al. Thermodynamic analysis of a modified system for a 1000 MW single reheat ultra-supercritical thermal power plant[J]. Energy, 2018, 145: 25-37.

[4] FEWSTER J. Mixed forced and free convective heat transfer to supercritical pressure fluids flowing in vertical pipes[M]. Manchester: The University of Manchester Press, 1976: 11-19.

[5] BRINGER R P, SMITH J M. Heat transfer in the critical region[J]. AIChE Journal, 1957, 3(1): 49-55.

[6] BOYKO L D, KRUZHILIN G N. Heat transfer and hydraulic resistance during condensation of steam in a horizontal tube and in a bundle of tubes[J]. International Journal of Heat and Mass Transfer, 1967, 10(3): 361-373.

[7] ADEBIYI G A, HALL W B. Experimental investigation of heat transfer to supercritical pressure carbon dioxide in a horizontal pipe[J]. International Journal of Heat and Mass Transfer, 1976, 19(7): 715-720.

[8] PIDAPARTI S R, MCFARLAND J A, MIKHAEIL M M, et al. Investigation of buoyancy effects on heat transfer characteristics of supercritical carbon dioxide in heating mode[J]. Journal of Nuclear Engineering and Radiation Science, 2015, 1(3): 031001.

[9] CHU X, LAURIEN E. Investigation of convective heat transfer to supercritical carbon dioxide with direct numerical simulation[C]// Proceedings of the High Performance Computing in Science and Engineering. Berlin: Springer, 2016: 315-331.

[10] 淮秀兰, KOYAMA S. 微通道内超临界二氧化碳的压降与传热特性[J]. 工程热物理学报, 2004, 25(5): 843-845.
HUAI X L, KOYAMA S. Heat transfer and pressure drop of supercritical carbon dioxide in multi-port channels[J]. Journal of Engineering Thermophysics, 2004, 25(5): 843-845 (in Chinese).

[11] PALKO D, ANGLART H. Theoretical and numerical study of heat transfer deterioration in high performance light water reactor[J]. Science and Technology of Nuclear Installations, 2008, 2008: 405072.

[12] 王军辉, 郭鹏程, 颜建国, 等. 水平小圆管内超临界二氧化碳对流传热特性的试验研究[J]. 西安理工大学学报, 2018, 34(3): 272-277.
WANG J H, GUO P C, YAN J G, et al. Experimental studies on convection heat transfer of supercritical carbon dioxide in a horizontal circular mini-tube[J]. Journal of Xi'an University of Technology, 2018, 34(3): 272-277 (in Chinese).

[13] 吴超, 李会雄, 张谦. 超临界流体水平管换热特性可视化研究[J]. 工程热物理学报, 2014, 35(5): 897-901.
WU C, LI H X, ZHANG Q. Visualization experimental investigations of the heat transfer characteristics in a horizontal tube under supercritical pressure[J]. Journal of Engineering Thermophysics, 2014, 35(5): 897-901 (in Chinese).

[14] KAO M T, LEE M, FERNG Y M, et al. Heat transfer deterioration in a supercritical water channel[J]. Nuclear Engineering and Design, 2010, 240(10): 3321-3328.

[15] JACKSON J D. Fluid flow and convective heat transfer to fluids at supercritical pressure[J]. Nuclear Engineering and Design, 2013, 264: 24-40.

[16] BAE Y Y, KIM H Y, KANG D J. Forced and mixed convection heat transfer to supercritical CO_2 vertically flowing in a uniformly-heated circular tube[J]. Experimental Thermal and Fluid Science, 2010, 34(8): 1295-1308.

[17] KIM D E, KIM M H. Experimental investigation of heat transfer in vertical upward and downward supercritical CO_2 flow in a circular tube[J]. International Journal of Heat and Fluid Flow, 2011, 32(1): 176-191.

[18] KIM D E, KIM M H. Two layer heat transfer model for supercritical fluid flow in a vertical tube[J]. The Journal of Supercritical Fluids, 2011, 58(1): 15-25.

[19] JIANG P X, ZHANG Y, SHI R F. Experimental and numerical investigation of convection heat transfer of CO_2 at super-critical pressures in a vertical mini tube[C]//Proceedings of the ASME 4th International Conference on Nanochannels, Microchannels, and Minichannels. New York: ASMEDC, 2006: 583-590.

[20] JIANG P X, SHI R F, XU Y J, et al. Experimental investigation of flow resistance and convection heat transfer of CO_2 at supercritical pressures in a vertical porous tube[J]. The Journal of Supercritical Fluids, 2006, 38(3): 339-346.

[21] JIANG P X, ZHANG Y, XU Y J, et al. Experimental and numerical investigation of convection heat transfer of CO_2 at supercritical pressures in a vertical tube at low Reynolds numbers[J]. International Journal of Thermal Sciences, 2008, 47(8): 998-1011.

[22] JIANG P X, LIU B, ZHAO C R, et al. Convection heat transfer of supercritical pressure carbon dioxide in a vertical micro tube from transition to turbulent flow regime[J]. International Journal of Heat and Mass Transfer, 2013, 56(1-2): 741-749.

[23] JIANG P X, XU Y J, LV J, et al. Experimental investigation of convection heat transfer of CO_2 at super-critical pressures in vertical mini-tubes and in porous media[J]. Applied Thermal Engineering, 2004, 24(8-9): 1255-1270.

[24] ZHAO C R, LIU Q F, ZHANG Z, et al. Investigation of buoyancy-enhanced heat transfer of supercritical CO_2 in upward and downward tube flows[J]. The Journal of Supercritical Fluids, 2018, 138: 154-166.

[25] ZHANG S J, XU X X, LIU C, et al. Experimental and numerical comparison of the heat transfer behaviors and buoyancy effects of supercritical CO_2 in various heating tubes[J]. International Journal of Heat and Mass Transfer, 2020, 149: 119074.

[26] LEMMON E W, HUBER M L, MCLINDEN M O. NIST standard reference database 23[S]. Washington, D. C. : National Institute of Standards and Technology, 2010.

[27] CHENG X, KUANG B, YANG Y H. Numerical analysis of heat transfer in supercritical water cooled flow channels[J]. Nuclear Engineering and Design, 2007, 237(3): 240-252.

[28] YANG J, OKA Y, ISHIWATARI Y, et al. Numerical investigation of heat transfer in upward flows of supercritical water in circular tubes and tight fuel rod bundles[J]. Nuclear Engineering and Design, 2007, 237(4): 420-430.

[29] DU Z X, LIN W S, GU A Z. Numerical investigation of cooling heat transfer to supercritical CO_2 in a horizontal circular tube[J]. The Journal of Supercritical Fluids, 2010, 55(1): 116-121.

[30] 陈玮玮. 管内超临界流动传热特性及应用研究[D]. 南京: 南京航空航天大学, 2016: 17-18.
CHEN W W. Research on the heat transfer characteristics and ap-

plication of In-tube supercritical flow[D]. Nanjing: Nanjing University of Aeronautics and Astronautics, 2016: 17-18 (in Chinese) .

[31] HE S, KIM W S, JIANG P X, et al. Simulation of mixed convection heat transfer to carbon dioxide at supercritical pressure[J]. Journal of Mechanical Engineering Science, 2004, 218(11): 1281-1296.

[32] 杨世铭, 陶文铨. 传热学[M]. 4 版. 北京: 高等教育出版社, 2006: 99-117.
YANG S M, TAO W Q. Heat transfer[M]. 4th ed. Beijing: Higher Education Press, 2006: 99-117 (in Chinese).

[33] PETUKHOV B S, POLYAKOV A F, KULESHOV V A, et al. Turbulent flow and heat transfer in horizontal tubes with substantial influence of thermogravitational forces[C]//Proceeding of the International Heat Transfer Conference 5. Connecticut: Begellhouse, 1974: 3-7.

[34] JACKSON J D. Influences of buoyancy on heat transfer to fluids flowing in vertical tubes under turbulent conditions[J]. Turbulent Forced Convection in Channels and Bundles, 1979, 2: 613-640.

[35] KIM D E, KIM M H. Experimental study of the effects of flow acceleration and buoyancy on heat transfer in a supercritical fluid flow in a circular tube[J]. Nuclear Engineering and Design, 2010, 240 (10): 3336-3349.

[36] 吕海财, 赵金乐, 潘辉, 等. 超临界 CO_2 水平管内浮升力和热加速效应评判准则[J]. 西安交通大学学报, 2018, 52(9): 140-147.
LYU H C, ZHAO J L, PAN H, et al. Evaluation criterion of buoyancy and thermal acceleration behavior of supercritical CO_2 in a horizontal circle tube[J]. Journal of Xi'an Jiaotong University, 2018, 52(9): 140-147 (in Chinese).

[37] WANG K Z, XU X X, WU Y Y, et al. Numerical investigation on heat transfer of supercritical CO_2 in heated helically coiled tubes[J]. The Journal of Supercritical Fluids, 2015, 99: 112-120.

[38] 孙健, 刘伟强. 高超声速飞行器热管冷却前缘结构一体化建模分析[J]. 物理学报, 2013, 62(7): 239-246.
SUN J, LIU W Q. Investigation on integral model of heat-pipe-cooled leading edge of hypersonic vehicle[J]. Acta Physica Sinica, 2013, 62(7): 239-246 (in Chinese).

Numerical study on flow and heat transfer of supercritical carbon dioxide under non-uniform heat flux influences

DAI Wei[1], SHI Shenzhou[2], FU Yanchen[2,*], ZUO Yang[2]

(1. Beijing Institute of Astronautical Systems Engineering, Beijing 100076, China;
2. Research Institute of Aero-Engine, Beihang University, Beijing 100191, China)

Abstract: In some extreme conditions, heat flux with high density is non-uniformly distributed on the surface. To address this issue, this paper analyzed the influence of non-uniform heat flux on the flow and heat transfer characteristics of supercritical carbon dioxide by employing numerical calculation and analysis methods. Numerical simulations were sequentially performed to analyze the flow and heat transfer of supercritical carbon dioxide in a horizontal circular tube with a diameter of 4 mm under two sets of linear non-uniform heat flux conditions including the same total heat flux with different slopes, as well as different total heat fluxes with the same slope. The influence of the distribution of thermophysical properties and buoyancy lift on the convective heat transfer characteristics under linear non-uniform heat flux conditions was obtained. The simulation results show that the heat transfer of the upper wall near the entrance of the heating section under non-uniform heat flux conditions severely deteriorates compared with that under uniform heat flux. The maximum tube wall temperature under linear non-uniform heat flux is 1.66 times that under uniform heat flux. Furthermore, numerical simulations of flow and heat flux of supercritical carbon dioxide on a given actual surface heat flux boundary are carried out, and the maximum tube wall temperature at the entrance of the heating section is 3.41 times that at the outlet, while the ratio decreases to 1.50 when the flow direction changes. Relevant suggestions are offered for cooling solutions using supercritical carbon dioxide as a working medium.

Keywords: non-uniform heat flux; supercritical fluids; carbon dioxide; convective heat transfer; buoyancy lift

Received: 2022-09-14; **Accepted**: 2022-12-18; **Published Online**: 2023-01-10 15:40
URL: link.cnki.net/urlid/11.2625.V.20230109.1825.009
Foundation item: National Natural Science Foundation of China (51906009)
* **Corresponding author**. E-mail: yanchenfu@buaa.edu.cn

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2022.0783

基于本体的软件安全漏洞模式

胡璇[1, 2]，陈俊名[1, 2, *]，李海峰[3]

(1. 工业和信息化部电子第五研究所 信息安全中心，广州 511370；2. 基础软硬件性能与可靠性测试工业和信息化部重点实验室，广州 511370；3. 北京航空航天大学 可靠性与系统工程学院，北京 100191)

摘 要：针对软件缺陷与软件安全漏洞研究中存在的概念混淆问题，对DevSecOps框架下的软件安全漏洞生存期进行研究。基于软件安全漏洞生存期引入漏洞的4种情况，结合漏洞的特点提出软件安全漏洞模式定义，并采用本体方法进行表示。本体是概念化明确的规范说明，能够解决软件安全漏洞研究领域存在的二义性、不一致性、难以共享，以及由分析知识分散所导致的对人员知识和经验过度依赖的问题。以软件安全漏洞模式的分析为基础，兼顾宏观事件表现，构造漏洞分析的3层模型，即事件表示层、行为动作层和漏洞技术层。实例应用所构模型的层次结构实施渗透测试，包括安全风险分析、威胁建模、漏洞分析及渗透攻击等。实验结果表明：基于所提软件安全漏洞模式本体库的改进渗透测试具有科学性和有效性。

关 键 词：软件缺陷；软件安全漏洞；模式；本体；渗透测试
中图分类号：TP311
文献标志码：A **文章编号**：1001-5965(2024)10-3084-16

美国计算机安全专家Denning[1]于1982年提出信息安全漏洞的定义：漏洞是导致系统执行操作与访问控制矩阵定义的安全策略之间相互冲突的因素。Longley等[2]从风险管理的角度描述了漏洞的含义。Bishop和Bailey[3]采用状态空间描述法定义漏洞。美国国家标准与技术研究院及相关标准化组织也对漏洞进行了定义[4-5]。文献[6]也给出了漏洞的定义，认为漏洞是系统设计、实施或操作和管理中的缺陷或弱点，可被利用以违反系统的安全准则。分析上述定义和描述可知：漏洞是在硬件、软件、协议的具体实现或系统安全策略上存在的缺陷，使攻击者能在未授权的情况下访问或破坏系统，例如：Intel Pentium芯片中存在的逻辑错误，网络文件系统(network file system, NFS)协议中认证方式上的弱点，Unix系统管理员设置匿名文件传输协议(file transfer protocol, FTP)服务时配置不当的问题等，都可能被攻击者利用，进而威胁系统安全，因此，这些都可被认为是系统中存在的安全漏洞。

随着互联网、移动互联网、工业控制系统、物联网(internet of things, IoT)的快速发展，软件呈现种类多样化、数量规模化、形态动态化的特点[7]。信息系统规模和复杂度日益增大，受到各类网络威胁和攻击的情况愈发严重，软件安全漏洞问题也不断呈现出各类新形态。软件中存在的漏洞已成为软件行业的主要关注点[8]。尽管人们正在尽力减少软件中的安全漏洞，包括升级开发流程和工具，但漏洞的数量及其引发的计算机安全事件的数量仍然很大[8]。2017年，WannaCry勒索软件利用未修复的漏洞在全球范围内关停了30多万台电脑。Equifax Apache服务器上的一个漏洞引发了灾难性的数据泄露，导致美国一半的社会保障号码被暴露。过去

收稿日期：2022-09-15；**录用日期**：2022-12-02；**网络出版时间**：2022-12-14 10:49
网络出版地址：link.cnki.net/urlid/11.2625.V.20221212.1833.006
基金项目：国家重点研发计划(2021YFB2012205)
* **通信作者**. E-mail：47507346@qq.com

引用格式：胡璇，陈俊名，李海峰. 基于本体的软件安全漏洞模式[J]. 北京航空航天大学学报，2024，50(10)：3084-3099.
HU X，CHEN J M，LI H F. Software security vulnerability patterns based on ontology[J]. Journal of Beijing University of Aeronautics and Astronautics，2024，50(10)：3084-3099 (in Chinese).

几十年间, 信息安全领域积累了大量知识, 但其中许多概念定义模糊, 经常在不同上下文中具有不同语义, 这种歧义常常导致不同利益相关者的误读[9]。首先, 是二义性问题。不同利益相关者对应问题域的差异造成的知识鸿沟导致软件漏洞研究中对同一概念有不同理解。例如: 漏洞分类问题, 不同利益相关者的分类标准是不同的, 按利用方法分、按触发类型分还是按表现形式分等; 漏洞利用的动作行为问题, 当获得 web 后台管理脚本 (webshell) 后, 对主机自身做端口探测和数据采集, 属于横向移动 (lateral movement) 还是信息收集 (information collection), 这些概念也存在争议。其次, 是不一致性问题。系统规模和复杂性的增加使得系统开发成为多团队协作活动, 使得多视角、多范型的开发方法被广泛使用。这导致对漏洞修复方案和整体加固方案都有不同的视角和理解, 不同团队间难以达成一致; 对安全人员来说, 发现问题与提出解决方案分属不同领域, 后者涉及的知识面更广, 这也体现了视角的不一致; 在成本限制下, 漏洞修复的优先级主要依据团队负责人自身具备的知识进行确定, 这可能产生与其他参与者不同的观点。此外, 安全工具的设计和开发又需要对安全概念和技术进行系统的分类和定义[9]。因此, 有必要使用明确定义的词汇表和标准化语言以更好地描述漏洞信息和对策。此外, 尽管学界与业界不断改进软件安全漏洞的研究方法, 但仍无法摆脱当前的软件安全漏洞分析过度依赖个人知识及经验的现状[7], 且相关知识难以共享。该问题一定程度上由二义性和不一致性问题所导致。而更深层次的原因在于软件安全漏洞研究缺乏体系化的方法, 知识碎片化问题严重。有必要将软件安全漏洞分析知识从具体的分析事例中提取出来, 实现对漏洞分析知识的传承, 降低知识的分散性, 减少对个人知识及经验的依赖。因此, 亟须构建一系列与软件安全漏洞分析相关的知识实体和关联, 为分析人员提供关联查询和推理分析功能, 并为高质量漏洞分析奠定基础。

本文以软件安全漏洞为研究对象, 研究其生存期, 进而提出软件安全漏洞模式定义, 并采用本体方法对软件安全漏洞模式相关知识进行表示。本体方法能够有效解决前述软件安全漏洞研究领域二义性、不一致性、难以共享及由分析知识的分散所导致的对人员知识和经验过度依赖的问题。基于本体方法的软件安全漏洞模式是结构化和可推理的数据模型, 具有继承和关联属性, 能够展示数据、对象、元素间的关系, 实现对漏洞的精确描述和准确分类, 以及在不同分析人员间的软件安全漏洞知识共享及知识和经验的固化。以此为基础能进一步对安全人员的渗透测试等活动进行指导, 并使得渗透测试所依据的知识更为完备, 进而提高测试成功率。

1　软件安全漏洞生存期

1.1　软件缺陷和软件安全漏洞

软件缺陷一词较早出现在 1945 年的 Annals 年报中, 并被称为“bug”。目前, 软件工程领域普遍认为软件缺陷是软件的一种固有属性, 并且可通过修改软件而消除。本文认为软件缺陷指存在于软件 (文档规约、数据、源代码) 中的那些不希望或不可接受的偏差, 其结果是软件运行于某一特定条件时出现软件故障。

一般认为, 软件安全漏洞是一类软件缺陷, 可被潜在的攻击者利用, 绕过授权访问计算机系统或执行操作。有时, 授权用户也会恶意利用系统中未修补的已知漏洞。对漏洞的利用会针对机密性、完整性、可用性中的一个或多个进行。因此, 从直观上看, 软件安全漏洞是一种可被攻击者恶意利用的软件缺陷[10]。但只是大部分的漏洞来源于软件缺陷, 并不是全部, 其之间有一个很大的交集。并非源于软件缺陷的软件安全漏洞实例用 Java script 实现的伪代码所示, 如下伪代码实例所示。该段伪代码用来完成人员身份信息查询功能。当用户第 1 次通过访问前端页面的请求查询时, 后端服务器一次性将所有人员信息伴随页面源码一并返回到浏览器, 通过“胖前端”形式分担后端压力。从软件功能实现的角度看, 这段代码是没有问题的, 因此, 不存在软件缺陷。但从信息安全的角度看, 由于后端服务器一次性将全部内容返回前端页面, 通过 F12 和查看源码的形式, 任何人都能看到所有人员的相关信息。这可能造成信息泄露, 违反了机密性安全准则, 因此, 其是一个软件安全漏洞。

伪代码实例:

```
name = $(“#input - name”).val();
if name = “姓名 A”{
 return “姓名 A, 身份证号 A, 其他资料 A...”;
} else if name = “姓名 B”{
 return “姓名 B, 身份证号 B, 其他资料 B...”;
} else if name = “姓名 C”{
 return “姓名 C, 身份证号 C, 其他资料 C...”;
} .....
```

一般情况下, 软件缺陷、软件安全漏洞和软件故障间的关系模型如图 1 所示。其中, 影响系统的安全属性指影响系统的机密性、完整性及可用性中

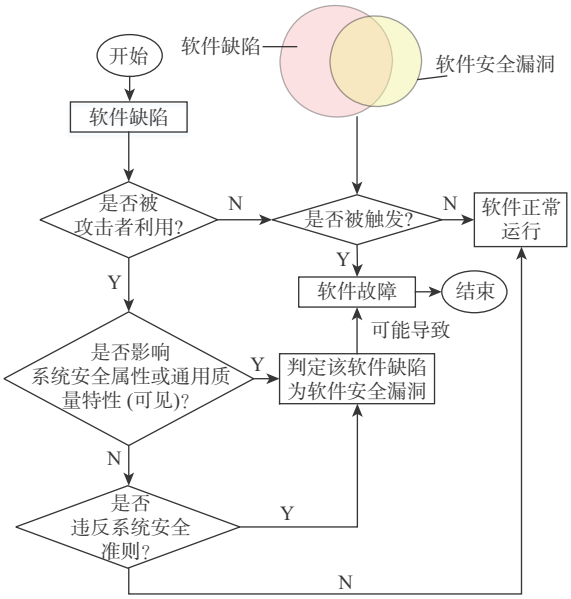


图 1 软件缺陷、软件安全漏洞和软件故障间的关系模型

Fig. 1 Relational model of software errors, software security vulnerabilities and software faults

的一个或多个；影响系统的通用质量特性指影响系统的可靠性、维修性、保障性、测试性、安全性及环境适应性中的一个或多个。

1.2 软件安全漏洞生存期模型

DevOps[11] 是软件开发生命周期从瀑布式到敏捷再到精益的演化，其生命周期始于需求定义，覆盖开发、构建、测试、部署和交付反馈等阶段[12]。2016 年 9 月，文献 [13] 提出了 DevSecOps 模型。DevSecOps 通过在 DevOps 生命周期中融入相关的安全实践，以帮助企业达成 DevSec+SecOps 的目标。因此，本文在 DevSecOps[13] 框架下对软件安全漏洞生存期进行研究。在 DevSecOps[13] 框架下，软件安全漏洞可能在过程中的任一阶段被引入，包括：需求/设计、编码、构建、测试、发布、部署、运维和监测等。

软件需求/设计阶段是引入软件安全漏洞的重要阶段之一，该阶段的软件安全漏洞属于软件开发中的早期漏洞，会在后续的实现中进行发散式传播，这将大大提高相应的修改成本，并可能造成巨大的经济损失。因此，本文以需求/概要设计阶段引入漏洞为例，描述了其生存期的全过程，如图 2 所示。图中，虚线表示上述过程是多次重复的。类似地，还可得到详细设计/编码阶段引入漏洞的生存期全过程。

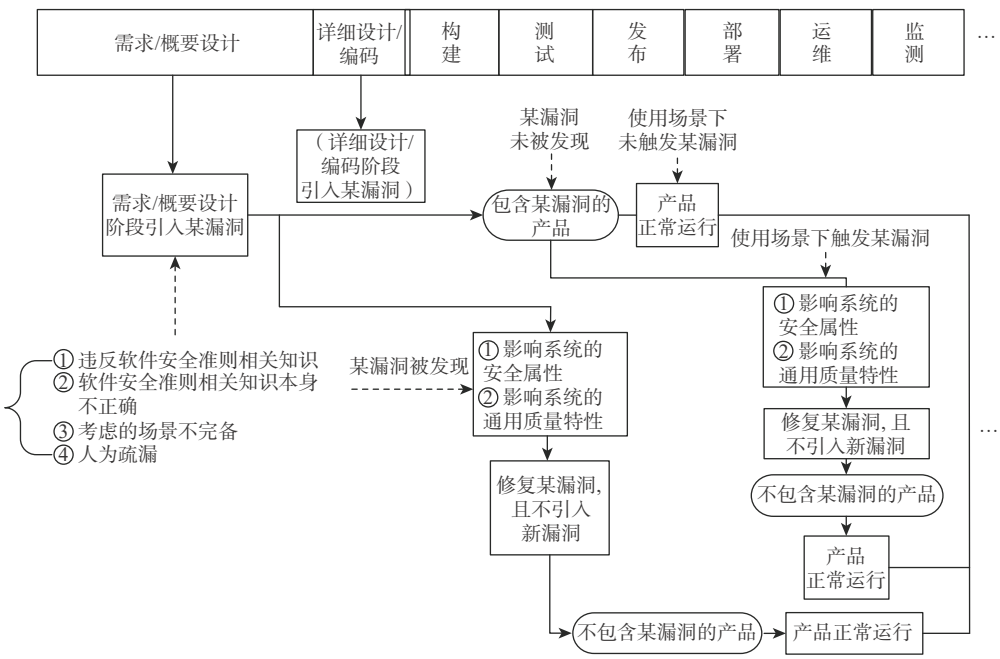


图 2 以需求/设计阶段引入漏洞为例的软件安全漏洞生存期模型

Fig. 2 Software security vulnerability lifetime model as an example of introduction of vulnerabilities in requirements/design phase

2 软件安全漏洞模式

2.1 软件安全漏洞模式定义

软件是用计算机语言表示人类知识和正确运用知识的序列[14]。对软件开发者而言，其重点关注领域包括业务逻辑实现、软件缺陷修复以满足质量要求等，而潜在攻击者重点关注领域包括系统和软件行为分析等。因此，软件作为人类知识和人类正确运用这些知识序列的产物[14]，其质量很大程度上受人的因素影响。同时，基于图 2 中引入漏洞的 4 种情况，本文认为软件安全漏洞是作为主体的人创造软件产品客体的过程中产生的与依据的软件

安全知识的偏差在软件产品客体中的反映, 表现为软件安全漏洞, 需要采用应对措施进行必需的调整。通过工程实践发现, 同一类软件安全漏洞是被反复引入软件的; 导致某类漏洞产生的语境, 即漏洞的存在环境通常是特定的; 软件漏洞可以被修正, 即存在满足正确意图的软件实现。这些特征启发研究人员借助模式[15]的概念对软件安全漏洞进行研究。

本文提出的软件安全漏洞模式 (software security vulnerability pattern, SSVP) 指产生于软件生存期各阶段内及阶段间, 在某种场景中反复出现, 可被处于非授权状态的人员或组织利用以违反系统的安全准则, 或影响系统的安全属性或通用质量特性的特定缺陷文档 (文档规约、数据、源代码) 的表现形式, 这种缺陷表现形式在特定的场景中, 具有一般性和共性, 并能够通过某种手段被修正, 可表示为 SSVP := <名称、场景、漏洞表现形式、解决方案、时间、通用漏洞评分系统 (common vulnerability scoring system, CVSS) 值、漏洞引入阶段、漏洞修复阶段>。

软件安全漏洞 (模式) 与场景 (具体环境) 和时间密切相关。其影响范围包括系统本身及其支撑软件、网络客户和服务器软件、网络路由器和安全防火墙等。在不同种类的软、硬件设备、同种设备的不同版本之间, 由不同设备构成的不同系统之间, 以及同种系统在不同的设置条件下, 都会存在各自不同的安全漏洞问题。此外, 一个系统随着测试或用户使用, 其包含的漏洞会被不断暴露, 并不断被补丁软件修补, 或在以后发布的新版系统中得以修正。而在新版系统修正了旧版本中漏洞的同时, 也会引入一些新的漏洞。随着时间推移, 旧的漏洞会不断消失, 新的漏洞会不断出现。漏洞问题也会长期存在。因而脱离具体场景和具体时间讨论漏洞问题是毫无意义的。只能针对目标系统的版本、其上运行的软件版本及服务运行设置等实际环境来具体谈论其中可能存在的漏洞及其可行的解决办法。CVSS 值可以用于比较漏洞的严重程度, 从而确定其处理优先级, 因此, 其也是刻画漏洞的一个重要指标。可见, SSVP 的核心是场景、漏洞表现形式、解决方案、时间和 CVSS 值。除了上述要素外, SSVP 还具有一些可剪裁的要素: 名称、漏洞引入阶段和漏洞修复阶段等。

漏洞表现形式可用集合形式表示为: 漏洞表现形式 = {源代码漏洞, 字节或对象代码漏洞, 设计错误}。图 3 为源代码漏洞下的若干子类, 其中, CWE 表示通用缺陷枚举。此分类参照通用缺陷/漏洞列表, 结合采集数据获得[16-17], 不同阶段漏洞模式类型可据此进行剪裁且该集合是可扩展的。漏洞引入阶段贯穿整个软件安全漏洞生存期, 漏洞发现阶段是测试和使用各阶段的集合。实例是模式的一个实际例子。上述 SSVP 各要素均是两两正交的。同时, 这些构成要素也反映了测试者、使用者 (包括攻击者) 和开发者的多个视点。

2.2　SSVP 实例

2.2.1　软件缺陷、漏洞和故障间关系模型改进

结合工程实践对图 1 进行改进, 得到改进后的软件缺陷、软件安全漏洞和软件故障间的关系模型如图 4 所示。图 1 仅以结果为导向对三者间关系进行描述, 而图 4 通过引入软件异常状态, 对软件处于不同状态的可能性进行了扩展, 使得对不同状态间关系的刻画更为完备, 也更符合工程实践的情况。其中, FMEA 表示故障模式与影响分析, VMEA 表示漏洞模式与影响分析。

软件异常情况大体可分为 2 类: 故意和过失。故意指以攻击为目的, 通过人工输入攻击性语句引发软件异常, 再由异常信息来暴露内部数据, 如报错注入。过失指不一定以攻击为目的 (但也是可被利用的), 但在软件运行过程中产生了异常, 由异常信息可进一步导致安全风险, 如 tomcat500 错误页面的程序调用堆栈信息打印 (包含若干包名、类名)。软件安全漏洞通过 SSVP 进行刻画。其中, 场景可采用漏洞利用链的形式进行描述; 解决方案从理论上可描述为: 至少将漏洞利用链中的任意一环切断, 即可将整个链条切断, 终止漏洞利用。

2.2.2　SSVP 实例表示

本文以工程实践中遇到的结构化查询语言 (structured query language, SQL) 执行和联合注入这 2 种软件安全漏洞对应漏洞模式为例给出实例。

1) SQL 执行漏洞模式。该模式通常在详细设计或编码阶段引入, 表现形式对应 CWE-676, 模式表示如下。

场景: 通过漏洞利用链表示 (见本节漏洞利用链实例)。

漏洞表现形式: 使用潜在危险的函数。

解决方案: 切断漏洞利用链 (见本节的代码实例)。

时间: 根据实际填写 (漏洞暴露时间、漏洞修复时间、软件版本号)。

CVSS 值: 高。

实例: 图 5 为 SQL 执行漏洞模式实例, 按照 SQL 语句进行查询, 可以合法获得数据库的版本号信息 “3.39.2”, 这可能会被攻击者利用来进行攻击活动。

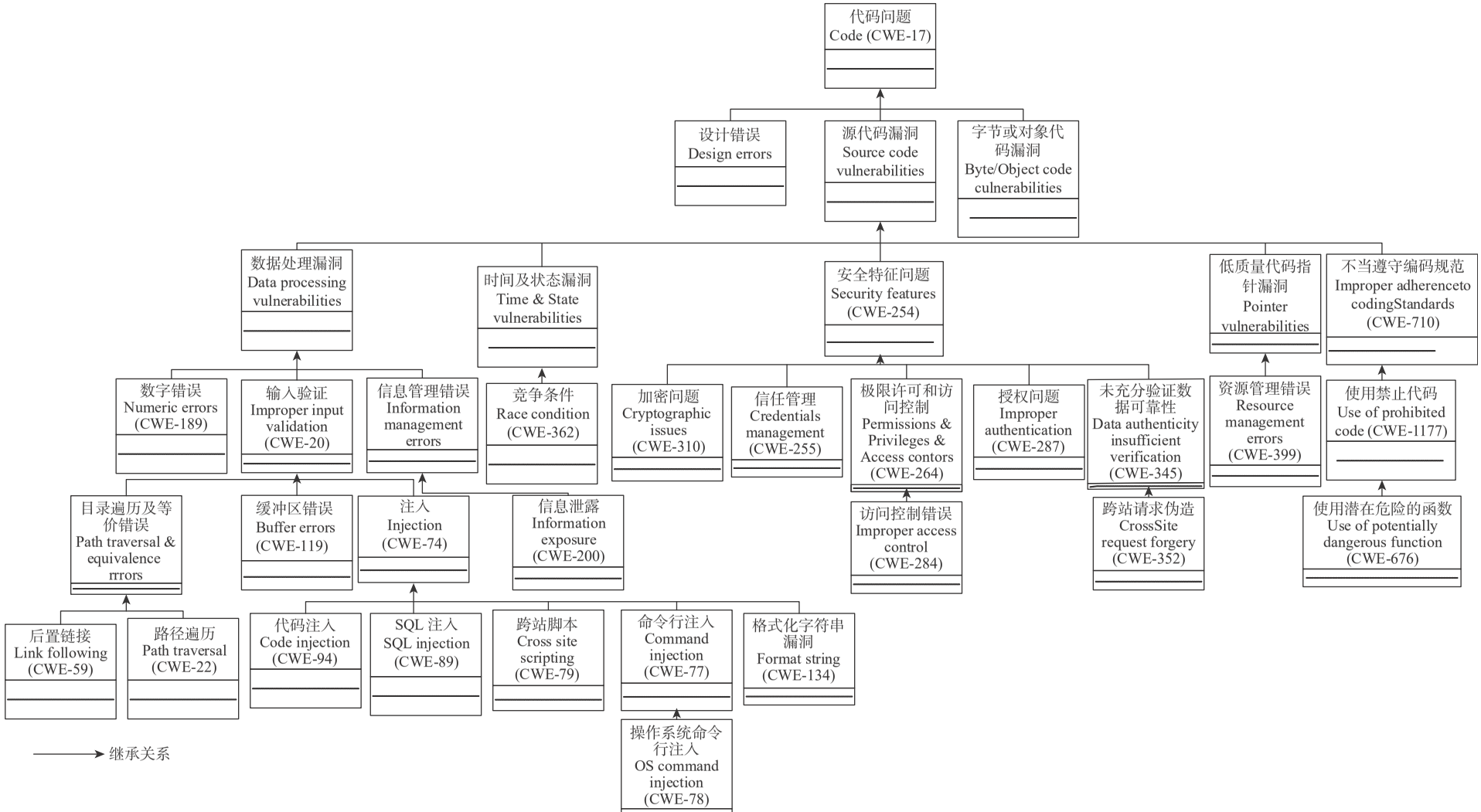


图 3　软件安全漏洞表现形式类层次结构

Fig. 3　Class hierarchy of software security vulnerabilities manifestation

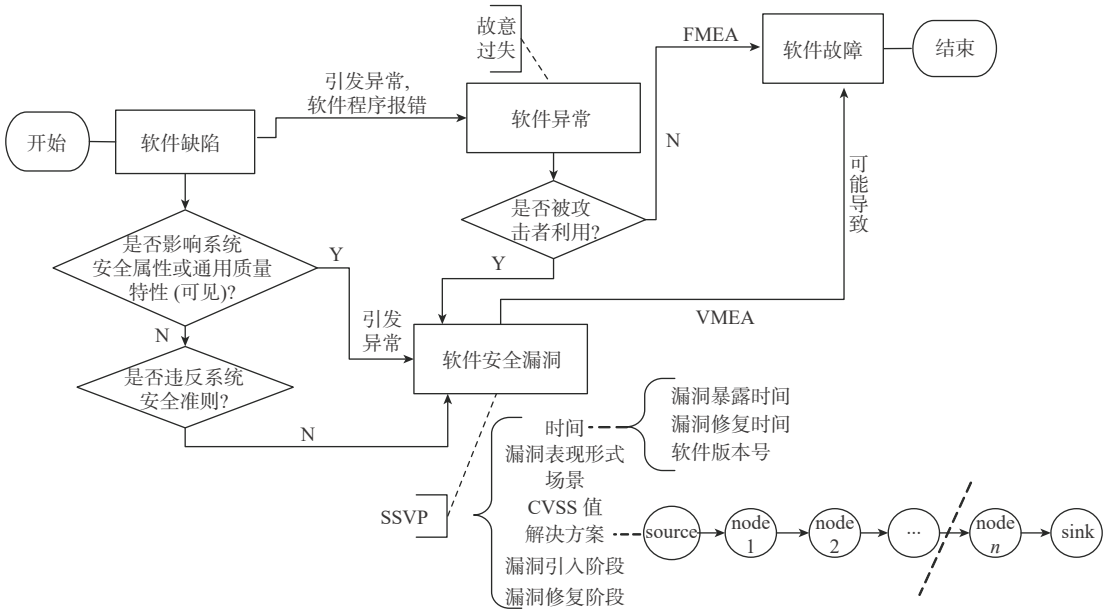


图 4　基于图 1 改进的关系模型

Fig. 4　Improved relational model based on fig.1

编程语言为 RUST 的 SQL 执行漏洞模式实例代码:

```
/**
漏洞利用链中节点的编号格式遵循以下规则:
<nodeID>-<gadgetID1[-gadgetID2-⋯-gadgetIDN-⋯]>
    链的分支序号(自顶向下)
**/
fn sql_execute(conn: &Connection) -> Result<()>{
  //SQL 执行
println!("输入要执行的 SQL 语句: ");
let mut stmt = String::new();   //CWE 的 weakness 来源
    的点(最初始的点)
stdin().read_line(&mut stmt).expect("参数类型错误");
  //攻击者进行漏洞利用的初始点
let stmt_parsed:Vec<&str> = stmt.split(" ").collect();
  //node1 //解决方案:在 node1 的位置切断漏洞利用链
if stmt_parsed[0].to_uppercase() == "SELECT" {
  //node2-1
      let mut stmt_prepared = conn.prepare(stmt.
trim())?; //node3-1
      match stmt_prepared.query([]) { //node4-1
        Ok(t) => { //node5-1-1
            println!("执行成功: ");
            let mut rows = t; //node6-1-1
              while let Some(row) = rows.next()? {
  //node7-1-1
                  let raw_row:String = row.get_
unwrap(0); //node8-1-1
                println!("{:?}", raw_row); //sink-1-1
            }
        },
        Err(err) => {//node5-1-2
            println!("执行错误: {:?}", err);   //sink-
1-2
        }
    };
} else { //node2-2
    match conn.execute(&stmt, ()) { //node3-2
        Ok(_t) => { //node4-2-1
            println!("执行成功"); //sink-2-1
        },
        Err(err) => { //node5-2-1
            println!("{:?}", err); //sink-2-2
        }
    };
}
println!("模块执行完毕!");
Ok(())
}
```

上述实例的场景, 即漏洞利用链如图 6 所示。

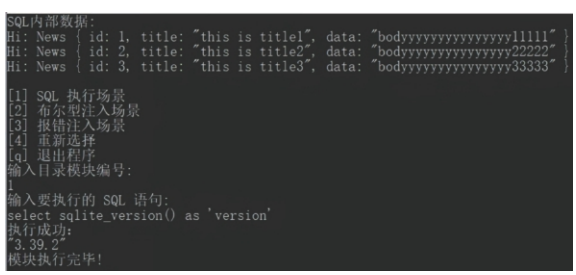


图 5　SQL 执行漏洞模式实例

Fig. 5　Instance of SQL execution vulnerability pattern

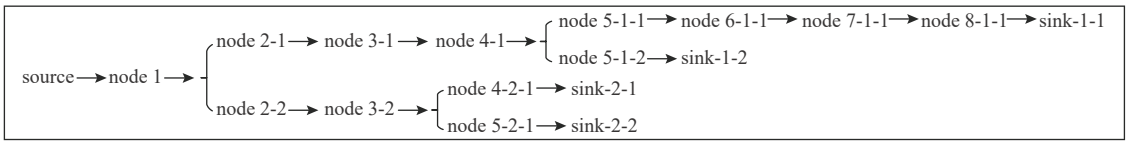


图 6 SQL 执行漏洞模式漏洞利用链

Fig. 6 Vulnerability exploitation chain of SQL execution vulnerability pattern

解决方案为切断漏洞利用链，尽管理论上在链上任一节点切断就可将其所在的整个链条切断，但由于在不同节点处切断的成本不同，因此，该实例中解决方案为在代码中 node1 处切断漏洞利用链为最优选择。

2）联合注入漏洞模式。联合注入（又称联合查询注入）漏洞模式通常在详细设计或编码阶段引入，表现形式对应 CWE-89(又称 SQL 注入，SQLInjection)，模式表示如下。

场景：通过漏洞利用链表示（见本节漏洞利用链实例）。

漏洞表现形式：SQL 注入。

解决方案：切断漏洞利用链（见本节的代码实例）。

时间：根据实际填写（漏洞暴露时间、漏洞修复时间、软件版本号）。

CVSS 值：高。

实例：图 7 为布尔注入漏洞模式实例，按照 SQL 语句进行查询，可以合法获得数据库的版本号信息“3.39.2”，这可能会被攻击者利用来进行攻击活动。

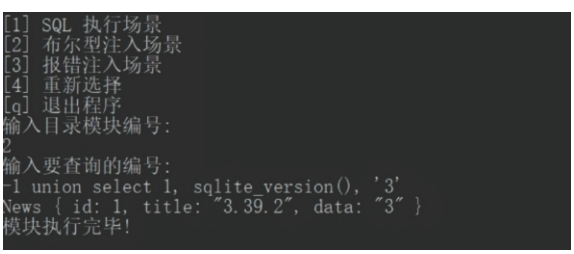


图 7 布尔注入漏洞模式实例

Fig. 7 Instance of Boolean injection vulnerability pattern

语言为 RUST 的布尔注入漏洞模式实例：

```
fn sql_boolean(conn: &Connection) -> Result<()>{
  //布尔注入
  println!("输入要查询的编号: ");
  let mut stmt = String::new(); //CWE 的 weakness 来源的点（最初始的点）
  stdin().read_line(&mut stmt).expect("参数类型错误");
    // 攻击者进行漏洞利用的初始点
  let sql = format!( "SELECT * FROM news WHERE id = {}", stmt.trim()); //node1 （拼接 SQL 语句）
  match conn.prepare(&sql) { //node2 //编译、解析上面的 SQL 语句
    // 解决方案：在 node1 或 node2 的位置切断漏洞利用链
    Ok(mut t) => {
      match t.query([]) {  //执行 SQL 语句
        Ok(mut t) => {
          match t.next() { //node3 //与上面的执行结果进行匹配
            Ok(Some(row)) => { //node4 //提取正常的执行结果
              let raw_item = News {
                id: row.get(0)?, title: row.get(1)?, data: row.get(2)?,
              }; //node5 基于提取结果构造一个正常的结构体
              println!("{:?}", raw_item); //sink 输出构造的结构体
            },
            Ok(None) => (),
            Err(_e) => (),
          }
        },
        Err(_e) => (),
      };
    },
    Err(_e) => (),
  }
  println!("模块执行完毕!");
    Ok(())
}
```

上述实例的场景，即漏洞利用链为：source→node1→node2→node3→node4→node5→sink

该实例中的解决方案为在代码中 node1 或 node2 处切断漏洞利用链。

2.2.3 软件安全漏洞分类方法的比较

目前几种主流的软件安全漏洞分类方法及本文提出的 SSVP 的特点和不足如表 1 所示，其中，CVE 表示通用漏洞和风险。由于 SSVP 包含对漏洞机理层（场景、解决方案）的分析内容，使其对漏洞分析人员具有重要的参考和指导作用，并可进一步通过对漏洞利用链的分析应用于渗透测试过程。

表 1　软件安全漏洞分类方法比较

Table 1　Comparison of software security vulnerabilities classification methods

分类方法	特点	不足
CWE	社区开发的常见软件和硬件安全弱点列表; 是弱点识别、缓解和预防工作的基准	①将软件漏洞和硬件漏洞混杂在一起, 降低了信息的有效性, 并容易造成二义性和不一致性问题; ②仅关注漏洞本身, 缺乏对整个漏洞利用链的描述和分析; ③缺乏对漏洞解决方案的描述
CVE	包含了已知漏洞和安全缺陷的标准化名称的概念字典表	①分类较为粗糙, 无法对软件安全漏洞进行精准识别和分类, 并容易造成二义性和不一致性问题; ②仅关注漏洞本身, 缺乏对整个漏洞利用链的描述和分析; ③针对漏洞的解决方案较为宏观, 缺乏漏洞机理层的分析
SSVP	对代码层面软件安全漏洞结构化的分类和描述; 组成元素的多视角(测试者、使用者(包括攻击者)、开发者), 提升了完备性; 通过将分类漏洞中的语义信息转化为度量, 能及时为开发人员提供反馈, 并能指导安全漏洞修正	①仅考虑了代码层面的软件安全漏洞, 范围较为局限; ②在未来的自动化分析过程中由于涉及到漏洞利用链分析, 可能导致路径爆炸问题, 从而影响资源占用率和性能

3　SSVP 本体表示

3.1　本体论与本体建模

本体是概念化明确的规范说明[18-19]。本体提供元素间关系、领域概念间层次、数据结构和异构信息的集成[20]。不同的本体类、关系、约束和公理定义了一个通用词汇表以共享知识[21]。

一个本体可形式化表示为如下六元组[22]:

$$O=(C,H,I,R,P,A) \tag{1}$$

式中: $C=C_{\mathrm{C}}\cup C_{\mathrm{I}}$为本体实体的集合, C_{C} 为类的集合, C_I 为实例集; $H=\{\mathrm{kind_of}(c_1,c_2)|c_1\in C_{\mathrm{C}},c_2\in C_{\mathrm{C}}\}$ 表示概念间分类关系的集合, 其定义了概念层次, 表示为 $\mathrm{kind_of}(c_1,c_2)$, 即 c_1 是 c_2 的子类; $I=\{\mathrm{is_a}(c_1,c_2)|c_1\in C_{\mathrm{I}}\wedge c_2\in C_{\mathrm{C}}\}\cup\{\mathrm{prop}_K(c_i,\mathrm{value})\ |c_i\in C_{\mathrm{I}}\}\cup\{\mathrm{rel}_K(c_1,c_2,\cdots,c_n)|\forall i,c_i\in C_{\mathrm{I}}\}$表示本体元素和实例间关系的集合; $R=\{\mathrm{rel}_K(c_1,c_2,\cdots,c_n)|\forall i,c_i\in C_{\mathrm{C}}\}$表示既不是 kind_of 又不是 is_a 关系的本体概念间关系的集合, 本体概念关系主要包含层次关系和非层次关系 2 类[23]; $P=\{\mathrm{prop}_K(c_i,\mathrm{datatype})|c_i\in C_{\mathrm{C}}\}$表示本体实体及其基本数据类型的属性集; $A=\{\mathrm{condition}_x\Rightarrow\mathrm{conclusion}_y(c_1,c_2,\cdots,c_n)|\forall j,c_j\in C_{\mathrm{C}}\}$是一组公理、规则, 其允许检查本体一致性并通过某种推理机制推断出新知识。$\mathrm{condition}_x$ 由$\mathrm{condition}_x=\{(\mathrm{cond}_1,\mathrm{cond}_2,\cdots,\mathrm{cond}_n)|\forall z,\mathrm{cond}_z\in H\cup I\cup R\}$给出。

上述本体构成元素与网络本体语言 (web ontology language, OWL) 高度兼容, 有利于使用工具进行本体编辑, 并为系统组织和推理自动化奠定基础。因此, 基于本体的 SSVP 是结构化、可计算的数据模型, 具有继承和关联属性, 能够展示数据、对象、元素间关系, 实现软件安全漏洞模式的无二义、一致和完备的描述及分类, 为其自动化推理和检测建立基础, 并在不同利益相关者间实现漏洞知识的共享和重用。

3.2　SSVP 构成要素类的表示

图 8 以类的形式给出基于经验和文献 [16] 所得 SSVP 及构成要素集 (不包含名称)。SSVP 及构成要素均可视作类, 其间为聚合关系。SSVP 类包含安全评估、指导软件开发和指导软件安全验证 3 种操作。漏洞表现形式、漏洞引入阶段、漏洞修复阶段、CVSS 值和时间类共有的操作是安全评估。这些构成要素的值即为相应类的属性。上述类的操作之一的安全评估的作用与正交缺陷分类属性的度量作用类似, 即通过分类漏洞中的语义信息并将其转化为度量, 如通过各阶段各类漏洞数量及危害程度的分布分析其变化趋势, 与预期情况的偏离都能说明存在问题。这样能及时为开发人员提供反馈, 并能指导安全漏洞修正。解决方案和漏洞表现形式类的指导软件开发操作从开发者视点提出; 而场景类的指导软件安全验证操作从测试者和使用者视点提出。

3.3　SSVP 本体构造

SSVP 本体 (SSVP ontology, SSVPO) := <Concepts, Object-property, Data-property, P^{R}, P^{C}, Inherit-hierarchies, Relations, Rules, Instances>。其中, Concepts 表示概念 (类), Object-property 表示对象属性, Data-property 表示数据属性、P^{R} 表示对属性的限制、P^{C} 表示属性特征、Inherit-hierarchies 表示概念类层次、Relations 表示关联关系、Rules 表示规则、Instances 表示实例。上述 SSVPO 的组成元素均以集合形式存在; 同时, 为叙述上的简洁, 其中一些元素后续提及时仅表示其语义上的意义。

SSVP 的概念 (类)、概念类层次 (继承关系) 及部分关联关系 (聚合关系) 如图 9 所示, 带“*”的为非终结概念类, 其余为终结概念类。SSVP 的所有构成要素与 SSVP 间为聚合关系。场景概念类与攻击链概念类间为 isDescribedBy 关系, 即前者通过后者来描述。环境概念类是静态属性概念类的一个子类, 两者间构成继承关系。可使用规则来约束信息或推导新信息, 还可用规则来表达更为丰富的概念关联, 如在 SSVPO 中将概念类环境、任务和对象

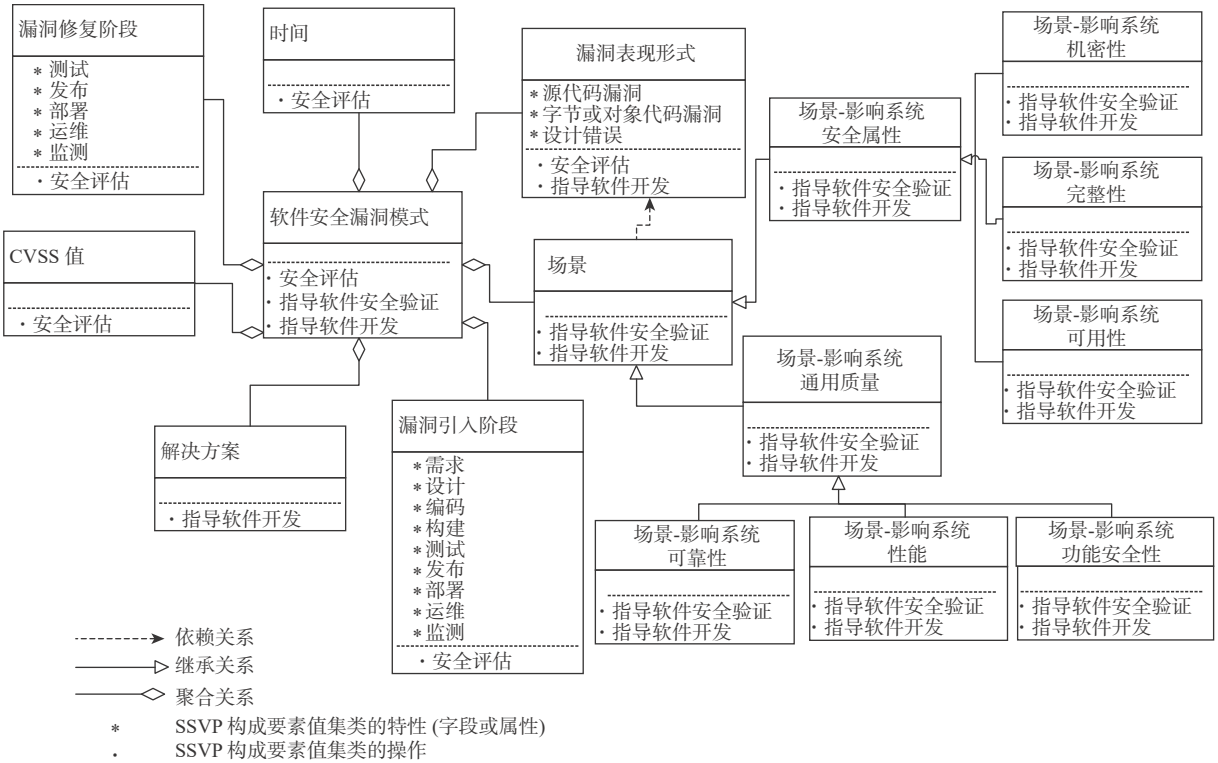


图 8 SSVP 及构成要素值集类的表示

Fig. 8 Class representation of SSVP and element value sets

定义为类互斥关系 (disjointWith)。实例是相对类而言的，通过定义类的实例，并给定属性值及限制条件，可以获得一个知识集合。Object-property 和 Data-property 是属性的 2 种类型：Object-property 是关联实例的桥梁，Data-property 是关联实例与数值的桥梁。Object-property 在内容上与关联关系相同，但两者含义不同。如前面举例的 isDescribedBy 关系，单独对场景或攻击链来说，isDescribedBy 则是 Object-property。为 CVSS 值概念类定义 hasValue，可取值为高、中、低，hasValue 即 Data-property。取值为“高”时的数值限定在 [7.0, 10.0]，这个限定即为 P^{R}。而如后续定理证明的继承关系具有的偏序性 (partial order) 即为 P^{C}。

若继承关系是一种 SSVPO 概念类层次中子类自动共享父类属性和结构的机制，那么非终结概念类下的子类与父类构成继承关系。这表明，可在一个现存概念类的基础上实现一个新类，将现存类定义的内容作为自身内容，并加入若干新内容。

定理 概念关联关系中的继承关系是偏序关系。

证明 由图 9 可知，概念类层次中的非终结概念类才具有子类，因此，在这些概念类间才具有继承关系。设 F 为一个由非终结概念类及其子类构成的非空集合，继承关系 H 是 F 上的一个关系。本节证明，继承关系满足偏序关系要求的自反性、反对称性和传递性 3 个特性。

1）$\forall a \in F, (a,a) \in H$

由面向对象基本原理可知，一个类可看作对其自身所有内容，即属性和结构的继承，因此，与其自身具有自反关系，故满足自反性。

2）$\forall a,b \in F, ((a,b) \in H) \cap ((b,a) \in H) \rightarrow (a = b)$

由继承关系定义知：$(a,b) \in H$ 说明 b 继承了 a 的所有内容；$(b,a) \in H$ 说明 a 继承了 b 的所有内容。这 2 个条件同时满足说明 a 和 b 的内容完全相同，包括属性和结构。这也说明 a 和 b 的外延与内涵均相同，故 $a = b$，即满足反对称性。

3）$\forall a,b,c \in F, ((a,b) \in H) \cap ((b,c) \in H) \rightarrow (a,c) \in H$

由继承关系定义知：$(a,b) \in H$ 说明 b 继承了 a 的所有内容；$(b,c) \in H$ 说明 c 继承了 b 的所有内容。故 c 必然继承 a 的所有内容，包括属性和结构，故 c 继承 a，即 $(a,c) \in H$，满足传递性。 **证毕**

由于 F 上的关系 H 满足上述 3 条性质，因此，称 H 为 F 上的一个偏序关系，表示为：$a \leqslant b$。带偏序关系的集合 F 称为偏序集 (partially ordered set, poset)。

在概念类层次基础上得到其他概念关联关系，进而得到 SSVPO 概念空间如图 10 所示。此即初始本体的非形式化表示。

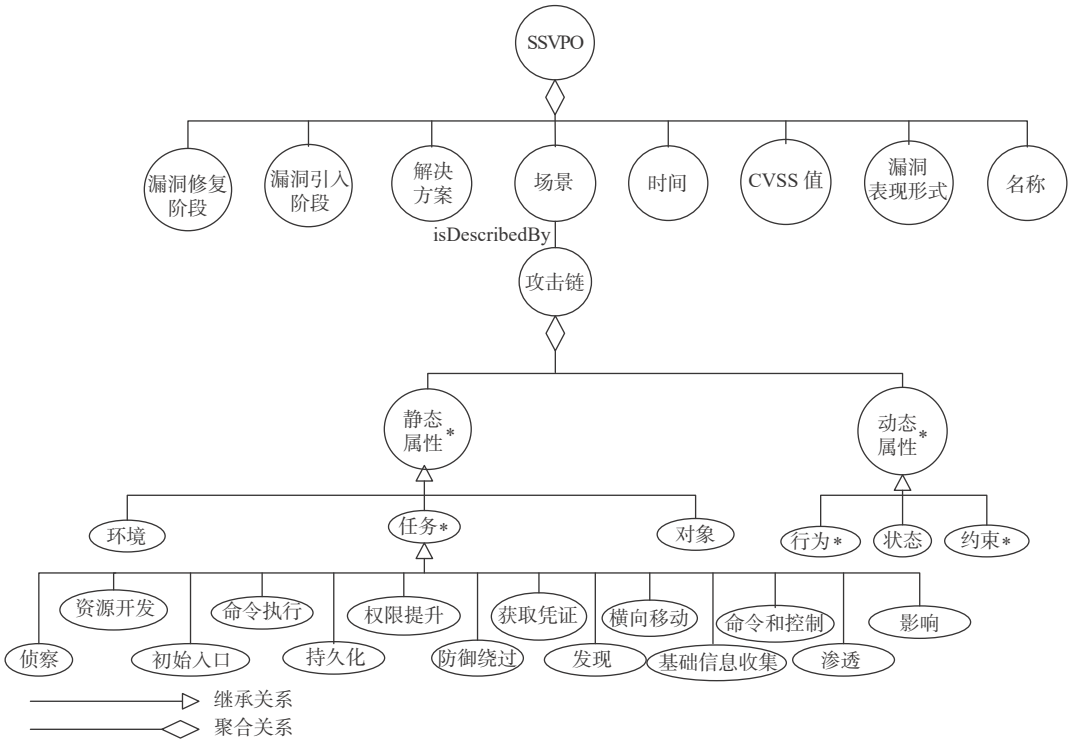


图 9　SSVPO 概念类层次

Fig. 9　Concept class hierarchy of SSVPO

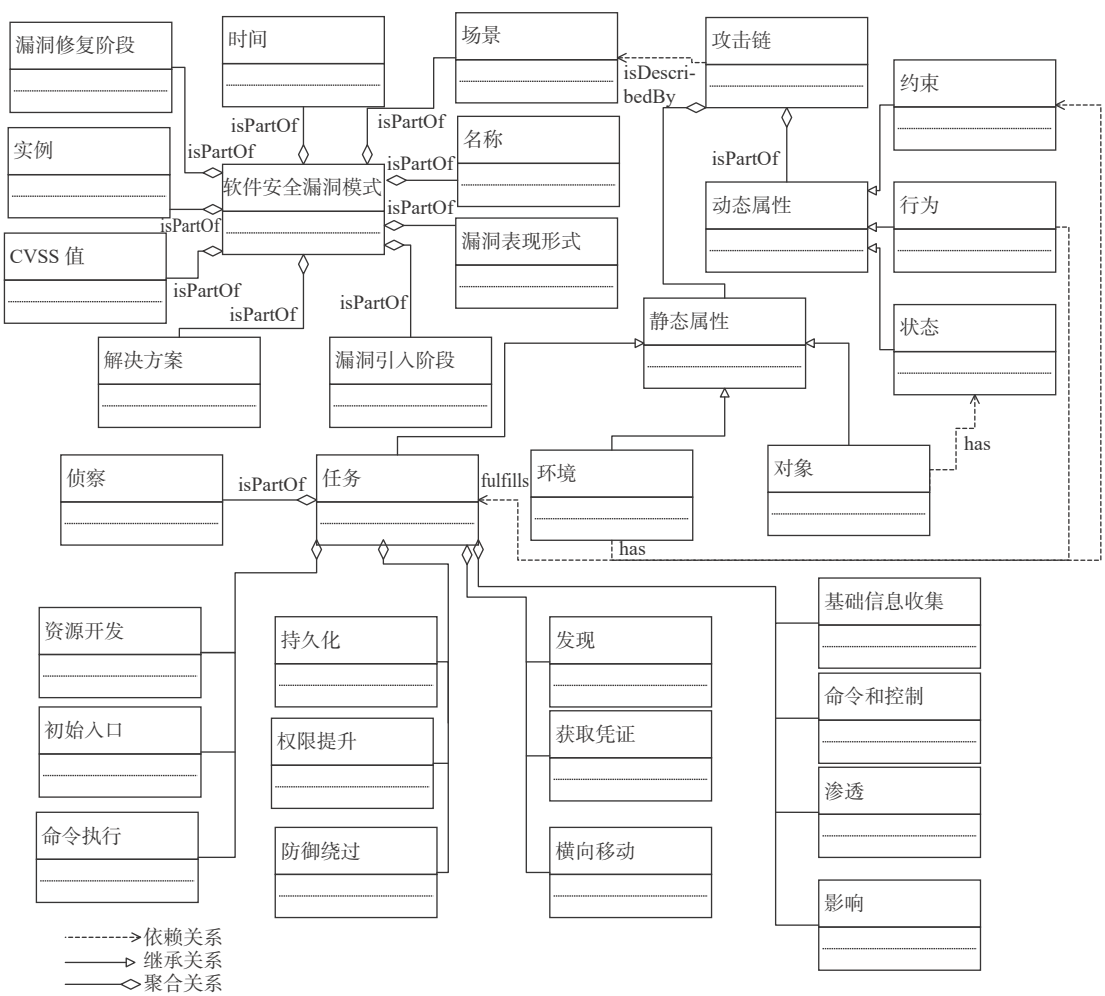


图 10　SSVPO 概念空间

Fig. 10　Concept space of SSVPO

4 实例验证

软件安全漏洞模式本体库能够对安全人员起到指导作用，这一点在渗透测试中体现得尤为明显。本节介绍相关应用，应用对象为IoT系统。

4.1 漏洞分析3层模型

以对SSVP的分析为基础，兼顾宏观事件表现，构造漏洞分析的3层模型，如图11所示，自顶向下表示该3层模型中的一个安全事件对应一条攻击链，一条攻击链对应多条漏洞利用链(一对一对多：少数特殊情况下也可以是一对多对多。即一个事件里，有多条同时并发执行的攻击链(多人)，在多个攻击链里又可交叉使用不同的链条中的漏洞)。基于该模型进行威胁建模和漏洞分析可将不同抽象层次的概念及其关联聚于一个模型中，更有利于渗透测试的实施。该模型中事件表示层里的事件链依据实际安全事件，按照攻击阶段进行分解而得。工程实践中进行安全风险分析、威胁建模、漏洞分析及渗透攻击即可按照该3层模型进行，并形成一个可追溯的流程。

4.2 改进的渗透测试过程

渗透测试通过在真实环境中模拟攻击者的攻击意图和行为来综合利用和评估系统，以帮助发现潜在的攻击链。其作为一项高度专业化的活动，要求测试人员具备多种安全知识和技能，需要通过反复实践来积累。基于本体方法可以实现对碎片化知识的系统收集和建模，并可以将结果固化以指导其他安全人员。

Nickerson等于2014年提出渗透测试执行标准[24]，将渗透测试过程划分为参与前交互、情报收集、威胁建模、漏洞分析、利用、利用后和报告。本文将软件安全漏洞模式本体库和安全风险分析本体库集成到威胁建模和漏洞分析中。改进后的渗透测试流程如图12所示。整个渗透测试过程都是面向漏洞的。其中，威胁建模和漏洞分析阶段的主要任务是确定风险类型、攻击点和攻击面，并生成尽可能多的备选攻击方案，以提高渗透测试的成功率。此外，图中的虚线箭头表示当最终攻击目标

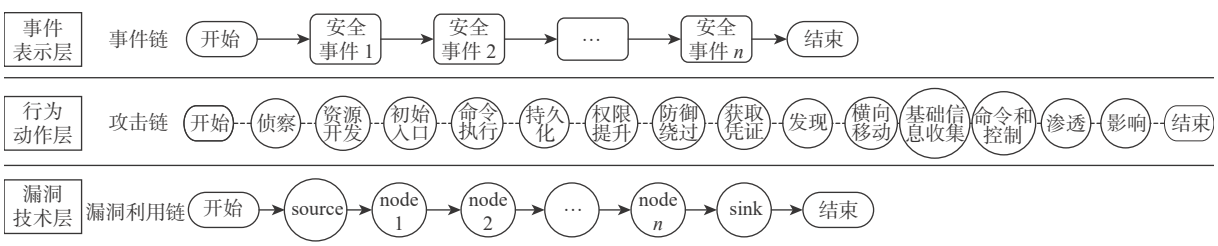


图11 漏洞分析的3层模型

Fig. 11 Three-layer model of vulnerability analysis

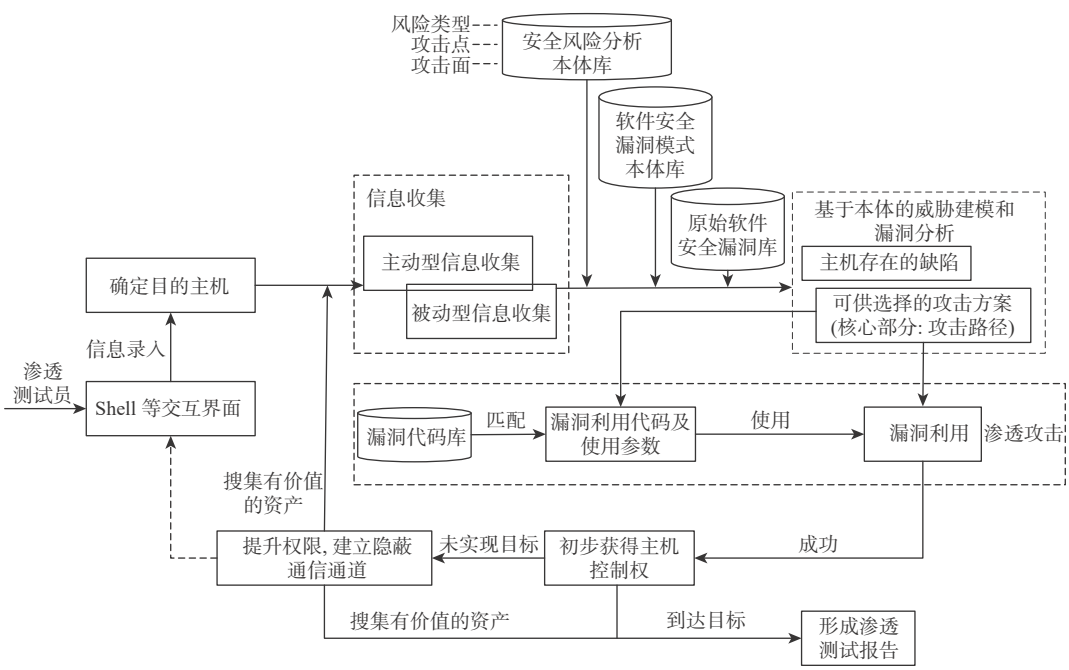


图12 改进的渗透测试过程

Fig. 12 Improved penetration testing process

尚未实现时, 需要返回目标确认阶段, 再次执行攻击过程。上述过程重复多次, 直到实现最终攻击目标。

4.3　IoT 系统渗透测试

当前, IoT 设备被越来越多地利用以导致网络安全事件频发[25]。本节以 IoT 系统为例说明渗透测试的开展。

1) 给出以“软件安全漏洞”为核心的概念和关联的本体表示, 如图 13 所示。图 12 中安全风险分析本体库和软件安全漏洞模式本体库可部分基于该图实现。

2)对 IoT 系统进行安全风险分析, 结果如表 2 所示[26-28]。其可作为安全风险分析本体库的另一来源。

3)对 IoT 系统进行基于漏洞分析 3 层模型的威胁建模 (漏洞分析) 和渗透攻击。图 14 为基于图 11 中事件链的 IoT 系统 (物联工厂 (factory of things, FoT)) 威胁建模 (漏洞分析) 和渗透攻击过程。整个区域分为 2 部分: 本地区域, 包括本地生产区和其他区域; 非本地生产区。物理连接如图 14 所示, 非本地生产区服务器和本地主服务器均存在软件安全漏洞。由于集成 Dell 远程控制卡 (integrated Dell remote access controller, iDRAC) 相应的 web 服务存在远程代码执行 (remote code execution, RCE) 漏洞 (如 CVE-2021-21571、CVE-2021-21572、CVE-2021-21573、CVE-2021-21574 等), 攻击者可以利用这些漏洞获得对服务器硬件资源的控制权, 从而控制服务器上运行的业务平台, 获取数据报告接口、密码等敏感信息, 进一步攻陷非本地生产区。继而攻击者使用类似的方法攻陷本地主服务器, 最终攻陷本地生产区。完整的事件链为: 攻陷非本地生产区→攻陷本地主服务器→攻陷本地生产区, 事件链中的任一事件都可对应一条基于对抗战术、技术和常识 (adversarial tactics, techniques and common knowledge, ATT&CK) 企业框架获得的攻击链, 攻击链由侦察、资源开发、初始入口、命令执行、持久化、权限提升、防御绕过、获取凭证、发现、横向移动、基础信息收集、命令和控制、渗透和影响组成的集合中的若干个元素组成。本节以集合中的元素侦察为例, 给出工具调用命令及参数的截图。

侦察阶段实例及证据:

从互联网收集暴露的 iDRAC 相关资产信息。这一步主要通过 T1590 收集受害者网络信息和 T1592 收集受害者主机信息 (详见 ATT&CK) 等方式从互联网上收集暴露的 iDRAC 相关资产信

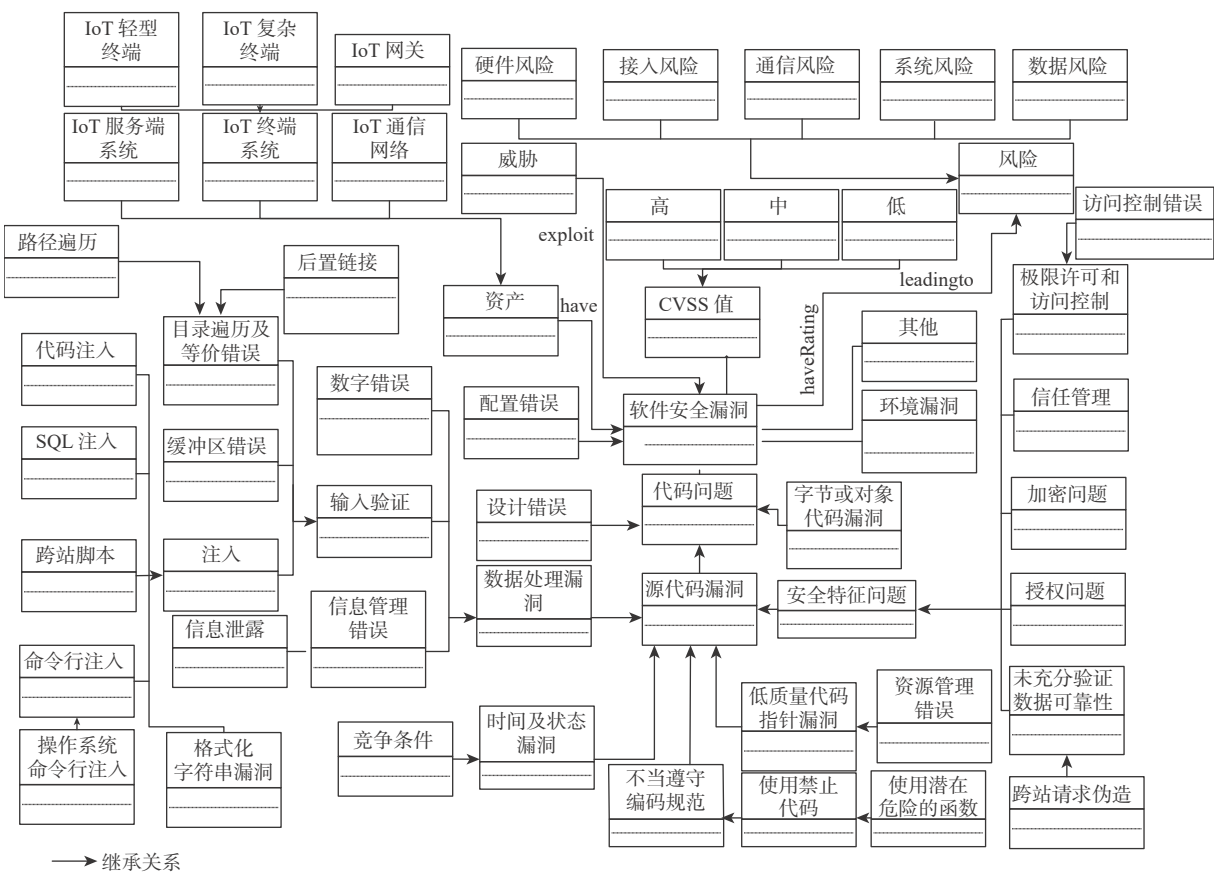


图 13　以软件安全漏洞为核心的概念和关联关系的本体表示

Fig. 13　Ontology representation of concepts and relationships with software security vulnerabilities as core

表 2 IoT 系统安全风险分析

Table 2 IoT system security risk analysis

风险类型	攻击点	攻击面
硬件风险	部署环境	攻击者进行盗窃、破坏；恶劣温、湿度环境；信号干扰或屏蔽
	供电能力	供电可靠性不足以支撑正常运行
	设备接口	存在闲置的外部接口，如联合测试工作组(joint test action group, JTAG)串口、通用异步收发器(universal asynchronous receiver/transmitter, UART)串口、晶体管-晶体管逻辑(transistor transistor logic, TTL)串口等，攻击者可以转储/重新编程闪存
接入风险	接入认证	基于媒体存取控制(media access control, MAC)或网络身份标识的身份鉴别机制失效
	访问控制	终端访问控制策略不严格，造成非授权访问或控制；绕过认证环节进行远程控制；篡改用户标识实现越权访问
通信风险	保密性	敏感信息/密钥明文传输或将敏感信息/密钥明文写在固件中；安全套接层(secure sockets layer, SSL)/传输层安全(transport layer security, TLS)不可用或配置不当；有线等效保密(wired equivalent privacy, WEP)密码破解；Wi-Fi网络安全接入(Wi-Fi protected access, WPA)/WPA2爆破；个人身份识别码(personal identification number, PIN)穷举攻击(Wi-Fi保护设置(Wi-Fi protected setup, WPS)破解)
	完整性	缺乏完整性校验，攻击者通过截取数据后，篡改重放
系统风险	身份鉴别	存在用户标识或身份冒用，身份鉴别存在弱口令或空口令
	权限控制	存在不同用户未严格限制数据的读、写、可执行权限
	固件升级	固件存在隐藏后门；远程发布恶意更新指令；固件未进行安全测试；无法验证固件合法性或真实性；固件不支持升级功能
数据风险	数据保密性	非授权用户非法取得信息导致终端本地存储的敏感或隐私信息泄露
	数据完整性	恶意篡改、破坏信息输入和传输过程导致数据源无法追溯、校验
	数据可用性	合法用户对信息和资源的使用被不当拒绝导致终端采集或监控的数据传输过程超出规定造成失效

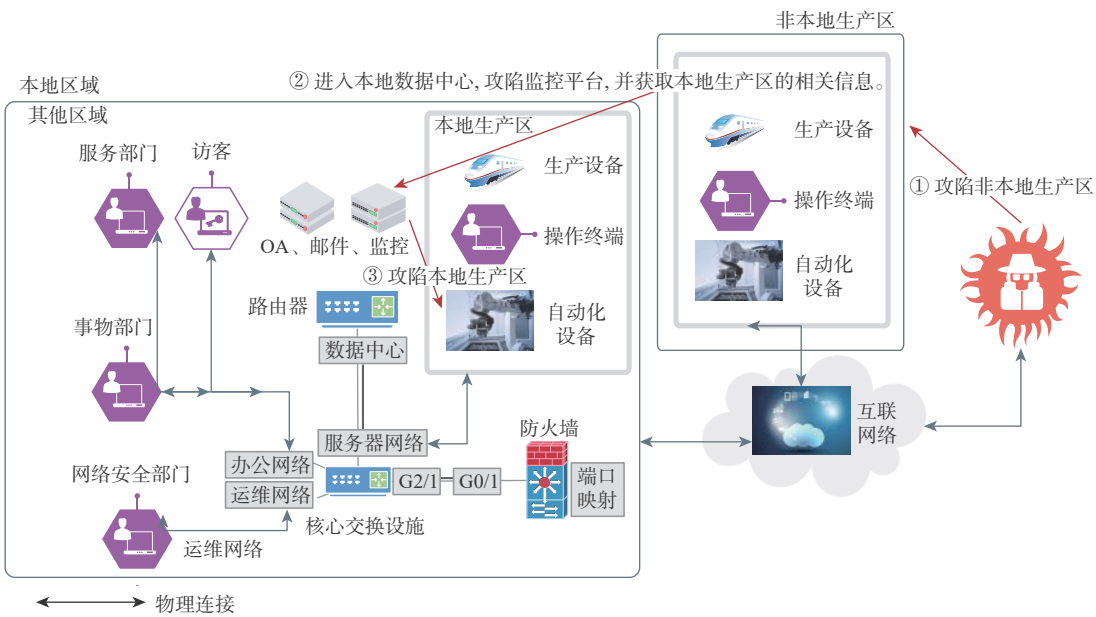


图 14 IoT 系统基于漏洞分析 3 层模型的威胁建模

Fig. 14 IoT system threat modeling based on a three-layer model of vulnerability analysis

息。图 15 为与图 14 对应的工具调用命令和参数截图。

除上述基于 3 层模型的分析，通过建立 CVE 与 CWE 的映射关系，即：CVE-2021-21571(CWE-295)、CVE-2021-21572(CWE-787、CWE-122)、CVE-2021-21573(CWE-787、CWE-121)、CVE-2021-21574(CWE-787、CWE-121)，还可向上追溯到表 2 所示的 IoT 系统安全风险分析中的内容。如 CVE-2021-21571(CWE-295) 可追踪至表 2 中的攻击点：接入认证和身份鉴别，并进一步追踪到对应的攻击面。

4)渗透测试完成后，填写渗透测试报告。渗透测试报告应当包含测试对象、测试时间、测试人，软件安全漏洞类型、软件安全漏洞描述、软件安全漏洞等级、风险评估、修复措施和结果、证明等多项要素。报告应重点对渗透测试过程中不同动作行为进行描述，并总结漏洞利用链。

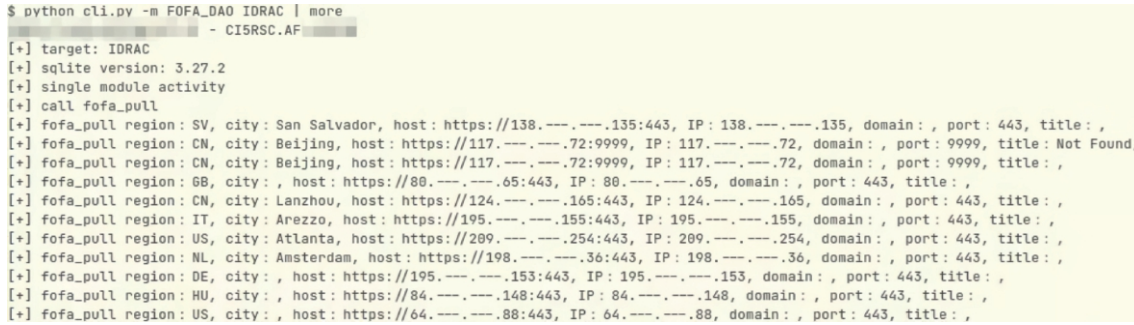


图 15　对应图 13 的工具调用命令和参数截图

Fig. 15　Screenshot of Tool call commands and parameters corresponding to Fig.13

5　结　论

1）在对 DevSecOps 框架下软件安全漏洞生存期研究的基础上，借助模式的概念对软件安全漏洞进行研究，给出 SSVP 定义。

2）采用本体方法对 SSVP 进行表示，解决了软件安全漏洞研究领域二义性、不一致性、难以共享及由分析知识的分散导致的对人员知识和经验过度依赖的问题。

3）构造漏洞分析的 3 层模型，即事件表示层、行为动作层和漏洞技术层。实例应用按照该模型的层次结构实施渗透测试，将不同抽象层次的概念及关联聚于一个模型中，形成一个可追溯的流程，有利于提高渗透测试的质量。实验结果表明，基于所提软件安全漏洞模式本体库的改进渗透测试具有科学性和有效性。

未来需要进一步研究自动化渗透测试的内容，针对漏洞利用链分析可能导致的路径爆炸问题，及对资源占用率和性能的影响问题提出解决方案。

参考文献（References）

[1] DENNING D E R. Cryptography and data security[M]. Boston: Addison-Wesley, 1982: 191-265.

[2] LONGLEY D, SHAIN M, CAELLI W. Information security: Dictionary of concepts, standards and terms[M]. New York: Macmillan, 1992: 9-23.

[3] BISHOP M, BAILEY D. A critical analysis of vulnerability taxonomies: CSE-96-11 [R]. Davis: Department of Computer Science at the University of California at Davis, 1996: 1-15.

[4] KISSEL R. Glossary of key information security terms: NIST IR 7298 Revision 2[R]. Gaithersburg: NIST, 2013: 212-213.

[5] 国家市场监督管理总局, 国家标准化管理委员会. 信息安全技术术语: GB/T25069—2010 [S]. 北京: 中国标准出版社, 2011: 42-43.
State Administration for Market Regulation, Standardization Administration of the People's Republic of China. Information security technology-glossary : GB/T25069—2010 [S]. Beijing: Standards Press of China, 2011: 42-43(in Chinese).

[6] SHIREY R. Internet security glossary, version 2: FYI, RFC 4949[R]. Virginia: Network Working Group, 2007: 333-334.

[7] 袁子牧, 肖扬, 吴炜, 等. 知识、探索与状态平面组织的软件漏洞分析架构研究[J]. 信息安全学报, 2019, 4(6): 10-33.
YUAN Z M, XIAO Y, WU W, et al. Research on the software vulnerability analysis architecture with the knowledge, exploration and state plane[J]. Journal of Cyber Security, 2019, 4(6): 10-33 (in Chinese).

[8] SHAHMEHRI N, MAMMAR A, DE OCA E M, et al. An advanced approach for modeling and detecting software vulnerabilities[J]. Information and Software Technology, 2012, 54(9): 997-1013.

[9] WANG J, GUO M M, CAMARGO J. An ontological approach to computer system security[J]. Information Security Journal: A Global Perspective, 2010, 19(2): 61-73.

[10] 慕冬亮. 软件漏洞发现、识别及诊断技术的研究[D]. 南京: 南京大学, 2019: 1-2.
MU D L. A research on vulnerability discovery, identification and diagnosis[D]. Nanjing: Nanjing University, 2019: 1-2(in Chinese) .

[11] BASS L, WEBER I, ZHU L. DevOps: A software architect's perspective[M]. New Jersey: Addison-Wesley Professional, 2015: 80-105.

[12] KIM G, HUMBLE J, DEBOIS P, 等. DevOps 实践指南[M]. 刘征, 王磊, 马博文, 等, 译. 北京: 人民邮电出版社, 2018: 4-22.
KIM G, HUMBLE J, DEBOIS P, et al. The DevOps handbook[M]. Translated by PLIU Z, WANG L, MA B W, et al, Beijing: osts&telecom Press, 2018: 4-22(in Chinese).

[13] MACDONALD N, HEAD I. DevSecOps: How to seamlessly integrate security into DevOps[EB/OL]. (2016-09-30)[2022-06-27]. https://www.gartner.com/en/documents/3463417/devsecops-how-to-seamlessly-integrate-security-into-devo.

[14] 徐仁佐. 基于软件知识的测试方法[J]. 武汉大学学报 (自然科学版), 2000, 46(1): 61-62.
XU R Z. The testing method based on software knowledge[J]. Wuhan University Journal (Natural Science Edition), 2000, 46(1): 61-62 (in Chinese).

[15] ALEXANDER C. The timeless way of building[M]. New York: Oxford University Press, 1979: 246-248.

[16] 国家市场监督管理总局, 国家标准化管理委员会. 信息安全技术网络安全漏洞分类分级指南: GB/T 30279—2020[S]. 北京: 中国标准出版社, 2020: 1-5.
State Administration for Market Regulation, Standardization Ad-

ministration of the People's Republic of China. Information security technology—Guidelines for categorization and classification of cybersecurity vulnerability: GB/T 30279—2020[S]. Beijing: Standards Press of China, 2020: 1-5(in Chinese).

[17] 张雪芹, 徐金瑜, 顾春华. 基于本体的信息安全漏洞关联分析[J]. 华东理工大学学报 (自然科学版), 2014, 40(1): 125-131.
ZHANG X Q, XU J Y, GU C H. Information security vulnerability association analysis based on ontology technology[J]. Journal of East China University of Science and Technology (Natural Science Edition), 2014, 40(1): 125-131 (in Chinese).

[18] GRUBER T R. A translation approach to portable ontology specifications[J]. Knowledge Acquisition, 1993, 5(2): 199-220.

[19] GRUBER T R. Toward principles for the design of ontologies used for knowledge sharing[J]. International Journal of Human-Computer Studies, 1995, 43(5-6): 907-928.

[20] KUREYCHIK V M, SAFRONENKOVA I B. Ontology-based approach to design problem formalization[C]//Proceedings of the International Seminar on Electron Devices Design and Production. Piscataway: IEEE Press, 2019: 1-5.

[21] 李霖, 王红. 基于形式化本体的基础地理信息分类[J]. 武汉大学学报 (信息科学版), 2006, 31(6): 523-526.
LI L, WANG H. Classification of fundamental geographic information based on formal ontology[J]. Geomatics and Information Science of Wuhan University, 2006, 31(6): 523-526 (in Chinese).

[22] HU X, LIU J. Ontology construction and evaluation of UAV FCMS software requirement elicitation considering geographic environment factors[J]. IEEE Access, 2020, 8: 106165-106182.

[23] HAMDULLA A, YILAHUN H, ABDURAHMAN K, et al. A hierarchical clustering based relation extraction method for domain ontology[C]//Proceedings of the 9th International Symposium on Parallel Architectures, Algorithms and Programming. Piscataway: IEEE Press, 2018: 36-40.

[24] NICKERSON C, KENNED D, SMITH E, et al. Penetration testing execution standard[EB/OL]. (2014-08-16)[2024-06-27]. http://www.pentest-standard.org/index.php/Main_Page.

[25] MORAIS S V, DE SÁ A O, RUST L F, et al. Malicious traffic description: Toward a data model for mitigating security threats to home IoT[J]. IEEE Communications Standards Magazine, 2021, 5(3): 48-55.

[26] 孙跃, 司冠林, 徐小天, 等. 电力物联网终端设备渗透测试技术研究[C]//全国智能用电工程建设经验交流会论文集. 湛江: 中国会议, 2020: 19-22.
SUN Y, SI G L, XU X T, et al. Research on penetration testing technology of power IoT terminal equipment[C]//Proceedings of the National Intelligent Electricity Engineering Construction Experience Exchange Conference. Zhanjiang: China Conference, 2020: 19-22(in Chinese).

[27] 中国信息通信研究院. 物联网安全态势综述[J]. 保密科学技术, 2018(9): 12-20.
China Academy of Information and Communications Technology. Overview of internet of things security situation[J]. Secrecy Science and Technology, 2018(9): 12-20(in Chinese).

[28] 王海峰, 李朝阳, 吕政权, 等. 泛在电力物联网环境下网络安全攻击研究[J]. 浙江电力, 2019, 38(12): 76-81.
WANG H F, LI Z Y, LYU Z Q, et al. Survey on cyber attack in the context of ubiquitous power Internet of Things[J]. Zhejiang Electric Power, 2019, 38(12): 76-81 (in Chinese).

Software security vulnerability patterns based on ontology

HU Xuan[1, 2], CHEN Junming[1, 2, *], LI Haifeng[3]

(1. Information Security Center, The Fifth Research Institute of Electronics, Ministry of Industry and Information Technology, Guangzhou 511370, China;
2. The Ministry of Industry and Information Technology Key Laboratory of Performance and Reliability Testing and Evaluation for Basic Software and Hardware, Guangzhou 511370, China;
3. School of Reliability and Systems Engineering, Beihang University, Beijing 100191, China)

Abstract: This paper studies the lifetime of software security vulnerabilities under the DevSecOps framework aiming at the conceptual confusion problem of research on software errors and software security vulnerabilities. This work provides a definition of software security vulnerability pattern together with vulnerability characteristics, and uses ontology to represent it. It is based on four scenarios of introducing vulnerabilities in the life cycle of software security vulnerabilities. An ontology is an explicit specification of a conceptualization, which can solve the problems of ambiguity, inconsistency, difficulty in sharing, and excessive dependence on personnel knowledge and experience caused by the dispersion of analysis knowledge in the field of software security vulnerability research. A three-layer model for vulnerability analysis is built, comprising the event representation layer, behavior action layer, and vulnerability technology layer, based on the study of software security vulnerability patterns and accounting for the macro event performance. The example application implements penetration testing according to the hierarchical structure of the bulit model, including security risk analysis, threat modeling, vulnerability analysis, and penetration attacks. The experimental results show that the improved penetration testing method based on the software security vulnerability pattern ontology library proposed in this paper is scientific and effective.

Keywords: software error; software security vulnerabilities; pattern; ontology; penetration testing

Received: 2022-09-15; **Accepted**: 2022-12-02; **Published Online**: 2022-12-14 10:49

URL: link.cnki.net/urlid/11.2625.V.20221212.1833.006

Foundation items: National Key Research and Development Program of China (2021YFB2012205)

* **Corresponding author**. E-mail: 47507346@qq.com

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2022.0811

基于视觉 Transformer 飞行员姿态估计

吴红兰*，刘豪，孙有朝
(南京航空航天大学 民航学院，南京 211106)

摘 要：人体姿态估计是行为感知领域中的一个重要环节，也是民用飞机驾驶舱智能交互方式的一项关键技术。为建立民用飞机驾驶舱复杂光照环境与飞行员姿态估计模型性能的可解释联系，提出基于视觉 Transformer 飞行员姿态 (ViTPPose) 估计模型，该模型在卷积神经网络（CNN）主干网络末端使用包含多层编码层的双支路 Transformer 模块，编码层联合 Transformer 和空洞卷积，在增大感受野的同时捕捉后期高阶特征的全局相关性。基于飞行机组标准操作程序，建立飞行模拟场景下的飞行员操纵行为关键点检测数据集，ViTPPose 估计模型在此数据集上完成飞行员坐姿估计，并通过与基准模型对比，验证了其有效性。在驾驶舱复杂光照的背景下，构建坐姿估计热图，分析 ViTPPose 估计模型对光照强度的偏好，测试其在不同光照等级下的性能，揭示其对不同光照强度的依赖关系。

关 键 词：民用飞机；智能驾驶舱；飞行员姿态估计；卷积神经网络；自注意力；可解释性
中图分类号：V249.1; TB553
文献标志码：A **文章编号**：1001-5965(2024)10-3100-11

人工智能等战略新兴技术正逐步与传统行业交融[1-2]，其在民用航空领域的应用逐渐成为民用飞机技术变革与创新的驱动力。智能化民机飞行专注于重新构建飞行过程中的人机交互模式，其成为行业中突破性变革的方向[3]。研究飞行过程中飞行员的操纵行为，对飞行操纵人员进行实时与准确的姿态估计、动作捕捉有很多重要的应用。其一方面可以实现驾驶舱内飞行员智能控制飞机某些功能，即人-机交互；另一方面可以实时识别飞行员行为，提供决策信息、减少人为操纵差错，实现机-人反馈。民用飞机飞行员姿态估计是体感交互的基本任务之一，然而由于遮挡、复杂光照环境和个体差异等因素，使得姿态估计具有挑战性。

在民机驾驶舱内进行飞行员姿态估计，对于现有的卷积神经网络（convolutional neural networks, CNN）模型，存在以下问题使其难以提供辅助信息进行飞行决策：①高度非线性关系促使神经网络中的每层卷积单元与激活函数高度耦合，将输出结果解耦并进行可视化是困难的[4-5]。②基于 CNN 的模型往往具有深层的结构。然而随着模型深度的增加，会使模型面临复杂的长期依赖问题，进一步阻碍了对各层功能的解释[6]。③由于 CNN 感受野大小有限及卷积层的固有性质，使其在捕获全局相关性方面存在局限性。这种局限性会导致复杂卷积结构在捕获高阶特征信息方面性能下降[7-8]。因此，在驾驶舱复杂光照环境背景下，CNN 很难捕获具体图像的依赖特征。虽然已经有一些相关的模型可视化的技术[9-10]，但其中大多数侧重于揭示影响被识别物体的整体结构信息，而非探究影响局部关键点识别的背景光照信息，如何针对驾驶舱具体的环境开发姿势估计模型，进一步辅助人们进行决策需要更深入的研究。

随着 CNN 的发展，二维人体姿态估计任务取得了稳定的进展，模型也变得实用[11-12]。越来越多

收稿日期：2022-09-29；**录用日期**：2023-03-26；**网络出版时间**：2023-04-20 16:26
网络出版地址：link.cnki.net/urlid/11.2625.V.20230420.1442.002
基金项目：国家自然科学基金-民航联合研究基金 (U2033202,U1333119)；国家自然科学基金 (52172387)
* **通信作者**. E-mail：wuhonglan@nuaa.edu.cn

引用格式：吴红兰，刘豪，孙有朝. 基于视觉 Transformer 飞行员姿态估计 [J]. 北京航空航天大学学报，2024，50（10）：3100-3110.
WU H L，LIU H，SUN Y C. Vision Transformer-based pilot pose estimation[J]. Journal of Beijing University of Aeronautics and Astronautics，2024，50（10）：3100-3110（in Chinese）.

先进的方法在诸如 MPII[13]、LSP[14] 及 MS COCO[15] 等关键点检测数据集测试中显示出良好性能。由于民机飞行员特殊的坐姿方式, 导致不能使用人体姿态估计领域主流数据集进行飞行员姿态估计任务。为克服数据集缺乏的问题, 在不同的飞行场景下, 采集飞行员操纵飞机时不同人体姿态图片, 进行人工标注, 建立飞行员关键点检测数据集。

视觉 Transformer(vision Transformer, ViT)[16] 在 ImageNet[17] 分类任务中显示出良好的性能。受 ViT 结构在视觉任务中的成功启发, 越来越多的学者使用 Transformer 改善 CNN 感受野局限性问题。一些后续研究通过知识蒸馏[18], 采用层次更复杂的体系结构[19] 等来提高任务精度。考虑到 CNN 在模型后期捕捉特征全局相关性方面的局限性, Transformer 架构在绘制成对交互方面具有优于 CNN 的天然优势[20], 本文提出一种 CNN 联合 Transformer 的模型—基于视觉 Transformer 飞行员姿态(visual Transformer-based pilot pose, ViTPPose)估计模型。鉴于 ViTPPose 估计模型前期使用卷积有利于提升性能[21], ViTPPose 的体系结构在模型前期使用 CNN 提取图像特征, 在模型后期使用双支路 Transformer 模块捕获全局相关信息。双支路模块中一条分支通过重塑操作将特征图展平为二维序列信息, 并将其作为 Transformer 的输入。在 Transfomer 中进行全局自注意力信息交互; 另一条分支使用空洞卷积来提取更大感受野的信息。最后, 2 条支路信息累加并将其输出序列通过形变操作恢复成图片结构, 并进行热图预测。

为探索具体的图像线索对模型关键点位置预测的显著作用, 本文进一步分析不同实验变量对模型的性能影响。本文把模型的最后输出层作为一个激活器, 通过置信概率分数形成热图中的关键点最大位置, 通过设置不同等级的光照强度, 以及对热图的置信概率进行可视化展示, 建立模型预测结果与驾驶舱复杂光照的联系, 改善驾驶舱复杂光照环境给姿态估计带来的问题。

1　相关的研究进展

1.1　人体姿态估计

二维人体姿态估计分为自上而下和自下而上两种方法。自上向下的姿态估计方法[4,22-23] 将人体节点检测过程解释为 2 个阶段, 即从图像中检测所有人体, 然后进行单人姿态估计, 该方法通常具有较高的识别准确率; 自下而上的方法[5,24-25] 直接预测所有人体关节点, 进一步将全部关节点分别组合成每个人的姿态, 该方法通常具有较高的推理速度。为提取特征的空间和纹理信息, 卷积姿态机(convolutional pose machine, CPM)[26] 使用顺序化体系结构。为使网络结构能够感知全局上下文特征信息, Hourglass[22] 使用沙漏类型体系结构。CPN(cascaded pyramid network)[23] 使用 Mask-RCNN[12] 的部分结构进行人体检测, 进一步实现关键点检测。Higher-HRNet[5] 前期遵循 HRNet 网络的设计模式, 在末端进一步使用反卷积的方法提升特征分辨率。

飞行员姿态估计任务为人体姿态估计任务中的一种, 是驾驶舱飞行员行为识别的基础任务。如图 1 所示, 飞行员坐姿估计与手部关键点检测相结合, 摄像头实时捕获飞行操纵行为。鉴于飞行员手部与正常人体手部较为统一, 手部节点识别使用正常人体手部数据集, 本文着重于检测飞行员坐姿关键节点。在飞行过程中, 考虑到飞行员姿态识别高准确率的需求, 本文模型采用自上而下的方法, 提升模型识别准确率。

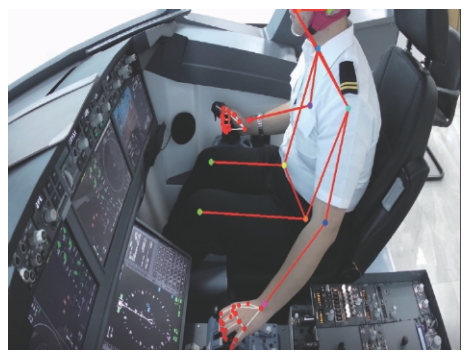


图 1　摄像头实时捕捉飞行员操纵行为

Fig. 1　Capturing pilot action in real time through camera

1.2　视觉 Transformer

Transformer 在自然语言处理(natural language processing, NLP)领域中的成功应用, 促使一些研究使用 Transformer 解决计算机视觉问题。Dosovitskiy 等[16] 用 ViT 证明了使用自注意结构进行图像特征提取的有效性, 以及在多个图像分类基准上获得优越的性能表现。Swin Transformer[27] 吸收了 ViT 优秀的设计思想, 采取滑动窗口的方式将自注意结构设计成扩大感受野的工具, 取得了具有竞争力的目标检测性能。此外, 部分学者使用 Transformer 改善更多的视觉任务, 如人体姿态估计任务、行人重识别任务[28]、图像分割任务。HRFormer[29] 通过多分辨率并行变换模块将高分辨率表示引入 ViT, 从而适应姿态估计任务。一些研究将 Transformer 与 CNN 结合起来用于人体关键点检测。Swin-Pose[30] 通过使用预训练的 Swin Transformer[27] 作为主干来提取特征, 从而实现了优异的性能。TransPose[31] 将注意力模块放在 CNN 末端, 以有效捕获长期依赖关系。

与以往的研究不同，ViTPPose估计模型在整体结构上联合CNN与Transformer的同时，进一步在卷积网路的末端使用双支路模块去联合Transformer与空洞卷积，在捕捉全局相关性的同时增大感受野信息。ViTPPose估计模型为CNN与Transformer的结合提供了一种高效的方法。

1.3 可解释性研究

神经网络的可解释性方法主要用于探究模型推理的机理，从而帮助人们更好地理解模型如何实施预测任务。Samek等[32]从端到端的角度，定义可解释性的目标是确定哪些输入与预测最相关。为判断哪些样本对模型训练结果产生何种影响，通常采取去除某些样本后重新训练的方式。Koh和Liang[33]提出基于影响函数的方法，在无须对模型重新训练的情况下，通过剔除某个输入样本来评估其对模型输出结果的影响。文献[34-35]通过可视化神经网络内部结构来揭示不同卷积层之间的相关联系。DeConvNet[34]被用于生成特征活动，以显示卷积层所学到的知识。文献[35]通过选择特定的神经元或通道来可视化特征结构图。还有一些研究，如分类激活映射(class activation mapping, CAM)[36]从神经网络中定位某个特定类别在输入图像上的显著性区域，从而实现弱监督的目标定位。为突出分类结果的作用区域，Grad-CAM[13]、Score-CAM[37]和Group-CAM[14]分别在CAM的基础上进行不同策略的改进。

与以往研究不同，本文更专注于揭示有助于在民机驾驶舱复杂照背景下关键点定位的线索，为民机飞行员姿态估计预测结果提供确切的证据。

2 飞行员姿态估计模型

本文ViTPPose估计模型由3个部分组成，如图2所示。HRNet-S-W32主干网络，用于提取早期图像特征；双支路Transformer模块(包含多层编码层)，用于捕获网络后期特征向量之间的空间交互信息；关键点热图，用于输出预测结果，其中，MutilHead Self-Attention为多头自注意力，FFN表示前向反馈网络。

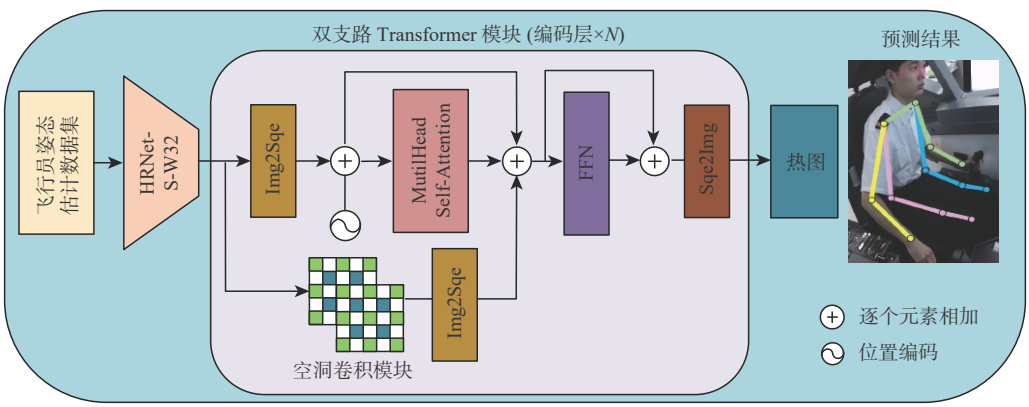


图2 ViTPPose估计模型姿态估计网络结构

Fig. 2 Structure of ViTPPose estimation model pose estimation network

2.1 HRNet-S-W32主干网络

首先，出于安全飞行的考虑，本文倾向于提升ViTPPose估计模型的识别性能；其次，鉴于民机驾驶舱狭小的空间，本文确保ViTPPose估计模型识别任务达到一定准确率的同时，进一步考虑其轻量化设计。由于HRNet[4]的高分辨率设计模式在解决人体姿态关键点检测中，展示出良好的性能。本文遵循HRNet[4]的设计模式，采用轻量化卷积主干HRNet-S-W32作为前期特征提取网络。如图3所示，模型包含4个阶段，分别为Stage 1、Stage 2、Stage 3和Stage 4。开始通过2个卷积核为3×3、步长为2的卷积模块，将分辨率降低到1/4。相较于原始HRNet[4]网络，为平衡ViTPPose估计模型的准确性和参数量，HRNet-S-W32的Stage 2、Stage 3和Stage 4分别包含1个、4个和2个并行分支。从第1个高分辨率分支开始逐渐并行添加一个由高到低分辨率的分支，新分支为当前分支中分辨率的1/2。各个分支中有4个3×3卷积核的残差卷积模块。

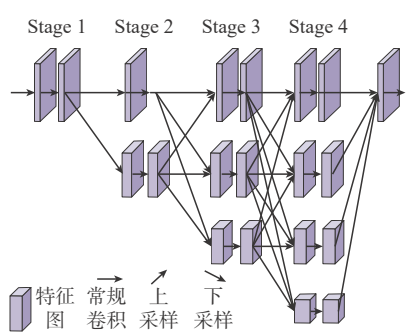


图3 CNN主干结构

Fig. 3 Structure of CNN backbone

跨分辨率特征通过多尺度融合模块重复进行信息交换以实现信息融合。

2.2　双支路 Transformer 模块

与局部卷积相比，全局特征信息交互是自注意机制的关键优势。如图 2 所示，本文使用联合 Transformer 与空洞卷积的双支路架构。为在网络结构后期捕捉全局相关信息，第 1 条支路将 Transformer 模块放置在 CNN 主干网络的末端进行全局信息编码。图片输入 $\boldsymbol{N}\in\mathbf{R}^{3\times H\times W}$（$H$ 为特征高度，W 为特征宽度）经过 CNN 主干网络得到输出 $\boldsymbol{I}\in\mathbf{R}^{C\times H\times W}$，$C$ 为通道数。对于 Transformer 的输入部分，经过实验测试得出，模型维度的增加带来识别准确率的提升。考虑到 ViTPPose 估计模型的准确率与计算量，先将 CNN 主干网络输出特征 $\boldsymbol{I}$ 进行 1×1 卷积激活并进行升维如下：

$$\boldsymbol{M}=\sigma(F_1(\boldsymbol{I})) \tag{1}$$

式中：$\boldsymbol{M}\in\mathbf{R}^{D\times H\times W}$，$D$ 为通道维度；$F_1(\cdot)$ 为 1×1 卷积变换操作；σ 为 Relu 函数。

升维后的特征 $\boldsymbol{M}$ 经过重塑操作，线性映射为二维序列 $\boldsymbol{X}\in\mathbf{R}^{L\times D}$，$L=H\times W$，如下：

$$\boldsymbol{X}=\text{Img2Sqe}(\boldsymbol{M}) \tag{2}$$

式中：Img2Sqe(·) 表示图像转换序列操作[16]。

如图 2 所示，为保留每个卷积单元的空间信息，将待训练位置信息[16] $\boldsymbol{P}\in\mathbf{R}^{L\times D}$ 嵌入添加到二维序列 $\boldsymbol{X}$ 中。

如图 4 所示，为从序列中捕捉到成对的依赖关系，序列 $\boldsymbol{X}$ 进行多头注意力学习，其中，Concat表示拼接操作，Attention 表示注意力操作，Scale Dot Product 表示逐点点积操作。对于给定的输入序列 $\boldsymbol{X}$，经过线性映射为 Query（$\boldsymbol{Q}\in\mathbf{R}^{L\times D}$）、$\boldsymbol{V}\in\mathbf{R}^{L\times D}$ 和 Key（$\boldsymbol{K}\in\mathbf{R}^{L\times D}$），可得到单头自注意力值：

$$\text{Attention}(\boldsymbol{Q},\boldsymbol{K},\boldsymbol{V})=\text{Softmax}(\boldsymbol{Q}\boldsymbol{K}^{\mathrm{T}}/\sqrt{D})\boldsymbol{V} \tag{3}$$

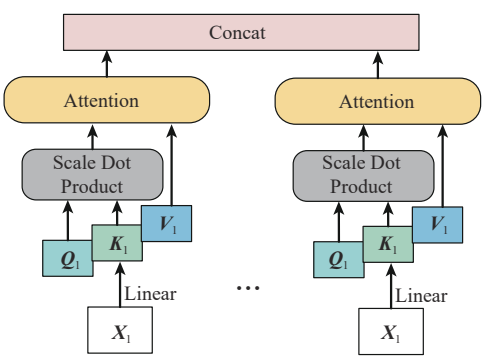


图 4　多头自注意机制结构

Fig. 4　Structure of multi-head self-attention mechanism

为得到多头注意力结果，把每个单头自注意力头的输出拼接起来，乘以一个权重矩阵 $\boldsymbol{W}$，得到多头自注意力输出结果：

$$\text{MultiheadAttention}(\boldsymbol{Q},\boldsymbol{K},\boldsymbol{V})=\text{Concat}(h_1,h_2,\cdots,h_h)\boldsymbol{W} \tag{4}$$

为捕捉到更多感受野的信息，ViTPPose 估计模型的第 2 条支路经过一个空洞率为 2 的空洞卷积模块。其中，具体空洞卷积模块信息如图 5 所示，该模块包含 2 个 3×3 残差卷积单元，其中，DiConv 为空洞卷积操作，BN 为批标准化操作，Relu 为激活函数。

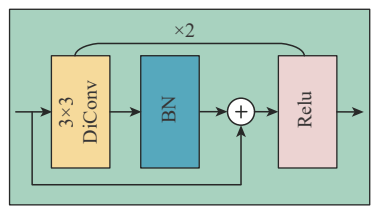


图 5　空洞卷积模块结构

Fig. 5　Structure of dilated convolution module

为减少特征信息损失，经过多头注意力的特征信息与初始序列 $\boldsymbol{X}$ 及空洞卷积结果进行融合，如下：

$$\boldsymbol{f}=\boldsymbol{X}+\text{MSA}(\boldsymbol{X})+\text{DiC}(\text{Img2Sqe}(\boldsymbol{I})) \tag{5}$$

式中：MSA(·) 表示多头自注意力运算；DiC(·) 表示空洞卷积运算。

如式(6)所示，融合信息 $\boldsymbol{f}$ 进入 FFN 完成特征信息交互以获得输出。FFN 表示双层感知器，该感知器使用非线性 Gelu 激活函数。对于信息交互的下一阶段，将序列经过 Sqe2Img(·) 序列转换图形操作获得下一个图像特征 $\boldsymbol{O}\in\mathbf{R}^{D\times H\times W}$。最后，经过多层编码层的特征信息进行 1×1 卷积变换维度与人体关键点个数一致，进一步生成姿态估计关键点热图。

$$\boldsymbol{O}=\text{Sqe2Img}(\boldsymbol{f}+\text{FFN}(\boldsymbol{f})) \tag{6}$$

2.3　模型参数设置

HRNet-S-W32 主干网络中，并行分支模型的通道数目分别为 32、64、128、256。为探究多层注意力对模型的增益，本文在双支路 Transformer 模块中重复堆叠 4 层编码层。式(1)中采用 1×1 卷积将通道维度从 32 升到 64，此时特征分辨率为 64×48，模型输出头部通道数目为 17。此外，在多头自注意力模块中，ViTPPose 估计模型设置的多头注意力数目为 2，双层感知器中扩张因子为 4。

3　飞行员姿态估计实验验证

3.1　实验验证方案

飞行员姿态估计实验验证方案如图 6 所示。

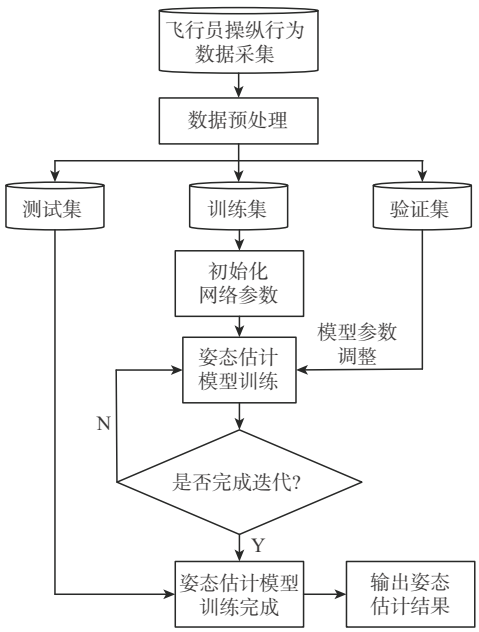


图6 飞行员姿态估计实验验证方案

Fig. 6 Experimental validation scheme for pilot pose estimation

3.2 飞行员关键点检测数据集

飞行员关键点检测数据集来源于智能驾驶舱飞行模拟平台，如图7所示。平台主要由操纵系统、高级仿真仪系统、图像采集系统、虚拟场景系统等组件构成。

基于《机组标准操作手册SOP第三版》飞行程序操作动作规范，开展五边飞行场景任务，从多个角度共采集3 000张飞行员操纵图片。图8为飞行员姿态估计数据集部分图片，包含背面、侧面等方向。该数据集包含推拉油门杆、打开自动导航按钮、打开起落架等多个操作动作。



图7 智能驾驶舱飞行模拟平台

Fig. 7 Intelligent cockpit flight simulation platform

经过数据预处理，图片数据集转化为高度为640像素，宽度为480像素。使用数据标注软件Labelme进行人工标注关键点。标注方式与MS COCO[8]关键点数据集相同，即3种标注方式：标注方式1为图中飞行员个体不存在某些关键节点，标注方式2为存在个体关键点但是被遮挡，标注方式3为人体关键节点清晰可见。每张图片标注一人或两人，对被标注飞行员身体标注17个关键点标签，分别为鼻子、左眼、右眼、左耳、右耳、左肩、右肩、左肘、右肘、左腕、右腕、左臀、右臀、左臀膝中部、右臀膝中部、左膝、右膝。数据集共设置训练集2 400张、验证集200张、测试集400张。



图8 飞行员关键点检测数据集样例

Fig. 8 Pilots keypoint detection dataset samples

3.3　模型训练环境配置

训练配置如下: Ubuntu 20.04 LST 64 位系统, GeForce RTX 3090 显卡, 采用 Pytorch 1.8.1 深度学习框架。在数据集上进行训练时, 将数据集中的图像裁剪后固定大小。采用 Adam 作为网络训练时的优化器, 初始学习率为 10^{-3}, 在第 150 轮时学习率衰减到 10^{-4}, 在第 200 轮时, 学习率衰减到 10^{-5}, 网络总共训练 220 轮。GPU 的最小批量大小为 32。本文在训练过程中使用随机的图像旋转和水平翻转进行数据增强。其中, 随机旋转角度在−45°～45°, 随机缩放比例在 0.65～1.35。

3.4　OKS 评估方法

目标关键点相似度 (object keypoint similarity, OKS)[8] 评估方式与 MS COCO[8] 关键点检测数据评估方式相似, 包括: AP^{50} 是在 D_{OKS}=0.5 时的检测关键点的准确率, AP^{75} 是在 D_{OKS}=0.75 时的检测关键点的准确率, mAP 表示在 D_{OKS}=0.50, 0.55, …, 0.95 时 10 个阈值之间所有预测关键点的准确率的平均值, AP^{M} 为中尺寸物体检测关键点的准确率, 以及 AR 在 D_{OKS}=0.50, 0.55, …, 0.95 时 10 个阈值点的平均召回率值。具体实现方法如下:

$$D_{\mathrm{OKS}}=\sum_{j}\frac{\exp(-d_j^2/2s^2k_j^2)\delta(v_j>0)}{\sum_{j}\delta(v_j>0)} \tag{7}$$

式中: j 为关节点的类型; d_j 为关节点人工标注坐标与模型预测坐标之间的欧氏距离; v_j 为关节点的可见信息; s 为关节点的尺度; k_j 为每种关键点的相关控制衰减常数。当 D_{OKS}=1 时, 表示精确的预测关键点, 当 D_{OKS}=0 时, 表示预测值与真实值差距太大。此外, 对于飞行员关键点数据集与 MS COCO[8] 数据集关键点不同之处, 各个关节点的 δ_j 分别为 0.026、0.025、0.025、0.035、0.035、0.079、0.079、0.072、0.072、0.062、0.062、0.107、0.107、0.100、0.100、0.087、0.087。

3.5　实验结果分析

为验证本文 ViTPPose 估计模型, 本文复现人体姿态估计领域的基准模型。ViTPPose 估计模型在飞行员关键点检测测试数据集上与基准模型进行了性能比较, 如表 1 所示。表中参数量和 GFLOPs (10^9 次/s 的浮点运算数)根据人体姿态估计网络计算得出。HRNet-S-W32 为 ViTPPose 估计模型的主干网络。ViTPPose 估计模型从输入像素尺寸为 256×192 开始训练, 达到 AP 为 92.3% 。相较于主干网路为 ResNet-50 的 Simple baseline[38] 模型, ViTPPose 估计模型预测的 AP 分数增加了 4.6%。当输入像素尺寸为 256×192 时, 与 HRNet-W32[34] 模型进行对比, 模型在参数量减少 5.0×10^6 的条件下, 预测 AP 提升了 3.0%。与 PoseUR[39] 模型相比, ViTPPose 估计模型在参数量减少 5.3×10^6 的条件下, 预测 AP 提升了 0.2%。与 ViTPose[40] 模型相比, 虽然 ViTPPose 估计模型预测 AP 减少了 3.3%, 但是 ViTPPose 估计模型参数量和计算量较少。如何在保证准确率的前提下减少模型参数量和复杂度, 需要进一步探究。图 9 为本文模型与表 1 部分模型的性能可视化效果。由于在卷积网络后期添加了双支路自注意力机制模块, 使得 ViTPPose 估计模型在后期能较好地感知高阶成对交互信息, 从而提升了 ViTPPose 估计模型的整体性能。部分飞行员关键点数据集可视化的效果, 如图 10 所示, 包含侧面、背面、遮挡环境。

表 1　模型在飞行员关键点检测测试数据集上的性能比较

Table 1　Performance comparison of models on pilot keypoint detection test dataset

模型	主干网络	输入尺寸/像素	参数量	GFLOPs	AP/%	AP^{50}/%	AP^{75}/%	AP^{M}/%	AR/%
SimpleBaseline[38]	ResNet-50	256×192	3.40×10^7	8.90	87.7	89.9	89.1	85.3	87.8
SimpleBaseline[38]	ResNet-152	384×288	6.86×10^7	35.6	89.4	91.0	91.1	87.8	88.2
HRNet-W32[4]	HRNet-W32	256×192	2.85×10^7	7.10	89.3	91.4	91.3	87.3	86.9
HRNet-W32[4]	HRNet-W32	384×288	2.85×10^7	16.0	90.6	92.5	93.6	90.9	92.5
PoseUR[39]	HRFormer-B	256×192	2.88×10^7	12.6	92.1	93.6	91.8	93.0	93.8
ViTPose[40]	ViTAE-H	256×192	6.32×10^8		95.6	95.2	94.3	95.6	96..4
ViTPPose	HRNet-S-W32	256×192	2.35×10^7	13.0	92.3	93.5	92.5	92.6	93.5

3.6　消融实验

3.6.1　双支路 Transformer 模块编码层

为探讨双支路 Transformer 模块编码层数目对模型参数数目和精度的影响, 本文采用不同层数在飞行员关键点数据集进行验证。为提高消融试验的准确性, 设计了 5 种编码层数目(0、1、2、3、4), 其中, 0 层表示不使用双支路 Transformer 模块。如表 2 所示, 3 层编码带来的增益效果较为明显, 相较于 2 层编码 AP 提升了 0.7%。相较于使用 3 层编码, 模型使用 4 层编码 AP 性能提升了 0.3%, 带来性能增益效果减少。通过比较 5 种不同数目的编码层, 可以得出使用合适的层数可以在增加少量参数的情况下获得优越的结果。图 11 为 ViTPPose 估计模型使用不同编码层数目对训练速度影响的

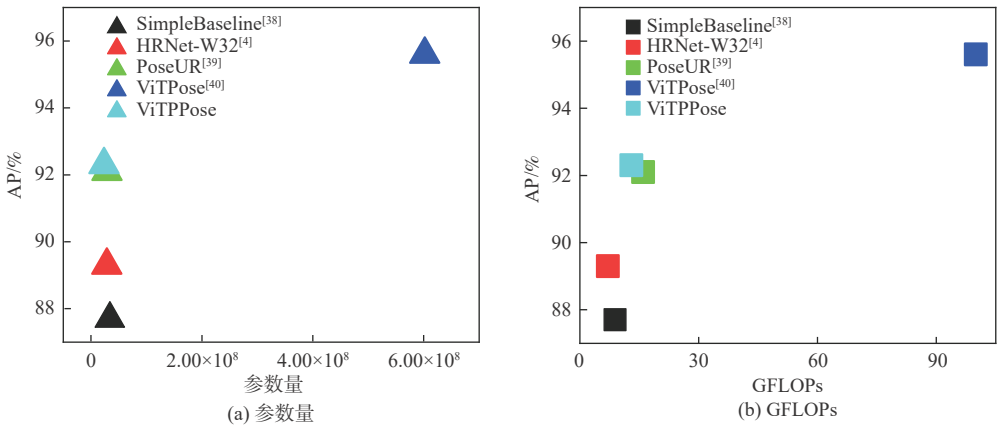


图 9 模型在飞行员关键点检测测试数据集上的性能可视化效果

Fig. 9 Performance visualization of model on pilot keypoint detection test dataset



图 10 飞行员姿态估计可视化效果

Fig. 10 Visualization of pilot pose estimation

表 2 模型使用不同编码层数目在飞行员关键点检测测试数据集上的性能比较

Table 2 Performance comparison of using different numbers of encoder layers on pilot keypoint detection test dataset

编码层数目	输入尺寸/像素	参数量	GFLOPs	AP/%
0	256×192	2.15×10^7	6.2	90.3
1	256×192	2.20×10^7	7.9	90.7
2	256×192	2.24×10^7	9.6	91.3
3	256×192	2.29×10^7	11.3	92.0
4	256×192	2.35×10^7	13.0	92.3

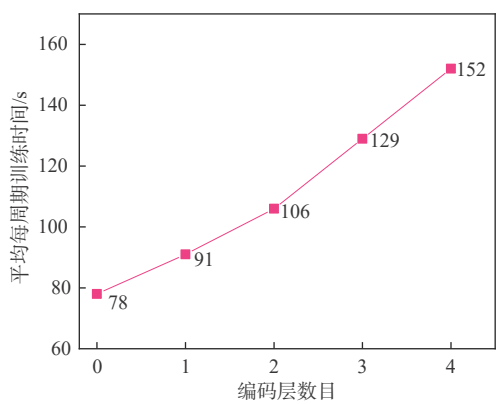


图 11 编码层数目对 ViTPPose 估计模训练速度影响的可视化效果

Fig. 11 Visualization of effect of number of coding layers on training speed of ViTPPose estimation model

可视化效果，图中每周期平均训练时间=总训练时长/总训练周期。由图 11 可知，随着编码层数目的增加，ViTPPose 估计模型的每周期平均训练时间逐渐增加。选取合适的编码层数，对平衡模型的准确率及训练速度是有效的。

图 12 为 ViTPPose 估计模型使用不同编码层数目对训练损失值影响的可视化效果。随着迭代次数的增加，ViTPPose 估计模型损失趋向稳定。

3.6.2 空洞卷积模块

为探究双支路 Transformer 模块编码层中空洞卷积模块对模型性能的影响。本文采用不同模型在飞行员关键点数据集进行了验证。如表 3 所示，ViTPPose* 模型表示 CNN 末端编码层中仅使用 Transformer 模块而不使用空洞卷积模块。ViTPPose$^{\&}$ 模型为空洞卷积模块的空洞卷积率为 1，ViTPPose$^{\$}$ 模型表示空洞卷积模块的空洞卷积率为 4。由表 3 可知，相较于 ViTPPose*估计模型，ViTPPose 估计模型在增加 0.7×10^6 参数的条件下，AP 提升了 0.5%。

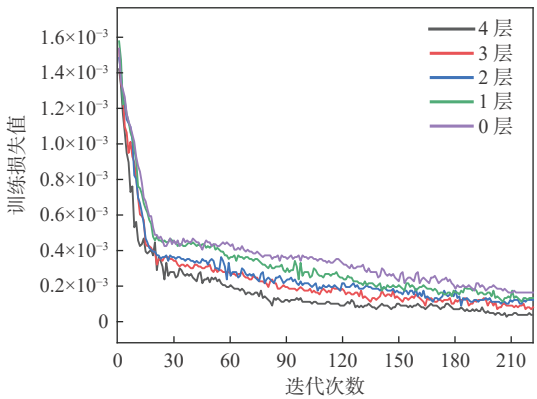


图 12　编码层数目对 ViTPPose 估计模型训练损失值影响可视化效果

Fig. 12　Visualization of effect of number of encoder layers on loss value of the ViTPPose estimation model training

表 3　模型使用空洞卷积模块在飞行员关键点测试数据集上的性能比较

Table 3　Performance comparison of models using dilated convolution module on pilot keypoint test dataset

模型	输入尺寸/像素	参数量	GFLOPs	AP/%
ViTPPose*	256×192	2.28×10^7	12.6	91.8
ViTPPose$^{\&}$	256×192	2.35×10^7	13.0	92.0
ViTPPose$^{\$}$	256×192	2.36×10^7	13.0	91.5
ViTPPose	256×192	2.35×10^7	13.0	92.3

相较于 ViTPPose$^{\&}$、ViTPPose$^{\$}$ 估计模型，ViTPPose 估计模型使用空洞卷积率为 2 时，取得最高 AP。在编码层中，通过使用 Transformer 联合具有合适空洞率的空洞卷积模块，从而感知更大感受野信息的策略是有效的。ViTPPose 估计模型在增加少许参数的条件下，获得了明显的性能提升。

3.7　光照强度可解释分析

为探究不同光照强度对模型性能的影响，如图 13 所示，本文设置 4 种不同的光照强度，4 种级别光照强度分别对应 50 000，20 000，500，0.5 lx。

本文对训练完成的 ViTPPose 估计模型解析输出结果，并对输出结果进行概率激活，形成民机驾驶舱飞行员关键点预测热图。如图 14 所示，从上而下表示 50 000，20 000，500，0.5 lx 这 4 种不同光照强度对 ViTPPose 估计模型输出产生的可视化效果图。色柱由上而下的色彩变化对应着不同的置信概率，其中，红色区域表示置信概率最大。椭圆中白色图形为被手工制作标签标记的位置。图中水平方向表示从不同关键点的类型，与飞行员关键点检测数据集相互对应。其中，白色区域与深红色区域越吻合表示模型识别效果越好，反之效果越差。



图 13　不同光照强度对比

Fig. 13　Comparison of different light strengths

对比图 14 淡红色椭圆区域可知，ViTPPose 估计模型在 20 000，500 lx 等级光照强度下识别能力较突出，在光照强度 50 000，0.5 lx 的条件下，ViTPPose 估计模型识别能力较弱。对比图 14 绿色矩形区域，ViTPPose 估计模型在 20 000 lx 光照强度下展示出了较好的性能。基于飞行员坐姿估计热图，定性地揭示 ViTPPose 估计模型对各个关键点的识别能力受不同光照强度度模型影响。

为进一步定量化地揭示驾驶舱复杂光照对 ViTPPose 估计模型识别关键节点性能的影响，本文把 400 张飞行员关键点检测测试集的光照等级依

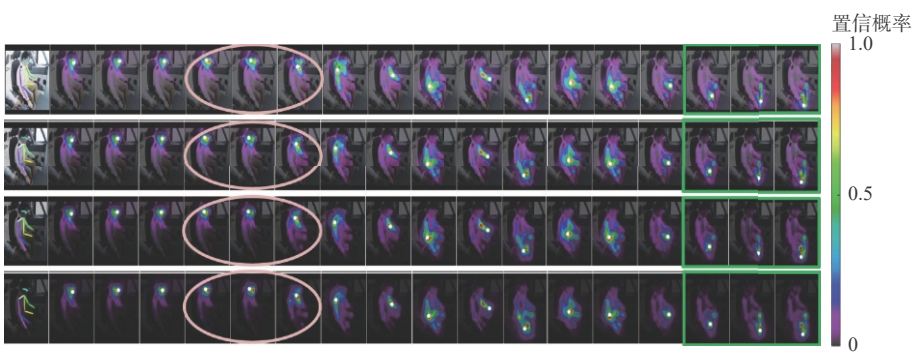


图 14　不同光照下飞行员姿态估计输出结果对比

Fig. 14　Comparison of pilot attitude estimation output results under different light intensities

次设置为4种，分别为50 000，20 000，500，0.5 lx。如表4所示，采用本文模型在4种级别光照强度等级下飞行员关键点测试数据集上进行识别准确率测验。

表4 本文模型在4种级别光照强度等级下飞行员关键点测试数据集上的性能比较

Table 4 Performance of the proposed model is compared on pilot keypoint test dataset under four levels of light intensity levels

光照强度/lx	输入尺寸/像素	AP/%
50 000	256×192	86.2
20 000	256×192	93.8
500	256×192	90.7
0.5	256×192	85.1

由表4可知，本文模型对不同光照强度表现出不同的识别准确率。当光照强度为20 000 lx时，AP为93.8%，为表中最大AP值。当光照强度为500 lx时，AP为90.7%。当光照强度为50 000，0.5 lx时，AP分别为86.2%和85.1%，模型关键点识别能力较弱。

通过定性与定量的比较不同光照强度下飞行员姿态估计输出结果对比图，可知ViTPPose估计模型对不同光照强度展现偏好性。利用ViTPPose估计模型的偏好性，在民机驾驶舱图片中进行光照增强与减弱，光照强度重塑为具体多少数值，从而提升姿态估计准确率有着积极的指导意义。

4 结 论

1）针对缺少民用飞机飞行员姿态估计数据集的问题，本文在飞行模拟场景下拍摄了飞行员模拟飞行时的姿态图片，经过人工标注人体关键点，制作飞行员操纵数据集。

2）本文提出ViTPPose估计模型。该模型在网络末端使用了包含多层编码层的双支路Transformer模块，定量地探究了不同编码层数目对模型性能的影响，以及定量的探究双支路空洞卷积策略的有效性。

3）在驾驶舱复杂光照的环境下，构建飞行员坐姿估计热力图，定性与定量地探究不同光照强度对模型性能的影响，并作出可解释分析。

参考文献（References）

[1] 中华人民共和国国务院. “十三五”国家战略性新兴产业发展规划67号[R]. 北京: 中华人民共和国国务院, 2016.
The State Council of the People’s Republic of China. “13th Five-Year” national strategic emerging industry development plan No 67[R]. Bejing: State Council of the People’s Republic of China, 2016(in Chinese).

[2] 中华人民共和国国务院. 中华人民共和国国民经济和社会发展第十四个五年规划和2035年远景目标纲要[R]. 北京: 中华人民共和国国务院, 2021.
The State Council of the People’s Republic of China. Outline of the People’s Republic of China 14th five-year plan for national economic and social development and long-range objectives for 2035[R]. Bejing: State Council of the People’s Republic of China, 2021(in Chinese).

[3] 杨志刚, 张炯, 李博, 等, 民用飞机智能飞行技术综述[J]. 航空学报, 2021, 42 (4): 525198.
YANG Z G, ZHANG J, LI B, et al. Reviews on intelligent flight technology of civil aircraft[J]. Acta Aeronautica et Astronautica Sinica, 2021, 42(4): 525198(in Chinese).

[4] TOSHEV A, SZEGEDY C. DeepPose: Human pose estimation via deep neural networks[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2014: 1653-1660.

[5] CHEN Y P, DAI X Y, LIU M C, et al. Dynamic ReLU[C]//Proceedings of the European Conference on Computer Vision. Berlin: Springer, 2020: 351-367.

[6] HE K M, ZHANG X Y, REN S Q, et al. Deep residual learning for image recognition[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2016: 770-778.

[7] HE K M, GKIOXARI G, DOLLÁR P, et al. Mask R-CNN[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2020, 42(2): 386-397.

[8] SELVARAJU R R, COGSWELL M, DAS A, et al. Grad-CAM: Visual explanations from deep networks via gradient-based localization[C]//Proceedings of the IEEE International Conference on Computer Vision. Piscataway: IEEE Press, 2017: 618-626.

[9] ZHANG Q L, YANG Y B. Group-CAM: Group score-weighted visual explanations for deep convolutional networks[EB/OL]. (2021-03-25)[2022-08-19]. http://arxiv.org/abs/2103.13859.

[10] LIU Z, MAO H Z, WU C Y, et al. A ConvNet for the 2020s[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2022: 11966-11976.

[11] SUN K, XIAO B, LIU D, et al. Deep high-resolution representation learning for human pose estimation[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2019: 5686-5696.

[12] CHENG B W, XIAO B, WANG J D, et al. HigherHRNet: Scale-aware representation learning for bottom-up human pose estimation[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2020: 5385-5394.

[13] ANDRILUKA M, PISHCHULIN L, GEHLER P, et al. 2D human pose estimation: New benchmark and state of the art analysis[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2014: 3686-3693.

[14] JOHNSON S, EVERINGHAM M. Clustered pose and nonlinear appearance models for human pose estimation[C]//Proceedings of the British Machine Vision Conference 2010. London: British Machine

Vision Association, 2010: 1-11.

[15] LIN T Y, MAIRE M, BELONGIE S, et al. Microsoft COCO: Common objects in context[C]//Proceedings of the European Conference on Computer Vision. Berlin: Springer, 2014: 740-755.

[16] DOSOVITSKIY A, BEYER L, KOLESNIKOV A, et al. An image is worth 16x16 words: Transformers for image recognition at scale [EB/OL]. (2020-10-22)[2022-08-20]. https://arxiv.org/abs/2010.11929.

[17] DENG J, DONG W, SOCHER R, et al. ImageNet: A large-scale hierarchical image database[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2009: 248-255.

[18] TOUVRON H, CORD M, DOUZE M, et al. Training data-efficient image transformers & distillation through attention[EB/OL]. (2020-12-23)[2022-08-20]. https://arxiv.org/abs/2012.12877.

[19] TOUVRON H, CORD M, SABLAYROLLES A, et al. Going deeper with Image Transformers[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. Piscataway: IEEE Press, 2021: 32-42.

[20] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need[C]//Proceedings of the 31st International Conference on Neural Information Processing Systems. New York: ACM, 2017: 6000-6010.

[21] XIAO T T, SINGH M, MINTUN E, et al. Early convolutions help transformers see better[EB/OL]. (2021-06-28)[2022-08-21]. https://arxiv.org/abs/2106.14881.

[22] NEWELL A, YANG K Y, DENG J. Stacked hourglass networks for human pose estimation[C]//Proceedings of the European Conference on Computer Vision. Berlin: Springer, 2016: 483-499.

[23] CHEN Y L, WANG Z C, PENG Y X, et al. Cascaded pyramid network for multi-person pose estimation[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2018: 7103-7112.

[24] NEWELL A, HUANG Z A, DENG J. Associative embedding: end-to-end learning for joint detection and grouping[C]//Proceedings of the 31st International Conference on Neural Information Processing Systems. New York: ACM, 2017: 2274–2284.

[25] CAO Z, SIMON T, WEI S H, et al. Realtime multi-person 2D pose estimation using part affinity fields[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2017: 1302-1310.

[26] WEI S H, RAMAKRISHNA V, KANADE T, et al. Convolutional pose machines[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2016: 4724-4732.

[27] LIU Z, LIN Y T, CAO Y, et al. Swin Transformer: Hierarchical vision transformer using shifted windows[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. Piscataway: IEEE Press, 2021: 9992-10002.

[28] 王静, 李沛橦, 赵容锋, 等. 融合卷积注意力和 Transformer 架构的行人重识别方法[J]. 北京航空航天大学学报, 2024, 50(2): 466-476.
WANG J, LI P T, ZHAO R F, et al. A person re-identification method for fusing convolutional attention and Transformer architecture[J]. Journal of Beijing University of Aeronautics and Astronautics, 2024, 50(2): 466-476 (in Chinese).

[29] YUAN Y H, FU R, HUANG L, et al. HRFormer: High-resolution transformer for dense prediction[EB/OL]. (2021-10-18)[2022-08-21]. https://arxiv.org/abs/2110.09408.

[30] XIONG Z N, WANG C X, LI Y, et al. Swin-Pose: Swin transformer based human pose estimation[C]//Proceedings of the IEEE 5th International Conference on Multimedia Information Processing and Retrieval. Piscataway: IEEE Press, 2022: 228-233.

[31] YANG S, QUAN Z B, NIE M, et al. TransPose: Keypoint localization via transformer[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. Piscataway: IEEE Press, 2021: 11782-11792.

[32] SAMEK W, MONTAVON G, LAPUSCHKIN S, et al. Explaining deep neural networks and beyond: A review of methods and applications[J]. Proceedings of the IEEE, 2021, 109(3): 247-278.

[33] KOH P W, LIANG P. Understanding black-box predictions via influence functions[EB/OL]. (2020-12-29)[2022-08-22]. http://arxiv.org/abs/1703.04730.

[34] ZEILER M D, FERGUS R. Visualizing and understanding convolutional networks[C]//Proceedings of the European Conference on Computer Vision. Berlin: Springer, 2014: 818-833.

[35] ZHANG S S, YANG J, SCHIELE B. Occluded pedestrian detection through guided attention in CNNs[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2018: 6995-7003.

[36] ZHOU B L, KHOSLA A, LAPEDRIZA A, et al. Learning deep features for discriminative localization[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2016: 2921-2929.

[37] WANG H F, WANG Z F, DU M N, et al. Score-CAM: Score-weighted visual explanations for convolutional neural networks [C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops. Piscataway: IEEE Press, 2020: 111-119.

[38] XIAO B, WU H P, WEI Y C. Simple baselines for human pose estimation and tracking[EB/OL]. (2018-08-21)[2022-08-22]. http://arxiv.org/abs/1804.06208.

[39] MAO W A, GE Y T, SHEN C H, et al. PoseUR: Direct human pose regression with transformers[C]//Proceedings of the European Conference on Computer Vision. Berlin: Springer, 2022: 72-88.

[40] XU Y F, ZHANG J, ZHANG Q M, et al. ViTPose: Simple vision transformer baselines for human pose estimation[EB/OL]. (2022-04-26)[2022-08-22]. http://arxiv.org/abs/2204.12484.

Vision Transformer-based pilot pose estimation

WU Honglan*, LIU Hao, SUN Youchao

(College of Civil Aviation, Nanjing University of Aeronautics and Astronautics, Nanjing 211106, China)

Abstract: Human pose estimation is an important aspect in the field of behavioral perception and a key technology in the way of intelligent interaction in the cockpit of civil aircraft. To establish an explainable link between the complex lighting environment in the cockpit of civil aircraft and the performance of the pilot pose estimation model, the visual Transformer-based pilot pose (ViTPPose) estimation model is proposed. In order to capture the global correlation of subsequent higher-order features while expanding the perceptual field, this model employs a two-branch Transformer module with several coding layers at the end of the convolutional neural networks (CNN) backbone network. The coding layers combine the Transformer and the dilated convolution. Based on the flight crew's standard operating procedures, a pilot maneuvering behavior keypoint detection dataset is established for flight simulation scenarios. ViTPPose estimation model completes the pilot seating estimation on this dataset and verifies its validity by comparing it with the benchmark model. The seating estimation heatmap is created in the context of the cockpit's complicated lighting to examine the model's preferred lighting intensity, evaluate the ViTPPose estimation model's performance under various lighting conditions, and highlight the model's reliance on various lighting intensities.

Keywords: civil aircraft; intelligent cockpit; pilot pose estimation; convolutional neural network; self-attention; explainability

Received: 2022-09-29; **Accepted**: 2023-03-26; **Published Online**: 2023-04-20 16:26

URL: link.cnki.net/urlid/11.2625.V.20230420.1442.002

Foundation items: Joint Fund of National Natural Science Foundation of China and Civil Aviation Administration of China (U2033202,U1333119); National Natural Science Foundation of China (52172387)

* **Corresponding author**. E-mail: wuhonglan@nuaa.edu.cn

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2022.0794

侧风对直升机沙盲特性的影响

谭剑锋[1,*]，杨宇霄[1]，张卫国[2]，王世雄[3]，吕琳[3]

(1. 南京工业大学 机械与动力工程学院，南京 211816；2. 中国空气动力研究与发展中心 低速研究所，绵阳 621000；
3. 陆军航空兵学院 第三飞行训练旅，临汾 041000)

摘 要：直升机旋翼与地面复杂干扰流场诱导沙床沙粒运动，形成沙盲现象，然而侧风改变旋翼与地面干扰流场，改变沙云形态和沙盲现象。因此，将侧风作用项嵌入基于黏性涡粒子的旋翼/地面气动干扰模型，体现侧风对旋翼与地面干扰流场的作用，并基于离散单元法的沙粒动力学模型，增加侧风引起的气动力，体现侧风对沙粒运动的作用。研究侧风对直升机前飞沙盲特性的影响，并分析侧风速度、风向对直升机沙盲特性的影响规律。结果表明：侧风对直升机沙盲沙云形态影响显著，且左侧风显著增强沙尘浓度，而迎风状态减小沙尘浓度；随着侧风速度的增加，左侧风状态下的沙尘浓度先增加后减小。

关 键 词：侧风；沙盲；旋翼/地面气动干扰；离散单元法；直升机
中图分类号：V221.52; TB553
文献标志码：A **文章编号**：1001-5965(2024)10-3111-12

直升机低空悬停、着陆、贴地飞行状态下，旋翼与地面干扰形成复杂的非定常流场，并与松软的沙床耦合，引起沙粒扬起，诱发直升机沙盲，产生弱视环境，威胁直升机飞行安全，引发大量飞行事故[1-2]。然而，直升机贴地飞行中通常出现较为明显的侧风，进而影响旋翼与地面干扰流场，改变直升机沙盲特性，因此，有必要开展侧风对直升机沙盲现象的影响研究。

针对直升机沙盲，国内外开展了部分试验。Wachspress 等采用 H-21 纵列式直升机开展悬停状态的沙盲试验[3]，但此试验中的直升机仅限于悬停状态。Whitehouse 等[4]开展基于模型旋翼的沙盲试验，研究旋翼参数对沙盲的影响，然而试验并未考虑复杂的旋翼桨叶特性，且未明确沙盲形成机制。Leishman 等[5]采用粒子图像测速 (particle image velocity, PIV) 测量模型旋翼沙盲环境下的两相流特性，并结合粒子示踪系统研究沙粒扬起机制[6]。Wong 和 Tanner[7]开展了 EH-60L 全尺寸直升机沙盲飞行试验，但试验仅限几种飞行状态，并未考虑侧风影响。

由于沙盲试验测试难度大，国内外开展了直升机沙盲现象数值分析方法研究。文献 [8-9] 采用基于雷诺平均 Navier-Stokes (reynolds average Navier-Stokes, RANS) 方程的计算流体力学 (computational fluid dynamics, CFD) 和沙粒运动模型分析直升机沙盲现象，但该方法将沙云简化为连续体，而非真实沙粒离散体。文献 [10-12] 将基于 RANS、自由混合法 -RANS (free-vortex method-RANS, FVM-RANS)、有限体积法的 CFD 软件与拉格朗日沙粒跟踪方法耦合，研究模型旋翼沙盲现象，采用基于经验的沙粒扬起速度罚值、起跳概率模型等。徐国华等[13]耦合 STAR CCM+ CFD 软件与 EDEM 软件开展两片桨叶直升机沙盲特性研究，但计算量限制了数值模拟时间和沙粒模拟数量。文献 [14] 通过耦合旋

收稿日期：2022-09-18；**录用日期**：2022-11-14；**网络出版时间**：2023-01-17 13:26
网络出版地址：link.cnki.net/urlid/11.2625.V.20230116.1731.002
基金项目：国家自然科学基金 (12172165)；江苏省自然科学基金 (BK20211259)；江苏省高校“青蓝工程”优秀青年骨干教师项目
*通信作者. E-mail：Jianfengtan@njtech.edu.cn

引用格式：谭剑锋，杨宇霄，张卫国，等. 侧风对直升机沙盲特性的影响 [J]. 北京航空航天大学学报，2024，50（10）：3111-3122.
TAN J F，YANG Y X，ZHANG W G，et al. Influence of crosswind on helicopter brownout [J]. Journal of Beijing University of Aeronautics and Astronautics，2024，50（10）：3111-3122 （in Chinese）.

翼自由尾迹模型、地面镜像模型及拉格朗日沙粒跟踪方法模拟直升机沙盲现象，但仍采用速度罚值模型等假设模型。谭剑锋等[15-16]将旋翼黏性涡粒子法、黏性地面气动模型、沙粒离散单元法 (discrete element method, DEM) 耦合，研究 EH-60L 全尺寸直升机悬停和着陆状态下的沙盲形成特性。旋翼与地面干扰流场耦合沙粒运动，导致复杂的沙云行为，形成沙盲现象。然而，上述研究均未考虑直升机着陆和飞行状态的侧风现象。

因此，本文基于黏性涡粒子和黏性地面气动模型的旋翼/地面气动干扰模型，增加侧风作用项，并将侧风引起的气动力作用项增加至沙粒动力学模型，从而分析侧风对 EH-60L 全尺寸直升机沙盲特性的影响，并研究侧风风向、风速对 EH-60L 全尺寸直升机前飞沙盲特性的影响规律。

1 计算方法

1.1 侧风状态的沙粒动力学模型

侧风状态的直升机沙云由沙粒复杂运动轨迹组成，根据 DEM[17-18] 计算得到沙粒复杂运动轨迹和沙云空间形态，则沙粒动力学模型表示为

$$m\frac{\mathrm{d}\boldsymbol{V}_\mathrm{p}}{\mathrm{d}t}=\boldsymbol{F}_\mathrm{i}+\boldsymbol{F}_\mathrm{f}+\boldsymbol{F}_\mathrm{g} \tag{1}$$

$$I\frac{\mathrm{d}\boldsymbol{\omega}_\mathrm{p}}{\mathrm{d}t}=\boldsymbol{T} \tag{2}$$

式中：m和$\boldsymbol{V}_\mathrm{p}$为沙粒的质量和速度；$I$和$\boldsymbol{\omega}_\mathrm{p}$为沙粒的惯量和角速度；$\boldsymbol{F}_\mathrm{i}$、$\boldsymbol{F}_\mathrm{f}$、$\boldsymbol{F}_\mathrm{g}$分别为沙粒内部力、旋翼与地面干扰流场和侧风对沙粒的气动力、沙粒重力；$\boldsymbol{T}$为沙粒受到的力矩。

沙粒内部力$\boldsymbol{F}_\mathrm{i}$由沙粒间接触碰撞力和沙粒-地面接触碰撞力构成，并将地面定义为刚性表面，两者的法向力$\boldsymbol{F}_\mathrm{i}^n$和切向力$\boldsymbol{F}_\mathrm{i}^t$由 Hertz-Mindlin 模型计算，即

$$\boldsymbol{F}_\mathrm{i}^n=\frac{4}{3}Y^*\delta_n^{3/2}\sqrt{R^*}\boldsymbol{n}-2\sqrt{5/6}\beta\sqrt{S_nm^*}\boldsymbol{v}_n^\mathrm{rel} \tag{3}$$

$$\boldsymbol{F}_\mathrm{i}^t=-S_t\delta_t\boldsymbol{t}-2\sqrt{5/6}\beta\sqrt{S_nm^*}\boldsymbol{v}_t^\mathrm{rel} \tag{4}$$

式中：Y^*、m^*、R^*、β、δ_n、$\boldsymbol{n}$、$\boldsymbol{v}_n^\mathrm{rel}$、$S_n$、$\delta_t$、$\boldsymbol{t}$、$\boldsymbol{v}_t^\mathrm{rel}$和$S_t$分别为沙粒模量、质量、半径、刚度、法向变形量、法矢量、法向速度、法向刚度、切向变形量、切矢量、切向速度和切向刚度。

沙云由直径小于 100 μm 的沙粒构成，相比于直升机旋翼流场尺度较小，因而忽略沙粒对直升机旋翼与地面干扰流场和侧风的影响，但考虑旋翼与地面干扰流场及侧风对沙粒运动的影响。旋翼与地面干扰流场和侧风作用下沙粒受到的气动力$\boldsymbol{F}_\mathrm{f}$为

$$\boldsymbol{F}_\mathrm{f}=\frac{1}{2}\rho_\mathrm{air}\left|\boldsymbol{v}_\mathrm{p}-\boldsymbol{v}_\mathrm{f}-\boldsymbol{v}_\mathrm{w}\right|(\boldsymbol{v}_\mathrm{p}-\boldsymbol{v}_\mathrm{f}-\boldsymbol{v}_\mathrm{w})C_DA \tag{5}$$

式中：A、ρ_air、$\boldsymbol{v}_\mathrm{f}$和$\boldsymbol{v}_\mathrm{w}$分别为沙粒投影面积、空气密度、流场速度和侧风速度；C_D为气动阻力系数：

$$C_D=\frac{24}{Re}\left(1+0.15Re^{0.687}\right) \tag{6}$$

式中：$Re=\left|\boldsymbol{v}_\mathrm{p}-\boldsymbol{v}_\mathrm{f}\right|d_\mathrm{e}/\nu$；$d_\mathrm{e}$为沙粒直径；$\nu$为黏性系数。

1.2 侧风状态的旋翼/地面气动干扰模型

侧风影响直升机旋翼与地面干扰流场，并影响直升机沙云形态。旋翼/地面气动干扰模型由旋翼非定常面元气动模型、基于黏性涡粒子的旋翼尾迹模型、基于涡面的黏性地面气动模型构成，并增加侧风作用项。流场速度势函数ϕ为[19]

$$\phi(x,y,z,t)=\frac{1}{4\pi}\int_{S_\mathrm{B}}\mu\boldsymbol{n}\cdot\nabla\left(\frac{1}{r_\mathrm{t}}\right)\mathrm{d}S-\frac{1}{4\pi}\int_{S_\mathrm{B}}\sigma\left(\frac{1}{r_\mathrm{t}}\right)\mathrm{d}S+\frac{1}{4\pi}\int_{S_\mathrm{w}}\mu\boldsymbol{n}\cdot\nabla\left(\frac{1}{r_\mathrm{t}}\right)\mathrm{d}S \tag{7}$$

式中：μ、σ、S_W、S_B和r_t分别为桨叶表面偶极子分布、汇分布、桨叶尾随涡面、桨叶表面和空间距离。

根据 Neumann 边界条件，并考虑侧风作用，计算得到桨叶汇分布为

$$-\sigma=\boldsymbol{n}\cdot(\boldsymbol{v}_\mathrm{B}+\boldsymbol{v}_\mathrm{w}) \tag{8}$$

式中：$\boldsymbol{v}_\mathrm{B}$为旋翼桨叶表面速度。

根据桨叶无穿透边界条件、后缘 Kutta 条件，将旋翼桨叶离散为四边形面元，则桨叶偶极子分布为

$$\sum_{k=1}^{N_\mathrm{B}}\mu_kA_k=-\sum_{k=1}^{N_\mathrm{B}}\sigma_kB_k \tag{9}$$

式中：N_B为桨叶面元数量；A_k和B_k为影响系数。

直升机旋翼尾迹模型采用黏性涡粒子方法，尾迹涡量分布为

$$\frac{\mathrm{d}\boldsymbol{y}}{\mathrm{d}t}=\boldsymbol{u}(\boldsymbol{y},t) \tag{10}$$

$$\frac{\mathrm{d}\boldsymbol{\alpha}}{\mathrm{d}t}=\nabla\boldsymbol{u}\cdot\boldsymbol{\alpha}+\nu\nabla^2\boldsymbol{\alpha} \tag{11}$$

式中：$\boldsymbol{u}$为流场速度；$\boldsymbol{y}$和$\boldsymbol{\alpha}$分别为涡粒子位置和涡量。

采用 4 阶 Runge-Kutta 时间推进格式、粒子强度交换法、直接求解法求解上式旋翼尾迹涡量方程，得到旋翼尾迹结构，并计算得到旋翼流场。地面干扰效应采用基于涡面的黏性地面气动模型[19]计算，地面涡面分布为

$$\frac{\boldsymbol{\gamma}\times\boldsymbol{n}}{2}-\int_S K_\mathrm{G}(\boldsymbol{z}-\boldsymbol{z}')\times\boldsymbol{\gamma}(\boldsymbol{z}')\mathrm{d}S=-\int_V\frac{1}{\varepsilon^3}K(\boldsymbol{z}-\boldsymbol{y}_i)\times\boldsymbol{\alpha}_i\mathrm{d}V+\nabla\phi+\boldsymbol{v}_\mathrm{w} \tag{12}$$

式中: $\boldsymbol{\gamma}$、z、K_{G}、ε和 K 分别为地面涡面、位置、Green 函数梯度、涡核半径和 Biot-Savart 核函数。

根据涡量扩散方程, 地面涡面将扩散到旋翼流场, 通过积分得到旋翼尾迹涡量增量, 并在旋翼流场中增加诱导项, 即旋翼流场速度为

$$\boldsymbol{u}(\boldsymbol{x}) = -\int_V \frac{1}{\varepsilon^3} K(\boldsymbol{x}-\boldsymbol{y}) \times \boldsymbol{\alpha} \mathrm{d}V + \nabla\phi + \int_S K_{\mathrm{G}}(\boldsymbol{x}-\boldsymbol{z}) \times \boldsymbol{\gamma}(z) \mathrm{d}S + \boldsymbol{v}_{\mathrm{w}} \tag{13}$$

旋翼尾迹涡量诱导地面速度变化, 从而构成旋翼尾迹模型与地面气动模型的相互耦合, 形成旋翼/地面气动干扰模型。侧风状态下旋翼与地面气动干扰导致的流场速度 $\boldsymbol{v}_{\mathrm{f}}$ 为

$$\boldsymbol{v}_{\mathrm{f}} = \nabla\phi - \int_V \frac{1}{\varepsilon^3} K(\boldsymbol{x}-\boldsymbol{y}) \times \boldsymbol{\alpha} \mathrm{d}V + \int_S K_{\mathrm{G}}(\boldsymbol{x}-\boldsymbol{z}) \times \boldsymbol{\gamma}(z) \mathrm{d}S \tag{14}$$

1.3　侧风状态的沙尘浓度

直升机着陆过程飞行员无需全方位视野, 仅需要较小的清晰视野 (field of view, FOV)。根据 EH-60L 直升机座舱参数, 以驾驶员双目为参考点, 驾驶舱可见典型视野区域为 124°×45°(其中, 水平视野区 124°, 俯仰视野区 45°), 如图 1 所示。飞行员典型的聚焦视野区域为 40°×25°, 此视野区域为移动扫掠区, 即聚焦视野区可在驾驶舱视野区移动扫掠, 飞行员任一聚焦视野区能清晰观测地面, 直升机就可安全着陆。飞行过程中用于操纵的最佳视野区为沙粒数量最少的飞行员扫掠聚焦视野区。因此, 直升机沙尘浓度定义需基于飞行员聚焦视野区, 即沙粒数量最少的飞行员扫掠聚焦视野区。由于地面沙粒并不影响飞行员视野, 只有扬起的沙尘阻碍飞行员视野, 沙尘浓度仍需考虑沙尘扬起因素。因此, 沙尘浓度ρ_{s}定义为扫掠的飞行员聚焦视野区内沙粒数量最小值与扬起沙粒数量的比值, 即

$$\rho_{\mathrm{s}} = \min\left[\sum_{\phi_{\mathrm{s}}=i}^{i+25^\circ}\sum_{\theta_{\mathrm{s}}=j}^{j+40^\circ} n_{\mathrm{s}}(\phi_{\mathrm{s}},\theta_{\mathrm{s}})\right] \Big/ N_{\mathrm{s}} \tag{15}$$

式中: $-21^\circ \leqslant i \leqslant -1^\circ$, $-62^\circ \leqslant j \leqslant 22^\circ$; n_{s}为位于飞行员双目位置极坐标夹角θ_{s}和ϕ_{s}的单位视角范围沙粒数量, 如图 2 所示; N_{s}为扬起沙粒数量。

第 1.1 节和 1.2 节基于地面全局坐标系计算得到的沙粒空间分布, 而沙尘浓度需基于飞行员双目视野极坐标, 因此, 需将沙粒空间分布从地面全局坐标系转换到飞行员双目视野极坐标系。通过坐标系转换, 获得沙粒在飞行员双目坐标系下水平方位角θ_{s}和俯仰角ϕ_{s}, 以及与飞行员双目距离r_{s}, 并用式 (15) 计算得到直升机飞行过程的沙尘浓度。

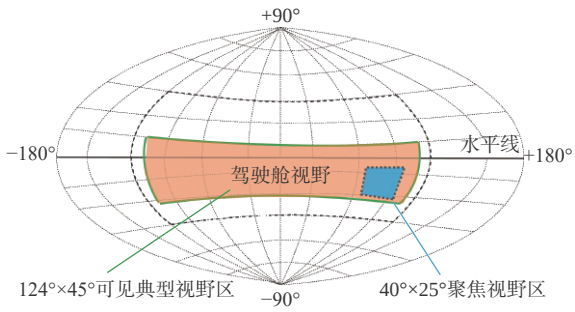


图 1　直升机驾驶舱视野

Fig. 1　Cockpit view of helicopter



图 2　飞行员视野坐标

Fig. 2　Frame of pilot's view

2　侧风状态直升机沙盲特性分析

2.1　无侧风直升机沙盲

计算得到 EH-60L 全尺寸直升机前飞沙云形成形态, 并与文献 [7] 飞行试验测量结果对比。EH-60L 直升机飞行各时间点的轨迹如图 3 所示。

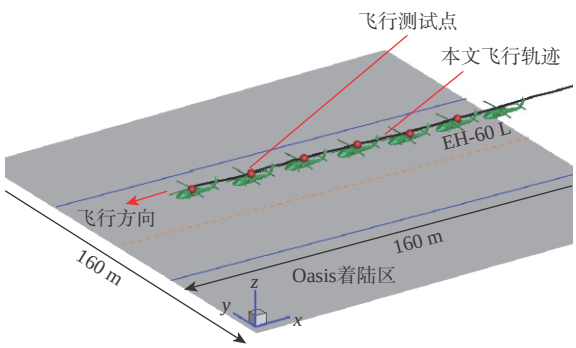


图 3　EH-60L 直升机前飞轨迹

Fig. 3　Trajectory of EH-60L in forward flight

EH-60L 质量、旋翼转速、桨叶数目、半径、弦长分别为 7 394 kg、258 r/min、4、8.18 m、0.527 4 m。旋翼采用 4 800 个面元, 时间步长为 5°。EH-60L 直升机飞行高度为 9.15 m、飞行速度为 37 km/h。地面计算模型覆盖试验测试跑道, 并离散为 10 000 个单元。依据试验测量的沙粒分布设置计算模拟中的沙床沙粒直径分布, 球性沙粒直径为 1～100 μm, 密度为 2 599 kg/m^3, 并随机分布于地面, 沙粒数量为 2.625×10^7。计算得到无侧风下沙云形态如图 4～图 6 所示。

图4～图6中地面扬起的连续面表示计算得到的沙云，而小球为飞行测试结果。各时间点，计算得到沙云俯视图、侧视图、正视图均与飞行测试结果基本一致。前飞状态下，直升机后方两侧卷起的沙云均覆盖飞行测试结果(见图4)，直升机后方沙云高度与飞行测试结果一致(见图5)，且直升机右侧卷起的沙云覆盖飞行测试的大部分观测点(见图6)，由此表明计算的卷起沙云形态与飞行测试结果吻合。此外，计算得到直升机后方不同位置处的沙云均与飞行测试结果重叠(见图4和图5)，由此表明直升机沙盲计算方法较好捕捉沙云形态，且对不同位置沙云具有较好的透视度捕捉能力。直升机前飞状态，旋翼尾迹向下方移动，并与地面干扰形成明显的两侧卷起桨尖涡结构，从而推到沙粒向两侧迁移、堆积、扬起、卷起，形成显著的两侧卷起沙云。旋翼与地面干扰形成弧形流场结构，导致沙云轮廓俯视图也具有明显的弧形特征(见图4(a)和图4(b))，且弧形包络线与飞行测试的观测点基本一致。由于直升机飞行速度较大，沙云形成起始点滞后于直升机旋翼(见图5(a)和图5(b))，因而沙云远离直升机，未阻碍直升机飞行员视野，但在直升机后方两侧，形成显著的卷起沙云(见图6(a)和图6(b))，且沙云形态受到旋翼桨尖涡影响而出现多个卷起结构，并构成起伏的非定常结构(见图6(c))。此外，相比前行侧，后行侧的沙云卷起更为明显(见图6)。

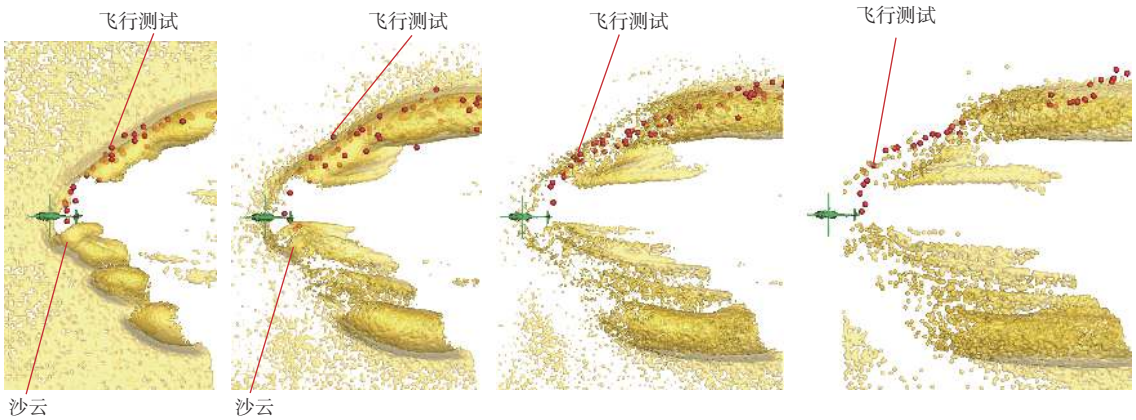


图4 前飞状态的直升机沙云俯视图

Fig. 4 Dust cloud of helicopter at top view in forward flight

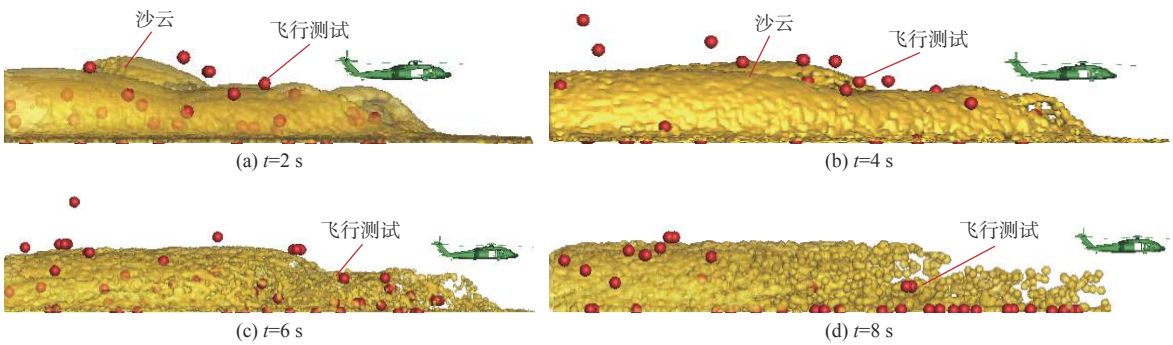


图5 前飞状态的直升机沙云侧视图

Fig. 5 Dust cloud of helicopter at side view in forward flight

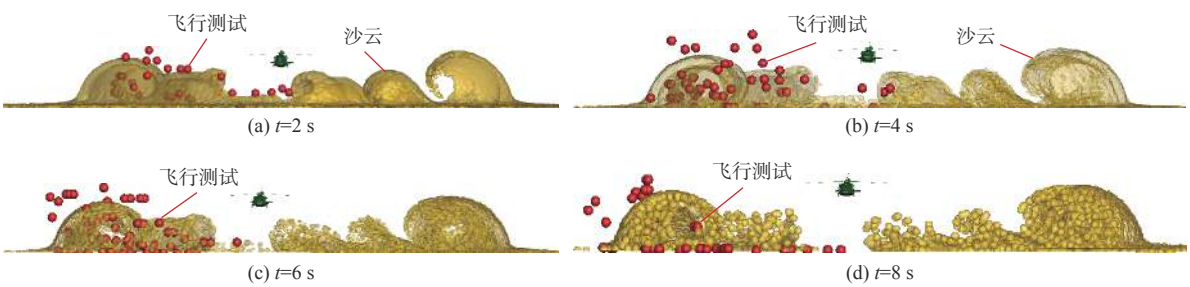
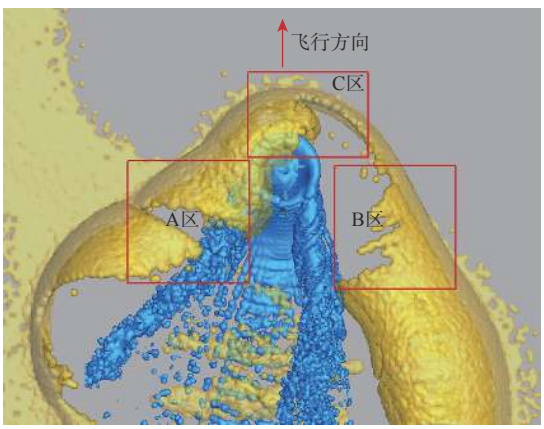


图6 前飞状态的直升机沙云正视图

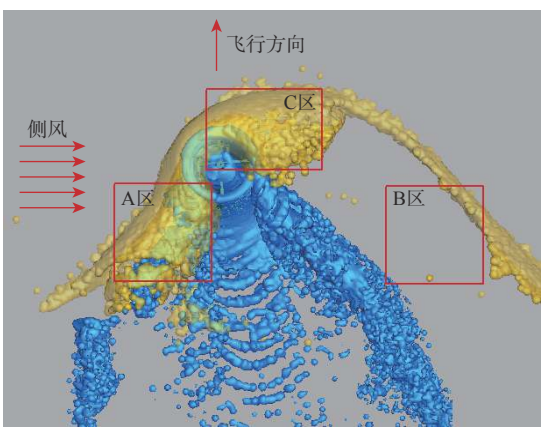
Fig. 6 Dust cloud of helicopter at front view in forward flight

2.2　侧风对直升机沙盲的影响

计算得到无侧风和左侧风状态下直升机沙云形态, 如图 7 和图 8 所示 (左侧风速度为 2.57 m/s), 其中, 旋翼脱出的连续点表示旋翼桨尖涡结构, 地面扬起的连续面表示扬起沙云轮廓。直升机 *xz* 截面和后方流场如图 9 和图 10 所示。由于旋翼桨尖涡诱导外洗流, 推动沙粒向外移动, 从而导致扬起沙云轮廓远离旋翼桨尖涡。此外, 为满足飞行要求, 无侧风和侧风状态下均通过旋翼配平获得旋翼操纵量。



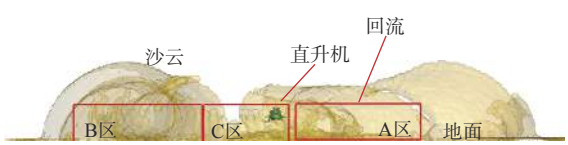
(a) 无侧风



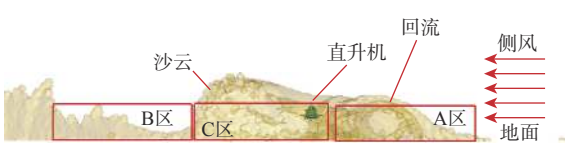
(b) 左侧风

图 7　有无侧风直升机沙云俯视图

Fig. 7　Dust clouds with and without the crosswind at top view



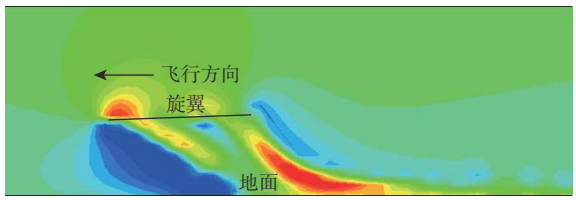
(a) 无侧风



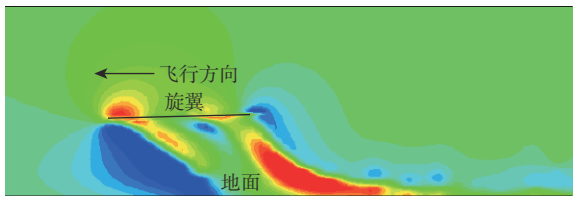
(b) 左侧风

图 8　有无侧风直升机沙云正视图

Fig. 8　Dust clouds with and without the crosswind at front view



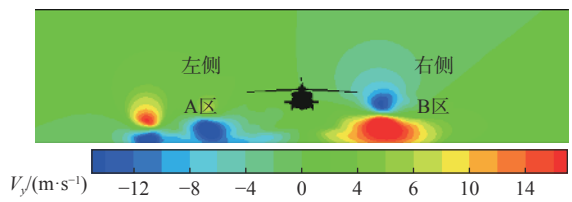
(a) 无侧风



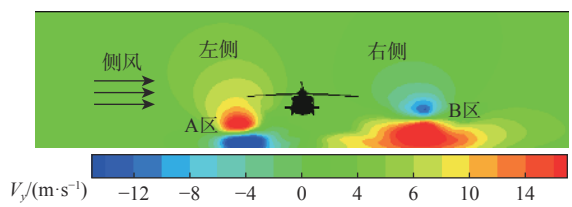
(b) 左侧风

图 9　有无侧风直升机 *xz* 截面流场

Fig. 9　Flow field of helicopter with and without crosswind (*xz* plane)



(a) 无侧风（后视图）



(b) 左侧风（后视图）

图 10　有无侧风直升机后方流场 (x=2R)

Fig. 10　Flow field behind of helicopter with and without crosswind (x=2R)

无侧风状态, 旋翼桨尖涡向后下方移动, 靠近地面, 并向直升机两侧扩张, 最终在直升机两侧形成显著的卷起桨尖涡结构, 诱发两侧外洗流和上洗流。由此, 沙粒受到旋翼流场作用力向直升机两侧迁移、堆积, 并在桨尖涡的诱导下起跳、扬起。在旋翼两侧卷起桨尖涡诱导上洗流作用下, 直径较小 (<50 μm) 的沙粒由于质量较小开始悬浮, 并随流场形成回流现象, 构成明显的两侧卷起沙云结构 (见图 7(a) 和图 8(a))。此外, 直升机正前方的旋翼桨尖涡受到地面干扰和前飞速度影响, 形成弧形地面涡, 并诱发地面射流 (见图 9(a), 其中, V_x 为 x 方向速度), 沙粒扬起、悬浮、回流, 并形成覆盖直升机前半部分的沙云 (见图 7(a))。因此, 卷起沙云高度大

于直升机飞行高度，最终覆盖直升机大部分 (见图 8(a))。

然而，左侧风状态下，旋翼桨尖涡和两侧的卷起桨尖涡向右侧移动 (见图 7(b))。相比于无侧风状态，左侧风的最大外洗速度对应位置点向右移动 (见图 10(b)A 区，其中，V_y 为 y 方向速度)，因而，更加靠近直升机旋翼，且最大外洗速度更大。直升机后行侧的沙粒迁移、堆积、扬起、回流更显著，形成明显沙云，且卷起沙云更靠近直升机旋翼 (见图 8(b)A 区)，卷起沙云半径更小。因此，直升机左侧 A 区的沙云更加明显 (见图 7(b) 和图 8(b))。相比于 A 区，左侧风直升机前行侧 B 区外洗流远离直升机 (见图 10(b))，外洗速度与侧风速度相同，导致沙粒迁移速度更大，沙粒迁移特征更为明显，并远离前行侧卷起桨尖涡，导致前行侧 B 区的沙云远离直升机 (见图 7(b) 和图 8(b))。由于沙粒迁移速度更大，卷起桨尖涡诱导上洗速度更小，因而，沙粒扬起程度减弱，悬浮和回流现象消失 (见图 8(b)B 区)。同时，左侧风也改变直升机正前方沙云形态。直升机正前方 (见图 7(b) C 区) 的扬起沙粒受到左侧风影响向右侧移动，并进入旋翼下洗流，返回地面，形成明显的回流沙云 (见图 7(b)C 区)。因此，相比于无侧风，左侧风导致直升机正前方 C 区沙云更靠近直升机机头，且沙云更为显著 (见图 8(b))，并覆盖直升机驾驶舱左、右两侧，从而完全覆盖驾驶员视野。

2.3 侧风风向对直升机沙盲的影响

风向角 θ 定义如图 11 所示。θ>0°为右侧风状态，而 θ<0°为左侧风状态。计算得到不同侧风风向下直升机飞行员视野沙尘浓度如图 12 所示 (侧风速度$|\boldsymbol{v}_w|$为 2.57 m/s)。

随着侧风风向角增加，沙尘浓度迅速降低，随后略微增加。风向角 θ 为−90°时，即左侧风状态，沙尘浓度最大。主要原因在于左侧风与旋翼后侧行桨尖涡干扰，导致后行侧卷起沙云更靠近直升机。风向角 θ 为 0°时，即迎风飞行时，沙尘浓度最小。主要原因为风向与飞行方向平行，相当于增加飞行速度，增加沙尘向直升机后方移动速度，减小直升机前方扬起沙粒数量。此外，相比于左侧风，右侧风下的沙尘浓度明显更小。主要原因为旋翼前行侧桨尖涡强度小于后行侧，外洗流也更小，因而在右侧风作用下，沙粒扬起程度更小，卷起沙云程度更轻。

不同侧风风向角下的沙云形态如图 13 所示。

当风向角 θ 为−90°时，风向与飞行方向垂直，左侧风导致两侧旋翼卷起桨尖涡向右侧移动 (见图 13(a))，受到后行侧卷起桨尖涡诱导的外洗流作

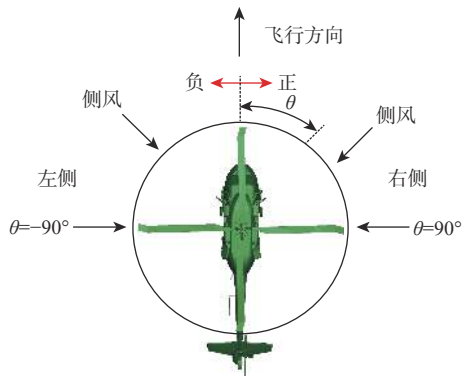


图 11 侧风风向角定义

Fig. 11 Direction of a crosswind angle

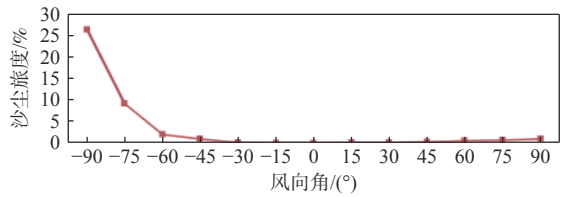


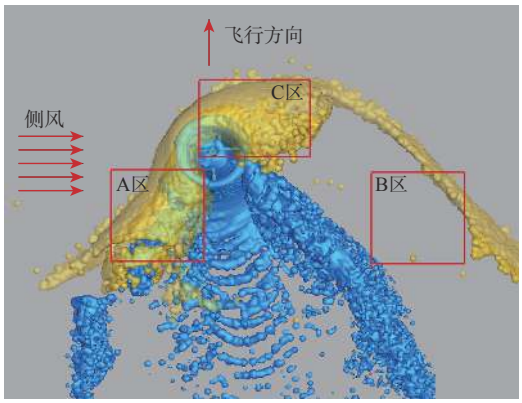
图 12 沙尘浓度随风向角的变化

Fig. 12 Variations of density of dust cloud with crosswinds at different directions

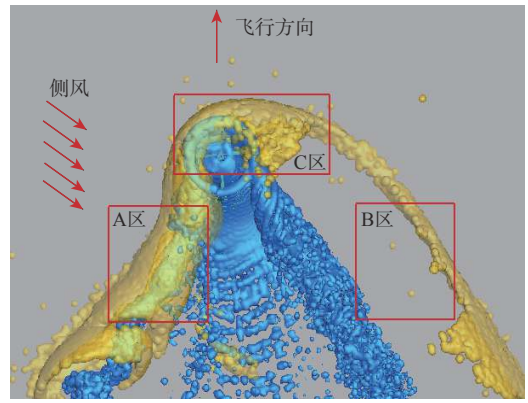
用 (见 A 区)，沙粒向外侧迁移，且由于左侧风方向与外洗流方向相反，导致沙粒沉积与聚集更为严重，并出现更为明显的扬起。由于卷起桨尖涡更靠近旋翼，扬起的沙尘被快速吸入旋翼尾流，重新返回地面，从而形成更为显著的回流沙云 (见图 13(a) 和图 14(a))。然而，前行侧 B 区的外洗流与风向相同，且旋翼桨尖涡向外侧移动，从而导致沙粒向外侧迁移速度更大，扬起更小 (见图 13(a))。因此，相比于前行侧，后行侧沙尘浓度更大。

相比于风向角−90°，风向角 θ 为−60°状态下的 y 方向 (横向) 速度分量减小，x 方向 (纵向) 速度分量增加，因此，直升机后行侧 A 区 (见图 13(b)) 的卷起桨尖涡靠近旋翼速度减小，沙粒扬起和回流程度减小，然而向后方移动速度增加，因此，A 区的卷起沙云减弱，并向直升机左后方延伸 (见图 13(b))。直升机前行侧 B 区，由于横向速度分量减小，右侧卷起桨尖涡与直升机旋翼距离减小，诱导速度增加，沙粒扬起和回流程度增强。因此，相比于风向角−90°，风向角 θ 为−60°状态下后行侧沙尘浓度减小，前行侧沙尘浓度增加 (见图 13(b) 和图 14(b))。此外，由于纵向速度分量增加，直升机前方 C 区 (见图 13(b)) 的卷起桨尖涡远离直升机机头位置，向前诱导射流速度减小，沙粒向后方迁移速度增加，扬起和回流程度减弱 (见图 13(b))。

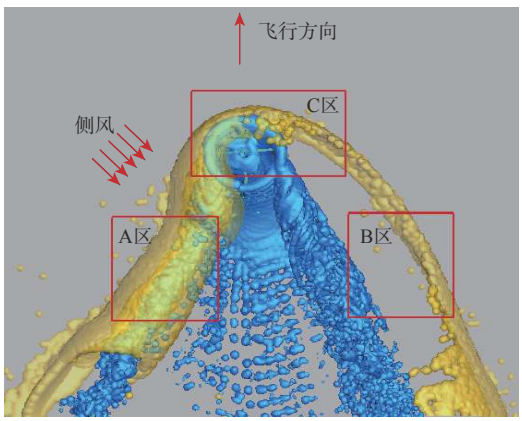
风向角 θ 从−90°增加到 0°，y 方向 (横向) 速度分量减小，而 x 方向 (纵向) 速度分量明显增加，因



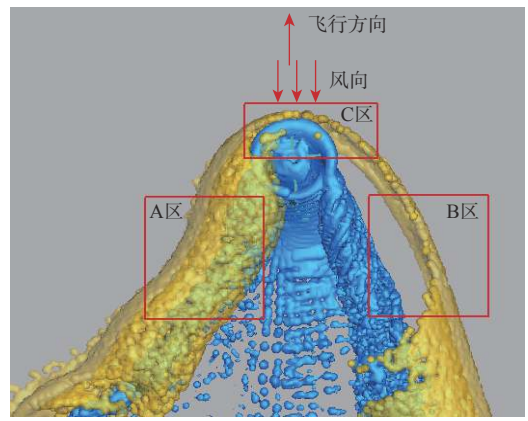
(a) θ=−90°



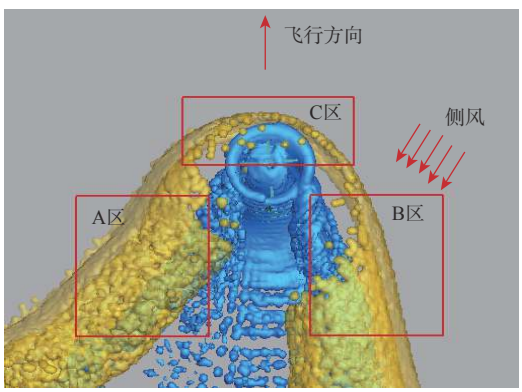
(b) θ=−60°



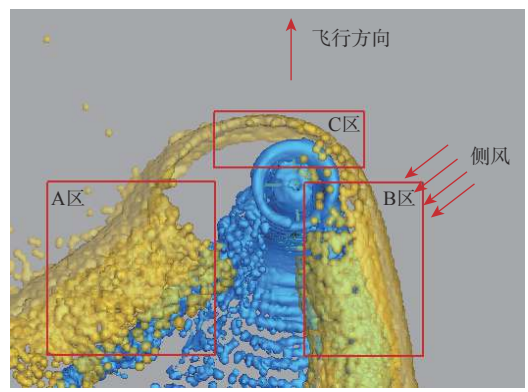
(c) θ=−30°



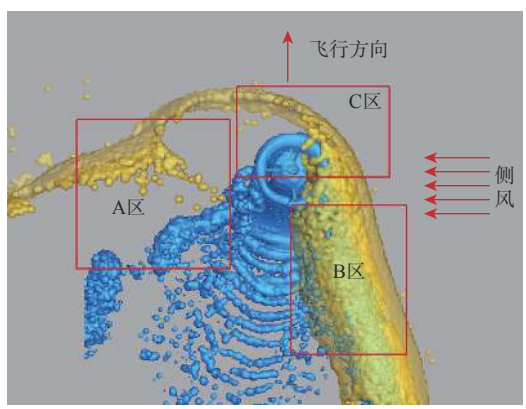
(d) θ=0°



(e) θ=30°



(f) θ=60°



(g) θ=90°

图 13　不同风向角的直升机沙云俯视图

Fig. 13　Dust cloud of helicopter with crosswinds at different directions at top view

此，直升机后行侧A区的扬起沙粒向后方移动，扬起沙粒与卷起桨尖涡和直升机的距离逐渐增加(见图13(a)～图13(d))。但直升机前行侧B区的扬起沙粒与卷起桨尖涡和直升机距离减小，因此，直升机后行侧卷起沙云减弱，前行侧卷起沙云增强(见图13(a)～图13(d)、图14(a)～图14(d))。此外，由于x方向(纵向)速度分量明显增强，沙粒向后方迁移速度增大，直升机前方C区卷起沙云程度减弱。

风向角θ从0°增加到90°，直升机后行侧A区卷起桨尖涡诱导的外洗与风向相同，沙粒向外侧迁移速度增加，扬起沙云与卷起桨尖涡和直升机距离增加(见图13(e)～图13(g))，沙云扬起程度逐渐减弱。前行侧B区诱导外洗速度与风向相反，沙粒向外迁移速度减小，扬起的沙粒向卷起桨尖涡和直升机靠近，并进入旋翼尾流，从而形成显著的卷起沙云(见图13(e)～图13(g)、图14(e)～图14(g))。同时，由于x方向(纵向)速度分量减小，直升机机头方向的沙粒扬起数量增加，沙云程度增加。

风向角θ为−90°时，直升机后行侧出现卷起沙云，而风向角θ为90°时，直升机前行侧出现卷起沙云。由于相同状态下直升机前行侧卷起涡强度小于后行侧，从而导致后行侧外洗和上洗速度更大，因此，风向角−90°的直升机后行侧沙尘浓度大于风向角90°的前行侧沙云，但风向角90°的直升机后行侧沙云扬起高度大于风向角−90°时前行侧沙云扬起高度。

各风向角下沙云正视图如图14所示。

风向角θ从−90°增加到0°，直升机后行侧A区形成卷起沙云，而后行侧B区的沙粒仅扬起。主要原因为风向与B区的外洗方向相同。风向角θ从−90°增加到0°，直升机后行侧A区卷起沙尘浓度减小，而前行侧B区扬起沙尘高度却增加，主要原因为横向速度分量减小。风向角θ从0°增加到90°，直升机后行侧A区的卷起沙云消失，并转变为扬起沙尘，而前行侧B区从扬起沙尘转变为卷起沙云。主要原因为风向改变导致前行侧沙粒与卷起桨尖涡和直升机距离减小，干扰更为显著。直升机后行侧A区扬起沙尘高度减小，而前行侧B区的卷起沙尘浓度增加。相比于左侧风，右侧风的卷起沙尘浓度更小。

2.4 侧风速度对直升机沙盲的影响

侧风风速$|\boldsymbol{v}_w|$分别为1.28, 2.57, 3.85, 5.14 m/s的直升机沙尘浓度如图15和图16所示。

相比于风向角−90°和90°状态，风向角−30°～75°之间的各侧风速度下沙尘浓度均明显减小，主要原因为左侧风与右侧风对直升机沙云形态影响较

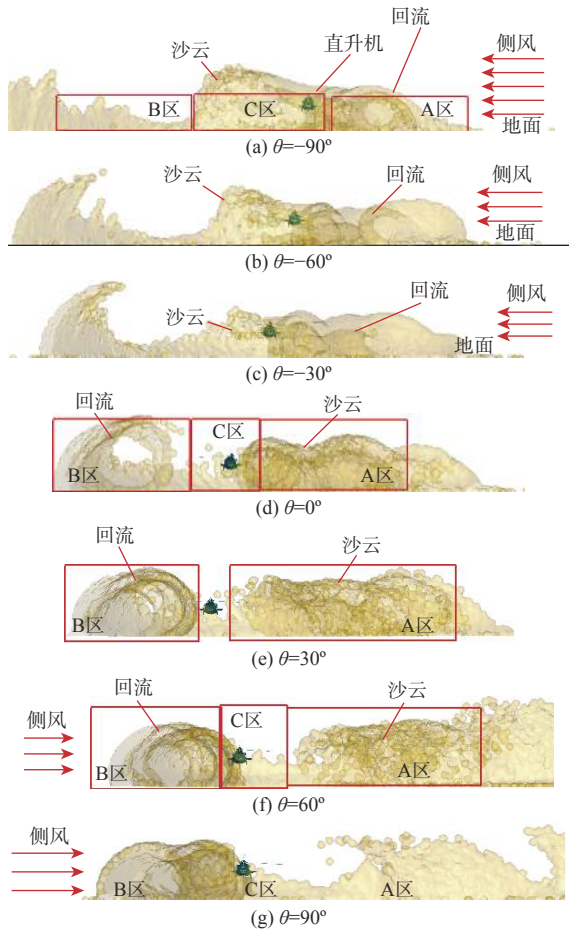


图14 不同风向角的直升机沙云正视图

Fig. 14 Dust cloud of the helicopter with crosswinds at different directions at front view

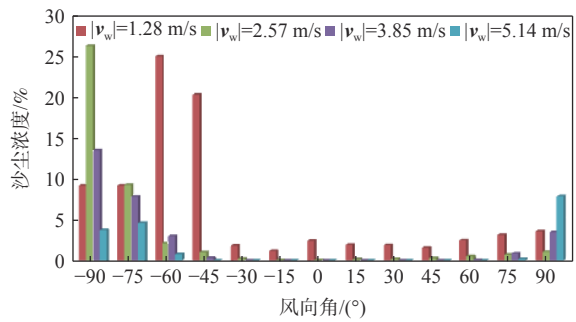


图15 沙尘浓度随风速的变化

Fig. 15 Variations of density of dust cloud at different velocities

大。风速$|\boldsymbol{v}_w|\geqslant$2.57 m/s时，随风向角的增加，沙尘浓度先快速减小，然后基本不变，最后明显增加。但风速$|\boldsymbol{v}_w|$<2.57 m/s时，风向角θ为−75°～−60°，沙尘浓度快速增加，之后随风向角的增加快速减小，最后略有增加。风速$|\boldsymbol{v}_w|$为1.28 m/s和2.57 m/s时的沙尘浓度最大值基本相同，均明显大于$|\boldsymbol{v}_w|$为3.85 m/s和5.14 m/s，但对应的风向角略有不同。此外，左侧风状态下，风速增加，沙尘浓度先增加后减小，由此表明，风速对沙尘浓度具有显著影响。

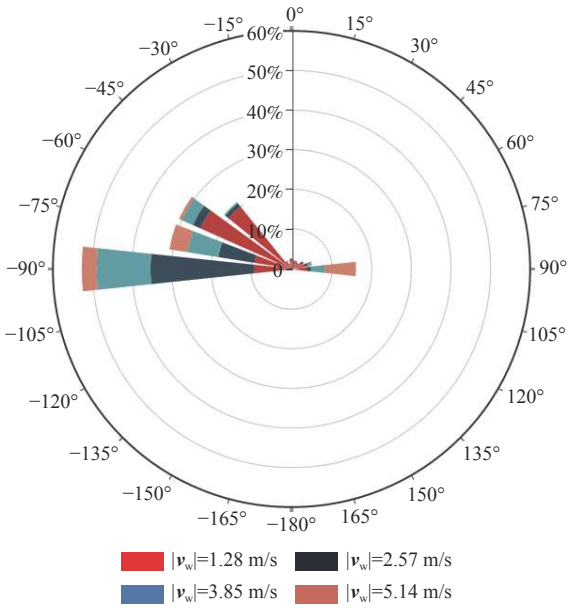


图 16　沙尘浓度雷达图

Fig. 16　Radar map of Density of dust cloud

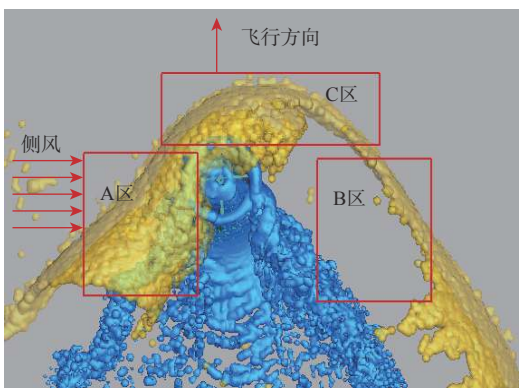
由于左侧风状态下的沙尘浓度较大, 计算得到左侧风不同风速下的沙云形态如图 17 所示。

随着风速的增加, 沙云覆盖直升机的程度先增加后减小, 并与图 15 所示的沙尘浓度规律相同。相比于风速为 1.28 m/s, 风速为 2.57 m/s 状态下的沙云覆盖直升机旋翼区域增大, 且基本覆盖了直升机驾驶舱左右两侧。主要原因为侧风风速 y 方向 (横向) 分量更大, 导致扬起的沙尘与直升机后行侧卷起桨尖涡更近, 回流现象更显著, 因而沙云更贴近直升机驾驶舱。随着风速增加 (3.85 m/s、5.14 m/s), 侧风风速 y 方向 (横向) 分量进一步增大, 沙粒向右侧迁移速度增加, 且扬起的沙粒数量减小, 并快速穿过直升机底部, 从而未进入旋翼尾流中, 沙云形成困难, 覆盖直升机驾驶舱的程度明显减轻。

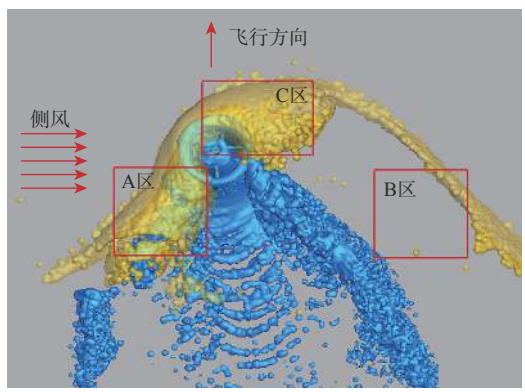
不同风速下的沙云正视图如图 18 所示。

随着风速增加, 直升机前行侧 B 区 (见图 18) 扬起沙云高度明显减小, 且扬起的沙云整体向 B 区移动, 然而, 直升机后行侧 A 区的沙云更靠近直升机驾驶舱, 产生更为显著的回流沙云, 而后随着风速的进一步增大, A 区沙云被侧风吹向 B 区, 且扬起的幅度明显减小, 最终 A 区沙云逐步消失。随着风速增加, 直升机驾驶舱 C 区被沙云包裹, 且包裹程度增加, 随后包裹程度迅速减低, 驾驶员视野逐步清晰。

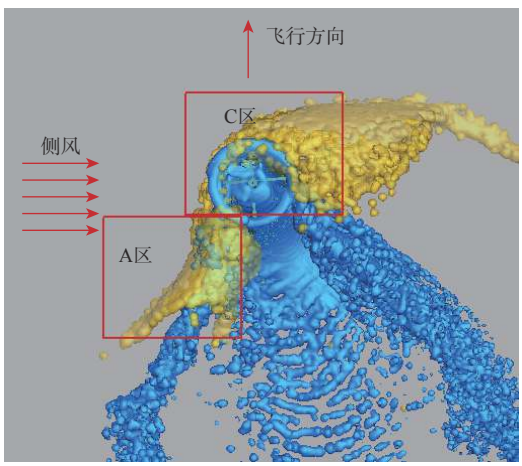
风向角 θ=−60º状态下, 不同风速的沙云正视图



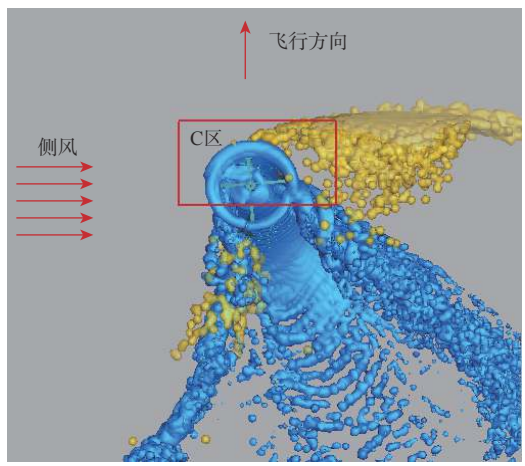
(a) $|v_w|$=1.28 m/s



(b) $|v_w|$=2.57 m/s



(c) $|v_w|$=3.85 m/s



(d) $|v_w|$=5.14 m/s

图 17　不同风速的直升机沙云俯视图 (θ=−90º)

Fig. 17　Dust cloud of helicopter with different velocities at top view (θ=−90º)

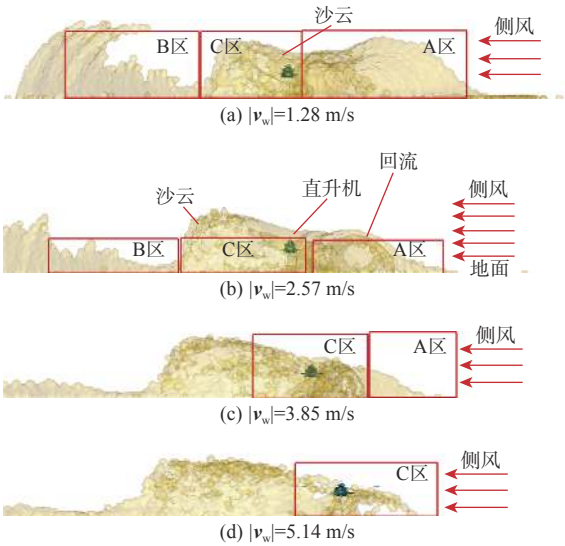


图 18 不同风速的直升机沙云正视图 (θ=−90º)

Fig. 18 Dust cloud of helicopter with different velocities at front view (θ=−90º)

如图 19 所示。

与风向角 θ=−90º不同，沙云程度最大所对应的风速为 1.28m/s。主要原因为 x 方向 (纵向) 存在侧风速度分量，直升机机头前方扬起沙粒受到风速作用，靠近直升机正前方卷起桨尖涡，并被吸入旋翼尾流返回地面，在直升机机头形成回流沙云，导致直升机机头 C 区沙尘浓度增加 (见图 19)。风速增加，侧风速度 x 方向 (纵向) 分量增加，直升机机头方向和后行侧方向沙粒迁移速度增加，扬起沙粒逐渐向直升机后侧和右侧移动，从而减小对直升机的覆盖。因此，直升机后行侧的 A 区沙云高度和强度逐渐减弱，C 区的沙云也逐渐减弱，沙尘浓度逐渐降低。此外，相比于风向角 θ=−90º状态 (见图 18)，

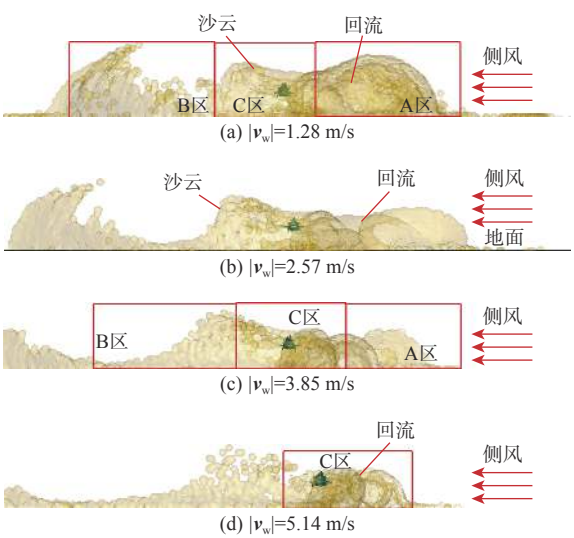


图 19 不同风速的直升机沙云正视图 (θ=−60º)

Fig. 19 Dust cloud of helicopter with different velocities at front view (θ=−60º)

由于侧风速度 x 方向 (纵向) 分量增加，y 方向 (横向) 分量减小，因此，沙云横向移动距离减小，但向后的纵向移动距离增加，导致沙尘浓度明显减弱。由此，飞行方向和风速对直升机沙云影响显著。

3 结 论

1) 将侧风作用项嵌入旋翼/地面气动干扰模型和沙粒动力学模型，建立侧风状态下直升机沙盲分析方法，计算得到全尺寸直升机前飞状态沙云形态俯视图、侧视图、正视图与飞行测试结果基本一致。

2) 相比于无侧风状态，左侧风下的旋翼桨尖涡、两侧卷起桨尖涡、前方弧形地面涡向右侧移动，直升机后行侧和正前方的沙粒迁移、堆积、扬起、回流，并靠近直升机，形成半径更小、浓度更大的卷起沙云；直升机前行侧外洗速度与侧风速度相同，沙粒迁移速度增大，沙云扬起程度减弱，并远离直升机，卷起现象消失。

3) 风向角增加，沙尘浓度迅速降低，随后略微增加。左侧风状态的沙尘浓度最大，迎风飞行状态，风向与飞行方向平行，沙尘向后方迁移速度增强，沙尘浓度最小。由于旋翼前行侧桨尖涡强度小于后行侧，相比于左侧风，右侧风下的沙尘浓度明显更小。

4) 风向角从−90º增加到 0º，直升机后行侧卷起沙尘浓度减小，而前行侧扬起沙尘高度增加；风向角从 0º增加到 90º，直升机后行侧卷起沙云消失，扬起沙尘高度减小，而前行侧从扬起沙尘转变为卷起沙云，且浓度增加。

5) 随着侧风速度的增加，左侧风状态下的沙尘浓度先增加后减小。侧风风速大于 2.57 m/s 时，风向角增加，沙尘浓度先快速减小，然后增加；风速小于 2.57 m/s 时，风向角为−75º～−60º，沙尘浓度快速增加，且随风向角增加而快速减小，然后增加。

参考文献 (References)

[1] SZOBOSZLAY Z, DAVIS B, FUJIZAWA B T, et al. Degraded visual environment mitigation (DVE-M) program, Yuma 2016 flight trials in brownout[C]//Proceedings of the AHS International 73rd Annual Forum&Technology Display. Alexandria: AHS, 2017: 1-20.

[2] MILLER J, GODFROY-COOPER M, SZOBOSZLAY Z. Degraded visual environment mitigation (DVE-M) program, bumper RADAR obstacle cueing flight trials 2020[C]//Proceedings of the Vertical Flight Society's 77th Annual Forum& Technology Display. Alexandria: AHS, 2021: 16747.

[3] WACHSPRESS D A, WHITEHOUSE G R, KELLER J D, et al. A high fidelity brownout model for real-time flight simulations and

trainers[C]//Proceedings of the American Helicopter Society 65th Annual Forum. Alexandria: AHS, 2009: 1281-1304.

[4] WHITEHOUSE G R, WACHSPRESS D A, QUACKENBUSH T R, et al. Exploring aerodynamic methods for mitigating brownout [C]//Proceedings of the American Helicopter Society 65th Annual Forum. Alexandria: AHS, 2009: 349-364.

[5] JOHNSON B, LEISHMAN J G, SYDNEY A. Investigation of sediment entrainment using dual-phase, high-speed particle image velocimetry[J]. Journal of the American Helicopter Society, 2010, 55(4): 1-16.

[6] SYDNEY A, LEISHMAN J G. Measurements of rotor/airframe interactions in ground effect over a sediment bed[J]. Annual Forum Proceedings - AHS International, 2013, 3: 1812-1836.

[7] WONG O D, TANNER P E. Photogrammetric measurements of an EH-60L brownout cloud[J]. Journal of the American Helicopter Society, 2016, 61(1): 1-10.

[8] RYERSON C, HAEHNEL R, KOENIG G, et al. Visibility enhancement in rotorwash clouds[C]//Proceedings of the 43rd AIAA Aerospace Sciences Meeting and Exhibit. Reston: AIAA, 2005: 263.

[9] HAEHNEL R B, MOULTON M A, WENREN W, et al. A model to simulate rotorcraft-induced brownout[C]//Proceedings of the 64th Annual Forum of the American Helicopter Society. Alexandria: AHS, 2008: 589-601.

[10] THOMAS S, LAKSHMINARAYAN V K, KALRA T S, et al. Eulerian-Lagrangian analysis of cloud evolution using CFD coupled with a sediment tracking algorithm[J]. Annual Forum Proceedings - AHS International, 2011, 1: 298-315.

[11] THOMAS S, AMIRAUS M, BAEDER J. GPU-accelerated FVM-RANS hybrid solver for simulating two-phase flow beneath a hovering rotor[C]//Proceedings of the 69th Annual Forum of the American Helicopter Society. Alexandria: AHS, 2013: 2462-2484.

[12] KUTZ B M, GUNTHER T, RUMPF A, et al. Numerical examination of a model rotor in brownout conditions[C]//Proceedings of the 70th Annual Forum of the American Helicopter Society. Montreal: Engineering, 2014: 2450-2461.

[13] 胡健平, 徐国华, 史勇杰, 等. 基于 CFD-DEM 耦合数值模拟的全尺寸直升机沙盲形成机理[J]. 航空学报, 2020, 41(3): 123363.
HU J P, XU G H, SHI Y J, et al. Formation mechanism of brownout in full-scale helicopter based on CFD-DEM couplings numerical simulation[J]. Acta Aeronautica et Astronautica Sinica, 2020, 41(3): 123363 (in Chinese).

[14] GOVINDARAJAN B M, LEISHMAN J G. Predictions of rotor and rotor/airframe configurational effects on brownout dust clouds[J]. Journal of Aircraft, 2016, 53(2): 545-560.

[15] 谭剑锋, 何龙, 于领军, 等. 基于黏性涡粒子/沙粒 DEM 的直升机沙盲建模[J]. 航空学报, 2022, 43(8): 125536.
TAN J F, HE L, YU L J, et al. Helicopter brownout modeling based on viscous vortex particle and sand particle DEM[J]. Acta Aeronautica et Astronautica Sinica, 2022, 43(8): 125536 (in Chinese).

[16] 谭剑锋, 韩水, 王畅, 等. 基于 DEM 的直升机沙盲加速计算方法[J]. 北京航空航天大学学报, 2023, 49(6): 1352-1361.
TAN J F, HAN S, WANG C, et al. Accelerated computational method of helicopter brownout based on DEM[J]. Journal of Beijing University of Aeronautics and Astronautics, 2023, 49(6): 1352-1361 (in Chinese).

[17] TAN J, GAO J W, BARAKOS G, et al. Novel approach to helicopter brownout based on vortex and discrete element methods[J]. Aerospace Science and Technology, 2021, 116: 106839.

[18] TAN J F, YON T, HE L, et al. Accelerated method of helicopter brownout with particle-particle collisions[J]. Aerospace Science and Technology, 2022, 124: 107511.

[19] 谭剑锋, 周天熠, 王畅, 等. 旋翼地面效应的气动建模与特性[J]. 航空学报, 2019, 40(6): 122602.
TAN J F, ZHOU T Y, WANG C, et al. Aerodynamic model and characteristics of rotor in ground effect[J]. Acta Aeronautica et Astronautica Sinica, 2019, 40(6): 122602 (in Chinese).

Influence of crosswind on helicopter brownout

TAN Jianfeng[1,*], YANG Yuxiao[1], ZHANG Weiguo[2], WANG Shixiong[3], LYU Lin[3]

(1. School of Mechanical and Power Engineering, Nanjing Tech University, Nanjing 211816, China;
2. Low-speed Institute, China Aerodynamics Research and Development Center, Mianyang 621000, China;
3. The third Training Brigade, Army Aviation Institute, Linfen 041000, China)

Abstract: The complex interaction flow between the helicopter rotor and ground induced the movement of sand particles in the sand bed and then yielded helicopter brownout. However, a crosswind will change the rotor-ground interaction flow field, and influence the brownout. Therefore, a crosswind is coupled into the helicopter rotor-ground aerodynamic model based on viscous vortex particles to consider the influence of the crosswind on the rotor-ground interaction flow field. In order to take into account how the crosswind affects sand movement, the air force caused by the crosswind is also incorporated to the sand particle dynamic model based on DEM. Then, the effect of the crosswind on the process of helicopter brownout at forward flight is analyzed, and the influences of velocity, and direction of the crosswind on the brownout are also investigated. Results show that the crosswind has a significant influence on the helicopter brownout, the density of dust cloud obviously increases with the crosswind at the port side, and it is restrained at forward flight. The density of the dust cloud firstly increases and then decreases with increasing the velocity of crosswind at the port side.

Keywords: crosswind; brownout; rotor-ground aerodynamic interaction; discrete element method; helicopter

Received: 2022-09-18; **Accepted**: 2022-11-14; **Published Online**: 2023-01-17 13:26

URL: link.cnki.net/urlid/11.2625.V.20230116.1731.002

Foundation items: National Natural Science Foundation of China (12172165); Natural Science Foundation of Jiangsu Province (BK20211259); Outstanding Young Backbone Teacher Project of Jiangsu Qinglan Project

* **Corresponding author**. E-mail: Jianfengtan@njtech.edu.cn

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2022.0764

采用温度梯度的干燥反应器 H_2O 穿透深度评估方法

张勇平，卞强*，杨润泽，张震，白攀峰
(中国航天员科研训练中心 人因工程重点实验室，北京 100094)

摘 要：针对空间站 CO_2 去除系统中干燥反应器对 H_2O 穿透深度缺乏直接评估手段的问题，提出利用干燥反应器内部温度传感器的温度梯度特性来间接评估的方法；根据硅胶材料吸附 H_2O 放热、脱附 H_2O 吸热的物理特性，结合温度传感器在干燥反应器中的深度位置，分析了不同边界条件下温度传感器数据曲线在吸附及解吸周期内的变化特性，提出反映干燥反应器 H_2O 穿透深度的指标集和评估方法；并对所提方法在系统的密闭舱试验中进行了验证，验证结果表明了指标集的合理性和所提方法的有效性。

关 键 词：空间站；CO_2 去除系统；反应器；吸附；解吸；水汽穿透深度
中图分类号：V11; TQ051.1
文献标志码：A **文章编号**：1001-5965(2024)10-3123-08

CO_2 去除系统是空间站环控生保分系统中一个重要的再生生保系统，其功能是去除舱内由人体代谢产生的 CO_2，维持 CO_2 分压在医学要求范围内。空间站 CO_2 去除系统中的干燥反应器与工业上很多反应器有相似和区别之处，很多文献把反应器看作一个黑盒子，利用输入输出及部分传感器参数，采用神经网络、支持向量机等方法进行故障诊断[1-7]。中国空间站正处于建造阶段，对于空间站 CO_2 去除系统故障诊断的文献较少，从可查阅的文献来看，有关于硅胶材料吸附性能衰减及与分子筛共吸附[8-9]等相关的文献，有关于 CO_2 去除系统状态综合评估[6]的文献，鲜有关于国际空间站 CO_2 去除系统干燥反应器穿透深度评估方面的文章。空间站 CO_2 去除系统故障会导致舱内 CO_2 分压升高，使用非再生 CO_2 去除设备会使得上行成本大幅增加，无法满足有人长期驻留任务需求。本文仅对干燥反应器的水汽穿透深度问题进行研究，从硅胶材料吸附 H_2O 的特性分析入手，对吸附周期和解吸周期内干燥反应器内部不同区域的硅胶温度变化规律进行研究，建立了基于温度梯度的干燥反应器内部 H_2O 穿透深度模型，提出水气穿透深度的评估方法。其重要意义不仅在于当前边界条件变化时，对干燥反应器的 H_2O 穿透深度进行评估，更重要的是可以结合进气温度变化情况，对硅胶吸附性能下降后干燥反应器的 H_2O 穿透深度进行评估，提前识别出干燥反应器性能衰减问题和舱内 H_2O 损失问题，为空间站 CO_2 去除系统故障诊断及在轨运营的备件生产、上行及更换提供决策依据。

1 空间站 CO_2 去除系统组成及原理

中国空间站 CO_2 去除系统主要由 2 个干燥反应器、2 个吸附反应器、风机、热交换器、切换阀及管路等组成，工作原理是利用吸附反应器内 5A 分子筛对 CO_2 气体的选择吸附特性进行吸附，利用热真空条件实现分子筛材料的解吸再生。系统组成及运行流程如图 1 所示，运行流程主要分 2 个阶段，阶段 1 中(见图 1)，气体从干燥反应器 B 进入，由于硅胶吸附放热，空气温度升高，干燥空气经过风机和热交换器后变成低温低湿气体(热交换器冷边连接舱内低温液路回路)，进入吸附反应器 A，实

收稿日期：2022-09-07；录用日期：2023-03-06；网络出版时间：2023-05-10 16:56
网络出版地址：link.cnki.net/urlid/11.2625.V.20230510.1356.004
*通信作者. E-mail：13810213687@139.com

引用格式：张勇平，卞强，杨润泽，等. 采用温度梯度的干燥反应器 H_2O 穿透深度评估方法 [J]. 北京航空航天大学学报，2024，50（10）：3123-3130.
ZHANG Y P，BIAN Q，YANG R Z，et al. Evaluation method of H_2O penetration depth of drying reactor based on temperature gradient [J]. Journal of Beijing University of Aeronautics and Astronautics，2024，50（10）：3123-3130 （in Chinese）.

现 CO_2 气体的吸附，吸附反应器 A 在前 1 个热真空解吸周期内存储的热量被吹入干燥反应器 A，实现干燥反应器 A 内部硅胶吸附 H_2O 性能的再生，吸附反应器温度降低后开始吸附 CO_2 气体；阶段 2 中，气体从干燥反应器 A 中进入，经过风机、热交换器、吸附反应器 B 和干燥反应器 B；2 个阶段交替运行，实现 CO_2 气体的连续吸附。干燥反应器内装填有硅胶，用以吸附系统进气中的水汽，满足进入吸附反应器内气体低露点要求[10]。由于 H_2O 分子的极性比 CO_2 分子的极性强，5A 分子筛材料对于 H_2O 分子的吸附力远大于对 CO_2 分子的吸附力，分子筛对 H_2O 和 CO_2 分子存在竞争吸附问题[11]。因此，进入吸附反应器的 H_2O 一方面会降低分子筛对 CO_2 的吸附性能，另一方面会随着吸附反应器热真空解吸排至外太空，影响系统的 H_2O 损失量。由于干燥反应器出口无露点传感器，因此，在系统进气温度和湿度变化及硅胶性能下降的情况下，如何评估干燥反应器内部的 H_2O 穿透深度，保证干燥反应器出口气体露点在规定值以下是中国空间站 CO_2 去除系统故障诊断中的一个棘手问题。

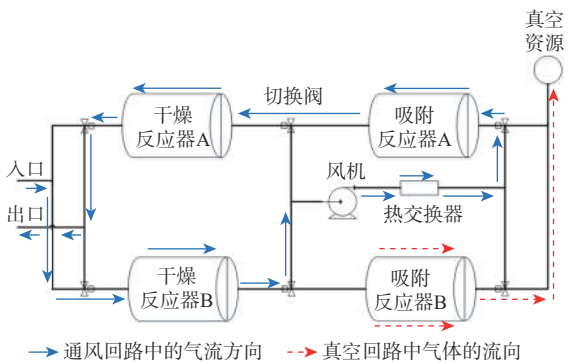


图 1 CO_2 去除系统组成及运行流程

Fig. 1 Composition and operation flow of CO_2 removal system

2 硅胶吸附及脱附 H_2O 特性

空间站 CO_2 去除系统干燥反应器内部装填有无机型硅胶，这是一种坚硬、无定形链状和网状结构的硅酸聚合物颗粒，由无定形的 SiO_2 组成，内部为纳米级微孔结构，具有很大的比表面积和孔容，其表面存在大量羟基，对于 H_2O 具有很强的吸附能力。

硅胶对 H_2O 的吸附作用来自于硅胶表面与 H_2O 分子之间的范德华力，是一个自发物理变化过程。由于 H_2O 分子吸附在硅胶表面，由三维运动变为二维运动，H_2O 气分子由自由运动状态到在吸附剂表面定向排列，其混乱度减小，是一个典型的熵减小过程[12-13]。由热力学中吉布斯函数可知，$G=H-TS$（G 为吉布斯自由能，T 为温度，S 为熵，H 为焓），等温下有公式：$\Delta G=\Delta H-T\Delta S$，由于此过程的 $\Delta G<0$，即 $\Delta H-T\Delta S\leqslant 0$，于是 $\Delta H\leqslant T\Delta S<0$，所以硅胶吸附 H_2O 的过程是一个放热过程[13]。

硅胶吸附 H_2O 分子后，在热风吹扫下，H_2O 分子获得足够的能量后从硅胶表面脱附。H_2O 分子从硅胶表面的定向排列到三维运动，其混乱度增加，是一个熵值增加的过程[14-15]。根据热力学中的吉布斯函数，$G=H-TS$，等温下有公式：$\Delta G=\Delta H-T\Delta S$，由于此过程 $\Delta G>0$，即 $\Delta H-T\Delta S\geqslant 0$，于是 $\Delta H\geqslant T\Delta S>0$，所以 H_2O 从硅胶表面脱附的过程是一个吸热过程[16-19]。

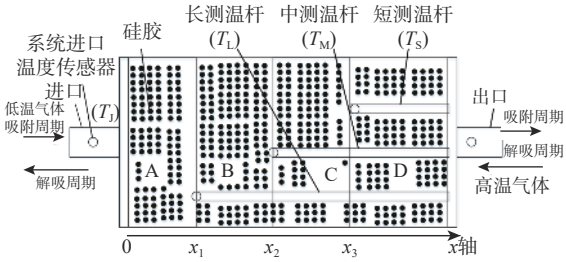
硅胶表面覆盖有自由硅羟基、粒子间缔合硅羟基、表面缔合硅羟基和内部硅羟基等各种硅羟基（官能团），长期高低温交变工况及外界杂质吸附后，硅胶内部部分硅羟基氢键断裂，硅羟基种类发生变化，数量下降，导致硅胶吸附性下降，同时产生的粉末也会影响吸附性能。此外，硅胶内部毛细凝结水会对吸附空间造成挤压，减小自由水的吸附量，化学水影响硅胶的表面特性，这些因素也会导致硅胶吸附性能下降[20]。

根据硅胶吸附 H_2O 放热和脱附 H_2O 吸热的特性，对干燥反应器内部硅胶吸附 H_2O 的升温过程及脱附 H_2O 吸热过程进行分析，建立 H_2O 穿透深度评估方法及判据。

3 吸附周期内 H_2O 穿透深度评估

3.1 吸附周期内硅胶升温趋势分析及 H_2O 穿透深度

为提高吸附容量，防止 H_2O 进入下游吸附反应器，中国空间站 CO_2 去除系统中的干燥反应器采用内部气流轴向流通的设计方案，内部设计有短、中、长 3 根测温杆（测温杆主体为不锈钢材质，温度传感器类型为 PT1000 铂电阻，测量范围 0～350 ℃，测量精度高于±2 ℃）。沿轴向，长、中、短测温杆布置位置如图 2 所示，在轴向断面截面上，3 根测温杆装配在圆中心位置。干燥反应器内部填充有硅胶材料，在干燥反应器进口和出口附近，由于直径突变，实际流场比较复杂，靠近端盖边缘的硅胶吸附 H_2O 的速率和中间部分差异较大。除两端外，反应器中间绝大部分区域的硅胶吸附 H_2O 的趋势是可以用沿反应器轴向断面逐层吸附的模型来近似。并且，短、中、长 3 根测温杆均布置在轴线附近，用该点附近的硅胶吸附和解吸 H_2O 的情况来代表沿轴向断面处的硅胶具有一定代表性。沿气流方向，从进口到出口，对硅胶装填区进行区域划分，划分为 A、B、C、D 共 4 个区域，如图 2 所示。



(a) 干燥反应器轴向剖面



(b) 干燥反应器径向剖面

图 2　干燥反应器测温杆布局及内部区域划分示意图

Fig. 2　Scheme diagram of temperature measuring rod layout and internal area division of drying reactor

在干燥反应器吸附周期, 气流从进口进入, 依次经过硅胶材料 A、B、C、D 区域, 最后, 从出口流出。吸附周期内, 低温气流从进口进入干燥反应器, 在一个吸附周期内, H_2O 穿透深度定义为Ⅰ、Ⅱ、Ⅲ、Ⅳ共计 4 级, Ⅰ和Ⅱ的定义及分析如下。

1) Ⅰ: H_2O 未穿透区域 A。区域 A 的硅胶吸附 H_2O 并放出热量, 热量随着气流被带入下游区域, 导致干燥短、中、长测温杆与系统进气温度差值持续上升。

2) Ⅱ: H_2O 到达区域 B, 未达区域 C。区域 A 的硅胶吸附 H_2O 达到饱和状态, 吸附放热反应停止; 区域 B 的硅胶开始吸附 H_2O 并放出热量, 热量随着气流被带入下游区域, 导致干燥长测温杆与系统进气温度的差值出现下降趋势, 干燥中、短测温杆与系统进气温度差值持续上升。

同理, Ⅲ和Ⅳ的穿透深度定义如表 1 所示。

表 1　干燥周期 H_2O 穿透深度定义（Ⅲ、Ⅳ）

Table 1　Definition of H_2O penetration depth in drying cycle (Ⅲ and Ⅳ)

穿透深度	区域A	区域B	区域C	区域D
III	穿透	穿透	到达, 未穿透	未到达
IV	穿透	穿透	穿透	到达, 是否穿透未知

3.2　吸附周期内 H_2O 气穿透深度评估方法及判据

根据 3.1 节分析, 建立吸附周期内干燥反应器内部 H_2O 穿透深度评估判据如下:

假设吸附周期内, 系统进气温度、干燥长、中、短测温杆温度分别为 T_J、T_L、T_M、T_S, 吸附时长 i 时刻, 上述温度之间差值定义为 $(T_L-T_J)_i$、$(T_M-T_L)_i$、$(T_S-T_M)_i$; 0～T_i 期间, 上述温度之间差值最大值定义为 $(T_L-T_J)_{max}$、$(T_M-T_L)_{max}$、$(T_S-T_M)_{max}$, 令

$$\nabla T_1=(T_L-T_J)_i-(T_L-T_J)_{max} \quad (1)$$

$$\nabla T_2=(T_M-T_L)_i-(T_M-T_L)_{max} \quad (2)$$

$$\nabla T_3=(T_S-T_M)_i-(T_S-T_M)_{max} \quad (3)$$

当 $\nabla T_1\geqslant 0$ 时, H_2O 未穿透区域 A, 穿透深度为Ⅰ级; 当 $\nabla T_1<0$ 时, 判断周期 i 时刻的 $(T_L-T_J)_i$ 的值, 稳定工况下, 由于系统外部环境温度比进气温度高, $(T_L-T_J)_i$ 会稳定在某一个略大于 0 的正值, 如果系统进气温度变化, 该值未有明显变化, 说明 H_2O 已经穿透区域 A, 进入区域 B, 穿透深度为Ⅱ级(或以上)。

同理, 根据上述分析, 可以确定 H_2O 穿透深度为Ⅱ级、Ⅲ级(或以上)的判据如表 2 所示。

表 2　吸附周期 H_2O 穿透深度判据

Table 2　H_2O penetration depth criteria in adsorption cycle

穿透深度	判据	备注
Ⅱ	$\nabla T_1<0, \nabla T_2\geqslant 0$	H_2O进入但未穿透区域B
Ⅲ(或以上)	$\nabla T_1<0, \nabla T_2<0$	H_2O进入区域C
Ⅲ	$\nabla T_1<0, \nabla T_2<0, \nabla T_3=0$	H_2O进入但未穿透区域C
Ⅳ(或级以上)	$\nabla T_1<0, \nabla T_2<0, \nabla T_3<0$	H_2O穿透区域C进入区域D

表 2 评估判据可以评估 H_2O 是否穿透区域 A、B、C, 但是当 H_2O 穿透区域 C 后, 无法评估 H_2O 是否穿透反应器。

4　解吸周期内 H_2O 穿透深度评估

4.1　解吸周期内硅胶升温趋势

空间站 CO_2 去除系统干燥反应器的解吸再生是采用热风吹扫方式实现的。解吸周期内, 在风机作用下, 冷空气经过吸附反应器内部高温状态分子筛加热后吹扫到干燥反应器内部, 热空气从干燥反应器出口进入, 从干燥反应器进口吹出, 气流示意图如图 1 所示。到达干燥反应器气流的温度变化规律是先上升到达最高温度, 后逐步下降的过程。

热空气首先到达区域 D 的硅胶材料, 该区域内硅胶在热空气作用下开始脱附 H_2O 分子, 硅胶材料本体的温度开始升高, 同时由于脱附吸热的物理反应及硅胶材料本身的吸热过程, 热空气的温度逐步下降。在每一个解吸周期内, 短测温杆的最高温度与气流上游吸附反应器内部的最高温度、热容及区域 D 内硅胶材料吸附 H_2O 的量有关。

热空气到达区域 C 后, 该区域内硅胶材料的吸热、升温机理同上所述。在每一个解吸周期内, 中

测温杆的最高温度与气流上游吸附反应器内部的最高温度、热容及区域C和区域D内硅胶材料吸附H_2O的量有关，其值比短测温杆的最高温度低。

热空气到达区域B后，该区域内硅胶材料的吸热、升温机理同上所述。在每一个解吸周期内，长测温杆的最高温度与气流上游吸附反应器内部的最高温度、热容及区域B、区域C和区域D内硅胶材料吸附H_2O的量有关，其值比中、短测温杆的最高温度都要低。

降温后的热空气经过区域A后，由于区域A中硅胶材料本体的吸热和H_2O脱附反应吸收热量，进一步降温后的气流从干燥反应器的进口吹出。由于解吸周期内，在干燥反应器气流出口无温度传感器，因此，在解吸周期内，从短、中、长测温杆的最高温度是无法衡量区域A中硅胶材料在吸附周期内吸附的H_2O的总量。

4.2 解吸周期内H_2O穿透深度评估方法及判据

假设解吸周期内长、中、短测温杆的最高温度分别为T_{Lmax}、T_{Mmax}和T_{Smax}，当系统进气温度和湿度、吸附反应器的热真空解吸温度稳定时，这3项最高温度会稳定在T_{LZmax}、T_{MZmax}和T_{SZmax}上。当系统进气温度和露点发生变化时，随着干燥反应器内部H_2O穿透深度的变化，T_{Lmax}、T_{Mmax}和T_{Smax}值会发生变化。

根据本节解吸周期中短、中、长测温杆最高温度与不同区域内硅胶材料脱附H_2O的反应程度之间的关系，建立解吸周期内干燥反应器中H_2O穿透深度的温度评估判据，系统进气温度、露点变化或硅胶吸附性能下降时，H_2O是否穿透区域A判据判据为：T_{Lmax}、T_{Mmax}和T_{Smax}均不变，说明H_2O未穿透区域A，穿透深度为Ⅰ级；无法评估区域A中的H_2O穿透深度。

H_2O是否穿透区域B、区域C、区域D的判据如表3所示。

表3 解吸周期H_2O穿透深度判据

Table 3 H_2O penetration depth criteria in desorption cycle

穿透深度	判据	备注
Ⅱ	T_{Lmax}波动，T_{Mmax}和T_{Smax}不变	$T_{Lmax}>T_{LZmax}$，穿透深度变浅 $T_{Lmax}<T_{LZmax}$，穿透深度变深
Ⅲ	T_{Lmax}波动，T_{Mmax}波动，T_{Smax}不变	$T_{Mmax}>T_{MZmax}$，穿透深度变浅； $T_{Mmax}<T_{MZmax}$，穿透深度变深
Ⅳ（或以上）	T_{Lmax}波动，T_{Mmax}波动，T_{Smax}波动	$T_{Smax}>T_{SZmax}$，穿透深度变浅； $T_{Smax}<T_{SZmax}$，穿透深度变深

表3中解吸周期内Ⅳ（或以上）的判据中，无法直接判断H_2O是否已经穿透区域D（需要对比H_2O完全穿透干燥反应器的数据才能判断H_2O是否已经穿透区域D）。

5 H_2O穿透深度综合评估方法及判据

综合吸附和解吸周期内的H_2O穿透深度评估判据可知，吸附周期内，基于不等长测温杆温差判断的评估判据可以评估H_2O是否进入或穿透区域A、区域B、区域C，无法评估H_2O是否进入或穿透区域D。解吸周期内，基于测温杆最高温度判断的评估判据可以评估H_2O是否穿透区域B、区域C及是否进入区域D，无法直接评估H_2O是否穿透区域D。综合吸附和解吸周期内的评估判据，就可以评估H_2O是否进入区域A、区域B、区域C、区域D，是否穿透区域A、区域B、区域C。由于区域D位于干燥反应器的出口最末端，当前评估判据还无法判断H_2O是否已经穿透该区域，还需要下游吸附反应器温度等其他参数进行综合判断，如表4所示。

表4 吸附及解吸周期H_2O进入及穿透深度评估方法汇总

Table 4 Summary of evaluation methods for H_2O entry and penetration depth in adsorption and desorption cycles

模式	进入				穿透			
	A	B	C	D	A	B	C	D
吸附	√	√	√	×	√	√	√	×
解吸	√	√	√	√	√	√	√	×

6 试验验证

空间站地面密闭舱内CO_2去除系统验证试验长期运行后，干燥反应器内硅胶的吸附性能会发生衰减。在相同边界条件下，其内部H_2O穿透深度会增加。由于短期内无法真实模拟出硅胶吸附性能实际衰减的情况，因此，通过调整系统进气温度来间接模拟硅胶吸附性能变化，进而创造出干燥反应器内部H_2O穿透深度变化的实际工况。根据吸附和解吸周期的H_2O穿透深度评估方法和判据，对CO_2去除系统在不同进气温度下的干燥反应器穿透深度进行评估。选取了CO_2去除系统进气温度在14 ℃、16 ℃和18 ℃附近的3段试验数据（舱内气体相对湿度约40%），用来模拟干燥反应器内部H_2O穿透深度不断增加的情况，系统进气温度曲线及干燥反应器A内部温度曲线如图3和图4所示。

由图3可知：①系统进气温度由低到高大致可以分为3个时间段：第1个时间段（进气温度在13.7～14.5 ℃之间），2018年4月2日～4月3日；第2个时间段（进气温度在15.7～16.9 ℃之间），2018年4月4日～4月5日；第3个时间段（进气温度在18.4～19.3 ℃之间），2018年4月6日～

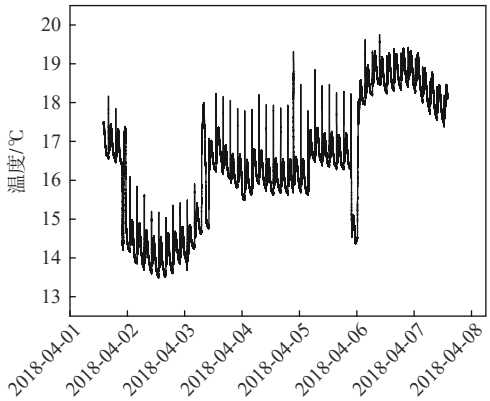


图 3　系统进气温度曲线

Fig. 3　System intake temperature curve

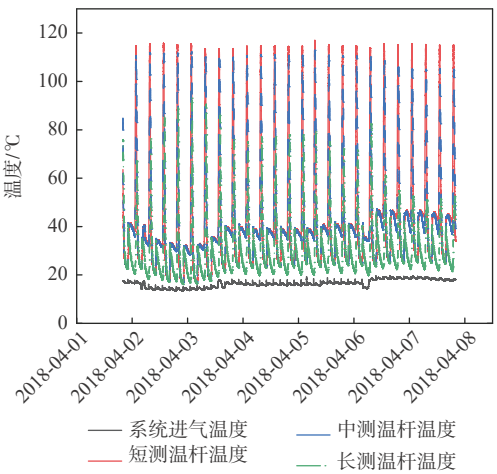


图 4　干燥反应器 A 长、中、短测温杆温度曲线

Fig. 4　Temperature curves of long, medium, and short temperature measuring rods of drying reactor A

4 月 7 日。另外, 在 2018 年 4 月 2 日之前, 进气温度在 16.5～17.5 ℃ 之间; 在 2018 年 4 月 6 日, 进气温度由 16.2 ℃ 下降到 14.5 ℃, 从 2018 年 4 月 6 日, 上升到 17.6 ℃, 该时段系统进气温度突然下降是由于密闭舱低温液路回路控制阀调试导致, 在进行数据处理时需要对该时段数据造成的异常点进行说明。②从干燥反应器内部长、中和短测温杆温度曲线无法判断出 H_2O 穿透深度。

在 OriginPro2021 数据处理软件中(编程语言 LabTalk, 集成开发环境为 OriginPro 自带的 Code Builder), 根据本文方法对 2018 年 4 月 2 日—2018 年 4 月 7 日之间的数据进行处理[21]。吸附周期评估方法中的 i 时刻为吸附周期 179 min, 得到吸附周期内评估判据 ∇T_1、∇T_2、∇T_3 的曲线如图 5 所示, 图 6 为 2018 年 4 月 6 日干燥反应器 A 内部短、中、长测温杆实际温度曲线。采用循环迭代法, 计算出每个解吸周期内干燥反应器 A 内部长、中、短最高温度 T_{Lmax}、T_{Mmax} 和 T_{Smax} 如图 7 所示。

由图 5 和图 6 可知: ①在第 1 个时间段: ∇T_1 远

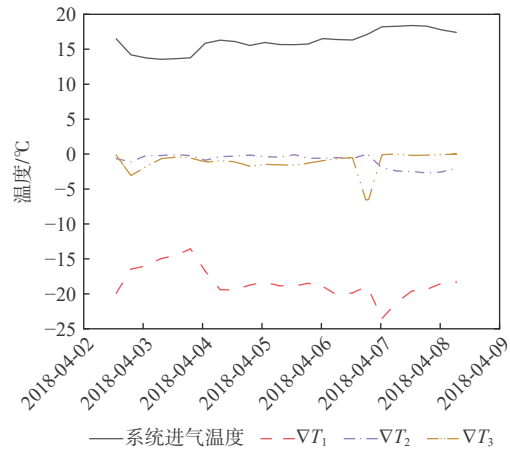


图 5　吸附周期干燥反应器 A 穿透深度评估曲线

Fig. 5　Evaluation curves of penetration depth for drying reactor A in adsorption cycle

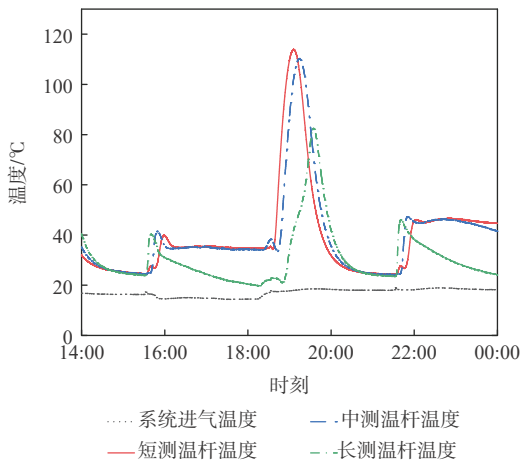


图 6　干燥反应器 A 内部长、中、短测温杆温度曲线(2018 年 4 月 6 日局部时间段)

Fig. 6　Temperature curves of long, medium, and short temperature measuring rods in drying reactor A (local time period on April 6, 2018)

小于 0, 其值都在−15 ℃ 以下, 说明 H_2O 已经穿透区域 A, 穿透深度为Ⅱ级(或以上); ∇T_2、∇T_3 接近于 0, 并且无明显波动, 说明在该时间段内, H_2O 进入但未穿透区域 B, 穿透深度为Ⅱ级。②在第 2 时间段: ∇T_1 远小于 0, 其值都在−17 ℃ 以下, 说明 H_2O 已经穿透区域 A, 穿透深度为Ⅱ级(或以上); ∇T_2、∇T_3 接近于 0, 并且无明显波动, 说明在第 2 时间段内, H_2O 进入但未穿透区域 B, 穿透深度为Ⅱ级。③在第 3 个时间段, ∇T_1 远小于 0, 其值都在−17 ℃ 以下, 说明 H_2O 已经穿透区域 A, 穿透深度为Ⅱ级(或以上); $\nabla T_2<0$, ∇T_3 接近于 0, 说明在该时间段内, H_2O 进入但未穿透区域 C, 穿透深度为Ⅲ级。④吸附周期评估判据曲线中(见图 4), ∇T_3 曲线在 2018 年 4 月 6 日 16:30 左右出现一个明显的下凹点。出现该异常点是由于该时刻进气温度从 14.5 ℃ 升高到 18 ℃(见图 6), 导致干燥中测温杆出现一个明

显的升温台阶(见图6),并且该时刻正好在系统转阶段前。由于吸附周期评估判据中的i选取的是179 min(吸附周期为180 min),刚好处于系统进气温度突变的时段,导致吸附周期的评估算法∇T_3出现一个明显的下凹点,在进行H_2O穿透深度评估时需要剔除这个异常点。这个异常点说明了在进气温度由14.5 ℃升高到18 ℃时,干燥反应器A内部中测温杆温度明显升高,说明区域B中的硅胶吸附放热反应加剧,也间接说明了在进气温度为14.5 ℃工况下,H_2O穿透深度为Ⅱ级,与吸附周期的评估结果吻合。

由图6和图7可知:①在第1个时间段,T_{Lmax}在波动,T_{Lmax}约为93 ℃,说明在该时间段内,H_2O已经穿透区域A,穿透深度在Ⅱ级(或以上);T_{Mmax}、T_{Smax}基本不变,说明在该时间段内,H_2O进入但未穿透区域B,H_2O穿透深度为Ⅱ级。②在第2时间段,T_{Lmax}在波动,T_{Lmax}约为80℃(低于第1个时间段),说明在该时间段内,H_2O已经穿透区域A,穿透深度在Ⅱ级(或以上),与第1个时间段相比,穿透深度增加;T_{Mmax}、T_{Smax}基本不变,说明在该时间段内,H_2O进入但未穿透区域B,H_2O穿透深度为Ⅱ级。②在第3个时间段,T_{Lmax}在波动,T_{Lmax}约为50 ℃(低于第2个时间段),说明在该时间段内,H_2O已经穿透区域A,穿透深度在Ⅱ级(或以上),与第2时间段相比,穿透深度在增加;T_{Mmax}出现明显下降趋势,说明在该时间段内,H_2O穿透深度为Ⅲ级(或以上),穿透深度增加;T_{Smax}一直未有明显波动,说明H_2O未进入区域D,H_2O穿透深度小于Ⅳ级。

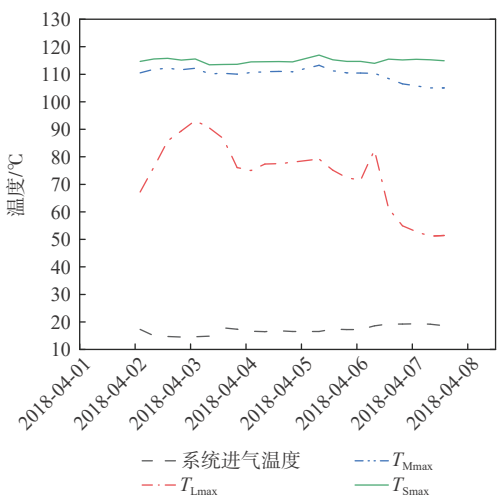


图7 解吸周期干燥反应器穿透深度评估曲线(测温杆最高温度)

Fig. 7 Evaluation curves of penetration depth for drying reactor in desorption cycle (maximum temperature of temperature measuring rod)

综合吸附和解吸周期的H_2O穿透深度评估结果,可以得出:①在第1和第2时间段内,H_2O已经穿透区域A进入区域B,H_2O穿透深度为Ⅱ级;②在第3个时间段内,H_2O已经穿透区域B进入区域C,H_2O穿透深度为Ⅲ级;③吸附和解吸周期的H_2O穿透深度评估结果吻合。

地面试验数据及H_2O穿透深度判别结果表明:CO_2去除系统连续运行1年后,在舱内相对湿度40%,进气温度低于19 ℃时,无H_2O穿透干燥反应器的风险。进气温度约为17 ℃时(目前空间站在轨数据),干燥反应器吸附容量有足够设计裕度,防止短时边界条件变化导致的H_2O穿透问题。随着运行时间增加,干燥反应器性能会下降,需要关注进气温度在19 ℃以下的H_2O穿透深度,以及进气温度升高或舱内湿度增加时H_2O穿透深度增加情况。

7 结 论

1)本文对干燥反应器内部硅胶材料在吸附和解吸周期内的热场变化情况进行了详细的分析,提出的吸附周期内基于不等长测温杆温度梯度和解吸周期内基于不等长测温杆最高温度变化的H_2O穿透深度评估方法实现了相互验证和补充,可以较好的评估在当前边界条件下干燥反应器内部的H_2O穿透深度。

2)本文采用调节系统进气温度的方式来模拟硅胶吸附性能衰减后干燥反应器H_2O穿透深度增加的情况,间接创造了系统性能衰减的表象,验证了评估方法和评估判据,可以用来预测CO_2去除系统长期运行过程中干燥反应器H_2O穿透深度变化问题。

3)本文方法是基于CO_2去除系统在正常工况稳定运行的先验数据基础上得出的,在入口气流温度和露点变化,以及系统长期运行后,硅胶吸附性能下降情况下可以进行提前识别。

本文方法不足之处是在基于软件数据处理后,还需要人工进行判读,并且H_2O是否穿透干燥反应器区域D及穿透深度与其出口气体露点变化之间的对应关系还需要多次拉偏试验进行验证;在此研究基础上,后续设计拉偏试验,结合寿命试验数据,对穿透深度的判据进行量化研究,得出穿透深度与干燥反应器出口气体露点的精确对应关系。

参考文献(References)

[1] HUANG Y H, WANG J L, JIANG H Q, et al. Research of polymerization reactor fault diagnosis based on acoustic emission[J]. Advanced Materials Research, 2011, 221: 550-554.

[2] ZHOU J, HUANG F N, SHEN W H, et al. Sub-period division strategies combined with multiway principle component analysis for fault diagnosis on sequence batch reactor of wastewater treatment process in paper mill[J]. Process Safety and Environmental Protection, 2021, 146: 9-19.

[3] LIU H R, BUVAT J C, ESTEL L, et al. Bayesian network method for fault diagnosis in a continuous tubular reactor[J]. Chemical Product and Process Modeling, 2010, 5(1): 1-9.

[4] SHEIBAT-OTHMAN N, LAOUTI N, VALOUR J P, et al. Support vector machines combined to observers for fault diagnosis in chemical reactors[J]. The Canadian Journal of Chemical Engineering, 2014, 92(4): 685-695.

[5] 刘敏. 人工神经网络在化学反应器故障诊断中的应用[J]. 聊城师院学报 (自然科学版), 2000, 13(1): 51-54.
LIU M. Fault detection in the chemical reactors using artificial neural networks[J]. Journal of Liaocheng Teachers College (Natural Science Edition), 2000, 13(1): 51-54 (in Chinese).

[6] LV M B, LI Y H, GUO R. Galerkin-based extended Kalman filter with application to CO_2 removal system[J]. Journal of Central South University, 2020, 27(6): 1780-1789.

[7] SHI L, WANG L J, WANG Z Z. The modeling and the sensor fault diagnosis of a continuous stirred tank reactor with a takagi-sugeno recurrent fuzzy neural network[J]. International Journal of Distributed Sensor Networks, 2009, 5(1): 37.

[8] CMARIK G, KNOX J, HUFF T. Analysis of performance degradation of silica gels after extended use onboard the ISS[C]//Proceedings of the 48th International Conference on Environmental Systems. Albuquerque: International Conference on Environmental Systems, 2018: 1-11.

[9] CMARIK G, KNOX J. Co-adsorption of carbon dioxide on zeolite 13X in the presence of preloaded water[C]//Proceedings of the 48th International Conference on Environmental Systems. Albuquerque: International Conference on Environmental Systems, 2018: 1-10.

[10] 韩永强, 吴宝治, 杨润泽, 等. 分子筛 CO_2 去除系统研究[J]. 航天医学与医学工程, 2013, 26(3): 181-184.
HAN Y Q, WU B Z, YANG R Z, et al. Study on carbon dioxide removal assembly using molecular sieve technology[J]. Space Medicine & Medical Engineering, 2013, 26(3): 181-184 (in Chinese).

[11] 管国锋, 赵汝溥. 化工原理[M]. 4 版. 北京: 化学工业出版社, 2015.
GUAN G F, ZHAO R P. Principles of chemical engineering[M]. 4th ed. Beijing: Chemical Industry Press, 2015 (in Chinese).

[12] SOLMUŞ İ, ANDREW S REES D, YAMALI C, et al. A two-energy equation model for dynamic heat and mass transfer in an adsorbent bed using silica gel/water pair[J]. International Journal of Heat and Mass Transfer, 2012, 55(19-20): 5275-5288.

[13] ISLAM M A, PAL A, SAHA B B. Experimental study on thermophysical and porous properties of silica gels[J]. International Journal of Refrigeration, 2020, 110: 277-285.

[14] YAMAMOTO E, KATSURAYAMA K, WATANABE F, et al. Heat and mass transfer characteristics in adsorption of water vapor for silica gel packed bed adsorber[J]. Journal of Chemical Engineering of Japan, 2000, 33(1): 12-18.

[15] SUN J, BESANT R W. Heat and mass transfer during silica gel-moisture interactions[J]. International Journal of Heat and Mass Transfer, 2005, 48(23-24): 4953-4962.

[16] ULRICH L, RICARDO C S J, RICHARD B, et al. Fundamental thermodynamic properties of sorbents for atmospheric water capture[J]. Chemical Engineering Journal, 2022, 431: 134058.

[17] BERNAL V, GIRALDO L, MORENO-PIRAJÁN J C. Understanding the solid-liquid equilibria between paracetamol and activated carbon: Thermodynamic approach of the interactions adsorbent-adsorbate using equilibrium, kinetic and calorimetry data[J]. Journal of Hazardous Materials, 2021, 419: 126432.

[18] 邓述波, 余刚. 环境吸附材料及应用原理[M]. 北京: 科学出版社, 2012.
DENG S B, YU G. Environmental adsorption materials and their application principles[M]. Beijing: Science Press, 2012 (in Chinese).

[19] 周立幸, 陈淑霞, 郑禄彬. 水/硅胶吸附体系的热化学研究[J]. 物理化学学报, 1985, 1(6): 566-570.
ZHOU L X, CHEN S X, ZHENG L B. Thermochemical study for water/silica adsorption system[J]. Acta Physico-Chimica Sinica, 1985, 1(6): 566-570 (in Chinese).

[20] 王德昌, 韩艳培, 张光, 等. 吸附制冷中硅胶吸附性能衰减特性研究[J]. 工程热物理学报, 2012, 33(2): 207-209.
WANG D C, HAN Y P, ZHANG G, et al. Investigation of adsorption performance deterioration of silica gel in water adsorption refrigeration[J]. Journal of Engineering Thermophysics, 2012, 33(2): 207-209 (in Chinese).

[21] 叶卫平. Origin9.1 科技绘图及数据分析[M]. 北京: 机械工业出版社, 2019.
YE W P. Origin9.1 Technology drawing and data analysis[M]. Beijing: China Machine Press, 2019 (in Chinese).

Evaluation method of H_2O penetration depth of drying reactor based on temperature gradient

ZHANG Yongping，BIAN Qiang*，YANG Runze，ZHANG Zhen，BAI Panfeng

(National Key Laboratory of Human Factor Engineering，Astronaut Research and Training Center of China，Beijing 100094，China)

Abstract：There is no direct evaluation method for the H_2O penetration depth of the drying reactor in the CO_2 removal system in the space station. Therefore, an indirect evaluation method based on the temperature gradient characteristics of the temperature sensor inside the drying reactor was proposed. Since silica gel has the physical characteristic of releasing heat while adsorbing H_2O and absorbing heat while desorbing H_2O, the variation characteristics of the temperature sensor data curve in the adsorption and desorption cycles under different boundary conditions were analyzed by considering the depth position of the temperature sensor in the drying reactor. The index set and evaluation method reflecting the H_2O penetration depth of the drying reactor were proposed, and the proposed method is verified in the closed chamber test of the system, the verification results show the rationality of the index set and the effectiveness of the evaluation method.

Keywords：space station; CO_2 removal system; reactor; adsorption; desorption; water penetration depth

Received：2022-09-07；**Accepted**：2023-03-06；**Published Online**：2023-05-10 16：56

URL：link.cnki.net/urlid/11.2625.V.20230510.1356.004

* **Corresponding author**. E-mail：13810213687@139.com

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2022.0801

基于时序知识图谱的电力设备缺陷预测

纪鑫[1,2]，武同心[1,*]，杨智伟[1]，余婷[1]，李君婷[1]，何禹德[1]

(1. 国家电网有限公司 大数据中心，北京 100031；2. 北京航空航天大学 计算机学院，北京 100191)

摘 要：随着电力设施规模的日渐增长，对于电力设备的运行监测成为电力系统管理的一个重要内容。电力设备的缺陷预测是电力系统运行监测的关键步骤。为解决对大规模电力系统中的电力设备进行缺陷预测的问题，结合人工智能技术，提出一种基于时序知识图谱的电力设备缺陷预测模型。通过注意力机制对多模态信息进行融合，利用关系感知的图神经网络和循环神经网络得到实体和关系的时序表示，基于时序表示对电力设备进行缺陷预测。提出基于图神经网络的时序知识图谱推理方法能够充分利用多模态信息，提升电力设备缺陷预测的准确率。实验表明：所提模型性能优于基线模型。

关 键 词：时序知识图谱；图神经网络；缺陷预测；电力系统；电力设备

中图分类号：TM933

文献标志码：A **文章编号**：1001-5965(2024)10-3131-08

电力网络的高速发展及其规模的逐渐增大对电力设备的可靠性提出了更高的要求，对于潜在的电力设备缺陷及早地发现显得尤为重要[1]。面向电网运行积累的大量设备运维数据，深入挖掘设备运维数据特征，感知设备运行态势，是新型电力系统精细化发展的重要方向。针对电力设备的缺陷预测需要全面地掌握电力设备的各种运行数据与环境参数，以便对其运行状态进行合理的评估，进而预测可能发生的故障。

电力系统中的各种数据和变量常以非结构化的形式存储，计算机难以理解其中的语义信息。知识图谱是一种将非结构化的知识和数据转化为结构化知识的数据库，其用图的形式表示知识，图上的每个节点表示实体，实体可能会有自己对应的属性和特征；图上的边表示实体之间的对应关系；整张知识图谱包含了知识库中所有知识的映射，并能很好地被计算机所理解。近年来，知识图谱被成功地应用于搜索[2]、推荐[3]和问答系统[4]等任务中。上述知识图谱只关注静态的信息，图谱中的关系和节点不会随着时间变化，然而在现实世界中时序信息非常重要，很多结构和关系只在特定时间段中有效。将时序信息融入到知识图谱中，这样的知识图谱称为时序知识图谱，也就是说，时序知识图谱将静态知识图谱的三元组扩充为四元组，在每一个时间片上都存在一个静态图谱。与普通知识图谱相比，时序知识图谱增加了一个时间维度，能够更好地建模现实世界中的知识。

针对电力设备的缺陷预测问题，本文将其转换为时序知识图谱上的链接预测(推理)问题。将电力设备和可能发生的缺陷抽象为知识图谱上的节点，将设备与其发生的历史缺陷之间的对应关系抽象为知识图谱之间的关系，另外，将电力设备的属性、运行状态等信息作为节点的多模态属性。通过对这样的知识图谱进行学习，预测电力设备节点在未来时刻与缺陷节点之间可能存在的链接，完成对相应电力设备的缺陷预测，另外，由于电力设备缺

收稿日期：2022-09-19；**录用日期**：2022-10-03；**网络出版时间**：2023-02-01 16:48
网络出版地址：link.cnki.net/urlid/11.2625.V.20230201.1047.002
基金项目：国家电网有限公司大数据中心科技项目 (52999021N005)
* **通信作者**. E-mail：tongxin-wu@sgcc.com.cn

引用格式：纪鑫，武同心，杨智伟，等. 基于时序知识图谱的电力设备缺陷预测 [J]. 北京航空航天大学学报，2024，50（10）：3131-3138.
JI X，WU T X，YANG Z W，et al. Power equipment defect prediction based on temporary knowledge graph[J]. Journal of Beijing University of Aeronautics and Astronautics，2024，50（10）：3131-3138（in Chinese）.

陷的出现存在时序性，因此，引入时序特征，以形成时序知识图谱，时序特征是离散的，在每一个时间片上都根据之前时间片的信息去预测缺陷。

知识图谱技术是以图形的方式展示实体与属性之间的关系，有效整合数据，形成知识驱动的管理和运行模式，将其应用在电力行业领域，将大量电力数据凭借知识抽取、推理、检索等功能转化为电力知识的应用，实际知识图谱技术就是人工智能与大数据结合的产物。知识图谱链接预测又称知识图谱推理。给定一个知识图谱，知识图谱推理的任务是预测出当前知识图谱缺失的三元组集合。对于时序知识图谱来说，时序知识图谱的推理可以视作在特定时间节点的静态知识图谱进行推理。

由于知识图谱的广泛应用及命名实体识别(named entity recognition, NER)在知识图谱构建中的基础性作用，目前对于NER的研究已经比较成熟，无论是基于规则等的传统NER方法，还是基于深度学习的NER方法，都具有一套相对充分和完整的理论和技术。

另一方面，在电力领域中的设备属性常具有多模态的特性[5-8]，例如，一个电力设备的图片属于视觉模态；电力设备的描述和作用等属于文本模态；电力设备的参数等属于数字模态。如何充分地利用这些多模态信息来得到更具辨识度的实体表示也是缺陷检测任务的关键点。

针对电力设备的缺陷检测任务，本文构造了基于多模态的电力设备缺陷时序知识图谱数据集。其中，包含多个时间段的多种电力设备实体和对应的缺陷类型。每个电力设备还有相对应的多模态属性信息。本文提出一个基于图神经网络的时序知识图谱推理模型，首先，利用注意力机制对实体节点的多模态信息进行融合；其次，利用图神经网络学习节点表示；最后，基于节点对的表示预测电力设备可能会出现的缺陷。实验结果表明，本文模型性能优于基线模型。

1 电力设备缺陷检测识别框架

1.1 多模态数据融合

如图1所示，本文针对多模态数据提出一种基于注意力机制的特征融合方法。

为获取实体节点的视觉特征，将实体节点的每张图片转换为特征向量。本文使用在Imagenet数据集上预训练过的ResNet50[9-10]网络来提取图片的特征。ResNet网络中引入了残差链接结构，使得网络可以更深，学习到的特征表达能力更强，本文使用层数为50的网络结构，将最后一层的输出作为图片的特征。总结来说，提取到的视觉特征为

$$\boldsymbol{E}_{\mathrm{v}}=\boldsymbol{R}_{50}\left(i_{\mathrm{mg}i}\right) \tag{1}$$

式中：$\boldsymbol{E}_{\mathrm{v}}$为图片的特征向量；$\boldsymbol{R}_{50}$为使用ResNet50网络提取图片特征；$i_{\mathrm{mg}i}$表示第$i$张图片。

实体节点也存在着许多文本类的特征，如属性描述等信息。为从文本中提取有用的信息，使用预训练语言模型Bert[11]作为编码器。Bert是2018年Google推出的大规模文本预训练模型。Bert由基于transformer架构的多个encoder block堆叠而成，其在许多nlp任务上的性能都提升明显。相较于传统的word2vec[12]等算法，经过海量文本预训练的Bert能够在得到的文本表征中引入更多的语义信息，表示能力更强。文本特征为

$$\boldsymbol{E}_{t}=\boldsymbol{B}_{\mathrm{ert}}\left(t_{\mathrm{ext}i}\right) \tag{2}$$

式中：$\boldsymbol{E}_{t}$表示文本的特征向量；$\boldsymbol{B}_{\mathrm{ert}}$表示使用Bert模型编码文本；$t_{\mathrm{ext}i}$表示第$i$段文本。

数值特征指的是电力设备参数等包含数值的离散信息。首先，需要将离散特征转换为连续的特征表示，径向基函数(radical basis function, RBF)[13]

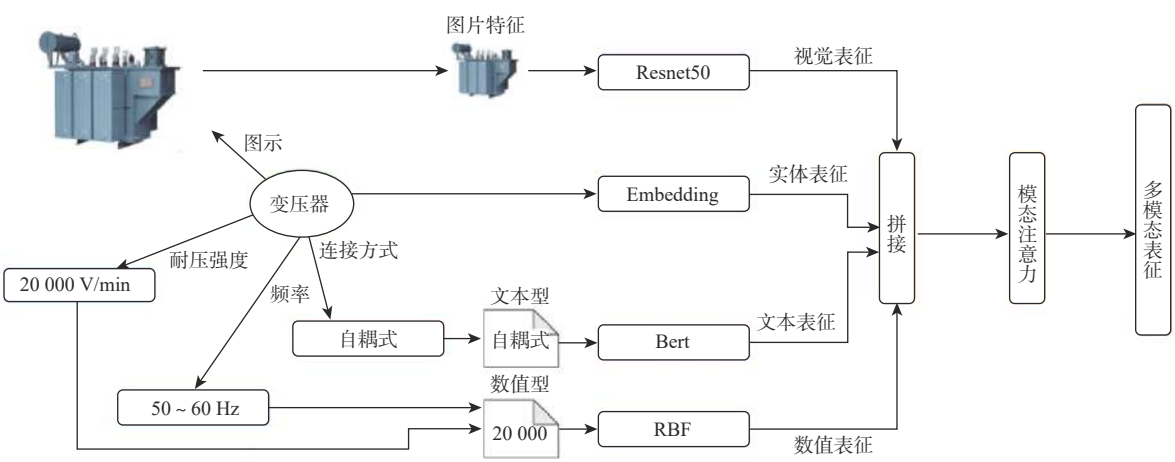


图1 多模态数据融合

Fig. 1 Multi-modal data fusion

正好满足本文的需求。RBF 可以拟合任意分布，具有收敛快、泛化能力强的优点。本文使用的是高斯径向基函数，设离散的数字特征为$\boldsymbol{x}_i$，则根据 RBF 得到的连续数值特征为

$$\boldsymbol{E}_{\mathrm{n}}=\phi(r)\boldsymbol{I}=\mathrm{e}^{-(\varepsilon r)^2}\boldsymbol{I} \tag{3}$$

$$r=\|\boldsymbol{x}-\boldsymbol{x}_i\| \tag{4}$$

式中：$\boldsymbol{E}_{\boldsymbol{n}}$为数值特征向量；$\phi(r)$表示 RBF；$r$为输入特征与中心特征之间的距离，表示当前特征值；$\boldsymbol{I}$为单位向量；$\boldsymbol{x}_i$为第$i$个中心特征值；$\varepsilon$为 RBF 的参数。

除众多的属性特征之外，实体节点本身也具有相应的特征表示，对于这一类的特征，使用随机初始化的特征嵌入$\boldsymbol{E}_{\mathrm{e}}$表示。相应的，也对关系设置了相应的嵌入向量$\boldsymbol{R}$。

在得到上述多模态特征之后，需要考虑如何有效地将多模态信息融合。考虑到不同模态的特征位于不同的表征空间中，因此，本文利用变换矩阵将不同模态的信息变换到同一表征空间中：

$$\boldsymbol{E}'_{\mathrm{v}}=\boldsymbol{W}_{\mathrm{v}}\boldsymbol{E}_{\mathrm{v}}+\boldsymbol{b}_{\mathrm{v}} \tag{5}$$

$$\boldsymbol{E}'_{\mathrm{t}}=\boldsymbol{W}_{\mathrm{t}}\boldsymbol{E}_{\mathrm{t}}+\boldsymbol{b}_{\mathrm{t}} \tag{6}$$

$$\boldsymbol{E}'_{\mathrm{n}}=\boldsymbol{W}_{\mathrm{n}}\boldsymbol{E}_{\mathrm{n}}+\boldsymbol{b}_{\mathrm{n}} \tag{7}$$

$$\boldsymbol{E}'_{e}=\boldsymbol{W}_{\mathrm{e}}\boldsymbol{E}_{\mathrm{e}}+\boldsymbol{b}_{\mathrm{e}} \tag{8}$$

式中：$\boldsymbol{E}'_{\mathrm{v}}$、$\boldsymbol{E}'_{\mathrm{t}}$、$\boldsymbol{E}'_{\mathrm{n}}$、$\boldsymbol{E}'_{e}$表示经过线性变换后的特征向量；$\boldsymbol{W}_{\mathrm{v}}$、$\boldsymbol{W}_{\mathrm{t}}$、$\boldsymbol{W}_{\mathrm{n}}$、$\boldsymbol{W}_{\mathrm{e}}$表示特征变换的权重矩阵；$\boldsymbol{b}_{\mathrm{v}}$、$\boldsymbol{b}_{\mathrm{t}}$、$\boldsymbol{b}_{\mathrm{n}}$、$\boldsymbol{b}_{\mathrm{e}}$表示特征变换的偏置项。

使用一个基于注意力[14]的融合机制聚合所有特征：

$$\left(\alpha'_{\mathrm{v}},\alpha'_{\mathrm{t}},\alpha'_{\mathrm{n}},\alpha'_{\mathrm{e}}\right)=$$

$$\mathrm{Softmax}(\boldsymbol{a}^{\mathrm{T}}\boldsymbol{E}'_{\mathrm{v}},\boldsymbol{a}^{\mathrm{T}}\boldsymbol{E}'_{\mathrm{t}},\boldsymbol{a}^{\mathrm{T}}\boldsymbol{E}'_{\mathrm{n}},\boldsymbol{a}^{\mathrm{T}}\boldsymbol{E}'_{\mathrm{e}}) \tag{9}$$

式中：α'_{v}、α'_{t}、α'_{n}、α'_{e}为注意力系数；$\boldsymbol{a}^{\mathrm{T}}$为可学习的注意力向量。最终的融合的特征为

$$\boldsymbol{E}=\alpha'_{\mathrm{v}}\boldsymbol{E}'_{\mathrm{v}}+\alpha'_{\mathrm{t}}\boldsymbol{E}'_{\mathrm{t}}+\alpha'_{\mathrm{n}}\boldsymbol{E}'_{\mathrm{n}}+\alpha'_{\mathrm{e}}\boldsymbol{E}'_{\mathrm{e}} \tag{10}$$

1.2　时序知识图谱推理

图 2 为时序知识图谱推理，主要包括节点表示学习和时序特征建模 2 个部分。

由于知识图谱是一个图结构，近年来，随着图神经网络的兴起，有很多关于知识图谱推理的研究将图神经网络应用于模型中[14-18]，本文使用关系图神经网络（relational graph convolutional network, RGCN）[18]来学习每个时间戳中静态知识图谱的实体表示。RGCN 相比于传统的图神经网络模型，增加了对关系数据的建模，其计算过程为

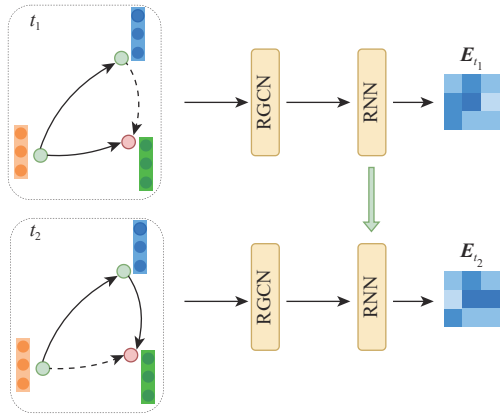


图 2　时序知识图谱推理

Fig. 2　Temporary knowledge graph completion

$$\boldsymbol{h}_i^{l+1}=\sigma(\boldsymbol{W}_0^l\boldsymbol{h}_i^l+\sum_{r\in M}\sum_{j\in P(i)}\frac{1}{|P(i)|}\boldsymbol{W}_r{}^l\boldsymbol{h}_j{}^l) \tag{11}$$

式中：$\boldsymbol{h}_i^{l+1}$为第$l+1$层的节点；$\boldsymbol{h}_j^l$为经过第l层 RGCN 计算之后的节点；σ为激活函数；$\boldsymbol{W}_0^l$和$\boldsymbol{W}_r{}^l$为第l层的权重矩阵；$\boldsymbol{W}_r^l$为对不同关系的映射；M为关系集合；$P(i)$为与节点i相邻的节点集合。

经过 RGCN 层之后，得到图谱中每个节点的更新表示。考虑到时序信息，本文选择序列模型循环神经网络（recurrent neural network, RNN）[19]来建模时序关系：

$$\boldsymbol{N}_t=R_{\mathrm{NN}}\left(\boldsymbol{N}_{t-1},h_t^e\right) \tag{12}$$

$$\boldsymbol{R}_t=R_{\mathrm{NN}}\left(\boldsymbol{R}_{t-1},h_t^r\right) \tag{13}$$

式中：$\boldsymbol{N}_t$和$\boldsymbol{R}_t$分别为第t个时间段的实体表示矩阵和关系表示矩阵。

基于成对的节点表示，本文将电力设备实体节点与缺陷实体节点之间的链接存在性的条件概率建模为

$$p(o|s,r,\boldsymbol{N}_t,\boldsymbol{R}_t)=\sigma(\boldsymbol{N}_t\phi(s_t,r_t)) \tag{14}$$

式中：s为头实体，即电力设备节点；$\phi(s_t,r_t)$表示三元组评分函数；o为尾实体，即缺陷节点；σ为 sigmoid 函数：

$$\sigma=\frac{1}{1+\mathrm{e}^{-x}} \tag{15}$$

1.3　损失函数

使用交叉熵损失函数来进行端到端的模型优化，具体的损失函数如下：

$$L=\sum_t\sum_{e\in\varepsilon}\sum_{i\in V}y_{t+1,i}\log_2 p_i(o|s,r,\boldsymbol{N}_t,\boldsymbol{R}_t) \tag{16}$$

式中：ε为时间片集合；V为节点集合；$y_{t+1,i}$为真实标签；$p_i(o|s,r,\boldsymbol{N}_t,\boldsymbol{R}_t)$为模型预测的条件概率。

2 数据集与实验设置

本文使用的数据集来自公司的电力设备缺陷事件，这其中存在着多模态数据，还有时间序列信息，因此，本文将其建模为时序的多模态图谱，每个时间片上都是一个静态的多模态知识图谱。

具体来说，在时间片 τ 对应的多模态知识图谱，包含电力设备的多模态信息（如设备图片、参数等）及设备缺陷等，任务目标是根据过去时间片的图谱信息去预测未来某个时间片上电力设备可能发生的缺陷。

2.1 数据预处理

数据集中大部分的数据都存在着数据格式不统一的问题。为后续的模型学习，需要对数据进行必要的清洗。

1）图像数据预处理。在电力领域的多模态知识图谱中，图片的尺寸和分辨率的分布常是不一致的[20-23]。另一方面，由于拍摄设备的限制，图片所表示的实体可能在图片中所占的比例很小。因此，先将所有图像统一裁剪到224像素×224像素大小，再将图片输入到目标检测网络中，将目标实体所在的图片区域作为最终的图像数据。

2）文本数据预处理。数据集中的文本数据存在着大量的噪声，如特殊符号等。先利用Nltk工具去除文本中的所有特殊字符，去除标点符号。由于预训练语言模型对于输入的长短有限制，所以还对所有文本进行截断处理，超出限制最长的文本只保留前缀。

3）数值数据预处理。针对电力设备的属性等离散的数值特征，先进行预处理，去除无关的文本字符，后对所有数值进行正规化，防止出现偏离过多的数值对模型学习造成不良的影响。

2.2 数据统计

经过预处理之后，本文构建了用于缺陷检测的时序多模态知识图谱，对数据集进行了初步的统计。如表1所示，本文的知识图谱包含7个时间段的2 106个三元组，其中，每个三元组包含电力设备机器及其产生的缺陷类型。

图谱的总体统计如表2所示，图谱共包含三元组4 488个，共包含2 463个实体节点（电力设备节点和缺陷节点），还包含3 077条多模态属性。

为方便模型学习与验证，将完整的数据集按照边划分为训练集、验证集和测试集，总的比例为6∶2∶2。

2.3 实验设置

本文所有模型与实验都是基于Pytorch，Pytorch是一个具有自动求导功能的开源深度学习框架[24-26]。

表1 电力设备及其缺陷类型

Table 1 Electrical equipment and type of defect event

序号	电力设备	缺陷类型	三元组数量
1	主变压器	堵塞，进水，污秽，滑档，生锈，喷油，油温过高，漏油等	1 147
2	变压器	堵塞，进水，污秽，滑档，生锈，喷油，油温过高，漏油等	534
3	联变变压器	生锈，无油位，无防雨罩，未绝缘包扎等	160
4	站用变压器	氧气含量超标	220
5	电抗器	连接处震动，压力低，数据未上传，封堵，硅胶变色，乙炔超标等	1 234
6	高压电抗器	乙炔、甲烷超标，漏油，操作不当等	456
7	干式电抗器	缺少绝缘试验报告	567
8	其他子部件	生锈，缺少实验报告，无防雨罩，脏污，堵塞，封堵不当等	170

表2 数据类型和数量

Table 2 Type and quantity of dataset

三元组数	节点数	属性数
4 488	2 463	3 077

Resnet网络采用的是Pytorch官方的实现，并加载了基于Imagenet数据集预训练的模型参数。Bert网络采用的是开源自然语言处理库Hugging Face，这个库提供了大量预训练语言模型的实现。知识图谱推理方面，本文模型基于官方的RGCN代码实现。其中，隐藏层的维度为256。RNN基于Pytorch的官方实现，隐藏层的维度为512。

在训练时，每个小批次的训练样本的大小为64，随机丢弃概率为0.1，RGCN的层数为4，RNN的层数为2，总的训练epoch数为100，优化器采用Adam，为防止过拟合，设置正则化系数为0.01。同时，本文还采用了early stop的策略，当模型在验证集上的表现连续3次无法提升时，就提前停止训练。

关于模型的评价指标，本文采用平均排名（mean rank，MR）、平均倒数排名（mean reciprocal ranking，MRR）、击中率排名第1的比例（Hits@1）、击中率排名前10的比例（Hits@10）这4个指标，4个指标都是在知识图谱推理中常见的评价指标。所有的实验都在NVIDIA GTX-1080 Ti GPU上进行训练。详细的模型参数如表3所示。

表3 模型主要参数

Table 3 Main parameters of model

参数名	参数介绍	参数值
Learning rate	初始学习率	0.05
Total epochs	总训练轮次	50
Dropout rate	随机丢弃概率	0.1
Regularizer	正则化系数	0.01
Batch size	每批次训练样本的大小	64
Num layers	网络层数	4
Hidden dim	隐藏层维度	256

3 实　验

3.1 基线模型

本文选择的基线模型主要包括以下 4 种类型:

1) DisMult 模型[27]、ConvE 模型[28] 和 RotatE 模型[29], 这些模型属于静态知识图谱推理方法, 在本文实验中, 被应用于每个时间戳对应的静态知识图谱, 然后做出预测;

2) TA-DistMult 模型[30] 和 HyTE 模型[31], 是基于插值设置的时序知识图谱推理模型;

3) Dyngraph2vecAE 模型[32]、tNodeEmbed 模型[33]、EvolveGCN 模型[34] 和 GCRN 模型[35], 都是基于同构图的外推推理模型;

4) Know-Evolve 模型[36]、DyRep 模型[37] 和 RE-Net 模型[38], 是外推式的时序知识图谱推理模型。

对于所有模型都使用了相同的多模态信息融合方法, 所有的隐藏层向量的维度大小设置为相同的 256, 以便进行公平的比较。

3.2 主 实 验

对于电力设备缺陷检测的主实验, 将其转换为时序知识图谱的链接预测任务。预处理后的知识图谱先经过基于注意力的信息融合得到初始实体表示, 然后知识图谱通过基线模型和本文模型, 本文使用 MR、MRR、Hits@1、Hits@10 来评测实验效果。

MR 的计算式如下:

$$M_{\mathrm{R}}=\frac{1}{|S|}\sum_{i=1}^{|S|}r_{\mathrm{ank}i} \tag{17}$$

式中: $|S|$为三元组集合的大小; $r_{\mathrm{ank}i}$为第i个三元组的排名。

MRR 的计算式如下:

$$M_{\mathrm{RR}}=\frac{1}{|S|}\sum_{i=1}^{|S|}\frac{1}{r_{\mathrm{ank}i}} \tag{18}$$

Hits@n 表示排名小于 n 的三元组的占比, 计算式如下:

$$H_n=\frac{1}{|S|}\sum_{i=1}^{|S|}\begin{cases}1 & r_{\mathrm{ank}i}\leqslant n\\ 0 & r_{\mathrm{ank}i}>n\end{cases} \tag{19}$$

为减少实验的偶然性误差, 本文对每一个模型都重复实验了 10 次, 报告 10 次实验的平均值, 如表 4 所示。

从实验结果可知, 相比于基线模型, 本文模型在 MR、MRR、Hits@1、Hits@10 指标上都显著优于其他基线模型, 其值分别为 21.34%、68.22%、57.82% 和 82.50%。通过对实验结果的分析可以得出, 本文模型通过基于注意力机制的多模态信息融合机制学习到多模态的实体信息, 再通过 RGCN 和 RNN 模型学习时序信息, 能够很好地对于电力设备可能发生的缺陷进行预测。

表 4　缺陷检测性能

Table 4　Defect prediction performance

基线模型	MR/%	MRR/%	Hits@1/%	Hits@10/%
DisMult[27]	51.37	61.83	50.92	72.38
ConvE[28]	33.83	52.11	36.83	69.24
RotatE[29]	42.82	44.33	48.85	70.76
TA-DistMult[30]	34.23	65.37	53.21	71.72
HyTE[31]	42.78	53.49	56.19	72.25
Dyngraph2vecAE[32]	36.72	50.02	59.72	70.82
tNodeEmbed[33]	41.83	62.38	50.48	71.63
EvolveGCN[34]	32.28	62.82	54.05	72.32
GCRN[35]	36.83	52.23	51.21	76.22
Know-Evolve[36]	44.32	66.73	51.92	75.38
DyRep[37]	35.83	54.35	44.83	69.24
RE-Net[38]	35.73	68.77	56.48	70.58
本文模型	21.34	68.22	57.82	82.50

3.3 消融实验

为验证本文模型中各个模块的有效性, 本文进行了消融实验。消融实验是指删除一个模型的一个或多个模块, 并与未消融地模型进行对比实验, 通过比较性能差异来研究深度学习模型中各个组件的必要性。

本文针对模型设置了以下消融实验。首先, 基于注意力的多模态融合, 分别验证图片、文字、数值等多模态信息对模型的影响。其次, 探讨在 RGCN 中是否对不同的关系设置不同的权重。最后, 对上述 2 个组件进行消融实验, 将该模型作为基线模型进行对比分析。

对于每个消融实验, 本文采取了相同的超参数, 每次实验同样重复进行 10 次, 取所有实验结果的平均值作为最终的结果。消融实验的结果如表 5 所示。

从实验结果来看, 视觉、文本、数值等多模态

表 5　消融实验结果

Table 5　Results of ablation experiments

消融模块	MR/%	MRR/%	Hits@1/%	Hits@10/%
基线模型	44.56	49.17	44.58	74.31
只消融视觉模态	42.61	51.23	45.18	75.21
只消融文本模态	28.88	65.17	52.58	79.31
只消融数值模态	29.43	64.76	51.62	78.66
消融RGCN关系矩阵	25.73	66.14	55.91	80.55
完整模型	21.34	68.22	57.82	82.50

信息无论是消融哪个模块，关于电力设备缺陷预测的MR、MRR、Hits@1、Hits@10指标均有明显下降，说明多模态信息的引入对于本文模型具有重要的作用。

从多模态信息的比较来看，文本模态对于模型性能的提升作用更加明显，说明了文本信息在本文的缺陷检测任务中更为重要，可能是因为电力设备的文本信息中包含了大量的关于运行状态的信息，这些信息对预测未来时间段内电力设备的预测非常有益，而本文提取文本信息使用的Bert预训练模型能够提取蕴含在文本中的有益信息。

在消融了RGCN的关系映射矩阵之后，缺陷检测任务的性能下降也比较明显，由于本文不对不同的关系设置不同的映射，那么模型就会对所有关系平等看待，这显然是不合理的，因为来自不同关系的特征对于当前节点的贡献应当不同，实验结果也验证了本文的假设。

4 结 论

1）本文提出基于图神经网络和时序知识图谱推理的模型在电力设备缺陷检测任务中表现出色。MR为21.34%、MRR为68.22%、Hits@1为57.82%、Hits@10为82.50%。

2）本文提出基于注意力机制的多模态信息融合方法能够有效提升节点表征的表达能力。融合图片、文字、数值等多模态信息后，节点特征更加丰富，有助于提升缺陷检测的准确性。

3）引入时序特征后，模型能够更准确地预测未来时间段内电力设备可能发生的缺陷。时序信息的引入使模型能够更好地建模电力设备缺陷的时间依赖性，从而提高预测的准确性。通过对比实验可以看出，包含时序特征的模型在各项评价指标上均显著优于不包含时序特征的基线模型。

参考文献（References）

[1] 蒲天骄, 乔骥, 韩笑, 等. 人工智能技术在电力设备运维检修中的研究及应用[J]. 高电压技术, 2020, 46(2): 369-383.
PU T J, QIAO J, HAN X, et al. Research and application of artificial intelligence in operation and maintenance for power equipment[J]. High Voltage Engineering, 2020, 46(2): 369-383 (in Chinese).

[2] 田萱, 陈杭雪. 推荐任务中知识图谱嵌入应用研究综述[J]. 计算机科学与探索, 2022, 16(8): 1681-1705.
TIAN X, CHEN H X. Survey on applications of knowledge graph embedding in recommendation tasks[J]. Journal of Frontiers of Computer Science and Technology, 2022, 16(8): 1681-1705 (in Chinese).

[3] 董天波. 基于知识图谱的智能问答系统实现与应用[J]. 数字技术与应用, 2022, 40(3): 165-167.
DONG T B. Realization and application of intelligent question answering system based on knowledge map[J]. Digital Technology & Application, 2022, 40(3): 165-167 (in Chinese).

[4] LI J, SUN A X, HAN J L, et al. A survey on deep learning for named entity recognition[J]. IEEE Transactions on Knowledge and Data Engineering, 2022, 34(1): 50-70.

[5] 田玲, 张谨川, 张晋豪, 等. 知识图谱综述: 表示、构建、推理与知识超图理论[J]. 计算机应用, 2021, 41(8): 2161-2186.
TIAN L, ZHANG J C, ZHANG J H, et al. Knowledge graph survey: Representation, construction, reasoning and knowledge hypergraph theory[J]. Journal of Computer Applications, 2021, 41(8): 2161-2186 (in Chinese).

[6] 付雷杰, 曹岩, 白瑀, 等. 国内垂直领域知识图谱发展现状与展望[J]. 计算机应用研究, 2021, 38(11): 3201-3214.
FU L J, CAO Y, BAI Y, et al. Development status and prospect of vertical domain knowledge graph in China[J]. Application Research of Computers, 2021, 38(11): 3201-3214 (in Chinese).

[7] 陈烨, 周刚, 卢记仓. 多模态知识图谱构建与应用研究综述[J]. 计算机应用研究, 2021, 38(12): 3535-3543.
CHEN Y, ZHOU G, LU J C. Survey on construction and application research for multi-modal knowledge graphs[J]. Application Research of Computers, 2021, 38(12): 3535-3543 (in Chinese).

[8] 张吉祥, 张祥森, 武长旭, 等. 知识图谱构建技术综述[J]. 计算机工程, 2022, 48(3): 23-37.
ZHANG J X, ZHANG X S, WU C X, et al. Survey of knowledge graph construction techniques[J]. Computer Engineering, 2022, 48(3): 23-37 (in Chinese).

[9] SIMONYAN K, ZISSERMAN A. Very deep convolutional networks for large-scale image recognition[EB/OL]. (2015-04-10) [2022-05-10]. http://arxiv.org/abs/1409.1556.

[10] HE K M, ZHANG X Y, REN S Q, et al. Deep residual learning for image recognition[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2016: 770-778.

[11] DEVLIN J, CHANG M W, LEE K, et al. BERT: Pre-training of deep bidirectional transformers for language understanding [EB/OL]. (2019-05-24) [2022-05-10]. http://arxiv.org/abs/1810.04805.

[12] MIKOLOV T, CHEN K, CORRADO G, et al. Efficient estimation of word representations in vector space[EB/OL]. (2013-09-17) [2022-05-10]. http://arxiv.org/abs/1301.3781.

[13] CHEN S, COWAN C F N, GRANT P M. Orthogonal least squares learning algorithm for radial basis function networks[J]. IEEE Transactions on Neural Networks, 1991, 2(2): 302-309.

[14] CAI L, YAN B, MAI G C, et al. TransGCN: Coupling transformation assumptions with graph convolutional networks for link prediction[C]//Proceedings of the 10th International Conference on Knowledge Capture. New York: ACM, 2019: 131–138.

[15] SHANG C, TANG Y, HUANG J, et al. End-to-end structure-aware convolutional networks for knowledge base completion[J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2019, 33(1): 3060-3067.

[16] NATHANI D, CHAUHAN J, SHARMA C, et al. Learning attention-based embeddings for relation prediction in knowledge graphs

[EB/OL]. (2019-06-04) [2022-05-10]. http://arxiv.org/abs/1906.01195.

[17] XU X R, FENG W, JIANG Y S, et al. Dynamically pruned message passing networks for large-scale knowledge graph reasoning [EB/OL]. (2020-04-07) [2022-05-10]. http://arxiv.org/abs/1909.11334.

[18] VOLODYMYR M, NICOLAS H, ALEX G. Recurrent models of visual attention[J]. Advances in Neural Information Processing Systems, 2014, 27(2): 2204-2212.

[19] ZAREMBA W, SUTSKEVER I, VINYALS O. Recurrent neural network regularization[EB/OL]. (2015-02-19) [2022-05-10]. http://arxiv.org/abs/1409.2329.

[20] CAO Y X, LIU Z Y, LI C J, et al. Multi-channel graph neural network for entity alignment[EB/OL]. (2019-08-26) [2022-05-10]. http://arxiv.org/abs/1908.09898.

[21] ZHANG Q H, SUN Z Q, HU W, et al. Multi-view knowledge graph embedding for entity alignment[EB/OL]. (2019-06-06) [2022-05-10]. http://arxiv.org/abs/1906.02390.

[22] GUO L B, SUN Z Q, HU W. Learning to exploit long-term relational dependencies in knowledge graphs[EB/OL]. (2019-05-13) [2022-05-10]. http://arxiv.org/abs/1905.04914.

[23] SUN Z Q, WANG C M, HU W, et al. Knowledge graph alignment network with gated multi-hop neighborhood aggregation[J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2020, 34(1): 222-229.

[24] ADAM P, SAM G, SOUMITH C, et al. Automatic differentiation in pytorch[J]. Advances in Neural Information Processing Systems, 2017, 3(6): 1401-1412.

[25] ADAM P, SAM G, FRANCISCO M, et al. Pytorch: An imperative style, high-performance deep learning library[J]. Advances in Neural Information Processing Systems, 2019, 32(13): 611-624.

[26] VISHNU S. Deep Learning with PyTorch: A practical approach to building neural network models using PyTorch[M]. Birmingham: Packt Publishing Ltd, 2018.

[27] SCHLICHTKRULL M, KIPF T N, BLOEM P, et al. Modeling relational data with graph convolutional networks[C]//Proceedings of the European Semantic Web Conference. Berlin: Springer, 2018: 593-607.

[28] DETTMERS T, MINERVINI P, STENETORP P, et al. Convolutional 2D knowledge graph embeddings[J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2018, 32(1): 1811-1818.

[29] SUN Z Q, DENG Z H, NIE J Y, et al. RotatE: Knowledge graph embedding by relational rotation in complex space[EB/OL]. (2019-02-26) [2022-05-10]. http://arxiv.org/abs/1902.10197.

[30] GARCÍA-DURÁN A, DUMANČIĆ S, NIEPERT M. Learning sequence encoders for temporal knowledge graph completion[C]// Proceedings of the Conference on Empirical Methods in Natural Language Processing. Stroudsburg: Association for Computational Linguistics, 2018: 4816-4821.

[31] DASGUPTA S S, RAY S N, TALUKDAR P. HyTE: Hyperplane-based temporally aware knowledge graph embedding[C]//Proceedings of the Conference on Empirical Methods in Natural Language Processing. Stroudsburg: Association for Computational Linguistics, 2018: 2001-2011.

[32] GOYAL P, CHHETRI S R, CANEDO A. Dyngraph2vec: Capturing network dynamics using dynamic graph representation learning[J]. Knowledge-Based Systems, 2020, 187: 104816.

[33] URIEL S, IDO G, KIRA R. Node embedding over temporal grapns[C]//Proceedings of the International Joint Conference on Artificial Intelligence. Amsterdam: Morgan Kaufmann, 2019: 4605-4612.

[34] PAREJA A, DOMENICONI G, CHEN J, et al. EvolveGCN: Evolving graph convolutional networks for dynamic graphs[J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2020, 34(4): 5363-5370.

[35] SEO Y, DEFFERRARD M, VANDERGHEYNST P, et al. Structured sequence modeling with graph convolutional recurrent networks[C]//Proceedings of the International Conference on Neural Information Processing. Berlin: Springer, 2018: 362-373.

[36] RAKSHIT T, HANJUN D, YICHEN W, et al. Know-Evolve: Deep temporal reasoning for dynamic knowledge graphs[C]//International Conference on Machine Learning. New York: ACM, 2017: 3462-3471.

[37] TRIVEDI R, FARAJTABAR M, BISWAL P, et al. Representation learning over dynamic graphs[EB/OL]. (2018-03-16) [2022-05-10]. https://arxiv.org/abs/1803.04051.

[38] JIN W, QU M, JIN X S, et al. Recurrent event network: Autoregressive structure inferenceover temporal knowledge graphs[C]// Proceedings of the Conference on Empirical Methods in Natural Language Processing. Stroudsburg: Association for Computational Linguistics, 2020: 6669-6683.

Power equipment defect prediction based on temporary knowledge graph

JI Xin[1,2], WU Tongxin[1,*], YANG Zhiwei[1], YU Ting[1], LI Junting[1], HE Yude[1]

(1. Big Data Center, State Grid Corporation of China, Beijing 100031, China;
2. School of Computer Science and Technology, Beihang University, Beijing 100191, China)

Abstract: Power system management now includes monitoring power equipment operation, which is crucial given the growing size of power facilities. Defect prediction of power equipment is a key step in power system operation monitoring. In order to solve the problem of defect prediction for power equipment in large-scale power systems, we propose a defect prediction model for power equipment based on a temporary knowledge graph. The attention mechanism fuses the multimodal data, and the relationship-aware graph neural network and recurrent neural network are then employed to extract the temporal representation of entities and relations. Finally, we perform defect prediction of power equipment based on the temporal representation. The method proposed in this paper can make full use of multimodal information to improve the accuracy of power equipment defect prediction. Experimental results show that the model has considerable performance improvement compared to the baseline model.

Keywords: temporary knowledge graph; graph neural network; defect prediction; power system; power equipment

Received: 2022-09-19; **Accepted**: 2022-10-03; **Published Online**: 2023-02-01 16:48

URL: link.cnki.net/urlid/11.2625.V.20230201.1047.002

Foundation item: State Grid Corporation of China Big Data Center Technology Project (52999021N005)

* **Corresponding author**. E-mail: tongxin-wu@sgcc.com.cn

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2022.0773

基于自适应参数优化 RSSD-CYCBD 的行星齿轮箱复合故障诊断

孙环宇[1]，杨志鹏[2]，王艺玮[2,*]，郭琦[2]

(1. 南京航空航天大学 自动化学院，南京 211106；2. 北京航空航天大学 机械工程及自动化学院，北京 100191)

摘 要：针对行星齿轮箱多振源耦合导致故障源辨识困难、较弱故障特征容易被噪声和较强故障特征掩盖，以及由传播路径引起的信号衰减导致的故障特征微弱等问题，提出一种自适应参数优化的共振稀疏分解 (RSSD) 和最大二阶循环平稳盲解卷积 (CYCBD) 的行星齿轮箱多故障耦合信号分离及诊断算法。根据轴承和齿轮故障的不同共振属性，用 RSSD 算法将多故障耦合信号分解为包含齿轮故障特征的高共振分量和主要包含轴承故障特征的低共振分量后，通过 CYCBD 算法分别对高、低分量进行解卷积，消除传播路径影响和噪声干扰，实现微弱故障特征的增强和提取。特别地，针对 RSSD 和 CYCBD 中参数优化困难、依赖人工经验和自适应差等问题，使用基于松鼠算法 (SSA) 对参数进行自适应优化选取，设计了融合包络谱峭度、自相关函数最大值均方根和特征频率比在内的复合指标作为优化目标。对解卷积后的信号进行包络解调提取故障特征频率，识别不同故障源。通过行星齿轮箱多故障模拟信号和实测信号验证了所提算法的有效性和可行性，进一步地，将所提算法集成在边缘计算设备中，为行星齿轮箱等旋转机械的状态检测诊断及远程运维提供解决方案。

关 键 词：多源故障分离；共振稀疏分解；最大二阶循环平稳盲解卷积；松鼠算法；行星齿轮箱

中图分类号：TH165.3; TP206.3

文献标志码：A **文章编号**：1001-5965(2024)10-3139-12

行星齿轮减速箱具有运行稳定、传动效率高、安全性能好等特点，作为传动系统的关键部件广泛应用于航空航天、兵器、船舶、制造等关乎国家安全与国民经济的工业领域。其工况恶劣、复杂多变且结构耦合性强，容易出现故障，一旦发生故障或性能劣化极易引起整个机械系统的性能下降甚至停机。减速箱内多部件发生故障时会引入多激励源，原始检测信号具有多振源耦合的特点，导致仅凭监测信号难以准确定位到故障源。此外，实际情况中，齿轮传动系统多被层层壳体包裹在复杂装置内部，受空间可达性影响，传感器难以安装在距故障点最敏感的位置，导致信号传播路径复杂、信号衰减、故障特征微弱等问题。因此，如何从混叠了背景噪声和多故障源信号的原始监测信号中识别出目标振源，自适应地分离出不同部件产生的信号并对微弱故障特征进行增强和提取，是故障诊断算法走向实际应用过程中亟待解决的问题。

针对多源故障信号分离与诊断，张文义等结合自适应参数优化的共振稀疏分解(resonance-based sparse signal decomposition, RSSD)和 Teager 能量算

收稿日期：2022-09-14；**录用日期**：2022-09-30；**网络出版时间**：2022-11-08 15:24
网络出版地址：link.cnki.net/urlid/11.2625.V.20221108.1421.003
基金项目：国家科学技术支持计划 (MKF20210012)；国家自然科学基金 (51805262)
* **通信作者**. E-mail：wangyiwei@buaa.edu.cn

引用格式：孙环宇，杨志鹏，王艺玮，等. 基于自适应参数优化 RSSD-CYCBD 的行星齿轮箱复合故障诊断 [J]. 北京航空航天大学学报，2024，50（10）：3139-3150. SUN H Y，YANG Z P，WANG Y W，et al. Compound fault diagnosis of planetary gearbox based on RSSD-CYCBD by adaptive parameter optimization [J]. Journal of Beijing University of Aeronautics and Astronautics，2024，50（10）：3139-3150（in Chinese）.

子解调,实现齿轮箱多故障的分离与诊断[1];黄文涛等通过遗传算法自适应选取RSSD的品质因子,实现行星齿轮与行星架轴承耦合故障的诊断[2];He等通过可调品质因子小波变换构建2个过完备字典,利用分裂增广拉格朗日收缩算法结合优化最小化(majorization–minimization, MM)算法,从齿轮箱振动信号中提取周期瞬态分量和高共振分量用于故障诊断[3];Yang等提出一种基于双字典和分裂增广拉格朗日收缩算法的信号分解算法,通过构建稳态调制字典和冲击调制字典,实现齿轮箱内稳态调制分量和冲击调制分量的分离[4];王霄等使用Symlet8小波字典和余弦字典对RSSD得到的低共振分量进行二次分解,实现对齿轮箱故障脉冲的增强和提取[5]。

上述稀疏分解算法能够有效解决齿轮箱多故障检测中的常见欠定问题,其中,RSSD是Selesnick提出的一种基于信号共振属性的稀疏分解算法[6],与传统的基于频带的分解算法,如短时傅里叶变换、经验模态分解[7](empirical mode decomposition, EMD)、固有时间尺度分解[8](intrinsic time scale decomposition, ITD)等相比,更适用于具有非线性和非平稳特性的振动信号。但是RSSD算法的分解效果随着系统噪声的增大而逐渐变差,因此,需要对RSSD后的信号做进一步处理以抑制噪声。

何群等利用最大相关峭度解卷积(maximum correlated kurtosis deconvolution, MCKD)算法处理RSSD得到的低共振分量,实现对齿轮箱内周期冲击成分的提取[9];齐咏生等利用布谷鸟算法选取RSSD的品质因子,结合多点峭度最优最小熵解卷积修正(multipoint kurtosis optimal minimum entropy deconvolution adjusted, MK-MOMEDA)对低共振分量进行解卷积,实现了单个轴承复合故障的诊断[10]。

上述研究大多只分析了RSSD分解得到的低共振分量中的故障信息,没有对高共振分量进行处理。对于行星齿轮减速箱内轴承和齿轮同时发生故障的情况,高共振分量中也包含大量的故障特征信息,十分值得分析与处理。本文利用RSSD算法对包含轴承和齿轮故障特征的行星齿轮减速箱多故障耦合信号进行分解,分别对高、低共振分量进行解卷积以消除传播路径的影响和强背景噪声的干扰。对于解卷积算法,Buzzoni等提出的最大二阶循环平稳盲解卷积(maximum second order cyclostationary blind deconvolution, CYCBD)算法,通过对比常用的最小熵解卷积(minimun entropy deconvolution, MED)、MCKD等算法在振动信号特征提取上的效果,表明CYCBD算法拥有更强的故障特征提取能力[11]。因此,本文融合RSSD与CYCBD算法,对多故障耦合信号实行先分离后解卷积以实现微弱特征增强的策略。

此外,RSSD和CYCBD算法的参数对最终分解和解卷积的效果有很大影响,目前参数选取依赖人工经验,存在自适应差、对变化敏感等问题。松鼠算法(squirrel search algorithm, SSA)是Jain等于2019年提出的一种新颖的全局优化算法,通过对比粒子群算法、人工蜂群算法、蝙蝠算法和萤火虫算法等群体优化算法在经典数值优化问题上的表现,显示出SSA在优化精度和收敛速度方面有了一定的提高[12]。本文采用SSA分别对RSSD和CYCBD算法的相应参数进行自适应选取,设计了融合包络谱峭度、自相关函数最大值均方根和特征频率比在内的复合指标作为SSA的优化目标。

综上所述,本文针对受强噪声和谐波等干扰的行星齿轮减速箱多故障耦合信号,提出一种自适应参数优化RSSD-CYCBD的多故障分离及诊断算法,用于对行星齿轮箱内齿轮及齿轮轴轴承等多部件故障产生的多激励源耦合信号进行分离与微弱特征增强。进一步地,将自适应参数优化RSSD-CYCBD算法部署到Jetson Nano边缘计算设备中,开发了便携式旋转机械故障监测诊断系统,在转子齿轮传动实验台上进行了验证。

1 自适应参数优化RSSD-CYCBD

1.1 RSSD及CYCBD算法理论

关于RSSD与CYCBD的算法原理可参考文献[6,11],此处仅做简要介绍。RSSD算法将信号中不同成分振荡程度的差异性作为分解依据,通过可调品质因子小波变换(tunable Q-factor wavelet transform, TQWT)分别建立原始信号$\boldsymbol{x}$的高、低品质因子小波字典$\boldsymbol{S}_1$和$\boldsymbol{S}_2$。再基于形态分量分析(morphological component analysis, MCA)算法假定信号$\boldsymbol{x}$可表示为信号$\boldsymbol{x}_1$、$\boldsymbol{x}_2$与残余分量$\boldsymbol{n}_{\mathrm{t}}$之和:

$$\boldsymbol{x}=\boldsymbol{x}_1+\boldsymbol{x}_2+\boldsymbol{n}_{\mathrm{t}}=\boldsymbol{S}_1\boldsymbol{W}_1+\boldsymbol{S}_2\boldsymbol{W}_2+\boldsymbol{n}_{\mathrm{t}} \tag{1}$$

为保证系数矩阵$\boldsymbol{W}_1$和$\boldsymbol{W}_2$的稀疏性,信号分解过程中的目标函数为

$$J(\boldsymbol{W}_1,\boldsymbol{W}_2)=\|\boldsymbol{x}-\boldsymbol{S}_1\boldsymbol{W}_1-\boldsymbol{S}_2\boldsymbol{W}_2\|_2^2+\lambda_1\|\boldsymbol{W}_1\|_1+\lambda_2\|\boldsymbol{W}_2\|_1 \tag{2}$$

式中:$\|\cdot\|_2$为l_2范数;$\|\cdot\|_1$为l_1范数;λ_1、λ_2为正则化参数。

通过分裂增广拉格朗日收缩算法(split augmented Lagrangian shrinkage algorithm, SALSA),求解

$J(\boldsymbol{W}_1,\boldsymbol{W}_2)$最小时的系数矩阵$\boldsymbol{W}_1^*$和$\boldsymbol{W}_2^*$。最后，通过逆 TQWT，得到高共振分量$\boldsymbol{x}_1^*=\boldsymbol{S}_1\boldsymbol{W}_1^*$和低共振分量$\boldsymbol{x}_2^*=\boldsymbol{S}_2\boldsymbol{W}_2^*$。

CYCBD 算法通过求解一个逆滤波器特征向量$\boldsymbol{h}$，使得信号$\boldsymbol{X}$解卷积后的信号$\boldsymbol{s}=\boldsymbol{X}\boldsymbol{h}$的二阶循环平稳性（second-order indicators of cyclostationary, ICS2）最大，其中，I_{CS2}的计算如下：

$$I_{\mathrm{CS2}}=\frac{\boldsymbol{h}^{\mathrm{H}}\boldsymbol{X}^{\mathrm{H}}\boldsymbol{W}\boldsymbol{X}\boldsymbol{h}}{\boldsymbol{h}^{\mathrm{H}}\boldsymbol{X}^{\mathrm{H}}\boldsymbol{X}\boldsymbol{h}}=\frac{\boldsymbol{h}^{\mathrm{H}}\boldsymbol{R}_{\boldsymbol{X}\boldsymbol{W}\boldsymbol{X}}\boldsymbol{h}}{\boldsymbol{h}^{\mathrm{H}}\boldsymbol{R}_{\boldsymbol{X}\boldsymbol{X}}\boldsymbol{h}} \tag{3}$$

式中：上标H表示矩阵的共轭转置；$\boldsymbol{W}$为加权矩阵，与故障特征频率ω有关，具体计算公式参见文献 [11]；$\boldsymbol{R}_{\boldsymbol{X}\boldsymbol{W}\boldsymbol{X}}$和$\boldsymbol{R}_{\boldsymbol{X}\boldsymbol{X}}$分别为加权相关矩阵和相关矩阵。

CYCBD 算法的求解过程可以描述为

$$\boldsymbol{h}_{\max}=\operatorname{argmax}(I_{\mathrm{CS2}}(\boldsymbol{h})) \tag{4}$$

这个过程等价于求解一个广义特征值问题，$\boldsymbol{R}_{\boldsymbol{X}\boldsymbol{W}\boldsymbol{X}}\boldsymbol{h}=\boldsymbol{R}_{\boldsymbol{X}\boldsymbol{X}}\boldsymbol{h}\lambda$的最大特征值$\lambda$所对应的特征向量$\boldsymbol{h}$，可以通过以下迭代过程获得：

步骤 1　初始化逆滤波器特征向量$\boldsymbol{h}$，根据信号$\boldsymbol{X}$计算$\boldsymbol{R}_{\boldsymbol{X}\boldsymbol{X}}$；

步骤 2　根据$\boldsymbol{h}$、$\boldsymbol{X}$和ω计算$\boldsymbol{W}$，进而计算$\boldsymbol{R}_{\boldsymbol{X}\boldsymbol{W}\boldsymbol{X}}$；

步骤 3　通过求解$\boldsymbol{R}_{\boldsymbol{X}\boldsymbol{W}\boldsymbol{X}}\boldsymbol{h}=\boldsymbol{R}_{\boldsymbol{X}\boldsymbol{X}}\boldsymbol{h}\lambda$，得到最大特征值$\lambda$和对应的特征向量$\boldsymbol{h}$；

步骤 4　返回步骤 2，更新逆滤波器$\boldsymbol{h}$，重复整个流程直至达到收敛条件（迭代次数或收敛精度）。

1.2　基于 SSA 的多参数优化

对于 RSSD 算法，从低共振分量的循环平稳性、高共振分量周期性脉冲成分的富集程度，以及高低共振分量中齿轮和轴承故障特征比重 3 个方面考虑，分别设计了包络谱峭度、自相关函数最大值均方根和特征频率比 3 个评价指标，融合成 SSA 优化 RSSD 算法的复合指标。对于 CYCBD 算法，考虑解卷积后信号的循环平稳性和故障特征比重 2 个方面，融合包络谱峭度和特征频率比形成 SSA 优化 CYCBD 算法的复合指标。

综上所述，根据设计的 2 个复合指标，通过 SSA 对 RSSD 和 CYCBD 进行参数的自适应选取。

1.2.1　RSSD 算法优化指标

RSSD 的核心是通过 TQWT 生成与不同故障特征匹配良好的 2 个过完备小波字典，涉及的参数包括小波字典的品质因子 Q、冗余因子r和分解层数J_{d}。2 个过完备小波字典的上述参数分别记为Q_1、r_1、J_1和Q_2、r_2、J_2。

对于一个长度为N_x的信号，在Q和r都确定的情况下，TQWT 的最大分解层数J_{dmax}为

$$J_{\mathrm{dmax}}=\left\lfloor\frac{\ln[N_x/(4Q+4)]}{\ln[(Q+1)/(Q+1-2/r)]}\right\rfloor \tag{5}$$

考虑到 SSA 会处理大量不同$[Q,r]$参数组合下信号的 RSSD，为减少计算量，文献 [13] 规定处于最大层数（第J_{d}层）分量的带宽大于等于 3 倍最大故障特征频率，以便在所需频带内保留足够的轴承或齿轮故障相关信号：

$$J_{\mathrm{d}}=\left\lfloor 1+\frac{\ln\left[6(Q+1)f_{\mathrm{imax}}/f_{\mathrm{s}}\right]}{\ln[1-2/(Qr+r)]}\right\rfloor \tag{6}$$

式中：$\lfloor\cdot\rfloor$表示向下取整；f_{s}为信号的采样频率；f_{imax}为轴承或齿轮的最大故障特征频率。

对于高共振分量，主要包含齿轮故障特征，f_{imax}取齿轮故障类型中故障特征频率最大值：

$$f_{\mathrm{imax}}=\max\left(f_{\mathrm{sun}},f_{\mathrm{planet}},f_{\mathrm{ring}}\right) \tag{7}$$

式中：f_{sun}、f_{planet}和f_{ring}分别为太阳轮、行星轮和内齿圈故障特征频率。对于行星齿轮减速箱，一般太阳轮局部故障特征频率f_{sun}最大。

对于低共振分量，主要包含轴承故障特征，f_{imax}取轴承内圈故障特征频率：

$$f_{\mathrm{imax}}=\max(f_{\mathrm{i}},f_{\mathrm{o}},f_{\mathrm{e}},f_{\mathrm{c}})=f_{\mathrm{i}} \tag{8}$$

式中：f_{i}、f_{o}、f_{e}和f_{c}分别对应轴承内圈、外圈、滚动体和保持架故障特征频率。

不同于针对单一故障的 RSSD 算法可以仅用一种指标衡量分离效果，本文针对多故障源信号的不同特征，从循环平稳性、周期性和故障特征比重 3 个方面综合考虑，设计了 3 个指标融合形成多目标优化复合指标，通过 SSA 来自适应选取 RSSD 算法的参数Q_1、r_1和Q_2、r_2。

根据循环平稳理论，滚动轴承故障信号的瞬态冲击是非平稳的，具有二阶循环平稳性。根据相关文献，包络谱的峭度指标可以用来表征轴承故障信号的循环平稳性[14]。因此，用低共振分量的包络谱峭度（kurtosis of envelope spectrum, KES）作为第 1 个指标，衡量低共振分量的循环平稳性，包络谱峭度计算如下：

$$K_{\mathrm{ES}}=\frac{\dfrac{1}{N_{\mathrm{A}}}\displaystyle\sum_{n=1}^{N_{\mathrm{A}}}\left(A(f_n)-\overline{A}(f)\right)^4}{\left[\dfrac{1}{N_{\mathrm{A}}}\displaystyle\sum_{n=1}^{N_{\mathrm{A}}}\left(A(f_n)-\overline{A}(f)\right)^2\right]^2} \tag{9}$$

式中：$A(f_n)$为包络谱中频率f_n处的幅值；$\overline{A}(f)$为包络谱的均值；N_{A}为包络谱的长度。

第 2 个指标是高共振分量的自相关函数最大值（autocorrelative function maximum, AFM）的均方根[15]，用来衡量高共振分量中的周期性故障脉冲成分。该指标是基于自相关函数提出，如果高共振分

量中含有的周期性故障脉冲成分越多，则指标值越大，计算式如下：

$$A_{\mathrm{FM}}=\sqrt{\sum_{u=1}^{U}\frac{p_u^2}{U}} \tag{10}$$

式中：p_u为自相关函数的最大值；U为自相关函数的延迟次数。

为使齿轮故障信息和轴承故障信息尽可能多的分解到高、低共振分量中，借鉴文献[16]，引入第3个指标，即特征频率比（characteristic frequency ratio, CFR）。该指标是一个无量纲参数，反映的是故障特征频率的能量在包络谱全频带中的能量比，指标值越大，意味着提取的故障信息越多，计算式如下：

$$C_{\mathrm{FR}}=\sum_{k=1}^{K}\sum_{i=1}^{I}A(kF_i)\Big/\sum_{n=1}^{N}A(f_n) \tag{11}$$

式中：$A(kF_i)$为包络谱中故障特征频率F_i的k级谐波分量的幅值；I为故障类型数量；K为故障特征频率的谐波数；N为频谱中的频率成分数。

对于单级行星齿轮箱，有太阳轮、行星轮和内齿圈3种齿轮。本文只考虑局部故障特征，有3种故障特征频率，$I=3$。所以高共振分量中齿轮故障特征频率比可以表示为

$$C_{\mathrm{FRg}}=\frac{\sum_{k=1}^{K}\left(A(kf_{\mathrm{sun}})+A\left(kf_{\mathrm{planet}}\right)+A\left(kf_{\mathrm{ring}}\right)\right)}{\sum_{n=1}^{N}A(f_n)} \tag{12}$$

式中：$A(kf_{\mathrm{sun}})$为包络谱中太阳轮故障频率f_{sun}的第k级谐波分量的幅值；$A\left(kf_{\mathrm{planet}}\right)$为包络谱中行星轮故障频率$f_{\mathrm{planet}}$的第$k$级谐波分量的幅值；$A\left(kf_{\mathrm{ring}}\right)$为包络谱中内齿圈故障频率$f_{\mathrm{ring}}$的第$k$级谐波分量的幅值。

同样的，低共振分量的轴承故障特征频率比可以表示为

$$C_{\mathrm{FRb}}=\frac{\sum_{k=1}^{K}\left(A(kf_{\mathrm{o}})+A(kf_{\mathrm{i}})+A(kf_{\mathrm{e}})+A(kf_{\mathrm{c}})\right)}{\sum_{n=1}^{N}A(f_n)} \tag{13}$$

式中：$A(kf_{\mathrm{o}})$为包络谱中轴承外圈故障特征频率f_{o}的第k级谐波分量的幅值；$A(kf_{\mathrm{i}})$为包络谱中内圈故障特征频率f_{i}的第k级谐波分量的幅值；$A(kf_{\mathrm{e}})$为包络谱中滚动体故障特征频率f_{e}的第k级谐波分量的幅值；$A(kf_{\mathrm{c}})$为包络谱中保持架故障特征频率f_{c}的第k级谐波分量的幅值。

将低共振分量的包络谱峭度K_{ESlow}、高共振分量的自相关函数最大值A_{FMhigh}的均方根及高低共振分量的齿轮和轴承故障特征频率比C_{FRg}和C_{FRb}融合形成SSA优化RSSD算法的复合指标R，可表示为

$$R=A_{\mathrm{FMhigh}}C_{\mathrm{FRg}}+K_{\mathrm{ESlow}}C_{\mathrm{FRb}} \tag{14}$$

通过SSA搜索解空间，求解复合指标R最大时的最优分解参数组合$[Q_1,\ r_1,\ Q_2,\ r_2]$，再结合式(6)构造与不同故障特征匹配良好的小波字典，从而提高RSSD算法对多故障信号的解耦能力。考虑算法的有效性和计算效率，分析文献[2,13,17]中的相关参数设置，将Q_1的搜索区间设置为[3,8]；Q_2的搜索区间设置为[1,3]；r_1和r_2的搜索区间设置为[2,4]。

1.2.2 CYCBD算法优化指标

为解决CYCBD算法参数选取困难的问题，同样应用SSA对滤波器h的长度L和故障特征频率ω这2个关键参数进行自适应的选取。

由于高、低共振分量包含不同故障特征，解卷积后信号s的包络谱峭度$K_{\mathrm{ES}s}$指标不能够很好地评价不同分量的解卷积效果，因此，引入特征频率比$C_{\mathrm{FR}\omega}$衡量不同故障特征，融合2个指标构建联合优化指标P，通过SSA自适应地选取P最大时CYCBD的最优解卷积参数组合$[L,\omega]$，复合优化指标为

$$P=C_{\mathrm{FR}\omega}K_{\mathrm{ES}s} \tag{15}$$

$$C_{\mathrm{FR}\omega}=\sum_{k=1}^{K}A(k\omega)\Big/\sum_{n=1}^{N}A(f_n) \tag{16}$$

式中：$A(k\omega)$为包络谱中故障特征频率ω的第k级谐波分量的幅值。对于高共振分量，ω为齿轮故障特征频率，对于低共振分量，ω为轴承故障特征频率。

考虑算法的有效性和计算效率，将滤波器长度L的搜索区间设置为[100,1 500][18]；故障特征频率ω的搜索区间限制在理论故障特征频率误差的2%以内[19]，即$[\omega(1-2\%),\omega(1+2\%)]$。

1.2.3 自适应RSSD-CYCBD算法

针对多故障源耦合信号，不同于单一故障信号只有一个故障特征频率，如果将CYCBD算法作为预处理放在RSSD前面，由于故障特征频率的不同，需要分别用不同故障特征频率进行解卷积，每一次解卷积都需要进行一次RSSD，这样算法的复杂性将大大的提高。因此，本文采用先RSSD算法分离再CYCBD算法解卷积的策略。

基于RSSD、CYCBD和SSA，自适应参数优化RSSD-CYCBD算法的流程如图1所示。具体流程为：①针对RSSD算法，利用SSA寻找最优参数Q_1、r_1和Q_2、r_2；②将最优参数Q_1、r_1和Q_2、r_2代入RSSD算法，得到高、低共振分量和残差分量；

③对于高共振分量, 包含齿轮故障特征较多, 利用 SSA 寻找 CYCBD 算法的最优参数 L_1、ω_1, 将最优参数带入 CYCBD, 对高共振分量进行解卷积, 实现齿轮故障特征的提取; ④对于低共振分量, 包含轴承故障特征较多, 利用 SSA 寻找 CYCBD 算法的最优参数 L_2、ω_2, 将最优参数带入 CYCBD, 对低共振分量进行解卷积, 实现轴承故障特征的提取; ⑤将通过步骤③和步骤④得到的 2 种解卷积信号分别进行包络解调, 观察包络谱中对应的齿轮故障、轴承故障特征频率是否存在且显著, 从而判断发生何种故障。

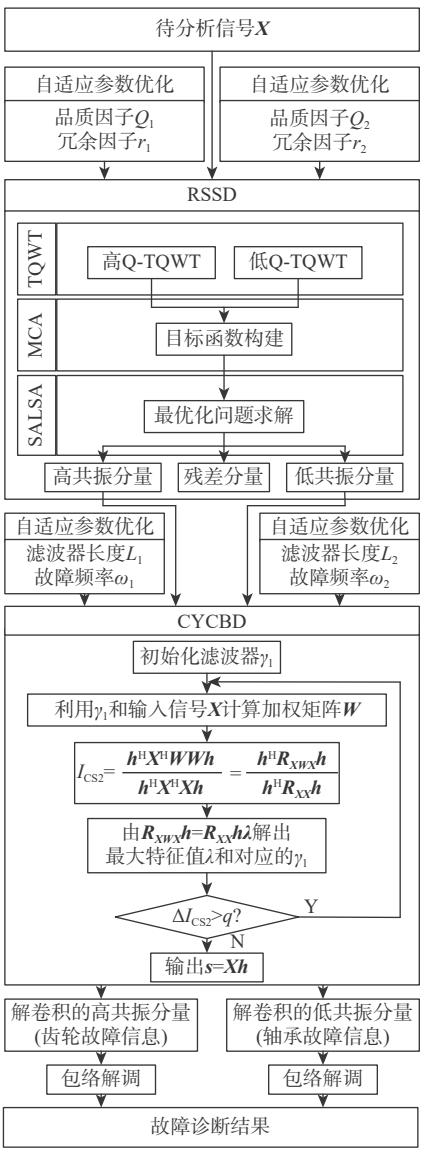


图 1　自适应参数优化 RSSD-CYCBD 算法流程

Fig. 1　Flow of RSSD-CYCBD algorithm by adaptive parameter optimization

2　仿真信号验证

2.1　仿真信号构建

本节通过建立行星齿轮箱多故障耦合信号的仿真模型, 验证本文自适应参数优化算法的可行性。仿真信号的具体形式表示为

$$x(t)=g(t)+b(t)+n(t) \tag{17}$$

式中: $g(t)$为太阳轮局部故障产生的调制信号; $b(t)$为轴承内圈故障激发的瞬态冲击信号; $n(t)$为信噪比为SNR的高斯白噪声, 用来模拟强背景噪声。

由文献 [20] 可知, 太阳轮局部故障的数学模型可表示为

$$\begin{cases} g(t)=S(t)A(t)\cos(2\pi f_{\mathrm{m}}t+B(t)+\theta) \\ A(t)=1+A\cos(2\pi f_{\mathrm{gear}}t+\phi) \\ B(t)=B\sin(2\pi f_{\mathrm{gear}}t+\varphi) \\ S(t)=1-\cos(2\pi f_{\mathrm{s}}^{(\mathrm{r})}t) \end{cases} \tag{18}$$

式中: $S(t)$为太阳轮转动对故障信号的调幅效应; $A(t)$和$B(t)$分别为太阳轮故障的调幅和调频函数; f_{m}为齿轮啮合频率; $f_{\mathrm{s}}^{(\mathrm{r})}$为太阳轮绝对转动频率; f_{gear}为太阳轮故障特征频率; θ为啮合频率的初始相位; A和B分别为调幅和调频幅值; ϕ和φ为啮合频率边带的相位。

轴承内圈故障的数学模型可表示为

$$\begin{cases} b(t)=\sum_{m=1}^{M}A_{\mathrm{m}}h(t-mT_{\mathrm{b}}) \\ h(t)=\exp(-2\pi\varepsilon f_{\mathrm{natural}}t)\sin(2\pi f_{\mathrm{natural}}t) \\ A_{\mathrm{m}}=(1+A_0\cos(2\pi f_{\mathrm{i}}^{(\mathrm{r})}t)) \end{cases} \tag{19}$$

式中: A_{m}为轴承内圈转动引起的调幅效应; A_0为振幅; $f_{\mathrm{i}}^{(\mathrm{r})}$为轴承内圈的转动频率; $h(t)$为指数衰减脉冲; ε为衰减系数; f_{natural}为谐振频率; T_{b}为 2 个连续故障脉冲之间的时间间隔, 则轴承故障特征频率为 $f_{\mathrm{bear}}=1/T_{\mathrm{b}}$; M为故障脉冲数。

仿真信号的参数设置如表 1 所示, 其中, 采样频率为 10 kHz, 数据长度为 $2^{14}=16\,384$, SNR 为信噪比。

将表 1 中的数值代入式 (17)～式 (19), 得到的仿真信号时域波形如图 2 所示。

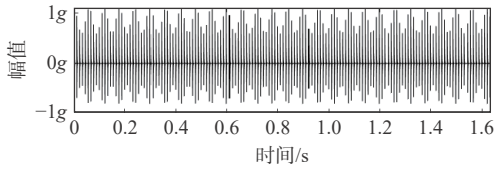
2.2　仿真信号验证

对多故障耦合仿真信号进行分析, 图 3 为其时域波形和包络谱。时域波形图上无明显冲击特征,

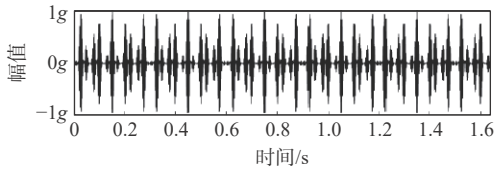
表 1　仿真信号参数

Table 1　Parameters of simulated signal

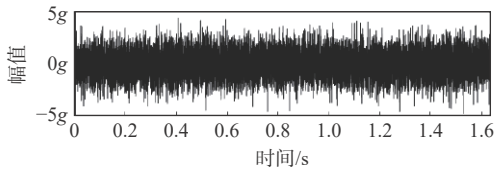
A	B	θ/rad	ϕ/rad	φ/rad	f_{m}/Hz	f_{gear}/Hz	$f_{\mathrm{s}}^{(\mathrm{r})}$/Hz	A_0	$f_{\mathrm{i}}^{(\mathrm{r})}$/Hz	f_{natural}/Hz	ε	M	T_{b}/s	f_{bear}/Hz	SNR/dB
1	1	0	0	0	280	40	16.67	0.3	16.67	2 500	0.1	150	0.011 1	90.091	−12



(a) 轴承内圈故障仿真信号



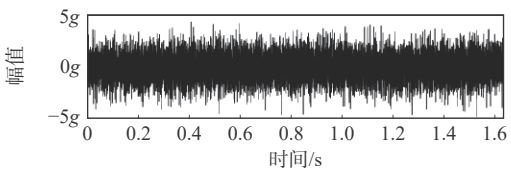
(b) 太阳轮局部故障仿真信号



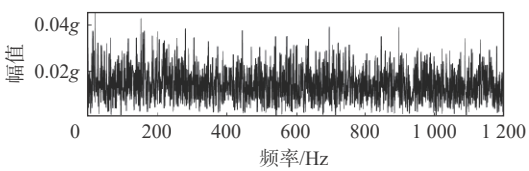
(c) 多故障耦合信号

图 2 仿真信号时域波形

Fig. 2 Time domain waveform of simulated signal



(a) 时域波形



(b) 包络谱

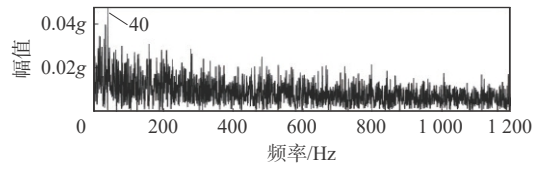
图 3 仿真信号时域波形和包络谱

Fig. 3 Time domain waveform and envelope spectrum of simulated signal

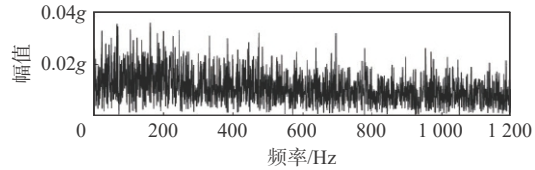
仅通过包络解调无法判断是否发生故障。

根据自适应参数优化 RSSD-CYCBD 算法对仿真信号进行处理。首先通过 RSSD 对图 3 中的故障信号进行稀疏分解，由 SSA 得到的最优品质因子和冗余因子为：$Q_1 = 3.48$，$Q_2 = 1.72$，$r_1 = 3.50$，$r_1 = 3.49$。稀疏分解结果如图 4 所示，可以看出仿真信号在高共振分量的包络谱中太阳轮故障特征频率（40Hz）已经较为明显。而低共振分量的包络谱中轴承和齿轮故障特征频率及倍频均不明显。

为进一步增强与故障相关的微弱特征，消除噪声干扰，采用 CYCBD 分别对高、低共振分量进行解卷积。由 SSA 得到的 2 组最优滤波器长度和故障特征频率为：$L_1 = 187$，$\omega_1 = 39.99$；$L_2 = 568$，$\omega_2 = 90.08$。将这 2 组参数分别代入 CYCBD，仿真信号解卷积后高、低共振分量的包络谱如图 5 所示，从



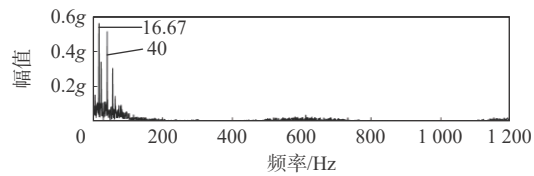
(a) 高共振分量包络谱



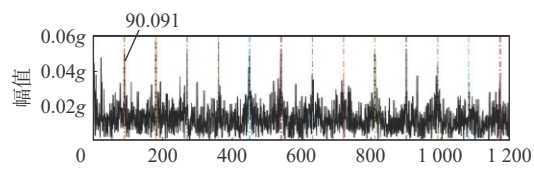
(b) 低共振分量包络谱

图 4 仿真信号高低共振分量的包络谱

Fig. 4 Envelope spectrum of high and low resonance components of simulated signal



(a) 解卷积后高共振分量包络谱



(b) 解卷积后低共振分量包络谱

图 5 仿真信号解卷积后高低共振分量的包络谱

Fig. 5 Envelope spectrum of high and low resonance components after deconvolution of simulated signal

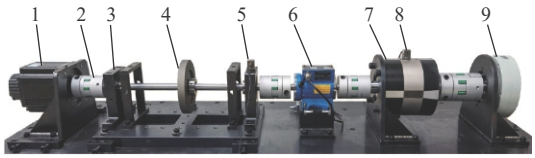
解卷积后高共振分量的包络谱中可以看出太阳轮故障特征频率（40 Hz）和太阳轮轴转动频率（16.67 Hz）处的峰值十分明显。同样，在解卷积后低共振分量的包络谱中轴承内圈故障特征频率（90.09 Hz）及其倍频占绝对的主导。

上述结果表明本文自适应参数优化 RSSD-CYCBD 算法对于受噪声干扰严重（SNR为−12 dB）的行星齿轮箱多故障仿真信号，能够有效地实现故障特征的分离和微弱故障特征增强。

3 实测数据验证及边缘设备部署

3.1 实验台描述

实测数据来自实验室的转子齿轮传动系统综合故障实验台，如图 6 所示，主要由电机、转子主轴、单级行星齿轮减速箱和磁粉加载器等组成，单行星齿轮减速箱齿轮参数如表 2 所示，其中，行星轮共有 3 个。本文所用数据来源网站为 https://gitee.com/phm720/Gearbox-Fault-Data。



1. 电机; 2. 联轴器; 3. 滚动轴承6 205; 4. 飞轮;
5. 振动加速度传感器 (测点2); 6. 转速转矩传感器;
7. 行星齿轮箱; 8. 振动加速度传感器 (测点1); 9.磁粉加载器。

图 6　转子齿轮传动系统综合故障实验台

Fig. 6　Comprehensive fault test bench of rotor-gear transmission system

表 2　单级行星齿轮减速箱齿轮参数

Table 2　Gear parameters of single planetary gear reducer

齿轮	齿数
太阳轮	21
行星轮	31
内齿圈	84

如图 6 所示, 在距离单级行星齿轮减速箱内故障点较近的测点 1 和距离故障点较远的测点 2 处分别采集单级行星齿轮减速箱的振动信号。实验过程中, 与太阳轮相连的输入轴转速为 1 789 r/min, 2 处振动加速度传感器测点的采样频率均为 10 kHz。

根据单级行星齿轮减速箱齿轮参数和输入转速, 计算得到齿轮的理论故障特征频率, 如表 3 所示。

表 3　单级行星齿轮减速箱理论故障特征频率

Table 3　Theoretical fault feature frequency of single planetary gear reducer

啮合频率/Hz	绝对旋转频率/Hz		局部故障特征频率/Hz		
	太阳轮	行星架	太阳轮	行星轮	齿圈
500.92	29.82	5.96	71.56	16.16	17.89

单级行星齿轮减速箱内太阳轮轴承和行星架轴承均为 6212 型号, 根据查得的轴承尺寸参数和工作条件列出轴承的理论故障频率, 如表 4 所示。

表 4　6212 轴承故障特征频率

Table 4　Fault feature frequency of 6212 bearing

故障类型	特征频率/Hz
内圈故障	175.68
外圈故障	122.49
滚动体故障	80.87

为得到单级行星齿轮减速箱中的多故障耦合信号, 分别在太阳轮和太阳轮轴承上加工局部故障。对太阳轮的某一轮齿进行切除, 模拟断齿故障; 在太阳轮轴承外滚道上加工局部故障点, 模拟轴承外圈故障。图 7 和图 8 分别为加工的故障轴承和故障太阳轮的实物照片。

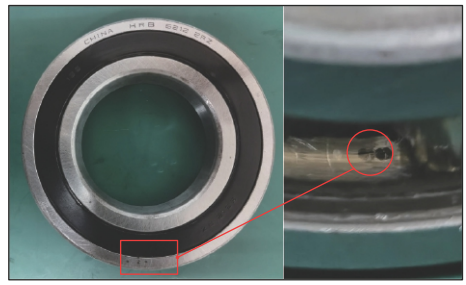


图 7　太阳轮轴承外圈故障实物照片

Fig. 7　Outer ring fault of sun gear bearing

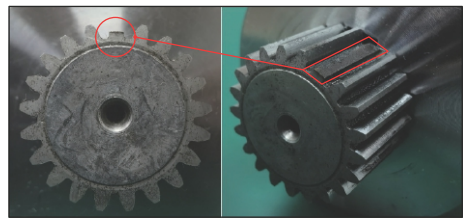


图 8　太阳轮断齿故障实物照片

Fig. 8　Broken tooth fault of sun gear

3.2　实测数据验证

在实际情况中, 传动系统多被层层壳体包裹, 受空间可达性影响, 传感器难以安装在距关键部件最接近、对故障最敏感的位置, 而在壳体外部安装传感器会导致振动信号传播路径长、信号衰减等问题。针对上述问题, 本节将分别用测点 1 和测点 2 这 2 处距离故障源不同距离的传感器采集到的振动数据来验证本文算法。

3.2.1　信号近距离传播下算法验证

测点 1 位于单级行星齿轮减速箱顶, 如图 6 所示, 在传感器不能深入单级行星齿轮减速箱内部的情况下, 是距离故障源最近的测点。

图 9 为测点 1 测得的故障信号时域波形和包络谱。可以看出, 时域波形有较为明显的冲击成分, 对其包络谱进行分析, 太阳轮故障特征频率(72.02 Hz)

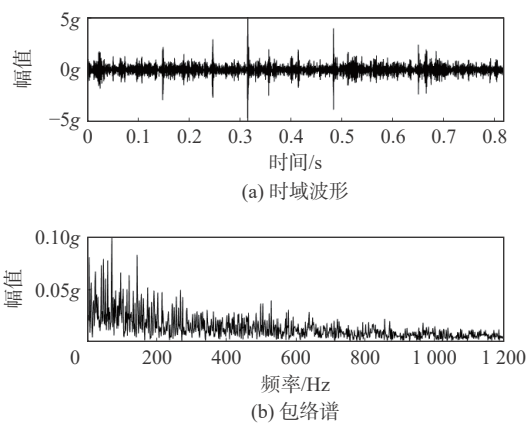
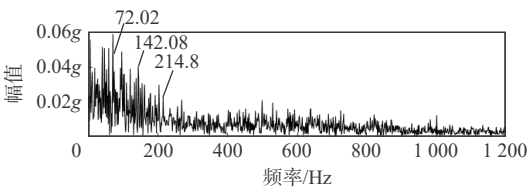


图 9　测点 1 实测信号时域波形图和包络谱

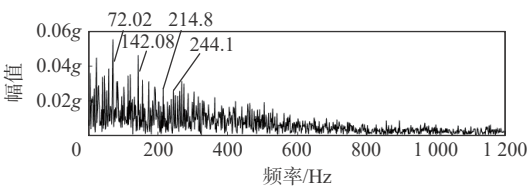
Fig. 9　Time domain waveform and envelope spectrum of measured signal at point 1

和其2倍频、3倍频处均出现较为明显的峰值；而轴承外圈故障特征频率的幅值很小，只在其2倍频(244.1 Hz)处幅值相对较大，表明轴承故障特征很微弱，被齿轮故障特征所掩盖，仅凭包络谱很难确定轴承发生了故障。

为分离多故障耦合信号中的太阳轮断齿故障特征和太阳轮轴承外圈故障特征，根据本文算法对故障信号进行处理。采用RSSD对图9中的多故障耦合信号进行分解，由SSA得到的最优品质因子和冗余因子为：$Q_1=7.03$，$Q_2=2.25$，$r_1=3.96$，$r_2=2.14$。分解结果如图10所示，高、低共振分量的包络谱中，太阳轮故障特征频率和其2倍频、3倍频处的峰值都比较明显，表明发生了齿轮故障。而低共振分量包络谱中轴承外圈故障特征频率2倍频(244.1 Hz)处的幅值相较于原始故障信号和高共振分量的包络谱已经得到初步增强，表明低共振分量包含更多轴承故障特征，但其特征频率(122.49 Hz)仍未能被显示出来，需要对分解后的信号进一步处理。



(a) 高共振分量包络谱



(b) 低共振分量包络谱

图10 测点1信号高低共振分量的包络谱

Fig. 10 Envelope spectrum of high and low resonance components of signal at point 1

采用CYCBD算法分别对高、低共振分量进行解卷积，由SSA得到的2组最优滤波器长度和特征频率分别为：$L_1=1\,131$，$\omega_1=71.677\,8$，$L_2=726$，$\omega_2=122.840\,3$。解卷积结果如图11所示，可以看出在解卷积后高共振分量的包络谱中太阳轮故障特征频率及其倍频占主导地位，表明发生了太阳轮故障。同样，从低共振分量包络谱中也清晰地辨识出轴承外圈故障特征频率。

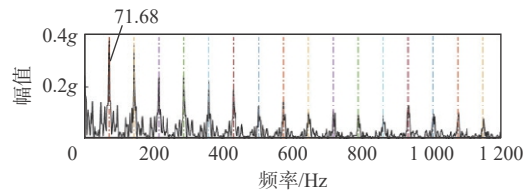
3.2.2 信号长距离传播下算法验证

为进一步验证信号在长距离传播衰减后，本文算法对多故障耦合信号分解的有效性，将传感器安装在图6测点2的位置，距离测点1传感器的水平距离约为170 cm。

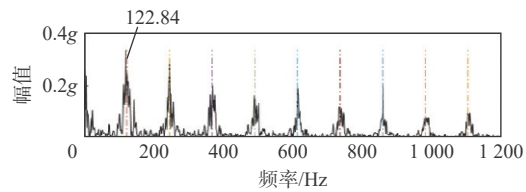
图12为测点2处测得信号的时域波形和包络谱。可以看出，由于测点2距离故障源较远，信号时域波形中的冲击成分相较于测点1处更加不明显。分析对应的包络谱，太阳轮故障特征频率和轴承外圈故障特征频率均无法辨识，根据包络解调的结果无法诊断出任何一种故障。

根据本文算法对测点2处采集的信号进行处理。由SSA得到RSSD算法的最优品质因子和冗余因子为：$Q_1=6.05$，$Q_2=2.44$，$r_1=2.86$，$r_1=2.12$，分解结果如图13所示。

由于传感器距离故障源较远，即使经过RSSD分解，高、低共振分量中的故障特征仍然不明显。采用CYCBD对高、低共振分量分别进行解卷积，由SSA得到的2组最优滤波器长度和特征频率分别为：$L_1=1\,060$，$\omega_1=72.430\,5$；$L_2=1\,420$，$\omega_2=122.436\,0$。解卷积结果如图14所示，解卷积后高共振分量的包络谱中太阳轮故障特征频率被清晰地识别出来。同样，从解卷积后低共振分量的包络谱也可以



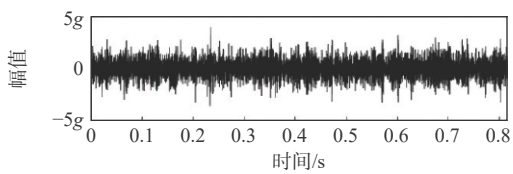
(a) 解卷积后高共振分量包络谱



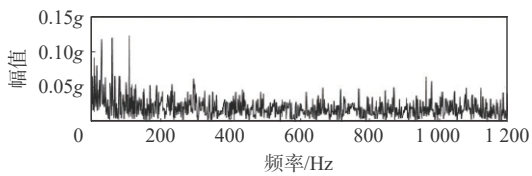
(b) 解卷积后低共振分量包络谱

图11 测点1处CYCBD解卷积后高低共振分量的包络谱

Fig. 11 Envelope spectrum of high and low resonance components after deconvolution by CYCBD at point 1



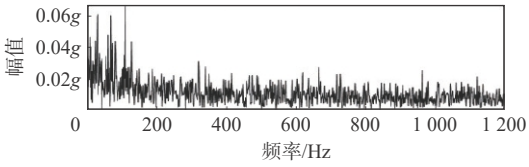
(a) 时域波形



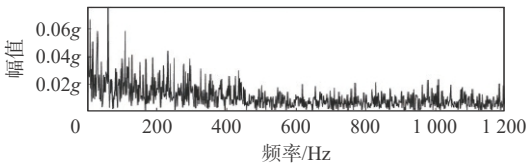
(b) 包络谱

图12 测点2实测信号时域波形图和包络谱

Fig. 12 Time domain waveform and envelope spectrum of measured signal at point 2



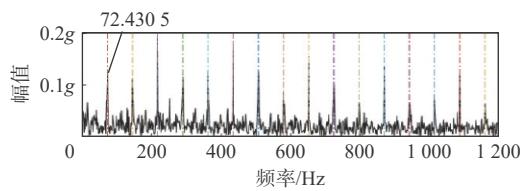
(a) 高共振分量包络谱



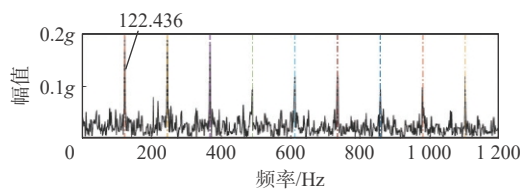
(b) 低共振分量包络谱

图 13　测点 2 信号高低共振分量的包络谱

Fig. 13　Envelope spectrum of high and low resonance components of signal at point 2



(a) 解卷积后高共振分量包络谱



(b) 解卷积后低共振分量包络谱

图 14　测点 2 处 CYCBD 解卷积后高低共振分量的包络谱

Fig. 14　Envelope spectrum of high and low resonance components after deconvolution by CYCBD at point 2

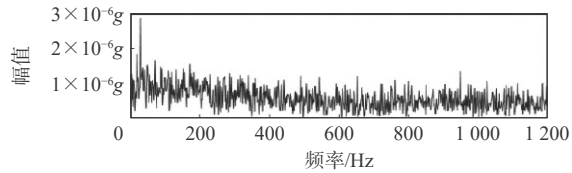
观察到十分明显的轴承外圈故障特征频率。

为说明 CYCBD 算法在解卷积方面的优越性，将其和常用的解卷积算法进行对比。用 MED 和 MCKD 算法分别去处理图 13 中的高、低共振分量，滤波器长度分别设置为 $L_1 = 1\,131$ 和 $L_2 = 726$，这与 CYCBD 算法的滤波器长度保持一致。MCKD 算法中的故障特征频率也同样设置为 $\omega_1 = 71.677\,8$ 和 $\omega_2 = 122.840\,3$。图 15 和图 16 分别为 MED 和 MCKD 这 2 种算法的解卷积结果，对比图 14 中 CYCBD 解卷积结果可以发现，CYCBD 算法解卷积能力更强。

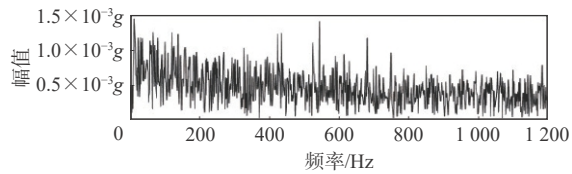
综上所述，测点 1 和测点 2 处实测信号的算法验证结果，特别是测点 2 处耦合信号的分离及诊断结果，表明本文算法为多故障源耦合信号分解、长距传播衰减导致故障特征微弱等实际故障诊断场景，提供了较好的解决方案。

3.3　算法的边缘计算设备部署

在旋转机械状态监测诊断与运维场景中，长时



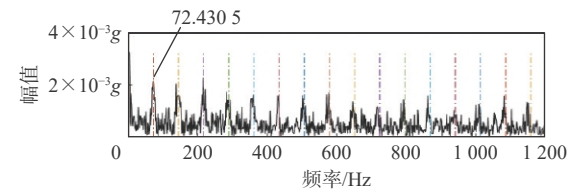
(a) 解卷积后高共振分量包络谱



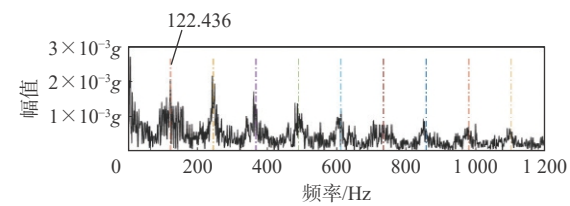
(b) 解卷积后低共振分量包络谱

图 15　测点 2 处 MED 解卷积后高低共振分量的包络谱

Fig. 15　Envelope spectrum of high and low resonance components after deconvolution by MED at point 2



(a) 解卷积后高共振分量包络谱



(b) 解卷积后低共振分量包络谱

图 16　测点 2 处 MCKD 解卷积后高低共振分量的包络谱

Fig. 16　Envelope spectrum of high and low resonance components after deconvolution by MCKD at point 2

间、多设备、多测点、多信号类型、高采样率的监测模式导致监测数据体量巨大，而边缘计算范式通过在靠近设备的边缘端部署低廉的计算设备如智能芯片，对监测数据就地处理，如冗余数据过滤、故障诊断等，进而得到设备的健康状态，再将设备健康状态及相应的运维决策发布至中控运维管理人员。这种模式能够在保障计算性能的同时有效减少边缘端数据传输带来的成本，并且隐私数据在设备端处理能够避免泄露风险，非常适用于批量大、地理位置分布广的旋转机械实时状态监测故障诊断与运维。

基于边缘计算思想，将本文算法集成到英伟达研发的 Jetson Nano 智能芯片上，开发了便携式旋转机械故障诊断系统，如图 17 所示。Jetson Nano 相比于一般单片机，集成了图形处理器(GPU)，具有更强大的计算力，能够运行多种算法和人工智能模型框架，其相比于上位机，价格更加低廉，适合大范围部署，结合数据采集仪、传感器等硬件，可有效地

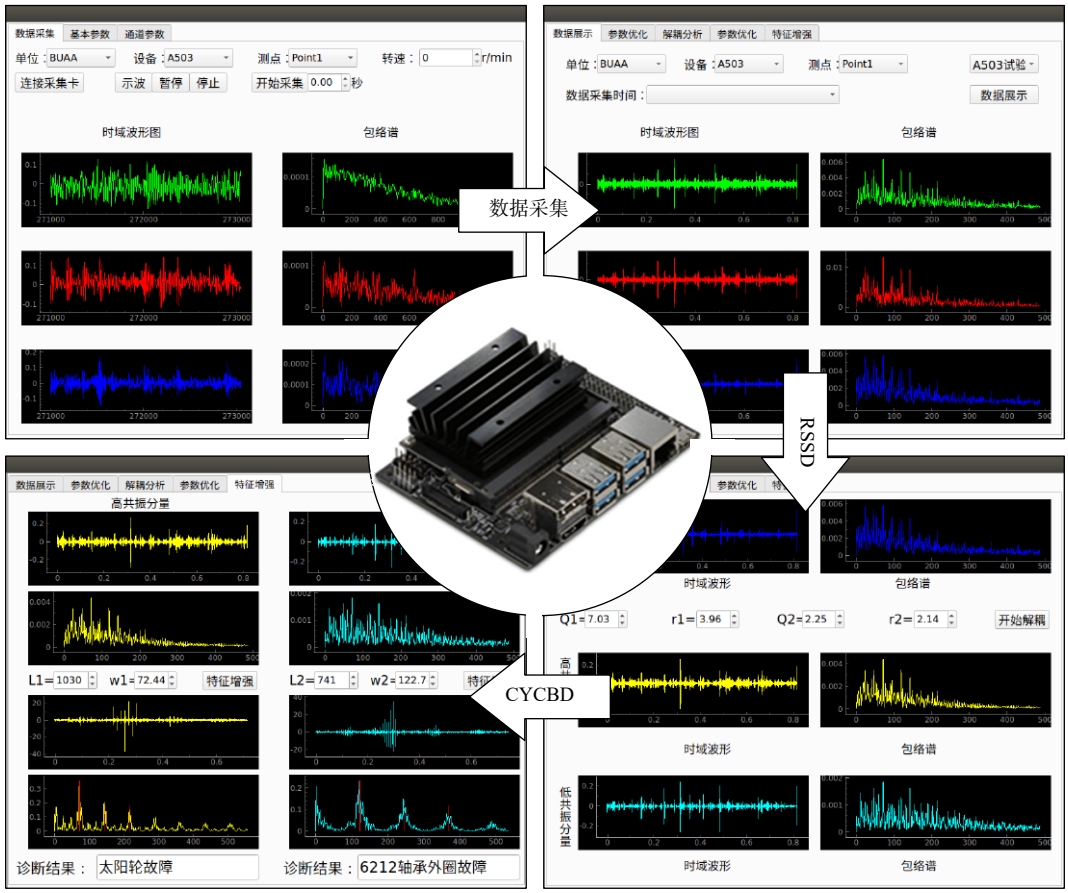


图 17 Jetson Nano 开发板数据采集与故障分离诊断界面

Fig. 17 Data acquisition and fault separation and diagnosis interface on Jetson Nano development board

在边缘端对设备进行数据采集、分析处理、故障诊断，只将处理结果上传给远程运维人员，为旋转机械设备的远程智能运维提供了一种边缘-云端结合的新思路。

4 结 论

针对行星齿轮箱多激励源耦合故障受噪声和谐波干扰严重及算法参数选择困难等问题，本文提出一种自适应参数优化 RSSD-CYCBD 的多故障耦合信号分离及诊断算法，得出如下结论。

1）对于单级行星齿轮减速箱内包含轴承和齿轮故障的多激励源耦合故障，自适应参数优化的 RSSD 算法分离的高、低共振分量，各自包含着较多的齿轮故障特征信息和轴承故障特征信息。

2）对于被齿轮故障特征所掩盖的轴承故障特征，经过 RSSD 算法分离之后，自适应参数优化的 CYCBD 算法可以将微弱的故障特征有效地提取出来。

3）基于自适应参数优化 RSSD-CYCBD 算法，在进行振动信号采集时，对测点的位置要求降低。

4）基于 SSA 的自适应参数优化，这一过程耗费较长时间，与诊断的时效性有所冲突，后续需要进一步优化。

5）基于边缘计算思想，将本文算法部署到体积小、算力强、成本低的智能芯片 Jetson Nano 上，为旋转机械的远程智能运维提供新的解决方案。

参考文献（References）

[1] 张文义, 于德介, 陈向民. 齿轮箱复合故障诊断的信号共振分量能量算子解调方法[J]. 振动工程学报, 2015, 28(1): 148-155.
ZHANG W Y, YU D J, CHEN X M. Energy operator demodulating of signal's resonance components for the compound fault diagnosis of gearbox[J]. Journal of Vibration Engineering, 2015, 28(1): 148-155 (in Chinese).

[2] 黄文涛, 付强, 窦宏印. 基于自适应优化品质因子的共振稀疏分解方法及其在行星齿轮箱复合故障诊断中的应用[J]. 机械工程学报, 2016, 52(15): 44-51.
HUANG W T, FU Q, DOU H Y. Resonance-based sparse signal decomposition based on the quality factors optimization and its application of composite fault diagnosis to planetary gearbox[J]. Journal of Mechanical Engineering, 2016, 52(15): 44-51(in Chinese).

[3] HE W P, CHEN B Q, ZENG N Y, et al. Sparsity-based signal extraction using dual Q-factors for gearbox fault detection[J]. ISA Transactions, 2018, 79: 147-160.

[4] YANG X Q, DING K, HE G L, et al. Double-dictionary signal decomposition method based on split augmented Lagrangian shrinkage algorithm and its application in gearbox hybrid faults

diagnosis[J]. Journal of Sound Vibration, 2018, 432: 484-501.

[5] 王霄, 谢平, 郭源耕, 等. 基于多字典-共振稀疏分解的脉冲故障特征提取[J]. 中国机械工程, 2019, 30(20): 2456-2462.
WANG X, XIE P, GUO Y G, et al. Impulse fault signature extraction based on multi-dictionary resonance-based sparse signal decomposition[J]. China Mechanical Engineering, 2019, 30(20): 2456-2462 (in Chinese).

[6] SELESNICK I W. Resonance-based signal decomposition: A new sparsity-enabled signal analysis method[J]. Signal Processing, 2011, 91(12): 2793-2809.

[7] 张琳, 黄敏. 基于 EMD 与切片双谱的轴承故障诊断方法[J]. 北京航空航天大学学报, 2010, 36(3): 287-290.
ZHANG L, HUANG M. Fault diagnosis approach for bearing based on EMD and slice bi-spectrum[J]. Journal of Beijing University of Aeronautics and Astronautics, 2010, 36(3): 287-290(in Chinese).

[8] 余建波, 吕靖香, 程辉, 等. 基于 ITD 和改进形态滤波的滚动轴承故障诊断[J]. 北京航空航天大学学报, 2018, 44(2): 241-249.
YU J B, LYU J X, CHENG H, et al. Fault diagnosis for rolling bearing based on ITD and improved morphological filter[J]. Journal of Beijing University of Aeronautics and Astronautics, 2018, 44(2): 241-249(in Chinese).

[9] 何群, 郭源耕, 王霄, 等. 基于信号共振稀疏分解和最大相关峭度解卷积的齿轮箱故障诊断[J]. 中国机械工程, 2017, 28(13): 1528-1534.
HE Q, GUO Y G, WANG X, et al. Gearbox fault diagnosis based on RB-SSD and MCKD[J]. China Mechanical Engineering, 2017, 28(13): 1528-1534(in Chinese).

[10] 齐咏生, 樊佶, 李永亭, 等. 一种改进的解卷积算法及其在滚动轴承复合故障诊断中的应用[J]. 振动与冲击, 2020, 39(21): 140-150.
QI Y S, FAN J, LI Y T, et al. An improved deconvolution algorithm and its application in compound fault diagnosis of rolling bearing[J]. Journal of Vibration and Shock, 2020, 39(21): 140-150(in Chinese).

[11] BUZZONI M, ANTONI J, D'ELIA G. Blind deconvolution based on cyclostationarity maximization and its application to fault identification[J]. Journal of Sound Vibration, 2018, 432: 569-601.

[12] JAIN M, SINGH V, RANI A. A novel nature-inspired algorithm for optimization: Squirrel search algorithm[J]. Swarm and Evolutionary Computation, 2019, 44(2): 148-175.

[13] 王晓龙, 唐贵基, 周福成. 自适应可调品质因子小波变换在轴承早期故障诊断中的应用[J]. 航空动力学报, 2017, 32(10): 2467-2475.
WANG X L, TANG G J, ZHOU F C. Application of adaptive tunable Q-factor wavelet transform on incipient fault diagnosis of bearing[J]. Journal of Aerospace Power, 2017, 32(10): 2467-2475(in Chinese).

[14] 顾晓辉, 杨绍普, 刘永强, 等. 基于多目标交叉熵优化的轮对轴承故障特征提取方法[J]. 机械工程学报, 2018, 54(4): 285-292.
GU X H, YANG S P, LIU Y Q, et al. Fault feature extraction of wheel-bearing based on multi-objective cross entropy optimization[J]. Journal of Mechanical Engineering, 2018, 54(4): 285-292 (in Chinese).

[15] CHEN D Y, LIN J H, LI Y P. Modified complementary ensemble empirical mode decomposition and intrinsic mode functions evaluation index for high-speed train gearbox fault diagnosis[J]. Journal of Sound and Vibration, 2018, 424: 192-207.

[16] 李永波. 滚动轴承故障特征提取与早期诊断方法研究[D]. 哈尔滨: 哈尔滨工业大学, 2017: 101-102.
LI Y B. Investigation of fault feature extraction and early fault diagnosis for rolling bearings[D]. Harbin: Harbin Institute of Technology, 2017: 101-102(in Chinese).

[17] 张守京, 慎明俊, 杨静雯, 等. 改进的共振稀疏分解方法及其在滚动轴承复合故障诊断中的应用[J]. 中国机械工程, 2022, 33(14): 1697-1706.
ZHANG S J, SHEN M J, YANG J W, et al. Improved RSSD and its applications to composite fault diagnosis of rolling bearings[J]. China Mechanical Engineering, 2022, 33(14): 1697-1706 (in Chinese).

[18] WANG C G, LI H K, OU J Y, et al. Identification of planetary gearbox weak compound fault based on parallel dual-parameter optimized resonance sparse decomposition and improved MOMEDA[J]. Measurement, 2020, 165: 108079.

[19] 黄包裕, 张永祥, 赵磊. 基于布谷鸟搜索算法和最大二阶循环平稳盲解卷积的滚动轴承故障诊断方法[J]. 机械工程学报, 2021, 57(9): 99-107.
HUANG B Y, ZHANG Y X , ZHAO L. Research on fault diagnosis method of rolling bearings based on cuckoo search algorithm and maximum second order cyclostationary blind deconvolution[J]. Journal of Mechanical Engineering, 2021, 57(9): 99-107(in Chinese).

[20] 冯志鹏, 赵镭镭, 褚福磊. 行星齿轮箱齿轮局部故障振动频谱特征[J]. 中国电机工程学报, 2013, 33(5): 119-127.
FENG Z P, ZHAO L L, CHU F L. Vibration spectral characteristics of localized gear fault of planetary gearboxes[J]. Proceedings of the CSEE, 2013, 33(5): 119-127(in Chinese).

Compound fault diagnosis of planetary gearbox based on RSSD-CYCBD by adaptive parameter optimization

SUN Huanyu[1]，YANG Zhipeng[2]，WANG Yiwei[2,*]，GUO Qi[2]

(1. College of Automation Engineering，Nanjing University of Aeronautics and Astronautics，Nanjing 211106，China；
2. School of Mechanical Engineering & Automation，Beihang University，Beijing 100191，China)

Abstract：The coupling of multiple vibration sources of planetary gearboxes results in difficulty in identifying fault sources, and weak fault features are easily masked by noise and strong fault features. In addition, signal attenuation caused by the propagation path causes weak fault features. To address these issues, a multi-fault coupled signal separation and diagnosis method for planetary gearboxes utilizing resonance-based sparse signal decomposition (RSSD) by adaptive parameter optimization and maximum second order cyclostationary blind deconvolution (CYCBD) was proposed. According to the different resonance properties of bearing faults and gear faults, the multi-fault coupled signal was divided into high resonance components containing gear fault features and low resonance components mainly containing bearing fault features by RSSD. Then, the two components were treated by the CYCBD to eliminate the influence of the propagation path and noise interference, so as to enhance and extract weak fault features. In particular, to solve the problems of difficulty in parameter optimization, dependence on artificial experience, and poor adaptation in RSSD and CYCBD, an adaptive parameter optimization method based on the squirrel search algorithm (SSA) was proposed, and a composite index integrating kurtosis of envelope spectrum, root mean square of autocorrelation function maximum, and characteristic frequency ratio was designed as an optimization objective. Finally, envelope demodulation was performed on the deconvolved signal to extract the fault feature frequency and identify different fault sources. The effectiveness and feasibility of the proposed algorithm were verified by the multi-fault simulation signal and the measured signal of the planetary gearbox. Moreover, the proposed method was integrated into edge computing equipment to provide solutions for state detection and diagnosis, as well as remote operation and maintenance of rotating machinery such as planetary gearboxes.

Keywords：multi-source fault separation；resonance-based sparse signal decomposition；maximum second order cyclostationary blind deconvolution；squirrel search algorithm；planetary gearbox

Received：2022-09-14；**Accepted**：2022-09-30；**Published Online**：2022-11-08 15：24

URL：link.cnki.net/urlid/11.2625.V.20221108.1421.003

Foundation items：Nation Science and Technology Support Program of China (MKF20210012); National Natural Science Foundation of China (51805262)

* **Corresponding author**. E-mail：wangyiwei@buaa.edu.cn

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2023.0049

面向突防的滑翔制导炮弹弹道规划方法

尹秋霖，陈琦*，王中原，王庆海
(南京理工大学 能源与动力工程学院，南京 210094)

摘 要：针对滑翔制导炮弹在不可避免的威胁区域内选择突防方案的问题，从量化威胁值的角度建立了敌方防御手段的数学模型，基于模型设计了全程综合威胁值最低的规划指标，提出考虑目标防御威胁的弹道规划方法。为实现滑翔制导炮弹全程飞行过程中初始弹道倾角、偏角、火箭点火时刻、滑翔启控时刻等各参数的最佳匹配，建立了多阶段全弹道规划模型，并采用 hp 自适应伪谱法将最优控制问题转换为非线性规划问题求解。通过仿真验证了在该指标下滑翔制导炮弹对目标防御的规避效果，分析了影响有效性的因素。与传统弹道规划方法进行对比，证明了所提方法的优越性。

关 键 词：滑翔制导炮弹；突防；威胁建模；全弹道规划；hp 自适应伪谱法
中图分类号：TJ413+.6
文献标志码：A **文章编号**：1001-5965(2024)10-3151-11

方案弹道规划是滑翔制导炮弹(简称滑翔弹)设计中至关重要的一环，方案弹道设计的好坏将直接影响滑翔弹的作战效果[1]。规划性能指标是评价方案弹道优劣的标准，针对此问题已有大量的文献研究。传统指标有最远射程、最大升阻比等[2]。随着现代战争的发展，对于制导炮弹新的作战需求催生出了新的性能指标。文献 [3] 以最短飞行时间、最优控制能量等为性能指标，使用序列二次规划(sequence quadratic program, SQP)算法规划最优滑翔轨迹。文献 [4]提出滑翔效率、机动效率及复合效率因子的概念，并在此基础上使用伪谱法进行方案弹道设计，取得了不错的效果。然而实际战场情况复杂多样，不仅要着眼于弹丸自身攻击特性，还要考虑目标的防御手段。当前防空体系日趋严密，为提高滑翔弹的突防成功率，对复杂的战场环境进行评估，规划出一条整体威胁最小的突防弹道，具有现实意义。

评估战场环境需要对弹丸所要经过的战场空间中各种威胁源进行建模。Xiao 等[5] 针对再入弹道导弹建立了威胁评估指标体系框架，并给出量化方法。Kumar 和 Tripathi[6] 采用贝叶斯网络进行威胁评估并应用于任务分配，但需要大量的先验信息，推演过程复杂。王勃等[7] 在文献 [6] 的基础上改进使用影响网建立了战场威胁强度模型。Lee 等[8] 通过马尔可夫链预测威胁值并使用神经网络进行建模，取得了不错的效果，但计算量巨大，不利于工程实践。

以上研究均侧重于威胁评估，未将威胁建模应用于方案弹道的规划。刘娟等[9-11] 将效能评估中的指数法引入威胁度计算模型，并基于仿真数据，描述了作战空域的威胁分布情况，为航迹规划提供了参考。在此基础上，文献 [12-13] 将威胁细分为软、硬威胁，利用势场的概念进行量化，生成了三维威胁空间。文献 [14-16] 对复杂的三维地形进行建模，规划了无人机的避障路径。文献 [17-18] 在地形威胁的基础上，对预警雷达进行了建模。文献 [19] 对雷达威胁进行了更深入的研究，指出最重要的影响因素是雷达与无人机之间的交战几何结构。在此基础上，文献 [20-22] 考虑了气象条件、地空导弹、电磁干扰等威胁，规划了飞机与无人机的三维

收稿日期：2023-02-14；**录用日期**：2023-06-02；**网络出版时间**：2023-06-26 14:17
网络出版地址：link.cnki.net/urlid/11.2625.V.20230626.1007.001
基金项目：国家自然科学基金 (52202475)；江苏省自然科学基金 (BK20200498)
* **通信作者**. E-mail：qichen@njust.edu.cn

引用格式：尹秋霖，陈琦，王中原，等. 面向突防的滑翔制导炮弹弹道规划方法 [J]. 北京航空航天大学学报，2024，50（10）：3151-3161.
YIN Q L，CHEN Q，WANG Z Y，et al. Trajectory programming method of gliding-guided projectiles for penetration [J]. Journal of Beijing University of Aeronautics and Astronautics，2024，50（10）：3151-3161（in Chinese）.

航迹。文献 [23] 为提高导弹的突防效率，提出安装干扰机的突防模型，并通过仿真验证了干扰的有效性。文献 [24] 基于威胁建模提出一种在线规划算法，文献 [25-26] 引入混沌学习与强化学习的思想，但若要获取高质量的弹道轨迹，计算量巨大。

对于滑翔弹的全程飞行弹道来说，除滑翔段准攻角、侧滑角的动态控制参数变化规律，初始弹道倾角、偏角、火箭点火时刻、滑翔启控时刻等静态参数对弹道规划也有重要影响，而且各参数之间存在耦合效应[3,27]。具体到实际的作战任务中，这些参数的匹配存在最佳组合。可见，为保证方案弹道的最优性，有必要进行全弹道规划。滑翔弹的全弹道规划是一个多阶段、多约束的最优控制问题。近年来，伪谱法在解决轨迹优化问题方面显示出较大的优势。伪谱法采用全局正交多项式逼近状态变量和控制变量，属于全局优化方法，具有快速收敛、对初值敏感度不高的特点。hp 自适应伪谱法在其基础上结合了有限元区间划分的思想，相较于其他伪谱法，对状态变量与控制变量的近似程度更高，在解的精度上更具优越性[28]。

基于上述分析，本文选取某型滑翔弹为研究对象，结合工程实际与相关参考文献，选取打击目标为拥有近程地面警戒雷达、防空高炮、近防密集阵的舰船或小型军事基地，建立了目标防御的威胁值模型，基于该模型提出考虑突防时间与威胁值的方案弹道性能指标，并提出考虑目标防御威胁的弹道规划方法。采用 hp 自适应伪谱法求解全弹道最优控制问题，并通过不同作战场景的仿真及相同场景下与传统弹道规划方法的对比，验证了本文方法的有效性与优越性。

1 面向突防的滑翔弹弹道规划模型

1.1 滑翔弹运动模型

在设计滑翔弹方案弹道时，为把握住问题的本质，以便研究飞行弹道的特性，可将其视作一个可操纵的质点。假设制导炮弹在飞行过程中满足瞬时平衡，气象为标准条件，控制系统理想工作，且不考虑外界因素的干扰[29]。

滑翔弹质心运动方程组为

$$\begin{cases} m\dot{V} = P\cos\alpha^*\cos\beta^* - X - mg\sin\theta \\ mV\dot{\theta} = P\sin\alpha^* + Y - mg\cos\theta \\ -mV\dot{\psi}_{\rm v}\cos\theta = -P\cos\alpha^*\sin\beta^* + Z \\ \dot{x} = V\cos\theta\cos\psi_{\rm v} \\ \dot{y} = V\sin\theta \\ \dot{z} = -V\cos\theta\sin\psi_{\rm v} \\ \dot{m} = -m_{\rm c} \end{cases} \tag{1}$$

式中：m为滑翔弹的质量；g为重力加速度；V为弹体速度；θ为弹道倾角；$\psi_{\rm v}$为弹道偏角；P为助推火箭发动机的平均推力；x、y、z分别为制导炮弹在地面坐标系中的位置；α^*和β^*分别为针对低速滚转制导炮弹而定义的准攻角和准侧滑角，详见文献 [2]；$m_{\rm c}$为制导炮弹单位时间内质量消耗，其值由火箭助推发动机性能确定。

制导炮弹所受阻力、升力和侧向力的表达式为

$$\begin{cases} X = qSC_{x0}\left[1 + k_{\rm c}\left((\alpha^{*^2}) + (\beta^{*^2})\right)\right] \\ Y = qSC_y^{\alpha}\alpha^* \\ Z = -qSC_y^{\alpha}\beta^* \end{cases} \tag{2}$$

式中：$q = \rho V^2/2$为动压，ρ为空气密度，取值参考文献 [29]；S为制导炮弹特征面积；C_{x0}为零升阻力系数，C_y^{α}为升力系数相对于攻角的偏导数，C_{x0}和C_y^{α}的取值由风洞试验数据插值得到，$k_{\rm c}$为与诱导阻力系数有关的参数。对于本文研究的制导炮弹，可视为轴对称飞行器，因此，侧向力系数使用$-C_y^{\alpha}$来表示，其负号是由α^*和β^*的定义决定的。

1.2 威胁模型

在制导炮弹的突防过程中，威胁来源随时间和攻防双方的实时状态而动态变化。在以往的文献中，威胁大多被简化，等效为圆柱、圆锥、半球等，这样的处理只能表示威胁的存在，无法体现出其大小变化，也无法展示多个威胁之间的耦合作用[30-31]。同时，已有文献在方案弹道规划过程中大多将威胁考虑为禁飞区约束的形式进行绕行，无法反映出可穿越的、不可避免的软威胁。本文针对滑翔弹突防过程中的对抗情况，将威胁源分为以下 3 类进行讨论。

1）近程地面警戒雷达。警戒雷达构成的威胁与其对目标的探测概率成正比。根据雷达方程[17]，结合雷达本身的性能和空间相对位置关系，定义雷达的威胁值函数为

$$P_{\rm r} = \begin{cases} 0 & R_{\rm ra} \geqslant R_{\rm rmax} \\ \dfrac{R_{\rm rmax} - R_{\rm ra}}{R_{\rm rmax} + R_{\rm ra}} \cdot \dfrac{R_{\rm rmax}^4}{R_{\rm ra}^4 + R_{\rm rmax}^4} & R_{\rm ra} < R_{\rm rmax} \end{cases} \tag{3}$$

式中：$R_{\rm ra}$为滑翔弹与雷达的径向距离；$R_{\rm rmax}$为雷达威胁覆盖的最大半径。

若同时存在多个雷达，则

$$P_{\rm r} = 1 - \prod_{i=1}^{n}(1 - P_{ri}) \qquad i = 1,2,\cdots,n \tag{4}$$

2）防空高炮。根据工程经验与现有文献可知，防空高炮存在攻击范围限制，在其射程范围之外，过远或过近都无法构成威胁。本文结合其自身的特性和空间相对位置关系，定义防空高炮的威胁值函数为

$$P_{\mathrm{g}}=\begin{cases}0 & R\in\left[0,R_{\mathrm{gmin}}\right]\cup\left(R_{\mathrm{gmax}},+\infty\right)\\ \dfrac{1}{2}\left\{\sin\left[\dfrac{4\pi}{\Delta R}\left(R-R_{\mathrm{gmin}}\right)-\dfrac{\pi}{2}\right]+1\right\} & R\in\left(R_{\mathrm{gmin}},R_{\mathrm{gmin}}+\dfrac{\Delta R}{4}\right]\\ 1 & R\in\left(R_{\mathrm{gmin}}+\dfrac{\Delta R}{4},R_{\mathrm{gmax}}-\dfrac{\Delta R}{4}\right]\\ \dfrac{1}{2}\left\{\cos\left[\dfrac{4\pi}{\Delta R}\left(R_{\mathrm{gmax}}-\dfrac{\Delta R}{4}-R\right)\right]+1\right\} & R\in\left(R_{\mathrm{gmax}}-\dfrac{\Delta R}{4},R_{\mathrm{gmax}}\right]\end{cases}\tag{5}$$

假设高炮位于防区中心或舰上, R为弹与防区中心的径向距离, R_{gmin}为高炮的攻击近界, R_{gmax}为攻击远界, $\Delta R=R_{\mathrm{gmax}}-R_{\mathrm{gmin}}$。

3) 近防密集阵。根据工程经验与现有文献可知, 近防密集阵同样存在攻击范围限制, 与防空高炮不同之处在于: 目标与之距离越近, 威胁值越大, 超出射程范围时, 则无法构成威胁。本文定义近防密集阵的威胁值函数为

$$P_{\mathrm{p}}=\begin{cases}1 & R\in\left[0,R_{\mathrm{pmin}}\right]\\ \dfrac{1}{2}\left[\cos\left(\pi\dfrac{R-R_{\mathrm{pmin}}}{R_{\mathrm{pmax}}-R_{\mathrm{pmin}}}\right)+1\right] & R\in\left(R_{\mathrm{pmin}},R_{\mathrm{pmax}}\right]\\ 0 & R\in\left(R_{\mathrm{pmax}},+\infty\right)\end{cases}\tag{6}$$

假设密集阵位于防区中心或舰上, R_{pmin}为密集阵的攻击近界, R_{pmax}为攻击远界。

1.3　全弹道规划问题模型

滑翔弹的飞行弹道可以分为 4 个阶段: 起飞段、助推段、爬升段和滑翔段。本文为寻求具体任务中动、静态参数的最优匹配组合, 基于式 (1) 三自由度弹道模型, 进行全弹道规划设计。

1.3.1　优化目标

通过无人机侦查等方式, 可以提前获取目标的防御手段。本文假设目标的防御情况已经事先探明, 且防空高炮和近防密集阵在雷达探测到的基础上进行攻击。为提高突防的成功率, 需要滑翔弹暴露在威胁之内的时间尽可能小且全程威胁值尽可能低。本文采用加权的形式, 选取如式 (7) 所示的综合威胁值作为优化目标函数:

$$\min J=\int_0^{t_{\mathrm{f}}}\left[\alpha+P_{\mathrm{r}}\left(\beta+\gamma P_{\mathrm{g}}+\eta P_{\mathrm{p}}\right)\right]\mathrm{d}t\tag{7}$$

式中: t_{f}为未知的弹道末段时刻; β、γ和η分别为雷达、高炮和密集阵威胁值在目标防御总威胁中所占的权重, $\beta+\gamma+\eta=1$, 具体取值参考工程经验; α为飞行时间项在综合威胁值中的权重。

本文目标函数考虑了飞行时间尽量短与全程威胁值尽量低 2 个部分, 为使这 2 个部分的重要性大致相同, 其数量级应大致相当, 故而参数α的取值并非随意, 需要在仿真过程中进行调整。

1.3.2　约束条件

为满足相关作战任务需求, 设置如下约束条件:

1) 边界条件。出炮口位置(x_0,y_0,z_0), 初速度V_0, 初始弹道倾角范围$[\theta_{0\min},\theta_{0\max}]$, 初始弹道偏角范围$[\psi_{v0\min},\psi_{v0\max}]$, 目标位置$(x_{\mathrm{T}},y_{\mathrm{T}},z_{\mathrm{T}})$, 最大飞行时间$t_{\mathrm{f}}\leqslant T_{\max}$, 末端存速$V(t_{\mathrm{f}})\geqslant V_{\mathrm{f}\min}$。

2) 过程约束。飞行阶段不同, 滑翔弹的过程约束也不同。除助推段之外, 火箭在其他阶段均不工作; 而除滑翔段外, 其他阶段均为无控飞行。因此, 滑翔弹各阶段的过程约束可以表示如下。

起飞段和爬升段:

$$\begin{cases}P=0,m_{\mathrm{c}}=0\\ \alpha^*=0,\beta^*=0\end{cases}\tag{8}$$

助推段:

$$\begin{cases}P=T_{\mathrm{p}},m_{\mathrm{c}}=m_{\mathrm{p}}/t_{\mathrm{b}}\\ \alpha^*=0,\beta^*=0\end{cases}\tag{9}$$

滑翔段:

$$\begin{cases}P=0,m_{\mathrm{c}}=0\\ |\alpha^*|\leqslant\alpha^*_{\max}\\ |\beta^*|\leqslant\beta^*_{\max}\end{cases}\tag{10}$$

式中: T_{p}为火箭的平均推力值; m_{p}为火箭装药质量; t_{b}为火箭工作时长。考虑到滑翔弹有限的控制能力, $\alpha^*_{\max}$和$\beta^*_{\max}$分别为所允许的最大准攻角和准侧滑角幅值。

3) 连接约束。为保证 4 个阶段之间弹道的连续性, 考虑如下连接约束:

$$\begin{cases}\boldsymbol{x}(t_{\mathrm{i}}^-)=\boldsymbol{x}(t_{\mathrm{i}}^+)\\ \boldsymbol{x}(t_{\mathrm{c}}^-)=\boldsymbol{x}(t_{\mathrm{c}}^+)\\ \boldsymbol{x}(t_{\mathrm{o}}^-)=\boldsymbol{x}(t_{\mathrm{o}}^+)\end{cases}\tag{11}$$

式中: $\boldsymbol{x}$为式 (1) 所示的滑翔弹状态变量; 符号 "–" 和 "+" 分别表示上一阶段的终端和下一阶段的起始; t_{i}、t_{c}和t_{o}分别表示火箭点火时刻、火箭停机时刻和滑翔启控时刻。由于火箭工作时长固定, t_{i}与t_{c}并不独立, $t_{\mathrm{c}}=t_{\mathrm{i}}+t_{\mathrm{b}}$, 在设计参数时仅需设计火箭点火时刻$t_{\mathrm{i}}$即可。

1.3.3　全弹道规划流程

面向突防的滑翔弹全弹道规划问题可以具体

描述为：在时间$[t_0,t_f]$内，确定出初始弹道倾角θ_0、初始弹道偏角ψ_{v0}、火箭点火时刻t_i、滑翔启控时刻t_o等静态参数及滑翔段的时变控制量α^*和β^*，使得由式(1)确定的动力学系统，在满足边界条件及约束条件式(8)～式(11)的基础上，式(7)所示目标函数最小。具体流程如图1所示。

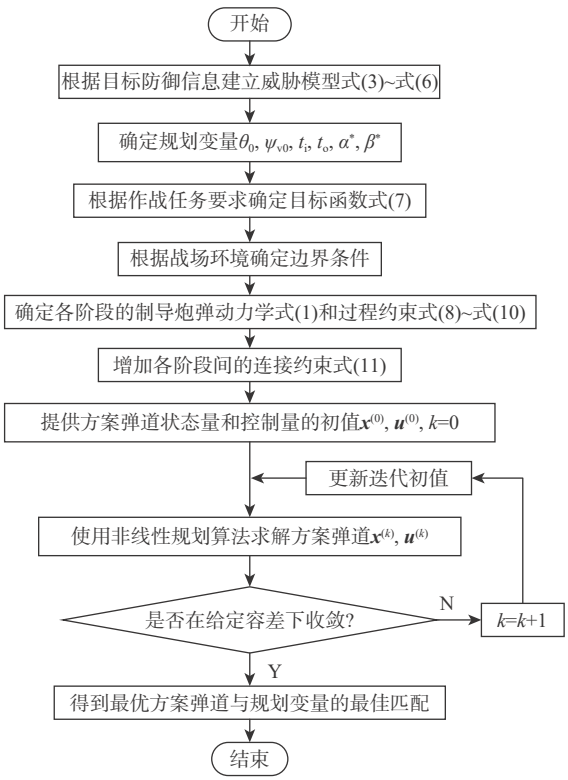


图1 全弹道规划流程

Fig. 1 Flow of full trajectory planning

2 最优控制问题的转化与hp自适应伪谱法

第1节提出的全弹道规划问题，从数学本质上看可抽象为求解包含微分-代数约束和不等式约束的一类非线性最优控制问题，本文采用基于hp自适应的多区间Radau伪谱法求解。

2.1 多区间伪谱离散

多区间非线性最优控制问题是指将时间区间$[t_0,t_f]$划分为K个网格子区间$[t_{k-1},t_k],k=1,2,\cdots,K$。通过如式(12)所示的时域转换关系，将每个子网格的时域转换到$\tau\in[-1,1]$：

$$\tau=\frac{2t}{t_k-t_{k-1}}-\frac{t_k+t_{k-1}}{t_k-t_{k-1}} \tag{12}$$

于是有

$$\frac{\mathrm{d}t}{\mathrm{d}\tau}=\frac{t_k-t_{k-1}}{2} \tag{13}$$

为保证各网格子区间的连续性，设$\boldsymbol{x}^{(k)}(\tau)$和$\boldsymbol{u}^{(k)}(\tau)$分别为第$k$个网格子区间的状态变量和控制变量，增加如式(14)所示的内点约束：

$$\boldsymbol{x}^{(k)}(1)=\boldsymbol{x}^{(k+1)}(-1)\qquad k=1,2,\cdots,K-1 \tag{14}$$

通过上述步骤变换，可将式(7)确定的目标函数转化为

$$\min J=\sum_{k=1}^{K}\frac{t_k-t_{k-1}}{2}\int_{-1}^{1}\left[\alpha+P_{\mathrm{r}}\left(\beta+\gamma P_{\mathrm{g}}+\eta P_{\mathrm{p}}\right)\right]\mathrm{d}\tau \tag{15}$$

对于第k个子区间，取N_k个勒让德-高斯-拉道(Legendre-Gauss-Radau, LGR)配置点$\tau_i^{(k)}(i=1,2,\cdots,N_k)$和$\tau_{N_k+1}^{(k)}=1$作为节点，构成$N_k$阶Lagrange插值多项式$L_i^{(k)}(\tau)$：

$$L_i^{(k)}(\tau)=\prod_{\substack{j=1\\ j\neq i}}^{N_k+1}\frac{\tau-\tau_j}{\tau_i-\tau_j}\qquad i=1,2,\cdots,N_k+1 \tag{16}$$

以$L_i^{(k)}(\tau)$为基函数，则状态量可近似为

$$\boldsymbol{x}^{(k)}(\tau)\approx\sum_{j=0}^{N_k}\boldsymbol{x}_j^{(k)}L_j^{(k)}(\tau) \tag{17}$$

控制量可近似为

$$\boldsymbol{u}^{(k)}(\tau)\approx\sum_{i=1}^{N_k}\boldsymbol{u}_i^{(k)}L_i^{(k)}(\tau) \tag{18}$$

对近似状态方程求微分，得

$$\frac{\mathrm{d}\boldsymbol{x}^{(k)}(\tau)}{\mathrm{d}\tau}=\sum_{j=0}^{N_k}\boldsymbol{x}_j^{(k)}\dot{L}_j^{(k)}(\tau) \tag{19}$$

则N_k个LGR点上的动力学微分方程约束变为代数方程约束，即

$$\sum_{j=0}^{N_k}\boldsymbol{x}_j^{(k)}D_{ij}^{(k)}-\left(\frac{t_k-t_{k-1}}{2}\right)\boldsymbol{f}_i^{(k)}=\boldsymbol{0}$$
$$i=1,2,\cdots,N_k;k=1,2,\cdots,K \tag{20}$$

式中：$D_{ij}^{(k)}=\dot{L}_j^{(k)}\left(\tau_i^{(k)}\right)$为第$k$个子区间的Radau伪谱微分矩阵；$\boldsymbol{f}$为式(1)所示的动态方程约束。

式(15)所示目标函数可由多区间LGR积分近似为

$$\min J=\frac{t_k-t_{k-1}}{2}\sum_{k=1}^{K}\sum_{j=1}^{N_k}w_j^{(k)}g_j^{(k)} \tag{21}$$

式中：$g_j^{(k)}$为第k个子区间内第j个配置点的待积目标函数值，同式(7)所示；$w_j^{(k)}$为相应配置点的积分权值，具体形式为[32]

$$\begin{cases}w_0^{(k)}=\dfrac{2}{(N_k+1)^2}\\ w_j^{(k)}=\dfrac{2}{(N_k+1)^2}\cdot\dfrac{1-\tau_j^{(k)}}{\left(L_{N_k}\left(\tau_j^{(k)}\right)\right)^2}\end{cases} \tag{22}$$

路径约束与边界条件可采用相同的方法近似, 至此将多区间连续时间的最优控制问题离散化, 并转变为相应的非线性规划(nonlinear programming, NLP)问题。

2.2　hp 自适应伪谱法

对于上述转化后的 NLP 问题, 使用 NLP 算法进行求解。若结果不满足精度要求, 根据 hp 自适应伪谱法, 可通过增加多项式维数或增加子区间个数的方式对问题进行更新。本文选取状态轨迹在采样点处的曲率比作为评判标准进行网格细化。

当第 k 个子区间的误差不满足精度要求时, 若区间的曲率比不符合要求, 则通过 h 法增加子区间的数量; 若曲率比符合要求, 则通过 p 法增加区间内插值多项式的维数。

1) h 法。增加子区间数量, 定义新的节点个数为 n_k, 表达式为

$$n_k = 2\mathrm{ceil}\left[\lg\left(\frac{\varepsilon^{(k)}}{\varepsilon_\mathrm{d}}\right)\right] - 1 \tag{23}$$

新节点的位置分布由曲率密度函数决定[28]。

2) p 法。增加插值多项式维数:

$$N_+^{(k)} = N_-^{(k)} + \mathrm{ceil}\left[\lg\left(\frac{\varepsilon^{(k)}}{\varepsilon_\mathrm{d}}\right)\right] + 1 \tag{24}$$

式中: $N_-^{(k)}$为更新前的多项式维数; $N_+^{(k)}$为更新后的多项式维数; ceil函数的返回值为不小于指定表达式的最小整数; $\varepsilon^{(k)}$为第 k 个区间内近似状态量的误差; ε_d为误差限值。

3　仿真和分析

本节针对某型滑翔弹, 在个人 PC 端 MATLAB 2022b 平台调用 GPOPS-Ⅱ 工具箱, 采用 hp 自适应伪谱法对第 1 节所提弹道规划问题进行求解。为验证所提性能指标的有效性, 对一系列不同防御场景进行仿真, 并与传统性能指标作用下的最优弹道进行对比分析以验证其优越性。

3.1　不同防御场景对比

基于第 1 节的假设, 高炮与密集阵均固定在目标位置处, 本节通过更改雷达位置坐标进行对比分析。模拟不同战场环境, 分别分析了单雷达位置在滑翔弹无控段、滑翔段及双雷达 3 种场景下的威胁度变化、不同情况对于待优化参数的影响, 并在每个场景下设置不同的算例, 来讨论雷达位置偏离基准弹道距离远近对威胁度变化及弹道特性的影响。仿真使用参数如表 1～表 3 所示。

3.1.1　场景 1: 单雷达无控段

以炮口位置为坐标原点, 固定雷达 x 方向坐标为 x_ra = 10 km, 改变 z 方向的坐标, 模拟 3 种战场环境, 具体坐标位置如表 4 所示。

表 1　滑翔弹参数

Table 1　Gliding-guided projectile parameter

参数	数值
m/kg	44.5
m_p/kg	7.23
S/m^2	0.013 3
T_p/N	1 219.2
t_b/s	14.068
k_c	35

表 2　目标防御参数

Table 2　Target defense parameter

参数	数值
R_rmax/km	50
R_gmin/km	2
R_gmax/km	7
R_pmin/km	1
R_pmax/km	3

表 3　仿真参数

Table 3　Simulation parameter

参数	数值	参数	数值
x_0/km	0	ψ_v0max/(°)	90
y_0/km	0	T_max/s	300
z_0/km	0	V_fmin/(m·s^{-1})	250
x_T/km	60	α^*_max/(°)	15
y_T/km	0	β^*_max/(°)	10
z_T/km	1	α	0.1
V_0/(m·s^{-1})	800	β	0.5
θ_0min/(°)	0	γ	0.2
θ_0max/(°)	90	η	0.3
ψ_v0min/(°)	−90		

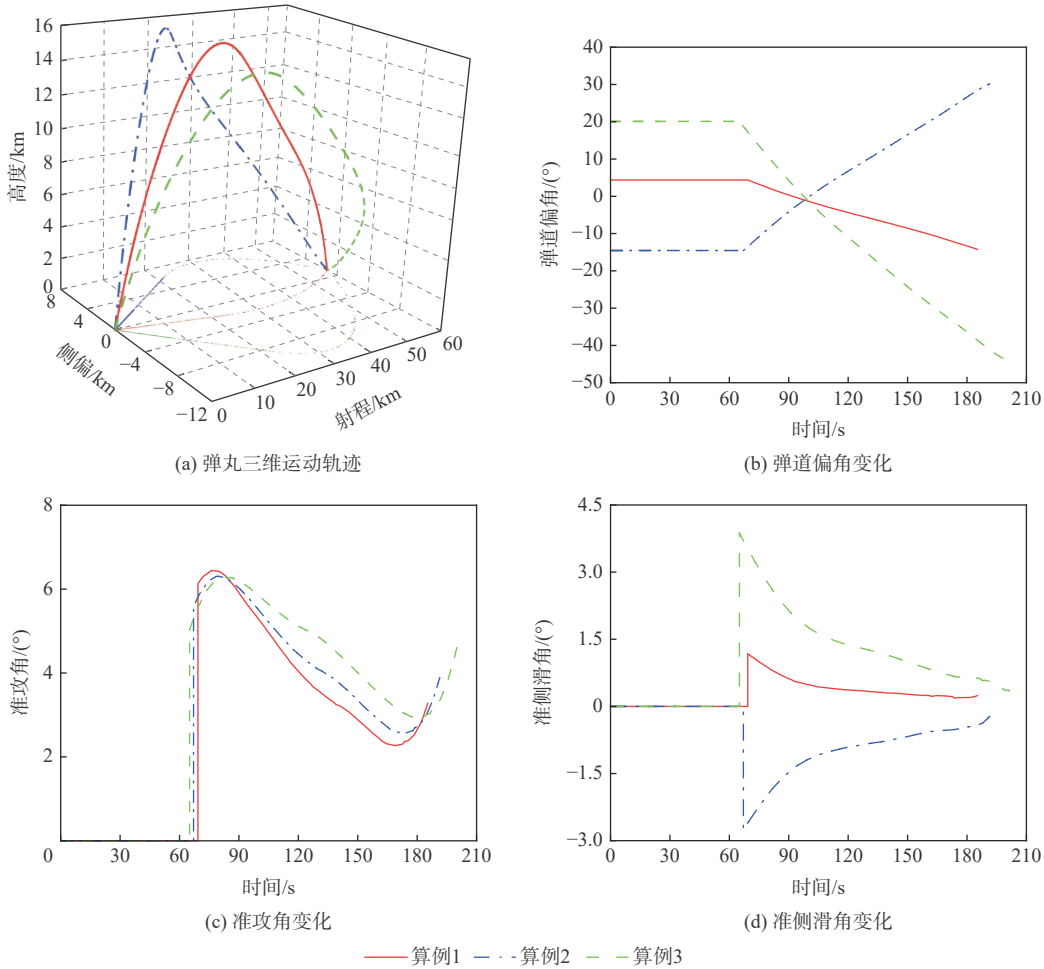
表 4　场景 1 雷达位置

Table 4　Radar position of Scene 1

算例	雷达位置坐标/km
1	(10,0,2)
2	(10,0,−5)
3	(10,0,10)

仿真结果如图 2 所示(为方便观察, 三维轨迹增加了侧向平面内的投影)。场景 1 全弹道规划结果如表 5 所示。

由图 2(a)、图 2(b)、图 2(d) 可知, 随着雷达位置变化, 弹丸轨迹及侧向平面内弹道诸元变化明显, 对目标雷达起到了不同程度的规避作用。由于雷达位于地面, 纵向平面内弹道诸元对雷达位置变化并不敏感。



(a) 弹丸三维运动轨迹　(b) 弹道偏角变化

(c) 准攻角变化　(d) 准侧滑角变化

图2　场景1仿真结果

Fig. 2　Simulation results of Scene 1

表5　场景1全弹道规划结果

Table 5　Full trajectory programming results of Scene 1

算例	θ_0/(°)	ψ_{v0}/(°)	t_i/s	t_o/s	t_f/s	威胁值
1	56.72	4.38	9.55	69.31	185.29	19.61
2	56.26	−14.53	9.80	67.10	191.34	16.13
3	56.02	20.09	10.51	65.09	201.38	12.25

由全弹道仿真结果可知，雷达位置变化对初始弹道偏角和弹丸飞行时间影响较大，对初始弹道倾角、火箭点火时刻、滑翔启控时刻影响较小。

3.1.2　场景2：单雷达滑翔段

以炮口位置为坐标原点，固定雷达x方向坐标为$x_{ra}=40$ km，具体坐标位置如表6所示。仿真结果如图3所示。场景2全弹道规划结果如表7所示。

表6　场景2雷达位置

Table 6　Radar position of Scene 2

算例	雷达位置坐标/km
1	(40,0,−1)
2	(40,0,2/3)
3	(40,0,−15)

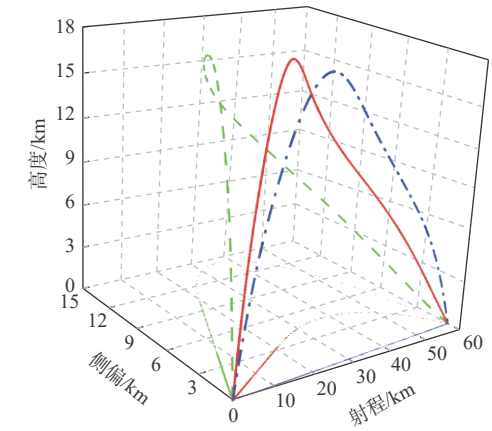
由仿真结果可知，当雷达位置处在基准弹道曲线上(算例2)时，侧向平面内弹道对雷达并未采取规避动作。从全弹道仿真结果分析，雷达距离基准弹道曲线越远(在最大威胁范围内)，对应弹道的威胁值越小，初始弹道偏角越大，飞行时间也越长。在特殊情况下(如算例3)，可以极大地度减小威胁值。

3.1.3　场景3：双雷达

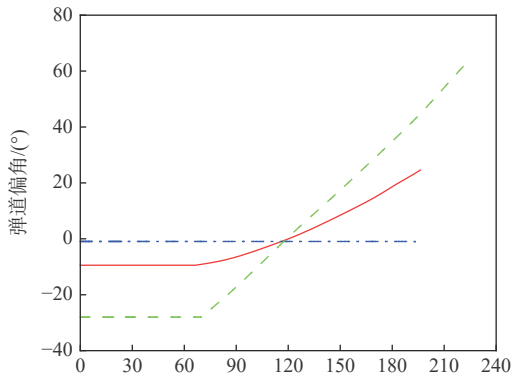
考虑目标具有多重防御或多目标协同防御的情况，以双雷达的场景进行初步探索。考虑到双雷达均位于基准弹道曲线同侧等效为单雷达的情况，算例中将雷达置于异侧。具体坐标位置如表8所示。仿真结果如图4所示。场景3全弹道规划结果如表9所示。

由图4(a)可知，当目标拥有多个预警雷达时，由于其互相的耦合作用，对比单雷达的情况，弹道轨迹更加多样。相较于前2个场景，弹道偏角的变化情况也更丰富，出现了不单调的情况(算例3和算例4)。

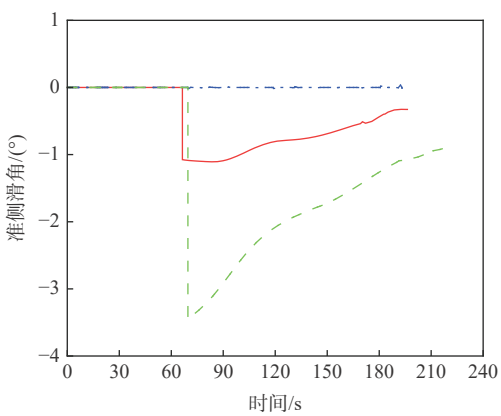
从全弹道仿真结果分析，若多雷达位置分布在



(a) 弹丸三维运动轨迹



(b) 弹道偏角变化



(c) 准侧滑角变化

算例1　算例2　算例3

图 3　场景 2 仿真结果

Fig. 3　Simulation results of Scene 2

表 7　场景 2 全弹道规划结果

Table 7　Full trajectory programming results of Scene 2

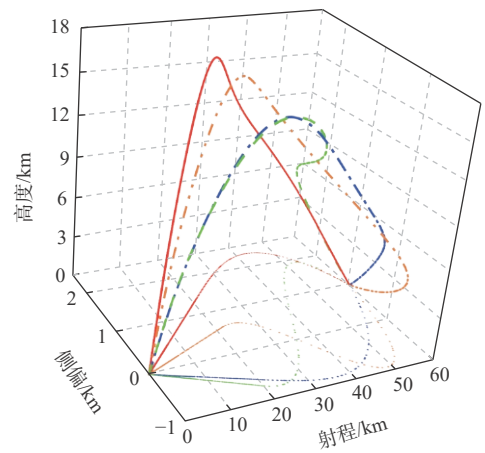
算例	θ_0/(°)	ψ_{v0}/(°)	t_i/s	t_o/s	t_f/s	威胁值
1	64.12	−9.44	16.34	66.42	196.29	29.40
2	64.13	−0.96	16.08	69.54	193.20	30.25
3	65.11	−27.99	22.71	69.60	221.87	8.85

基准弹道曲线两侧, 其对弹道的影响呈相互补偿作用, 相比于单雷达的情况(见表 5 和表 7), 初始弹道偏角的变化范围更小, 整体飞行时间也相对减少。

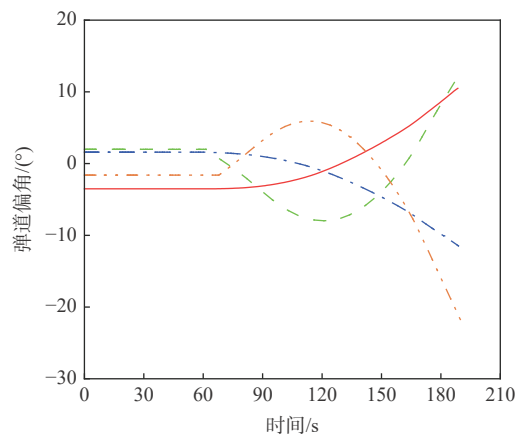
表 8　场景 3 雷达位置

Table 8　Radar position of Scene 3

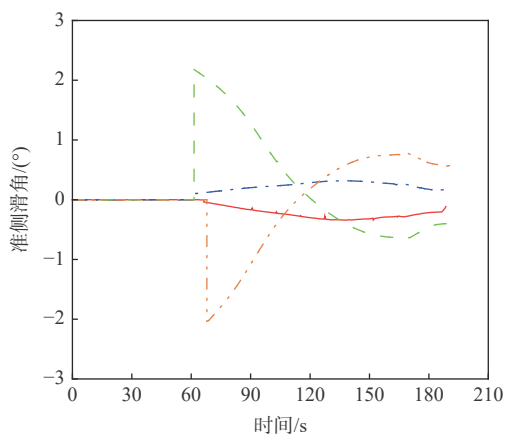
算例	雷达位置坐标/km
1	(10,0,2), (40,0,−1)
2	(10,0,−1), (40,0,2)
3	(15,0,20), (50,0,−5)
4	(10,0,−20), (50,0,10)



(a) 弹丸三维运动轨迹



(b) 弹道偏角变化



(c) 准侧滑角变化

算例1　算例2　算例3　算例4

图 4　场景 3 仿真结果

Fig. 4　Simulation results of Scene 3

表 9 场景 3 全弹道规划结果

Table 9 Full trajectory programming results of Scene 3

算例	θ_0/(°)	ψ_{v0}/(°)	t_i/s	t_o/s	t_f/s	威胁值
1	60.18	−3.51	10.80	66.33	188.56	45.93
2	60.21	1.59	10.87	61.53	188.83	45.87
3	61.99	1.99	13.65	61.42	191.05	35.39
4	61.69	−1.61	13.48	67.89	190.33	32.25

3.2 与传统弹道规划方法对比

根据作战任务需求的不同，评价滑翔弹方案弹道好坏的指标也不同，比较常见的有最大射程、最大升阻比、最小控制能量消耗、最短全程飞行时间等。本节设计不同算例，在相同作战环境下，将本文方法同 2 种典型的传统弹道规划方法进行仿真对比，分析各自优劣势及程度。仿真参数设置如表 10 所示。

表 10 雷达位置

Table 10 Radar positions

算例	雷达位置坐标/km
1	(10,0,2)
2	(40,0,−1)
3	(60,0,1)
4	(15,0,20), (50,0,−5)

1）方法 1。综合威胁最低（本文方法，同式 (7)）。

2）方法 2。控制能量最优[27]：

$$\min J=\int_0^{t_f}\left[\left(\alpha^*/\alpha_{\max}^*\right)^2+\left(\beta^*/\beta_{\max}^*\right)^2\right]\mathrm{d}t \tag{25}$$

3）方法 3。飞行时间最短[4]：

$$\min T_f=\int_0^{t_f}1\mathrm{d}t \tag{26}$$

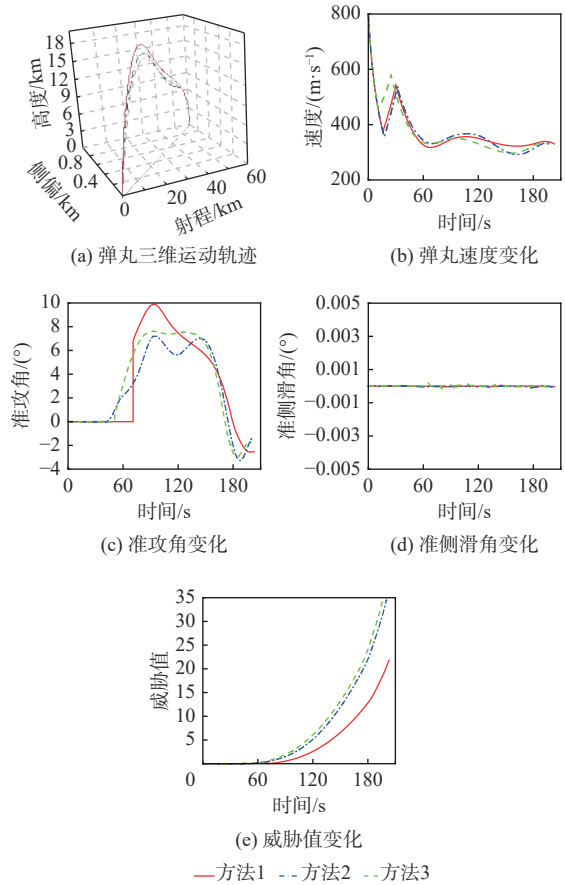
由于传统的弹道规划方法（方法 2 和方法 3）并不考虑目标防御的位置，在不同算例下仿真曲线相同，全弹道规划结果也相同，如表 11 所示，只有威胁值的区别。故本节只展示雷达在目标处 (算例 3) 的图线，如图 5 所示，其他算例的仿真结果如表 12 所示。

表 11 传统方法全弹道规划结果

Table 11 Full trajectory programming results of traditional methods

方法	θ_0/(°)	ψ_{v0}/(°)	t_i/s	t_o/s	t_f/s	能量消耗/(rad²·s)
2	63.07	−0.86	17.48	43.42	200.32	17.02
3	58.57	−0.95	10.77	50.83	196.32	22.24

由图 5 结果可知，相比于传统方法，本文方法理论弹道的弹道高更大，速度变化更平缓，威胁值更低。但由于考虑了对目标防御的规避，飞行时间更长，控制能量消耗更多。



(a) 弹丸三维运动轨迹 (b) 弹丸速度变化

(c) 准攻角变化 (d) 准侧滑角变化

(e) 威胁值变化

图 5 与传统方法对比结果（算例 3）

Fig. 5 Results compared with traditional methods (Example 3)

表 12 不同算例中与传统方法对比结果

Table 12 Results compared with traditional methods in different examples

算例	威胁值			方法1的能量消耗/(rad²·s)	方法1的飞行时间/s
	方法1	方法2	方法3		
1	19.51	29.65	28.69	22.51	197.76
2	28.35	41.95	42.25	24.31	202.86
3	21.79	34.46	34.51	24.93	203.07
4	34.58	53.33	53.05	24.29	200.78

更直观的结果可由表 11 与表 12 的对比分析得到。在本节算例下，同方法 3（飞行时间最短）相比，本文方法的飞行时间相差无几，最大不超过 4%，控制能量的消耗方面，差距也不明显，最大不超过 11%。但在威胁值方面，优势明显，平均可以达到 34%；同方法 2（控制能量最优）相比，飞行时间的劣势更小，最大不超过 2%，控制能量具有劣势（均值 29%），是因为对目标防御手段的规避造成的，但威胁值的优势（均值 35%）更大。

4 结 论

本文针对滑翔制导炮弹最优突防方案弹道规

划问题, 设计了考虑综合威胁值的性能指标, 并通过全弹道仿真进行了验证, 结论如下:

1) 在单雷达防御情况下, 理论弹道对目标防御手段均有良好的规避效果, 雷达距离基准弹道曲线越远(在最大威胁范围内), 对应弹道的威胁值越小, 飞行时间也越长; 对于全弹道而言, 雷达位置所处的弹道阶段(无控段或滑翔段)对初始弹道倾角和火箭点火时刻影响较为明显。

2) 当目标拥有多个预警雷达时, 其对弹道的影响呈相互耦合作用, 相比于单雷达的情况, 初始弹道偏角的变化范围更小, 弹道偏角的变化情况更复杂, 整体飞行时间也相对减少。

3) 同传统弹道规划方法相比, 本文方法的飞行时间相差不超过 4%, 控制能量的消耗方面相差不超过 11%; 在降低威胁值方面, 优势平均可以达到 34%。

综上可知, 本文方法在降低威胁方面具有突出的优势。同传统弹道规划方法相比, 在缩减最大飞行时间方面表现良好, 在考虑控制能量消耗方面存在劣势(规避威胁不可避免的消耗), 可根据实际作战任务要求进行性能指标的选择。

参考文献（References）

[1] GUO F S, LIU X X, LIU D. Review on the trajectory design of the dynamic-glider missile[J]. Procedia Engineering, 2011, 15: 1191-1195.

[2] 陈琦. 滑翔增程制导炮弹弹道优化及制导控制方法研究[D]. 南京: 南京理工大学, 2017.
CHEN Q. Investigation on trajectory optimization and guidance & control scheme for a type of gliding guided projectiles[D]. Nanjing: Nanjing University of Science and Technology, 2017(in Chinese).

[3] CHEN Q, WANG Z Y. Optimal trajectory for time-on-target of a guided projectile using direct collocation method[C]//Proceedings of the International Conference on Mechatronic Sciences, Electric Engineering and Computer. Piscataway: IEEE Press, 2013: 2803-2806.

[4] 徐秋坪. 滑翔制导炮弹弹道规划及其自抗扰控制系统研究[D]. 南京: 南京理工大学, 2018.
XU Q P. Trajectory optimization and active-disturbance-rejection control system for gliding guided projectiles[D]. Nanjing: Nanjing University of Science and Technology, 2018(in Chinese).

[5] XIAO J K, WANG G, GUO X K, et al. Modeling on threat assessment system of reentry-course ballistic missile[J]. Advanced Materials Research, 2013, 722: 301-305.

[6] KUMAR S, TRIPATHI B K. Modelling of threat evaluation for dynamic targets using Bayesian network approach[J]. Procedia Technology, 2016, 24: 1268-1275.

[7] 王勃, 姚佩阳, 税冬东, 等. 基于影响网的战场威胁强度评估建模[J]. 计算机工程与设计, 2014, 35(8): 2836-2840.
WANG B, YAO P Y, SHUI D D, et al. Modeling of battlefield threat intensity evaluation based on influence nets[J]. Computer Engineering and Design, 2014, 35(8): 2836-2840(in Chinese).

[8] LEE H, CHOI B J, KIM C O, et al. Threat evaluation of enemy air fighters via neural network-based Markov chain modeling[J]. Knowledge-Based Systems, 2017, 116: 49-57.

[9] LIU J, LI X B, LEI Y L, et al. An index based threat modeling method for path planning[C]//Proceedings of the 2009 First International Conference on Advances in System Simulation. Piscataway: IEEE Press, 2009: 1-5.

[10] 刘娟, 赵新, 雷永林, 等. 基于指数法的导弹航迹规划空间威胁建模研究[J]. 系统仿真学报, 2009, 21(17): 5353-5356.
LIU J, ZHAO X, LEI Y L, et al. Modeling of threat space for missile route planning based on index-method[J]. Journal of System Simulation, 2009, 21(17): 5353-5356(in Chinese).

[11] 刘娟, 王维平. 基于仿真数据的航迹规划威胁建模方法研究[J]. 计算机仿真, 2010, 27(9): 24-28.
LIU J, WANG W P. A simulation data based space threat modeling method for path planning[J]. Computer Simulation, 2010, 27(9): 24-28(in Chinese).

[12] 黄俊, 于雷, 李战武, 等. 对地任务规划系统威胁空间综合量化建模与仿真[J]. 系统仿真学报, 2010, 22(4): 1033-1036.
HUANG J, YU L, LI Z W, et al. Integrated quantization modeling and simulation of threat space for air-to-surface mission planning system[J]. Journal of System Simulation, 2010, 22(4): 1033-1036 (in Chinese).

[13] 申舟, 谢文俊, 赵晓林, 等. 基于人工势场的无人机战场威胁建模研究[J]. 计算机仿真, 2014, 31(2): 60-64.
SHEN Z, XIE W J, ZHAO X L, et al. Modeling of UAV battlefield threats based on artificial potential field[J]. Computer Simulation, 2014, 31(2): 60-64(in Chinese).

[14] 郭志鸿. 多无人机协同的实时任务分配与航迹规划研究[D]. 重庆: 重庆理工大学, 2022.
GUO Z H. Research on real-time task assignment and tracks planning for multi-UAV collaboration[D]. Chongqing: Chongqing University of Technology, 2022(in Chinese).

[15] ZHANG R, LI S, DING Y M, et al. UAV path planning algorithm based on improved Harris Hawks optimization[J]. Sensors, 2022, 22(14): 5232.

[16] GUO Y C, LIU X X, ZHANG W G, et al. An improved algorithm of UAV 3D path planning[C]//Proceedings of the International Conterence on Guidance, Vavigation and Control. Berlin: Springer, 2021: 849-859.

[17] 李晓萌, 王道波, 郭继凯, 等. 基于某种生物启发式算法的无人机航迹规划[J]. 机械与电子, 2018, 36(11): 15-19.
LI X M, WANG D B, GUO J K, et al. Path planning of UAV based on a biological heuristic algorithm[J]. Machinery & Electronics, 2018, 36(11): 15-19(in Chinese).

[18] 彭建亮, 孙秀霞, 蔡满意, 等. 基于人工势场的防空威胁建模与仿真[J]. 系统工程与电子技术, 2010, 32(2): 338-341.
PENG J L, SUN X X, CAI M Y, et al. Modeling and simulation of air defense threat based on artificial potential field[J]. Systems Engineering and Electronics, 2010, 32(2): 338-341(in Chinese).

[19] ALPDEMIR M N. Tactical UAV path optimization under radar threat using deep reinforcement learning[J]. Neural Computing and Applications, 2022, 34(7): 5649-5664.

[20] 温乃峰. 低空复杂环境下小型无人机的在线航迹规划算法研究[D]. 哈尔滨: 哈尔滨工业大学, 2016.
WEN N F. Research on the small UAV online path planning algorithms in complex and low-altitude environments[D]. Harbin: Harbin Institute of Technology, 2016(in Chinese).

[21] 张志伟, 蒋道刚, 袁坤刚. 低空突防航线规划研究[J]. 飞行力学, 2019, 37(4): 62-67.
ZHANG Z W, JIANG D G, YUAN K G. Study on low altitude penetration route planning[J]. Flight Dynamics, 2019, 37(4): 62-67(in Chinese).

[22] 周鑫, 李新洪, 王谦. 基于威胁建模的 PSO 在 UAV 3 维航路规划中的应用[J]. 兵工自动化, 2017, 36(4): 73-76.
ZHOU X, LI X H, WANG Q. Application of PSO based on threat modeling in 3D route planning of UAV[J]. Ordnance Industry Automation, 2017, 36(4): 73-76(in Chinese).

[23] LIU J S, JIANG W Z, LIU T. Penetration efficiency evaluation of missile with jammer[J]. Applied Mechanics and Materials, 2015, 719-720: 369-375.

[24] 唐嘉钰, 李相民, 代进进. 基于威胁建模的 UCAV 武器投放在线航迹规划[J]. 西北工业大学学报, 2021, 39(2): 382-391.
TANG J Y, LI X M, DAI J J. Online trajectory planning for UCAV performing ground attacks based on threat modeling[J]. Journal of Northwestern Polytechnical University, 2021, 39(2): 382-391(in Chinese).

[25] MA M X, WU J, SHI Y, et al. Chaotic random opposition-based learning and cauchy mutation improved moth-flame optimization algorithm for intelligent route planning of multiple UAVs[J]. IEEE Access, 2022, 10: 49385-49397.

[26] GUO J G, HU G J, GUO Z Y, et al. Evaluation model, intelligent assignment, and cooperative interception in multimissile and multitarget engagement[J]. IEEE Transactions on Aerospace and Electronic Systems, 2022, 58(4): 3104-3115.

[27] CHEN Q, WANG Z Y, CHANG S J, et al. Multiphase trajectory optimization for gun-launched glide guided projectiles[J]. Proceedings of the Institution of Mechanical Engineers, Part G: Journal of Aerospace Engineering, 2016, 230(6): 995-1010.

[28] 邱文杰, 孟秀云. 基于 hp 自适应伪谱法的飞行器多阶段轨迹优化[J]. 北京理工大学学报, 2017, 37(4): 412-417.
QIU W J, MENG X Y. Multi-phase trajectory optimization of vehicle based on hp-adaptive pseudospectral method[J]. Transactions of Beijing Institute of Technology, 2017, 37(4): 412-417(in Chinese).

[29] 韩子鹏. 弹箭外弹道学[M]. 北京: 北京理工大学出版社, 2008.
HAN Z P. External ballistics of projectile and arrow[M]. Beijing: Beijing Insititute of Technology Press, 2008 (in Chinese).

[30] 任天祥, 贺建良, 邹杰. 基于改进 Theta*算法的直升机三维航路规划[J]. 电光与控制, 2022, 29(2): 45-48.
REN T X, HE J L, ZOU J. 3D route planning for helicopters based on improved Theta* algorithm[J]. Electronics Optics & Control, 2022, 29(2): 45-48(in Chinese).

[31] 卞强, 孙齐, 童余德. 一种新的改进 A*算法无人机三维路径规划[J]. 武汉理工大学学报, 2022, 44(7): 80-88.
BIAN Q, SUN Q, TONG Y D. A new improved A* algorithm for UAV 3D path planning[J]. Journal of Wuhan University of Technology, 2022, 44(7): 80-88(in Chinese).

[32] 王丽英, 张友安, 黄诘. 带约束的末制导律与伪谱法轨迹优化[M]. 北京: 国防工业出版社, 2015.
WANG L Y, ZHANG Y A, HUANG J. Constrained terminal guidance law and pseudospectral trajectory optimization[M]. Beijing: National Defense Industry Press, 2015(in Chinese).

Trajectory programming method of gliding-guided projectiles for penetration

YIN Qiulin，CHEN Qi*，WANG Zhongyuan，WANG Qinghai

(School of Energy and Power Engineering，Nanjing University of Science and Technology，Nanjing 210094，China)

Abstract: Aiming at the problem of gliding-guided projectile penetration schemes in inevitable threat areas, mathematical models of the enemy's defense are established from the perspective of quantifying the threat value. Based on the model, a cost function with the lowest comprehensive threat value in the whole process is designed. To accomplish the best possible matching of the initial trajectory inclination, deflection angle, rocket ignition time, gliding start time, and other parameters over the course of the flight, a multi-stage full trajectory planning model is created. Then, the hp adaptive pseudo-spectral method is used to transform the optimal control problem into a nonlinear programming problem. The evading effect of the projectile on target defense is verified by simulation, and the factors affecting the effectiveness are analyzed. The proposed method's advantages over conventional trajectory programming techniques are also examined.

Keywords: gliding-guided projectile; penetration; threat modeling; full trajectory programming; hp adaptive pseudo-spectral method

Received：2023-02-14；**Accepted**：2023-06-02；**Published Online**：2023-06-26 14:17

URL：link.cnki.net/urlid/11.2625.V.20230626.1007.001

Foundation items：National Natural Science Foundation of China (52202475); Natural Science Foundation of Jiangsu Province (BK20200498)

* **Corresponding author**. E-mail：qichen@njust.edu.cn

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2022.0766

基于SINS/OD/高度计的车载重力异常测量方法

李新宇[1,2]，张志利[1,2]，周召发[1,2,*]，常振军[1,2]，赵芝谦[1,2]

(1. 火箭军工程大学 导弹工程学院，西安 710025；2. 火箭军工程大学 兵器发射理论和技术国家重点学科实验室，西安 710025)

摘 要：车载重力测量沿地球表面实施，其较慢的速度和机动灵活的特点为高精度地面重力测量提供有利条件。传统动基座重力测量严重依赖全球定位系统(GPS)而无法实现完全自主性，基于捷联惯导(SINS)/里程计(OD)/高度计组合导航提出一种新的自主无源车载重力异常测量方法。组合导航系统提供了高精度的比力、姿态和载体加速度信息，气压高度计用于抑制高度通道的发散，通过直接差分法即可实现重力信息的获取；对比了以航位推算位置(SINS/DR)、里程计速度(SINS/OD)、位置速度(SINS/DR/OD)为量测和传统SINS/GPS这4种方法的测量精度，分析量测选择对所提方法测量效果的影响。重复测线仿真及单测线跑车试验结果表明：所提方法可以实现完全自主的车载重力异常测量，相比传统SINS/GPS方法具有普遍的精度优势，SINS/OD与SINS/DR/OD测量精度相差不大，且优于SINS/DR测量方法。

关 键 词：车载重力测量；重力异常；直接差分法；里程计；气压高度计
中图分类号：P223+.7
文献标志码：A **文章编号**：1001-5965(2024)10-3162-10

车载重力测量可以实现重力连续测量或快速定点停车测量，显著提高重力加密测量的效率[1]。车载重力测量相比于卫星和航空重力测量更加灵活机动，较低的运行速度保证了更高的空间分辨率，沿地面测量信噪比也会得到增强[2]。车载重力测量无须进行高度延拓，时间和经济成本都较低[3]。自俄亥俄州立大学进行首次车载重力测量试验[4]，至今已沉淀发展了近20年。全球定位系统(global positioning system, GPS)在传统的重力测量中对于定位、测速、载体加速度的获取具有不可替代的作用[5]，但GPS数据的使用受到复杂环境的限制，容易被干扰、遮蔽甚至欺骗，无源性和自主性是一大挑战[6-7]。

近年来，车载捷联惯导(strapdown inertial navigation system, SINS)和里程计(odometer, OD)的组合导航SINS/OD替代GPS用于精确定位的方法得到快速发展[8-9]。Yu等提出GPS/SINS/OD/航位推算(dead reckoning, DR)的组合系统可以满足高精度定位要求，但未实现完全自主性[10]。Gao等提出以方位角误差作为量测的组合系统可以进一步提升定位精度，但未分析测速精度的变化[11]。在无GPS条件下，刘鹏飞证实了车载SINS与OD采用Kalman滤波可以实现载体位置、速度、加速度和姿态的高精度估计[12]。Yu等提出SINS/测速仪(velocimeter, VEL)组合的方法用于车载重力测量，试验结果表明，该方法具有与SINS/GPS相当的测量精度[13]，但定位精度尤其是高度上的误差会影响重力测量的精度。Yu等在SINS/GPS和SINS/VEL的基础上又提出一种SINS/GPS/VEL的测量方法[14]，但自主性又是一大问题。

鉴于重力测量领域大部分研究工作严重依赖GPS，本文基于SINS/OD/高度计提出一种新的车载重力异常测量方法以解决自主性问题，并建立了不同量测的数学模型，对比分析量测选择对重力异常

收稿日期：2022-09-07；**录用日期**：2023-01-18；**网络出版时间**：2023-02-27 13:53
网络出版地址：link.cnki.net/urlid/11.2625.V.20230227.1100.001
基金项目：航空科学基金(201808U8004)
*通信作者. E-mail：zzftxy@163.com

引用格式：李新宇，张志利，周召发，等. 基于SINS/OD/高度计的车载重力异常测量方法[J]. 北京航空航天大学学报，2024，50(10)：3162-3171. LI X Y，ZHANG Z L，ZHOU Z F，et al. Vehicle gravity anomaly measurement methods based on SINS/OD/altimeter[J]. Journal of Beijing University of Aeronautics and Astronautics，2024，50(10)：3162-3171 (in Chinese).

测量精度的影响。仿真及试验结果表明, 本文方法具有可行性与实用性, 满足自主性要求的同时, 测量精度也会有所提升, 为实现完全自主动基座高精度车载重力测量提供参考。

1　重力测量理论基础

1.1　基本原理

重力测量的基本原理符合地球重力场的牛顿运动定律[15]。静基座下, 载体运动加速度为 0, 通过调整重力传感器方向并取模值, 就可以感知重力加速度。静态重力测量可以达较高的精度, 但操作起来费时费力、效率较低。重力传感器搭载运载体处于运动状态的动基座重力测量是节约成本、提高效率的有效措施。

爱因斯坦广义相对论的等效原理指出, 在封闭的系统内, 加速度所造成的“重量感”和万有引力是等效的[16]。因此, 动基座重力测量重力传感器测量的信息中既包含重力加速度信息, 又包含载体运动加速度, 这对于重力传感器是区分不开的。直接差分法是使用两套测量系统将重力传感器的比力信息和载体运动加速度分别求出, 将两套系统的输出在同一坐标系下做差值处理, 从而消除载体加速度的影响, 在经过对各误差项的估计和补偿, 实现重力信息的求解[17-19]。

1.2　数学模型

选取东北天坐标系为导航坐标系 n, 动基座重力测量的表达式在 n 系下可表示为

$$\boldsymbol{g}^n = \dot{\boldsymbol{v}}_e^n - \boldsymbol{C}_b^n \boldsymbol{f}^b + \left(2\boldsymbol{\omega}_{ie}^n + \boldsymbol{\omega}_{en}^n\right) \times \boldsymbol{v}_e^n \tag{1}$$

式中: $\boldsymbol{g}^n$为导航坐标系下的重力矢量; $\boldsymbol{v}_e^n$和$\dot{\boldsymbol{v}}_e^n$分别为载体速度和载体加速度; $\boldsymbol{C}_b^n$为载体坐标系 b 到导航坐标系 n 的方向余弦矩阵; $\boldsymbol{f}^b$为 b 系中比力测量值; $\boldsymbol{\omega}_{ie}^n$为地球自转加速度在 n 系下的投影; $\boldsymbol{\omega}_{en}^n$为 n 系相对 e 系的转动角速度。

实际重力是扰动重力和正常重力之和[20-21]:

$$\boldsymbol{g}^n = \delta\boldsymbol{g}^n + \boldsymbol{\gamma}^n \tag{2}$$

则扰动重力矢量$\delta\boldsymbol{g}^n$可表示为

$$\delta\boldsymbol{g}^n = \dot{\boldsymbol{v}}_e^n - \boldsymbol{f}^n + \left(2\boldsymbol{\omega}_{ie}^n + \boldsymbol{\omega}_{en}^n\right) \times \boldsymbol{v}_e^n - \boldsymbol{\gamma}^n \tag{3}$$

式中: $\boldsymbol{\gamma}^n$为导航坐标系下正常重力矢量。

本文主要针对重力异常展开研究, 将式 (3) 表示为地理坐标系下的分量形式, 天向通道表达式为

$$\delta g_U = \dot{v}_U - f_U + 2\omega_{ie} v_E \cos L + \frac{v_E^2}{R_{Nh}} + \frac{v_N^2}{R_{Mh}} - \gamma \tag{4}$$

式 (4) 为重力标量测量的基本原理。式中: δg_U为重力扰动矢量垂向分量; v_E、v_N和v_U分别为载体速度的东向、北向和天向分量; $\dot{v}_U$为垂向加速度; f_U为比力测量值的垂向分量; L 为纬度; R_{Mh}、R_{Nh}分别为子午圈半径、卯酉圈半径与高度的和; γ为正常重力矢量的垂向分量。正常重力采用 WGS-84 模型下的正常重力公式计算, 详见参考文献 [22]。

2　测量方法

为获得精确的比力和姿态信息, 将外部观测的速度和位置作为量测, 通过 Kalman 滤波进行组合导航解算[23], 并借助气压高度计抑制高度发散。动基座重力测量与组合导航是一体化的[24], 组合导航得到的导航参数按如图 1 所示的流程进一步处理即可实现重力异常测量, 其中, FIR 为有限脉冲响应, EGM 为地球重力场模型。

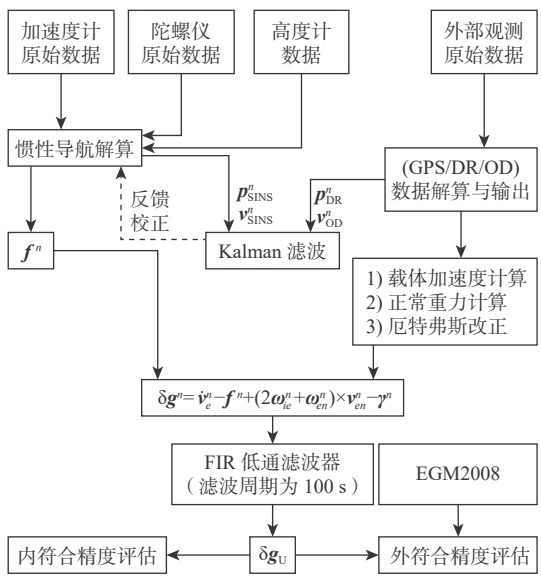


图 1　重力测量流程

Fig. 1　Flow of gravimetry

2.1　SINS/GPS 测量方法

选择 SINS 的姿态误差$\boldsymbol{\phi}^n$、速度误差$\delta\boldsymbol{v}^n$、位置误差$\delta\boldsymbol{p}^n$为系统状态量, 同时把加速度计零偏∇^b和陀螺仪零偏$\boldsymbol{\varepsilon}^b$加入状态变量中, 有

$$\boldsymbol{X}_1 = \left[\left(\boldsymbol{\phi}^n\right)^{\mathrm{T}} \quad \left(\delta\boldsymbol{v}^n\right)^{\mathrm{T}} \quad \left(\delta\boldsymbol{p}^n\right)^{\mathrm{T}} \quad \left(\boldsymbol{\varepsilon}^b\right)^{\mathrm{T}} \quad \left(\nabla^b\right)^{\mathrm{T}}\right]^{\mathrm{T}} \tag{5}$$

惯性器件的误差微分方程为[25]

$$\dot{\boldsymbol{\varepsilon}}^b = \boldsymbol{0} \tag{6}$$

$$\dot{\nabla}^b = \boldsymbol{0} \tag{7}$$

选取 GPS 位置、速度与 SINS 位置、速度之差作为量测信息, 有

$$\boldsymbol{Z}_1 = \begin{bmatrix} \delta\boldsymbol{p} \\ \delta\boldsymbol{v} \end{bmatrix} = \begin{bmatrix} \boldsymbol{p}_{\mathrm{GNSS}}^n - \boldsymbol{p}_{\mathrm{SINS}}^n \\ \boldsymbol{v}_{\mathrm{GNSS}}^n - \boldsymbol{v}_{\mathrm{SINS}}^n \end{bmatrix} \tag{8}$$

基于式 (6)、式 (7) 及 SINS 姿态、速度和位置误差方程[22], 可推导出 15 维的 Kalman 状态方程和量

测方程为

$$\dot{\boldsymbol{X}}_1 = \boldsymbol{F}_{\mathrm{SINS}}\boldsymbol{X}_1 + \boldsymbol{W} \tag{9}$$

$$\boldsymbol{Z}_1 = \boldsymbol{H}_1\boldsymbol{X}_1 + \boldsymbol{V} \tag{10}$$

$$\boldsymbol{H}_1 = \begin{bmatrix} \boldsymbol{0}_{3\times3} & \boldsymbol{0}_{3\times3} & \boldsymbol{I}_{3\times3} & \boldsymbol{0}_{3\times3} & \boldsymbol{0}_{3\times3} \\ \boldsymbol{0}_{3\times3} & \boldsymbol{I}_{3\times3} & \boldsymbol{0}_{3\times3} & \boldsymbol{0}_{3\times3} & \boldsymbol{0}_{3\times3} \end{bmatrix} \tag{11}$$

式中：$\boldsymbol{W}$、$\boldsymbol{V}$分别为等效系统噪声矩阵和量测噪声矩阵，满足零均值白噪声的要求；$\boldsymbol{F}_{\mathrm{SINS}}$为状态量系数矩阵，表达式详见参考文献[26]。

2.2 SINS/OD 测量方法

从测量原理的角度分析，速度及加速度误差是影响车载重力测量的一个重要因素[27]。本节以OD提供的速度信息与SINS速度信息之差为量测，可以有效提高速度及加速度精度，进而提高重力测量的精度。

2.2.1 系统状态方程

SINS状态量同2.1节的15维状态量$\boldsymbol{X}_1$，OD提供的速度状态量为载体坐标系测量误差$\delta\boldsymbol{v}^b_{\mathrm{OD}}$，共3个状态变量，即

$$\boldsymbol{X}_{\mathrm{OD}} = \delta\boldsymbol{v}^b_{\mathrm{OD}} = \begin{bmatrix} \delta v^b_{\mathrm{OD}x} & \delta v^b_{\mathrm{OD}y} & \delta v^b_{\mathrm{OD}z} \end{bmatrix}^{\mathrm{T}} \tag{12}$$

SINS/OD重力测量方法的系统状态方程为

$$\dot{\boldsymbol{X}}_2 = \boldsymbol{F}\boldsymbol{X}_2 + \boldsymbol{W} \tag{13}$$

式中：$\boldsymbol{X}_2 = [\boldsymbol{X}_1^{\mathrm{T}} \quad \boldsymbol{X}_{\mathrm{OD}}^{\mathrm{T}}]^{\mathrm{T}}$为18维系统状态变量；$\boldsymbol{F}$为18×18维状态量系数矩阵，表示为

$$\boldsymbol{F} = \begin{bmatrix} \boldsymbol{F}_{\mathrm{SINS}} & \boldsymbol{0}_{15\times3} \\ \boldsymbol{0}_{3\times15} & \boldsymbol{0}_{3\times3} \end{bmatrix} \tag{14}$$

其中：$\boldsymbol{F}_{\mathrm{SINS}}$为15×15维SINS状态量系数矩阵，根据SINS的误差方程确定。

2.2.2 速度量测方程

由于姿态误差的存在，SINS数学平台对应的坐标系为n'系，通过姿态解算得到的姿态矩阵为$\boldsymbol{C}^{n'}_b$，实际的姿态矩阵为$\boldsymbol{C}^n_b$，n'系与导航坐标系n的转换矩阵为$\boldsymbol{C}^{n'}_n$。SINS解算的速度为$\hat{\boldsymbol{v}}_{\mathrm{SINS}}$，OD测量的速度为$\hat{\boldsymbol{v}}_{\mathrm{OD}}$，取两者之差作为系统量测：

$$\boldsymbol{Z}_2 = \hat{\boldsymbol{v}}_{\mathrm{SINS}} - \hat{\boldsymbol{v}}_{\mathrm{OD}} \tag{15}$$

根据速度误差的定义，可得

$$\hat{\boldsymbol{v}}^n_{\mathrm{SINS}} = \boldsymbol{v}^n_{\mathrm{SINS}} + \delta\boldsymbol{v}^n_{\mathrm{SINS}} \tag{16}$$

计算n'系相对n系的姿态误差角为$\boldsymbol{\phi}^n$，考虑$\boldsymbol{\phi}^n$为小角度，则

$$\boldsymbol{C}^{n'}_n = \boldsymbol{I} - (\boldsymbol{\phi}^n\times) \tag{17}$$

则在计算n'系中OD提供的速度信息为

$$\boldsymbol{v}^{n'}_{\mathrm{OD}} = \boldsymbol{C}^{n'}_n\boldsymbol{C}^n_b\hat{\boldsymbol{v}}^b_{\mathrm{OD}} = \left(\boldsymbol{I} - (\boldsymbol{\phi}^n\times)\right)\boldsymbol{C}^n_b\left(\boldsymbol{v}^b_{\mathrm{OD}} + \delta\boldsymbol{v}^b_{\mathrm{OD}}\right) \tag{18}$$

将式(18)展开去掉二阶小量，并将$\boldsymbol{v}^b_{\mathrm{OD}} = \boldsymbol{v}^b_{\mathrm{SINS}}$代入，可得

$$\hat{\boldsymbol{v}}^n_{\mathrm{OD}} = \boldsymbol{v}^n_{\mathrm{SINS}} - \boldsymbol{\phi}^n\times\boldsymbol{v}^n_{\mathrm{SINS}} + \boldsymbol{C}^n_b\delta\boldsymbol{v}^b_{\mathrm{OD}} \tag{19}$$

由式(16)和式(19)相减可得

$$\delta\boldsymbol{v}^n = \hat{\boldsymbol{v}}^n_{\mathrm{SINS}} - \hat{\boldsymbol{v}}^n_{\mathrm{OD}} = \delta\boldsymbol{v}^n_{\mathrm{SINS}} - \boldsymbol{v}^n_{\mathrm{SINS}}\times\boldsymbol{\phi}^n - \boldsymbol{C}^n_b\delta\boldsymbol{v}^b_{\mathrm{OD}} \tag{20}$$

系统的量测方程设为

$$\boldsymbol{Z}_2 = \boldsymbol{H}_2\boldsymbol{X}_2 + \boldsymbol{V} \tag{21}$$

$$\boldsymbol{H}_2 = \begin{bmatrix} \boldsymbol{H}_{21} & \boldsymbol{I}_{3\times3} & \boldsymbol{0}_{3\times9} & \boldsymbol{H}_{22} \end{bmatrix} \tag{22}$$

$$\boldsymbol{H}_{21} = \begin{bmatrix} 0 & v^n_U & -v^n_N \\ -v^n_N & 0 & v^n_E \\ v^n_N & -v^n_E & 0 \end{bmatrix} \tag{23}$$

$$\boldsymbol{H}_{22} = -\boldsymbol{C}^n_b \tag{24}$$

2.3 SINS/DR 测量方法

针对上述方法没有外部位置观测而导致位置积累影响长距离重力测量精度的问题，本节提出以DR提供的位置与SINS位置信息误差作为量测的重力测量方法，可以有效控制定位误差影响重力测量结果：

$$\boldsymbol{Z}_3 = \hat{\boldsymbol{p}}_{\mathrm{SINS}} - \hat{\boldsymbol{p}}_{\mathrm{DR}} = \delta\boldsymbol{p}_{\mathrm{SINS}} - \delta\boldsymbol{p}_{\mathrm{DR}} \tag{25}$$

将DR的位置误差加入状态量，即$\boldsymbol{X}_3 = \begin{bmatrix} \boldsymbol{X}_1^{\mathrm{T}} & \boldsymbol{X}_{\mathrm{DR}}^{\mathrm{T}} \end{bmatrix}^{\mathrm{T}}$，其中，$\boldsymbol{X}_1$同2.1节。

$$\boldsymbol{X}_{\mathrm{DR}} = \begin{bmatrix} \delta p^n_{\mathrm{DR}E} & \delta p^n_{\mathrm{DR}N} & \delta p^n_{\mathrm{DR}U} \end{bmatrix}^{\mathrm{T}} \tag{26}$$

式中：$\delta p^n_{\mathrm{DR}E}$、$\delta p^n_{\mathrm{DR}N}$、$\delta p^n_{\mathrm{DR}U}$分别为DR的位置误差在导航坐标系下的3个分量。

则量测方程可写为

$$\boldsymbol{Z}_3 = \boldsymbol{H}_3\boldsymbol{X}_3 + \boldsymbol{V} \tag{27}$$

$$\boldsymbol{H}_3 = \begin{bmatrix} \boldsymbol{0}_{3\times6} & \boldsymbol{I}_{3\times3} & \boldsymbol{0}_{3\times6} & -\boldsymbol{I}_{3\times3} \end{bmatrix} \tag{28}$$

状态方程的建立同2.2节。

2.4 SINS/DR/OD 测量方法

在SINS/DR和SINS/OD的基础上，提出SINS/DR/OD测量方法，同时将位置和速度信息作为量测，可以更好地控制误差发散，得到更高精度的重力测量结果。将DR的位置误差、OD速度误差列入状态量，有

$$\boldsymbol{X}_4 = \left[(\boldsymbol{X}_{\mathrm{SINS}})^{\mathrm{T}} \quad (\delta\boldsymbol{p}^n_{\mathrm{DR}})^{\mathrm{T}} \quad (\delta\boldsymbol{v}^n_{\mathrm{OD}})^{\mathrm{T}} \right]^{\mathrm{T}} \tag{29}$$

式中：$\boldsymbol{X}_{\mathrm{SINS}}$同SINS状态$\boldsymbol{X}_1$；$\delta\boldsymbol{p}^n_{\mathrm{DR}}$和$\delta\boldsymbol{v}^n_{\mathrm{OD}}$分别为DR的位置误差和OD速度误差。

则量测及量测方程为

$$\boldsymbol{Z}_4 = \begin{bmatrix} \delta\boldsymbol{p}^n_{\mathrm{SINS}} - \delta\boldsymbol{p}^n_{\mathrm{DR}} \\ \delta\boldsymbol{v}^n_{\mathrm{SINS}} - \delta\boldsymbol{v}^n_{\mathrm{OD}} \end{bmatrix} \tag{30}$$

$$\boldsymbol{Z}_4 = \boldsymbol{H}_4\boldsymbol{X}_4 + \boldsymbol{V} \tag{31}$$

$$\boldsymbol{H}_4=\begin{bmatrix}\boldsymbol{0}_{3\times6} & \boldsymbol{I}_{3\times3} & \boldsymbol{0}_{3\times6} & -\boldsymbol{I}_{3\times3} & \boldsymbol{0}_{3\times3}\\ \boldsymbol{H}_{21} & \boldsymbol{I}_{3\times3} & \boldsymbol{0}_{3\times6} & \boldsymbol{0}_{3\times6} & \boldsymbol{H}_{22}\end{bmatrix} \tag{32}$$

式中: $\boldsymbol{H}_{21}$和$\boldsymbol{H}_{22}$的表达式同 2.2.2 节所述。

3　仿真与试验

为对比上述 4 种方法的测量精度, 本文进行了仿真和跑车试验。以 EGM2008 模型计算所得的重力异常为真值, 计算内符合精度和外符合精度作为测量结果精度评估指标, 计算公式详见参考文献 [28]。

3.1　仿真条件

从 (N34.246°, E108.908°, 380 m) 点出发, 模拟试验车往返行驶轨迹, 共 3 个来回 6 条重复测线。往返测线间距约为 20 m, 可视为重复测线, 且单测线稳定运行后保持匀速直线运动状态。仿真轨迹高度保持不变, 叠加 1 m 以内的随机误差模拟气压高度计测量值。重复测线仿真参数如表 1 所示, 仿真轨迹如图 2 所示。

表 1　重复测线仿真参数

Table 1　Repeated measurement line simulation parameters

参数	数值
载体速度/(m·s^{-1})	15
单测线距离/km	28
测线间距/m	20
测量方向	南北
加速度计常值零偏/10^{-5} (m·s^{-2})	50
加速度计随机游走/10^{-5} (m·s^{-2})	5
陀螺仪常值零偏/((°)·h^{-1})	0.01
陀螺仪随机游走/((°)·h^{-1})	0.001
位置精度/m	0.1
速度精度/(m·s^{-1})	0.05
臂杆误差/m	[0;0;−1]
时间不同步/s	0.01
安装误差	[30′;10′;20′]
标度因数	1
测高精度/m	1
滤波周期/s	100
通带截止频率/Hz	0.01
阻带截止频率/Hz	0.05

3.2　跑车试验

为保证器件测量精度, 试验所选道路的路况较好, 车流量较少, 且车速不宜超过 100 km/h。本次跑车试验选择在晴朗无风的天气下进行, 尽量保持 50 km/h 匀速行驶, 减少急转弯、急刹车等情况。由 (N34.171°, E108.948°, 451 m) 出发, 至 (N34.085°, E109.017°, 432 m) 结束, 总里程共 28.7 km, 用时 34 min。试验所用车如图 3 所示, 该试验车配备有高精度 SINS、轮式 OD、GPS 接收机和气压高度计。OD 选用型号为 ZLC-512/H-J, 转轴转动一周, OD 输出 512 个脉冲, 脉冲输出精度小于 1×10^{-4}, 即旋转转轴 50 圈, 输出脉冲 (25 600±2) 个。标定后,

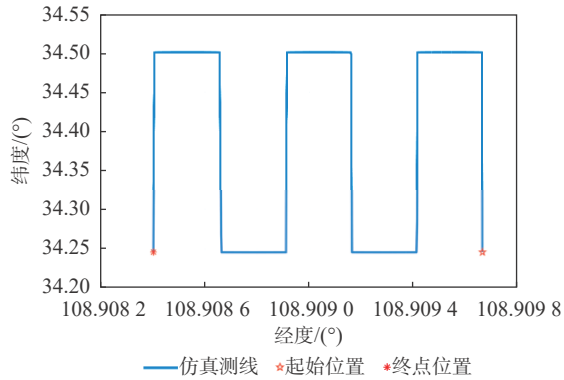


图 2　车载重力测量重复测线仿真轨迹

Fig. 2　Simulation track of repeated measurement lines for vehicle gravimetry



图 3　动基座重力测量试验用车

Fig. 3　Vehicle for gravimetry test on moving pedestal

刻度系数为 0.005 1, 安装误差为 [29.91′;0;23.38′], 由 OD 测速原理可知, 该型号实际使用时测速精度可达 0.001 1 m/s。气压高度计选用型号为 MG11-1000, 输出频率为 10 Hz, 在−200～5 000 m 内测量精度不大于 1 m。车载重力测量试验路线如图 4 所示, SINS 和 OD 组合导航的高度误差是发散的, 故在跑车试验中引入气压高度计辅助 SINS 和 OD, 抑制高度通道误差的发散。该试验进行了单测线测量, 通过计算外符合精度评估重力测量效果。

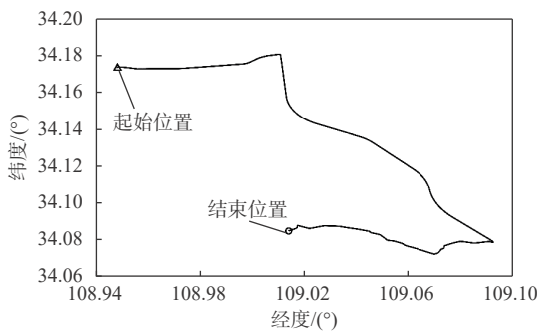


图 4　车载重力测量试验路线

Fig. 4　Roadmap of vehicle gravimetry test

4　结果与讨论

试验行进过程中, 环境比较复杂, 重力信息具有低频特性, 采用 FIR 低通滤波器将高频噪声去除后, 再对重力异常信息作内外符合精度评估[29]。

4.1　仿真结果

SINS/GPS、SINS/DR、SINS/OD、SINS/DR/OD

这 4 种测量方法的仿真结果经低通滤波处理后如图 5 所示，图中，Line1～Line 6 分别表示 6 条重复测线上的重力异常测量值。以 Line 2 为例，4 种方法测量结果的对比如图 6 所示。

4 种测量方法仿真试验计算得到的重复测线重力异常值的均值、标准差等数字特征和各测线内符合精度、外符合精度及总符合精度结果如表 2～表 5 所示。

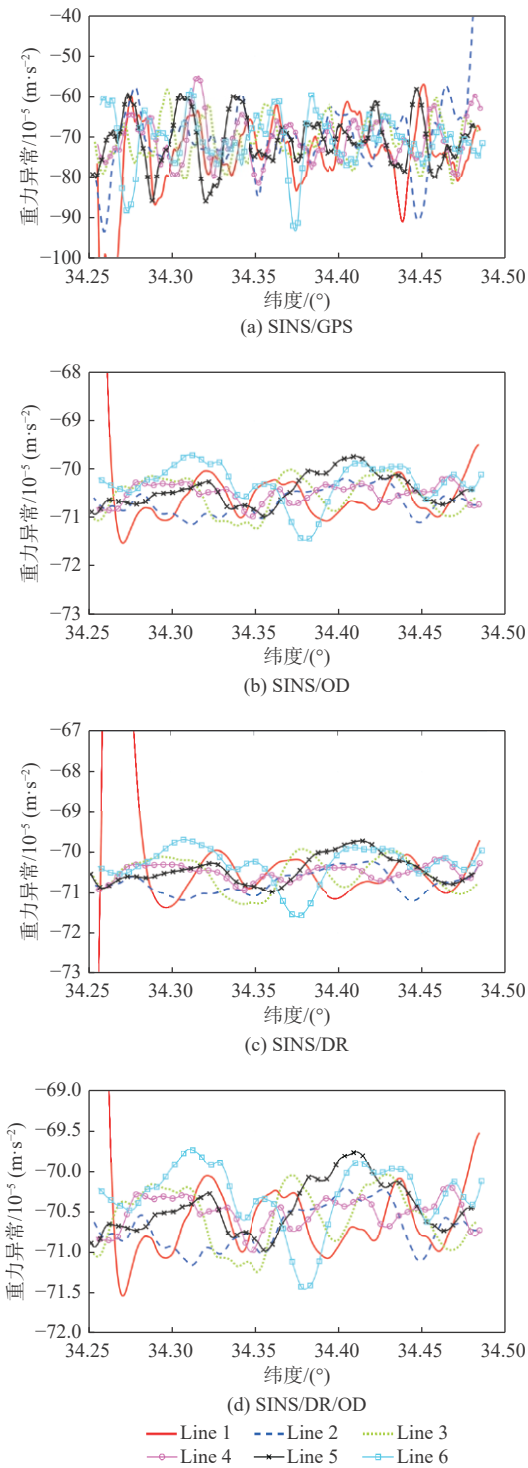


图 5　4 种方法重力异常测量结果

Fig. 5　Gravity anomaly measurement results of four methods

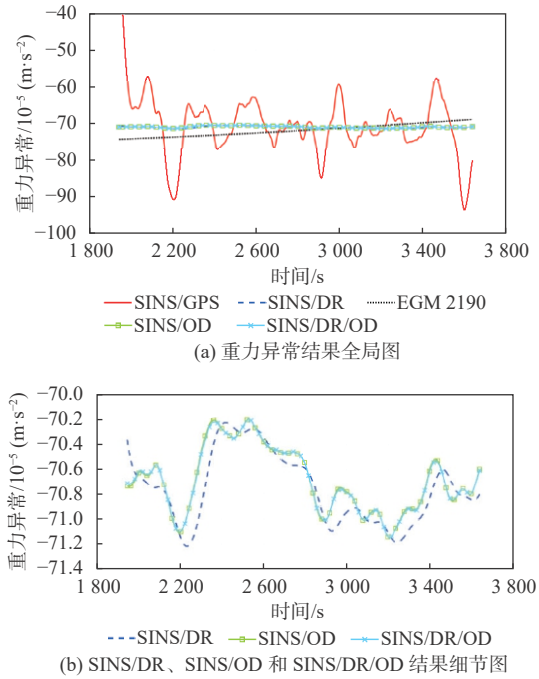


图 6　4 种方法重力异常仿真结果对比

Fig. 6　Simulation results comparison of gravity anomalies of four methods

在本文仿真试验中，SINS/DR 相比 SINS/GPS，内符合精度从 6.61×10^{-5} m/s^2 提升到 1.09×10^{-5} m/s^2，精度提升了 83.51%，外符合精度从 7.47×10^{-5} m/s^2 提升到 2.27×10^{-5} m/s^2，精度提升了 69.61%；SINS/OD 将符合精度提升到 7.00×10^{-6} m/s^2 和 2.16×10^{-5} m/s^2，精度分别提升了 89.41% 和 71.08%；SINS/DR/OD 可以将精度进一步提升。

4.2　跑车试验结果

跑车试验的组合导航结果及定位误差如图 7 和图 8 所示，总里程为 28.7 km，可以将定位误差控制在 20 m 以内。定位误差对重力异常测量结果不会产生显著性影响[30]，300 m 的纬度误差造成的重力扰动误差不会大于 2.0×10^{-6} m/s^2[26]，相比于 GPS 定

表 2　SINS/GPS 仿真精度统计

Table 2　Accuracy statistics of SINS/GPS simulation

精度	测线	均值/ 10^{-5} (m·s^{-2})	标准差/ 10^{-5} (m·s^{-2})	符合精度/ 10^{-5} (m·s^{-2})	总符合精度/ 10^{-5} (m·s^{-2})
内符合精度	Line 1	−3.20	8.19	8.79	6.61
	Line 2	1.17	8.09	8.16	
	Line 3	0.46	5.41	5.42	
	Line 4	0.71	5.19	5.24	
	Line 5	0.42	5.16	5.17	
	Line 6	0.45	5.89	5.90	
外符合精度	Line 1	−3.05	10.24	10.67	7.47
	Line 2	1.40	9.28	9.37	
	Line 3	0.79	5.62	5.67	
	Line 4	1.19	5.67	5.79	
	Line 5	0.95	6.57	6.63	
	Line 6	1.19	6.57	6.67	

表 3　SINS/DR 仿真精度统计

Table 3　Accuracy statistics of SINS/DR simulation

精度	测线	均值/ 10^{-5} (m·s^{-2})	标准差/ 10^{-5} (m·s^{-2})	符合精度/ 10^{-5} (m·s^{-2})	总符合精度/ 10^{-5} (m·s^{-2})
内符合精度	Line 1	0.53	2.29	2.36	1.09
	Line 2	−0.35	0.54	0.64	
	Line 3	−0.14	0.58	0.59	
	Line 4	−0.10	0.47	0.48	
	Line 5	−0.04	0.59	0.59	
	Line 6	0.10	0.55	0.56	
外符合精度	Line 1	1.64	2.51	2.99	2.27
	Line 2	0.83	1.70	1.89	
	Line 3	1.16	1.61	1.98	
	Line 4	1.34	1.61	2.10	
	Line 5	1.44	1.74	2.26	
	Line 6	1.79	1.61	2.41	

表 4　SINS/OD 仿真精度统计

Table 4　Accuracy statistics of SINS/OD simulation

精度	测线/ 10^{-5} (m·s^{-2})	均值/ 10^{-5} (m·s^{-2})	标准差/ 10^{-5} (m·s^{-2})	符合精度/ 10^{-5} (m·s^{-2})	总符合精度/ 10^{-5} (m·s^{-2})
内符合精度	Line 1	0.13	1.44	1.44	0.70
	Line 2	−0.22	0.39	0.45	
	Line 3	−0.08	0.44	0.45	
	Line 4	−0.03	0.35	0.35	
	Line 5	0.03	0.40	0.41	
	Line 6	0.17	0.38	0.42	
外符合精度	Line 1	1.17	2.01	2.32	2.16
	Line 2	0.90	1.71	1.93	
	Line 3	1.15	1.63	1.99	
	Line 4	1.34	1.61	2.09	
	Line 5	1.45	1.74	2.26	
	Line 6	1.79	1.58	2.39	

表 5　SINS/DR/OD 仿真精度统计

Table 5　Accuracy statistics of SINS/DR/OD simulation

精度	测线/ 10^{-5} (m·s^{-2})	均值/ 10^{-5} (m·s^{-2})	标准差/ 10^{-5} (m·s^{-2})	符合精度/ 10^{-5} (m·s^{-2})	总符合精度/ 10^{-5} (m·s^{-2})
内符合精度	Line 1	0.13	1.43	1.44	0.69
	Line 2	−0.22	0.39	0.45	
	Line 3	−0.08	0.44	0.44	
	Line 4	−0.03	0.35	0.35	
	Line 5	0.03	0.40	0.40	
	Line 6	0.17	0.38	0.41	
外符合精度	Line 1	1.17	2.01	2.32	2.15
	Line 2	0.90	1.71	1.93	
	Line 3	1.15	1.62	1.98	
	Line 4	1.34	1.61	2.09	
	Line 5	1.45	1.74	2.24	
	Line 6	1.79	1.58	2.38	

位误差增大 20 m 造成的重力测量异常误差微乎其微, 可以忽略。

高度计辅助 SINS 和 OD 的测高结果如图 9 所示。SINS/OD/高度计在总里程 30 km 内时, 高度定位精度优于 0.5 m, 按照地球重力垂向梯度约 3.086×10^{-6} (m·s^{-2})/m 计算, 高度引起的重力异常约

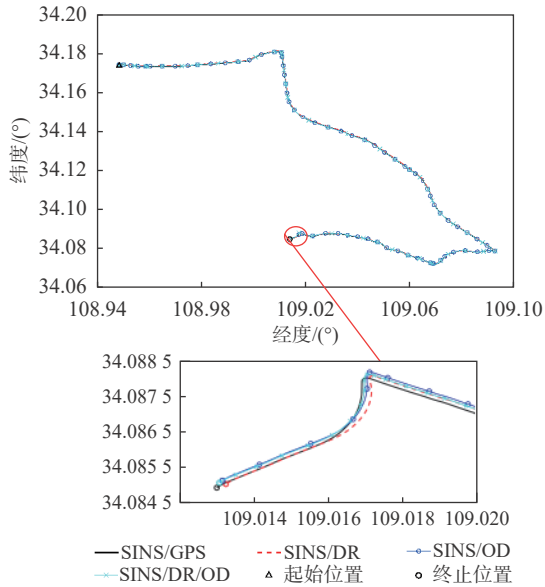


图 7　组合导航定位结果

Fig. 7　Positioning result of integrated navigation

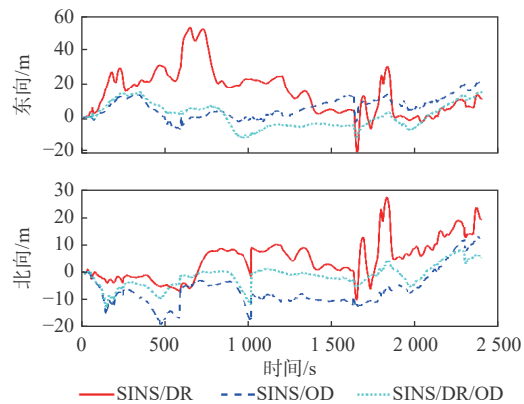


图 8　组合导航定位误差

Fig. 8　Positioning errors of integrated navigation

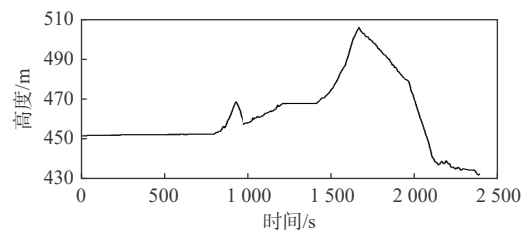


图 9　高度通道定位结果

Fig. 9　Height channel positioning results

为 1.5×10^{-6} m/s^2, 该量级在实际测量中可以忽略。

本文试验中, 实际测线 4 种测量方法得到的重力异常值及 EGM 2008 模型计算的标准值如图 10 所示, 符合精度统计结果如表 6 所示。

跑车试验的外符合精度统计结果为: SINS/GPS 为 2.89×10^{-5} m/s^2, SINS/DR 为 0.97×10^{-5} m/s^2, SINS/OD 为 0.53×10^{-5} m/s^2, SINS/DR/OD 为 0.49×10^{-5} m/s^2。SINS/GPS 传统方法测量结果中段存在较大波动, 导致精度较差。本文的 3 种方法精度分别提升了

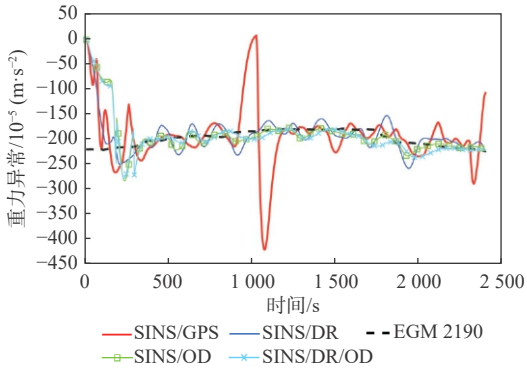


图 10 4 种方法重力异常实验结果对比

Fig. 10 Experiment results comparison of gravity anomalies of four methods

表 6 4 种测量方法精度统计

Table 6 Accuracy statistics of four measurement methods

精度	测量方法	均值/ 10^{-5} (m·s^{-2})	方差/ 10^{-5} (m·s^{-2})	符合精度/ 10^{-5} (m·s^{-2})
外符合精度	SINS/GPS	4.12	1.31	2.89
	SINS/DR	2.13	0.11	0.97
	SINS/OD	1.11	0.02	0.53
	SINS/DR/OD	0.73	0.01	0.49

66.44%、81.66% 和 83.04%, 跑车试验所得结果同仿真结论一致。

4.3 结果分析

结合式 (4) 可知, 测速精度和加速度测量精度直接影响重力异常的测量效果。试验结果表明, 本文方法相比 SINS/GPS 重力测量精度有明显提升, 其原因是: GPS 辅助情况下定位精度虽然很高, 但测速及加速度测量精度不及 OD 辅助的测量方法。以仿真 Line 1 为例, 载体速度由组合导航系统给出, 载体加速度由速度差分后得到。由于稳定运行后载体处于匀速运动状态, 即加速度真值为 0, 解算得到的加速度值即为加速度误差, 由此可得 SINS/GPS 方法与 SINS/OD 方法测速和加速度精度对比如图 11 和图 12 所示。经计算, OD 辅助的测速精度可达 0.002 m/s, 加速度测量误差经低通滤波后不大于 10^{-5} m/s^2, 该量级与实际使用时精度级别相符。

为进一步分析本文方法测量效果与器件性能参数之间的具体关系, 选取 SINS/OD/高度计以速度为量测的组合测量方案, 改变 OD 标定因数及高度计测高精度的仿真条件, 进行多次仿真试验, 并以内符合精度为指标绘制对应关系, 如图 13 所示。标度因数 K_{od} 表示实际行驶里程与 OD 测量里程的比值, K_{od}=1 表示没有标度因数误差, 且标度因数误差一般在 0.1 以内。图 13 中, δh 表示气压高度计测量精度。

由图 13 可知, 标度因数越接近于 1, 即标度因数误差越小, 气压高度计测高精度越高, 本文方法的测量精度也就越高。根据标度因数的定义, 易推导 OD 的测量误差 d_{err} 为

$$d_{err} = d_{mea} - d_{real} = \frac{1-K_{od}}{K_{od}} d_{real} = \frac{\delta K_{od}}{K_{od}} d_{real} \tag{33}$$

式中: d_{mea} 和 d_{real} 分别为 OD 的测量值和实际值; δK_{od} 为标度因数误差。故在同等标度因数误差条件下 (如 K_{od}=0.9 和 K_{od}=1.1), 标度因数大于 1 时的测量精度优于标度因数小于 1 的情况。本文仿真及试

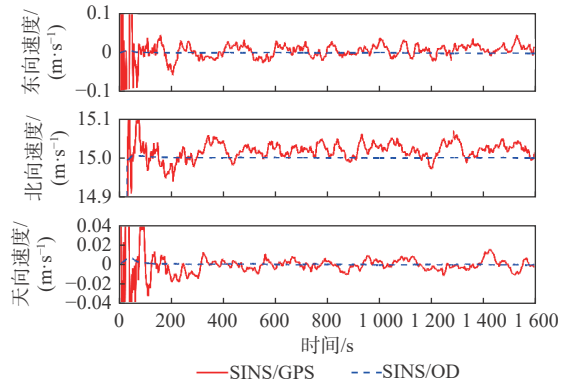


图 11 测速精度对比

Fig. 11 Comparison of velocity measurement accuracy

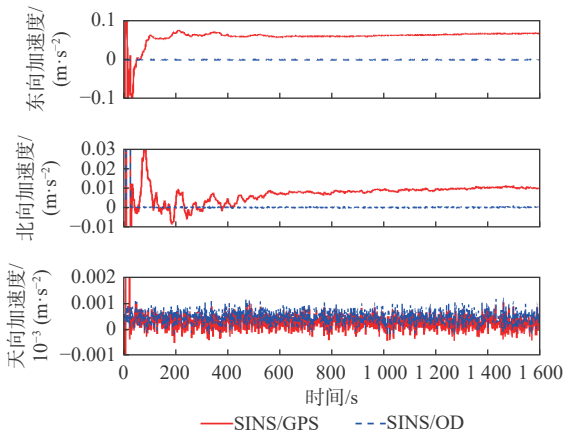


图 12 加速度精度对比

Fig. 12 Comparison of accelerometer accuracy

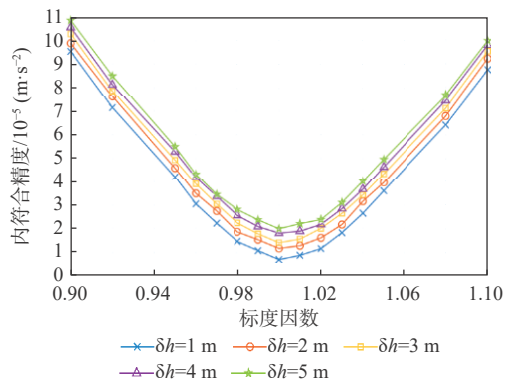


图 13 器件性能参数与测量精度对应关系

Fig. 13 Correspondence between device performance parameters and measurement accuracy

验中的 OD、高度计和 GPS 均为具备典型性能参数的型号, 可见, 当环境状况及器件性能满足试验条件时, 本文方法具备精度优势。

由标准差的对比可以得出, 本文方法的测量稳定性也得到了提升, 且 SINS/OD 优于 SINS/DR 测量方法。对比 Line 1～Line 6 发现, Line 2～Line 5 的测量结果普遍优于 Line 1 和 Line 6, 这是由于 Line 1 处于起步阶段, Line 6 处于停车阶段, 加速度和速度不及其他测线平稳行驶的测量精度。

图 10 中, SINS/GPS 结果曲线在 1 000 s 附近出现较大起伏, 测线在这段时间内 GPS 由于受到高楼、树木等遮挡存在较大测量误差。该时间段内测速结果如图 14 所示。由于 GPS 信号受环境影响大导致载体速度计算不稳定, 对重力测量精度造成不利影响, 从而引起重力结果的较大误差。

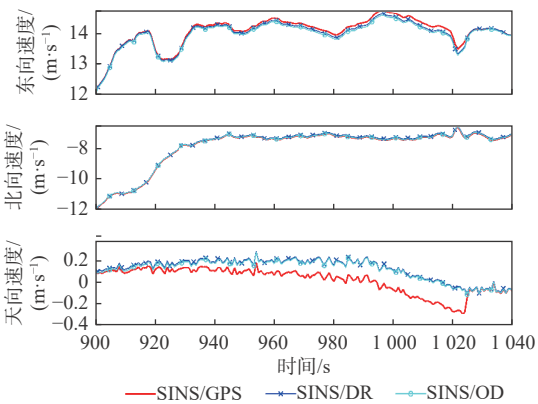


图 14　起伏较大时段速度精度对比

Fig. 14　Comparison of velocity accuracy in period with great fluctuations

5　结　论

1) 通过重复测线对比可知, 车载重力测量在加速、减速和转向时精度较差, 在匀速平稳近直线行驶状态下测量精度较高。

2) 借助里程计提供的速度信息, 结合航位推算和 Kalman 组合导航技术, 可以实现无 GPS 条件下的自主车载动态重力测量, 且本文方法的测量精度与所用器件的性能参数密切相关。

3) GPS 信息易受环境干扰进而导致结果中会出现野值, 里程计辅助的测量方法不仅具有良好的环境适应性, 且稳定性更好。

4) 一般情况下, 本文方法测量效果优于传统的 SINS/GPS 方法, 其中, SINS/OD 测量精度与 SINS/DR/OD 方法相差不大, 且均优于 SINS/DR 方法。

参考文献（References）

[1] YU R H, CAI S K, WU M P, et al. A SINS/GNSS ground vehicle gravimetry test based on SGA-WZ02[J]. Sensors, 2015, 15(9): 23477-23495.

[2] KWON J H. Airborne vector gravimetry using GPS/INS[D]. Columbus: The Ohin State University, 2000.

[3] 黄谟涛, 邓凯亮, 吴太旗, 等. 重力异常向上延拓严密改化模型及向下延拓应用[J]. 测绘学报, 2022, 51(1): 41-52.
HUANG M T, DENG K L, WU T Q, et al. Rigorous modification model of upward continuation and its applications on the downward continuation of gravity anomaly[J]. Acta Geodaetica et Cartographica Sinica, 2022, 51(1): 41-52(in Chinese).

[4] LI X P. Moving base INS/GPS vector gravimetry on a land vehicle [D]. Columbus: The Ohio State University, 2007.

[5] DAI D K, WANG X S, ZHAN D J, et al. Dynamic measurement of high-frequency deflections of the vertical based on the observation of INS/GNSS integration attitude error[J]. Journal of Applied Geophysics, 2015, 119(1):89-98.

[6] ZHANG X H, ZHENG K, LU C X, et al. Acceleration estimation using a single GPS receiver for airborne scalar gravimetry[J]. Advances in Space Research, 2017, 60(10): 2277-2288.

[7] XIONG Z M, CAO J L, LIAO K X, et al. A new method for underwater dynamic gravimetry based on multisensor integrated navigation[J]. Geophysics, 2020, 85(3):69-80.

[8] OUYANG W, WU Y X, CHEN H Y. INS/odometer land navigation by accurate measurement modeling and multiple-model adaptive estimation[J]. IEEE Transactions on Aerospace and Electronic Systems, 2021, 57(1): 245-262.

[9] YANG B, XUE L, FAN H D, et al. SINS/odometer/Doppler radar high-precision integrated navigation method for land vehicle[J]. IEEE Sensors Journal, 2021, 21(13): 15090-15100.

[10] YU Z P, HU Y J, HUANG J M. GPS/INS/odometer/DR integrated navigation system aided with vehicular dynamic characteristics for autonomous vehicle application[J]. IFAC-PapersOnLine, 2018, 51(31): 936-942.

[11] GAO J X, LI K, CHEN J Y. Research on the integrated navigation technology of SINS with couple odometers for land vehicles[J]. Sensors, 2020, 20(2): 546.

[12] 刘鹏飞. 里程计辅助的高精度车载 GNSS/INS 组合导航系统[J]. 光学 精密工程, 2020, 28(4): 979-987.
LIU P F. High-precision vehicle GNSS/INS integrated navigation system aided by odometer[J]. Optics and Precision Engineering, 2020, 28(4): 979-987(in Chinese).

[13] YU R H, WU M P, ZHANG K D, et al. A new method for land vehicle gravimetry using SINS/VEL[J]. Sensors, 2017, 17(4): 766.

[14] YU R H, QIU X T, CAO J L, et al. Improving land vehicle gravimetry using a new SINS/GNSS/VEL method[J]. IOP Conference Series: Earth and Environmental Science, 2020, 513: 012054.

[15] ZHANG Z Q, LI J C, ZHANG K D, et al. Experimental study on underwater moving gravity measurement by using strapdown gravimeter based on AUV platform[J]. Marine Geodesy, 2021, 44(2): 108-135.

[16] 张开东. 基于 SINS/DGPS 的航空重力测量方法研究[D]. 长沙: 国防科学技术大学, 2007.
ZHANG K D. Research on the methods of airborne gravimetry based on SINS/DGPS[D]. Changsha: National University of De-

fense Technology, 2007(in Chinese).

[17] LI X P, JEKELI C. Ground-vehicle INS/GPS vector gravimetry[J]. Geophysics, 2008, 73(2): 11-21.

[18] CAI S K, ZHANG K D, WU M P. Improving airborne strapdown vector gravimetry using stabilized horizontal components[J]. Journal of Applied Geophysics, 2013, 98(3): 79-89.

[19] CAI S K, LU S H, ZHAO L, et al. Improving strapdown airborne gravimetry by yaw tracking rotation modulation[J]. Journal of Coastal Research, 2020, 99(S1): 373-381.

[20] 郝诗文, 张志利, 周召发, 等. 重力扰动对惯性导航系统初始对准的影响[J]. 系统工程与电子技术, 2020, 42(7): 1575-1581.
HAO S W, ZHANG Z L, ZHOU Z F, et al. Influence of gravity disturbance on initial alignment of inertial navigation system[J]. Systems Engineering and Electronics, 2020, 42(7): 1575-1581(in Chinese).

[21] TIE J B, CAO J L, CHANG L B, et al. A model of gravity vector measurement noise for estimating accelerometer bias in gravity disturbance compensation[J]. Sensors, 2018, 18(3): 883-908.

[22] 严恭敏, 翁浚. 捷联惯导算法与组合导航原理[M]. 西安: 西北工业大学出版社, 2019.
YAN G M, WENG J. Strapdown inertial navigation algorithm and integrated navigation principle[M]. Xi'an: Northwestern Polytechnical University Press, 2019(in Chinese).

[23] JOSEFSSON R. New results from strapdown airborne gravimetry using temperature stabilization[J]. Remote Sensing, 2019, 11(22): 2682-2700.

[24] WANG W, GAO J Y, LI D M, et al. Measurements and accuracy evaluation of a strapdown marine gravimeter based on inertial navigation[J]. Sensors, 2018, 18(11): 3902-3914.

[25] KWON J H, JEKELI C. A new approach for airborne vector gravimetry using GPS/INS[J]. Journal of Geodesy, 2001, 74(10): 690-700.

[26] 于瑞航. 捷联式车载重力测量关键技术研究[D]. 长沙: 国防科技大学, 2017.
YU R H. Research on key technologies for strapdown ground vehicle gravimetry[D]. Changsha: National University of Defense Technology, 2017(in Chinese).

[27] 吴美平, 蔡劭琨, 于瑞航, 等. 捷联式重力测量技术研究进展[J]. 导航与控制, 2020, 19(4/5): 161-169.
WU M P, CAI S K, YU R H, et al. Research progress of strapdown gravimetry technique[J]. Navigation and Control, 2020, 19(4/5): 161-169(in Chinese).

[28] WANG M H, WU M P, CAO J L, et al. Strapdown airborne gravimetry quality assessment method based on single survey line data: A study by SGA-WZ02 gravimeter[J]. Sensors, 2018, 18(2): 360-376.

[29] 郎骏健, 梁星辉, 柳林涛, 等. 航空重力傅里叶基追踪低通滤波方法研究[J]. 地球物理学报, 2018, 61(12): 4737-4745.
LANG J J, LIANG X H, LIU L T, et al. Study on Fourier basis tracking low pass filtering method for airborne gravity[J]. Chinese Journal of Geophysics, 2018, 61(12): 4737-4745(in Chinese).

[30] 蔡劭琨. 航空重力矢量测量及误差分离方法研究[D]. 长沙: 国防科学技术大学, 2014.
CAI S K. The research about airborne vector gravimeter and methods of errors separation[D]. Changsha: National University of Defense Technology, 2014(in Chinese).

Vehicle gravity anomaly measurement methods based on SINS/OD/altimeter

LI Xinyu[1, 2], ZHANG Zhili[1, 2], ZHOU Zhaofa[1, 2, *], CHANG Zhenjun[1, 2], ZHAO Zhiqian[1, 2]

(1. School of Missile Engineering, Rocket Force University of Engineering, Xi'an 710025, China;
2. State Key Discipline Laboratory of Armament Launch Theory and Technology, Rocket Force University of Engineering, Xi'an 710025, China)

Abstract: Vehicle gravimetry is carried out along the surface of the earth. Its slow speed and flexible maneuverability can provide favorable conditions for high-precision ground gravimetry. Due to its heavy dependence on the global positioning system (GPS), traditional gravimetry on a moving pedestal cannot achieve complete autonomy. A new autonomous passive method based on strapdown inertial navigation system (SINS)/odometer (OD)/altimeter integrated navigation was proposed to measure vehicle gravity anomaly. The integrated navigation system provided high-precision specific force, attitude, and carrier acceleration. The barometric altimeter was used to suppress the divergence of the altitude channel, and the gravity information could be obtained by the direct difference method. Then, the measurement accuracy of the four methods, namely, position (SINS/DR), velocity (SINS/OD), position velocity (SINS/DR/OD), and traditional SINS/GPS, was compared, and the influence of measurement selection on the measurement effect of the method proposed was analyzed. The results of repeated measurement line simulation and single measurement line vehicle experiment show that the method proposed can achieve fully autonomous vehicle gravity anomaly measurement, and it has general accuracy advantages over the traditional SINS/GPS method. In addition, the SINS/OD and SINS/DR/OD measurement accuracy is not much different and is better than the SINS/DR measurement method.

Keywords: vehicle gravimetry; gravity anomaly; direct difference method; odometer; barometric altimeter

Received: 2022-09-07; **Accepted**: 2023-01-18; **Published Online**: 2023-02-27 13:53
URL: link.cnki.net/urlid/11.2625.V.20230227.1100.001
Foundation item: Aeronautical Science Foundation of China (201808U8004)
* **Corresponding author**. E-mail: zzftxy@163.com

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2022.0772

连续变弯度后缘飞机的滚转机动载荷减缓

雷朝辉[1]，杨超[1]，宋晨[1,2,*]，金天燚[3]，吴志刚[1]

(1. 北京航空航天大学 航空科学与工程学院，北京 100191；2. 北京航空航天大学 无人系统研究院，北京 100191；
3. 上海机电工程研究所，上海 201109)

摘　　要：变弯度后缘机翼具有变形连续、阻力较小、气动噪声较低等优势，越来越多地应用于新概念飞行器的设计之中。基于此，提出一种基于变弯度后缘的飞行器刚弹耦合动力学建模方法，并针对变弯度后缘飞机缩比模型开展了滚转机动仿真分析与风洞试验。结果表明：变弯度后缘可以操纵飞机在2 s内进行180°滚转机动。相较于外侧后缘单独变形，通过内外后缘协同变形可以降低30%以上的滚转机动附加载荷。另外，对比表明滚转角仿真结果与风洞试验数据误差小于6%，验证了所提方法的准确性。

关 键 词：变弯度后缘；刚弹耦合；滚转机动；载荷减缓；仿真分析；风洞试验
中图分类号：V212.1
文献标志码：A　　**文章编号**：1001-5965(2024)10-3172-11

飞机的机动性是指飞机在一定时间内改变飞行速度、高度和方向的能力[1]。飞机在机动过程中由于气动力和惯性力的作用，会产生一定的附加机动载荷，对于小展弦比的战斗机而言，随着机动性设计指标的不断提高，更大的机动载荷带来的结构安全与增重等问题也日渐明显。

不同于传统飞机设计中加强机体结构刚度和强度的方式来对抗更高的机动载荷，随着主动控制技术的出现和不断发展，多舵面联合偏转来改变气动力分布从而减缓关键部位附加载荷的机动载荷减缓控制，可以在不明显增加结构质量的前提下，改变结构中的载荷分布，有效减缓机动过程中的附加载荷。

目前，针对常规舵面飞机机动过程的研究已经取得了一定的进展。国外学者在20世纪后期就开展了一系列应用主动控制技术的飞机机动载荷减缓研究[2-3]。近年来，国内学者也相继开展了有关飞机机动过程的仿真计算方法[4]、滚转机动载荷减缓缩比模型设计分析[5]及风洞试验[6]等研究工作，并取得了相应成果。

飞机机动过程的研究大多以传统飞机为对象，对于变体飞机这类新概念飞行器的研究主要停留在方案设计和演示验证阶段，对其机动过程的研究涉及较少。

变体飞行器是一种在飞行过程中拥有自适应能力的新概念飞行器[7]，其按变形方式可分为面内变形和面外变形两大类，而变弯度则属于后者。

变弯度机翼可以通过改变机翼弯度来调整翼面的升力分布，从而达到提高飞行性能和控制飞机的目的[8]。研究表明，通过调整巡航状态飞机的机翼弯度分布，其升阻比可以增加3%～9%[9]；而机翼弯度分布的改变也可以调整气动力分布，从而有效降低阵风引起的载荷[10-11]；另外，相比传统刚性舵面，变弯度机翼由于没有缝隙，能够降低飞行时的阻力和噪声[12]。

对于变弯度机翼飞机的机动过程，国外开展了

收稿日期：2022-09-14；录用日期：2022-12-19；网络出版时间：2023-01-10 16:01
网络出版地址：link.cnki.net/urlid/11.2625.V.20230109.1359.005
基金项目：国家自然科学基金(11402013)
*通信作者. E-mail：songchen@buaa.edu.cn

引用格式：雷朝辉，杨超，宋晨，等. 连续变弯度后缘飞机的滚转机动载荷减缓[J]. 北京航空航天大学学报，2024，50(10)：3172-3182.
LEI C H，YANG C，SONG C，et al. Rolling maneuver load alleviation of aircraft with continuously variable-camber trailing edge[J]. Journal of Beijing University of Aeronautics and Astronautics，2024，50(10)：3172-3182 (in Chinese).

一些研究, 文献 [13] 对后缘变弯度机翼飞机开展了飞行试验, 研究了其在实际飞行过程中的滚转特性; 文献 [14] 对后缘变弯度机翼开展了阵风载荷和机动载荷减缓风洞试验, 研究了变弯度后缘主动控制对机翼阵风载荷及机动载荷的影响。然而, 国内有关变弯度机翼飞机的机动过程研究仍处于起步阶段, 因此, 从变体飞机等新概念飞行器的设计及应用角度而言, 对变弯度机翼飞机机动过程的研究, 特别是风洞试验研究, 是非常有必要的。

本文在传统飞机机动过程研究的基础上, 通过模态构造法提出一种基于变弯度后缘的飞机刚弹耦合动力学建模方法, 并针对某小展弦比飞机缩比模型开展了理论模型建模、仿真分析及风洞试验, 分析变弯度后缘飞机的机动过程特点, 验证所设计控制律的载荷减缓效果及所建立理论模型的准确性。

1　弹性飞机滚转机动过程建模

根据弹性飞机滚转机动过程特点, 采用时域分析方法搭建弹性飞行滚转机动仿真框架, 如图 1 所示。

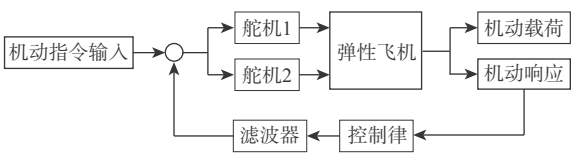


图 1　弹性飞机滚转机动仿真框架

Fig. 1　Framework of rolling maneuver simulation of elastic aircraft

1.1　弹性飞机状态空间建模

1.1.1　气动力计算

采用频域非定常气动力方法, 建立非定常机动过程的气动力模型与运动方程。由偶极子格网法可知, 当计算弹性机体和舵面对应的非定常气动力时, 气动网格控制点的下洗速度与广义坐标之间有如下关系:

$$\boldsymbol{\omega}_{\mathrm{bo}}=-\left(\frac{\partial \boldsymbol{\Phi}_{\mathrm{bo}}}{\partial x}+\mathrm{i}\frac{k}{b}\boldsymbol{\Phi}_{\mathrm{bo}}\right)\boldsymbol{q} \tag{1}$$

$$\boldsymbol{\omega}_{\mathrm{co}}=-\left(\frac{\partial \boldsymbol{\Phi}_{\mathrm{co}}}{\partial x}+\mathrm{i}\frac{k}{b}\boldsymbol{\Phi}_{\mathrm{co}}\right)\boldsymbol{\delta} \tag{2}$$

式中: $\boldsymbol{\omega}$ 为气动网格控制点的下洗速度; $\boldsymbol{\Phi}$ 为模态矩阵; 下标 bo 表示机体, co 表示舵面; $\boldsymbol{q}$ 为机体广义坐标; $\boldsymbol{\delta}$ 为舵面广义坐标; k 为减缩频率; b 为参考长度; x 为位置。

根据广义气动力的定义, 频域下其矩阵表达形式为

$$\boldsymbol{P}=\boldsymbol{\Phi}_P^{\mathrm{T}}\boldsymbol{S}\Delta\boldsymbol{p} \tag{3}$$

式中: $\boldsymbol{\Phi}_P$ 为气动力作用点处的模态矩阵; $\boldsymbol{S}=\mathrm{diag}(\Delta S_1,\Delta S_2,\cdots,\Delta S_n)$ 为面积加权矩阵, 其对角项表示对应气动网格面积; $\Delta\boldsymbol{p}$ 为压强。

将式 (1) 和式 (2) 代入式 (3) 中, 可得广义非定常气动力的表达式为

$$\boldsymbol{P}=\frac{1}{2}\rho V^2(\boldsymbol{Q}_{\mathrm{bo}}\boldsymbol{q}+\boldsymbol{Q}_{\mathrm{co}}\boldsymbol{\delta}) \tag{4}$$

式中: ρ 为空气密度; V 为来流速度; $\boldsymbol{Q}_{\mathrm{bo}}$ 和 $\boldsymbol{Q}_{\mathrm{co}}$ 分别为弹性机体和舵面对应的广义非定常气动力系数矩阵。

为后续便于进行时域分析与仿真, 对偶极子格网法计算得到的频域下的气动力, 采用最小状态 (minimum state, MS) 法进行有理函数拟合, 将频域下的广义气动力延拓到拉氏域:

$$\bar{\boldsymbol{Q}}(h)=\boldsymbol{A}_0+\boldsymbol{A}_1h+\boldsymbol{A}_2h^2+\boldsymbol{D}(\boldsymbol{I}h-\boldsymbol{R})^{-1}\boldsymbol{E}h \tag{5}$$

式中: $\bar{\boldsymbol{Q}}(h)$为拉氏域下的气功力, $h=sb/V$, s 为拉氏变量; $\boldsymbol{A}_0$、$\boldsymbol{A}_1$、$\boldsymbol{A}_2$、$\boldsymbol{D}$、$\boldsymbol{E}$ 分别为有理函数拟合得到的常系数矩阵; $\boldsymbol{I}$ 为单位矩阵; $\boldsymbol{R}=-\mathrm{diag}(\gamma_1,\gamma_2,\cdots,\gamma_{n_r})$ 为滞后根对角矩阵。

式(5)中的常系数矩阵 $\boldsymbol{A}_n(n=0,1,2)$ 和 $\boldsymbol{E}$ 可按照机体和舵面写成分块矩阵形式:

$$\begin{cases}\boldsymbol{A}_n=\begin{bmatrix}\boldsymbol{A}_{\mathrm{ss}n} & \boldsymbol{A}_{\mathrm{sc}n}\end{bmatrix}\\ \boldsymbol{E}=\begin{bmatrix}\boldsymbol{E}_{\mathrm{s}} & \boldsymbol{E}_{\mathrm{c}}\end{bmatrix}\end{cases}\quad n=0,1,2 \tag{6}$$

式中: 下标 ss 表示机体相关项, sc 表示机体与舵面耦合项, s 表示机体相关项, c 表示舵面相关项。

引入气动力滞后项:

$$\boldsymbol{x}_{\mathrm{a}}=(\boldsymbol{I}h-\boldsymbol{R})^{-1}\boldsymbol{E}h\left[\boldsymbol{\xi}(s)+\boldsymbol{\delta}(s)\right] \tag{7}$$

式中: $\boldsymbol{\xi}$ 表示机体广义坐标。

1.1.2　弹性飞机拉氏域方程

自由-自由弹性飞机广义坐标系下气动弹性运动方程形式为

$$\boldsymbol{M}s^2\boldsymbol{\xi}(s)+\boldsymbol{C}s\boldsymbol{\xi}(s)+\boldsymbol{K}\boldsymbol{\xi}(s)=\boldsymbol{P}_{\mathrm{a}}(s)+\boldsymbol{P}_0 \tag{8}$$

式中: $\boldsymbol{M}$、$\boldsymbol{C}$、$\boldsymbol{K}$ 分别为广义质量、阻尼、刚度矩阵; $\boldsymbol{P}_{\mathrm{a}}$ 为广义气动力; $\boldsymbol{P}_0$ 为与重力和配平气动力有关项。

将式 (8) 写成刚弹耦合的矩阵形式:

$$\begin{bmatrix}\boldsymbol{M}_{\mathrm{ss}}s^2 & \boldsymbol{M}_{\mathrm{sc}}s^2\end{bmatrix}\begin{bmatrix}\boldsymbol{\xi}(s)\\ \boldsymbol{\delta}(s)\end{bmatrix}+\begin{bmatrix}\boldsymbol{C}_{\mathrm{ss}}s & \boldsymbol{0}\end{bmatrix}\begin{bmatrix}\boldsymbol{\xi}(s)\\ \boldsymbol{\delta}(s)\end{bmatrix}+\begin{bmatrix}\boldsymbol{K} & \boldsymbol{0}\end{bmatrix}\begin{bmatrix}\boldsymbol{\xi}(s)\\ \boldsymbol{\delta}(s)\end{bmatrix}=q_\infty\boldsymbol{Q}(s)\begin{bmatrix}\boldsymbol{\xi}(s)\\ \boldsymbol{\delta}(s)\end{bmatrix}+\boldsymbol{P}_0 \tag{9}$$

式中: 机体结构的广义质量、阻尼、刚度矩阵均为对角矩阵, 广义质量矩阵刚体模态部分对应的对角元素分别为模型的质量和绕 3 个坐标轴的转动惯量, 弹性模态部分对应的对角元素为 1, 广义阻尼矩

阵对角元素分别为该阶模态的阻尼系数、圆频率与广义质量之积的2倍，广义刚度矩阵对角元素分别为该阶模态的圆频率平方乘其广义质量；机体和舵面结构模态的耦合质量矩阵，为机体模态矩阵、机体质量矩阵与舵面模态矩阵之积；$\boldsymbol{Q}(s)$为与结构和舵面模态有关的广义气动力影响系数（aerodynamic influence coefficient, AIC）矩阵；q_∞为当地动压。

1.1.3 弹性飞机状态空间方程

将式(5)～式(7)代入式(9)，整理可得拉氏域广义坐标下的飞机气动弹性运动方程：

$$\left(\bar{\boldsymbol{M}}_{\mathrm{s}}s^2-\bar{\boldsymbol{C}}_{\mathrm{s}}s-\bar{\boldsymbol{K}}_{\mathrm{s}}\right)\boldsymbol{\xi}(s)=\left(\bar{\boldsymbol{M}}_{\mathrm{sc}}s^2+\bar{\boldsymbol{C}}_{\mathrm{sc}}s+\bar{\boldsymbol{K}}_{\mathrm{sc}}\right)\boldsymbol{\delta}(s)+q_\infty\boldsymbol{D}\boldsymbol{x}_{\mathrm{a}}+\boldsymbol{P}_0 \tag{10}$$

式中：$\bar{\boldsymbol{M}}_{\mathrm{s}}=\boldsymbol{M}_{\mathrm{ss}}+\frac{1}{2}\rho b^2\boldsymbol{A}_{\mathrm{ss2}}$；$\bar{\boldsymbol{C}}_{\mathrm{s}}=\frac{1}{2}\rho bV\boldsymbol{A}_{\mathrm{ss1}}-\boldsymbol{C}_{\mathrm{ss}}$；$\bar{\boldsymbol{K}}_{\mathrm{s}}=\frac{1}{2}\rho V^2\boldsymbol{A}_{\mathrm{ss0}}-\boldsymbol{K}_{\mathrm{ss}}$；$\bar{\boldsymbol{M}}_{\mathrm{sc}}=\frac{1}{2}\rho b^2\boldsymbol{A}_{\mathrm{sc2}}-\boldsymbol{M}_{\mathrm{sc}}$；$\bar{\boldsymbol{C}}_{\mathrm{sc}}=\frac{1}{2}\rho bV\boldsymbol{A}_{\mathrm{sc1}}$；$\bar{\boldsymbol{K}}_{\mathrm{sc}}=\frac{1}{2}\rho V^2\boldsymbol{A}_{\mathrm{sc0}}$。

由于飞机滚转机动仿真过程考虑的是平衡位置，可以忽略方程中的配平项$\boldsymbol{P}_0$。对式(10)进行反拉氏变换，可以得到弹性飞机时域状态空间输入方程：

$$\dot{\boldsymbol{x}}_{\mathrm{ae}}=\boldsymbol{A}_{\mathrm{ae}}\boldsymbol{x}_{\mathrm{ae}}+\boldsymbol{B}_{\mathrm{ae}}\boldsymbol{u}_{\mathrm{ae}} \tag{11}$$

式中：

$$\boldsymbol{x}_{\mathrm{ae}}=\begin{bmatrix}\boldsymbol{\xi}\\ \dot{\boldsymbol{\xi}}\\ \boldsymbol{x}_{\mathrm{a}}\end{bmatrix}$$

$$\boldsymbol{u}_{\mathrm{ae}}=\begin{bmatrix}\boldsymbol{\delta}\\ \dot{\boldsymbol{\delta}}\\ \ddot{\boldsymbol{\delta}}\end{bmatrix}$$

$$\boldsymbol{A}_{\mathrm{ae}}=\begin{bmatrix}\boldsymbol{0} & \boldsymbol{I} & \boldsymbol{0}\\ \bar{\boldsymbol{M}}_{\mathrm{s}}^{-1}\bar{\boldsymbol{K}}_{\mathrm{s}} & \bar{\boldsymbol{M}}_{\mathrm{s}}^{-1}\bar{\boldsymbol{C}}_{\mathrm{s}} & q_\infty\bar{\boldsymbol{M}}_{\mathrm{s}}^{-1}\boldsymbol{D}\\ \boldsymbol{0} & \boldsymbol{E}_{\mathrm{s}} & \boldsymbol{R}V/b\end{bmatrix}$$

$$\boldsymbol{B}_{\mathrm{ae}}=\begin{bmatrix}\boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0}\\ \bar{\boldsymbol{M}}_{\mathrm{s}}^{-1}\bar{\boldsymbol{K}}_{\mathrm{sc}} & \bar{\boldsymbol{M}}_{\mathrm{s}}^{-1}\bar{\boldsymbol{C}}_{\mathrm{sc}} & \bar{\boldsymbol{M}}_{\mathrm{s}}^{-1}\bar{\boldsymbol{M}}_{\mathrm{sc}}\\ \boldsymbol{0} & \boldsymbol{E}_{\mathrm{c}} & \boldsymbol{0}\end{bmatrix}$$

其中：下标ae表示气动弹性系统。

对于飞机滚转机动过程，主要考虑滚转角位移、滚转角速度和翼尖加速度3种响应。以上3种响应均可通过模态法求解得到：

$$\begin{cases}\phi=\boldsymbol{\varphi}_{\mathrm{phi}}\boldsymbol{\xi}\\ p=\boldsymbol{\varphi}_{\mathrm{p}}\dot{\boldsymbol{\xi}}\\ a_{\mathrm{wingtip}}=\boldsymbol{\varphi}_{\mathrm{ac}}\ddot{\boldsymbol{\xi}}\end{cases} \tag{12}$$

式中：$\boldsymbol{\varphi}_{\mathrm{phi}}$、$\boldsymbol{\varphi}_{\mathrm{p}}$和$\boldsymbol{\varphi}_{\mathrm{ac}}$分别为角位移、角速度和加速度传感器处的模态矩阵。

对于飞机滚转机动过程中产生的机动载荷，可以采用模态位移（mode displacement, MD）法计算指定有限元单元的内载荷[15]：

$$\boldsymbol{F}_{\mathrm{load}}=\boldsymbol{K}_{\mathrm{a}}\boldsymbol{\varphi}_{\mathrm{load}}\boldsymbol{\xi} \tag{13}$$

式中：$\boldsymbol{K}_{\mathrm{a}}$为单元刚度矩阵；$\boldsymbol{\varphi}_{\mathrm{load}}$为载荷监测位置的模态矩阵。

联立式(12)与式(13)，整理可得弹性飞机时域状态空间输出方程：

$$\boldsymbol{y}_{\mathrm{ae}}=\boldsymbol{C}_{\mathrm{ae}}\boldsymbol{x}_{\mathrm{ae}} \tag{14}$$

式中：$\boldsymbol{y}_{\mathrm{ae}}=\begin{bmatrix}\phi\\ p\\ \boldsymbol{F}_{\mathrm{load}}\end{bmatrix}$；$\boldsymbol{C}_{\mathrm{ae}}=\begin{bmatrix}\boldsymbol{\varphi}_{\mathrm{phi}} & \boldsymbol{0} & \boldsymbol{0}\\ \boldsymbol{0} & \boldsymbol{\varphi}_{\mathrm{p}} & \boldsymbol{0}\\ \boldsymbol{K}_{\mathrm{a}}\boldsymbol{\varphi}_{\mathrm{load}} & \boldsymbol{0} & \boldsymbol{0}\end{bmatrix}$。

将式(11)与式(14)联立，可得弹性飞机时域状态空间方程：

$$\begin{cases}\dot{\boldsymbol{x}}_{\mathrm{ae}}=\boldsymbol{A}_{\mathrm{ae}}\boldsymbol{x}_{\mathrm{ae}}+\boldsymbol{B}_{\mathrm{ae}}\boldsymbol{u}_{\mathrm{ae}}\\ \boldsymbol{y}_{\mathrm{ae}}=\boldsymbol{C}_{\mathrm{ae}}\boldsymbol{x}_{\mathrm{ae}}\end{cases} \tag{15}$$

1.2 舵系统传递函数

舵系统主要由舵机、舵面及两者的连接组件组成，在仿真过程中，为模拟舵系统的动态特性，一般采用三阶传递函数环节模拟舵系统动态特性，其动力学模型传递函数可以表示为

$$\frac{\delta_{\mathrm{a}}(s)}{u_{\mathrm{aer}}(s)}=\frac{a_4}{s^3+a_1s^2+a_2s+a_3} \tag{16}$$

式中：$u_{\mathrm{aer}}(s)$为舵机的输入电压；a_n为传递函数系数；$\delta_{\mathrm{a}}(s)$为舵面的等效偏角。

1.3 控制环节

为便于风洞试验的实施及试验结果与仿真的对比分析，采用PID方法对飞机滚转机动过程进行控制，其原理框图如图2所示。

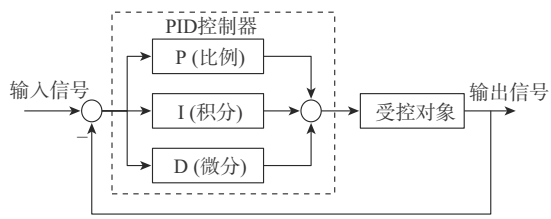


图2 PID控制器原理框图

Fig. 2 Schematic diagram of PID controller

由于真实的气动弹性系统不可避免地存在环境干扰，噪声信号往往比系统响应变化更为剧烈，而微分环节会降低系统的信噪比，不利于系统的稳定性控制。因此，在实际控制中，一般仅使用PI控制，并在反馈回路中添加低通滤波器环节过滤高频信号。另外，由于纯积分环节具有实部为0的极点，控制系统是临界稳定的，此时微小的非零均值即可导致系统失稳。因此，积分环节通常由惯性环节代替。

2　仿真分析

2.1　仿真对象

仿真对象为某飞翼布局飞机缩比模型，根据该飞机的实际结构建立简化梁架模型，如图 3 所示。该有限元模型全部单元均采用铝合金材料，弹性模量为 70 GPa，泊松比为 0.33，密度为 2 820 kg/m^3。

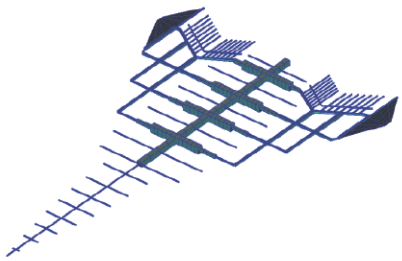


图 3　仿真对象结构有限元模型

Fig. 3　Structural finite element model of simulation object

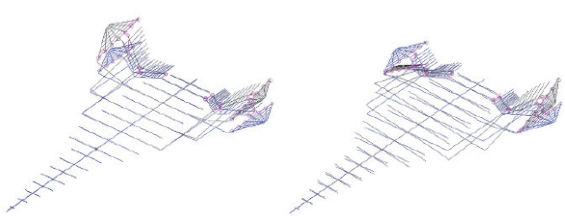
2.2　结构动力学特性

根据实际边界条件，仅放开滚转自由度，使用 MSC.NASTRAN 对全机结构有限元模型进行模态分析，转动惯量为 0.51 kg·m^2，计算得到全机结构固有模态频率，如表 1 所示，其中，典型模态阵型如图 4 所示。

表 1　全机结构动力学特性

Table 1　Structural dynamics characteristics of aircraft

序号	频率/Hz	模态描述
1	0	全机滚转
2	18.53	机翼对称一弯
3	27.64	机翼反对称一弯
4	39.51	面内模态
5	47.14	机翼对称一扭



(a) 机翼对称一弯　　(b) 机翼反对称一弯

图 4　典型模态振型

Fig. 4　Vibration forms of typical modes

2.3　连续变弯度后缘

飞机机翼后缘内侧与外侧分别布置一对连续变弯度后缘舵面，如图 5 所示，其中，变形翼肋部分采用拓扑优化方法进行设计[16]。通过舵机驱动变形翼肋变形，改变后缘弯度来调整机翼升力分布，从而操纵飞机进行滚转机动。

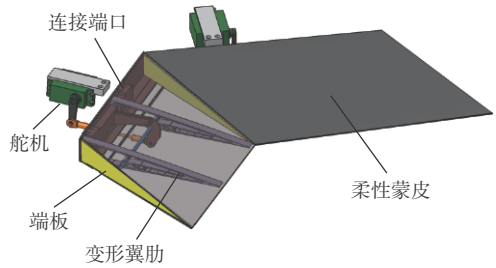


图 5　连续变弯度后缘模型

Fig. 5　Model of continuously variable-camber trailing edge

由于飞机在滚转机动过程中所需的舵面偏转角较小(<5°)，可以使用偶极子格网法并通过构造舵面偏转模态来计算舵面的气动力。

文献 [17] 中采用多项式的方式构造舵面模态：

$$y_c(x) = -\frac{\tan\beta}{(c-x_c)^{n-1}}(x-x_c)^n \qquad x\in[x_c,c];n=1,2,3,4 \tag{17}$$

式中：β 为舵面等效偏转角，即后缘点在竖直方向上位移 z_{te} 与舵面弦长 c_f 比值的反正切值，如图 6 所示；x_c 为舵面转轴位置坐标，c 为弦长；n 为舵面阶数，n=1 时为常规刚性舵面，n>1 时为连续变弯度后缘。

为更加真实地模拟连续变弯度后缘的气动特性，本文采用静力法构造变弯度后缘模态，即通过有限元方法计算变形翼肋在给定等效偏转角下的变形，如图 7 所示。

提取变形前后翼肋中弧线上点的位移构造舵面模态，并与式 (17) 的方法进行对比，结果如图 8 所示。

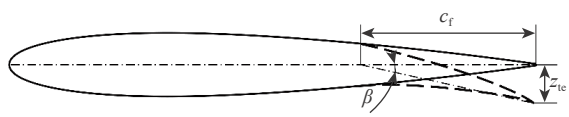


图 6　舵面等效偏转角

Fig. 6　Equivalent deflection angle of control surface

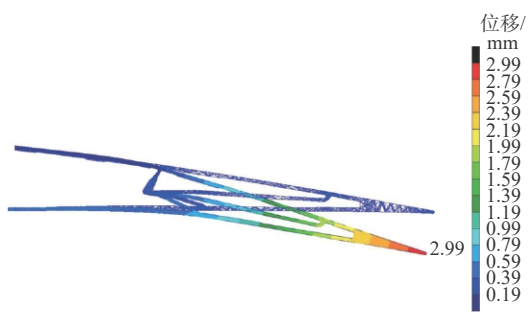


图 7　变形翼肋结构有限元模型变形分布

Fig. 7　Deformation distribution of structural finite element model of flexible rib

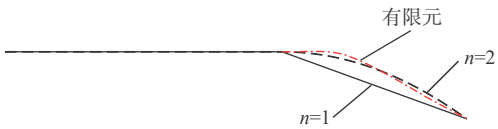
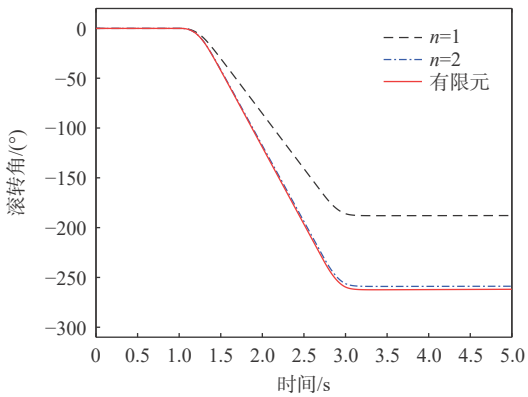


图 8 2种舵面模态构造方法对比

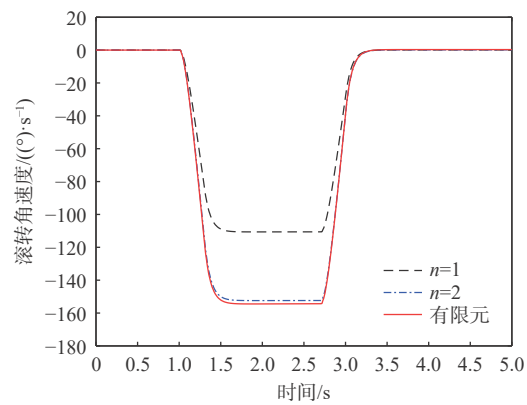
Fig. 8 Comparison of two control surface mode construction methods

2.4 弹性飞机滚转机动开环仿真

分别使用图 8 中的 3 种方法构造舵面模态，与 2.2 节中 MSC.NASTRAN 计算得到的有限元模型模态结果代入到第 1 节的理论模型中，计算给定工况下该模型的非定常气动力数据，并建立弹性飞机状态空间模型。风速 40 m/s 下给定相同的舵面输入信号，对该模型滚转机动的开环响应进行仿真，以飞机向右滚转为正方向，结果如图 9 所示。



(a) 滚转角



(b) 滚转角速度

图 9 不同舵面模态下滚转机动开环响应随时间变化曲线

Fig. 9 Variation of open-loop response of rolling maneuver with time in different control surface modes

由图 9 可知，2 种后缘连续变弯度模态构造方法（n=2 及有限元）的舵效均高于常规舵面模态（n=1），而有限元静力法的舵效略高于多项式法。说明相同面积的情况下，连续变弯度后缘的操纵效率明显高于传统刚性舵面。

2.5 弹性飞机滚转机动闭环仿真

飞机在滚转机动过程中，由于气动力和惯性力的作用，会产生一定的附加机动载荷，特别是滚转机动的开始和结束阶段，机动载荷会达到顶峰，从而对飞机结构产生较大影响。因此，为保证飞机的飞行安全，需要通过主动控制降低峰值机动载荷。

2.5.1 控制律设计

对于弹性飞机滚转机动控制律，包含滚转角保持回路和滚转机动载荷减缓回路，如图 10 所示。

以机翼外侧变弯度后缘单独变弯度作为基准控制方案；以机翼外侧和内侧变弯度后缘联合偏转外加翼尖加速度反馈回路为滚转机动载荷减缓控制方案。另外，为防止变弯度后缘偏转指令超出结构允许范围，在内外侧变弯度后缘前添加限幅环节。

2.5.2 控制律参数优化

针对基准控制方案及滚转机动载荷减缓控制方案的控制律参数，本文采用遗传优化算法进行优化设计，该优化问题的数学表达形式如下：

$$\begin{cases} \text{Find } K_{\text{p1}}, K_{\text{p2}}, K_{\text{p3}}, K_{\text{i}}, K_1, K_2 \\ \min\ M_{\text{wr}} \\ \text{s.t. } \sigma \leqslant \sigma_{\max}, t \leqslant t_{\max} \end{cases} \tag{18}$$

式中：M_{wr}为翼根处弯矩；σ为飞机滚转机动中实际滚转角较机动指令的超调量；t为飞机滚转角到达机动指令花费的时间；K_{pn}为增益参数，K_{i}为积分参数，K_n为舵面比例参数。即通过遗传优化算法确定给定机动指令下飞机最小滚转机动峰值载荷（翼根弯矩）对应的控制律参数。

2.5.3 闭环系统仿真

风速 40 m/s 下，以 1.5 s 飞机向一侧滚转 180°为机动指令信号，飞机实际滚转角较机动指令超调量为 1%，飞机滚转角到达机动指令时间为 1.75 s，采

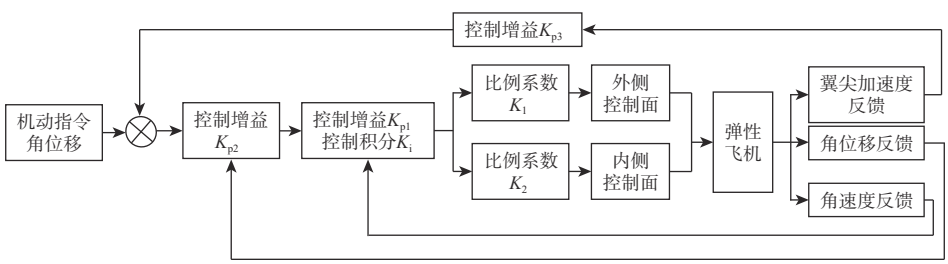


图 10 弹性飞机滚转机动控制系统架构

Fig. 10 Architecture of rolling maneuver control system of elastic aircraft

用遗传优化算法分别计算基准控制方案和滚转机动载荷减缓控制方案下相应的控制律参数, 如表 2 所示。

使用优化计算得到的控制律参数进行 2 种控制方案下的飞机滚转机动闭环仿真, 结果如图 11 所示。可知, 相比基准控制方案, 即机翼外侧变弯度后缘单独偏转, 采用滚转机动载荷减缓控制方案, 即机翼外侧和内侧变弯度后缘联合偏转, 可以有效降低飞机滚转机动过程中机翼翼根的峰值弯矩。

此外, 通过改变不同的实际滚转角较机动指令的超调量约束, 可以优化得到不同的峰值载荷减缓率, 如图 12 所示。可知, 在基准控制方案约束条件不变的情况下, 实际滚转角较机动指令的超调量越大, 机动载荷减缓控制方案较基准控制方案的峰值

表 2　控制律参数优化结果

Table 2　Parameter optimization results of control law

控制律参数	K_{p1}	K_{p2}	K_{p3}	K_i	K_1	K_2
基准控制	6.517 6	0.025 3	0	0.332 3	−1	0
滚转机动载荷减缓控制	6.389 5	0.032 1	0.097 3	0.518 9	−0.675 0	−0.091 6

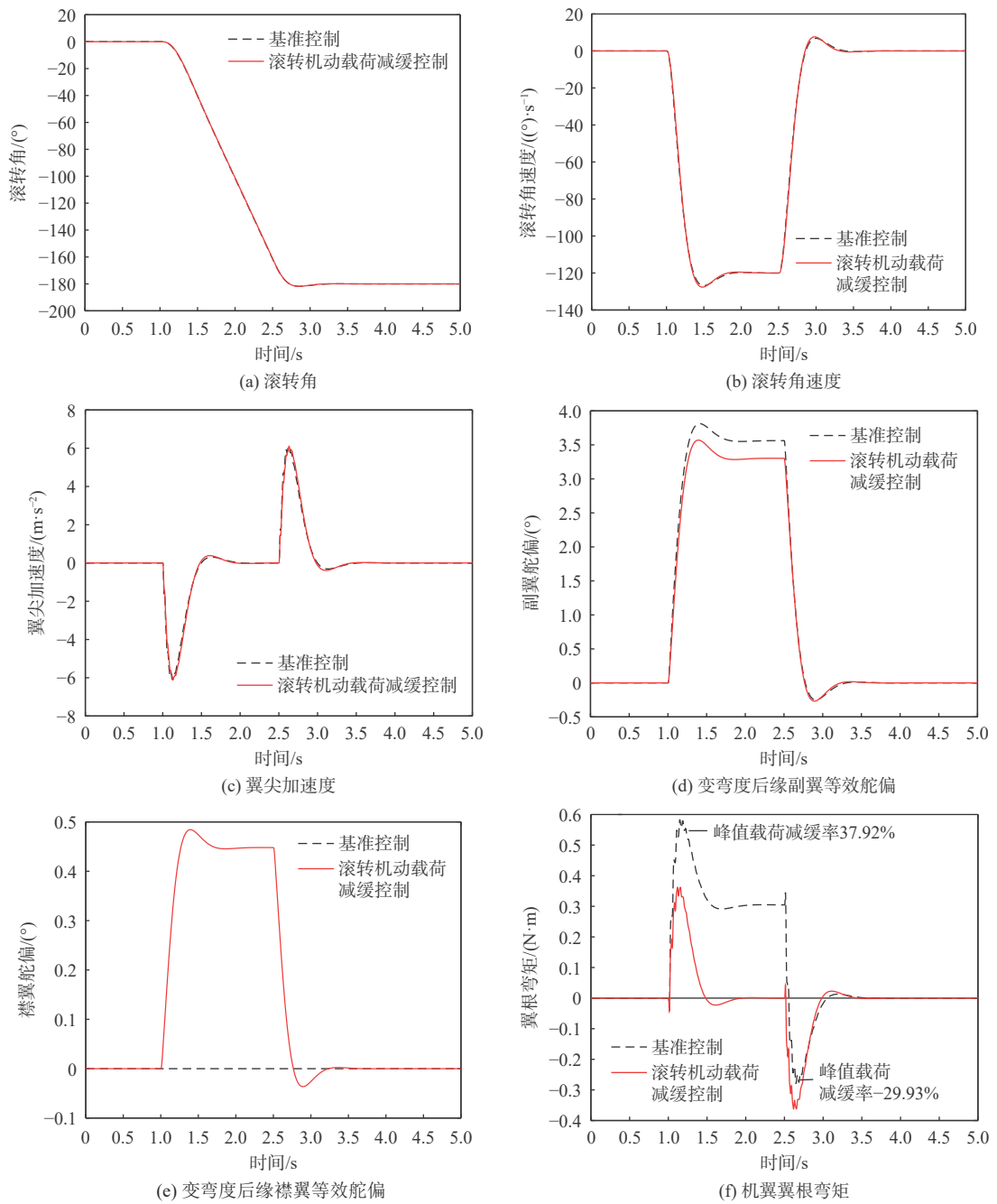


图 11　滚转机动闭环响应及载荷随时间变化曲线

Fig. 11　Variation of close-loop response of rolling maneuver and load with time

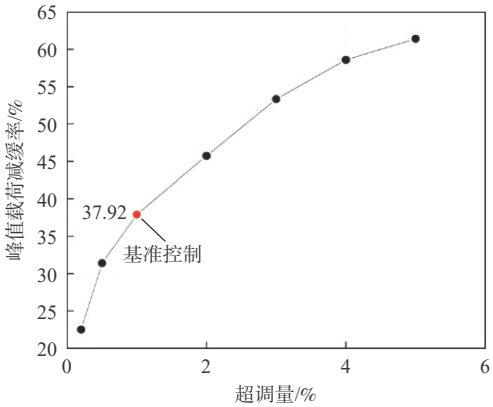


图 12 不同超调量下的峰值载荷减缓率

Fig. 12 Peak load alleviation rate under different overshoots

载荷减缓率越大。因此，可以通过适当调整滚转机动载荷减缓控制方案的约束条件来进一步降低飞机滚转机动过程中的峰值载荷，从而为飞机的结构设计提供优化空间。

3 试验结果

风洞试验模型为某飞翼布局飞机缩比模型，翼展为 1.6 m，机长为 2.3 m，其平面布局如图 13 所示。

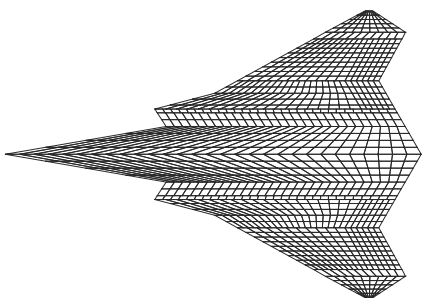


图 13 风洞试验模型平面布局

Fig. 13 Plane layout of wind tunnel test model

3.1 地面试验

为获取风洞试验模型的结构动力学特性及变弯度后缘系统的动态特性，在风洞试验前对模型进行地面模态试验及舵系统扫频试验。

3.1.1 地面模态试验

为获取风洞试验模型的结构动力学特性，以便于对仿真模型进行修正，并将仿真结果与试验数据对比，对风洞试验模型进行地面模态试验，如图 14 所示。

为与风洞试验状态保持一致，地面模态试验中，模型通过尾椎进行固定，除滚转自由度放开，其他方向自由度全部进行约束；另外，为测量模型滚转的转动惯量，分别将模型两侧的机翼翼尖与地面用已知刚度的弹簧进行固定，测得转动惯量为 1.59 kg·m^2。地面模态试验使用激振器法进行，即通过激振器对模型进行激励振动，同时测量模型上各

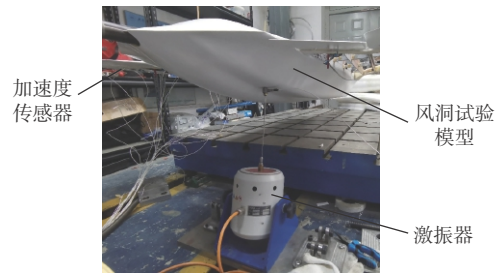


图 14 风洞试验模型地面模态试验

Fig. 14 Ground modal test of wind tunnel test model

点的加速度，从而解算出模型的结构固有频率。试验结果如表 3 所示，其中，典型模态振型如图 15 所示。

表 3 地面模态试验结果

Table 3 Results of ground modal test

序号	频率/Hz	模态描述
1	1.52	全机滚转
2	11.09	机翼对称一弯
3	15.88	机翼反对称一弯
4	64.86	机翼对称一扭



(a) 机翼对称一弯 (b) 机翼反对称一弯

图 15 地面模态试验振型

Fig. 15 Vibration forms of ground modal test

3.1.2 舵系统扫频试验

为获取舵系统的动态特性，以及舵机电信号与变弯度后缘等效偏转角的直接关系，对风洞试验模型舵系统进行扫频试验，如图 16 所示。

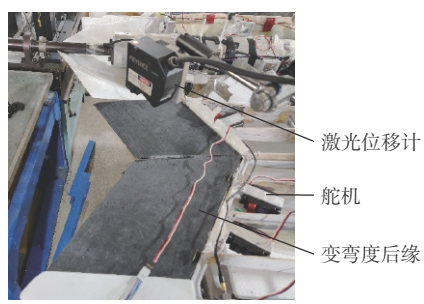


图 16 舵系统扫频试验

Fig. 16 Frequency sweep test of rudder system

舵系统扫频试验输入为步进正弦信号，起始频率为 0.1 Hz，终止频率为 4 Hz，步进频率为 0.3 Hz；输出为激光位移计测得的变弯度后缘后侧法向位移。将输入信号与输出信号进行快速傅里叶变换，并通过三阶传递函数辨识可得机翼内侧与

外侧变弯度后缘等效偏转角与舵机信号之间的传递函数：

$$G_{\text{in}}(s)=\frac{11\,880}{s^3+23.58s^2+1\,100s+12\,380} \tag{19}$$

$$G_{\text{out}}(s)=-\frac{2\,489}{s^3+8.36s^2+677s+4\,632} \tag{20}$$

3.2　风洞试验

风洞试验在中国航天空气动力技术研究院 FD-09 的 3 m×3 m 风洞进行，试验模型通过尾部锥杆与风洞固支端连接，锥杆内部布置有轴承以放开模型的滚转自由度，风洞模型安装情况如图 17 所示。



图 17　风洞中的试验模型

Fig. 17　Test model in wind tunnel

3.2.1　风洞试验系统

缩比模型滚转机动风洞试验系统主要分为试验模型、测量系统和控制系统 3 个部分，如图 18 所示。

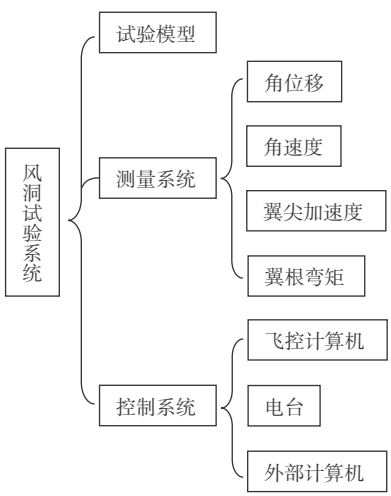


图 18　风洞试验系统框架

Fig. 18　Framework of wind tunnel test system

试验模型通过尾部的锥杆与风洞支撑杆固接；模型滚转机动风洞试验测量系统测量参数主要分为滚转机动响应与滚转机动载荷 2 部分，滚转机动响应分为滚转角位移、滚转角速度和翼尖加速度，滚转机动载荷为翼根弯矩，以上参数均通过相应传感器测量，传感器布置如图 19 所示。

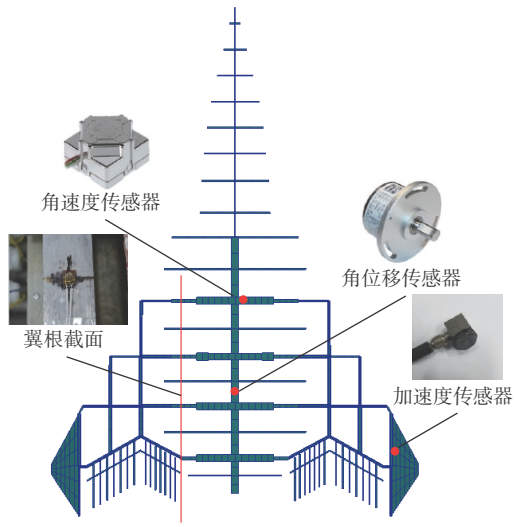


图 19　传感器布置示意图

Fig. 19　Sensor placement

翼根弯矩通过布置在机翼与机身连接梁上下表面的应变片测量，再通过解算得出翼根弯矩。风洞试验采用无线测控系统，过程中所有传感器测量得到的数据均可通过电台传输到风洞外的外部计算机中，另外控制系统的相应参数也可以通过电台反向传输到机载飞控计算机上。

3.2.2　风洞试验结果

缩比模型滚转机动风洞试验采用与 3.2.1 节中仿真分析相同的控制律，即以机翼外侧变弯度后缘单独变弯度为基准控制，机翼外侧与内侧变弯度后缘联合偏转外加翼尖加速度反馈回路为滚转机动载荷减缓控制。

以 1.5 s 向一侧滚转 180°为机动指令信号，分别在风速 35, 40, 45 m/s 下进行滚转机动基准控制及滚转机动载荷减缓控制试验，对比相同风速条件下模型机动响应及机动载荷变化情况，试验结果如图 20 所示。

从试验结果可以看出，在 35, 40, 45 m/s 这 3 种风速条件下，变弯度后缘均可有效驱动风洞模型在 2 s 内进行 180°滚转机动，且 3 种风速下模型的响应及载荷趋势基本相同；通过对比不同风速下的翼根弯矩可以看出，随着风速的增大，由于动压的增大，翼根弯矩的峰值也在增大，在 3 种风速下，滚转机动载荷减缓控制方案较基准控制方案均可以在保持近似相同机动过程的情况下有效减缓翼根的峰值弯矩，减缓率分别达到 43.01%、32.58% 和 32.46%，表明本文采用的滚转机动载荷减缓控制方案可以有效降低飞机机动过程中的峰值载荷，从而提高飞机结构强度裕度，为飞机结构设计提供优化空间。

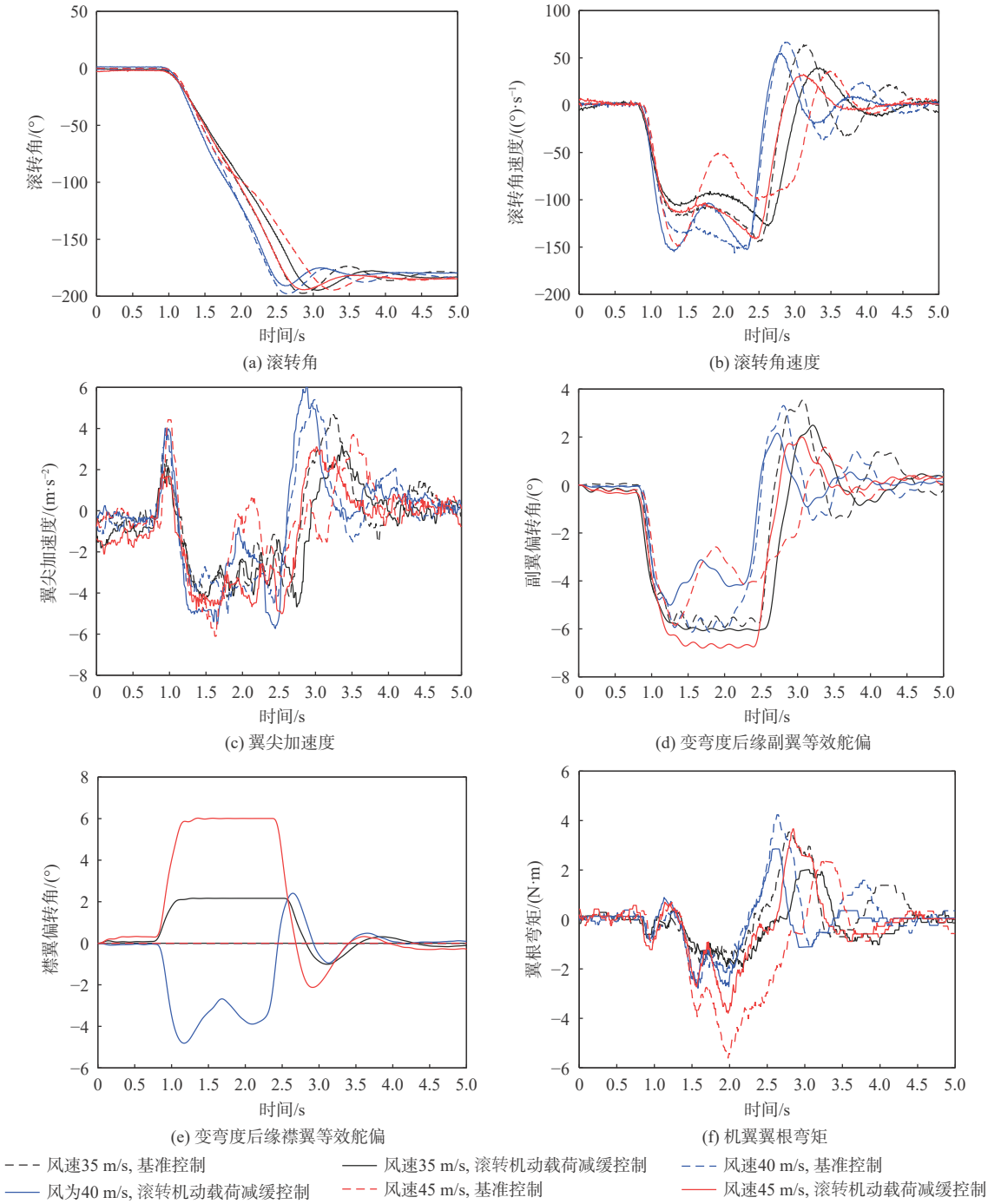


图 20 风洞试验结果

Fig. 20 Results of wind tunnel test

4 仿真与试验对比分析

4.1 仿真模型修正

在模型滚转机动风洞试验过程中，发现由于模型加工与设备安装存在一定误差，因此，风洞模型的重心位于尾部锥段，即飞机滚转机动转轴的下方，导致模型在滚转机动过程中重力会产生附加力矩，从而对滚转机动响应产生影响。

为使仿真模型更加接近实际试验模型，引入重力修正项，即

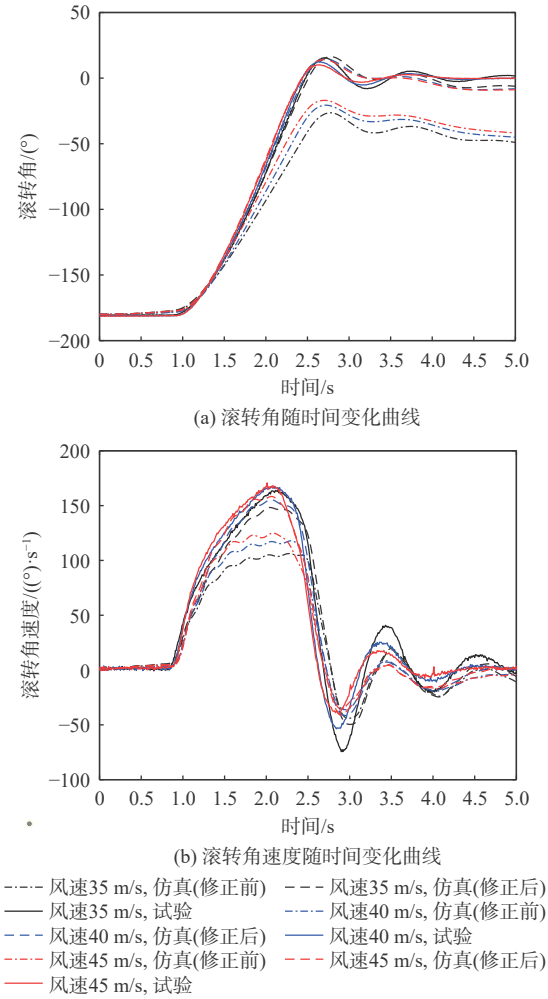
$$\boldsymbol{P}_0 = mgl\boldsymbol{T}\sin\boldsymbol{\xi}(s) \tag{21}$$

式中：m 为模型质量；g 为重力加速度；$\boldsymbol{T}$ 为转换矩阵；l 为模型重心与滚转机动转轴的竖直距离。将式(21)代入式(10)中，即可得到修正后的仿真模型。

4.2 仿真分析与试验数据对比

将模型滚转机动风洞试验得到的变弯度后缘等效偏转角数据分别代入修正前与修正后的仿真模型，进行仿真计算，并与风洞试验测量得到的机动响应进行对比，结果如图21所示。

由图21可知，在35, 40, 45 m/s这3种风速下，



(a) 滚转角随时间变化曲线

(b) 滚转角速度随时间变化曲线

图 21　风洞试验数据与仿真结果对比

Fig. 21　Comparison of wind tunnel test data and simulation results

由于机动指令相同, 模型的滚转机动响应变化趋势基本相同, 但由于模型加工和设备安装难以避免带来系统误差, 产生重力附加力矩, 因此, 3 种不同风速下风洞试验测得的飞机滚转角与滚转角速度均大于修正前理论模型的仿真结果, 且修正前的理论模型由于不考虑重心偏离转轴的影响, 3 种风速下的响应结果在数值上存在一定差异; 通过在理论模型中添加重力附加力矩进行修正, 3 种不同风速下的滚转角仿真结果与风洞试验数据的最大误差(滚转角差值与指令之比)分别为 5.49%、4.60% 和 3.85%, 吻合性均较好, 说明修正后的飞机滚转机动理论模型准确性较高, 可以对实际模型的滚转机动响应进行较好的预测。

5　结　论

1) 缩比模型滚转机动风洞试验结果表明, 连续变弯度后缘能够实现光滑连续变形, 其最大等效偏转角速度达到 47(°)/s, 可以操纵模型在 2 s 内进行 180°滚转机动。

2) 缩比模型滚转机动仿真和风洞试验结果表明, 本文采用的滚转机动载荷减缓控制律相较基准控制律可以降低 30% 以上的滚转机动附加载荷, 从而可以为结构优化提供设计空间。

3) 本文提出一种用于连续变弯度后缘飞机的滚转机动建模分析方法, 仿真结果与风洞试验数据的最大误差小于 6%, 表明本文方法能够较好地预测应用连续变弯度后缘飞机的实际滚转机动响应, 对连续变弯度后缘飞机的设计及仿真分析具有一定参考价值。

4) 针对风洞试验缩比模型出现的加工与设备安装误差, 本文提出一种用于重力偏轴的修正方法, 对比分析表明, 该方法可以有效解决试验模型误差对仿真的影响, 提高理论模型的精度。

致谢　感谢杨昊霖和张海波在本文风洞试验开展过程中提供的帮助。

参考文献（References）

[1] 方振平, 陈万春, 张曙光. 航空飞行器飞行动力学[M]. 北京: 北京航空航天大学出版社, 2005: 92-134.
FANG Z P, CHEN W C, ZHANG S G. Flight dynamics of aircraft[M]. Beijing: Beihang University Press, 2005: 92-134(in Chinese).

[2] WHITE R J. Improving the airplane efficiency by use of wing maneuver load alleviation[J]. Journal of Aircraft, 1971, 8(10): 769-775.

[3] WOODS-VEDELER J A, POTOTZKY A S, HOADLEY S T. Rolling maneuver load alleviation using active controls[J]. Journal of Aircraft, 1995, 32(1): 68-76.

[4] 张建刚, 何康乐, 金鑫. 飞机方案设计阶段机动载荷快速计算方法研究[J]. 力学与实践, 2020, 42(6): 726-730.
ZHANG J G, HE K L, JIN X. A fast calculation method of maneuvering load in aircraft scheme design stage[J]. Mechanics in Engineering, 2020, 42(6): 726-730(in Chinese).

[5] YIN H W, WU Z G, YANG C. Design and analysis of a wind tunnel test model system for rolling maneuver load alleviation of flying wings[C]//Proceedings of the 56th AIAA/ASCE/AHS/ASC Structures, Structural Dynamics, and Materials Conference. Reston: AIAA, 2015: 1860.

[6] 秦航远, 吴志刚, 杨超, 等. 滚转机动载荷减缓风洞试验[J]. 北京航空航天大学学报, 2016, 42(9): 2008-2016.
QIN H Y, WU Z G, YANG C, et al. Wind tunnel test of rolling maneuver load alleviation[J]. Journal of Beijing University of Aeronautics and Astronautics, 2016, 42(9): 2008-2016(in Chinese).

[7] CHU L L, LI Q, GU F, et al. Design, modeling, and control of morphing aircraft: A review[J]. Chinese Journal of Aeronautics, 2022, 35(5): 220-246.

[8] GOMEZ J C, GARCIA E. Morphing unmanned aerial vehicles[J]. Smart Material Structures, 2011, 20(10): 103001.

[9] HETRICK J, OSBORN R, KOTA S, et al. Flight testing of mission adaptive compliant wing[C]//Proceedings of the 48th AIAA/ASME/

ASCE/AHS/ASC Structures, Structural Dynamics, and Materials Conference. Reston: AIAA, 2007: 1709.

[10] DECAMP R, HARDY R. Mission adaptive wing advanced research concepts[C]//Proceedings of the 11th Atmospheric Flight Mechanics Conference. Reston: AIAA, 1984: 2088.

[11] GILBERT W. Development of a mission adaptive wing system for a tactical aircraft[C]//Proceedings of the Aircraft Systems Meeting. Reston: AIAA, 1980: 1886.

[12] SATTI R, LI Y B, SHOCK R, et al. Computational aeroacoustic analysis of a high-lift configuration[C]//Proceedings of the 46th AIAA Aerospace Sciences Meeting and Exhibit. Reston: AIAA, 2008: 34.

[13] CHANZY Q, KEANE A J. Analysis and experimental validation of morphing UAV wings[J]. The Aeronautical Journal, 2018, 122(1249): 390-408.

[14] WANG X R, MKHOYAN T, MKHOYAN I, et al. Seamless active morphing wing simultaneous gust and maneuver load alleviation [J]. Journal of Guidance, Control, and Dynamics, 2021, 44(9): 1649-1662.

[15] 杨超, 吴志刚, 万志强, 等. 飞行器气动弹性原理[M]. 北京: 北京航空航天大学出版社, 2011: 70-80.
YANG C, WU Z G, WAN Z Q, et al. Aeroelasticity principle of aircraft[M]. Beijing: Beihang University Press, 2011: 70-80(in Chinese).

[16] 张桢锴, 贾思嘉, 宋晨, 等. 柔性变弯度后缘机翼的风洞试验模型优化设计[J]. 航空学报, 2022, 43(3): 226071.
ZHANG Z K, JIA S J, SONG C, et al. Optimum design of wind tunnel test model for compliant morphing trailing edge[J]. Acta Aeronautica et Astronautica Sinica, 2022, 43(3): 226071(in Chinese).

[17] FORSTER E, SANDERS B, EASTEP F. Modelling and sensitivity analysis of a variable geometry trailing edge control surface[C]//Proceedings of the 44th AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics, and Materials Conference. Reston: AIAA, 2003: 1807.

Rolling maneuver load alleviation of aircraft with continuously variable-camber trailing edge

LEI Chaohui[1], YANG Chao[1], SONG Chen[1, 2, *], JIN Tianyi[3], WU Zhigang[1]

(1. School of Aeronautic Science and Engineering, Beihang University, Beijing 100191, China;
2. Institute of Unmanned System, Beihang University, Beijing 100191, China;
3. Shanghai Electro-Mechanical Engineering Institute, Shanghai 201109, China)

Abstract: The variable-camber trailing edge is frequently used in new concept aircraft design because of its advantages of continuous deformation, less resistance, and low aerodynamic noise. A rigid-elastic coupling dynamics modeling method based on the variable-camber trailing edge was proposed, and rolling maneuver simulation analysis and wind tunnel test were carried out using a scale model of aircraft with the variable-camber trailing edge. The results show that the variable-camber trailing edge can maneuver the aircraft to roll 180° within 2 s. Additional rolling maneuver load can be reduced by more than 30% through cooperative inner and outer trailing edge deformation compared to outer trailing edge deformation alone. In addition, a comparative study shows that the error between rolling maneuver simulation results and wind tunnel test data is less than 6%, which confirms the accuracy of the established theoretical model.

Keywords: variable-camber trailing edge; rigid-elastic coupling; rolling maneuver; load alleviation; simulation analysis; wind tunnel test

Received: 2022-09-14; **Accepted**: 2022-12-19; **Published Online**: 2023-01-10 16:01
URL: link.cnki.net/urlid/11.2625.V.20230109.1359.005
Foundation item: National Natural Science Foundation of China (11402013)
* **Corresponding author**. E-mail: songchen@buaa.edu.cn

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2022.0778

点接触共轭渐开线蜗杆副啮合理论及性能分析

任雯[1]，李杰[1,2]，王天羽[1,2]，李海涛[1,2,*]

(1. 中国农业大学 工学院，北京 100083；2. 中国农业大学 机械工程训练中心，北京 100083)

摘 要：点接触共轭渐开线蜗杆传动可以降低蜗杆副对制造、装配等误差的敏感性，并能保证传动精度。基于微分几何和齿轮啮合原理，构建了含有制造和安装误差的点接触共轭渐开线蜗杆传动数学模型，研究了螺旋角、模数、压力角等主要设计参数对啮合性能的影响规律，并进行了传动副对制造和安装误差敏感性的研究。通过仿真实验验证了接触区的分布情况。研究结果表明，含有制造和安装误差时，传动副的瞬时传动比恒定；选取适中的螺旋角、较小的模数及较小的压力角会增大接触区的面积，提高传动副的重合度；传动副对制造和安装误差均不敏感，但应尽量减小或避免制造和安装轴交角造成的误差。通过仿真分析得到接触区和重合度与理论计算结果一致，验证了所提理论的正确性。

关 键 词：点接触；共轭；渐开线蜗杆；啮合性能；误差分析

中图分类号：TH132.4

文献标志码：A **文章编号**：1001-5965(2024)10-3183-13

圆柱蜗杆传动具有传动比大、传动平稳性好、结构紧凑等优点，是一种可传递交错空间的运动和动力的传动，其符合减速器轻量化的发展趋势，在各种机械设备和仪表中都有广泛应用[1-3]。渐开线蜗杆传动是圆柱蜗杆传动的一种，蜗杆齿面为渐开螺旋面，相当于一个齿数等于蜗杆头数的渐开线斜齿轮[4]，其传动精度好、效率高，是圆柱蜗杆传动中较理想的传动[5]。蜗轮由蜗轮滚刀滚切而成，其基本蜗杆为渐开线蜗杆，但滚切出的蜗轮齿形精度无法保证。另外，渐开线蜗杆传动的形式为线接触，由于存在制造误差和安装误差，传动中的瞬时接触线只在理论上存在，需要经过长时间的跑合才能形成线接触，以及实现蜗轮蜗杆齿形的精确啮合，从而保证啮合质量[5-7]。

相关研究表明，将蜗杆传动的线接触改为点接触，对降低传动副对误差的敏感性、减少跑合时间从而提高生产效率有显著效果[8-10]。Simon[11-13]提出了一种非失配和失配的受载齿面接触计算机仿真方法，采用超大滚刀加工蜗轮，降低了蜗杆传动误差的敏感性。秦大同等[14-15]分别分析了失配点啮合环面蜗杆传动、锥面环面蜗杆与直廓环面蜗轮失配啮合传动的基本原理；Meng 等[16-17]和 Chi 等[18]分别建立了失配点接触包络锥蜗杆、失配点接触ZC1蜗杆传动和失配的环面蜗杆和直齿圆柱齿轮传动的啮合理论，以上理论均能实现良好的点接触。赵超飞和魏冰阳[19]提出了一种直廓环面蜗杆-圆柱斜齿轮啮合传动形式，齿面接触形式为失配的点接触。综上所述，目前多采用失配方法实现蜗杆副点接触啮合，虽然该方法可实现齿面的连续点啮合，但只能保证两齿面在参考点满足共轭条件，除参考点外，其他接触点的传动比与设计的传动比不一致，存在传动误差，因此传动精度无法保证[20]。

Ren 等[21]提出一种双产形面展成点接触共轭环面蜗杆传动，其瞬时传动比恒定且不存在传动误差，双产形面均为渐开螺旋面，由于渐开螺旋面的特性形成了点接触共轭齿面，其形成点接触共轭齿面的方法为实现齿面点接触共轭提供了思路。基于该方法，本文建立了含有制造和安装误差的点接

收稿日期：2022-09-14；录用日期：2022-11-04；网络出版时间：2023-01-10 13:50
网络出版地址：link.cnki.net/urlid/11.2625.V.20230109.1305.004
基金项目：直升机传动技术国家重点实验室基金 (HTL-O-19G02)
*通信作者. E-mail：h.li@cau.edu.cn

引用格式：任雯，李杰，王天羽，等. 点接触共轭渐开线蜗杆副啮合理论及性能分析[J]. 北京航空航天大学学报，2024，50(10)：3183-3195.
REN W，LI J，WANG T Y，et al. Meshing theory and performance analysis of point-contact conjugate involute worm gear pair[J]. Journal of Beijing University of Aeronautics and Astronautics，2024，50(10)：3183-3195 (in Chinese).

触共轭渐开线蜗杆传动数学模型，其中，产形面采用渐开螺旋面，在此基础上，研究主要涉及参数、误差对啮合性能的影响规律。基于三维模型开展有限元分析，将传动副啮合的理论分析结果与仿真分析结果对比，验证理论的正确性。

1 点接触共轭渐开线蜗杆副啮合理论

建立右旋直角坐标系，如图1所示，产形轮4齿面与斜齿轮一致，双自由度展成的实体为渐开线蜗杆1。取静坐标系 $\sigma_{04}(O_4;\boldsymbol{i}_{04},\boldsymbol{j}_{04},\boldsymbol{k}_{04})$ 和 $\sigma_{01}(O_1;\boldsymbol{i}_{01},\boldsymbol{j}_{01},\boldsymbol{k}_{01})$ 分别与机架固连，$\boldsymbol{k}_{04}$ 和 $\boldsymbol{k}_{01}$ 分别与产形轮4和渐开线蜗杆1的轴线重合。取动坐标系 $\sigma_4(O_4;\boldsymbol{i}_4,\boldsymbol{j}_4,\boldsymbol{k}_4)$ 和 $\sigma_1(O_1;\boldsymbol{i}_1,\boldsymbol{j}_1,\boldsymbol{k}_1)$ 与产形轮4和渐开线蜗杆1固连，$\boldsymbol{k}_4=\boldsymbol{k}_{04}$，$\boldsymbol{k}_1=\boldsymbol{k}_{01}$。$O_4$ 和 O_1 为坐标原点，切齿啮合中心距 $a_{14}=|O_1O_4|+\Delta a_{14}$，$\Delta a_{14}$ 为中心距误差，轴交角 $\Sigma_{14}=90°+\Delta\Sigma_{14}$，$\Delta\Sigma_{14}$ 为轴交角误差。当产形轮4转角为 φ_4 时，渐开线蜗杆的转角为 $\varphi_1=i_{14}\varphi_4+s_1/p_1$，其中，$i_{14}$ 为产形轮4和渐开线蜗杆的传动比，s_1/p_1 为渐开线蜗杆轴向移动 s_1 需要的附加转角，通过右手定则确定对应角速度矢量 $\boldsymbol{\omega}_4$ 和 $\boldsymbol{\omega}_1$。此时，产形轮4双自由度展成的实体为渐开线圆柱蜗杆。

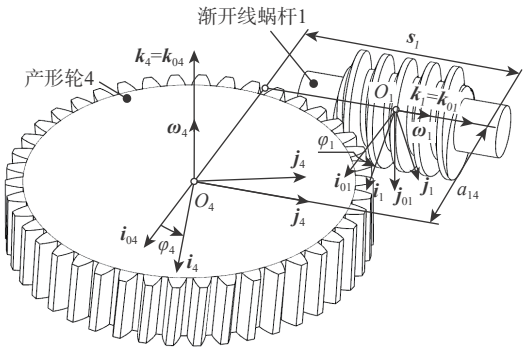


图1 双自由度展成渐开线蜗杆坐标系

Fig. 1 Coordinate system of generation of involute worm with two degrees of freedom

如图2所示，产形轮3螺旋面也是一个渐开线蜗杆螺旋面（可以与渐开线蜗杆1螺旋面一致或不一致），单自由度展成的实体为蜗轮2。取静坐标系 $\sigma_{03}(O_3;\boldsymbol{i}_{03},\boldsymbol{j}_{03},\boldsymbol{k}_{03})$ 和 $\sigma_{02}(O_2;\boldsymbol{i}_{02},\boldsymbol{j}_{02},\boldsymbol{k}_{02})$ 分别与机架固连，$\boldsymbol{k}_{03}$ 和 $\boldsymbol{k}_{02}$ 分别与产形轮3和蜗轮2的轴线重合，取动坐标系 $\sigma_3(O_3;\boldsymbol{i}_3,\boldsymbol{j}_3,\boldsymbol{k}_3)$ 和 $\sigma_2(O_2;\boldsymbol{i}_2,\boldsymbol{j}_2,\boldsymbol{k}_2)$ 与产形轮3和蜗轮2固连，$\boldsymbol{k}_3=\boldsymbol{k}_{03}$，$\boldsymbol{k}_2=\boldsymbol{k}_{02}$。$O_3$ 和 O_2 为坐标原点，切齿啮合中心距 $a_{23}=|O_2O_3|+\Delta a_{23}$，$\Delta a_{23}$ 为中心距误差，轴交角 $\Sigma_{23}=90°+\Delta\Sigma_{23}$，$\Delta\Sigma_{23}$ 为轴交角误差。当产形轮3转角为 φ_3 时，蜗轮的转角为 $\varphi_2=i_{23}\varphi_3$，其中，$i_{23}$ 为产形轮3展成蜗轮2的传动比，对应的角速度矢量为 $\boldsymbol{\omega}_3$ 和 $\boldsymbol{\omega}_2$。

产形轮3和产形轮4的下标用 j 表示，蜗杆1和蜗轮2的下标用 i 表示，当 j=4时，i=1；当 j=3时，i=2。根据微分几何向量的回转，$\sigma_j\to\sigma_i$ 的坐标变换矩阵分别为

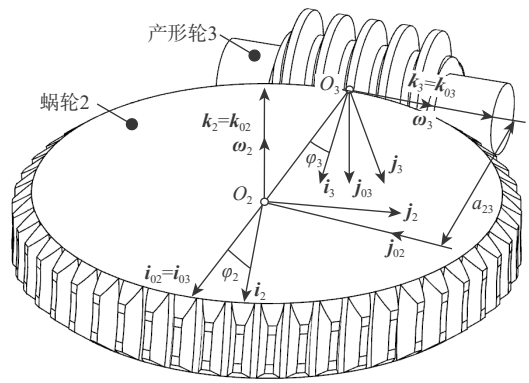


图2 单自由度展成蜗轮坐标系

Fig. 2 Coordinate system of generation of worm gear with single degree of freedom

$$\boldsymbol{R}\left[\boldsymbol{k}_j,\varphi_j\right]=\begin{bmatrix}\cos\varphi_j & -\sin\varphi_j & 0\\ \sin\varphi_j & \cos\varphi_j & 0\\ 0 & 0 & 1\end{bmatrix}$$

$$\boldsymbol{R}\left[\boldsymbol{i}_{0i},\Sigma_{ij}\right]=\begin{bmatrix}1 & 0 & 0\\ 0 & \cos\Sigma_{ij} & (-1)^{j-1}\sin\Sigma_{ij}\\ 0 & (-1)^{j}\sin\Sigma_{ij} & \cos\Sigma_{ij}\end{bmatrix}$$

$$\boldsymbol{R}\left[\boldsymbol{k}_i,-\varphi_i\right]=\begin{bmatrix}\cos\varphi_i & \sin\varphi_i & 0\\ -\sin\varphi_i & \cos\varphi_i & 0\\ 0 & 0 & 1\end{bmatrix}$$

1.1 由产形面 Σ_4 双自由度展成渐开线蜗杆螺旋面 Σ_1

1.1.1 渐开线蜗杆螺旋面 Σ_1 的方程

利用圆矢量函数和球矢量函数[20]得到坐标系 σ_4 中产形面 Σ_4 的方程为

$$(\boldsymbol{r}_4)_4=r_{b4}\boldsymbol{e}(\theta_4)-r_{b4}\theta_4\boldsymbol{g}(\theta_4)+u_4\boldsymbol{m}\left(90^\circ+\theta_4,\beta_{b4}\right)\tag{1}$$

式中：$\boldsymbol{e}(\theta_4)$、$\boldsymbol{g}(\theta_4)$ 为均圆矢量函数；$\boldsymbol{m}(\theta,\beta_4)$ 为球矢量函数；θ_4、u_4 为产形面 Σ_4 的参数；r_{b4} 为基圆半径。

进行 $\sigma_4\to\sigma_{04}$ 坐标变换，在 σ_{04} 中，Σ_4 的方程为

$$\begin{aligned}(\boldsymbol{r}_4)=&\,r_{b4}\boldsymbol{e}(\theta_4+\varphi_4)-r_{b4}\theta_4\boldsymbol{g}(\theta_4+\varphi_4)+\\&u_4\boldsymbol{m}(90^\circ+\theta_4+\varphi_4,\beta_{b4})\end{aligned}\tag{2}$$

在 σ_1 中，渐开线蜗杆螺旋面 Σ_1 的方程为

$$\begin{cases}(\boldsymbol{r}_1)_1=\boldsymbol{R}[\boldsymbol{k}_{01},-\varphi_1]\{\boldsymbol{R}[\boldsymbol{i}_{01},\Sigma_{14}](\boldsymbol{r}_4)_{04}+a_{14}\boldsymbol{i}_{04}+s_1\boldsymbol{j}_{04}\}\\ u_4=\sin\beta_{b4}\left[\dfrac{r_{b4}\cos(\theta_4+\varphi_4)+a_{14}}{\sin(\theta_4+\varphi_4)}+r_{b4}\theta_4\right]+\\ \qquad\dfrac{a_{14}\cos\beta_{b4}}{\tan\Sigma_{14}\sin(\theta_4+\varphi_4)}+\dfrac{-r_{b4}\cos\beta_{b4}[1-i_{14}\cos\Sigma_{14}]}{i_{14}\sin\Sigma_{14}\sin(\theta_4+\varphi_4)}\end{cases}\tag{3}$$

Σ_1 上任一点处的法矢量在 σ_1 中的方程为

$$(\boldsymbol{n}_1)_1=\boldsymbol{R}[\boldsymbol{k}_1,-\varphi_1]\boldsymbol{R}[\boldsymbol{i}_{01},\Sigma_{14}]\boldsymbol{R}\left[\boldsymbol{k}_4,\varphi_4\right](\boldsymbol{n}_4)_4\tag{4}$$

式中：$(\boldsymbol{n}_4)_4$ 为 Σ_4 在 σ_4 中法矢量的方程，且 $(\boldsymbol{n}_4)_4=-\boldsymbol{n}(90°+\theta_4,\beta_{b4})$。在 σ_{04} 中，法矢量为 $(\boldsymbol{n}_4)_{04}=-\boldsymbol{n}(90°+\theta_4+\varphi_4,\beta_{b4})$。

1.1.2 产形轮4与渐开线蜗杆的相对角速度和相对速度

产形轮4与渐开线蜗杆的相对运动包括转角 φ_4 和轴向移动 s_1。相对角速度矢量为

$$(\boldsymbol{\omega}_{41})_{04}=(\boldsymbol{\omega}_{41\varphi})_{04}\frac{\mathrm{d}\varphi_4}{\mathrm{d}t}+(\boldsymbol{\omega}_{41s})_{04}\frac{\mathrm{d}s_1}{\mathrm{d}t} \tag{5}$$

$$(\boldsymbol{\omega}_{41\varphi})_{04}\frac{\mathrm{d}\varphi_4}{\mathrm{d}t}=-i_{14}(\sin\Sigma_{14}\boldsymbol{j}_{04}+\cos\Sigma_{14}\boldsymbol{k}_{04})+\boldsymbol{k}_{04} \tag{6}$$

$$(\boldsymbol{\omega}_{41s})_{04}\frac{\mathrm{d}s_1}{\mathrm{d}t}=-\frac{1}{p_1}(\sin\Sigma_{14}\boldsymbol{j}_{04}+\cos\Sigma_{14}\boldsymbol{k}_{04}) \tag{7}$$

相对速度矢量为

$$(\boldsymbol{v}_{41})_{04}=(\boldsymbol{v}_{41\varphi})_{04}\frac{\mathrm{d}\varphi_4}{\mathrm{d}t}+(\boldsymbol{v}_{41s})_{04}\frac{\mathrm{d}s_1}{\mathrm{d}t} \tag{8}$$

$$\begin{aligned}(\boldsymbol{v}_{41\varphi})_{04}\frac{\mathrm{d}\varphi_4}{\mathrm{d}t}=&-i_{14}(\sin\Sigma_{14}\boldsymbol{j}_{04}+\cos\Sigma_{14}\boldsymbol{k}_{04})(\boldsymbol{r}_4)_{04}+\\&\boldsymbol{k}_{04}(\boldsymbol{r}_4)_{04}-i_{14}a_{14}(\sin\Sigma_{14}\boldsymbol{j}_{04}+\cos\Sigma_{14}\boldsymbol{k}_{04})\boldsymbol{i}_{04}\end{aligned} \tag{9}$$

$$\begin{aligned}(\boldsymbol{v}_{41s})_{04}\frac{\mathrm{d}s_1}{\mathrm{d}t}=&-\frac{1}{p_1}(\sin\Sigma_{14}\boldsymbol{j}_{04}+\cos\Sigma_{14}\boldsymbol{k}_{04})(\boldsymbol{r}_4)_{04}-\\&\frac{a_{14}}{p_1}(\sin\Sigma_{14}\boldsymbol{j}_{04}+\cos\Sigma_{14}\boldsymbol{k}_{04})\boldsymbol{i}_{04}-\\&(\sin\Sigma_{14}\boldsymbol{j}_{04}+\cos\Sigma_{14}\boldsymbol{k}_{04})\end{aligned} \tag{10}$$

1.1.3　产形轮 4 展成渐开线蜗杆的共轭条件函数和共轭条件方程

关于 φ 的共轭条件函数:

$$\begin{aligned}\Phi_{41\varphi}=&-i_{14}[(\sin\Sigma_{14}\boldsymbol{j}_{04}+\cos\Sigma_{14}\boldsymbol{k}_{04}),(\boldsymbol{r}_4)_{04},(\boldsymbol{n}_4)_{04}]-\\&i_{14}a_{14}[(\sin\Sigma_{14}\boldsymbol{j}_{04}+\cos\Sigma_{14}\boldsymbol{k}_{04}),\boldsymbol{i}_{04},(\boldsymbol{n}_4)_{04}]-\\&r_{\mathrm{b}4}\cos\beta_{\mathrm{b}4}\end{aligned} \tag{11}$$

关于 s 的共轭条件函数:

$$\begin{aligned}\Phi_{41s}=&\frac{1}{i_{14}p_1}(\Phi_{41\varphi}+r_{\mathrm{b}4}\cos\beta_{\mathrm{b}4})-\\&(\sin\Sigma_{14}\boldsymbol{j}_{04}+\cos\Sigma_{14}\boldsymbol{k}_{04})(\boldsymbol{n}_4)_{04}\end{aligned} \tag{12}$$

在双自由度展成过程中, 接触点满足共轭条件方程 $\Phi_{41\varphi}=0$ 和 $\Phi_{41s}=0$。将 $\Phi_{41\varphi}=0$ 代入式 (12), 化简后可得到接触点参数满足:

$$\theta_4+\varphi_4=\alpha_{\mathrm{t}4}+\pi \tag{13}$$

式中: $\alpha_{\mathrm{t}4}$ 为产形轮 4 分度圆端面压力角。

1.1.4　渐开螺旋面 Σ_1 曲率参数

取直母线方向和齿廓方向为 Σ_4 的主方向, 表示为 $\boldsymbol{\alpha}_\xi$ 和 $\boldsymbol{\alpha}_\eta$, 在 σ_4 中用微分几何公式, 可得

$$(\boldsymbol{\alpha}_\xi)_4=\boldsymbol{m}[90°+\theta_4,\beta_{\mathrm{b}4}] \tag{14}$$

$$(\boldsymbol{\alpha}_\eta)_4=-\boldsymbol{e}(\theta_4) \tag{15}$$

Σ_4 沿 $\boldsymbol{\alpha}_\xi$ 方向的法曲率和短程挠率及沿 $\boldsymbol{\alpha}_\eta$ 方向的法曲率分别为

$$\begin{cases}\kappa_{4\xi}=0\\ \tau_{4\xi}=0\\ \kappa_{4\eta}=\dfrac{\cos\beta_{\mathrm{b}4}}{u_4\sin\beta_{\mathrm{b}4}-r_{\mathrm{b}4}\theta_4}\end{cases} \tag{16}$$

根据 Σ_1 的方程, 按文献 [22] 中求解由双自由度展成齿面曲率的方法, 求得 Σ_1 沿 $\boldsymbol{\alpha}_\xi$ 方向的诱导法曲率、诱导短程挠率 $\kappa_{41\xi}$, $\tau_{41\xi}$ 及沿 $\boldsymbol{\alpha}_\eta$ 方向的诱导法曲率 $\kappa_{41\eta}$。由此, 可求出齿面 Σ_1 沿 $\boldsymbol{\alpha}_\xi$ 方向的法曲率和短程挠率及沿 $\boldsymbol{\alpha}_\eta$ 方向的法曲率:

$$\begin{cases}\kappa_{1\xi}=\kappa_{4\xi}-\kappa_{41\xi}\\ \tau_{1\xi}=\tau_{4\xi}-\tau_{41\xi}\\ \kappa_{1\eta}=\kappa_{4\eta}-\kappa_{41\eta}\end{cases} \tag{17}$$

1.2　由产形面 Σ_3 单自由度展成蜗轮齿面 Σ_2

1.2.1　蜗轮齿面 Σ_2 的方程

坐标系 σ_3 中, 产形面 Σ_3 的方程为

$$(\boldsymbol{r}_3)_3=r_{\mathrm{b}3}\boldsymbol{e}(-\theta_3)+r_{\mathrm{b}3}\theta_3\boldsymbol{g}(\theta_3)-u_3\boldsymbol{m}(90^\circ-\theta_3,\ \beta_{\mathrm{b}3}) \tag{18}$$

式中: θ_3 和 u_3 为产形面 Σ_4 的参数; $r_{\mathrm{b}3}$ 为基圆半径。

进行 $\sigma_3\rightarrow\sigma_{03}$ 坐标变换, σ_{03} 中 Σ_3 的方程为

$$\begin{aligned}(\boldsymbol{r}_3)_{03}=&r_{\mathrm{b}3}\boldsymbol{e}(\varphi_3-\theta_3)+r_{\mathrm{b}3}\theta_3\boldsymbol{g}(\varphi_3-\theta_3)-\\&u_3\boldsymbol{m}(90^\circ+\varphi_3-\theta_3,\ \beta_{\mathrm{b}3})\end{aligned} \tag{19}$$

在 σ_2 中, 蜗轮齿面 Σ_2 的方程为

$$\begin{cases}(\boldsymbol{r}_2)_2=\boldsymbol{R}[\boldsymbol{k}_{02},-\varphi_2]\{\boldsymbol{R}[\boldsymbol{i}_{02},\Sigma_{23}](\boldsymbol{r}_3)_{03}+a_{23}\boldsymbol{i}_{03}\}\\ u_3=\sin\beta_{\mathrm{b}3}\left[\dfrac{r_{\mathrm{b}3}\cos(\varphi_3-\theta_3)-a_{23}}{\sin(\varphi_3-\theta_3)}+r_{\mathrm{b}3}\theta_3\right]+\\ \qquad\dfrac{a_{23}\cos\beta_{\mathrm{b}3}}{\tan\Sigma_{23}\sin(\varphi_3-\theta_3)}+\dfrac{r_{\mathrm{b}3}\cos\beta_{\mathrm{b}3}[1-i_{14}\cos\Sigma_{23}]}{i_{23}\sin\Sigma_{23}\sin(\varphi_3-\theta_3)}\end{cases} \tag{20}$$

Σ_2 上任一点处的法矢量在 σ_2 中的方程为

$$(\boldsymbol{n}_2)_2=\boldsymbol{R}[\boldsymbol{k}_2,-\varphi_2]\boldsymbol{R}[\boldsymbol{i}_{02},-\Sigma_{23}]\boldsymbol{R}\left[\boldsymbol{k}_3,\varphi_3\right](\boldsymbol{n}_3)_3 \tag{21}$$

式中: $(\boldsymbol{n}_3)_3$ 为 Σ_3 在 σ_3 中法矢量的方程, 且 $(\boldsymbol{n}_3)_3=\boldsymbol{n}(90°-\theta_3,\beta_{\mathrm{b}3})$。在 σ_{03} 中, 法矢量为 $(\boldsymbol{n}_3)_{03}=-\boldsymbol{n}(90°-\theta_3+\varphi_3,\beta_{\mathrm{b}3})$。

1.2.2　产形轮 3 与蜗轮的相对角速度和相对速度

产形轮 3 与蜗轮的相对角速度矢量和相对速度矢量分别为

$$(\boldsymbol{\omega}_{32})_{03}=i_{23}(\sin\Sigma_{23}\boldsymbol{j}_{03}-\cos\Sigma_{23}\boldsymbol{k}_{03})+\boldsymbol{k}_{03} \tag{22}$$

$$\begin{aligned}(\boldsymbol{v}_{32})_{03}=&[i_{23}(\sin\Sigma_{23}\boldsymbol{j}_{03}-\cos\Sigma_{23}\boldsymbol{k}_{03})+\boldsymbol{k}_{03}]\times(\boldsymbol{r}_3)_{03}+\\&i_{23}a_{23}(\sin\Sigma_{23}\boldsymbol{j}_{03}-\cos\Sigma_{23}\boldsymbol{k}_{03})\times\boldsymbol{i}_{03}\end{aligned} \tag{23}$$

1.2.3　产形轮 3 展成蜗轮的共轭条件函数和共轭条件方程

产形轮 3 单自由度展成蜗轮的共轭条件函数为

$$\begin{aligned}\Phi_{32}=&i_{23}[(\sin\Sigma_{23}\boldsymbol{j}_{03}-\cos\Sigma_{23}\boldsymbol{k}_{03}),(\boldsymbol{r}_3)_{03},(\boldsymbol{n}_3)_{03}]+\\&i_{23}a_{23}[(\sin\Sigma_{23}\boldsymbol{j}_{03}-\cos\Sigma_{23}\boldsymbol{k}_{03}),\boldsymbol{i}_{03},(\boldsymbol{n}_3)_{03}]+\\&r_{\mathrm{b}3}\cos\beta_{\mathrm{b}3}\end{aligned} \tag{24}$$

由 $\Phi_{32}=0$ 可得共轭条件方程。

1.2.4　齿面 Σ_2 曲率参数

取直母线方向和齿廓方向为 Σ_3 的主方向, 表示为 $\boldsymbol{\beta}_\xi$ 和 $\boldsymbol{\beta}_\eta$, 在 σ_3 中用微分几何公式, 可得

$$(\boldsymbol{\beta}_\xi)_3=-\boldsymbol{m}[90°-\theta_3,\beta_{\mathrm{b}3}] \tag{25}$$

$$(\boldsymbol{\beta}_\eta)_3=-\boldsymbol{e}(-\theta_3) \tag{26}$$

Σ_3 沿 $\boldsymbol{\beta}_\xi$ 方向的法曲率和短程挠率及沿 $\boldsymbol{\beta}_\eta$ 方向的法曲率分别为

$$\begin{cases}\kappa_{3\xi}=0\\ \tau_{3\xi}=0\\ \kappa_{3\eta}=\dfrac{\cos\beta_{b3}}{u_3\sin\beta_{b3}-r_{b3}\theta_3}\end{cases} \tag{27}$$

根据 Σ_2 的方程，按文献 [20] 中求解单自由度展成齿面曲率的方法，求得 Σ_2 沿 $\boldsymbol{\beta}_\xi$ 方向的诱导法曲率、诱导短程挠率 $\kappa_{32\xi}$、$\tau_{32\xi}$ 及沿 $\boldsymbol{\beta}_\eta$ 方向的诱导法曲率 $\kappa_{32\eta}$。由此，可求出齿面 Σ_2 沿 $\boldsymbol{\beta}_\xi$ 方向的法曲率和短程挠率及沿 $\boldsymbol{\beta}_\eta$ 方向的法曲率分别为

$$\begin{cases}\kappa_{2\xi}=\kappa_{3\xi}-\kappa_{32\xi}\\ \tau_{2\xi}=\tau_{3\xi}-\tau_{32\xi}\\ \kappa_{2\eta}=\kappa_{3\eta}-\kappa_{32\eta}\end{cases} \tag{28}$$

1.3 点接触共轭渐开线蜗杆副啮合分析

将蜗杆和蜗轮按轴线交错安装，构成点接触共轭渐开线蜗杆传动，如图 3 所示。取静坐标系 $\sigma_{01c}(O_{1c};\boldsymbol{i}_{01c},\boldsymbol{j}_{01c},\boldsymbol{k}_{01c})$ 和 $\sigma_{02c}(O_{2c};\boldsymbol{i}_{02c},\boldsymbol{j}_{02c},\boldsymbol{k}_{02c})$ 分别与机架固连，$\boldsymbol{k}_{01c}$ 和 $\boldsymbol{k}_{02c}$ 分别与蜗杆和蜗轮的轴线重合。取动坐标系 $\sigma_{1c}(O_{1c};\boldsymbol{i}_{1c},\boldsymbol{j}_{1c},\boldsymbol{k}_{1c})$ 和 $\sigma_{2c}(O_{2c};\boldsymbol{i}_{2c},\boldsymbol{j}_{2c},\boldsymbol{k}_{2c})$ 与蜗杆和蜗轮固连，$\boldsymbol{k}_{1c}=\boldsymbol{k}_{01c}$，$\boldsymbol{k}_{2c}=\boldsymbol{k}_{02c}$。$O_{1c}$ 和 O_{2c} 为坐标原点，安装中心距 $a_{12}=|O_{1c}O_{2c}|+\Delta a_{12}$，$\Delta a_{12}$ 为中心距安装误差，轴交角 $\Sigma_{12}=90°+\Delta\Sigma_{12}$，$\Delta\Sigma_{12}$ 为轴交角安装误差。蜗杆和蜗轮沿各自轴线转过 φ_{1c} 和 φ_{2c}，对应的角速度矢量为 $\boldsymbol{\omega}_{1c}$ 和 $\boldsymbol{\omega}_{2c}$，二者的啮合运动以转角为唯一的运动参数，为单自由度运动。

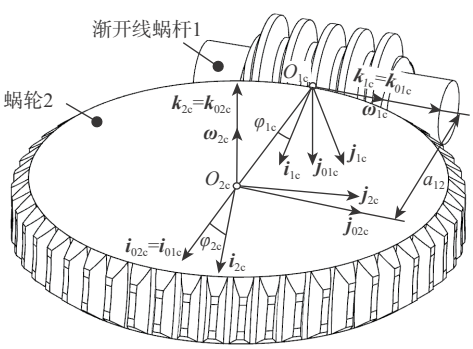


图 3 点接触共轭渐开线蜗杆副坐标系

Fig. 3 Coordinate system of point-contact conjugate involute worm gear pair

1.3.1 齿面接触点

在传动过程中，齿面 Σ_1 和 Σ_2 必须连续相切。连续相切的条件是：在任意时刻，Σ_1 和 Σ_2 的法向量在其共同接触点重合。根据齿面接触的几何条件，在 σ_{01c} 中得到接触点的欠定方程组：

$$\begin{cases}\boldsymbol{R}[\boldsymbol{i}_{01c},\Sigma_{12}]\boldsymbol{R}[\boldsymbol{k}_{02c},\varphi_{2c}](\boldsymbol{r}_2)_2=\boldsymbol{R}[\boldsymbol{k}_{01c},\varphi_{1c}](\boldsymbol{r}_1)_1+a_{12}\boldsymbol{i}_{01c}\\ \boldsymbol{R}[\boldsymbol{i}_{01c},\Sigma_{12}]\boldsymbol{R}[\boldsymbol{k}_{02c},\varphi_{2c}](\boldsymbol{n}_2)_2=\boldsymbol{R}[\boldsymbol{k}_{01c},\varphi_{1c}](\boldsymbol{n}_1)_1\end{cases} \tag{29}$$

坐标变换矩阵为

$$\boldsymbol{R}[\boldsymbol{i}_{01c},\Sigma_{12}]=\begin{bmatrix}1&0&0\\0&\cos\Sigma_{12}&-\sin\Sigma_{12}\\0&\sin\Sigma_{12}&\cos\Sigma_{12}\end{bmatrix}$$

$$\boldsymbol{R}[\boldsymbol{k}_{0ic},\varphi_{ic}]=\begin{bmatrix}\cos\varphi_{ic}&\sin\varphi_{ic}&0\\\sin\varphi_{ic}&\cos\varphi_{ic}&0\\0&0&1\end{bmatrix}\quad i=1,2$$

式 (29) 包含 5 个方程，6 个未知量 θ_4、φ_4、θ_3、φ_3、φ_{1c}、φ_{2c}，取其中 1 个为原始参数，迭代求解非线性方程组，求出其余 5 个参数。对于原始参数的每一个定值，可得到一组解，确定一个接触点。接触点的接触椭圆计算可运用文献 [20] 中的通式。

由 Σ_4 双自由度展成 Σ_1，接触点在静坐标系中组成啮合线 Γ_{41}，啮合线上的点满足式 (13)。根据文献 [21]，为求解点接触共轭渐开线蜗杆副的啮合线 Γ_{12}，将 Σ_3 单自由度展成 Σ_2 的运动增加一个虚拟自由度的运动，即蜗轮的轴向移动 s_2，得到啮合面 S_{32} 上的一条虚拟啮合线 Γ_{32}，线上的点满足：

$$\varphi_3-\theta_3=\alpha_{t3}+2\pi \tag{30}$$

式中：α_{t3} 为轮 3 分度圆端面压力角。

Γ_{41} 与 Γ_{32} 重合，根据双产形面展成共轭齿面的条件[20-21]，Γ_{41} 与 S_{32} 相交，交线即为渐开线蜗杆与蜗轮的啮合线 Γ_{12}，啮合线上的点同时满足式 (13) 和式 (30)。由于未知量 θ_4、φ_4 和 θ_3、φ_3 线性相关，未知量的个数由 6 个化简为 4 个，即 θ_4、φ_{1c}、θ_3、φ_{2c}，从而降低了求解非线性方程组的难度。

1.3.2 含制造和安装误差的瞬时传动比 i_{12e} 求解

蜗杆副的瞬时传动比为

$$i_{12e}=-\frac{[\boldsymbol{k}_{02c},(\boldsymbol{r}_2^*)_{02c},(\boldsymbol{n}_2^*)_{02c}]}{[\boldsymbol{k}_{01c},(\boldsymbol{r}_1^*)_{01c},(\boldsymbol{n}_1^*)_{01c}]}=-\frac{r_{b3}\cos\beta_{b3}}{r_{b4}\cos\beta_{b4}}\cdot\frac{i_{14}}{i_{23}} \tag{31}$$

式中：$(\boldsymbol{r}_1^*)_{01c}$ 为蜗杆齿面在 σ_{01c} 坐标系下接触点的方程；$(\boldsymbol{r}_2^*)_{02c}$ 为蜗轮齿面在 σ_{02c} 下接触点的方程；$(\boldsymbol{n}_1^*)_{01c}$ 为蜗杆齿面在 σ_{01c} 坐标系下接触点的法矢量；$(\boldsymbol{n}_2^*)_{02c}$ 为蜗轮齿面在 σ_{01c} 坐标系下接触点的法矢量。

将 r_{b3}、r_{b4}、β_{b3}、β_{b4}、i_{14}、i_{23} 代入式 (31)，化简得到瞬时传动比的值为蜗轮齿数和渐开线蜗杆头数之比，即 $i_{12e}=z_2/z_1$，因此，蜗杆副在传动中的瞬时传动恒定，不受制造和安装误差的影响，传动精度高。

蜗杆副的重合度可表示为

$$\varepsilon=\frac{z_1|\varphi_{1c\text{-}in}-\varphi_{1c\text{-}out}|}{2\pi} \tag{32}$$

式中：$\varphi_{1c\text{-}in}$ 和 $\varphi_{1c\text{-}out}$ 分别为蜗杆啮入点和啮出点的转角。

2 主要设计参数对啮合性能的影响

根据正确啮合条件，产形轮 4、渐开线蜗杆 1、

产形轮 3 及蜗轮 2 之间的参数需要满足以下条件: 产形轮 4 的螺旋角 β_4、蜗杆 1 的导程角 γ_1、产形轮 3 的导程角 γ_3 及蜗轮 2 的螺旋角 β_2 相等, 即 $\beta_4=\gamma_1=\gamma_3=\beta_2$; 产形轮 4 的端面模数 m_{t4}、蜗杆 1 的轴面模数 m_{x1}、产形轮 3 的轴面模数 m_{x3} 及蜗轮 2 的端面模数 m_{t2} 相等, 即 $m_{t4}=m_{x1}=m_{x3}=m_{t2}$。

分析主要设计参数对蜗杆副宏观和微观啮合性能的影响规律, 获得主要设计参数的设计原则, 可为点接触共轭渐开线蜗杆副的设计提供参考。影响蜗杆副啮合性能的主要设计参数为螺旋角、模数和压力角。基本几何参数具体如下: 安装中心距 a_{12}=80 mm, 蜗轮端面模数 m_{t2}=3 mm, 蜗杆头数 z_1=1, 蜗轮齿数 z_2=40, 产形轮 4 齿数 z_4=46, 产形轮 3 头数 z_3=3, 压力角 α_n=20°, 蜗杆有效长度 L=33 mm, 蜗轮齿宽 b_2=24 mm, 旋向为右旋。

改变蜗轮螺旋角 β_2、端面模数 m_{t2} 和压力角 α_n, 分析主要设计参数对齿面接触迹线、接触椭圆长半轴 a_e 及蜗杆副重合度的影响规律。

2.1　螺旋角对啮合性能的影响

以螺旋角为变量计算 4 组算例, 如表 1 所示。

表 1　螺旋角几何参数

Table 1　Geometric parameters of helix angle

算例	螺旋角β_2/(°)
算例Ⅰ-A	3.0
算例Ⅰ-B	3.6
算例Ⅰ-C	4.2
算例Ⅰ-D	4.8

传动副接触迹线在蜗轮齿面上的变化规律如图 4 所示。图中, $R_2=\sqrt{x_2^2+y_2^2}$, 实线为接触迹线。从蜗轮的啮入端到啮出端平均选取 10 个接触点, 接触时, 由于受力引起弹性变形, 接触点扩展为椭圆形接触面, 每一瞬时的椭圆接触面叠加成齿面上的接触区。

从图 4 可以看出, 随着螺旋角的增大, 接触迹线沿着蜗轮齿面轴 Z_2 负向一侧移动, 接触区发生轻微偏移。当螺旋角 β_2=4.2°时, 接触迹线位于蜗轮齿面中间位置; 当螺旋角 β_2=4.8°时, 接触椭圆短轴长度相较于其他 3 个算例明显减小。因此, 在点接触共轭渐开线蜗杆副中, 螺旋角主要影响接触区位置的分布, 即适中的螺旋角会获得较好的接触区分布。

基于传动副啮合分析模型, 进一步研究螺旋角对接触椭圆长轴 a_e 及传动副重合度的影响规律, 如图 5 所示。分析结果表明, 随着螺旋角的增大, 椭圆长轴逐渐增大, 齿面接触区变宽, 重合度略有减小。从整体来看, 螺旋角的选取应在保证接触区宽度的同时尽可能提高重合度, 因此, 算例Ⅰ-B 和Ⅰ-C 的参数可作为设计参考。

2.2　模数对啮合性能的影响

以蜗轮端面模数 m_{t2} 为变量计算 4 组算例, 其余基本几何参数均与前文相同。保持传动中心距和各轮齿数不变, 增大端面模数 m_{t2}, 蜗轮分度圆直径增大, 蜗杆分度圆直径相应减小, 模数的取值如表 2 所示。

蜗轮齿面接触迹线随模数的变化规律如图 6 所示。可以看出, 随着模数的增大, 接触迹线相对于蜗轮端面的倾角、在蜗轮齿面上的分布范围均变大, 故在点接触共轭渐开线蜗杆副中, 模数主要影响接触区的分布范围。

椭圆长轴和重合度随模数的变化如图 7 所示。分析结果表明, 随着模数的增大, 椭圆长轴逐渐减小, 齿面接触区变窄, 重合度略有增大, 且 m_{t2}=3.6 mm 时, 重合度达到最大值 1.908 9, 相较于算例Ⅱ-A~算

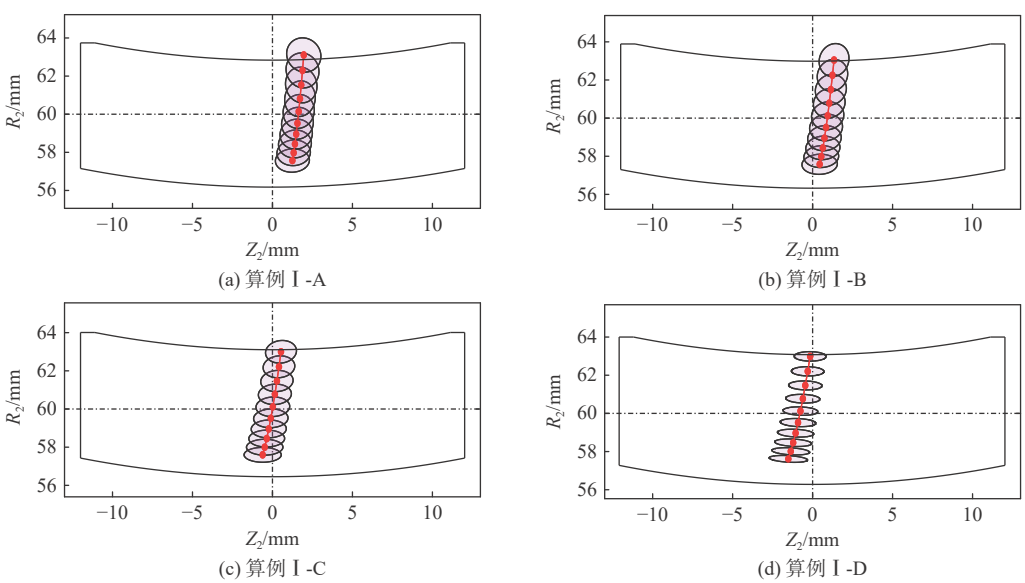
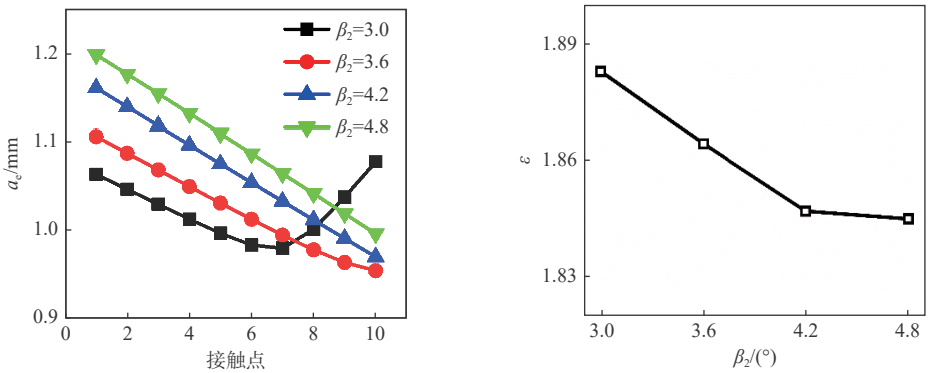


图 4　接触迹线与接触区随螺旋角的变化情况

Fig. 4　Variation of contact line and area with helix angle



(a) 螺旋角对椭圆长轴的影响 (b) 螺旋角对重合度的影响

图 5 螺旋角对椭圆长轴和重合度的影响

Fig. 5 Influence of helix angle on major axis of ellipse and contact ratio

表 2 模数几何参数

Table 2 Geometric parameters of modulus

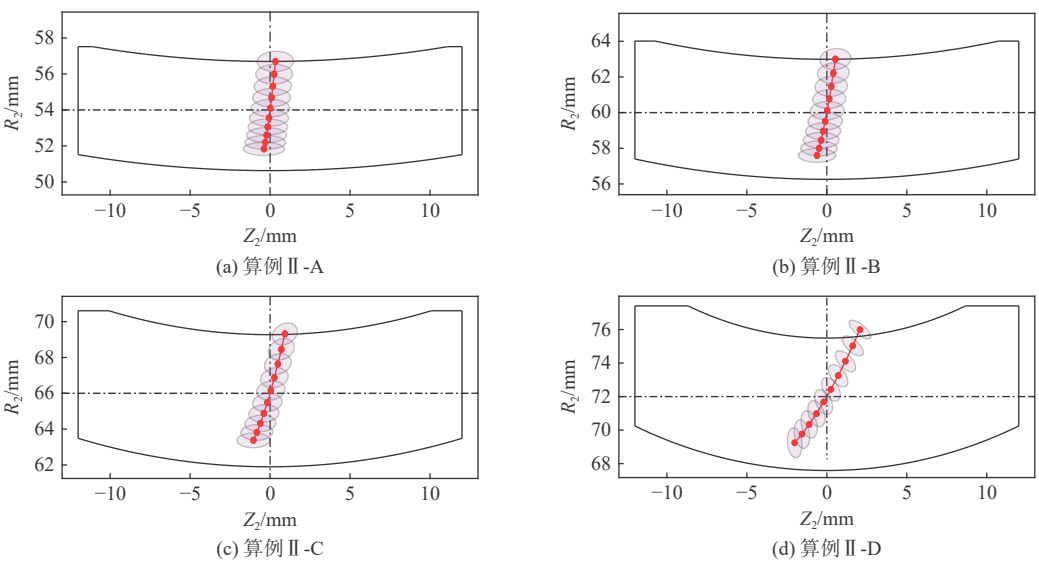
算例	端面模数m_{t2}/ mm
算例Ⅱ-A	2.7
算例Ⅱ-B	3.0
算例Ⅱ-C	3.3
算例Ⅱ-D	3.6

例Ⅱ-C, 重合度增大较多的原因是: 接触迹线相对于蜗轮端面的倾角越大, 其在蜗轮齿面的分布范围越大。

2.3 压力角对啮合性能的影响

以压力角α_n为变量计算 4 组算例, 其余基本几何参数均与前文相同。压力角的几何取值参数如表 3 所示。

蜗轮齿面接触迹线的变化规律如图 8 所示。

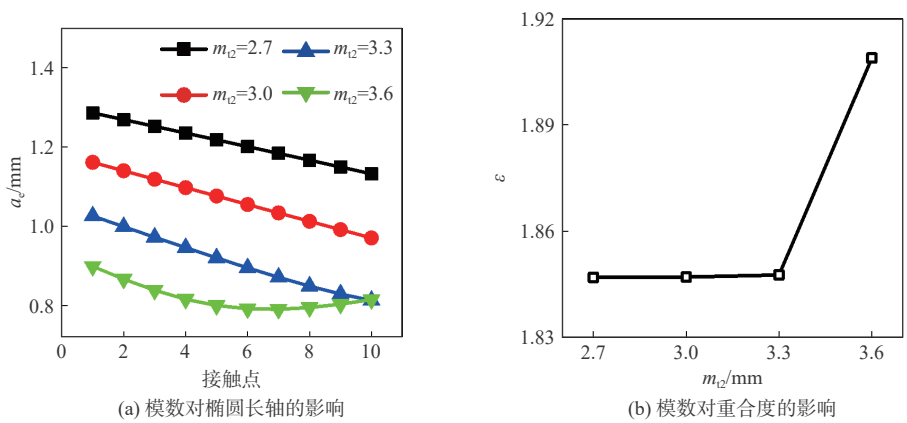


(a) 算例Ⅱ-A (b) 算例Ⅱ-B

(c) 算例Ⅱ-C (d) 算例Ⅱ-D

图 6 接触迹线与接触区随模数的变化情况

Fig. 6 Variation of contact line and area with modulus



(a) 模数对椭圆长轴的影响 (b) 模数对重合度的影响

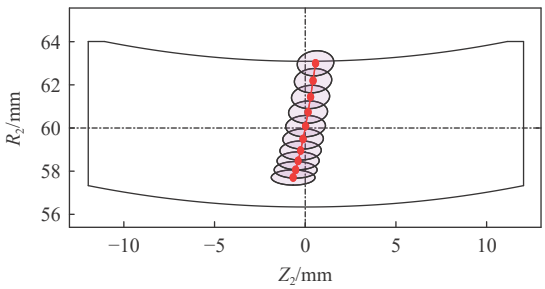
图 7 模数对椭圆长轴和重合度的影响

Fig. 7 Influence of modulus on major axis of ellipse and contact ratio

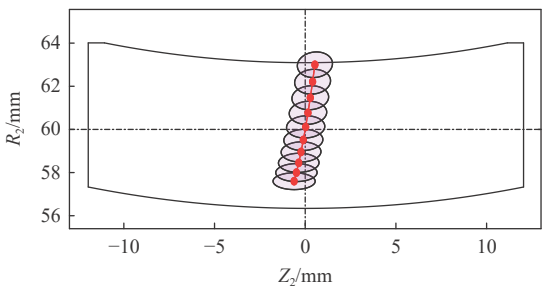
表 3　压力角几何参数

Table 3　Geometric parameters of pressure angle

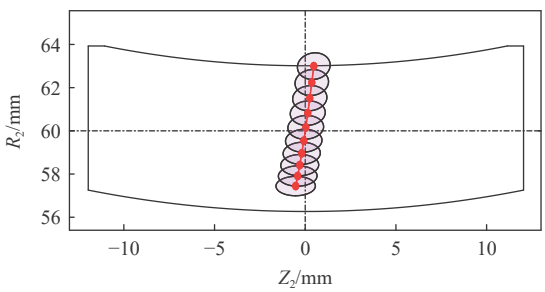
算例	压力角α_n/(°)
算例Ⅲ-A	18.5
算例Ⅲ-B	20.0
算例Ⅲ-C	21.5
算例Ⅲ-D	23.0



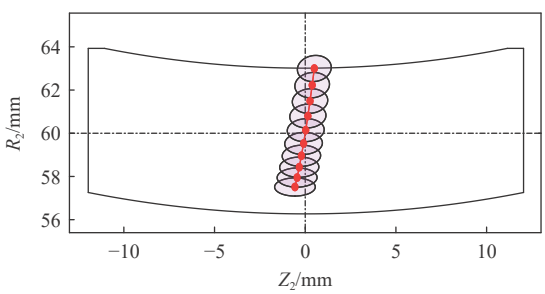
(a) 算例Ⅲ-A



(b) 算例Ⅲ-B



(c) 算例Ⅲ-C



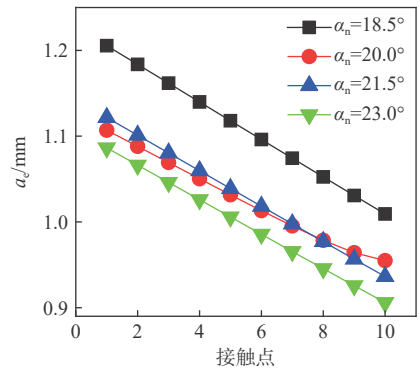
(d) 算例Ⅲ-D

图 8　接触迹线与接触区随压力角的变化情况

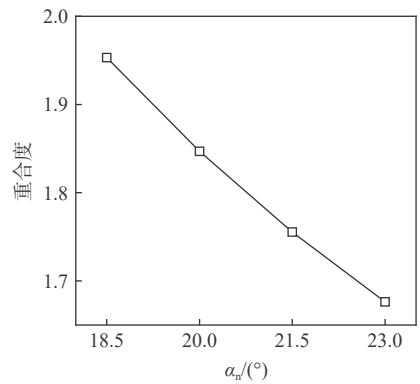
Fig. 8　Variation of contact line and area with pressure angle

可以看出, 随着压力角的增大, 接触迹线基本无变化。

椭圆长轴和重合度随压力角的变化如图 9 所示。分析结果表明, 随着压力角的增大, 椭圆长轴及传动副的重合度均有不同程度的减小。重合度降低的原因是: 给定分度圆和齿顶圆、中心距和传



(a) 压力角对椭圆长轴的影响



(b) 压力角对重合度的影响

图 9　压力角对椭圆长轴和重合度的影响

Fig. 9　Influence of pressure angle on major axis of ellipse and contact ratio

动比, 压力角与啮合角相等, 当压力角(啮合角)增大时, 基圆直径减小, 实际啮合线变短, 导致重合度降低。从整体来看, 压力角的选取应在保证接触区宽度的同时尽可能提高重合度, 因此, 压力角选取较小值。

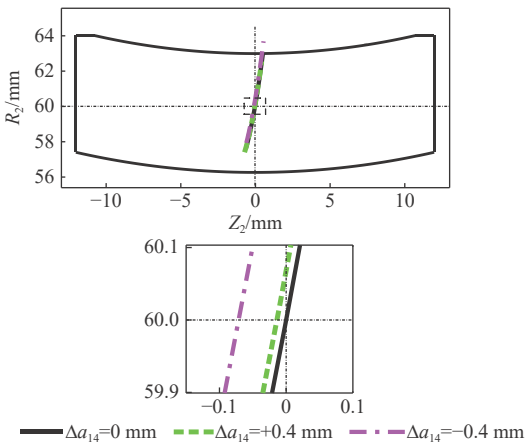
2.4　制造和安装误差对啮合性能的影响规律

根据含有误差的点接触共轭渐开线蜗杆副的数学模型, 与无误差时的啮合性能进行比较, 以研究误差对啮合性能的影响。误差包括制造误差和安装误差。制造误差为产形轮 4 展成 ZI 蜗杆 1 和产形轮 3 展成蜗轮 2 时产生的加工误差(轮 4 展成轮 1 用下标 14 表示, 轮 3 展成轮 2 用下标 23 表示), 安装误差为被展成的 ZI 蜗杆 1 和蜗轮 2 啮合时产生的装配误差(下标用 12 表示)。因此, 本文分别研究了中心距制造误差Δa_{14}、Δa_{23}, 中心距安装误差Δa_{12}, 轴交角制造误差$\Delta\Sigma_{14}$、$\Delta\Sigma_{23}$, 轴交角安装误差$\Delta\Sigma_{12}$, 产形轮沿轴向移动产生的制造误差Δc_{14}、Δc_{23}, 被展成轮沿轴向移动产生的制造误差Δb_{14}、Δb_{23}, ZI 蜗杆沿轴向移动产生的安装误差Δc_1 和蜗轮沿轴向移动产生的安装误差Δc_2 对啮合性能的影响, 误差取值参数如表 4 所示。根据 2.1 节~2.3 节的分析结果, 选取主要设计参数 β_2=4.2°、m_{t2}=3.0 mm、α_n=20°, 其余基本几何参数均与前文相同。中心距误差对蜗轮齿面接触迹线的变化规律如图 10 所示。图

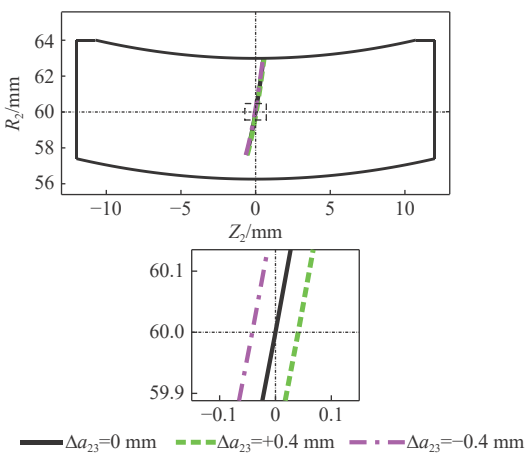
表4 误差几何参数

Table 4 Geometric parameters of errors

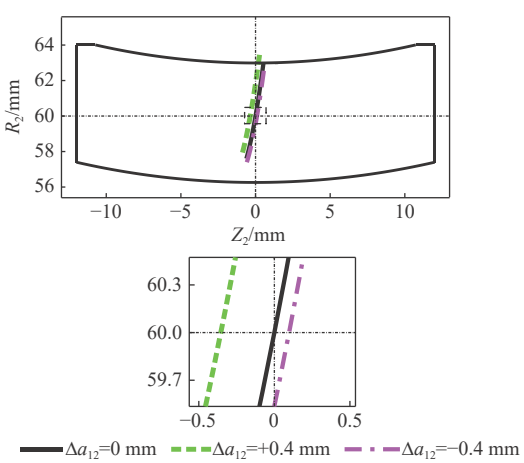
参数	误差值	参数	误差值
Δa_{14}/mm	±0.4	$\Delta\Sigma_{14}$/(°)	±0.8
Δa_{23}/mm	±0.4	$\Delta\Sigma_{23}$/(°)	±0.8
Δa_{12}/mm	±0.4	$\Delta\Sigma_{12}$/(°)	±0.8
Δb_{14}/mm	±0.4	Δc_{14}/mm	±0.4
Δb_{23}/mm	±0.4	Δc_{23}/mm	±0.4
Δc_1/mm	±0.4	Δc_2/mm	±0.4



(a) 接触迹线随中心距误差Δa_{14}变化



(b) 接触迹线随中心距误差Δa_{23}变化



(c) 接触迹线随中心距误差Δa_{12}变化

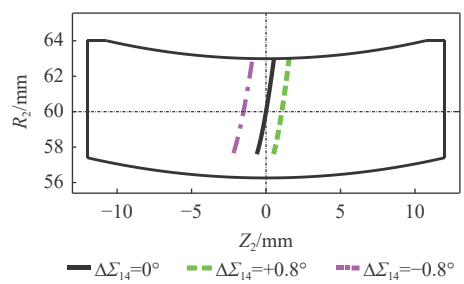
图10 接触迹线随中心距误差的变化情况

Fig. 10 Variation of contact line with center distance error

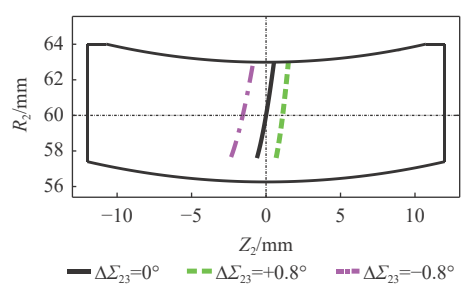
中，无误差的接触迹线用实线表示，含有正向误差的接触迹线用虚线表示，含有负向误差的接触迹线用点线表示。分析结果表明，中心距误差Δa_{14}、Δa_{23}、Δa_{12}仅会导致接触迹线的微小变化，几乎可以忽略不计。因此，该传动对中心距制造误差和中心距安装误差均不敏感。

轴交角误差对蜗轮齿面接触迹线的变化规律如图11所示。分析结果表明，轴交角误差$\Delta\Sigma_{14}$、$\Delta\Sigma_{23}$、$\Delta\Sigma_{12}$会导致接触迹线沿蜗轮轴线的偏移，相较于中心距误差，该偏移量略大，但从整体来看，接触迹线的偏移量仍在可控范围内。因此，该传动对轴交角制造误差和轴交角安装误差均不敏感。

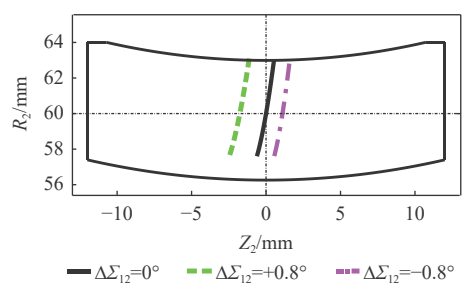
各轮沿轴向移动产生的制造误差和安装误差对蜗轮齿面接触迹线的变化规律如图12所示。分析结果表明，轴向移动产生的制造和安装误差Δb_{14}、Δb_{23}、Δc_{14}、Δc_{23}、Δc_1、Δc_2仅会导致接触迹线的微小变化，几乎可以忽略不计。因此，该传动对各轮沿轴向移动产生的制造误差和安装误差均不敏感。



(a) 接触迹线随轴交角误差$\Delta\Sigma_{14}$变化



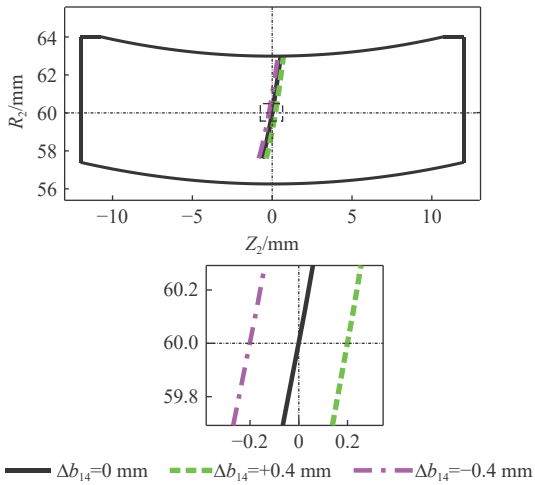
(b) 接触迹线随轴交角误差$\Delta\Sigma_{23}$变化



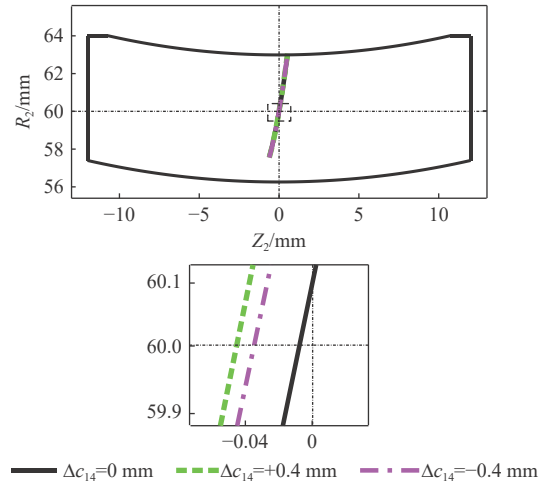
(c) 接触迹线随轴交角误差$\Delta\Sigma_{12}$变化

图11 接触迹线随轴交角误差的变化情况

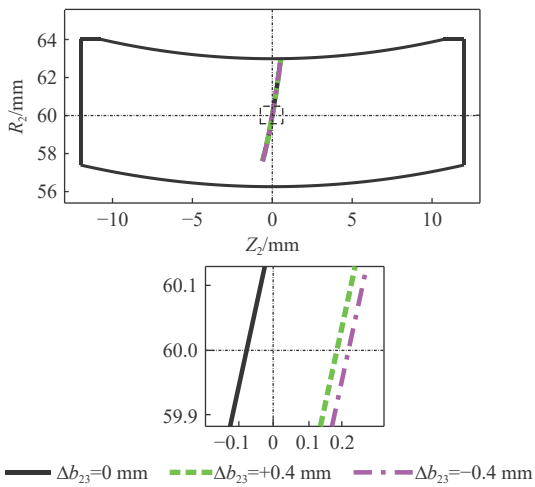
Fig. 11 Variation of contact line with crossed axis angle error



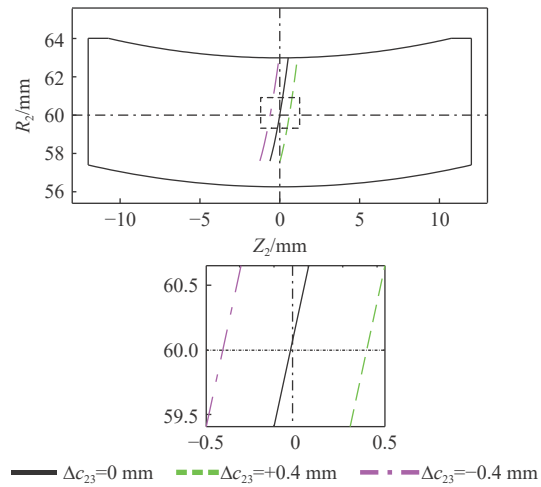
(a) 接触迹线随轴向移动误差Δb_{14}变化



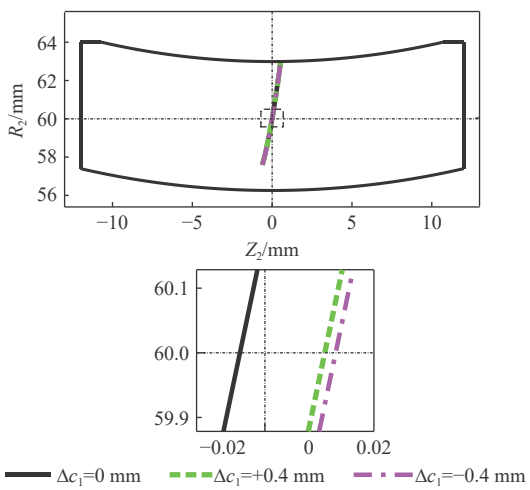
(b) 接触迹线随轴向移动误差Δc_{14}变化



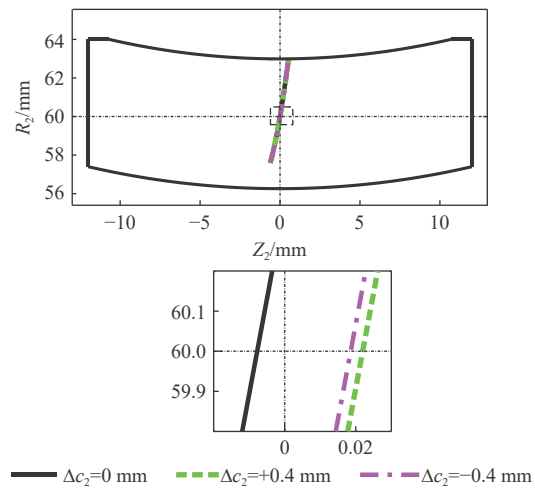
(c) 接触迹线随轴向移动误差Δb_{23}变化



(d) 接触迹线随轴向移动误差Δc_{23}变化



(e) 接触迹线随轴向移动误差Δc_1变化



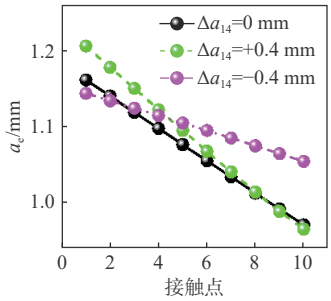
(f) 接触迹线随轴向移动误差Δc_2变化

图 12　接触迹线随轴向移动误差的变化情况

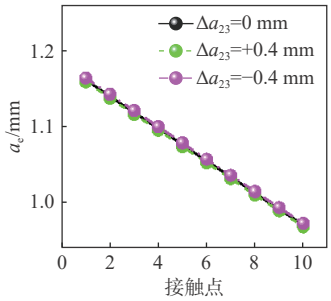
Fig. 12　Variation of contact line with axial movement error

制造误差和安装误差对接触椭圆长轴的影响规律如图 13 所示。分析结果表明, 制造误差和安装误差仅会导致接触椭圆长轴的微小变化, 几乎可以忽略不计。因此, 制造误差和安装误差对接触区的影响也几乎可以忽略。

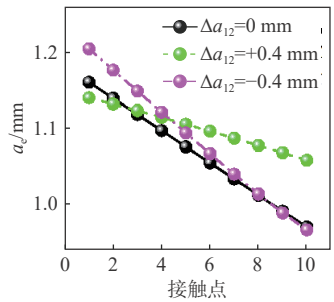
从整体来看, 相较于误差Δa_{14}、Δa_{23}、Δa_{12}、Δb_{14}、Δb_{23}、Δc_1 和Δc_1 对接触迹线的影响, 轴交角误差$\Delta\Sigma_{14}$、$\Delta\Sigma_{23}$ 和$\Delta\Sigma_{12}$ 的影响更大, 因此, 在控制制造和安装误差时, 应尽量减小或避免轴交角造成的误差。



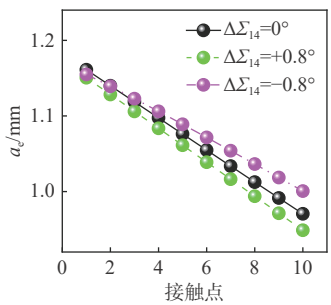
(a) 中心距误差Δa_{14}对椭圆长半轴的影响



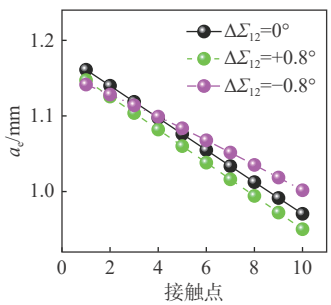
(b) 中心距误差Δa_{23}对椭圆长半轴的影响



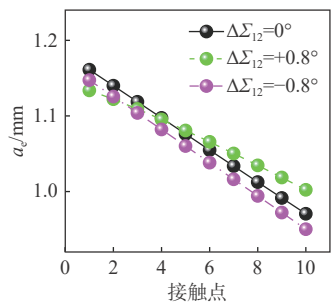
(c) 中心距误差Δa_{12}对椭圆长半轴的影响



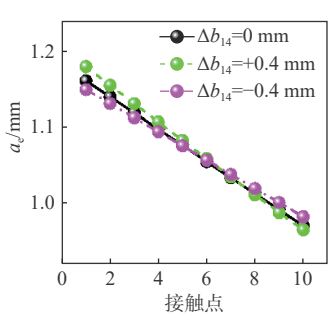
(d) 轴交角误差$\Delta\Sigma_{14}$对椭圆长半轴的影响



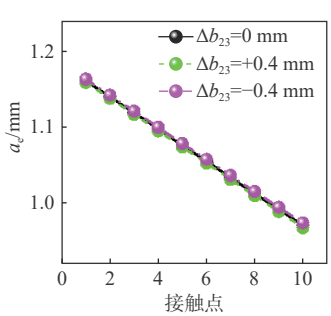
(e) 轴交角误差$\Delta\Sigma_{23}$对椭圆长半轴的影响



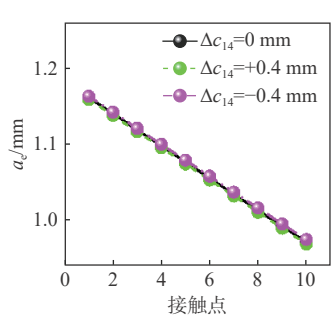
(f) 轴交角误差$\Delta\Sigma_{12}$对椭圆长半轴的影响



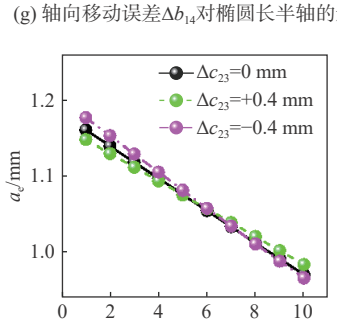
(g) 轴向移动误差Δb_{14}对椭圆长半轴的影响



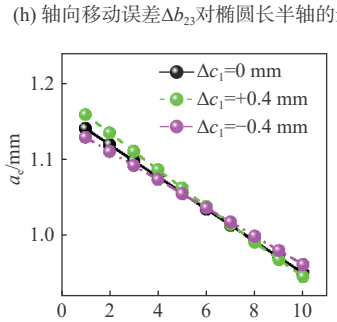
(h) 轴向移动误差Δb_{23}对椭圆长半轴的影响



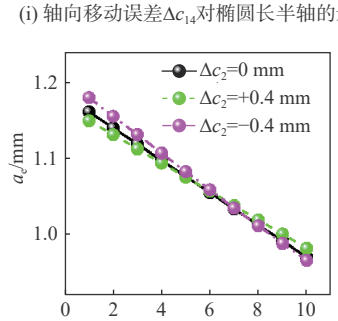
(i) 轴向移动误差Δc_{14}对椭圆长半轴的影响



(j) 轴向移动误差Δc_{23}对椭圆长半轴的影响



(k) 轴向移动误差Δc_1对椭圆长半轴的影响



(l) 轴向移动误差Δc_2对椭圆长半轴的影响

图 13 误差对椭圆长半轴的影响

Fig. 13 Influence of error on major axis of ellipse

3 仿真实例

综合考虑以上各参数对啮合性能的影响，选取主要设计参数 β_2=4.2°、m_{t2}=3.0 mm、α_n=20°，同算例Ⅲ-B，建立点接触共轭渐开线蜗杆副的数字化模型，如图 14 所示。

建立传动副的有限元分析模型，渐开线蜗杆材料 45 号钢的弹性模量取 2.06×10^{11} Pa、泊松比为 0.277、蜗轮材料锡青铜 ZQSn6-6-3 弹性模量取 1.24×10^{11} Pa、泊松比为 0.3，对蜗杆施加 10 N·m 扭矩进行静态接

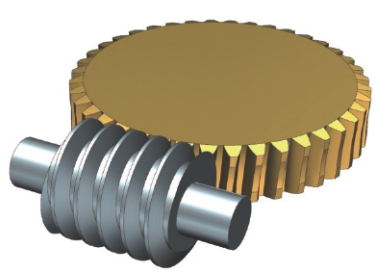
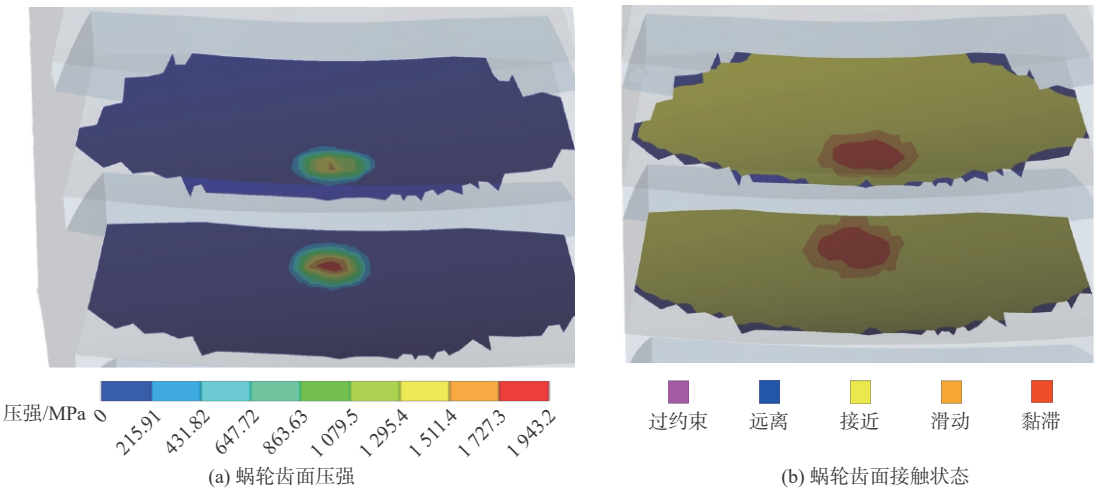


图 14 点接触共轭渐开线蜗杆副三维模型

Fig. 14 Three-dimensional model of point-contact conjugate involute worm gear pair

触分析, 传动副啮合时的齿面接触压强及状态如图 15(a) 和图 15 (b) 所示。可以看出, 蜗轮齿面的接触区形状近似椭圆, 符合 Hertz 接触理论, 接触区与图 8 中Ⅲ-B 所示的理论计算结果一致; 传动副同时有 2 个齿接触, 与理论重合度 ε=1.864 2 计算结果一致。

安装误差Δa_{12}、$\Delta\Sigma_{12}$、c_1、c_2 对传动副啮合时齿面压强的影响规律如图 16 所示。结果表明, 安装误差对蜗轮齿面压强的影响几乎可以忽略不计。



(a) 蜗轮齿面压强　　(b) 蜗轮齿面接触状态

图 15　蜗轮齿面接触压强及状态

Fig. 15　Contact pressure and state of worm gear tooth surface

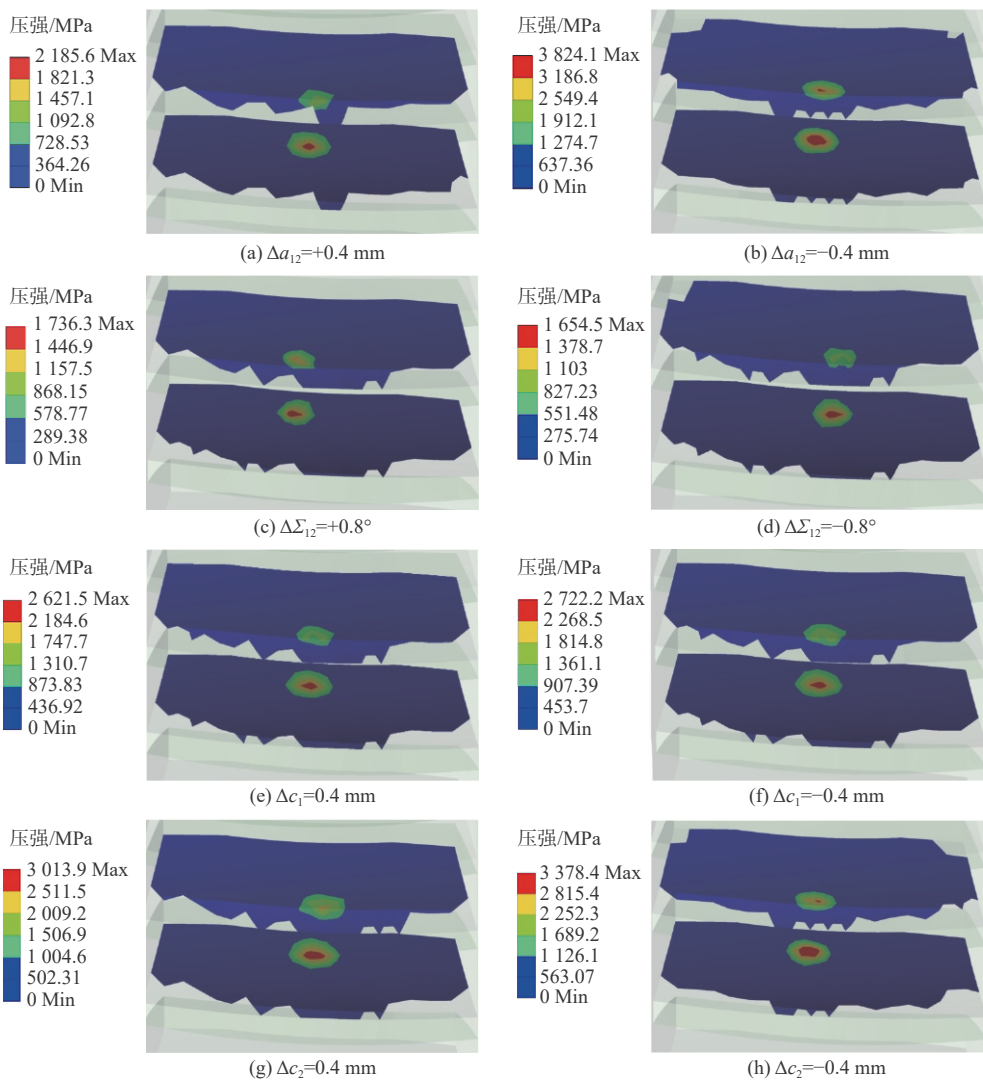


图 16　安装误差对蜗轮齿面压强的影响

Fig. 16　Influence of error on pressure of worm gear tooth surface

因此，该传动对安装误差不敏感，有限元分析的结果也验证了该理论分析的正确性。

4 结 论

1）本文构建了含有制造和安装误差的点接触共轭渐开线蜗杆传动的数学模型，通过理论推导证明了蜗杆副的瞬时传动比恒定，不受误差影响。

2）螺旋角主要影响接触区位置的分布，适中的螺旋角会获得较好的接触区分布；端面模数主要影响接触迹线的倾角和接触区面积，设计时应在合理范围内选择较小的模数；压力角的变化对接触区几乎无影响，但影响传动副的重合度，且呈负相关关系，因此，设计时压力角不宜过大，以确保增大接触区面积、提高传动副的重合度。

3）传动副对制造和安装误差均不敏感，但对轴交角误差的敏感程度要大于其他类型的误差，因此，在加工制造和安装传动副时，应尽量减小或避免轴交角造成的误差。

4）建立了传动副的数字化模型和有限元分析模型，通过仿真分析得到的接触区和重合度与理论计算结果一致，安装误差对接触区的影响规律也与理论分析结果一致，从而验证了点接触共轭渐开线蜗杆副啮合理论的正确性。

参考文献（References）

[1] 王树人, 刘平娟. 圆柱蜗杆传动啮合原理[M]. 天津: 天津科学技术出版社, 1982.
WANG S R, LIU P J. Meshing principle of cylindrical worm drive [M]. Tianjin: Tianjin Scientific & Technical Publishers, 1982 (in Chinese).

[2] LITVIN F L, FUENTES A. Gear geometry and applied theory[M]. 2nd ed. New York: Cambridge University Press, 2004.

[3] 吴鸿业, 张亚雄, 齐麟. 蜗杆传动设计[M]. 北京: 机械工业出版社, 1986.
WU H Y, ZHANG Y X, QI L. Worm drive design[M]. Beijing: China Machine Press, 1986(in Chinese).

[4] 申永胜. 机械原理教程[M]. 2版. 北京: 清华大学出版社, 2005: 162-165.
SHEN Y S. Theory of machines and mechanisms[M]. 2nd ed. Beijing: Tsinghua University Press, 2005: 162-165(in Chinese).

[5] 田培棠. 齿轮刀具设计与选用手册[M]. 北京: 国防工业出版社, 2011: 73-77.
TIAN P T. Handbook of gear cutter design and selection[M]. Beijing: National Defense Industry Press, 2011: 73-77(in Chinese).

[6] 彭泽良. 渐开线圆柱蜗杆传动承载啮合理论与实验研究[D]. 重庆: 重庆大学, 2003: 34-39.
PENG Z L. Theoretical and experimental research on loaded meshing of involute cylinder worm-drive[D]. Chongqing: Chongqing University, 2003: 34-39(in Chinese).

[7] 赵迪, 方舟, 范庆明, 等. 渐开线蜗轮蜗杆参数化设计及装配误差分析[J]. 西安工业大学学报, 2019, 39(2): 152-158.
ZHAO D, FANG Z, FAN Q M, et al. Parametric design and assembly error analysis of an involute worm gear[J]. Journal of Xi'an Technological University, 2019, 39(2): 152-158(in Chinese).

[8] 沈谦. 点接触圆柱蜗杆传动的设计原理[J]. 南京航空航天大学学报, 1991, 23(2): 48-54.
SHEN Q. A design principle of cylindrical worm drive with point contact tooth surface[J]. Journal of Nanjing University of Aeronautics & Astronautics, 1991, 23(2): 48-54(in Chinese).

[9] 苏代忠, 陈兵奎, 秦大同, 等. 失配渐开线圆柱蜗杆传动的弹性啮合分析、加工仿真及相关技术的理论与试验研究[J]. 机械工程学报, 2002, 38(增刊 1): 56-60.
SU D Z, CHEN B K, QIN D T, et al. Elastic meshing analysis, machining simulation and theoretical and experimental research on related technologies of mismatched involute cylindrical worm drive[J]. Chinese Journal of Mechanical Engineering, 2002, 38(Sup 1): 56-60(in Chinese).

[10] YE X X, CHEN Y H, LU B B, et al. Study on a novel backlash-adjustable worm drive via the involute helical beveloid gear meshing with dual-lead involute cylindrical worm[J]. Mechanism and Machine Theory, 2022, 167: 104466.

[11] SIMON V. Load distribution in cylindrical worm gears[J]. Journal of Mechanical Design, 2003, 125(2): 356-364.

[12] SIMON V. The influence of gear hobbing on worm gear characteristics[J]. Journal of Manufacturing Science and Engineering, 2007, 129(5): 919-925.

[13] SIMON V. Computer aided loaded tooth contact analysis in cylindrical worm gears[J]. Journal of Mechanical Design, 2005, 127(5): 973-981.

[14] 秦大同, 张光辉, 加藤正名. 失配点啮合环面蜗杆传动的研究[J]. 机械工程学报, 1995, 31(3): 79-83.
QIN D T, ZHANG G H, KATO M. Research on mismatch point-contact hourglass worm[J]. Chinese Journal of Mechanical Engineering, 1995, 31(3): 79-83(in Chinese).

[15] 秦大同, 张光辉, 加藤正名. 锥面包络环面蜗杆与直廓环面蜗轮失配啮合传动的研究[J]. 重庆大学学报 (自然科学版), 1995, 18(4): 15-20.
QIN D T, ZHANG G H, KATO M. Study on mismatched engagement of hourglass worm gearing with the worm generated by a cone and the wheel generated by a straight profile hob[J]. Journal of Chongqing University (Natural Science Edition), 1995, 18(4): 15-20 (in Chinese).

[16] MENG Q X, ZHAO Y P, YANG Z Y, et al. Meshing theory and error sensitivity of mismatched conical surface enveloping conical worm pair[J]. Mechanism and Machine Theory, 2020, 145: 103681.

[17] MENG Q X, ZHAO Y P, CUI J, et al. Meshing theory of mismatched ZC1 worm drive[J]. Mechanism and Machine Theory, 2020, 150: 103869.

[18] CHI Y F, ZHAO Y P, ZHU X Y, et al. Mismatched gearing composed of hourglass worm and spur gear: Meshing theory, tooth contact simulation, comprehensive design[J]. Mechanism and Machine Theory, 2022, 174: 104883.

[19] 赵超飞, 魏冰阳. 直廓环面蜗杆-圆柱斜齿轮传动的几何建模与接触特性分析[J]. 机械传动, 2018, 42(11): 123-126.

ZHAO C F, WEI B Y. Geometric modeling and contact characteristic analysis of hindley worm and cylindrical helical gear transmission[J]. Journal of Mechanical Transmission, 2018, 42(11): 123-126 (in Chinese).

[20] 董学朱. 齿轮啮合理论基础[M]. 北京: 机械工业出版社, 1989.
DONG X Z. Foundation of gear engagement theory[M]. Beijing: China Machine Press, 1989(in Chinese).

[21] REN W, LI H T, XU Z K, et al. A point-contact conjugate hourglass worm drive based on the meshing theory of conjugate tooth surfaces generated by two generating surfaces[J]. Mechanism and Machine Theory, 2022, 174: 104877.

[22] 董学朱, 李海清, 魏文军. 双自由度齿轮啮合理论及应用[M]. 北京: 机械工业出版社, 2011.
DONG X Z, LI H T, WEI W J. Theory and application of gear meshing with two degrees of freedom[M]. Beijing: China Machine Press, 2011(in Chinese).

Meshing theory and performance analysis of point-contact conjugate involute worm gear pair

REN Wen[1], LI Jie[1, 2], WANG Tianyu[1, 2], LI Haitao[1, 2, *]

(1. College of Engineering, China Agricultural University, Beijing 100083, China;
2. Mechanical Engineering Training Center, China Agricultural University, Beijing 100083, China)

Abstract: Point-contact conjugate involute worm drive can reduce the sensitivity of the worm gear pair to manufacturing and assembly errors and ensure transmission accuracy. On the basis of differential geometry and gear meshing theory, this paper constructed a mathematical model of point-contact conjugate involute worm drive containing manufacturing and assembly errors, studied the influence of main design parameters such as helix angle, modulus, and pressure angle on the meshing performance, and analyzed the sensitivity of the worm gear pair to manufacturing and assembly errors. Finally, the distribution of the contact area was verified through simulation. The results show that the instantaneous drive ratio of the worm gear pair remains constant with manufacturing and assembly errors. The selection of a moderate helix angle, a small modulus, and a small pressure angle will increase the contact area and raise the contact ratio of the worm gear pair. The worm gear pair is insensitive to manufacturing and assembly errors, but the manufacturing and assembly errors of crossed axis angle should be minimized or avoided. The contact area and contact ratio obtained from the simulation analysis agree with the theoretical calculations, verifying the correctness of the theory.

Keywords: point contact; conjugate; involute worm; meshing performance; error analysis

Received: 2022-09-14; **Accepted**: 2022-11-04; **Published Online**: 2023-01-10 13:50
URL: link.cnki.net/urlid/11.2625.V.20230109.1305.004
Foundation item: Fund of National Key Laboratory of Science and Technology on Helicopter Transmission (HTL-O-19G02)
* **Corresponding author**. E-mail: h.li@cau.edu.cn

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2022.0848

基于行为特征的雷达辐射源威胁评估

王俊迪[1]，王星[1]，田元荣[2,*]，陈游[1]
(1. 空军工程大学 航空工程学院，西安 710038；2. 国防科技大学 电子对抗学院，合肥 230037)

摘 要：针对当前雷达辐射源威胁评估对精确侦察数据依赖性较大的问题，提出一种基于行为特征的雷达辐射源威胁评估算法。从辐射源目标行为特征和数据融合理论出发，建立基于行为特征的辐射源威胁评估体系，并采用模糊理论和Vague数据集对各子行为进行表示；考虑到指标间的耦合性和空战的高动态性，利用改进的区间灰色关联度修正初始权重，建立以距离为自变量的态势状态函数，为各子行为动态赋权；采用改进的雷达图法计算威胁目标的威胁程度。仿真结果表明：所提算法具有较好的准确性和适应性。

关 键 词：数据融合；威胁评估；行为特征；动态权重；区间灰色关联度；雷达图法
中图分类号：TN974
文献标志码：A **文章编号**：1001-5965(2024)10-3196-12

在以超视距攻击为主要作战模式的空战中，导弹的瞄准和发射依赖自身的雷达辐射源，因此，通过评估目标的雷达辐射源威胁程度可以更加直接地反映目标的空战威胁程度。快速、准确的辐射源评估有利于提升载机的战场生存能力，并帮助飞行员采取恰当的电子对抗手段或战术规避。

威胁评估是各国学者研究的重点，主要算法包括多属性决策理论[1-2]、直觉模糊集(intuitionistic fuzzy sets, IFS)[3-4]、贝叶斯网络(Bayesian network, BN)[5-6]、多目标排序[7]、机器学习[8-9]等。Zhao等[10]提出了一种基于可能性理论的扩展累积前景交互式多标准决策威胁评估算法。寇英信团队利用灰主成分[11]、线性回归分析[12]、动态变权[13]等算法改进了CRITIC或TOPSIS算法。Zhang等[14]将信息量和信息的不确定性结合起来设计了一种新颖的直觉模糊熵。Zhang等[15]将众多专家提供的广义参数矩阵引入广义直觉模糊软集，形成了群广义直觉模糊软集。Ma等[16]提出了基于云模型的目标威胁评估技术。杨远志等[17]将粗糙集理论引入信息熵理论，构建了一套完备的辐射源威胁评估算法。张晓雷等[18]利用云重心评价法和排队法解决了未来联合防空作战雷达辐射源的威胁评估问题。张莹等[19]建立了基于改进组合赋权的雷达图辐射源威胁评估法。战场中，电磁脉冲密度的急剧增加和新体制雷达的广泛应用给单独依靠雷达辐射源信号信息进行威胁评估的算法带来了挑战。针对以上问题，王星团队[20-23]提出了将自身雷达探测信息和告警器截获的辐射源信号信息相融合的威胁评估思想。改进的辐射源威胁评估算法融合了雷达探测信息提供的距离、方位等信息，扩展了威胁评估指标的维度[24-25]，从而可以更加全面地反映战场的威胁评估。

为减少辐射源威胁评估对侦察参数数据准确度的依赖，本文在基于数据融合的辐射源威胁评估的思想基础上，建立了基于行为特征的威胁评估体系，将复杂空战态势的描述从具体转变为抽象，从而提升了算法的容错能力。根据建立的行为特征体系，计算各子行为指标的Vague数据集隶属度，利用动态变权算法和灰色关联度算法优化威胁评

收稿日期：2022-10-10；**录用日期**：2022-12-25；**网络出版时间**：2023-01-03 17:39
网络出版地址：link.cnki.net/urlid/11.2625.V.20230103.1419.002
基金项目：国家自然科学基金(62001489)；陕西省自然科学基金(2021JM-225)
*通信作者. E-mail：tianyuanrong20@nudt.edu.cn

引用格式：王俊迪，王星，田元荣，等. 基于行为特征的雷达辐射源威胁评估[J]. 北京航空航天大学学报，2024，50(10)：3196-3207.
WANG J D，WANG X，TIAN Y R，et al. Threat assessment of radar radiation sources based on behavioral characteristics[J]. Journal of Beijing University of Aeronautics and Astronautics，2024，50(10)：3196-3207 (in Chinese).

估权重，并使用改进的雷达图法计算威胁数值，以得到快速准确的辐射源来进行威胁评估。

1　辐射源行为特征的建立

目前，对于辐射源威胁评估的研究主要以解析截获的辐射源信号参数为评价指标，并在此基础上开展威胁等级评定。参数化的评价体系科学、严谨，能够细致准确地计算出辐射源威胁等级。但是，对实际运用而言，参数体系的威胁评估算法增加了对数据准确性的要求。针对以上问题，根据空中目标辐射源行为和态势行为的层次特点，本文建立了基于行为特征的空中目标辐射源威胁评估的层次划分算法，如图 1 所示。算法为 4 层划分结构，自下而上逐层抽象，这 4 层分别为：脉冲描述、状态描述、雷达/态势子行为描述、威胁度描述。

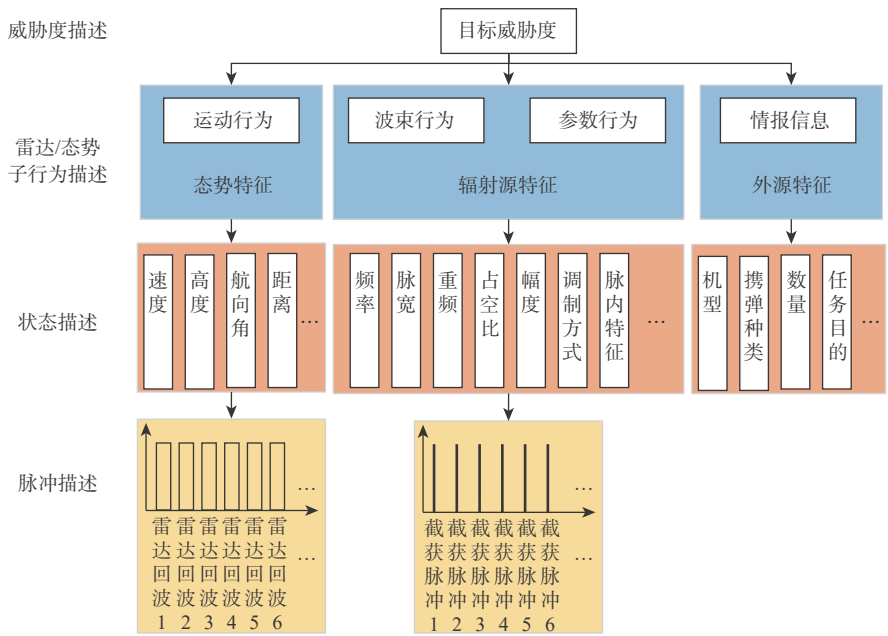


图 1　基于行为特征的威胁评估体系

Fig. 1　Threat assessment system based on behavioral characteristics

1）脉冲描述。脉冲描述是对雷达脉冲的完整描述，包含了脉冲信号的参数特征和样式特征。脉冲描述数据被储存到机载设备中，为后期分析处理提供数据支撑。

2）状态描述。状态描述是对雷达回波脉冲或侦察截获脉冲的解析，以此目标的态势信息和雷达辐射源信息。其同时还结合数据链中的其他外源信息，为威胁评估做支撑。

3）雷达/态势子行为描述。子行为描述是在状态描述的基础上对目标进行更加抽象的描述，从而更直观地表达目标的行动意图。

从行为层面上计算辐射源威胁度实际上是对目标从具体转变为抽象的描述，其聚焦于能够反映威胁等级的核心特征。相较于辐射源的参数化描述，行为描述降低了对测量数据精度的要求[24]，允许数据少量缺失和模糊。作为抽象的表征方式，一个行为特征通常包括多种不同的参数特征，以有效减小计算的规模，并且具备较强的泛化能力和通用性[25]。

2　基于 Vague 数据集的行为隶属度计算

利用目标的行为特征计算威胁度需要获得各子行为的威胁数值。由于各子行为是抽象概念，无法对其直接进行量化，因此本文使用模糊理论和 Vague 数据集对各子行为进行表示。模糊理论能够反映指标的不确定性，不需要进行精确的描述。Vague 数据集通过隶属度和非隶属度对模糊概念进行描述，可以体现出指标的模糊性和不确定性。概念定义如下：

定义 1　在论域 X 中，对模糊集 $\tilde{A}$ 用 Vague 数据集表示，$\mu_{\tilde{A}}(x)$ 为由支持证据所导出的肯定隶属度下界，$v_{\tilde{A}}(x)$ 为由反对证据所导出的否定隶属度下界，且 $\mu_{\tilde{A}}(x)+v_{\tilde{A}}(x)\leqslant 1$，$\pi_{\tilde{A}}=1-\mu_{\tilde{A}}(x)-v_{\tilde{A}}(x)$ 表示集合 A 的不确定或犹豫度，表示为 $\tilde{A}=\{\langle x,\mu_{\tilde{A}}(x),v_{\tilde{A}}(x)\rangle|x\in X\}$ 或 $[\mu_{\tilde{A}}(x),1-v_{\tilde{A}}(x)]$。

定义 2　设 $\tilde{A}$、$\tilde{B}$ 为论域 X 上的 2 个用 Vagues 数据集表示的模糊集合，令 $\tilde{A}\cup\tilde{B}$、$\tilde{A}\cap\tilde{B}$ 分别表示模糊集合 $\tilde{A}$ 与 $\tilde{B}$ 的交集、并集，对应的隶属度函数分别为

$\mu_{\tilde{A}\cup\tilde{B}}$、$\mu_{\tilde{A}\cap\tilde{B}}$，对于$X$的任意元素$x$，定义为

$$\begin{cases}\mu_{\tilde{A}\cup\tilde{B}}\overset{\text{def}}{==}\max(\mu_{\tilde{A}}(x),\mu_{\tilde{B}}(x))\\ v_{\tilde{A}\cup\tilde{B}}\overset{\text{def}}{==}\min(v_{\tilde{A}}(x),v_{\tilde{B}}(x))\\ \mu_{\tilde{A}\cap\tilde{B}}\overset{\text{def}}{==}\min(\mu_{\tilde{A}}(x),\mu_{\tilde{B}}(x))\\ v_{\tilde{A}\cap\tilde{B}}\overset{\text{def}}{==}\max(v_{\tilde{A}}(x),v_{\tilde{B}}(x))\end{cases}\tag{1}$$

定义3 由于数据在不同指标中的数值大小相差较大，因此为统一各指标下数值大小，对数据进行归一化处理：

$$x'=\frac{x-x_{\max}}{x_{\max}-x_{\min}}\tag{2}$$

式中：$x_{\max}=\max(x)$，$x_{\min}=\min(x)$。

2.1 运动行为隶属度计算

目标的运动行为是在速度、高度、距离、角度等雷达探测信息参数的基础上抽象而来的行为特征。结合超视距空战战术的特点和战机运动规律，将运动行为的模糊评价语言划分为5个等级。设目标运动行为的行为论域$N_{\mathrm{d}}=[0,1]$，其包括高速远离模糊子集N_{d1}、中低速远离模糊子集N_{d2}、盘旋模糊子集N_{d3}、中低速接近模糊子集N_{d4}、高速接近模糊子集N_{d5}，隶属度函数为

$$\begin{cases}\mu_{N_{\mathrm{d}i}}(l,s)=(\mu_{L(l)}+\mu_{S(s)})/2\\ \mu_{L(l)}=\exp\left\{-\dfrac{((1-l)-(i-1)/2)^2}{2\delta_1^2}\right\}\quad i\in(1,2,3)\\ \mu_{S(s)}=\exp\left\{-\dfrac{(s-(i-1))^2}{2\delta_2^2}\right\}\quad i\in(1,2)\end{cases}\tag{3}$$

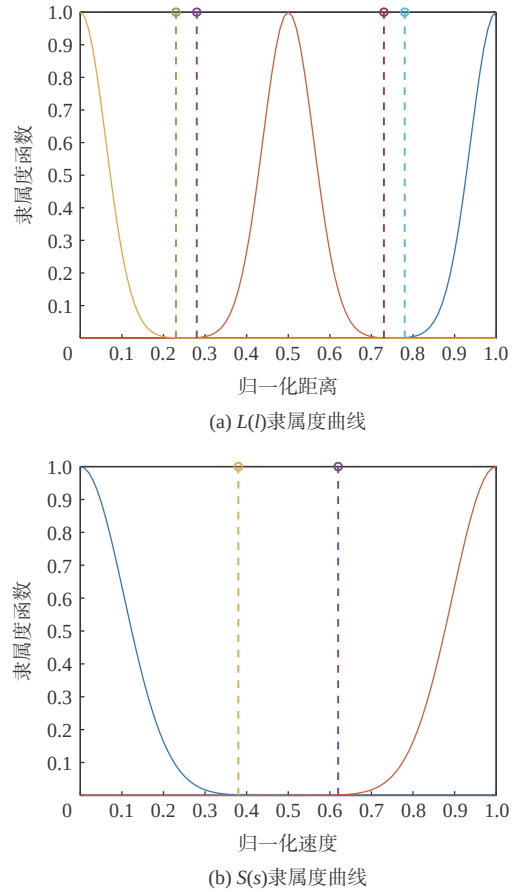
式中：$L(l)$和$S(s)$为2个模糊集，$l\in[0,1]$为归一化距离，$s\in[0,1]$为归一化速度。根据文献[11-13, 26]中对速度、距离、高度各隶属度函数的分析，求出式(3)中的$\delta_1=0.073$、$\delta_2=0.105$，$L(l)$和$S(s)$的隶属度曲线如图2所示。

根据式(3)可以得出目标运动行为的各模糊子集对应的Vague数据集数值，如表1所示。

2.2 辐射源行为隶属度计算

雷达行为被定义为一部雷达外显出来的所有特征及规律，包括信号的变化特征及雷达工作状态的转移特点和规则等。常见的雷达子行为包括波形行为、波束行为、天线扫描方式和占空比行为等。在现代空战中，机载雷达辐射源的波束行为和参数行为与火力方面的联系最为紧密，因此，本文选取这2种行为作为雷达行为的描述特征。

波束行为主要指目标机载雷达窄波束在指定空域的扫描方式。辐射源波束的变化方式通常有3种：方位不变、方位瞬变和方位慢变。辐射源波束方位不变，表示已锁定或跟踪到目标；方位慢变和瞬变表示敌攻击意图不明确或处于搜索状态。



(a) $L(l)$隶属度曲线

(b) $S(s)$隶属度曲线

图2 $L(l)$和$S(s)$隶属度曲线

Fig. 2 The membership curves of $L(l)$ and $S(s)$

表1 运动行为模糊评价语言与Vague转换

Table 1 Fuzzy evaluation language of motor behavior and Vague transition

威胁等级	Vague数据集数值
高速远离	[0,0.305]
中低速远离	[0.31,0.615]
盘旋	[0.28,0.73]
中低速接近	[0.39,0.69]
高速接近	[0.7,1]

根据这一规律，将雷达辐射源波束行为划为3个模糊等级。设目标辐射源波束行为论域$N_{\mathrm{b}}=[0,1]$，其包括短暂驻留模糊子集N_{b1}、循环驻留模糊子集N_{b2}、长时间驻留模糊子集N_{b3}。为客观计算波束行为隶属度，对雷达处理过程进行简要分析。

在雷达恒虚警处理中，只有当接收单元中的回波功率超过门限，才会检测到目标。令P_{Tc}表示测试单元中的功率，T_{CACFAR}表示自适应门限，k_2用于设定所需虚警率的因子，P_m表示第m个距离单元的功率，目标的检测准则可表示为

$$P_{\mathrm{Tc}}>T_{\mathrm{CACFAR}}=\frac{k_2}{M}\sum_{m=1}^{M}P_m\tag{4}$$

P_m与积累脉冲数有关，在改善因子0.8 dB下的经验表达为

$$I(n_{\mathrm{p}})=6.79(1+0.253P_{\mathrm{D}})\left[1+\frac{\lg(1/P_{\mathrm{fa}})}{46.6}\right]\lg n_{\mathrm{p}}\cdot$$
$$[1-0.14\lg n_{\mathrm{p}}+0.018\,310(\lg n_{\mathrm{p}})^2] \tag{5}$$

式中：P_{D}为检测概率；P_{fa}为虚警率；n_{p}为脉冲数。

改善因子与积累脉冲数的关系如图 3 所示。

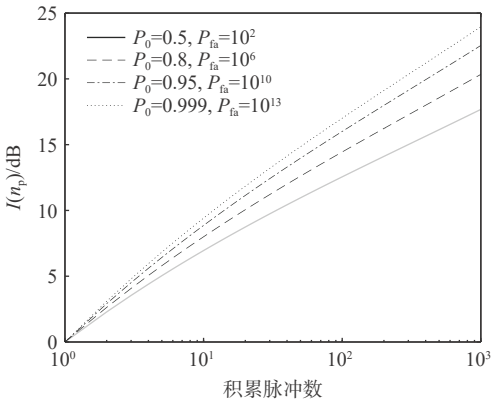


图 3　改善因子与积累脉冲数的关系曲线

Fig. 3　Relationship between improvement factor and accumulated pulse number

对于半球形空域覆盖，需要的独立波束位置数n_{p}为

$$n_{\mathrm{p}}=\frac{2\pi}{\theta_{\mathrm{Az}}\theta_{\mathrm{EL}}} \tag{6}$$

式中：$\theta_{\mathrm{AZ}}=3$ dB表示方位角波束带宽，rad；$\theta_{\mathrm{EL}}=3$ dB表示俯仰角波束带宽，rad。

对于恒定脉冲重频雷达，波束驻留时间为

$$T_{\mathrm{d}}=1/F_s\cdot n_{\mathrm{p}} \tag{7}$$

式中：F_{s}为信号脉冲频率。

根据式(4)～式(7)，目标辐射源波束行为模糊集隶属度函数为

$$\mu_{N_{\mathrm{bi}}(n)}=\exp\left\{-\frac{[t-(i-1)/2]^2}{2\delta_3^2}\right\}\qquad i\in(1,2,3) \tag{8}$$

式中：$\delta_3=0.078\,5$；$t\in[0,1]$为归一化驻留时间，N_{bi}函数隶属度曲线如图 4 所示。

根据式(8)可以得出目标波束行为 d 的各模糊子集所对应的 Vague 数据集数值，如表 2 所示。

表 2　波束行为模糊评价语言与 Vague 转换

Table 2　Fuzzy evaluation language of beam behavior and Vague transition

威胁等级	Vague数据集数值
短暂驻留	[0,0.28]
循环驻留	[0.22,0.78]
长时间驻留	[0.72,1]

参数行为主要指与导弹发射息息相关的雷达参数的变化，主要包括信号载频(radio frequency, RF)、信号脉宽(pulse width, PW)和信号重频(pulse repetition interval, PRI)。载频影响目标速度信息的获取，雷达系统主要通过多普勒频移求解探测目标的速度和距离信息，速度为V_{r}的探测目标的多普勒频移f_{d}为

$$f_{\mathrm{d}}=\frac{2V_{\mathrm{r}}}{\lambda}=\frac{2V_{\mathrm{r}}f_{tx}}{c} \tag{9}$$

式中：c为光速，λ为信号波长；f_{tx}为信号载频。

根据式 (9)，可以看出，多普勒频移f_{d}与载频成正比。目标速度分辨率ΔV_{r}与载频的关系为

$$\Delta V_{\mathrm{r}}=\frac{\lambda\Delta f_{\mathrm{d}}}{2}=\frac{\lambda c}{2T_{\mathrm{d}}f_{tx}} \tag{10}$$

速度分辨率与载频成反比，给定相参周期时，雷达的载频越高，得到的速度分辨率越好。而给定速度分辨率时，载频越高，相参处理周期越短，可获得更高的数据率。因此，雷达的载频越高，对于探测目标的速度信息掌握越准确，对探测目标而言，威胁程度越大。设目标辐射源载频行为论域为$N_{\mathrm{f}}=[0,1]$，其低载频模糊子集为N_{f1}、中载频模糊子集为N_{f2}、高载频模糊子集为N_{f3}，其隶属度函数为

$$\mu_{N_{\mathrm{fi}}}=\exp\left\{-\frac{[f-(i-1)/2]^2}{2\delta_4^2}\right\}\qquad i\in(1,2,3) \tag{11}$$

式中：$\delta_4=0.059\,8$；$f\in[0,1]$为归一化频率。

N_{fi}隶属度曲线如图 5 所示。

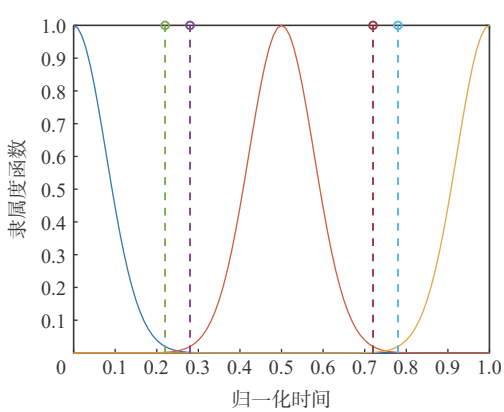


图 4　$N_{\mathrm{bi}}(i=1,2,3)$隶属度曲线

Fig. 4　The membership curve of $N_{\mathrm{bi}}(i=1,2,3)$

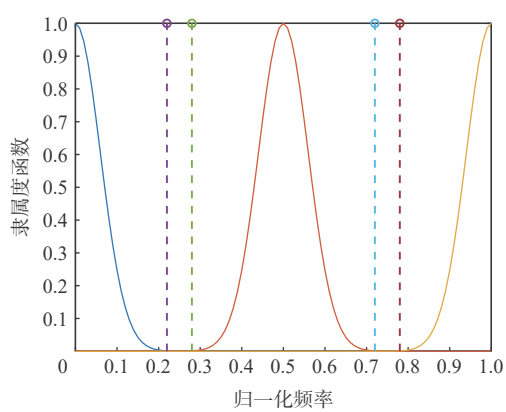


图 5　$N_{\mathrm{fi}}(i=1,2,3)$隶属度曲线

Fig. 5　The membership curve of $N_{\mathrm{fi}}(i=1,2,3)$

雷达另2个重要的参数是脉宽和重频，其影响着雷达测距精度ΔR和最大可测距离$R_{\max}$，关系式为

$$\begin{cases}\Delta R=\dfrac{cT_{PW}}{2}\\ R_{\max}=\dfrac{c}{2F_S}=R_{\max(l)}\displaystyle\prod_{i=2}^{k}\dfrac{R_{\max(i)}}{\Delta R_{\max(i)}}\end{cases}\tag{12}$$

为综合考虑脉宽和重频对威胁程度的影响，本文引入雷达信号占空比(duty ratio, DR)，表示为

$$D=T_{PW}F_s\tag{13}$$

设目标占空比行为论域$N_{\mathrm{DR}}=[0,1]$，其低占空比模糊子集N_{DR1}、高占空比模糊子集N_{DR2}，隶属度函数为

$$\mu_{N_{\mathrm{DR}i}(\tau)}=\exp\left\{-\frac{[\tau-(i-1)/2]^2}{2\delta_5^2}\right\}\qquad i\in(1,2)\tag{14}$$

式中：$\delta_5=0.15$；$\tau\in[0,1]$为归一化时间。

$N_{\mathrm{DR}i}$隶属度曲线如图6所示。

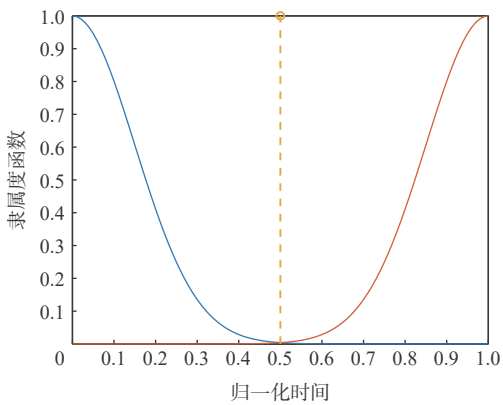


图6 $N_{\mathrm{DR}i}(i=1,2)$隶属度曲线

Fig. 6 The membership curve of $N_{\mathrm{DR}i}(i=1,2)$

辐射源参数行为的模糊集可由载频数据集和占空比数据集来表达：

$$N_{\mathrm{c}(i\times j)}=N_{\mathrm{f}i}N_{\mathrm{DR}j}\qquad i\in(1,2,3),j\in(1,2)\tag{15}$$

根据式(15)可以得出目标参数行为各模糊子集对应的Vague数据集数值，如表3所示。

表3 参数行为模糊评价语言与Vague转换

Table 3 Fuzzy evaluation language of parameter behavior and Vague transition

威胁等级	Vague数据集数值
低载频低占空比	[0,0.11]
低载频高占空比	[0,0.22]
中载频低占空比	[0,0.36]
中载频高占空比	[0.14,0.72]
高载频低占空比	[0,0.5]
高载频高占空比	[0.39,1]

2.3 情报信息

情报信息主要是指其他数据源传输的关于目标类型、数量等先验数据，这些数据通过数据链传输威胁评估中心。在空战中，目标的机型属性和载弹类型是威胁评估的重要因素。设机载类型论域$N_{\mathrm{J}}=[0,1]$，根据目前的机型划分其随队支援飞机模糊子集N_{J1}、三代战机模糊子集N_{J2}、四代战机模糊子集N_{J3}，其隶属度函数为

$$\mu_{N_{\mathrm{J}i}}=\kappa_{\mathrm{c}}\kappa_{\mathrm{e}}\exp\left\{-\frac{[\kappa_{\mathrm{m}}-(i-1)/2]^2}{2\delta_6^2}\right\}\qquad i\in(1,2,3)\tag{16}$$

式中：κ_{m}为战机的机动性能；κ_{c}为各类战机探测能力；κ_{e}为各类战机的电子对抗能力。

$N_{\mathrm{J}i}$隶属度曲线如图7所示。

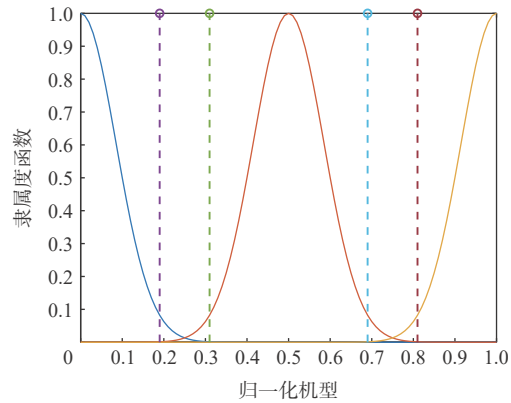


图7 $N_{\mathrm{J}i}(i=1,2,3)$隶属度曲线

Fig. 7 The membership curve of $N_{\mathrm{J}i}(i=1,2,3)$

设导弹类型论域$N_k=[0,1]$，根据导弹类型划分其对地制导炸弹模糊子集N_{k1}、反辐射导弹模糊子集N_{k2}、红外制导导弹模糊子集N_{k3}、雷达制导导弹模糊子集N_{k4}，其隶属函数为

$$k\mu_{N_{ki}}=\begin{cases}0.9 & \text{雷达制导导弹}\\ 0.7 & \text{红外制导导弹}\\ 0.2 & \text{反辐射导弹}\\ 0.1 & \text{对地制导炸弹}\end{cases}\tag{17}$$

不同类型的战机可以加装不同类型的导弹，此时，其威胁评估模糊集满足：

$$N_{\mathrm{U}(i\times j)}=N_{ki}N_{\mathrm{J}j}\qquad i\in(1,2,3,4),j\in(1,2,3)\tag{18}$$

3 子行为权重的确定

定义4 设$\tilde{A}=\{\langle x,\mu_{\tilde{A}}(x),v_{\tilde{A}}(x)\rangle|x\in X\}$为论域$X$上的Vague数据集，对于$\forall q_{\mathrm{a}},p_{\mathrm{a}}>0$，称式(19)为$\tilde{A}$的广义模糊熵[27]：

$$\begin{aligned}E_{p_{\mathrm{a}},q_{\mathrm{a}}}(\tilde{A})=&\frac{1}{n}\sum_{i=1}^{n}(1-|\mu_{\tilde{A}}(x_i)-v_{\tilde{A}}(x_i)|^{p_{\mathrm{a}}})\cdot\\ &\ln[\mathrm{e}+|\mu_{\tilde{A}}(x_i)-v_A(x_i)|^{p_{\mathrm{a}}}(1-\mu_{\tilde{A}}(x_i)-v_{\tilde{A}}(x_i))^{q_{\mathrm{a}}}]=\\ &\frac{1}{n}\sum_{i=1}^{n}(1-|\mu_{\tilde{A}}(x_i)-v_{\tilde{A}}(x_i)|^{p_{\mathrm{a}}})\cdot\\ &\ln[e+|\mu_{\tilde{A}}(x_i)-v_{\tilde{A}}(x_i)|^{p_{\mathrm{a}}}\pi_{\tilde{A}}^{q_{\mathrm{a}}}(x_i)]\end{aligned}\tag{19}$$

式中：q_{a}和p_{a}为广义模糊熵系数。可以根据决策者的风险态度设定q_{a}和p_{a}的值，本文选取决策者风险态度中立且Vague模糊数据集差异化信息凸显的系数$p_{\mathrm{a}}=0.455\,9$和$q_{\mathrm{a}}=0.099\,7$。

熵权法是一种只与决策信息有关的客观赋权算法, 其根据各属性提供的信息量的大小确定各属性权值。本文采用熵权法求解子行为权重, 同时采用式(18)计算 Vague 数据集的模糊熵。

在决策问题中, 应遵循的一个重要原则是减少整体系统的不确定性。根据该原则, 本文建立以下非线性规划模型求解最优权重:

$$\begin{cases} \min\sum_{i=1}^{n}(w_i^0)^2(1-E_{ip_a,q_a}(\tilde{A}))^{-1} \\ \text{s.t.}\quad \sum_{i=1}^{n}w_i^0=1 \end{cases} \tag{20}$$

通过对式(19)求解得出目标的初始属性权重:

$$w_i^0=\frac{(1-E_{jp_a,q_a}(\tilde{A}))}{\sum_{j=1}^{n}(1-E_{jp_a,q_a}(\tilde{A}))} \tag{21}$$

在建立评估体系的过程中, 不可避免地会引入主观因素, 且建立的各子行为之间存在一定的关联性。为减少各指标间的关联性, 保证各子行为的独立性, 本文利用区间灰色关联度算法减少了各子行为间的关联度。为准确度量模糊集之间的距离, 本文采用改进的相似度表达式, 如下:

$$d_{0sij}(t)=\frac{1}{2}\left|\mu_{a_{ij}^0}(x,t)-\mu_{a_{ij}^s}(x,t)\right|+\frac{1}{2}\left|v_{a_{ij}^0}(x,t)-v_{a_{ij}^s}(x,t)\right|+\frac{1}{6}\left|\pi_0(x,t)+\pi_s(x,t)\right| \tag{22}$$

式中: $\mu_{a_{ij}^0}$表示隶属度矩阵 u 的第 i 行, 第 j 列的数值; $(\mu_{a_{ij}^s})$表示参考隶属度矩阵 u 的第 i 行, 第 j 列的数值; $(v_{a_{ij}^0})$表示隶属度矩阵 v 的第 i 行, 第 j 列的数值; $(v_{a_{ij}^s})$表示参考隶属度矩阵 v 的第 i 行, 第 j 列的数值。

对求得的相似度 $d_{0sij}(t)$进行归一化处理, 得到 $D_{0sij}(a_{ij}^0(t),a_{ij}^s(t))$, 则灰色关联系数 $\gamma_{ij}(A_0(t),A_1(t))$为

$$\gamma_{ij}(A_0(t),A_1(t))=\left(\min_i\min_j\min_k\left\{D_{0sij}(a_{ij}^0(t),a_{ij}^s(t))\right\}+\rho\max_i\max_j\max_k\left\{D_{0sij}(a_{ij}^0(t),a_{ij}^s(t))\right\}\right)\Big/\left(D_{0sij}(a_{ij}^0(t),a_{ij}^s(t))+\rho\max_i\max_j\max_k\left\{D_{0sij}(a_{ij}^0(t),a_{ij}^s(t))\right\}\right) \tag{23}$$

式中: $\rho\in[0,1]$为分辨系数, 一般取 0.5。

将 2 个因素的关联度假设成 2 个集合的交集, 引入集合运算。用 w_i、w_j表示两两联系的因素权重, 则修正后的权重表示为

$$w_i+w_j-r_{ij}=\left(w_i-\frac{1}{2}r_{ij}\right)+\left(w_j-\frac{1}{2}r_{ij}\right) \tag{24}$$

通常, 威胁评估属性指标权重不变, 考虑到空战态势变化与权重之间的关联性, 本文采用动态权重理论。动态权重是指通过实时的战场态势改变属性指标权重, 主要算法是构造状态变权重向量 $\boldsymbol{S}=(x_1,x_2,\cdots,x_n)$与初始权重$\boldsymbol{w}$的 Hadamard 乘积, 可得

$$\boldsymbol{w}'=\frac{\boldsymbol{wS}}{\sum_{i=1}^{m}w_iS_i}=\frac{[w_1S_1,w_2S_2,\cdots,w_mS_m]}{\sum_{i=1}^{m}w_iS_i} \tag{25}$$

状态权重向量$\boldsymbol{S}=(x_1,x_2,\cdots,x_n)$主要由空战态势构建的均衡函数$\boldsymbol{B}(\boldsymbol{x})$求导得出, 其计算公式为

$$\boldsymbol{S}=\frac{\partial\boldsymbol{B}(\boldsymbol{x})}{\partial\boldsymbol{x}} \tag{26}$$

式中: $\boldsymbol{B}(\boldsymbol{x})$为构建空战场的威胁目标均衡函数及 j动态权重的核心, 本文根据现代空战的规律和飞行员判断准则构建了基于距离因素的均衡函数。

对于情报行为, 距离因素对其数值大小的影响, 采用一阶线性函数表达其均衡函数, 则第 j个威胁目标情报行为与距离 L_j的均衡函数 $B_{\mathrm{U}j}$为

$$B_{\mathrm{U}j}=L_j \tag{27}$$

对于运动行为, 随着距离的缩进, 其影响程度越来越小, 主要作用在于构成攻击的先决条件, 其均衡函数为非线性函数。则第 j个威胁目标的运动行为与距离 L_j的均衡函数 $B_{\mathrm{d}j}$为

$$B_{\mathrm{d}j}=2L_j+\frac{1}{2}\alpha_1(L_j/L_{jm\max})^2+\alpha_2\ln(L_j/L_{jm\max}) \tag{28}$$

式中: α_1和α_2为运动行为的变权权重, 可以根据指挥系统或飞行员实时确定; $L_{jm\max}$为第 j个威胁目标导弹最大攻击距离。

作战双方从远距不断到近距的过程中, 逐渐进入空空导弹发射范围内, 此时, 雷达辐射行为中的波束行为和参数行为充分体现了威胁目标的攻击意图, 对威胁程度的影响较大。则第 j个威胁目标辐射源行为与距离 L_j的均衡函数 $B_{\mathrm{R}j}$为

$$B_{\mathrm{R}j}=-\frac{\varepsilon_jL_{jm\max}}{\alpha_3}\mathrm{e}^{\alpha_3(1-L/L_{jm\max})} \tag{29}$$

式中: α_3为雷达辐射源行为的变权权重, 可以根据指挥系统或飞行员实时确定; ε_j为第 j个的威胁目标的电子设备作战性能。

综上所述, 各指标的动态权重表示为

$$w'(L)=\begin{cases} \dfrac{w_i^0[2+\alpha_1(L_j/L_{jm\max})+\alpha_2(L_j/L_{jm\max})]}{\sum_{i=1}^{n}w_i^0S_i} & \text{运动行为} \\ \dfrac{w_i^0[\varepsilon_j\mathrm{e}^{\alpha_3(1-L/L_{jm\max})}]}{\sum_{i=1}^{n}w_i^0S_i} & \text{辐射源行为} \\ \dfrac{w_i^0}{\sum_{i=1}^{n}w_i^0S_i} & \text{情报行为} \end{cases} \tag{30}$$

4 基于改进雷达图法的威胁数值计算

根据各子行为的权重划分不同指标扇形区域大小，为避免指标顺序的差异导致连接的多边形形状不同，本文用扇形区域取代原始的三角区域，算法的具体实现步骤如下：

步骤 1 根据属性指标权重，按照$\theta_j=360°\cdot w_j(l),j\in(1,2,\cdots,n)$计算各指标的扇形角度。

步骤 2 确定指标轴。以平行于水平线为起始轴，根据n个指标的角度，逆时针画出其余$n-1$条射线，如图 8 所示。

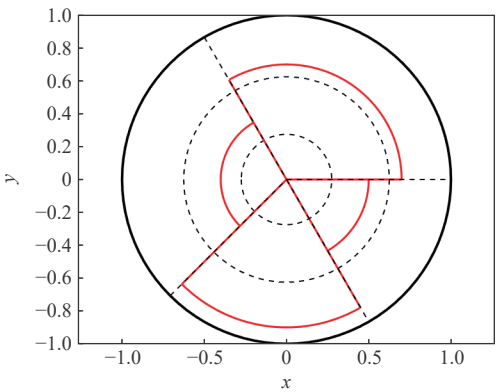


图 8 多扇形雷达

Fig. 8 Multi-sector radar chart

步骤 3 绘制多扇形雷达图。根据 Vague 数据集的距离式 (22) 计算目标的各指标属性到圆心的距离d_k。以圆心O为起点、以距离d_k为半径逆时针画扇形，分别计算各扇形弧长总和C'和面积总和S'：

$$\begin{cases} S'=\sum\limits_{i=1}^{n}\pi d_i^2\dfrac{\theta_i}{2\pi}=\sum\limits_{i=1}^{n}\pi d_i^2 w_i \\ C'=\sum\limits_{i=1}^{n}2\pi d_i\dfrac{\theta_i}{2\pi}=\sum\limits_{i=1}^{n}2\pi d_i w_i \end{cases} \tag{31}$$

步骤 4 计算辐射源威胁评估值K，参照评价思想和相似原理，构建面积评价指标和周长评价指标：

$$\begin{cases} \beta_1=S'/\pi \\ \beta_2=1+\dfrac{\left|L(2\pi\sqrt{S'\pi})^{-1}-S'\right|}{L\left(2\pi\sqrt{S'/\pi}\right)^{-1}} \\ K=\sqrt{\beta_1\beta_2} \end{cases} \tag{32}$$

5 仿真实验

5.1 某空战态势场景构建

构建超视距作战的典型空空对抗作战场景：蓝方 6 架战机对抗红方带有信息支援情况下的 5 架战机，如图 9 所示。红方根据卫星侦察、地面预警探测组网系统等获取蓝方情报信息，如表 4 所示。

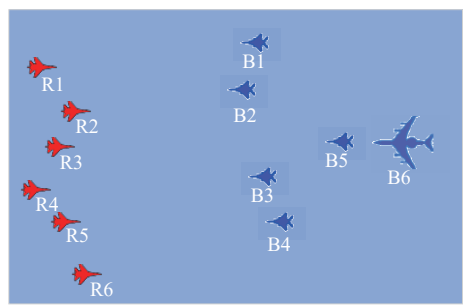


图 9 空空作战场景示意图

Fig. 9 Schematic diagram of air-to-air combat scenario

表 4 蓝方情报信息

Table 4 Blue intelligence information

编号	飞机型号	载弹类型	空空弹最大射程/km
B1	F-22战机	空对空红外制导导弹、空对空雷达制导导弹	150
B2	F-15战机	空对空红外制导导弹、空对空雷达制导导弹	100
B3	F-15战机	空对空红外制导导弹、空对空雷达制导导弹	100
B4	F-22战机	空对空红外制导导弹、空对空雷达制导导弹	150
B5	F-22战机	空对地反辐射导弹、空对空雷达制导导弹	150
B6	预警机		

5.2 基于行为特征的单机威胁评估仿真

红方某一战机在某一时间段内的雷达告警接收机接收到蓝方的辐射源特征参数行为和机载雷达探测的运动行为信息，如表 5 所示，表中，T_s为辐射源照射开始时间，T_e为辐射源照射结束时间，N为积累脉冲个数。

根据表 5 中的数据和式(3)～式(18)，可得蓝方各目标行为 Vague 数据集，如表 6 所示。

根据式(21)～式(23)，可得各属性指标的关联系数矩阵γ_{ij}，如图 10 所示。可以看出，波束行为和参数行为的关联度较大，情报行为和其他属性关联度较小。

根据式(24)和表 6，可计算出修正后的初始权重，如表 7 所示。

表 5　蓝方行为指标数据

Table 5　Blue square behavior indicator data

编号	运动行为	载频/MHz	占空比/%	T_s/μs	T_e/μs	N
B1	中低速接近	9 800	40	1	3 071	512
B2	中低速接近	9 200	20	7 910.00	9 446.00	1 024
B3	高速接近	9 100	20	4 510	6 097.2	512
B4	高速接近	9 500	40	3	3 071	1 024
B5	盘旋	9 000	20	1 340.2	2 799.4	512
B6	盘旋	1 500	2	4 670	11 070.00	64

表 6　蓝方行为模糊评价语言与 Vague 转换

Table 6　Fuzzy evaluation language of blue square behavior and Vague transition

编号	运动行为	波束行为	参数行为	情报行为
B1	[0.39,0.69]	[0.22,0.78]	[0.39,1]	[0.82,1]
B2	[0.39,0.69]	[0.72,1]	[0.39,1]	[0.56,0.76]
B3	[0.7,1]	[0.22,0.78]	[0,0.5]	[0.56,0.76]
B4	[0.7,1]	[0.72,1]	[0.39,1]	[0.82,1]
B5	[0.28,0.73]	[0,0.28]	[0,0.5]	[0.72,0.86]
B6	[0.28,0.73]	[0,0.28]	[0,0.36]	[0,0.23]

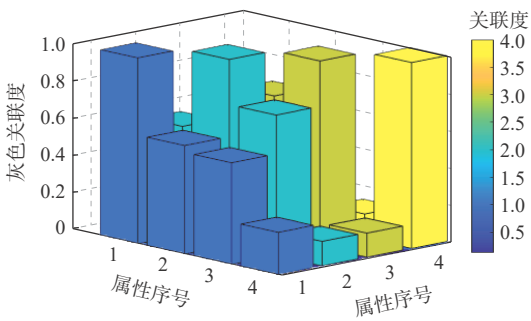


图 10　属性指标间的关联程度

Fig. 10　The degree of correlation between attribute indicators

表 7　行为评估初始权重

Table 7　Initial weights of behavior evaluation

行为	权重
运动行为	0.251 3
波束行为	0.260 6
参数行为	0.312 3
情报行为	0.175 8

根据变权理论和各类型战机的性能参数及先验情报信息，其状态函数的变化趋势如图 11 所示。

根据式(30)，权重随距离的变化趋势如图 12 所示。

由图 11 和图 12 可以看出，大于 150 km 时，B1 和 B4 的运动行为的指标权重大于辐射源行为指标权重。小于 150 km 时，B1 和 B4 的运动行为的指标权重小于辐射源行为指标权重战机运动行为的指标权重和辐射源行为指标权重大小转折点在 150 km 处。对于威胁目标 B2 和 B3 战机，运动行为

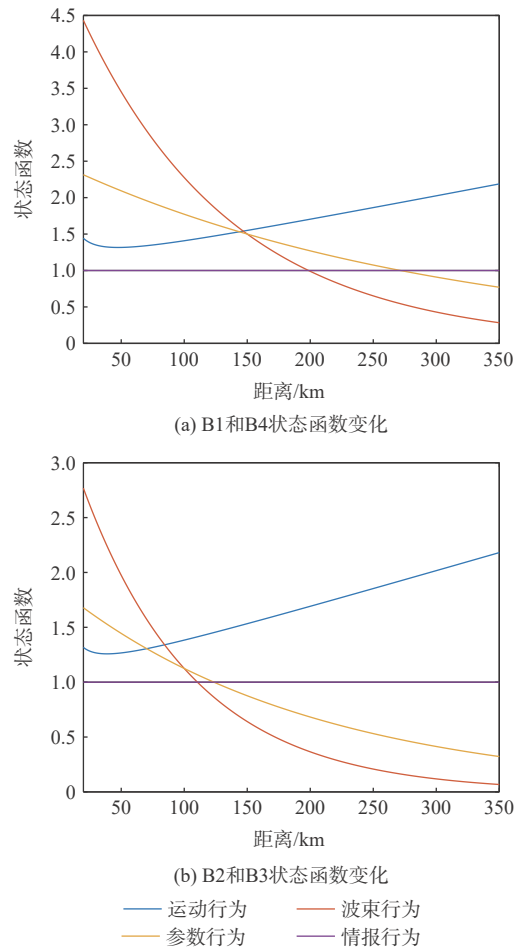


图 11　状态函数的变化趋势

Fig. 11　Change trend of state function

指标权重和辐射源指标权重的转折点大约在 100 km 处。与情报信息相比，属性权重的变化转折距离与各战机挂载导弹的最大发射距离大致相同。

分别选取距离威胁目标 350 km 和 50 km 处的属性权重，画出改进雷达图，如图 13 和图 14 所示。

根据式(21)和式(32)计算出各威胁目标的威胁度，如表 8 所示，威胁度排序示意图如图 15 所示。

由表 8 和图 15 可以看出，在 350 km 时，B3、B4 整体编队威胁度高于 B1、B2 整体编队，与态势信息相比，B3、B4 整体编队的运动行为表现高速接近。在 50 km 时，B4 的威胁度高于同编队的 B3，B2 的威胁度高于同编队的 B1，与态势信息相比，

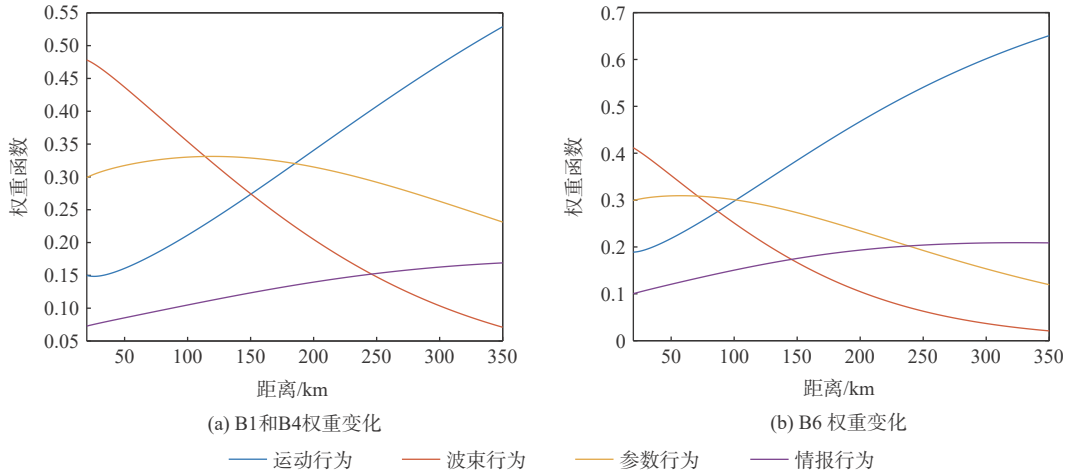


图 12 子行为权重的变化趋势

Fig. 12 Change trend of sub-behavior weights

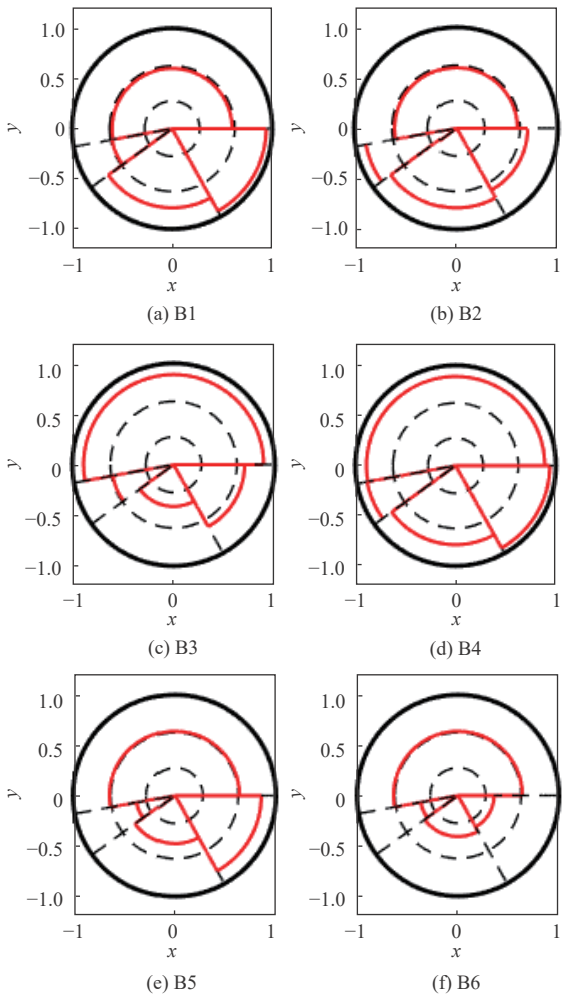


图 13 350 km 处各威胁目标雷达图

Fig. 13 Radar map of each threat target at 350 km

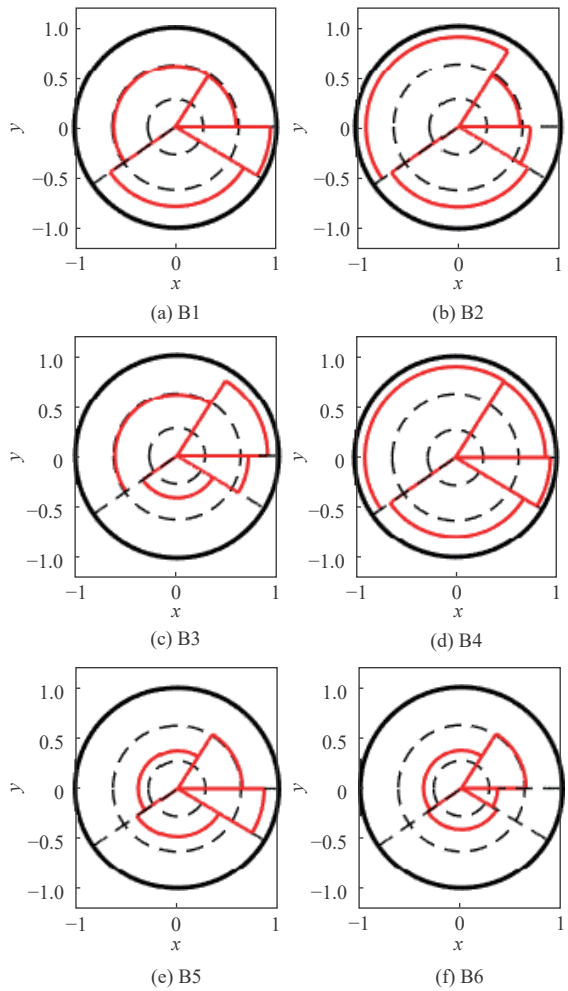


图 14 50 km 处各威胁目标雷达图

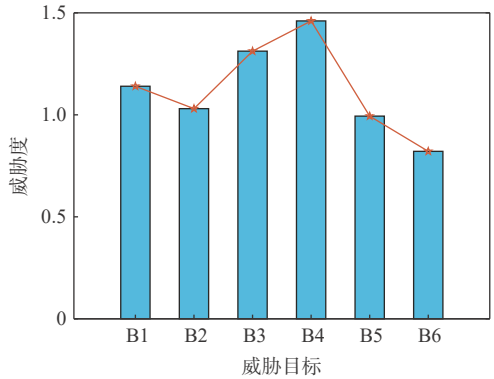
Fig. 14 Radar map of each threat target at 50 km

B4、B2 的辐射源行为极有可能进入跟踪状态。对比 350 km 和 50 km 的排序结果，发现在空对空制导导弹攻击范围之外时，威胁目标的威胁度主要取决于运动行为；当在空空导弹攻击范围之内时，威胁目标的威胁度主要取决于辐射源行为。

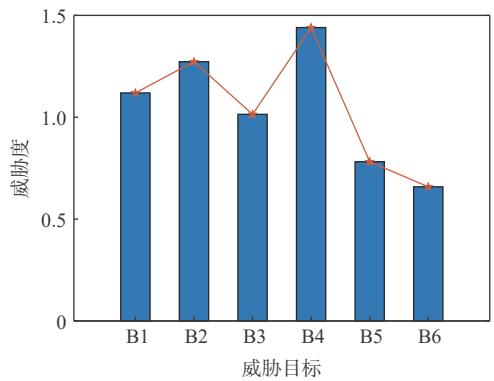
表 8 各威胁目标威胁度

Table 8 Threat level of each threat target

目标距离/km	威胁度					
	B1	B2	B3	B4	B5	B6
350	1.14	1.03	1.31	1.46	0.99	0.821
50	1.119	1.272	1.014	1.44	0.781	0.658



(a) 350 km威胁度柱状图



(b) 50 km威胁度柱状图

图 15　目标威胁度柱状图

Fig. 15　Histogram of target threat degree

为更好地验证本文算法的正确性和合理性，将其与常权初始权重算法[20]、贝叶斯算法[22]和 TOPSIS 算法[28]进行对比，对比结果如图 16 所示。

从结果中可以看出，常权初始数量算法计算出的 B1 和 B2 的威胁程度相同，TOPSIS 算法和本文算法在 50 km 处的威胁排序相同。这是因为：TOPSIS 算法主要计算数据的相对位置关系，因此容易受到指标间的关联影响。由于本文中有 2 个属性指标属于辐射源因素，所以在计算中加大了辐射源的权重。相较于 TOPSIS 算法和贝叶斯算法，本文算法能够使用精确度较低的数据得到相同的威胁评估，同时还可以根据态势的变化更加准确地反映出各目标的威胁程度。

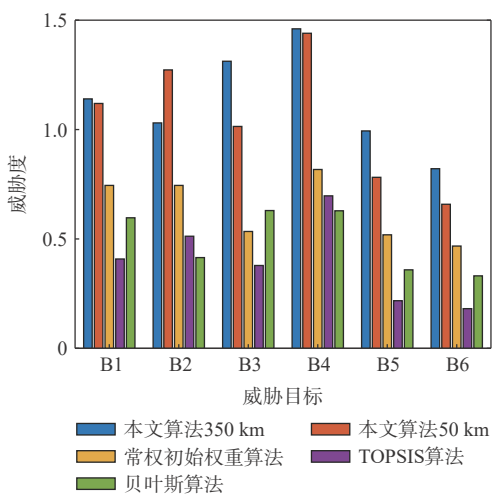


图 16　4 种威胁评估算法对比

Fig. 16　Comparison of four threat assessment algorithms

6　结　论

1）本文建立了基于行为特征的威胁评估层次指标模型，将对威胁目标的速度、高度、航向角等参数描述转化为态势行为、辐射源行为和情报行为的抽象描述，具备较强的泛化能力和通用性。

2）本文建立了各子行为特征的模糊子集和隶属度函数，通过仿真可知，隶属度函数能够准确地定量描述每个行为特征的威胁程度，相较于传统算法，本文算法减小了对数据精确度的依赖。

3）本文建立的特征权重计算算法能够消除特征之间的耦合性，使得威胁度计算结果更为精确。同时，能够准确反映现代空战的规律和飞行员判断准则，可以根据态势的变化更加准确地反映各目标的威胁程度。

4）本文算法为雷达辐射源威胁评估提供了新的思路，并具有一定的普适性。本文主要针对空空作战场景，在应用于其他作战场景时需要进一步优化和修正，以保证算法的准确性。

参考文献（References）

[1] ZHANG K, KONG W R, LIU P P, et al. Assessment and sequencing of air target threat based on intuitionistic fuzzy entropy and dynamic VIKOR[J]. Journal of Systems Engineering and Electronics, 2018, 29(2): 305-310.

[2] DENG Y. A threat assessment model under uncertain environment [J]. Mathematical Problems in Engineering, 2015, 2015: 878024.

[3] ZHAO R J, YANG F B, JI L N, et al. Dynamic air target threat assessment based on interval-valued intuitionistic fuzzy sets, game theory, and evidential reasoning methodology[J]. Mathematical Problems in Engineering, 2021, 2021: 6652706.

[4] LU Y L, LEI X L, ZHOU Z P, et al. Approximate reasoning based on IFRS and DS theory with its application in threat assessment[J]. IEEE Access, 2020, 8: 160558-160568.

[5] DI R H, GAO X G, GUO Z G, et al. A threat assessment method for unmanned aerial vehicle based on Bayesian networks under the condition of small data sets[J]. Mathematical Problems in Engineering, 2018, 2018: 8484358.

[6] HUANG L D, CAI G, YUAN H Y, et al. A hybrid approach for identifying the structure of a Bayesian network model[J]. Expert Systems with Applications, 2019, 131: 308-320.

[7] GAO Y, LI D S, ZHONG H. A novel target threat assessment method based on three-way decisions under intuitionistic fuzzy multi-attribute decision making environment[J]. Engineering Applications of Artificial Intelligence, 2020, 87: 103276.

[8] 翟翔宇. 基于全连接神经网络的空战目标威胁评估方法研究[D].

太原: 中北大学, 2020.
ZHAI X Y. Research on threat assessment method of air combat turget based on fully connected neural network[D]. Taiyuan: North University of China, 2020(in Chinese).

[9] 郭辉, 徐浩军, 刘凌. 基于回归型支持向量机的空战目标威胁评估[J]. 北京航空航天大学学报, 2010, 36(1): 123-126.
GUO H, XU H J, LIU L. Target threat assessment of air combat based on support vector machines for regression[J]. Journal of Beijing University of Aeronautics and Astronautics, 2010, 36(1): 123-126(in Chinese).

[10] ZHAO R J, YANG F B, JI L N. An extended fuzzy CPT-TODIM model based on possibility theory and its application to air target dynamic threat assessment[J]. IEEE Access, 2022, 10: 21655-21669.

[11] 奚之飞, 徐安, 寇英信, 等. 基于改进 GRA-TOPSIS 的空战威胁评估[J]. 北京航空航天大学学报, 2020, 46(2): 388-397.
XI Z F, XU A, KOU Y X, et al. Air combat threat assessment based on improved GRA-TOPSIS[J]. Journal of Beijing University of Aeronautics and Astronautics, 2020, 46(2): 388-397(in Chinese).

[12] 胡涛, 王栋, 孙曜, 等. 基于改进 CRITIC-LRA 和灰色逼近理想解排序法的空战威胁评估[J]. 兵工学报, 2020, 41(12): 2561-2569.
HU T, WANG D, SUN Y, et al. Air combat threat assessment of improved CRITIC-LRA and grey TOPSIS[J]. Acta Armamentarii, 2020, 41(12): 2561-2569(in Chinese).

[13] 杨爱武, 李战武, 李宝, 等. 基于动态变权重的空战态势评估[J]. 兵工学报, 2021, 42(7): 1553-1563.
YANG A W, LI Z W, LI B, et al. Air combat situation assessment based on dynamic variable weight[J]. Acta Armamentarii, 2021, 42(7): 1553-1563(in Chinese).

[14] ZHANG H, XIE J, SONG Y, et al. A novel ranking method for intuitionistic fuzzy set based on information fusion and application to threat assessment[J]. Iranian Journal of Fuzzy Systems, 2020, 17: 91-104.

[15] ZHANG Q, HU J H, FENG J F, et al. Air multi-target threat assessment method based on improved GGIFSS[J]. Journal of Intelligent & Fuzzy Systems, 2019, 36(5): 4127-4139.

[16] MA S D, ZHANG H Z, YANG G Q. Target threat level assessment based on cloud model under fuzzy and uncertain conditions in air combat simulation[J]. Aerospace Science and Technology, 2017, 67: 49-53.

[17] 杨远志, 王红卫, 索中英, 等. 基于粗糙集-逼近理想解排序的辐射源威胁排序方法[J]. 兵工学报, 2016, 37(5): 945-952.
YANG Y Z, WANG H W, SUO Z Y, et al. An emitter threat evaluation method based on rough set and TOPSIS[J]. Acta Armamentarii, 2016, 37(5): 945-952(in Chinese).

[18] 张晓雷, 单洁, 王刚. 基于云重心评价法的雷达辐射源威胁评估[J]. 火力与指挥控制, 2017, 42(8): 10-14.
ZHANG X L, SHAN J, WANG G. Research on threat assessment of radar radiation source based on cloud gravity center evaluation method[J]. Fire Control & Command Control, 2017, 42(8): 10-14(in Chinese).

[19] 张莹, 王红卫, 陈游. 基于 ICW-RCM 的辐射源组合威胁评估[J]. 系统工程与电子技术, 2018, 40(3): 557-562.
ZHANG Y, WANG H W, CHEN Y. Combined emitter threat assessment based on ICW-RCM[J]. Systems Engineering and Electronics, 2018, 40(3): 557-562(in Chinese).

[20] 徐宇恒, 程嗣怡, 周一鹏, 等. 基于 Shapley-TOPSIS 的辐射源威胁评估[J]. 空军工程大学学报 (自然科学版), 2020, 21(2): 91-96.
XU Y H, CHENG S Y, ZHOU Y P, et al. Radiator threat evaluating based on Shapley-TOPSIS[J]. Journal of Air Force Engineering University (Natural Science Edition), 2020, 21(2): 91-96(in Chinese).

[21] 董鹏宇, 王红卫, 陈游. 基于博弈论的 GRA-TOPSIS 辐射源威胁评估方法[J]. 北京航空航天大学学报, 2020, 46(10): 1973-1981.
DONG P Y, WANG H W, CHEN Y. GRA-TOPSIS emitter threat assessment method based on game theory[J]. Journal of Beijing University of Aeronautics and Astronautics, 2020, 46(10): 1973-1981 (in Chinese).

[22] 董鹏宇, 王红卫, 陈游. 区间条件下基于 GRA 和 TOPSIS 的辐射源威胁评估[J]. 控制与决策, 2021, 36(6): 1516-1522.
DONG P Y, WANG H W, CHEN Y. Combining TOPSIS and GRA for emitter threat evaluation with interval number[J]. Control and Decision, 2021, 36(6): 1516-1522(in Chinese).

[23] 陈振坤, 程嗣怡, 徐宇恒, 等. 数据缺失下基于 IOWA-TOPSIS 的辐射源威胁评估[J]. 空军工程大学学报 (自然科学版), 2021, 22(1): 105-111.
CHEN Z K, CHENG S Y, XU Y H, et al. A radiator threat assessment based on IOWA-TOPSIS under conditions of missing data[J]. Journal of Air Force Engineering University (Natural Science Edition), 2021, 22(1): 105-111(in Chinese).

[24] 黄旭佳. 自适应雷达行为分析与辨识技术研究[D]. 成都: 电子科技大学, 2015.
HUANG X J. Research on analysis and identification techniques for adaptive radar behavior[D]. Chengdu: University of Electronic Science and Technology of China, 2015(in Chinese).

[25] 欧健. 多功能雷达行为辨识与预测技术研究[D]. 长沙: 国防科技大学, 2017.
OU J. Research on behavior recognition and prediction techniques against multi-function radar[D]. Changsha: National University of Defense Technology, 2017(in Chinese).

[26] 肖冰松, 方洋旺, 胡诗国, 等. 一种新的超视距空战威胁评估方法[J]. 系统工程与电子技术, 2009, 31(9): 2163-2166.
XIAO B S, FANG Y W, HU S G, et al. New threat assessment method in beyond-the-horizon range air combat[J]. Systems Engineering and Electronics, 2009, 31(9): 2163-2166(in Chinese).

[27] 熊升华, 吴胜, 陈振颂, 等. 广义直觉模糊熵及其在权重确定中的应用[J]. 控制与决策, 2017, 32(5): 845-854.
XIONG S H, WU S, CHEN Z S, et al. Generalized intuitionistic fuzzy entropy and its application in weight determination[J]. Control and Decision, 2017, 32(5): 845-854(in Chinese).

[28] 张莹, 王红卫, 郭晓陶, 等. IFS-BN 结合的辐射源威胁评估方法[J]. 空军工程大学学报 (自然科学版), 2017, 18(1): 63-68.
ZHANG Y, WANG H W, GUO X T, et al. An assessment method of emitter threat based on intuitional fuzzy sets and Bayesian network[J]. Journal of Air Force Engineering University (Natural Science Edition), 2017, 18(1): 63-68(in Chinese).

Threat assessment of radar radiation sources based on behavioral characteristics

WANG Jundi[1]，WANG Xing[1]，TIAN Yuanrong[2,*]，CHEN You[1]

(1. Aviation Engineering School，Air Force Engineering University，Xi'an 710038，China；
2. School of Electronic Countermeasure，National University of Defense Technology，Hefei 230037，China)

Abstract：This paper studies the problem that the current radar radiation source threat assessment algorithm relies heavily on accurate reconnaissance data. Firstly, a radiation source threat assessment system based on behavior characteristics is constructed, starting from the radiation source target behavior characteristics and data fusion theory. Also, fuzzy theory and Vague datasets are employed to represent each sub-behavior. A situational state function with distance as an independent variable is constructed by using the modified starting weight of interval gray correlation degree to solve the coupling between indicators and the high dynamics of air warfare. Finally, an improved radar map method is adopted to calculate the threat level of the threatening target. The simulation results show that the algorithm in this paper has good accuracy and adaptability.

Keywords：data fusion；threat assessment；behavioral characteristics；dynamic weights；interval gray correlation degree；radar map method

Received：2022-10-10；**Accepted**：2022-12-25；**Published Online**：2023-01-03 17：39

URL：link.cnki.net/urlid/11.2625.V.20230103.1419.002

Foundation items：National Natural Science Foundation of China (62001489); China Postdoctoral Science Foundation (2021JM-225)

* **Corresponding author**. E-mail：tianyuanrong20@nudt.edu.cn

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2022.0790

振镜电机系统高性能数模混合驱动控制方法

赵家康，郭宏，徐金全*

(北京航空航天大学 自动化科学与电气工程学院，北京 100191)

摘 要：为满足振镜电机系统高精度和高动态性能的控制需求，提出数模混合架构和扩张状态观测器(ESO)-滑模控制(SMC)复合的高性能驱动控制方法。采用基于数字式控制器和模拟式驱动器的数模混合架构设计，一方面便于实现高性能的非线性控制算法，另一方面能够消除数字式驱动器中脉冲宽度调制(PWM)技术导致的高频电流纹波。基于ESO和SMC复合的振镜电机位置环控制方法，通过ESO估计系统的各种扰动，利用SMC实时消除扰动影响，显著地提升了振镜电机系统的动态性能和鲁棒性。仿真和实验结果表明：相较于比例-积分-微分(PID)控制器，所提方法能够有效地提升系统的鲁棒性和动态响应性能，其中动态响应性能提升了29.4%。

关 键 词：振镜电机；滑模控制；扩张状态观测器；高动态性能；运算放大器；跟踪微分器
中图分类号：V242.4; TM351
文献标志码：A **文章编号**：1001-5965(2024)10-3208-11

振镜电机系统作为一种优良的矢量扫描器件，能够快速、高精度地控制镜片偏转，在航空航天领域得到广泛应用，如航空零件加工、航空相机成像技术及激光瞄准平台等。在航空零件的加工中，振镜电机系统常被应用于增材制造领域，即将三维实体转换为二维平面的叠加，再采用二维扫描振镜进行逐层雕刻加工。航空零件一般都是复杂曲面，这对振镜电机系统的准确性提出了较高的要求，同时为避免加工过程中出现问题，一般不允许所控制系统的阶跃响应产生超调[1-3]。在航空相机成像领域中，振镜电机系统主要应用于相机的成像补偿，通过在飞行方向放置以一定角速度旋转的单轴振镜电机，能够解决在相机扫摆过程中产生的像移问题，从而使航空相机能够获取更高质量的图像[4]；而在激光设备领域中，振镜电机系统主要应用于激光设备的瞄准平台。激光设备瞄准平台采用双轴或多轴转台组合瞄准，并通过转台上的反射镜将激光光束发射至目标。由于激光设备要求发射的激光能够稳定在目标的某一位置，并且当所瞄准目标移动时，激光也应具有较好的跟踪能力。因此，激光设备的瞄准平台需要具备高动态响应及高精度[5-7]。

随着航空航天领域的不断发展，以振镜电机为核心的控制系统得到了大量的研究与广泛的应用。目前，振镜电机系统的控制器与驱动器分别具有数字式与模拟式2种类型，根均据控制器与驱动器类型的不同，形成了数字-数字架构、模拟-模拟架构及数模混合架构3种拓扑。其中，模拟式控制器由于无法采用PID算法以外的高性能控制算法已经逐渐被淘汰。数字式驱动器基于脉冲宽度调制(pulse width modulation, PWM)技术控制，其优点为系统效率高、发热小，适用于高功率应用场合，缺点是由于PWM斩波引起的电流纹波，会导致振镜电机系统的控制精度下降；模拟式驱动器以运算放大器为核心，其优点是输出电流信号光滑无纹波，便于提升系统控制性能，缺点是工作时系统损耗高、发热大，适用于小功率应用场合。

国内外对振镜电机系统进行了较为丰富的研

收稿日期：2022-09-18；**录用日期**：2022-10-10；**网络出版时间**：2022-11-04 09:47
网络出版地址：link.cnki.net/urlid/11.2625.V.20221103.1719.003
* **通信作者**. E-mail：xujinquan@buaa.edu.cn

引用格式：赵家康，郭宏，徐金全. 振镜电机系统高性能数模混合驱动控制方法[J]. 北京航空航天大学学报，2024，50(10)：3208-3218. ZHAO J K，GUO H，XU J Q. High performance control method for galvanometer laser scanner system[J]. Journal of Beijing University of Aeronautics and Astronautics，2024，50(10)：3208-3218 (in Chinese).

究, 文献 [8] 设计了一类采用电压放大器作为驱动器的控制系统, 但其响应时间较慢, 小信号阶跃响应时间仅能达到 2 ms 左右; 文献 [9] 设计了一种模糊滑模变结构算法, 仿真实验显示, 振镜系统的小阶跃响应时间约为 200 μs, 但没有进行实物验证; 文献 [10] 设计了一类扫描用振镜电机系统, 并采用自抗扰控制算法得到了正向全行程响应时间为 15.24 ms, 该系统具有一定的抗干扰能力; 文献 [11] 设计了一类分数阶鲁棒控制器, 这对于系统中摩擦导致的参数不确定性有较好的鲁棒性, 但是其阶跃响应有一定程度的超调; 文献 [12] 设计了一类数字式振镜电机系统, 采用自抗扰控制算法提升了系统的抗干扰能力, 通过实验得出 0.3°阶跃响应时间为 300 μs; 文献 [13] 在传统的 PID 控制算法的基础上额外加入了 PID-L_1 控制器, 提高了系统的响应速度及抗干扰能力, 但是只进行了仿真实验且阶跃响应有一定程度的超调。

本文提出一种数模混合振镜电机系统高性能驱动控制方法。首先, 对振镜电机系统的数学模型及架构进行分析, 提出了采用跨导放大器作为驱动器的数模混合驱动控制架构; 其次, 研究振镜电机系统双闭环高性能控制方法, 提出位置环采用基于扩张状态观测器 (extended stata observer, ESO)-滑模控制 (sliding mode control, SMC) 复合方法的控制策略; 最后, 通过仿真和实验验证了本文方法的有效性。

1　振镜电机系统的数学模型及架构分析

为满足振镜电机系统高精度和高动态性能的控制需求, 本文提出一种基于数模混合架构, 以及 ESO 与 SMC 复合的高性能驱动控制方法, 振镜电机系统框架如图 1 所示。

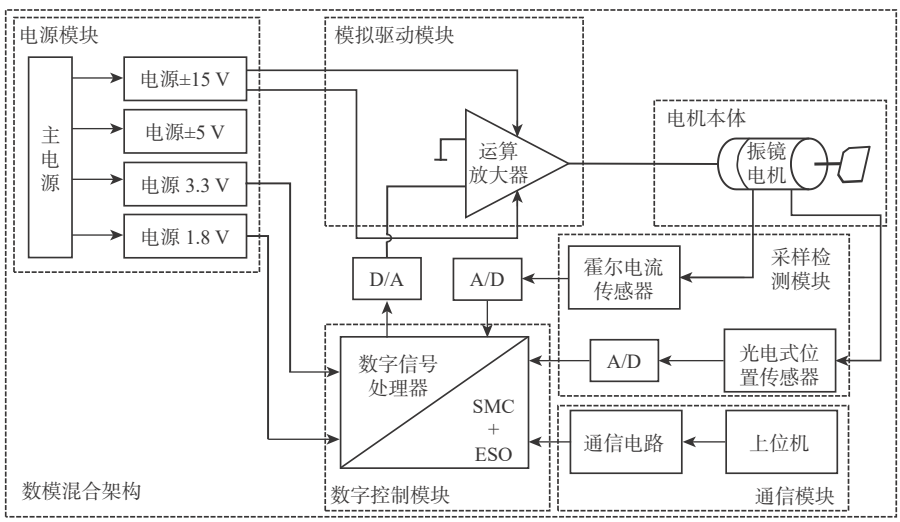


图 1　振镜电机系统框架

Fig. 1　Frame of galvanometer laser scanner

1.1　振镜电机数学模型

振镜电机作为振镜电机系统的执行机构, 是系统中最为重要的部分之一, 其结构、性能的不同都会对振镜系统整体的精度、速度等参数产生直接影响。本文选择动磁式结构的振镜电机, 该结构外圈为定子铁芯, 由铁磁材料构成, 内圈为转子, 由高磁能密度的稀土永磁体加工而成, 通过改变线圈中通过的电流来控制磁场的强弱。动磁式结构的优点在于能够不受温度、湿度变化的影响, 同时, 由于转子本身为永磁体, 可以通过磁场直接控制线圈的偏转, 因此可以解决铁芯饱和的问题[14-15]。

此类振镜电机可以视为一个带有角度限位的直流电机, 根据基尔霍夫电压定律, 可以得到动磁式振镜电机的电枢方程为

$$U_a = L\frac{\mathrm{d}i_a}{\mathrm{d}t} + Ri_a + E \tag{1}$$

式中: U_a为电枢绕组两端的电压; i_a为流过振镜电机的电流; R为振镜电机的电阻; L为振镜电机的电感; E为振镜电机的反电动势:

$$E = K_e\frac{\mathrm{d}\theta}{\mathrm{d}t}\cos\theta \tag{2}$$

式中: K_e为振镜电机的反电动势系数; θ为振镜电机转子的位置; 由于本文所使用的振镜电机的最大行程为 30°, 即 $\theta\in[-15°,+15°]$, $\cos\theta\in[0.965\,9,1]$。因此, 可以基本忽略振镜电机位置变化对反电动势的影响, 即令$\cos\theta=1$, 反电动势公式可以简化为

$$E = K_e\frac{\mathrm{d}\theta}{\mathrm{d}t} \tag{3}$$

振镜电机的转矩平衡方程为

$$T_e = J\frac{d^2\theta}{dt^2} + B_m\frac{d\theta}{dt} + T_d \tag{4}$$

式中：T_d为振镜电机的外部干扰力矩；J为振镜电机的转动惯量；B_m为振镜电机的粘性阻尼系数；T_e为振镜电机的电磁转矩，根据直流电机中电磁转矩与电枢电流的关系，还可以表示为

$$T_e = K_t i_a \tag{5}$$

式中：K_t为振镜电机的转矩系数。为便于后续分析，将振镜电机系统化为状态方程，令$x_1 = \theta$，$x_2 = \omega$，其中，ω为振镜电机的角速度。将式(1)、式(4)、式(5)联立，可得

$$\begin{cases} \dot{x}_1 = x_2 \\ \dot{x}_2 = \dfrac{K_t}{J}i_a - \dfrac{B_m}{J}x_2 - \dfrac{T_d}{J} \\ y = x_1 \end{cases} \tag{6}$$

式中：y为振镜电机控制系统的输出，即振镜电机的实际位置。

1.2 振镜电机驱动系统数学模型

在振镜电机系统中，驱动模块能够将控制器计算出的控制信号转化为输入电机的电压信号或电流信号，从而实现对振镜电机的控制。振镜电机系统的驱动器一般有数字式和模拟式2种类型，数字式驱动器以PWM为核心，其优点在于功率损耗较小，但是，该斩波驱动方式将高频谐波引入系统，会造成电机侧的力矩波动。同时，由于电流脉动与死区效应，采用数字式驱动方式的输出电流会产生偏差，从而导致系统精度较低。模拟式驱动器以功率线性运算放大器为核心，具有信号质量高、电流线性度好、纹波小、对外界电磁干扰小等优点，也不会存在数字驱动电路中常见的高频干扰问题。但是由于其发热较为严重，因此仅适用于功率较小的场合。

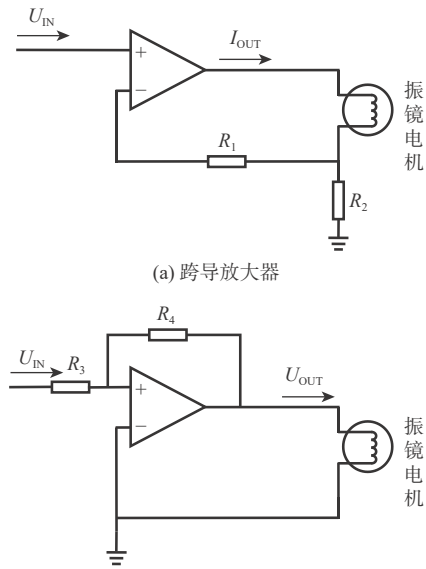
本文所设计的振镜电机系统为小功率器件，并选用功率线性运算放大器作为驱动器。功率线性运算放大器一般采用负反馈的连接方式，能够形成电压串联式、电压并联式、电流串联式和电流并联式4种基本组态，基本信息如表1所示。

如表1所示，这4种组态分别能够实现电压放大器、跨导放大器、跨阻放大器和电流放大器的功能。振镜电机系统控制器输出的数字信号经过A/D转换器转换为模拟电压信号作为驱动器的输入，因此，选用电压串联式或电压并联式的连接方法以实现电压放大器或跨导放大器的功能。这2类运算放大器的拓扑结构如图2所示。

表1 运算放大器的4种组态

Table 1 Four forms of operational amplifier

组态	输入信号	输出信号	放大公式
电压串联式	U_i	U_o	$K_v = U_o/U_i$
电压并联式	U_i	I_o	$K_s = I_o/U_i$
电流串联式	I_i	U_o	$K_\Omega = U_o/I_i$
电流并联式	I_i	I_o	$K_i = I_o/I_i$



(a) 跨导放大器

(b) 电压放大器

图2 驱动器的拓扑结构

Fig. 2 Topology of the drivers

图2(a)的跨导放大器将输入电压U_{IN}转换为恒定电流I_{OUT}并输出至振镜电机，其放大倍数取决于电阻R_2的阻值，设K_s为电压并联式运算放大器的放大倍数，则放大公式为

$$K_s = \frac{1}{R_2} \tag{7}$$

图2(b)的电压放大器将输入电压U_{IN}转换为U_{OUT}以实现电压放大功能，设K_v为电压串联式运算放大器的放大倍数，则放大公式为

$$K_v = \frac{R_4}{R_3} \tag{8}$$

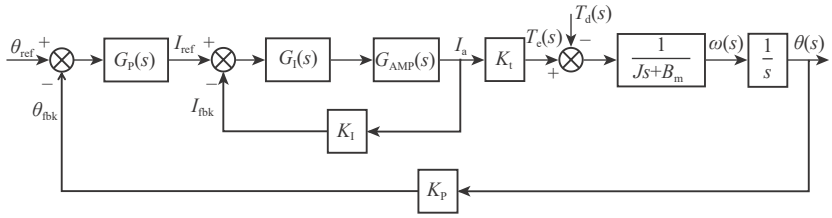
图2所示的2类运算放大器的传递函数可简化为一个比例环节：

$$G_v(s) = K_v \tag{9}$$

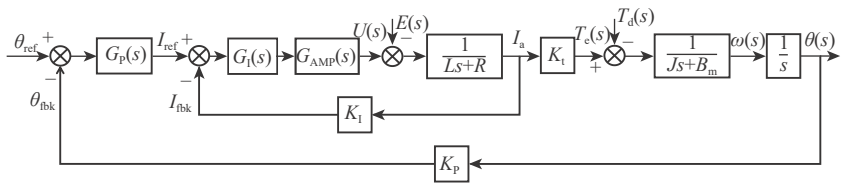
$$G_s(s) = K_s \tag{10}$$

本文所设计的振镜电机系统采用双闭环结构，内环为电流环、外环为位置环，由式(1)～式(8)可以得到如图3所示的以跨导放大器或电压放大器作为驱动器的振镜电机系统结构框图。

图3中，K_P为位置传感器反馈系数；K_I为电流传感器反馈系数。根据这2类振镜电机系统的结构框图，设$K_P = 1$、$K_I = 1$，外部无干扰力矩，即$T_d = 0$，则将电压放大器作为驱动器的振镜电机系



(a) 以跨导放大器作为驱动器



(b) 以电压放大器作为驱动器

图 3　振镜电机系统的结构框图

Fig. 3　Structural block diagram of the galvanometer laser scanner

统传递函数为

$$\begin{cases}\dfrac{\theta_{\mathrm{fbk}}}{\theta_{\mathrm{ref}}}=\dfrac{H_{\mathrm{v}}(s)}{LJs^3+G_1(s)s^2+G_2(s)B_{\mathrm{m}}s+H_{\mathrm{v}}(s)}\\ G_1(s)=JK_{\mathrm{v}}G_{\mathrm{I}}(s)+RJ+LB_{\mathrm{m}}\\ H_{\mathrm{v}}(s)=K_{\mathrm{t}}K_{\mathrm{v}}G_{\mathrm{I}}(s)G_{\mathrm{P}}(s)\\ G_2(s)=R+K_{\mathrm{v}}G_{\mathrm{I}}(s)\end{cases}\tag{11}$$

以跨导放大器作为驱动器的振镜电机系统传递函数为

$$\begin{cases}\dfrac{\theta_{\mathrm{fbk}}}{\theta_{\mathrm{ref}}}=\dfrac{H_{\mathrm{s}}(s)}{[1+K_{\mathrm{s}}G_{\mathrm{I}}(s)](Js^2+B_{\mathrm{m}}s)+H_{\mathrm{s}}(s)}\\ H_{\mathrm{s}}(s)=K_{\mathrm{t}}K_{\mathrm{s}}G_{\mathrm{I}}(s)G_{\mathrm{P}}(s)\end{cases}\tag{12}$$

式中: $G_{\mathrm{P}}(s)$为位置环控制器; $G_{\mathrm{I}}(s)$为电流环控制器。根据图 3, 对采用 2 类驱动器搭建的系统进行仿真实验, 电流环与位置环均采用 PID 控制器, 并分别将其控制参数调至最优。当位置环的输入为 0.085°的阶跃信号时, 2 类驱动器的位置环响应如图 4 所示。

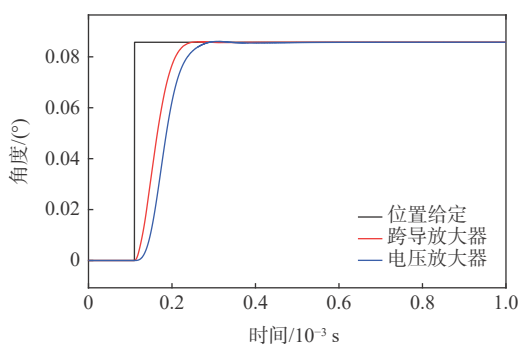


图 4　2 类驱动器的位置环响应

Fig. 4　Position loop response of two types of drivers

由图 4 可知, 相较于采用电压放大器作为驱动器, 将跨导放大器作为驱动器的小量程阶跃响应时间减少了 18.7%。

除了能够对电机电流进行直接控制从而提升振镜电机系统的动态性能, 外以跨导放大器为驱动器的驱动方式还能够提高振镜电机系统对振镜电机中参数变化的鲁棒性。由于在实际工程应用中, 振镜电机的电机电阻会随温度变化、电机电感及磁路饱和状态而发生变化, 因此, 对电机电阻及电机电感进行了参数摄动实验。所采用振镜电机为金海创公司的 7210X 型电机, 该电机电阻为3.55 Ω, 电机电感为 170 μH。对振镜电机电阻进行 50% 参数摄动实验, 即令电机电阻为1.775 Ω、对振镜电机电阻进行 50% 参数摄动实验, 即令电机电感为85 μH, 仿真结果分别如图 5 和图 6 所示。

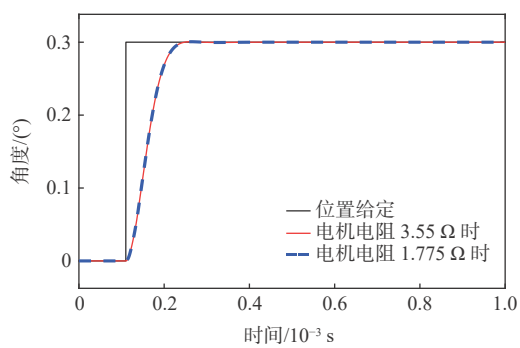


图 5　电机电阻参数摄动

Fig. 5　Perturbation of motor resistance parameters

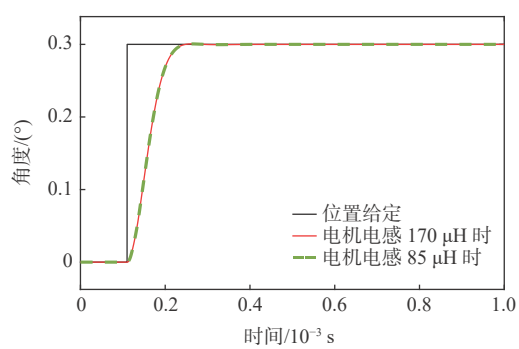


图 6　电机电感参数摄动

Fig. 6　Perturbation of motor inductance parameters

由式(12)可得，跨导放大器在实际应用中可以等效为一恒流源，因此，当振镜电机电阻阻值随温度的变化而变化时，或者电机电感随磁路的饱和状态发生变化时，虽然会对驱动器的输出电流产生影响，但不会对振镜电机系统的位置环响应产生影响。由于将跨导放大器作为驱动器时的动态性能与对电机本体参数鲁棒性的优势，因此选择跨导放大器作为本文振镜电机系统的驱动器。

2 本文方法

随着振镜技术的不断发展，对于振镜电机动态性能的要求也不断提升。本文所设计的振镜电机系统为双闭环控制系统，由于在实际工况中，振镜电机系统为一个非理想系统，因此各类干扰因素均会对振镜电机系统的性能产生影响。该方法通过带ESO估计系统各种扰动，利用SMC实时消除扰动的影响，其结构框图如图7所示。

2.1 SMC设计

SMC在物理上实现简单，响应快速，对系统的参数变化不灵敏，能够有效提升系统的动态响应性能及鲁棒性[16-19]。根据1.1节所获得的振镜电机系统的数学模型，定义振镜电机镜片位置的给定值与实际值的误差为

$$e=\theta_{\mathrm{d}}-\theta_{\mathrm{r}} \tag{13}$$

式中：θ_{d}为振镜电机从上位机得到的镜片位置的给定值；θ_{r}为通过位置传感器所测得的镜片位置的实际值；根据式(13)分别对时间进行一次求导及二次求导后可以得到振镜电机角速度的误差与加速度的误差分别为

$$\begin{cases}\dot{e}=\dot{\theta}_{\mathrm{d}}-\dot{\theta}_{\mathrm{r}}\\ \ddot{e}=\ddot{\theta}_{\mathrm{d}}-\ddot{\theta}_{\mathrm{r}}\end{cases} \tag{14}$$

根据式(13)和式(14)，设计滑模面函数为

$$s_{\mathrm{smc}}=ce+\dot{e} \tag{15}$$

式中：c为滑模面的调节参数，且$c>0$，对滑模面函数求导后，可得

$$\dot{s}_{\mathrm{smc}}=c\dot{e}+\ddot{e}=c(\dot{\theta}_{\mathrm{d}}-\dot{\theta}_{\mathrm{r}})+(\ddot{\theta}_{\mathrm{d}}-\ddot{\theta}_{\mathrm{r}}) \tag{16}$$

为使系统状态能够在有限时间内快速趋近滑模面，并缓解SMC中固有的抖振问题，本文选择采用改进指数趋近律的SMC算法，趋近律的表达式为

$$\dot{s}_{\mathrm{smc}}=c\dot{e}+\ddot{e}=-\varepsilon\mathrm{sat}(s_{\mathrm{smc}})-k_{\mathrm{s}}s_{\mathrm{smc}} \tag{17}$$

$$\mathrm{sat}(s_{\mathrm{smc}},\zeta)=\begin{cases}1 & s_{\mathrm{smc}}>\zeta\\ \dfrac{s_{\mathrm{smc}}}{\zeta} & -\zeta\leqslant s_{\mathrm{smc}}\leqslant\zeta\\ -1 & s_{\mathrm{smc}}<-\zeta\end{cases} \tag{18}$$

式中：k_{s}为决定系统状态达到滑模面前趋近速度的参数，$k_{\mathrm{s}}>0$；ε为决定系统状态达到滑模面时趋近速度的参数，$\varepsilon>0$；$\mathrm{sat}(s_{\mathrm{smc}},\zeta)$为饱和函数；$\zeta$为饱和函数的线性段参数。

相较于传统的指数趋近律，改进方法采用饱和函数$\mathrm{sat}(s_{\mathrm{smc}},\zeta)$替代了符号函数$\mathrm{sgn}(s_{\mathrm{smc}})$，该函数使系统的状态在到达滑模面时能够更平滑地靠近，从而缓解SMC中的抖振问题。定义u_0为振镜电机系统的SMC律，则根据式(6)、式(15)～式(18)，可得

$$u_0=\frac{J}{K_{\mathrm{t}}}\left(\varepsilon\mathrm{sat}(s_{\mathrm{smc}})+k_{\mathrm{s}}s_{\mathrm{smc}}+c\dot{e}+\ddot{\theta}_{\mathrm{d}}+\frac{T_{\mathrm{d}}}{J}+\frac{B_{\mathrm{m}}}{J}x_2\right) \tag{19}$$

2.2 ESO设计

SMC虽然对作用于系统的外部扰动有一定的抑制作用，但是这种作用仅限于小干扰信号，当突加扰动不匹配时，仅靠SMC不能达到很好的抑制效果。ESO以系统中可能对输出产生影响的扰动信号作为一个新的状态变量，并对其进行实时跟踪并反馈至原系统中，以实现对扰动的实时补偿，从而有效提高系统的抗干扰能力[20-22]。因此，在本文中选择非线性ESO对系统中的综合扰动进行观测，可以进一步提升系统的抗干扰能力。

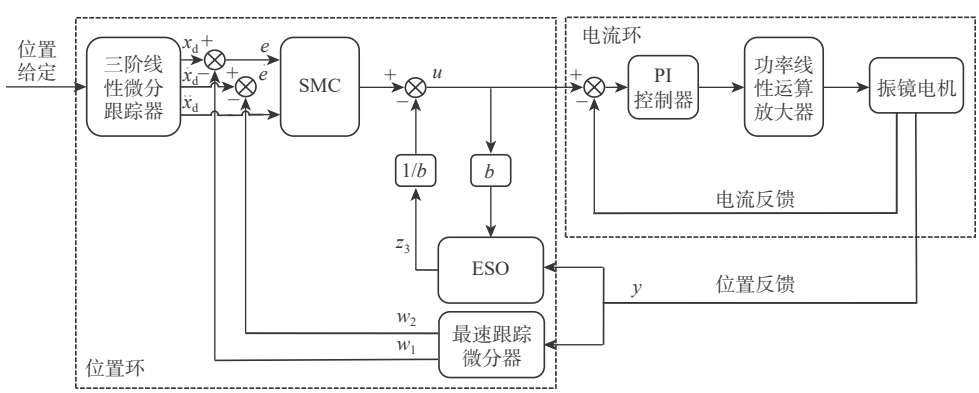


图7 本文方法结构框图

Fig. 7 Structure block diagram of the proposed method

定义系统中由模型误差、参数变化、扰动力矩等引起的扰动总和为

$$f(t)=-\frac{T_{\mathrm{d}}}{J}+d(t) \tag{20}$$

式中: $f(t)$为振镜电机系统中的扰动总和; $d(t)$为系统中的未知扰动。将式 (20) 代入式 (6), 可得

$$\begin{cases}\dot{x}_1=x_2\\ \dot{x}_2=bi_{\mathrm{a}}-ax_2+f(t)\\ y=x_1\end{cases} \tag{21}$$

式中, $b=K_{\mathrm{t}}/J$、$a=B_{\mathrm{m}}/J$为 ESO 的参数。将系统内外的扰动总和定义为一新变量x_3, 并定义$\dot{x}_3=w(t)$, 可将式 (15) 扩充为一个新的线性系统:

$$\begin{cases}\dot{x}_1=x_2\\ \dot{x}_2=bi_{\mathrm{a}}-ax_2+f(t)\\ \dot{x}_3=w(t)\\ y=x_1\end{cases} \tag{22}$$

定义$u=i_{\mathrm{a}}$为系统的控制量, 可以针对振镜电机系统设计三阶非线性 ESO 以对系统中的位置信号及角速度信号进行估计, 并对综合扰动进行补偿, 观测器的表达式为

$$\begin{cases}e=z_1-y\\ \dot{z}_1=z_2-\beta_1 e\\ \dot{z}_2=-az_2+z_3+bu-\beta_2\mathrm{fal}_1(e)\\ \dot{z}_3=-\beta_3\mathrm{fal}_2(e)\end{cases} \tag{23}$$

$$\mathrm{fal}(e,\alpha,\delta)=\begin{cases}\dfrac{e}{\delta^{\alpha-1}} & |e|\leqslant\delta\\ |e|^{\alpha}\mathrm{sgn}(e) & |e|>\delta\end{cases} \tag{24}$$

式中: z_1为振镜电机位置的观测值; z_2为振镜电机角速度的观测值; z_3为系统内外扰动总和的观测值; e为振镜电机位置的实际值与观测值的误差; β_1, β_2、β_3为非线性 ESO 的参数; $\mathrm{fal}(e,\alpha,\delta)$为原点附近具有线性段的连续幂次函数; α、δ为待调参数, 其中, α为一个0 ~ 1的常数, 取值越小, 观测器的跟踪速度越快, 但跟踪性能会变差, 而取值较大时, 观测器的跟踪速度会变慢, 但跟踪性能会变好; δ为一个与跟踪性能有关的参数, 一般取采样步长的整数倍。引入非线性函数$\mathrm{fal}(e,\alpha,\delta)$, 可以较好地抑制尖峰效应, 从而达到更好的观测效果。

由于本文所设计的振镜电机系统采用数字式控制器, 因此, 在实际应用时应采用离散式非线性 ESO:

$$\begin{cases}e(k)=z_1(k)-y(k)\\ z_1(k+1)=z_1(k)+h[z_2(k)-\beta_1 e(k)]\\ z_2(k+1)=z_2(k)+h[-az_2(k)+\\ \qquad z_3(k)+bu(k)-\beta_2\mathrm{fal}_1(e)]\\ z_3(k+1)=z_3(k)+h[-\beta_3\mathrm{fal}_2(e)]\end{cases} \tag{25}$$

式中: k为采样时刻; h为采样步长。将所设计的 ESO 代入式 (19) 后, 可以得到系统改进的控制律u_1为

$$u_0=\frac{1}{b}(\varepsilon\mathrm{sat}(s_{\mathrm{smc}})+k_{\mathrm{s}}s_{\mathrm{smc}}+c\dot{e}+\ddot{\theta}_{\mathrm{d}}+ax_2-z_3) \tag{26}$$

2.3　跟踪微分器设计

由 2.2 节中所设计的控制律可知, 式 (20) 需要得到给定信号θ_{d}的一阶微分信号$\dot{\theta}_{\mathrm{d}}$与二阶微分信号$\ddot{\theta}_{\mathrm{d}}$。

由于纯微分环节在物理上无法实现, 因此, 采用跟踪微分器对其进行近似微分[23], n阶线性跟踪微分器的表达式为

$$y=\frac{r^n}{(s+r)^n} \tag{27}$$

n阶线性跟踪微分器可以对给定信号、给定信号的一阶微分信号直至$n-1$阶微分信号进行跟踪。对于本文中的振镜电机系统, 可以选择三阶线性跟踪微分器对给定信号进行跟踪:

$$\begin{cases}\dot{v}_1=v_2\\ \dot{v}_2=v_3\\ \dot{v}_3=-r^3v_3-r^2v_2-rv_1+r^3\theta_{\mathrm{d}}\end{cases} \tag{28}$$

式中: θ_{d}为上位机给定的位置指令; v_1为给定位置指令的跟踪信号; v_2为给定位置指令一阶微分的跟踪信号; v_3为给定位置指令二阶微分的跟踪信号; r为控制参数, 决定了跟踪器的跟踪性能。因此在实际的振镜电机系统中, 微分跟踪器同样应采用离散形式:

$$\begin{cases}v_1(k+1)=v_1(k)+hv_2(k)\\ v_2(k+1)=v_2(k)+hv_3(k)\\ v_3(k+1)=v_3(k)+h(-r^3v_3(k)-r^2v_2(k)-\\ \qquad rv_1(k)+r^3\theta_{\mathrm{d}}(k))\end{cases} \tag{29}$$

线性三阶跟踪微分器能够较为有效地跟踪给定信号、一阶微分信号和二阶微分信号。但是, 由于振镜电机系统是非理想系统, 其位置反馈信号常会受到噪声污染, 当以被噪声污染的信号作为输入时, 前文所设计的跟踪微分器的跟踪效果也会变差, 甚至会将噪声放大并引入控制系统, 最终导致系统的性能变差。

为解决此问题, 并以最快的速度跟踪微分信号, 选择最速跟踪微分器来代替传统的线性跟踪微分器[18]。最速跟踪微分器的原理是采用最速综合控制函数$\mathrm{fhan}(\cdot)$, 使跟踪信号与输入信号的差值能够快速到达原点。定义等价公式如下:

$$\mathrm{fsg}(x,y)=\frac{\mathrm{sgn}(x+y)-\mathrm{sgn}(x-y)}{2} \tag{30}$$

函数fhan(x_1,x_2,r_1,h)的表达式为

$$\begin{cases} d=r_1h_0^2 \\ a_0=h_0x_2 \\ y_0=x_1+a_0 \\ a_1=\sqrt{d(d+8|y_0|)} \\ a_2=a_0+\mathrm{sgn}(y_0)(a_1-d)/2 \\ a_3=(a_0+y_0)\mathrm{fsg}(y_0,d)+a_2(1-\mathrm{fsg}(y_0,d)) \\ \mathrm{fhan}(x_1,x_2,r_1,h)=-r_1\left(\dfrac{a_3}{d}\right)\mathrm{fsg}(a_3,d)- \\ \qquad r_1\mathrm{sgn}(a_3)(1-\mathrm{fsg}(a_3,d)) \end{cases} \tag{31}$$

式中：d、a_0、y_0、a_1、a_2、a_3为求解最速综合控制函数时的中间变量；r_1为跟踪系数，其决定了跟踪速度；h_0为滤波因子；sgn(·)为符号函数。对于离散系统，二阶最速跟踪微分器的表达式为

$$\begin{cases} F_f(k)=\mathrm{fhan}(w_1(k)-u_{\mathrm{fhan}}(k),w_2(k),r_0,h_0) \\ w_1(k+1)=w_1(k)+hw_2(k) \\ w_2(k+1)=w_2(k)+hF_f(k) \end{cases} \tag{32}$$

式中：F_f为最速跟踪器微分器的控制函数；u_{fhan}为最速跟踪微分器的输入信号；w_1为u_{fhan}的跟踪信号，w_2为输入信号一阶微分的跟踪信号；h为采样步长。

根据工程经验，由于减小h或增大h_0可以有效地增强滤波效果，但是也会增加输出跟踪值的延迟相位，因此需根据实际应用合理地设置参数。将所设计的最速跟踪微分器代入式(26)可以得到系统的最终控制律u_{ed}为

$$u_{\mathrm{ed}}=\frac{1}{b}(\varepsilon\mathrm{sat}(s_{\mathrm{smc}})+k_{\mathrm{s}}s_{\mathrm{smc}}+c(v_2-w_2)+\ddot{\theta}_{\mathrm{d}}+ ax_2+v_3-z_3) \tag{33}$$

2.4 稳定性证明

针对本文方法，采用李雅普诺夫判据对其进行稳定性证明。取李雅普诺夫函数为

$$V(s_{\mathrm{smc}})=\frac{1}{2}{s_{\mathrm{smc}}}^2 \tag{34}$$

对式(34)求导，可得

$$\dot{V}(s_{\mathrm{smc}})=s_{\mathrm{smc}}\dot{s}_{\mathrm{smc}} \tag{35}$$

将式(15)和式(16)代入式(35)，可得

$$\begin{aligned} \dot{V}(s_{\mathrm{smc}})&=s_{\mathrm{smc}}\dot{s}_{\mathrm{smc}}=s_{\mathrm{smc}}(c\dot{e}+\ddot{e})=s_{\mathrm{smc}}(c\dot{e}+\ddot{\theta}_{\mathrm{d}}-\dot{x}_2)= \\ &s_{\mathrm{smc}}(c\dot{e}+\ddot{\theta}_{\mathrm{d}}-bu_0-f(t))=-s_{\mathrm{smc}}\varepsilon\mathrm{sat}(s_{\mathrm{smc}})- \\ &k_{\mathrm{s}}{s_{\mathrm{smc}}}^2+s_{\mathrm{smc}}(z_3-f(t))\leqslant-s_{\mathrm{smc}}\varepsilon\mathrm{sat}(s_{\mathrm{smc}})- \\ &k_{\mathrm{s}}{s_{\mathrm{smc}}}^2+|s_{\mathrm{smc}}||z_3-f(t)| \end{aligned} \tag{36}$$

式中：$-s_{\mathrm{smc}}\varepsilon\mathrm{sat}(s_{\mathrm{smc}})\leqslant 0$，$-k_{\mathrm{s}}{s_{\mathrm{smc}}}^2\leqslant 0$，$f(t)=x_3$且$z_3\to x_3$，故$|z_3-f(t)|\to 0$，因此，在有限时间内，$\dot{V}(s_{\mathrm{smc}})\leqslant 0$，故$\dot{V}(s_{\mathrm{smc}})$负定。同时，当$s_{\mathrm{smc}}\to\infty$时，$V(s_{\mathrm{smc}})\to\infty$，根据李雅普诺夫稳定性判据可得，所设计的控制器是大范围渐近稳定的。

2.5 仿真分析

根据2.1～2.4节的研究内容，对采用本文方法中的振镜电机系统进行仿真实验。

对所设计的振镜电机系统进行1%量程阶跃响应实验，由于本文选用的振镜电机最大摆角为±15°，故阶跃信号的给定幅值为0.3°，实验结果如图8所示。

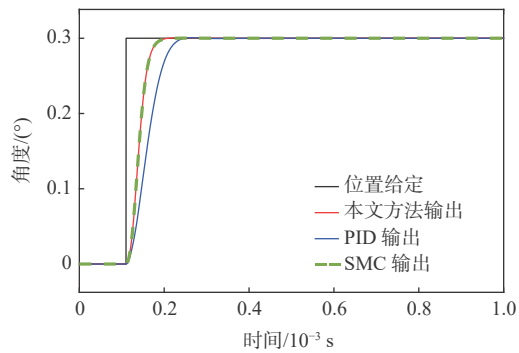


图8 1%量程阶跃响应

Fig. 8 1% span step response

由图8可知，采用SMC算法与采用本文方法中的1%量程阶跃的响应性能没有区别，响应时间均为120 μs，采用PID算法时响应时间为170 μs，可知，采用SMC算法与采用本文方法均比采用传统PID算法响应速度提高了约29.4%。

对振镜电机系统的正弦响应进行研究，将幅值0.3°、频率500 Hz的正弦信号作为给定信号时，实验结果如图9所示。

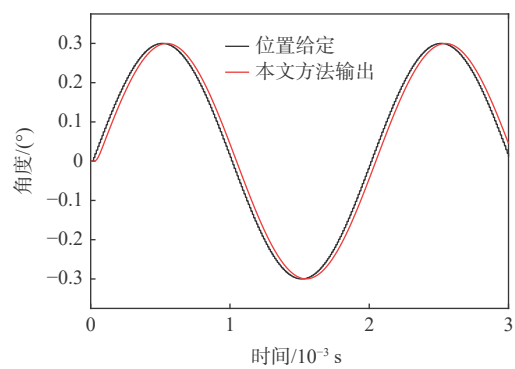


图9 位置环正弦响应

Fig. 9 Sinusoidal response of position loop

由图9可知，所设计系统对500 Hz的正弦信号具有较好的跟踪性能。完成阶跃响应实验及正弦响应实验后，对以上3种控制算法进行抗干扰性能测试。在阶跃响应稳定后进行突加负载实验，实验结果如图10所示。

由图10可知，采用PID算法与SMC算法时，在稳定后突加负载会使位置环响应无法跟随给定信号，但是在采用本文方法后，位置环响应虽然会

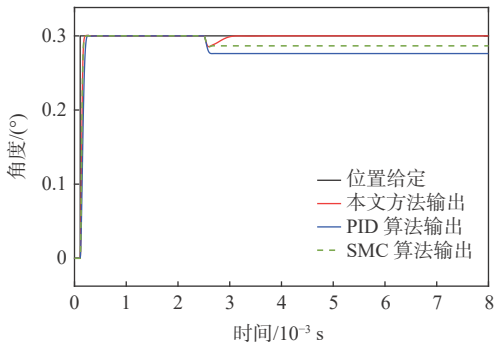


图 10　抗干扰能力测试

Fig. 10　Anti-interference capability test

产生波动, 但仍能在较短时间内继续跟踪给定信号, 具有更强的抗干扰能力。

接下来对振镜电机驱动器放大倍数进行参数摄动实验, 以测试本文方法的鲁棒性。对采用 PID 算法时的驱动器放大倍数进行 20% 参数摄动实验, 即令图 2(a) 中的 $R_2 = 0.8\ \Omega$; 对采用本文方法时的驱动器放大倍数分别进行 20% 及 40% 的参数摄动仿真, 即令 $R_2 = 0.8\ \Omega$ 及 $R_2 = 0.6\ \Omega$, 仿真结果如图 11 所示。

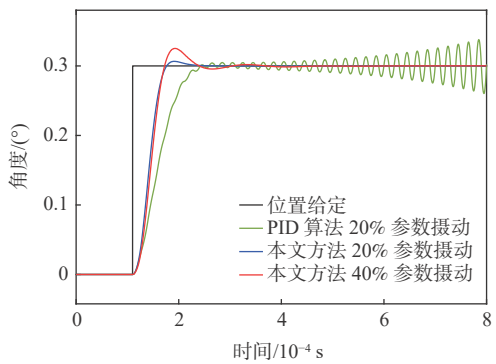


图 11　鲁棒性能力测试

Fig. 11　Robustness capability test

由图 11 可知, 采用 PID 算法时驱动器的放大倍数产生误差从而对系统的性能产生极大的影响, 但是采用本文方法时, 系统在驱动器的放大倍数产生较大变化的情况下并不会出现不稳定的现象, 可知整体振镜电机系统的鲁棒性得到了较大的提升。

通过仿真结果可以看出, 采用本文方法不但能够有效提高系统的动态响应性能, 还能够提高系统的抗干扰能力及鲁棒性。

3　实验验证

3.1　实验平台搭建

为验证本文方法的有效性, 搭建了基于模数混合架构的高性能振镜电机系统实验平台, 如图 12 所示。该实验平台主要包括上位机、直流电源、控制板、驱动器电源、仿真器、CAN 通信、示波器及振镜电机等部分。微处理器采用 TI 公司的 TMS320C28346, 其具有 300 MHz 的主频, 88 个可复用的 GPIO 引脚, 以及 CAN、ePWM、SPI 等片上外设。其丰富的片上外设与低延迟的内核能够较好地满足振镜电机系统高性能的控制需求。振镜电机采用金海创公司的 7210X 型电机, 其主要参数如表 2 所示。

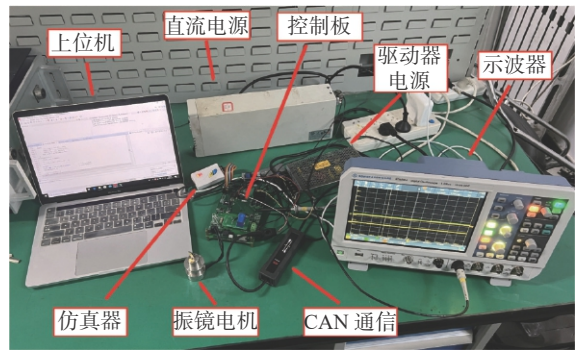


图 12　振镜电机系统实验平台

Fig. 12　Experimental Platform of Galvanometer laser scanner

表 2　振镜电机基本参数

Table 2　Basic Parameters of Galvo Motor

参数	数值
电机电阻/Ω	3.55
电机电感/μH	170
额定电压/V	30
额定电流/A	3
最大电流/A	6
力矩系数/(Nm·A^{-1})	0.017
反电势常数/(V·(rad·s^{-1})$^{-1}$)	8.3×10^{-3}
转动惯量/(kg·m^3)	1.25×10^{-8}
最大摆角/(°)	30

仿真实验得到的振镜电机系统位置环 1% 量程阶跃响应时间为 120 μs, 根据阶跃响应上升时间与系统带宽的公式, 可大体得出系统位置环的带宽为 3 kHz。根据工程经验可知, 系统的采样频率应取系统带宽的 10 倍以上, 所以系统的采样频率应至少为 30 kHz。根据选用处理器 TMS320C28346 的处理速度, 选定系统的采样频率为 100 kHz。

3.2　实验结果

为验证本文方法的控制性能, 对所设计的振镜电机系统进行了 1% 量程阶跃响应实验。给定 0.3°连续阶跃信号, 位置环反馈如图 13 所示。根据示波器的光标测量功能, 可得振镜电机系统的 1% 量程阶跃响应时间为 228 μs。

为提高振镜电机的控制精度, 采用了低通滤波器对反馈得到的位置反馈信号进行软件滤波, 因此, 实际响应速度与仿真实验结果有一定差距。

对本文设计的振镜电机系统进行全行程阶跃响应实验, 给定信号为幅值 30°的方波信号, 实验结

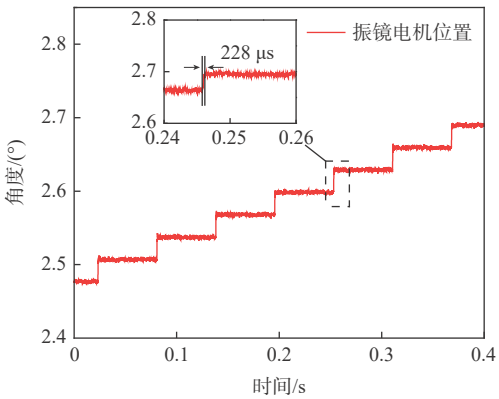


图 13 1% 量程阶跃响应实验

Fig. 13 1% step response experiment

果如图 14 所示，根据示波器的光标测量功能，可得振镜电机系统全行程阶跃响应时间为3.2 ms。

为测试振镜电机系统的随动能力，对其进行正弦信号响应实验。给定信号为幅值均为2°，频率分别为 100 Hz 和 500 Hz 的正弦信号，其响应结果如图 15 所示。

图 13～图 15 验证了本文方法的有效性。

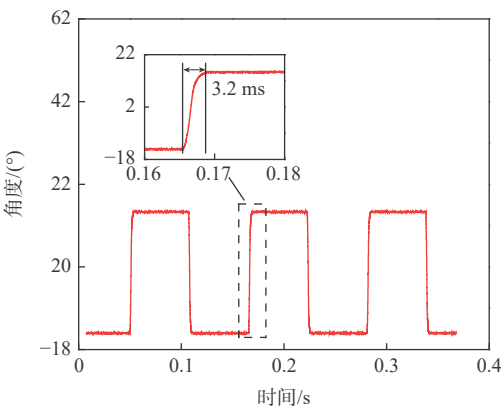


图 14 全行程阶跃响应实验

Fig. 14 Full range step response experiment

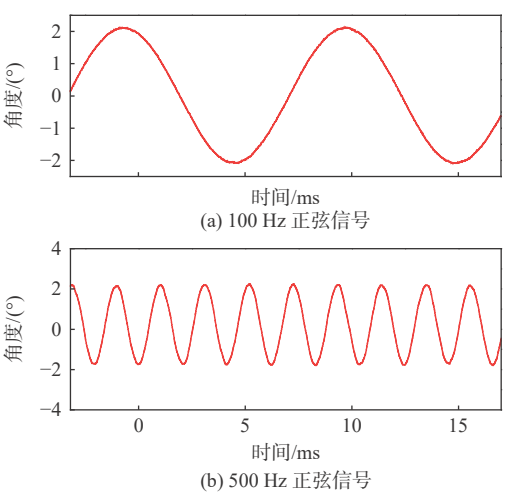


图 15 位置环正弦响应实验

Fig. 15 Sinusoidal response experiment of position loop

4 结 论

1）本文所设计的采用跨导放大器作为驱动器的数模混合控制架构对振镜电机自身电阻、电感数值的变化具有较强的鲁棒性。

2）本文方法能够有效提升振镜电机系统的动态响应性能，实验结果表明，其 1% 量程阶跃响应时间为228 μs，动态性能提升了 29.4%。

3）本文方法能够有效提升系统的抗干扰能力及鲁棒性，相较于传统的 PID 控制算法及 SMC 算法，本文方法能够在系统受到外部扰动后或驱动器参数发生变化后的较短时间内继续可靠跟踪给定信号。

参考文献（References）

[1] 陈波, 孟正, 马程远, 等. 扫描振镜激光 TC4 钛合金焊接性能及熔池流动行为[J]. 航空学报, 2022, 43(4): 525223.
CHEN B, MENG Z, MA C Y, et al. Welding properties and molten pool flow behavior of TC4 titanium alloy by oscillating galvanometer laser[J]. Acta Aeronautica et Astronautica Sinica, 2022, 43(4): 525223 (in Chinese).

[2] 李昊岳, 刘永江, 赵振兴, 等. 航空传感元件振镜激光钎焊界面组织及连接机理[J]. 航空学报, 2022, 43(2): 625020.
LI H Y, LIU Y J, ZHAO Z X, et al. Interface microstructure and formation mechanism of oscillating laser brazed joints for aerial sensing elements[J]. Acta Aeronautica et Astronautica Sinica, 2022, 43(2): 625020 (in Chinese).

[3] 朱小伟. 面向航空复合材料构件维修的激光分层雕刻工艺软件设计[D]. 温州: 温州大学, 2021: 9-12.
ZHU X W. The design of a laser layered engraving software for CFRP aviation component maintenance[D]. Wenzhou: Wenzhou University, 2021: 9-12(in Chinese).

[4] 乔明蕊. 基于可变论域模糊 PID 的振镜控制技术研究[D]. 西安: 中国科学院大学 (中国科学院西安光学精密机械研究所), 2016: 1-11.
QIAO M R. Research on galvanometer control technology base on variable universe fuzzy-PID[D]. Xi'an: Xi'an Institute of Optics and Precision Mechanics, Chinese Academy of Sciences, 2016: 1-11(in Chinese).

[5] 李乐. 针对低慢小目标的激光武器粗瞄控制系统研究与设计[D]. 哈尔滨: 哈尔滨工业大学, 2018: 1-4.
LI L. Research and design of aiming control system of laser weapon for low slow small target[D]. Harbin: Harbin Institute of Technology, 2018: 1-4(in Chinese).

[6] 黄勇, 邓建辉. 高能激光武器的跟瞄精度要求分析[J]. 电光与控制, 2006, 13(6): 86-88.
HUANG Y, DENG J H. On tracking & pointing accuracy requirement of high energy laser weapon[J]. Electronics Optics & Control, 2006, 13(6): 86-88(in Chinese).

[7] 刘小强, 杨修林, 陆培国, 等. 激光武器控制系统研究[J]. 激光与红外, 2022, 52(8): 1238-1245.

LIU X Q, YANG X L, LU P G, et al. Analysis of control system for high energy laser system[J]. Laser & Infrared, 2022, 52(8): 1238-1245 (in Chinese).

[8] 叶乔. 高速振镜理论研究及实践[D]. 武汉: 华中科技大学, 2004: 48-52.
YE Q. A study of high speed galvanometer and practise[D]. Wuhan: Huazhong University of Science and Technology, 2004: 48-52 (in Chinese).

[9] 井峰. 数字式振镜控制系统的研究[D]. 西安: 中国科学院研究生院 (西安光学精密机械研究所), 2012: 60-62.
JING F. Research on digital galvanometer control system[D]. Xi'an: Xi'an Institute of Optics and Precision Mechanics, Chinese Academy of Sciences, 2012: 60-62 (in Chinese).

[10] 权晓. 激光扫描振镜控制系统研究[D]. 太原: 中北大学, 2021: 53-60.
QUAN X. Research on control system of laser scanning galvanometer[D]. Taiyuan: North University of China, 2021: 53-60(in Chinese).

[11] YUE Z Y, YAN P. Fractional-order robust controller for high precision galvanometer scanner[C]//Proceedings of the IEEE International Conference on Manipulation, Manufacturing and Measurement on the Nanoscale. Piscataway: IEEE Press, 2019: 118-123.

[12] QIN W Y, GUO H, XU J Q, et al. High precision position control based on active disturbance rejection control for galvanometer scanner system[C]//Proceedings of the 22nd International Conference on Electrical Machines and Systems. Piscataway: IEEE Press, 2019: 1-4.

[13] MNERIE C A, PREITL S, DUMA V F. Performance enhancement of galvanometer scanners using extended control structures[C]//Proceedings of the 8th International Symposium on Applied Computational Intelligence and Informatics. Piscataway: IEEE Press, 2013: 127-130.

[14] JEAN M. Moving magnet galvanometers having a varied density winding distribution coil for a desired performance characteristic: US5225770[P]. 1993-07-06.

[15] JEAN M . Torque motor or moving-magnet galvanometer - useful for optical scanner e. g. for guiding laser beam: DE4205725B4[P]. 1992-04-20.

[16] 王丰尧. 滑模变结构控制[M]. 北京: 机械工业出版社, 1995: 244-270.
WANG F Y. Sliding mode variable structure control[M]. Beijing: China Machine Press, 1995: 244-270(in Chinese).

[17] 魏科鹏, 胡健, 姚建勇, 等. 航空机电作动器神经网络快速终端滑模控制[J]. 航空学报, 2021, 42(6): 624540.
WEI K P, HU J, YAO J Y, et al. Fast terminal sliding mode control of neural networks for aeromechanical actuators[J]. Acta Aeronautica et Astronautica Sinica, 2021, 42(6): 624540(in Chinese).

[18] 汪文涛, 刘正江, 李新民. 用于离心式作动器的变滑模面滑模控制[J]. 航空学报, 2019, 40(12): 323210.
WANG W T, LIU Z J, LI X M. Variable sliding mode sliding mode controller for centrifugal harmonic force generator[J]. Acta Aeronautica et Astronautica Sinica, 2019, 40(12): 323210(in Chinese).

[19] 高为炳, 程勉. 变结构控制系统的品质控制[J]. 控制与决策, 1989, 4(4): 1-6.
GAO W B, CHENG M. Quality control of variable structure control systems[J]. Control and Decision, 1989, 4(4): 1-6(in Chinese).

[20] 韩京清. 自抗扰控制技术: 估计补偿不确定因素的控制技术[M]. 北京: 国防工业出版社, 2008: 183-207.
HAN J Q. Active disturbance rejection control technique[M]. Beijing: National Defense Industry Press, 2008: 183-207(in Chinese).

[21] 张科, 姜海旭, 王靖宇. 基于滑动最速跟踪微分器的遥测数据滤波方法[J]. 西北工业大学学报, 2020, 38(3): 515-522.
ZHANG K, JIANG H X, WANG J Y. A sliding window optimal tracking differentiator filtering method for satellite telemetry data[J]. Journal of Northwestern Polytechnical University, 2020, 38(3): 515-522(in Chinese).

[22] 李海涛, 房建成. 基于扩张状态观测器的 DGMSCMG 框架伺服系统振动抑制方法[J]. 航空学报, 2010, 31(6): 1213-1219.
LI H T, FANG J C. Study on system vibration suppression method based on ESO used in gimbal servo system of DGMSCMG[J]. Acta Aeronautica et Astronautica Sinica, 2010, 31(6): 1213-1219 (in Chinese).

[23] 路遥. 基于跟踪微分器的高超声速飞行器 Backstepping 控制[J]. 航空学报, 2021, 42(11): 524737.
LU Y. Backstepping control for hypersonic flight vehicles based on tracking differentiator[J]. Acta Aeronautica et Astronautica Sinica, 2021, 42(11): 524737(in Chinese).

High performance control method for galvanometer laser scanner system

ZHAO Jiakang，GUO Hong，XU Jinquan*

(School of Automation Science and Electrical Engineering，Beihang University，Beijing 100191，China)

Abstract：To improve the control performance, the digital-analog hybrid control with the extended state observer (ESO) and sliding mode control (SMC) is proposed for the galvanometer laser scanner in this paper. The digital-analog hybrid control scheme is first proposed, which adopts the digital controller and analog driver for the high-performance control algorithm implementation and eliminates the ripple current due to the pulse width modulation (PWM) chopping. Then the ESO and SMC based nonlinear control is proposed to enhance the control accuracy and dynamic performance of the galvanometer laser scanner regardless of various disturbances. According to the mathematical model of the galvanometer laser scanner, the stability of the system is verified by the Lyapunov stability criterion. The findings of the simulation and testing demonstrate that the suggested control can enhance the step response time of the 1% range by roughly 29.4% when compared to the proportion-integral-derivative (PID) controller, hence improving the system's dynamic performance. At the same time, the anti-interference ability and robustness can also be improved.

Keywords：galvanometers；sliding mode control；extended state observer；high dynamic performance；operational amplifiers；tracking differentiator

Received：2022-09-18；**Accepted**：2022-10-10；**Published Online**：2022-11-04 09：47

URL：link.cnki.net/urlid/11.2625.V.20221103.1719.003

* **Corresponding author**. E-mail：xujinquan@buaa.edu.cn

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2022.0797

基于 MDk-DPC 的空中目标自动分群方法

马钰棠，孙鹏*，张杰勇，闫云飞，赵亮
(空军工程大学 信息与导航学院，西安 710077)

摘 要：空中目标分群本质上是一个类数未知的聚类问题，也是战场态势估计领域中的研究热点。针对未知的空战场环境，从聚类角度提出一种基于流形距离和 k 近邻采样密度的 MDk-DPC 算法。引入流形距离代替欧氏距离，以增加同一流形中目标的相似性；利用 k 近邻计算目标的局部密度，使其能更真实地反映目标周围分布；通过自适应选取聚类中心方法确定聚类中心，并运用密度峰值算法指定剩余点类别完成分群。仿真实验表明，所提方法在人工合成数据集和 UCI 真实数据集上均有更好的聚类性能，同时通过对空战场仿真数据进行分群验证了所提方法的可行性和有效性。

关 键 词：态势估计；目标分群；流形距离；k 近邻；密度峰值聚类
中图分类号：V247.5
文献标志码：A **文章编号**：1001-5965(2024)10-3219-11

新一轮的科技革命推动着现代战争由信息化向智能化演变，因此必将产生新型战争形态和作战体系[1]。然而面对复杂多变、交叉立体的战场环境，指挥人员很难快速准确地分析当前局势，因此，需要一种高效可行的态势估计方法。而对空中目标分群能去除大量冗余信息，有效降低战场的复杂度，从而辅助指挥人员进行决策。

战场目标分群本质上是聚类问题，现有方法主要是基于划分[2-4]、层次[5]、密度[6-9]、模型[10-12]等。基于划分的分群方法是根据簇的数量选择各簇的初始中心点，通过迭代重置使聚类结果中的类内距离近、类间距离远。该方法在球型数据中表现较好，但在聚类前需确定聚类数且其对初始中心点的选择较敏感。其中，较为典型的是 K 均值(K-means)和近邻传播(affinity propagation, AP)算法。文献 [2-3] 利用改进后的 K-means 算法和 K-prototype 算法来实现目标分群，但是算法性能建立在准确的目标群数之上，且算法在非球形数据上的分群效果不佳。文献 [4] 提出一种结合鲸鱼优化算法的 AP 算法，分群效果较好，但算法需要进行多次迭代，难以满足快速分群的要求。基于层次的聚类方法分为 2 步：①将距离最近的点合并到同一类中；②根据类与类之间的距离不断合并小类以合成大类。因此，此类方法无须指定聚类数，且能发现类之间的层级关系，但计算复杂度高、对噪声敏感。Chameleon 算法是较常见的层次聚类算法，文献 [5] 采用基于约束的 Chameleon 算法能在未知群数的条件下对敌目标进行空间分群，但数据量大时算法的实时性也会较差。基于密度的分群方法是将密度较高、距离较近的数据点合并为一类，能完成任意形状和大小的数据集聚类，并且不受噪声影响，但其聚类结果受参数选择影响较大。其中包括具有噪声的基于密度的聚类 [6](density-based spatial clustering of applications with noise, DBSCAN)方法和密度峰值聚类 [7](clustering by fast search and find of density peaks, DPC)算法。文献 [8] 提出改进基于 cell 的 DBSCAN 分群方法，能提高分群效率，但其参数难以确定。文献 [9] 在 DPC 算法的基础上提出基于流行距离的密度

收稿日期：2022-09-19；录用日期：2022-11-30；网络出版时间：2022-12-16 15:45
网络出版地址：link.cnki.net/urlid/11.2625.V.20221215.1606.002
*通信作者. E-mail：281126096@qq.com

引用格式：马钰棠，孙鹏，张杰勇，等. 基于 MDk-DPC 的空中目标自动分群方法 [J]. 北京航空航天大学学报，2024，50（10）：3219-3229. MA Y T，SUN P，ZHANG J Y，et al. Aerial target automatic grouping method based on MDk-DPC[J]. Journal of Beijing University of Aeronautics and Astronautics，2024，50（10）：3219-3229 （in Chinese）.

峰值快速搜索聚类 (manifold-clustering by fast search and find of density peaks, M-CFSFDP) 算法，并利用流形距离代替传统的欧氏距离，以提高同一流形目标间的相似性，但聚类中心的选择需要人工指定，因此分群效果会受主观影响。基于模型的分群方法主要依据神经网络模型描述数据分布，将数据集和模型进行拟合完成分群。文献 [10-12] 分别利用支持向量机、模糊自适应谐振理论（fuzzy adaptive resonance theory, ART）、深度自编码网络来完成目标分群，且实时性较好，但其训练数据难以获取且鲁棒性相对较差。

由以上分析可知，基于密度的分群方法能快速处理任意形状的数据，但在参数选择上过度依赖先验信息。因此，本文针对空中目标分群问题，在DPC算法的基础上引入流形距离和k近邻采样密度（manifold distance and k-nearest neighbor sampling density, MDK），提出一种基于MDk-DPC的空中目标自动分辨方法。首先，将两点间的流形距离作为相似性度量，再利用k近邻采样密度重新定义各点的局部密度和计算高密度近邻点距离；其次，在此基础上计算出各点的综合权重；最后，通过自适应选取聚类中心方法确定聚类中心并分配剩余点从而实现目标分群。将本文方法与多种经典算法进行对比，以纯度、兰德系数（Rand index, RI）、调整兰德系数（adjusted Rand index, ARI）和标准化互信息（normalized mutual information, NMI）为评估指标[13-14]，仿真结果证明：MDk-DPC算法在人工数据集和UCI真实数据集上能有效提升分群效果，同时验证了MDk-DPC算法在对空中目标分群时的可行性和有效性。

1 空中目标分群问题描述

空中目标分群是根据目标的速度、高度、航向角等特征信息将战场中某个空域内的所有目标划分成一定数量的空间群，且同一空间群中的目标有相似的特征状态。假设某个空域内有 n 个目标，每个目标有 m 个特征，则可表示为

$$T=\{x_1,x_2,\cdots,x_i,\cdots,x_n\} \tag{1}$$

$$x_i=\{x_{i1},x_{i2},\cdots,x_{ij},\cdots,x_{im}\} \tag{2}$$

式中：T 为所有目标的集合；x_i 为集合中的第 i 个目标；x_{ij} 为第 i 个目标的第 j 个特征。

目标集合在进行分群后形成 N 个互不相交的空间群，且每个空间群中至少包含一个目标：

$$T=\bigcup_{i=1}^{N}A_i \qquad \forall i\neq j, A_i\neq\varnothing, A_i\cap A_j=\varnothing \tag{3}$$

式中：A_i 为第 i 个空间群。

空战中的目标通常以楔队、梯队和横队等编队的形式出现，呈现出非球型的特点。基于作战指挥的需要，对空中目标进行分群的实时性要求较高，因此，应在DPC算法的基础上进行改进，以实现对空中目标的实时，准确的分群。

2 理论基础

2.1 DPC算法

DPC算法基于2个基本假设：①聚类中心的局部密度大于其近邻的局部密度；②不同的聚类中心相距较远。

首先，对于数据集 $\boldsymbol{T}=\{x_1,x_2,\cdots x_i,\cdots x_n\}$ 中每个点 x_i 求其局部密度和高密度近邻点距离：

$$\rho_i=\sum_{j\neq i}\chi\left(d_{ij}-d_c\right) \qquad \chi(x)=\begin{cases}1, x<0\\0, x\geqslant 0\end{cases} \tag{4}$$

$$\delta_i=\begin{cases}\min_{j:\rho_j>\rho_i}\left(d_{jj}\right) & \rho_i\neq\max_k(\rho_k)\\ \max_j\left(d_{ij}\right) & \rho_i=\max_k(\rho_k)\end{cases} \tag{5}$$

式中：ρ_i 和 δ_i 分别为 x_i 的局部密度和高密度近邻点距离；d_{ij} 为 x_i 和 x_j 两点间的欧氏距离；d_c 为截断距离。

根据每个点的 ρ_i 和 δ_i 绘制以 ρ_i 为横轴、δ_i 为纵轴的决策图，选择 ρ_i 和 δ_i 都较高的点为聚类中心，并将剩余点分配到密度比其大的最近邻点所在的类。

虽然DPC算法能快速搜索到聚类中心并完成聚类，但将其用于目标分群仍存在以下不足：

1）在计算两点之间的距离时使用的欧氏距离，不适用于空中目标流形的特点，因其影响分群结果。

2）在局部密度的计算过程中，截断距离 d_c 难以确定，且局部密度计算公式是阶跃函数，值域不连续，在目标密度不均匀的场景下的区分度不高，从而导致分群效果不佳。

3）采用手动选取聚类中心，在聚类数未知的条件下难以确定合适的聚类中心。

2.2 流形距离

当数据点分布为多个球状簇时，利用欧氏距离聚类的效果较好，而对于非凸数据集，欧氏距离则可能导致聚类效果不佳。如图1所示，a、b、c 为其中的3个数据点，从数据点的分布上可以看出，b 点与 c 点为同一簇，而与 a 点不同簇，所以 b 点与 c 点的相似性应大于 b 点与 a 点的相似性。若以欧式距离为聚类的相似性度量，由于 b 点与 a 点之间的欧氏距离小于 b 点与 c 点的距离，所以 b 点更有可能被分配到 a 点所在的簇，从而影响聚类算法的性能。因此，为提高同一流形中数据点间的相似性，本文采用流形距离作为相似性度量，具体定义如下：

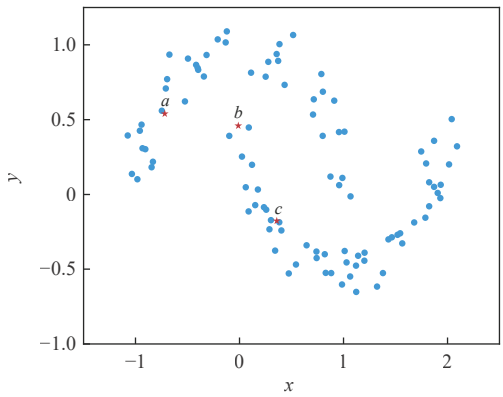


图 1　数据集中各点二维分布

Fig. 1　Two-dimensional distribution of points in the dataset

流形结构中任意两点之间线段长度定义为

$$L(x_1,x_2)=\rho^{d(x_i,x_j)}-1 \quad \rho>1 \tag{6}$$

式中: $d(x_i,x_j)$ 为点 x_i 和点 x_j 之间的欧氏距离; ρ 为伸缩因子, 用于控制两点间线段的长度。

假设一个加权无向图 $G=(V,E)$ 中的顶点 V 为所有数据点, 边集合 E 为任意两点之间定义的线段长度[15]。在 V 中连接任意两点 x_i 和 x_j 的路径集合表示为 P_{ij}, 其中, 一条路径为 $p=\{p_1,p_2,\cdots,p_l\}$, 边 $(p_k,p_{k+1})\in E, 1\leqslant k<l-1$。定义 x_i 和 x_j 之间的流形距离为

$$D\left(x_i,x_j\right)=\min_{p\in P_{ij}}\sum_{k=1}^{l-1}L(p_k,p_{k+1}) \tag{7}$$

如图 2 所示, b 点与 c 点的流形距离为 1.834 7, 明显小于 b 点与 a 点间的流形距离, 括号中的数字即为流形距离。因此, 采用流形距离聚类时能明显改善聚类结果。而在空中目标分群问题中, 由于空中目标大多以编队的形式行动, 具有流形的性质, 因此采用流形距离作为相似性度量能提升聚类算法的性能。

由流形距离的公式可知, 流形距离是从起始数据点开始, 搜索近邻点的连线来寻找下一个最近邻点直至终点, 并最终将连线中各线段的长度相加。因此, 流形距离使不同流形中的数据点距离放大, 相同流形间的点间距离缩小, 而其中的伸缩因子决定了放大缩小的程度。不同伸缩因子下的聚类结果对比如表 1 所示。

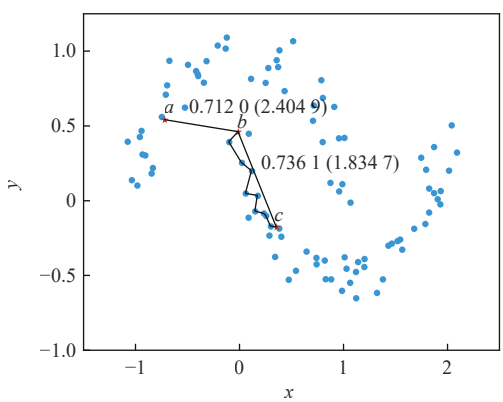


图 2　$\rho=e^2$ 时, ab 与 bc 间的欧氏距离和流形距离

Fig. 2　When $\rho=e^2$ the Euclidean distance and manifold distance between ab and bc

由表 1 可知, 伸缩因子过小会对聚类结果造成一定的影响, 使流形距离对点间距离的放大缩小效果不明显, 从而导致聚类结果不佳。而伸缩因子过大则会增加计算成本, 使聚类的实时性下降, 因此, 本文将伸缩因子取值为 e^3。

表 1　不同伸缩因子下的 Jain 数据集聚类结果对比

Table 1　Comparison of clustering results of Jain dataset with different scaling factors

ρ	纯度
$e^{0.5}$	0.924 9
e^3	1
e^{10}	1

2.3　k 近邻采样密度

密度计算本质上是对数据点周围分布的采样, 采样方法可以分为静态采样和动态采样[16]。在标准 DPC 算法中, 通过固定半径 d_c 在数据点周围形成固定大小的采样空间来计算局部密度 ρ_i, 这属于静态采样。静态采样计算所需的参数较少, 复杂度低, 因此不适用于多密度、高维的数据集。而动态采样则可以根据数据点周围的数据分布调整采样空间的形状和大小。其中, k 近邻是一种典型的动态采样方法[17], 其通过寻找数据点周围最近的 k 个点形成采样空间, 且在数据不均匀的场景下形成的采样空间是不规则的, 能更真实地反映实际数据分布。

结合核密度估计和 k 近邻, 重新定义局部密度:

$$\rho_i=\sum_{x_j\in N(x_i)}\mathrm{e}^{-D(x_i,x_j)} \tag{8}$$

式中: $N(x_i)$ 为 x_i 的 k 近邻集合; $D(x_i,y_i)$ 为 x_i 与 x_j 间的流形距离。

式(8)能在密度不均匀的情况下有效衡量数据点的局部密度, 但 k 值很难确定。若 k 值过小则会导致采样空间太小而难以得到准确的数据分布; 若 k 值过大, 不同数据点的采样空间都近似相同, 则会造成各点的局部密度的区分度较小, 不利于后续的聚类。

为保证局部密度能客观准确地反映数据点周围的分布, 本文根据所有数据点周围的真实近邻数量, 自适应计算 k 值。真实近邻由社交网络中的真实朋友的概念发展而来, 一个人的真实朋友应是被其当成朋友同时对方也将此人当成朋友的人[18]。

同理，若2个数据点互相在对方的k近邻集合中，则这对数据点即为真实近邻关系[19]。由此可知，x_i的真实近邻集合$N_R(x_i)$可表示为

$$N_R(x_i)=\{x_j|x_j\in N(x_i)\cap x_i\in N(x_j)\} \tag{9}$$

基于以上分析，提出一种自适应计算k值的方法，步骤如下：

步骤1 令t=1，数据集T中数据点的个数为n。

步骤2 计算数据集T中每个数据点x_i的t个近邻点集合。

步骤3 根据式(9)得到每个数据点x_i的真实近邻集合$N_R(x_i)$。

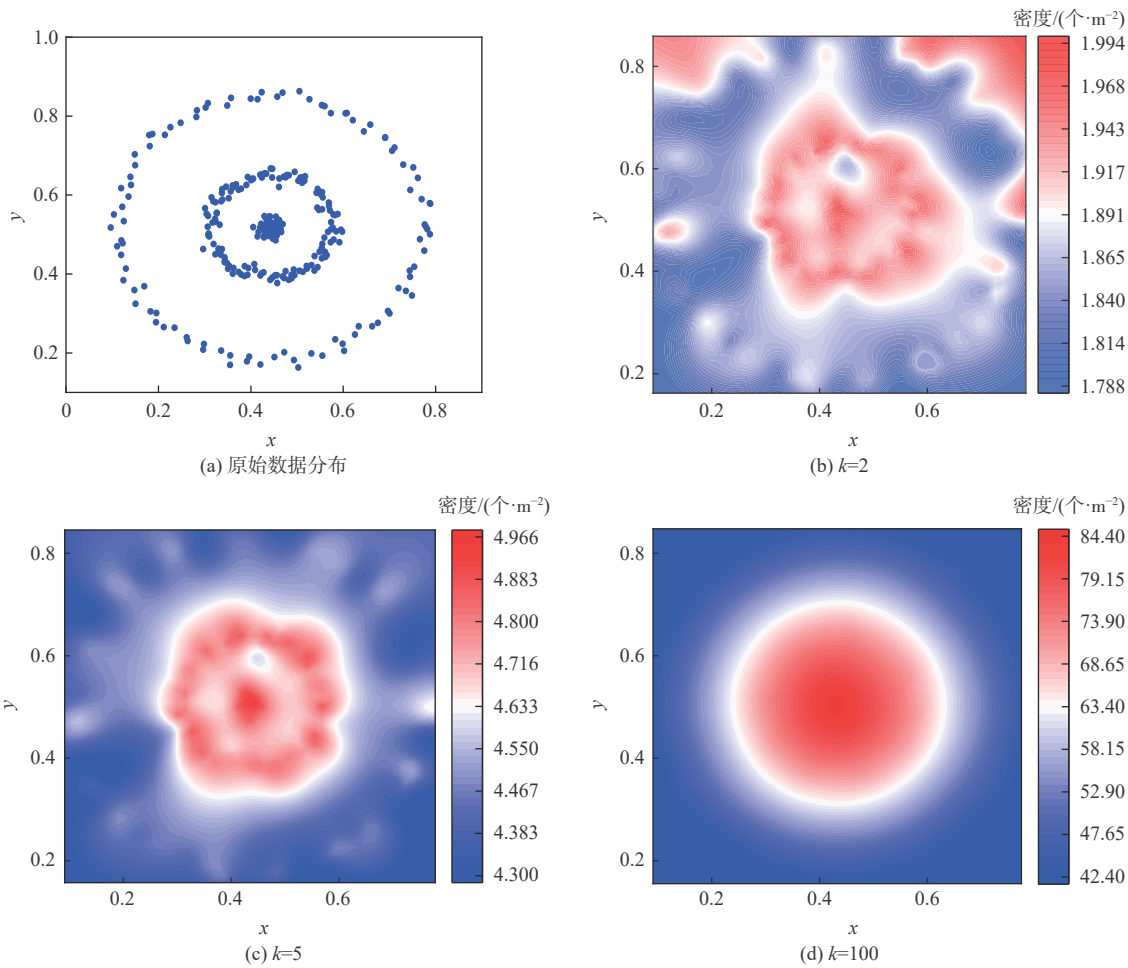
步骤4 将所有点的$RN(x_i)$中均没有包含的数据点放入集合N_N中。

步骤5 如果N_N中的元素发生变化，则t加1，并转到步骤2。

步骤6 根据式(10)计算k值。

$$k=\left\lfloor\frac{1}{n}\sum_{i=1}^{n}|N_R(x_i)|\right\rfloor \tag{10}$$

以Threecircles数据集为例，在相同密度度量下不同的k值产生的热分布如图3所示。图3(b)中k值过小导致与数据点分布明显不符，而图3(d)中k值过大使各簇之间的分界线模糊，进而影响聚类结果。通过自适应计算k值得到理想值，从而生成图3(c)中合理的密度结果。



(a) 原始数据分布　(b) k=2　(c) k=5　(d) k=100

图3 不同k值下的密度热分布图

Fig. 3 Density heat distribution diagram at different k values

2.4 聚类中心选取

聚类中心的选取直接决定了聚类的效果，选择合适的聚类中心能有效提高聚类的准确性[20]。在标准DPC算法中，局部密度ρ_i和高密度近邻点距离δ_i都是被选作聚类中心的较高的点，但聚类中心的选取并没有一个明确量化的标准。因此，为增大聚类中心与其他数据点之间的差异，结合ρ_i和δ_i定义数据点的综合权重w_i，计算式为

$$w_i=\rho_i\delta_i \tag{11}$$

数据点的综合权重w_i由ρ_i和δ_i相乘得到，ρ_i和δ_i越大，综合权重的值也越高，其被选作聚类中心的概率也越大。为合理选取聚类中心，需要确定聚类中心的个数。因此，本文使用轮廓系数来衡

量聚类效果的优劣, 以此来确定聚类中心的数量[21]。轮廓系数的定义如下:

$$S(x_i)=\frac{b(x_i)-a(x_i)}{\max\{a(x_i),b(x_i)\}} \tag{12}$$

$$S=\overline{S(x_i)}=\frac{1}{n}\sum_{i=1}^{n}S(x_i) \tag{13}$$

式中: $a(x_i)$ 为 x_i 和 x_i 所属类中其他数据点间的平均相异度; $b(x_i)$ 为 x_i 和其他某一类数据点的平均相异度中的最小值, n 为数据集中数据点的个数。本文采用 2.2 节中的流形距离表示相异度。由式 (12) 和式 (13) 可以得出轮廓系数的范围为 [−1, 1], 而通常情况下轮廓系数越大, 聚类效果越好。

基于以上定义, 本文提出一种自适应选取聚类中心的方法, 方法流程如图 4 所示。其中, c 为聚类中心数量。

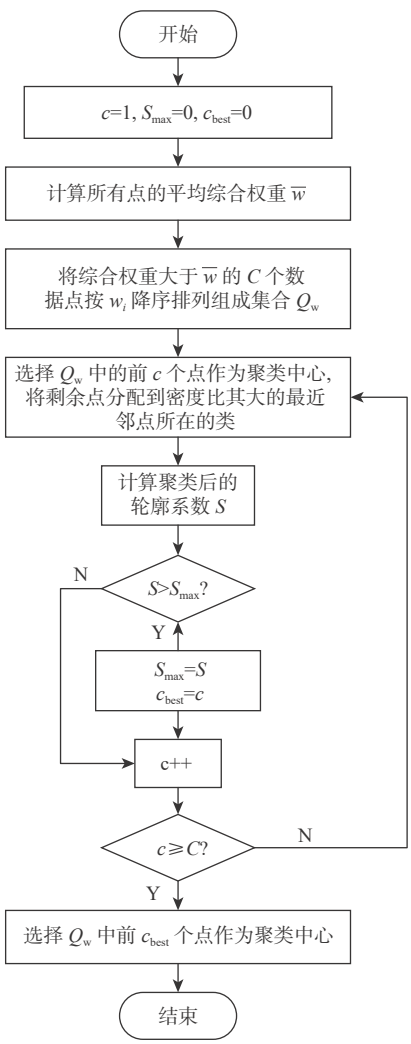


图 4　自适应选取聚类中心方法流程

Fig. 4　Flow of adaptive selection of cluster centers

2.5　MDk-DPC 算法

首先, MDk-DPC 算法使用流形距离代替欧氏距离计算各数据点间的相似度; 其次, 计算出各点的高密度近邻点距离和局部密度, 引入综合权重衡量各点被选作聚类中心的可能性; 最后, 进行自适应选取聚类中心, 并将其他数据点分配到各类中, 从而得到聚类结果。具体的算法步骤如下:

步骤 1　流形距离的计算。根据式(7)计算所有数据点之间的流形距离。

步骤 2　k 值的确定。依据自适应 k 值计算方法, 按式(10)得到 k 值。

步骤 3　局部密度和高密度近邻点距离的计算。将 k 值和流形距离代入式(8)计算得到局部密度 ρ_i, 再根据式(5)计算得到各点的高密度近邻点距离 δ_i。

步骤 4　聚类中心的选取。通过各点的 ρ_i 和 δ_i, 计算出各数据点的综合权重, 并利用自适应选取聚类中心方法选取聚类中心。

步骤 5　非聚类中心点的分类。在各个非中心点周围寻找局部密度比其大的近邻点, 选择其中最近邻的点, 将非中心点分配到最近邻点所属的类。

3　仿真实验分析

3.1　实验设置

实验数据采用人工合成数据集、UCI 真实数据集及战场空中目标仿真数据。将 MDk-DPC 算法与经典的 K-means、DBSCAN 及近几年提出的 DPC、M-CFSFDP 进行比较, 以验证 MDk-DPC 算法的有效性、可行性和优越性。其中, DPC 的局部密度采用高斯核计算得出, 其他算法中的参数采用插值法进行参数调优, 选择最优的结果。

实验评估指标选用纯度、RI、ARI、NMI, 兰德系数、调整兰德系数和标准化互信息, 其中, 纯度、RI、NMI 取值范围为 [0,1], ARI 取值范围为 [−1,1], 值越大聚类效果越好。

3.2　人工合成数据集实验

实验选用 6 个人工合成数据集, 基本信息如表 2 所示。实验结果如表 3 和图 5～图 7 所示。

表 2　人工数据集基本信息

Table 2　Basic information of artificial dataset

数据集	数据数量	类数	维数
Spiral	312	3	2
Flame	240	2	2
Jain	373	2	2
Lineblobs	266	3	2
Threecircles	299	3	2
R15	600	15	2

由表 3 可知, MDk-DPC 算法在 Spiral、Flame、Jain、Lineblobs、Threecircles 这 5 个数据集中均能得

表 3 人工数据集上的分群结果比较

Table 3 Comparison of clustering results on artificial datasets

数据集	算法	纯度	RI	ARI	NMI
Spiral	K-means	0.368 6	0.555 3	−0.003 0	0.003 1
	DBSCAN	1	1	1	1
	DPC	1	1	1	1
	M-CFSFDP	1	1	1	1
	MDk-DPC	1	1	1	1
Flame	K-means	0.837 5	0.726 7	0.453 5	0.419 3
	DBSCAN	0.966 7	0.920 0	0.840 8	0.756 6
	DPC	1	1	1	1
	M-CFSFDP	1	1	1	1
	MDk-DPC	1	1	1	1
Jain	K-means	0.836 5	0.725 7	0.448 3	0.442 8
	DBSCAN	1	0.969 8	0.937 3	0.855 0
	DPC	0.903 5	0.825 1	0.643 8	0.596 0
	M-CFSFDP	0.943 7	0.893 5	0.768 4	0.685 5
	MDk-DPC	1	1	1	1
Lineblobs	K-means	0.744 4	0.728 4	0.407 7	0.535 8
	DBSCAN	1	1	1	1
	DPC	0.751 9	0.734 5	0.420 1	0.545 0
	M-CFSFDP	0.853 4	0.831 8	0.628 3	0.743 0
	MDk-DPC	1	1	1	1
Threecircles	K-means	0.625 4	0.575 8	0.159 8	0.229 8
	DBSCAN	1	0.976 6	0.948 8	0.944 8
	DPC	0.571 9	0.456 5	0.065 9	0.182 3
	M-CFSFDP	0.652 2	0.545 2	0.169 7	0.308 5
	MDk-DPC	1	1	1	1
R15	K-means	0.993 3	0.998 3	0.985 7	0.989 3
	DBSCAN	0.731 7	0.911 2	0.551 1	0.885 2
	DPC	0.995 0	0.998 7	0.989 3	0.992 2
	M-CFSFDP	0.996 7	0.999 1	0.992 8	0.994 2
	MDk-DPC	0.996 7	0.999 1	0.992 8	0.994 2

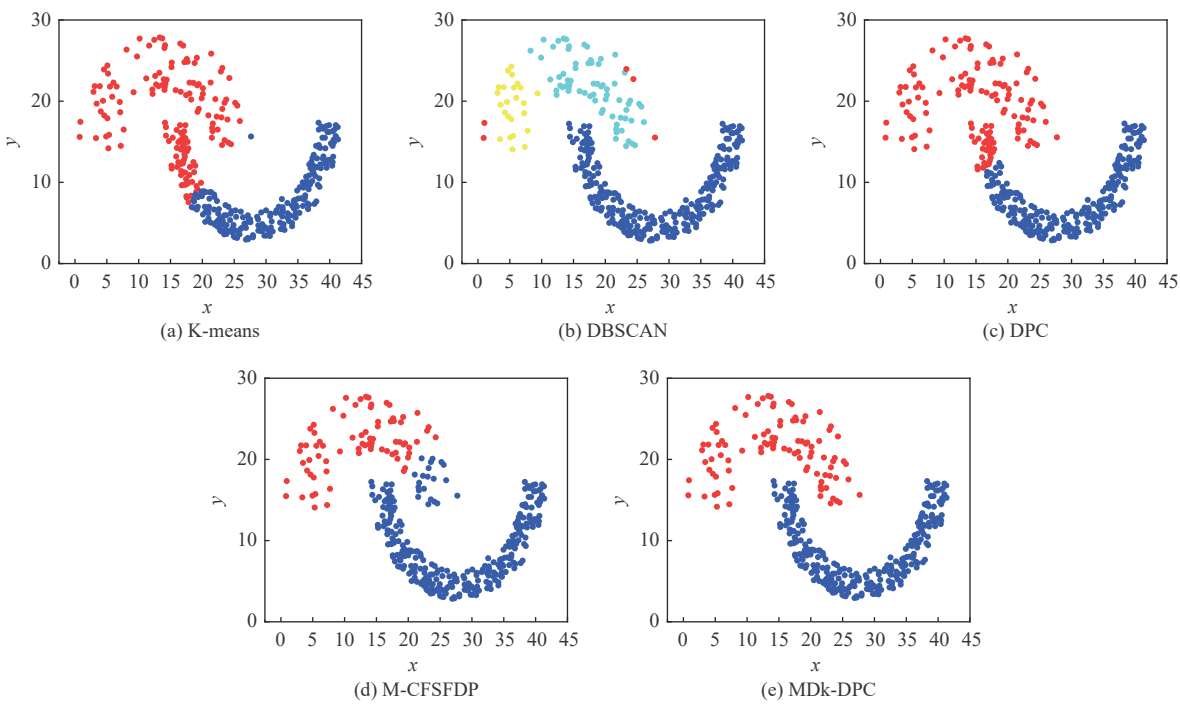


图 5 5 种算法在 Jain 数据集上的结果比较

Fig. 5 Comparison of the results of five algorithms on the Jain dataset

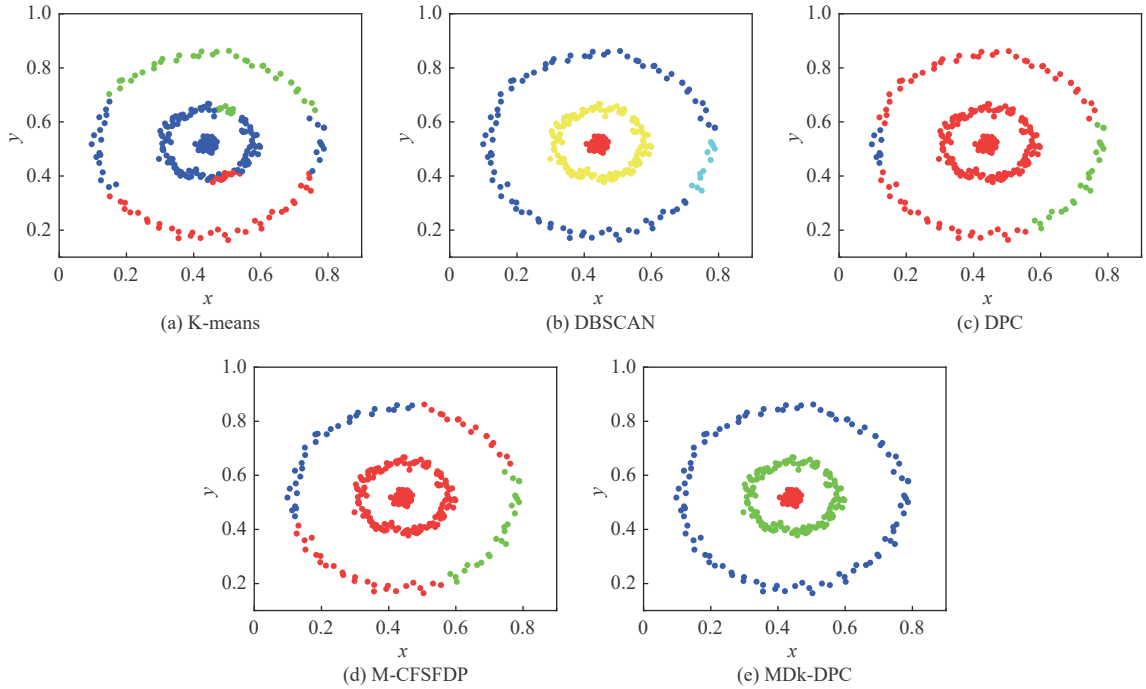


图 6　5 种算法在 Threecircles 数据集上的结果比较

Fig. 6　Comparison of the results of five algorithms on the Threecircles dataset

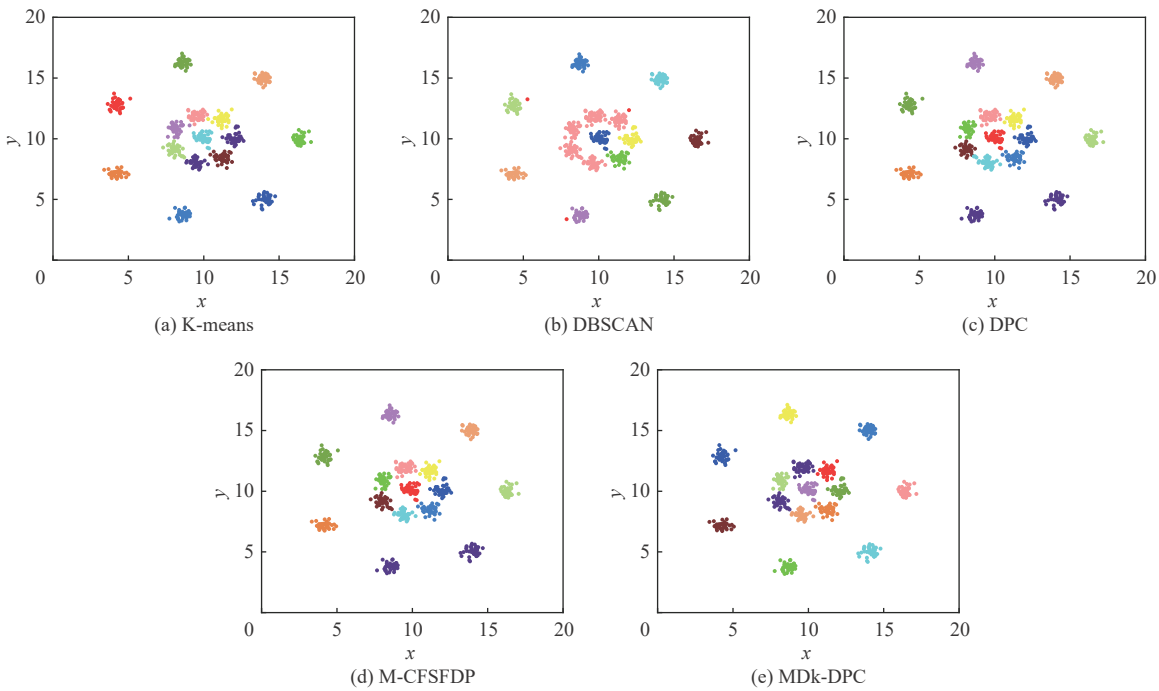


图 7　5 种算法在 R15 数据集上的结果比较

Fig. 7　Comparison of the results of five algorithms on the R15 dataset

到正确分群, 对 R15 数据集的分群结果在纯度、RI、ARI、NMI 这 4 项指标上也比其他 4 种算法表现得更好, 由此验证了 MDk-DPC 算法的有效性和优越性。

将图 5～图 7 进行对比可知, 相比其他 4 种算法, MDk-DPC 算法在不同数据集中的分群结果都较好。从图 5(a)、图 6(a) 和图 7(a) 中可以看出 K-means 在球形数据集中表现较好, 但不适用于流形数据集。而对比图 5(b)、图 6(b) 和图 7(b) 可知 DBSCAN 能处理非球形数据, 但在密度不均匀的 Jain 数据集上的分群效果不佳。图 6(c) 和图 6 (d) 中 DPC 和 M-CFSFDP 的分群结果和 MDk-DPC 有较大差距, 经分析应是没有正确选取聚类中心导致后续分群错误, 说明这 2 种算法在先验信息较少时分群表现不佳。

3.3　UCI 真实数据集实验

为验证 MDk-DPC 算法在真实高维环境中的分

群效果，实验采用6个UCI真实数据集，基本信息如表4所示。其中，WDBC数据集维数较高，Ecoli数据集类数较多。5种算法在UCI真实数据集上的结果如表5所示。

表4 UCI真实数据集基本信息

Table 4 Basic information of UCI real dataset

数据集	数据数量	类数	维数	各类数量
WDBC	569	2	30	357,212
Iris	150	3	4	50,50,50
Wine	178	3	13	59,71,48
Seeds	210	3	7	70,70,70
Vowel	871	6	3	72,89,172,151,207,180
Ecoli	336	8	8	143,77,2,2,259,20,5,52

由于真实数据集的数据都有一定的实际意义，通过计算平均值得到的聚类中心并不一定真实存在，不具有代表性，因此，适用于球形数据的K-means算法在真实数据集上并不一定能得到理想的分群结果，如Vowel数据集。

由表5可知，DBSCAN在Iris、Vowel数据集上表现较好，但在Wine和Ecoli数据集上表现不佳，这是由于未采用数据集的先验知识来确定参数，因而导致分群效果不稳定。DPC在Seeds和Ecoli数据集上能得到较高的纯度和ARI，但在WDBC和Vowel数据集上表现较差，可能是d_c选择不合理导致决策图中聚类中心过于靠近而难以确定。M-CFSFDP在多数数据集上的分群结果都优于DPC，这说明流形距离能改善分群效果，但在手动选取聚类中心时易出错，且实用性较差。而MDk-DPC通过自适应选取聚类中心的方法确定聚类中心，使其在大多数数据集上都能得到更高的指标值，以及更好的分群结果，从而证明了MDk-DPC算法在处理真实高维数据时的有效性。

3.4 战场仿真数据实验

战场空中目标数据由仿真平台获得，目标基本

表5 UCI真实数据集上的分群结果比较

Table 5 Comparison of clustering results on UCI real datasets

数据集	算法	纯度	RI	ARI	NMI
WDBC	K-means	0.850 6	0.745 4	0.481 1	0.456 7
	DBSCAN	0.799 6	0.655 8	0.319 1	0.285 1
	DPC	0.789 1	0.666 6	0.313 3	0.340 0
	M-CFSFDP	0.866 4	0.768 1	0.528 2	0.507 0
	MDk-DPC	0.882 2	0.791 9	0.577 3	0.545 9
Iris	K-means	0.880 0	0.867 9	0.702 8	0.727 7
	DBSCAN	0.933 3	0.878 6	0.706 3	0.713 0
	DPC	0.833 3	0.832 2	0.633 4	0.717 4
	M-CFSFDP	0.906 7	0.892 3	0.759 2	0.805 7
	MDk-DPC	0.960 0	0.949 5	0.885 8	0.870 5
Wine	K-means	0.702 2	0.718 7	0.371 1	0.428 8
	DBSCAN	0.646 1	0.625 8	0.289 7	0.402 9
	DPC	0.719 1	0.735 0	0.407 0	0.391 3
	M-CFSFDP	0.792 1	0.773 9	0.505 4	0.564 6
	MDk-DPC	0.893 3	0.865 4	0.699 0	0.726 2
Seeds	K-means	0.876 2	0.857 3	0.679 3	0.702 3
	DBSCAN	0.747 6	0.686 1	0.214 6	0.408 3
	DPC	0.890 5	0.871 5	0.722 7	0.712 6
	M-CFSFDP	0.895 2	0.874 8	0.717 0	0.674 4
	MDk-DPC	0.895 2	0.876 6	0.710 8	0.724 5
Vowel	K-means	0.551 1	0.801 3	0.336 0	0.469 2
	DBSCAN	0.695 8	0.816 5	0.288 5	0.486 4
	DPC	0.598 2	0.785 2	0.321 4	0.483 9
	M-CFSFDP	0.670 5	0.810 3	0.393 6	0.547 5
	MDk-DPC	0.740 5	0.832 8	0.407 2	0.536 0
Ecoli	K-means	0.800 6	0.834 0	0.538 6	0.577 1
	DBSCAN	0.663 7	0.760 1	0.393 7	0.451 8
	DPC	0.806 5	0.819 7	0.500 0	0.603 9
	M-CFSFDP	0.767 9	0.865 9	0.691 3	0.671 0
	MDk-DPC	0.788 7	0.885 2	0.731 6	0.697 1

状态信息如表 6 所示。其中, 1～5 号目标为群组 A, 6～8 号目标为群组 B, 9～10 号目标为群组 C。假设在 t 时刻, 由于收到临时命令, 4～5 号离开群组 A 前往群组 C 右侧, 6 号目标离开群组 B 与 1～3 号合编, 7～8 号加速在群组 C 行进路线左侧飞行。1～8 号目标对 9～10 号组成的群组 C 形成半包围之势。三维模拟场景如图 8 和图 9 所示。

使用 MDk-DPC 算法对战场仿真数据进行目标分群, 结果如图 10 所示。每个空中目标的运动状

表 6　空中目标状态信息

Table 6　Air targets status information

序号	x/m	y/m	z/m	速度/(km·h^{-1})	航向角/(°)	群组
1	4 597	9 856	4 221	959	90	A
2	4 098	10 182	4 197	934	90	A
3	3 092	11 210	4 213	967	90	A
4	3 785	9 045	4 203	980	90	A
5	2 890	8 002	4 208	942	90	A
6	9 013	24 345	4 127	931	150	B
7	9 854	24 521	3 951	964	150	B
8	9 470	23 698	4 055	954	150	B
9	11 013	10 132	4 195	980	120	C
10	10 311	9 780	4 089	979	120	C

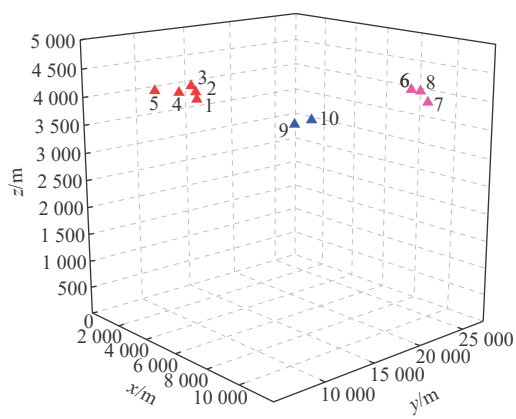


图 8　初始时刻空中目标分布

Fig. 8　Air targets distribution at the initial moment

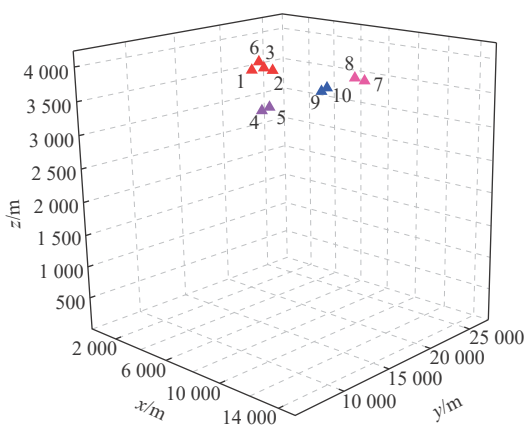


图 9　t 时刻空中目标分布

Fig. 9　Air targets distribution at time t

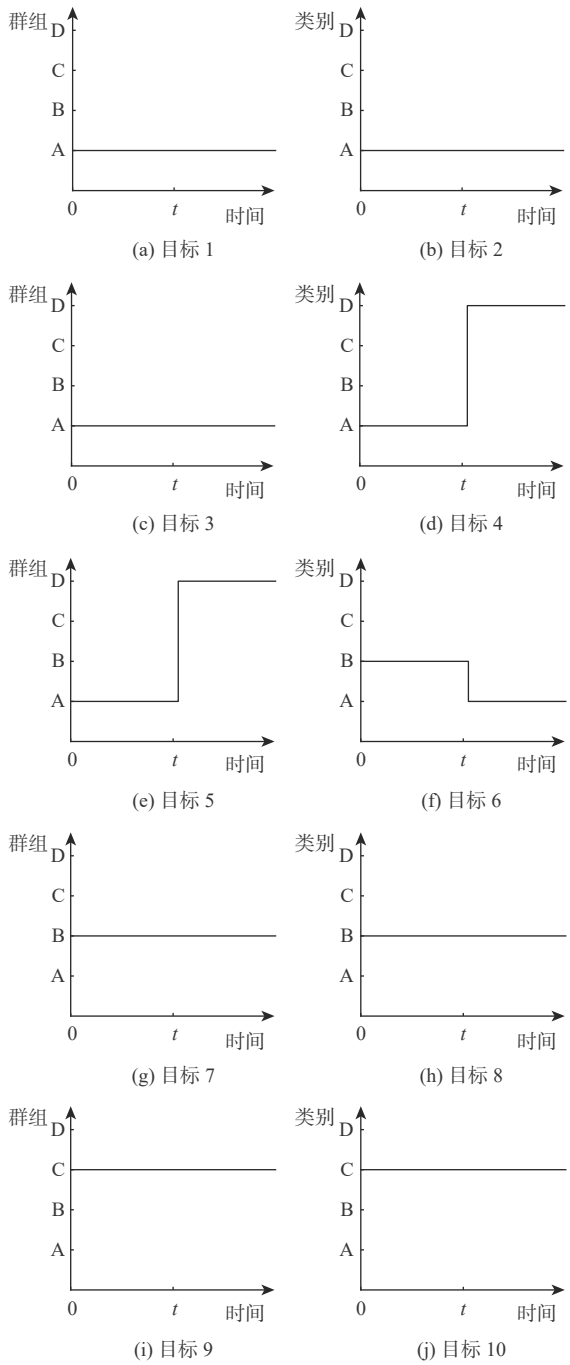


图 10　空中目标分群结果

Fig. 10　Aerial targets grouping results

态在 t 时刻都发生了变化, 4～6 号目标从原群组中分离, 因此, 分群结果也有所改变, 4～5 号目标生成独立的群组 D。

由图 9 和 10 可知, t 时刻后群组 A、群组 B、群组 D 对群组 C 形成半包围之势, 与事实相符, 从而证明了 MDk-DPC 算法的可行性和有效性。

4　结　论

1) 根据空中飞行群组通常呈现流形的特点, 采用流形距离代替欧氏距离, 这增强了同一流形中的目标相似性。

2）本文方法利用自适应选取聚类中心方法取代手动选取，提高了算法的运行效率。

3）MDk-DPC算法能对空中目标进行自动分群并动态输出结果，与其他算法相比有着更高的准确率。

参考文献（References）

[1] 郭明. 关于智能化战争的基本认知[J]. 人民论坛·学术前沿, 2021(10): 14-21.
GUO M. Basic understandings of the intelligent wars[J]. Frontiers, 2021(10): 14-21(in Chinese).

[2] 张绪亮, 张宏军, 綦秀利, 等. 基于改进K-means算法的陆战场机动目标分群方法[J]. 信息技术, 2016, 40(3): 128-131.
ZHANG X L, ZHANG H J, QI X L, et al. A clustering method of land battlefield maneuvering targets based on improved K-means algorithm[J]. Information Technology, 2016, 40(3): 128-131(in Chinese).

[3] 詹环, 宋爱斌, 罗俊芝, 等. 基于混合属性无监督聚类的作战目标分群[J]. 火力与指挥控制, 2021, 46(12): 166-170.
ZHAN H, SONG A B, LUO J Z, et al. Operational system target grouping method based on mixed attribute unsupervised clustering[J]. Fire Control & Command Control, 2021, 46(12): 166-170(in Chinese).

[4] DUAN Y X, LIU C Y, LI S. Battlefield target grouping by a hybridization of an improved whale optimization algorithm and affinity propagation[J]. IEEE Access, 2021, 9: 46448-46461.

[5] 袁德平, 郑娟毅, 史浩山, 等. 一种多作战编队下的目标编群算法[J]. 计算机科学, 2016, 43(2): 235-238.
YUAN D P, ZHENG J Y, SHI H S, et al. Target grouping algorithm based on multiple combat formations[J]. Computer Science, 2016, 43(2): 235-238(in Chinese).

[6] ESTER M, KRIEGEL H, SANDER J, et al. A density-based algorithm for discovering clusters in large spatial databases with noise[C]// Proceedings of the Second International Conference on Knowledge Discovery and Data Mining, 1996: 226–231.

[7] RODRIGUEZ A, LAIO A. Clustering by fast search and find of density peaks[J]. Science, 2014, 344(6191): 1492-1496.

[8] 闫孟达, 杨任农, 王新, 等. 改进cell密度聚类算法在空战目标分群中的应用[J]. 国防科技大学学报, 2021, 43(4): 108-117.
YAN M D, YANG R N, WANG X, et al. Air combat target grouping based on improved CBSCAN algorithm[J]. Journal of National University of Defense Technology, 2021, 43(4): 108-117(in Chinese).

[9] 李伟楠, 章卫国, 史静平, 等. 基于M-CFSFDP算法的战场目标分群方法[J]. 西北工业大学学报, 2018, 36(6): 1121-1128.
LI W N, ZHANG W G, SHI J P, et al. A battlefield target grouping method based on M-CFSFDP algorithm[J]. Journal of Northwestern Polytechnical University, 2018, 36(6): 1121-1128(in Chinese).

[10] 段同乐, 张冬宁. 二叉树多分类SVM在目标分群中的应用[J]. 无线电工程, 2015, 45(6): 88-91.
DUAN T L, ZHANG D N. Application of multiclass SVM based on binary tree in target grouping[J]. Radio Engineering, 2015, 45(6): 88-91(in Chinese).

[11] 樊振华, 师本慧, 陈金勇, 等. 基于模糊ART划分的目标分群算法[J]. 无线电工程, 2017, 47(9): 27-31.
FAN Z H, SHI B H, CHEN J Y, et al. A fuzzy ART based target clustering algorithm[J]. Radio Engineering, 2017, 47(9): 27-31(in Chinese).

[12] 陶宇, 蒋序平. 基于深度自编码网络的智能目标分群算法[J]. 指挥控制与仿真, 2020, 42(6): 52-58.
TAO Y, JIANG X P. Intelligent target clustering algorithm based on deep auto-encoder network[J]. Command Control & Simulation, 2020, 42(6): 52-58(in Chinese).

[13] WU Z H, LI C, ZHOU F, et al. A new weighted fuzzy C-means clustering approach considering between-cluster separability[J]. Journal of Intelligent & Fuzzy Systems, 2021, 40(1): 1017-1024.

[14] REHMAN A U, BELHAOUARI S B. Divide well to merge better: A novel clustering algorithm[J]. Pattern Recognition, 2022, 122: 108305.

[15] 马萌. 基于流形距离的聚类算法研究及其应用[D]. 西安: 西安电子科技大学, 2010: 9-21.
MA M. The study and the application of clustering algorithm based on manifold distance[D]. Xi'an: Xidian University, 2010: 9-21(in Chinese).

[16] CAO X F, QIU B Z, LI X L, et al. Multidimensional balance-based cluster boundary detection for high-dimensional data[J]. IEEE Transactions on Neural Networks and Learning Systems, 2019, 30(6): 1867-1880.

[17] SAMET H. K-nearest neighbor finding using MaxNearestDist[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2008, 30(2): 243-252.

[18] DAI Q Z, XIONG Z Y, XIE J, et al. A novel clustering algorithm based on the natural reverse nearest neighbor structure[J]. Information Systems, 2019, 84: 1-16.

[19] XIE J, XIONG Z Y, ZHANG Y F, et al. Density core-based clustering algorithm with dynamic scanning radius[J]. Knowledge-Based Systems, 2018, 142: 58-70.

[20] TONG W N, LIU S, GAO X Z. A density-peak-based clustering algorithm of automatically determining the number of clusters[J]. Neurocomputing, 2021, 458: 655-666.

[21] ZHANG R L, SONG X H, YING S R, et al. CA-CSM: A novel clustering algorithm based on cluster center selection model[J]. Soft Computing, 2021, 25(13): 8015-8033.

Aerial target automatic grouping method based on MDk-DPC

MA Yutang，SUN Peng*，ZHANG Jieyong，YAN Yunfei，ZHAO Liang

(Information and Navigation College，Air Force Engineering University，Xi'an 710077，China)

Abstract：Air target grouping is a popular topic for research in the area of combat scenario assessment and can be thought of as essentially an uncountable class clustering issue. Aiming at the unknown air battlefield environment, a MDk-DPC algorithm based on manifold distance and k-nearest neighbor sampling density is proposed from the perspective of clustering. First, manifold distance is introduced to replace Euclidean distance to increase the similarity of objects in the same manifold. Secondly, the target's local density is determined using the k-nearest neighbors method, allowing the local density to more accurately represent the distribution surrounding the targets. Finally, an adaptive cluster center selection method is proposed to automatically determine cluster centers, and the DPC algorithm is used to specify the remaining point categories to complete the clustering. Simulation experiments show that the proposed method has better clustering performance on both artificial synthetic datasets and UCI real datasets. At the same time, the feasibility and effectiveness of the method are verified by clustering the simulated air battlefield data.

Keywords：situation assessment；target grouping；manifold distance；k-nearest neighbors；density peak clustering

Received：2022-09-19；**Accepted**：2022-11-30；**Published Online**：2022-12-16 15：45

URL：link.cnki.net/urlid/11.2625.V.20221215.1606.002

* **Corresponding author**. E-mail：281126096@qq.com

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2022.0759

机载导航设备辅助的LEO卫星定位方法与仿真验证

徐鉴名*，黄智刚，李锐
(北京航空航天大学 电子信息工程学院，北京 100191)

摘 要：复杂电磁环境下全球导航卫星系统(GNSS)容易被欺骗和干扰导致不可使用，而近地轨道(LEO)通导一体化卫星系统具有落地功率高和星座几何形状变化快的优势，是一种很好的应急备份导航源。但大多数LEO卫星系统在中低纬度地区可见卫星的数量仅有1～2颗，无法实现实时动态定位。因此，提出一种基于机载惯性导航系统(INS)、气压计辅助的组合定位算法，将INS、气压计当作伪卫星使用，通过配置原子钟联合来解决可观性不足的问题。所提算法利用LEO卫星观测数据修正INS的误差，同时针对可观性不能定量分析的问题提出一种衡量滤波器稳定的可估计性指标。仿真验证表明：所提算法利用LEO观测数据在9 min内将INS误差降低了60%左右，定位误差达到50 m以下，能够实现实时、稳定定位，提出的可估计性指标也能够衡量系统的稳定性。

关 键 词：低轨卫星；扩展卡尔曼滤波；可观性；组合定位；惯性导航系统
中图分类号：V249.3; U666.134
文献标志码：A **文章编号**：1001-5965(2024)10-3230-09

近年来，近地轨道(low earth orbit, LEO)卫星通信系统将融合导航功能，可自主播发导航测距信号来服务于定位。LEO通导一体化卫星具有落地功率高、身份识别等优点，在未来复杂的电磁环境中的抗干扰性和抗欺骗性强，这为定位和导航带来了新契机[1]。

目前，LEO通导一体化卫星系统的星座设计优先解决通信问题，星座卫星数量少导致低纬度地区可见卫星仅有1～2颗，而使用1～2颗卫星实现定位的方法主要有2种。第1种方法是在接收机静止不动的条件下实现，如在货物监控等应用场景中，接收机在一段时间内按照相同的采样间隔接收同一颗卫星不同时刻的信号得到接收机的位置，但该方法的不足之处是定位时间长和精度较差[2-3]。1964年美国军方投入使用的子午仪卫星导航系统(Transit)使用该算法后的定位精度在100 m左右[4-5]。文献[6,7]研究了单颗全球卫星(Globalstar)自主定位的可行性。2010年，澳大利亚新南威尔士大学的学者将来自2颗全球导航卫星系统(global navigation satellite system, GNSS)的卫星使用该算法，发现2～3 h内定位误差可以收敛到20 m以下[8]。

第2种方法是接收机接收周围环境中已经存在的导航信号源或使用惯性导航系统(inertial navigation system, INS)、气压计、原子钟、测距仪/甚高频信标系统(distance measure equipment/very high frequency omnidirectional range, DME/VOR)等导航设备输出的信息作为观测信息，通过最小二乘(least square, LS)、扩展卡尔曼滤波(extended Kalman filter, EKF)算法解析出接收机的位置。常见的信号有Wi-Fi信号[9]、电视信号[10]、全球移动通信信号[11]和广播信号[12]等。因为这些信号具有机会存在性、机会可利用性和机会组合性，所以又称为机会信号

收稿日期：2022-09-07；**录用日期**：2022-09-16；**网络出版时间**：2022-11-19 16:01
网络出版地址：link.cnki.net/urlid/11.2625.V.20221117.1714.002
基金项目：国家重点研发计划(2020YFB0505602)
*通信作者. E-mail：3449421355@qq.com

引用格式：徐鉴名，黄智刚，李锐. 机载导航设备辅助的LEO卫星定位方法与仿真验证[J]. 北京航空航天大学学报，2024，50(10)：3230-3238. XU J M，HUANG Z G，LI R. LEO satellite positioning method and simulation verification aided by airborne navigation equipment[J]. Journal of Beijing University of Aeronautics and Astronautics，2024，50(10)：3230-3238 (in Chinese).

(signal of opportunity, SOP)[13]。然而, SOP 大多不服务于定位, 因此不可避免地会出现测量精度差及信号发射点的位置不精确等问题, 这对定位性能造成了较大影响[14]。文献 [15] 提出一种使用 2 颗全球定位系统 (global posetioning system, GPS) 卫星、多普勒导航系统和气压高度计的组合定位方法, 因为该方法能够提高定位精度。文献 [16] 提出一种 LEO 通信卫星与 INS 系统组合的方法, 该方法能够抑制 INS 误差的发散, 但存在动态定位误差大的问题。

简单来说, 第 1 种定位方法是用接收机在静止条件下观测同一颗卫星不同时刻的信号得到位置解; 第 2 种定位方法采用的是增加观测信息的数量。结合国内外研究现状可以发现, 关于使用 LEO 卫星实现进行实时定位的研究较少。因此, 本文提出一种基于 EKF 算法的利用机载设备辅助单颗 LEO 卫星实时定位的算法, 并使用电子计算机仿真, 对本文算法的可行性和稳定性进行了研究。

1　LEO 单星定位原理

LEO 卫星未来的发展趋势是播发具有时间同步特性的合作信号, 从而使接收机能够接收卫星信号并计算得到伪距和多普勒测量值。

1.1　伪距和多普勒定位原理

接收机计算出的伪距测量值主要由几何真实距离、大气传播延迟、接收机钟差、卫星钟差及测量误差组成。其中, 大气传播延迟误差可以用数学模型[17-18]估算得到, 卫星钟差和卫星的位置、速度可由卫星播发导航电文中的星历数据推导得出。而多普勒频移测量值 f_r 与伪距率 $\dot{\rho}$ 之间关系如下:

$$f_r = f_0 - \frac{\dot{\rho}}{\lambda} \tag{1}$$

式中: λ 为与卫星信号发射频率对应的波长; f_0 为卫星发射信号的频率。因为大气传播延迟在短时间内变化较小, 所以其对时间的求导可以认为是 0, 从而得到伪距和多普勒的定位方程组:

$$\begin{cases} \rho = \sqrt{(x_u - x^s)^2 + (y_u - y^s)^2 + (z_u - z^s)^2} + \delta t_u + \varepsilon_\rho \\ f_r = f_0 - \dfrac{1}{\lambda}\left(\dfrac{\boldsymbol{X}\boldsymbol{V}^{\mathrm{T}}}{r} + \dot{\delta}_{t_u}\right) + \varepsilon_f \end{cases} \tag{2}$$

式中: δ_{t_u} 和 $\dot{\delta}_{t_u}$ 分别为接收机钟差和钟差率; ε_ρ 和 ε_f 分别为伪距测量误差和频率测量误差; $\boldsymbol{X}=[x_u-x^s\ y_u-y^s\ z_u-z^s]$ 和 $\boldsymbol{V}=[v_{xu}-v_x^s\ v_{yu}-v_y^s\ v_{zu}-v_z^s]$ 为接收机与卫星之间的位置矢量和速度矢量, 其中, 带有上标 s 的参数为卫星在地心地固坐标系 (earth centered earth fixed, ECEF) 下的位置和速度, 带有下标 u 的参数表示接收机在 ECEF 坐标系下的真实位置。

通过对式 (2) 的进一步分析可以发现, 伪距方程的等面是一个球面, 而多普勒定位方程的等面是一个圆锥面, 卫星位置是圆锥的锥顶, 速度的方向是圆锥的轴线方向。伪距和多普勒定位原理如图 1 所示。

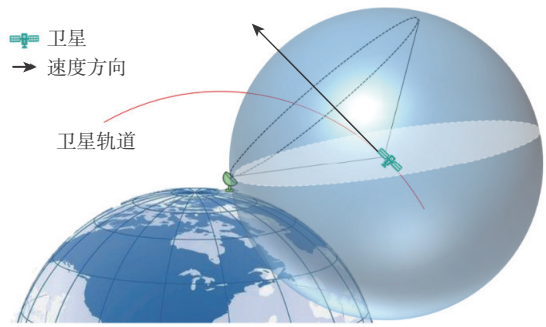


图 1　伪距和多普勒定位原理

Fig. 1　Pseudo-range and Doppler positioning principle

从图 1 可以看出, 接收机位于球面和圆锥面的相交线上。

1.2　EKF 滤波器设计

Kalman 最初提出的滤波理论只适用于包含线性状态方程和线性测量方程的线性控制系统, 但在实际工程中遇到的问题往往是非线性的。非线性卡尔曼滤波算法通常包括 EKF、无迹滤波 (unscented Kalman filter, UKF)、粒子滤波 (particle filter, PF) 等, 到目前为止, EKF 是应用最为广泛的一种非线性滤波方法, 其预测方程和更新方程如图 2 所示[19]。

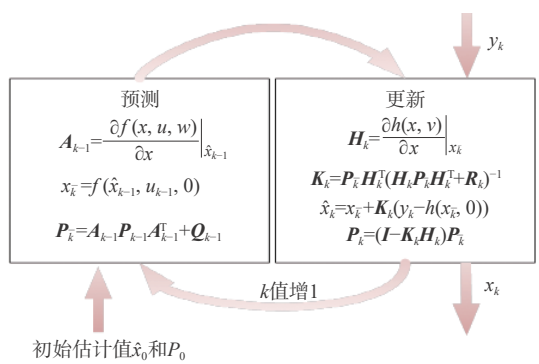


图 2　EKF 递推算法

Fig. 2　EKF recursive algorithm

在实际工程中, 如果只知道接收机的大致运动模型, 而不知道其具体的递推方程, 则一般使用白噪声模型或马尔可夫模型去匹配状态接收机的真实运动信息[20]。在进行 Kalman 滤波器设计时, 主要需要考虑不同模型的时间复杂度和空间复杂度: 时间复杂度是指算法运行的时间, 空间复杂度是指运行算法所需要使用的内存资源。由于 Kalman 滤

波的时间复杂度为n^3(n为状态向量的维数),因此,复杂度决定于其阶数,即状态向量的维数,其中,二阶恒速度(constant velocity, CV)模型的复杂度最低。因为本文单星静态定位时CV模型准确度更高,而动态定位的速度来源于INS的输出值,在已知速度采样间隔为1 s的情况下CV模型外推与高阶模型外推的结果相差较小,所以,可以同时兼顾准确度和复杂度2个方面,使用CV模型性能达到最优。

EKF的设计一般由4个步骤组成,分别包括:状态变量和观测量的选取;状态方程、状态转移矩阵和驱动噪声阵的确定;观测方程、观测矩阵和观测噪声协方差阵的确定;初值的设置,具体流程如图3所示。

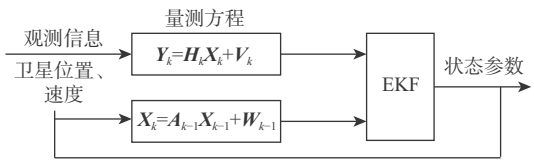


图3 EKF在状态估计中的流程框

Fig. 3 Flow chart of EKF in state estimation

步骤1 状态变量和观测量的选取。由最优估值理论可知,只有滤波器选取的状态变量包含系统中的所有变量,系统状态方程才能准确地描述系统变化规律,因此,将这种滤波器称为最优滤波器。本文状态变量$\boldsymbol{X}_s$可以选取如下:

$$\boldsymbol{X}_s=\begin{bmatrix}x & \dot{x} & y & \dot{y} & z & \dot{z} & \delta_t & \dot{\delta}_t\end{bmatrix} \tag{3}$$

式中:x和$\dot{x}$分别为在ECEF坐标系中x方向接收机的估计位置和估计速度;同理y、z及其导数为y、z方向上的分量;δ_t和$\dot{\delta}_t$分别为接收机的钟差和钟差率。

步骤2 状态转移矩阵和驱动噪声协方差阵的确定。CV模型的状态转移矩阵和驱动噪声协方差矩阵如下所示:

$$\boldsymbol{\Phi}=\begin{bmatrix}1 & T & & & & & & \\ 0 & 1 & & & & & & \\ & & 1 & T & & & & \\ & & 0 & 1 & & & & \\ & & & & 1 & T & & \\ & & & & 0 & 1 & & \\ & & & & & & 1 & T \\ & & & & & & 0 & 1\end{bmatrix} \tag{4}$$

$$\boldsymbol{Q}=\begin{bmatrix}\sigma_x^2\boldsymbol{q} & & & \\ & \sigma_y^2\boldsymbol{q} & & \\ & & \sigma_z^2\boldsymbol{q} & \\ & & & \sigma_{\delta_t}^2\boldsymbol{q}\end{bmatrix} \tag{5}$$

式中:T为离散化时间间隔,与接收机采样率有关;$\boldsymbol{q}=\begin{bmatrix}1/3T^3 & 1/2T^2\\ 1/2T^2 & T\end{bmatrix}$;$\sigma_x^2$、$\sigma_y^2$、$\sigma_z^2$及$\sigma_{\delta_t}^2$分别为对应方向上的过程噪声,一般会被设置为零均值高斯白噪声。

步骤3 观测矩阵和观测噪声协方差阵的确定。如图2所示,EKF的观测矩阵是在状态预测值处进行一阶线性化得到的,即观测方程分别对状态变量求偏导,即

$$\boldsymbol{H}_k=\begin{bmatrix}l_x & 0 & l_y & 0 & l_z & 0 & 1 & 0\\ m_x & n_x & m_y & n_y & m_z & n_z & 0 & -\dfrac{1}{\lambda}\end{bmatrix} \tag{6}$$

矩阵$\boldsymbol{H}_k$中的每一行即代表一个观测方程,其中,l_x为伪距方程对x求偏导数之后在x预测值处的值,即

$$l_x=\frac{X(1)}{\bar{r}} \tag{7}$$

式中:$X(1)$为多元方程在变量x处的一阶偏导;$\bar{r}$为接收机预测位置与卫星位置之间的几何距离。同理可求得l_x、l_y的值。m_x和n_x分别为多普勒方程对x和v_x求偏导:

$$m_x=\frac{V(1)}{\bar{r}}-\frac{X(1)\boldsymbol{X}\boldsymbol{V}^{\mathrm{T}}}{\bar{r}^3} \tag{8}$$

式中:$V(1)$为多普勒方程在x处的一阶偏导。

$$n_x=\frac{X(1)}{\bar{r}} \tag{9}$$

同理可求出m_y、n_y、m_z和n_z的值。

假设每个观测值的观测噪声互不相关,则可将观测噪声v的协方差阵$\boldsymbol{R}_k$设置为

$$\boldsymbol{R}_k=\begin{bmatrix}\sigma_\rho^2 & \\ & \sigma_f^2\end{bmatrix} \tag{10}$$

矩阵中的对角线元素从上到下分别为伪距、多普勒测量误差的方差。

步骤4 初值的设置。滤波的开始需要有状态初值$\hat{\boldsymbol{X}}_0$和初始误差协方阵$\boldsymbol{P}_0$,比较理想的滤波器初值应该设计为

$$\hat{\boldsymbol{X}}_0=E[\boldsymbol{X}_0]=\boldsymbol{m}_{x0} \tag{11}$$

$$\boldsymbol{P}_0=E[(\boldsymbol{X}_0-\boldsymbol{m}_{x0})(\boldsymbol{X}_0-\boldsymbol{m}_{x0})^{\mathrm{T}}] \tag{12}$$

但在实际工程中,很难得到状态初值的统计信息。根据线性控制理论可知,若系统随机可控及随机可观测的话,则盲目选取的滤波初值对滤波稳定值的影响将会逐渐变小。初始值虽然不影响滤波器稳定值,却会影响收敛速度,因此,初始值要尽量设置准确。

2 可观测性分析

系统随机可控和随机可观测性能越强,滤波器

的初值、随机噪声对滤波稳定值的影响越小。而本文所面临的最大问题是可观测性差, 无法实现稳定。因此, 很有必要对可观性的定性和定量进行分析。

2.1　线性系统可观测性理论

假设线性连续时间系统的状态方程和观测方程为

$$\dot{\boldsymbol{x}}(t)=\boldsymbol{A}(t)\boldsymbol{x}(t) \tag{13}$$

$$\boldsymbol{z}(t)=\boldsymbol{c}(t)\boldsymbol{x}(t) \tag{14}$$

方程有唯一解的充分必要条件是式(15)可逆, 该式也被称为 $\boldsymbol{G}$ 矩阵, 即

$$\boldsymbol{G}=\int_{t_0}^{t}\boldsymbol{\Phi}_{t,t_0}^{\mathrm{T}}\boldsymbol{c}^{\mathrm{T}}(t)\boldsymbol{c}(t)\boldsymbol{\Phi}_{t,t_0}\mathrm{d}t \tag{15}$$

由 $\boldsymbol{G}$ 矩阵及线性代数基础知识, 可以推导出可观的充分条件为矩阵 $[\boldsymbol{M}(t)\ \boldsymbol{M}(\cdot)(t),\cdots,\boldsymbol{M}(N-1)(t)]^{\mathrm{T}}$ 的秩为 n, 其中, $\boldsymbol{M}(t)=\boldsymbol{c}(t)\boldsymbol{\Phi}_{t_i,t_0}$, $\boldsymbol{M}(k)$ 表示 $\boldsymbol{M}(t)$ 对 t 的 k 阶导数。

当系统在离散时间的情况下, 其可观测的充分必要条件是 $\boldsymbol{G}$ 矩阵 $\sum_{i=0}^{N-1}\boldsymbol{\Phi}_{t_i,t_0}^{\mathrm{T}}\boldsymbol{c}^{\mathrm{T}}(t_i)\boldsymbol{c}(t_i)\boldsymbol{\Phi}_{t_i,t_0}$ 可逆。

2.2　可观测性定量分析

2.1 节的推导只是回答了系统在没有噪声干扰或干扰相差不大的情况下是否可观测的问题, 即使完全满足可观测条件, 系统也可能受到测量误差的影响而变得不稳定[21]。因此, 本文需要对可观测性的定量进行分析, 也即可估计性[22]。

将线性系统的测量误差考虑进来, 可观测性也就变成了随机可观测性, 通过相同的推理步骤, $\boldsymbol{G}$ 矩阵变为式(16), 称为 Fisher 信息矩阵, 记为 $\boldsymbol{F}$, 即

$$\boldsymbol{F}=\int_{t_0}^{t}\boldsymbol{\Phi}_{t,t_0}^{\mathrm{T}}\boldsymbol{c}^{\mathrm{T}}(t)\boldsymbol{R}^{-1}\boldsymbol{c}(t)\boldsymbol{\Phi}_{t,t_0}\mathrm{d}t \tag{16}$$

式中: $\boldsymbol{R}$ 为测量误差的协方差阵, 同理可以得到离散时间线性系统的信息矩阵:

$$\boldsymbol{F}=\sum_{i=0}^{N-1}\boldsymbol{\Phi}_{t_i,t_0}^{\mathrm{T}}\boldsymbol{c}^{\mathrm{T}}(t_i)\boldsymbol{R}^{-1}\boldsymbol{c}(t_i)\boldsymbol{\Phi}_{t_i,t_0} \tag{17}$$

当 $\boldsymbol{R}$ 为正定矩阵时, 可观测性的判别方法与 $\boldsymbol{G}$ 矩阵一样。

可采用 3 种比较常见的方法来表示系统的可观测程度:

1) 采用最小奇异值表示, 该数值越小表示在系统最差的观测方向上的可观测程度越差。

2)采用可观测矩阵的逆矩阵的迹为 $\dfrac{n}{\mathrm{trace}(\boldsymbol{F}^{-1})}$, 其中: n 为状态变量的维数。其值越大, 表示系统整体的可观性越强。

3) 采用可观测矩阵的行列式值为 $\det(\boldsymbol{F})^{1/n}$, 当该值为 0 时, 表示可观测矩阵为奇异阵, 系统是不可观测的。

从以上 3 种方法中可以发现, 从本质上讲可估计性是使用 $\boldsymbol{F}$ 矩阵的特征值中的奇异值来描述。本文考虑到无法判断整体不可观但部分稳定的可观系统的可估计性, 提出一种描述可估计性的指标为 $\left(\sum_{i=1}^{n}\sigma_i\right)^{\frac{1}{n}}$, 其中: σ 为信息矩阵的奇异值。

3　机载设备辅助 LEO 卫星定位原理

基于机载设备辅助 LEO 卫星定位算法框架如图 4 所示。

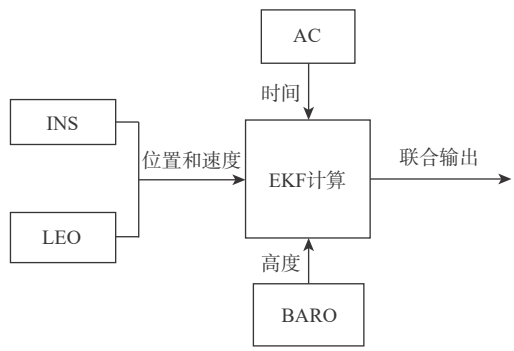


图 4　机载设备辅助 LEO 定位算法框

Fig. 4　Block diagram of airborne equipment-assisted LEO positioning algorithm

气压计根据托里拆利的实验原理而制成, 是一种测量大气压强的仪器, 其还可应用于测量接收机高度。气压计的应用方法之一是利用气压测高值增加一个关于接收机的近似椭球方程, 气压高度计的测量误差一般是均值且为真实高度的 1%, 方差为 30 m^2 的高斯噪声。需要注意的是椭球法的测量误差不再具有高斯特性, 会导致滤波器有偏。

普通接收机钟的预测精度差, 且能预测的时间很短, 在 3 min 后预测误差达到 40 ns[23] 因而不适用于本文定位。文献 [24] 利用与芯片原子钟相似的钟, 对 10 天内的钟差数据进行统计得到 99% 的预测误差在 20 ns 以内, 86.5% 的预测误差在 10 ns 以内, 均方根误差为 6.9 ns。

将原子钟估计得到的钟差距离量当作滤波系统的测量值 T_{N-1}, 测量方程可以写为

$$T_{N-1}=T_{\mathrm{r}}+\varepsilon_T \tag{18}$$

式中: T_r 为接收机钟差的真实值; ε_T 为观测误差。测量方程对状态变量求导得到滤波器的观测矩阵 $\boldsymbol{H}_T$:

$$\boldsymbol{H}_T=\begin{bmatrix}0 & 0 & 0 & 0 & 0 & 0 & 1 & 0\end{bmatrix} \tag{19}$$

INS 系统主要由陀螺仪和加速度计组成。该系

统先根据陀螺仪的输出建立导航坐标系，再根据加速度的输出算出运载体在导航坐标系中的位置和速度。

由于缺少真实的INS输出信息，所以本文通过推导INS传播方程中的位置误差方程和速度误差方程，然后根据方程仿真INS求出在任意时刻的位置和速度输出值。表1、表2分别列出了INS器件的精度、初始对准误差，这均是按照航空飞行的要求设置。

表1 陀螺仪、加速度计精度

Table 1 Accuracy of gyroscope and accelerometer

激光陀螺NV-G300/((°)·s^{-1})	加速度计/(m·s^{-2})
3.0×10^{-6}	9.8×10^{-6}

表2 INS初始对准误差

Table 2 Initial alignment error of INS

经度误差/(′)	纬度误差/(′)	高度误差/m	东向速度误差/(m·s^{-1})	北向速度误差/(m·s^{-1})	天向速度误差/(m·s^{-1})
0.1	0.1	10	0.1	0.1	0.1

INS的组合方法是以INS在ECEF坐标系下的初始对准位置(x_u,y_u,z_u)为基准点，并将其在第i时刻的位置输出值(x_i,y_i,z_i)与基准点之间的几何距离看作接收机伪距观测量ρ_I，即

$$\rho_I=\sqrt{(x_i-x_u)^2+(y_i-y_u)^2+(z_i-z_u)^2} \tag{20}$$

用伪距观测量对状态变量求偏导数得到滤波器的观测矩阵，即

$$\boldsymbol{H}_I=\left[\frac{x_i-x_u}{\rho_I}\quad 0\quad \frac{y_i-y_u}{\rho_I}\quad 0\quad \frac{z_i-z_u}{\rho_I}\quad 0\quad 0\quad 0\right] \tag{21}$$

观测信息Z为

$$Z=\rho_I-\sqrt{(\bar{x}_i-x_u)^2+(\bar{y}_i-y_u)^2+(\bar{z}_i-z_u)^2} \tag{22}$$

因为INS初始对准后的误差最小比较接近接收机的真实位置，所以将滤波器的每一时刻的观测噪声设置成当前INS输出值和初始对准接收机位置之间的模，即为ρ_I。

4 仿真分析

仿真条件与参数设置：LEO星座是按照铱星星座设计。图5为接收机静止在赤道卫星可见时间段内的星空图，可见总时长为8.8 min。

滤波器中的过程噪声是根据信号大气扰动所取的经验值，伪距和多普勒测量噪声是通过北斗接收机实测计算得到的经验值，如表3和表4所示。

当接收机只能接收到1颗LEO卫星信号的时候，滤波系统能够得到的观测信息有3维速度、伪距及多普勒。而滤波器的状态变量有8维，即使将所有的信息都正确使用，滤波器初值(状态初值和协方差阵初值)对定位误差也会产生很大的影响。

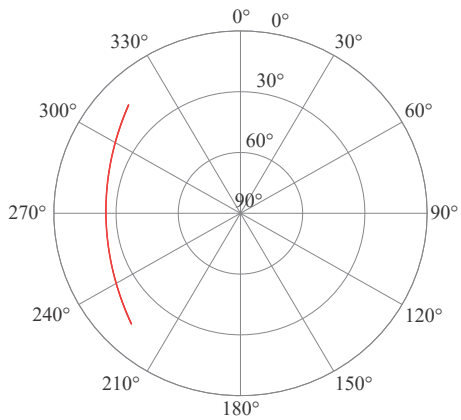


图5 LEO卫星星空图

Fig. 5 LEO satellite sky map

为了分析初值对LEO卫星定位性能的影响，定义2个初值：初值1，位置初值偏离真值1 km，协方差矩阵初值如式(12)所示；初值2，位置初值偏离真值10 km，协方差矩阵初值设置成10^9乘单位矩阵。LEO卫星不同初值定位性能的仿真结果如图6所示。

表5统计了6.6～8.8 min时，95%的水平、95%垂直及95%距离的均方根误差(RMS)。

由表5可知，在接收机静止的条件下，初值设置的理想定位误差能达到20 m以下，而初值较差时定位误差为1 km左右。由此可见，LEO单星静止定位对初值要求苛刻，且无法实现动态定位。

由式(17)计算信息矩阵的奇异值，得出最小值为2×10^{-4}，滤波器虽然可观但可估计性很差。由式(20)计算得到的可估计性为10.036。

图7为由表1和表2仿真得到的INS定位误差曲线，有了INS输出的位置和速度就可以实现LEO单星动态定位。本文将INS的初始对准位置当作伪卫星、速度当作观测量使用，如第3节所述。因为高精度INS仿真的时候使用了气压计阻尼，所以INS需要和气压计一起使用。图8为INS、

表3 滤波器过程噪声方差

Table 3 Noise variance of filter process m^2

σ_x^2	σ_y^2	σ_z^2	$\sigma_{\delta_t}^2$
10^{-4}	10^{-4}	10^{-4}	10^{-9}

表4 滤波器观测噪声方差

Table 4 Observed noise variance of filter

σ_ρ^2/ m^2	σ_f^2/Hz^2
16	0.04

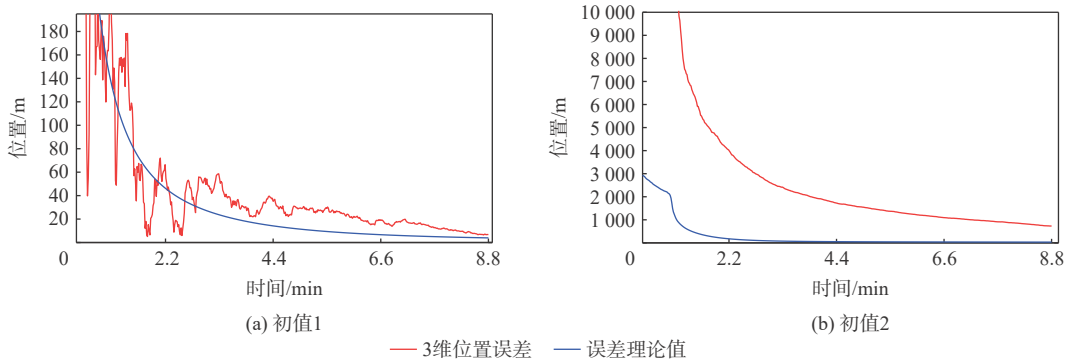


图 6　不同初值 LEO 卫星定位性能

Fig. 6　LEO satellite positioning performance with different initial values

表 5　不同初值 LEO 卫星定位误差统计

Table 5　Statistics of positioning errors of LEO satellites with different initial values　　m

初值	95%水平的RMS	95%垂直的RMS	95%距离的RMS
初值1	10.413 7	9.343 0	13.990 6
初值2	658.631 6	848.423 4	10^3

气压计辅助 LEO 单星的定位性能曲线, 表 6 为误差统计。

由图 6 和图 8 可知, INS 定位误差为 230～250 m, 而 INS 辅助 LEO 单星定位误差为 40～110 m, 由此可见, LEO 单星在 9 min 内将 INS 误差改善了 60% 左右。同时由表 6 可知, 组合之后的稳定性比

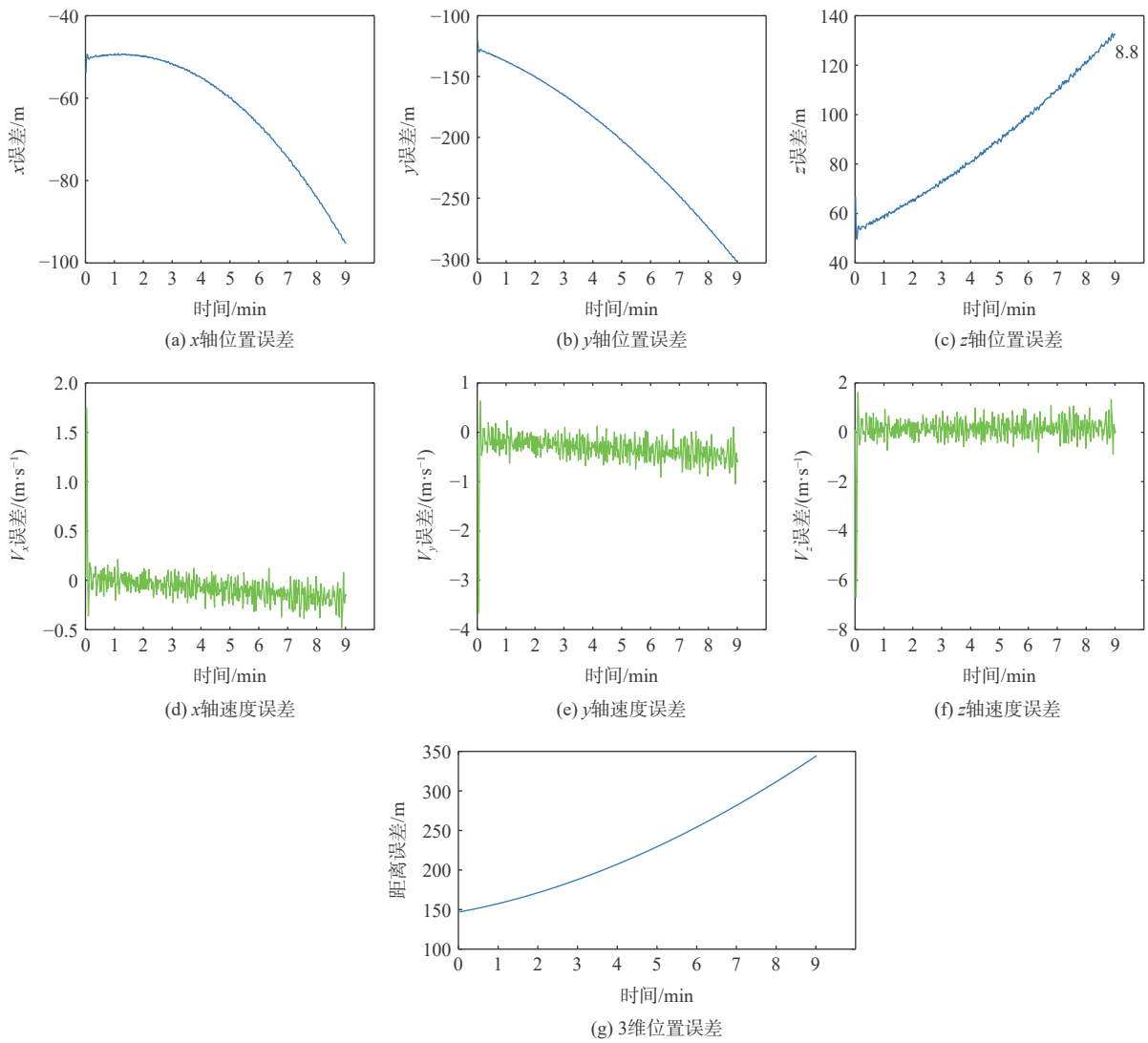


图 7　INS 定位性能曲线

Fig. 7　INS positioning performance curve

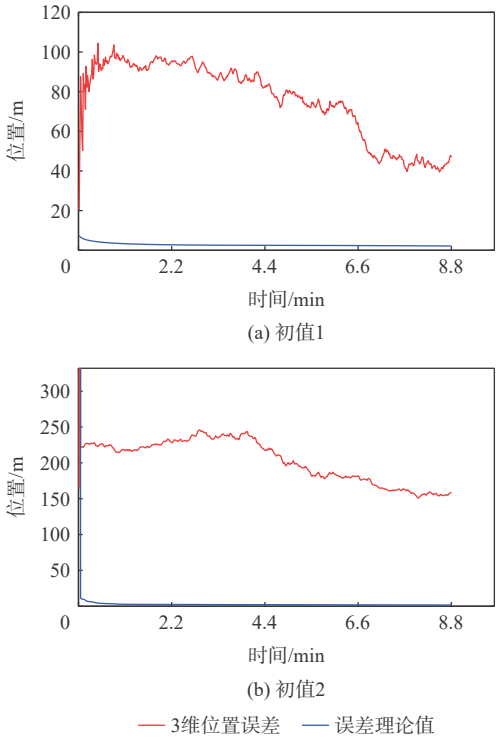


图 8 INS、气压计辅助 LEO 卫星定位性能曲线

Fig. 8 Performance curve of INS and barometer-assisted LEO satellite positioning

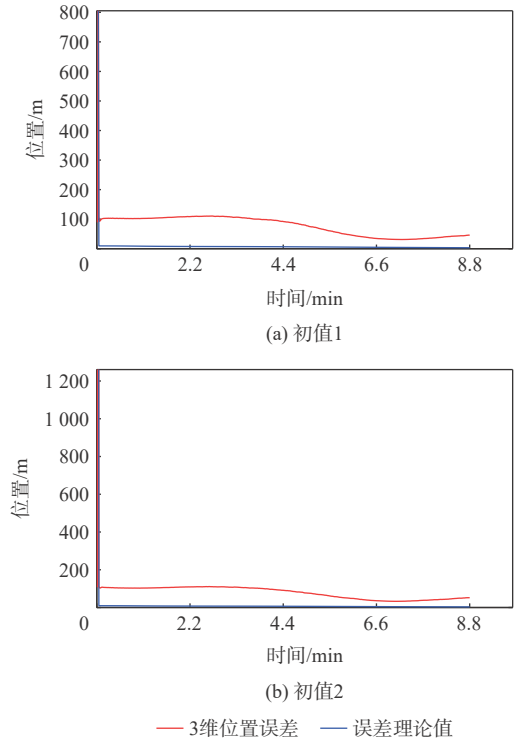


图 9 INS、气压计、原子钟辅助 LEO 卫星定位性能曲线

Fig. 9 Performance curve of INS, barometer, and atomic clock-assisted LEO satellite positioning

表 6 INS、气压计辅助 LEO 卫星定位误差统计

Table 6 Error statistics of INS and barometer-assisted LEO satellite positioning m

初值	95%水平的RMS	95%垂直的RMS	95%距离的RMS
初值1	22.510 7	46.429 4	48.802 9
初值2	148.219 8	52.421 7	156.259 2

表 7 INS、气压计、原子钟辅助 LEO 卫星定位误差统计

Table 7 Error statistics of INS, barometer, and atomic clock- assisted LEO satellite positioning m

初值	95%水平的RMS	95%垂直的RMS	95%距离的RMS
初值1	32.005 1	31.050 4	44.344 8
初值2	34.036 2	37.525 3	50.661 7

LEO 定位强。

同理，计算 INS 辅助 LEO 单星定位信息矩阵的奇异值，可得最小值为 11，计算得到的可估计性为 10.563 4。这比 LEO 单星定位的，可估计性更大。

从上述仿真结果可以发现，引入 INS、气压计辅助之后虽然可以实现动态定位，但系统依然容易受初值和噪声的影响，且定位误差较大。因此，本文假设接收机配置原子钟。图 9 为 INS、气压计、原子钟辅助 LEO 单星定位性能曲线，表 7 为误差统计。

由表 7 可知，在引入原子钟之后，不同初值的定位误差都在 50 m 左右，说明该算法能实现实时稳定定位。

同理，INS、气压计、原子钟辅助 LEO 单星定位信息矩阵的奇异值最小值为 1，计算得到的可估计性为 10.952 7。

由上述计算机仿真结果可知：在 9 min 内 LEO 卫星将 INS 误差改善了 60%；该算法能实现实时稳定定位，定位误差在 50 m 以下；佐证了可估计性与稳定性之间的关系，即可估计性越大、稳定性越强。

5 结 论

本文在 LEO 可见星仅有 1～2 颗无法实现稳定定位的条件下，提出一种基于机载惯导、气压计辅助的组合定位算法，通过配置原子钟联合定位解决了 LEO 卫星不能实现实时稳定定位的问题。计算机仿真表明：

1）在 9 min 内 LEO 单星将 INS 误差改善了 60%左右。

2）INS、气压计、原子钟辅助 LEO 单星能实现实时动态稳定定位，定位误差能达到 50 m 以下。

3）可估计性指标可用于衡量滤波系统的稳定性。

参考文献（References）

[1] SU Y T, LIU Y Q, ZHOU Y Q, et al. Broadband LEO satellite communications: Architectures and key technologies[J]. IEEE Wireless Communications, 2019, 26(2): 55-61.

[2] TAN Y K, SCOTT C A, DEMPSTER A G. Detecting object movement through the use of two GNSS satellites[C]//Ignss Symp,

JejuL Tan Y K, 2009: 103-114.

[3] YE L , YANG Y , JING X , et al. Single-satellite integrated navigation algorithm based on broadband leo constellation communication links[J]. Remote Sensing, 2021, 13(7): 703.

[4] DANCHIK R J. An overview of transit development[J]. Johns Hopkins Apl Technical Digest, 1998, 19(3): 18-26.

[5] STANSELL Jr, Thomas A. Transit, the navy navigation satellite system[J]. Navigation, 1971, 18(1): 93-109.

[6] LEVANON N. Instant active positioning with one LEO satellite[J]. Navigation, 1999, 46(2): 87-95.

[7] LEVANON N. Quick position determination using 1 or 2 LEO satellites[J]. IEEE Transactions on Aerospace and Electronic Systems, 1998, 34(3): 736-754.

[8] TAN Y K. Positioning techniques with two GNSS satellites over time[C]//Proceedings of International Technical Meeting of the Satellite Division of the Institute of Navigation. Porfland: Oregon lonvention Center, 2010: 207-216.

[9] ALSHAMAA D, MOURAD-CHEHADE F, HONEINE P. Tracking of mobile sensors using belief functions in indoor wireless networks[J]. IEEE Sensors Journal, 2018, 18(1): 310-319.

[10] RABINOWITZ M, SPILKER J J. A new positioning system using television synchronization signals[J]. IEEE Transactions on Broadcasting, 2005, 51(1): 51-61.

[11] BORENOVIC M, NESKOVIC A, NESKOVIC N. Vehicle positioning using GSM and cascade-connected ANN structures[J]. IEEE Transactions on Intelligent Transportation Systems, 2013, 14(1): 34-46.

[12] WEBB T A, GROVES P D, CROSS P A, et al. A new differential positioning method using modulation correlation of signals of opportunity[C]// IEEE/ION Position, Location and Navigation Symposium. Piscataway: IEEE Press, 2010: 972-981.

[13] 黄高明, 景桐, 田威. 机会信号导航综述[J]. 控制与决策, 2019, 34(6): 1121-1131.
HUANG G M, JING T, TIAN W. Survey on navigation via signal of opportunity[J]. Control and Decision, 2019, 34(6): 1121-1131 (in Chinese).

[14] MORALES J J, KHALIFE J J, KASSAS Z M. Information fusion strategies for collaborative inertial radio SLAM[J]. IEEE Transactions on Intelligent Transportation Systems, 2022, 23(8): 12935-12952.

[15] 孙国良, 沈士团, 丁子明, 等. 一种无源双星与多普勒导航系统组合的实现方法[J]. 航空学报, 2006, 27(4): 682-686.
SUN G L, SHEN S T, DING Z M, et al. An integration method for passive RDSS and DNS[J]. Acta Aeronautica et Astronautica Sinica, 2006, 27(4): 682-686 (in Chinese).

[16] YE L Y, YANG Y K, JING X L, et al. Single-satellite integrated navigation algorithm based on broadband LEO constellation communication links[J]. Remote Sensing, 2021, 13(4): 703.

[17] KATARZYNA S, WIELGOSZ P, PAZIEWSKI J. Accuracy analysis of the Klobuchar ionosphere model transmitted by the GPS system[C]//Proceedings of the 9th International Conference on Environmental Engineering. Vilnius: J Paziewski, 2014: 246-252.

[18] Leandro, Rodrigo , M. Santos , and R. B. Langley . UNB neutral atmosphere models: Development and performance[C]//Proceedings of the National Technical Meeting of the Institute of Navigation Ntm. Florida: IOW, 2006: 215-226.

[19] Hargrave P J. A tutorial introduction to Kalman filtering[C]//IEE Colloquium on Kalman Filters: Introduction, Applications and Future Developments. Pissataway: IEEE Press, 1989: 1-6.

[20] Li X R , Jilkov V P . Survey of maneuvering target tracking: decision-based methods[C]//International Symposium on Optical Science & Technology. International Society for Optics and Photonics. Paris: SPIE, 2000: 368-383.

[21] FAWCETT J A. Effect of course maneuvers on bearings-only range estimation[J]. IEEE Transactions on Acoustics, Speech, and Signal Processing, 1988, 36(8): 1193-1199.

[22] LE CADRE J E, JAUFFRET C. Discrete-time observability and estimability analysis for bearings-only target motion analysis[J]. IEEE Transactions on Aerospace and Electronic Systems, 1997, 33(1): 178-201.

[23] 滕云龙, 师奕兵, 郑植. 恶劣环境下 GPS 接收机定位算法研究[J]. 仪器仪表学报, 2011, 32(8): 1879-1884.
TENG Y L, SHI Y B, ZHENG Z. Research on GPS receiver positioning algorithm under bad conditions[J]. Chinese Journal of Scientific Instrument, 2011, 32(8): 1879-1884 (in Chinese).

[24] 李博, 徐超, 李孝辉, 等. 遮挡环境下原子钟和气压测高仪辅助北斗定位方法研究[J]. 电子与信息学报, 2018, 40(9): 2212-2218.
LI B, XU X, LI X H, et al. BeiDou navigation satellite system in challenge environment using an atomic clock and barometric altimeter[J]. Journal of Electronics & Information Technology, 2018, 40(9): 2212-2218.

LEO satellite positioning method and simulation verification aided by airborne navigation equipment

XU Jianming*，HUANG Zhigang，LI Rui

(School of Electronic Information Engineering，Beihang University，Beijing 100191，China)

Abstract：Global navigation satellite system (GNSS) is easy to be deceived and jammed in a complex electromagnetic environment, while the low earth orbit (LEO) integrated communication and navigation satellite system has the advantages of high landing power and rapid constellation geometry changes, and it is a good emergency backup navigation source. However, most of the LEO satellite systems only have 1–2 satellites visible in the middle and low latitudes, which fail to realize real-time dynamic positioning. Therefore, a combined positioning algorithm based on airborne inertial navigation system (INS) and barometer assistance was proposed. The INS and barometer were used as pseudolites, and the problem of insufficient observability was solved by configuring atomic clocks. The algorithm used the LEO satellite observation data to correct the error of the INS. At the same time, an estimability index was proposed to measure the stability of the filter for solving the problem that the observability cannot be quantitatively analyzed. The simulation verification shows that the algorithm uses the LEO observation data to reduce the INS error by about 60% within nine minutes, and the positioning error can reach up to 50 meters, realizing real-time stable positioning. The proposed estimability index can measure the stability of the system.

Keywords：low earth orbit satellite；extended Kalman filter；observability；combined positioning；inertial navigation system

Received：2022-09-07；**Accepted**：2022-09-16；**Published Online**：2022-11-19 16：01

URL：link.cnki.net/urlid/11.2625.V.20221117.1714.002

Foundation item：National Key Research and Development Program of China (2020YFB0505602)

* **Corresponding author**. E-mail：3449421355@qq.com

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2022.0761

变弯度柔性后缘机翼动响应减缓试验

张育鸣，戴玉婷*，尉濡恺，胡雅婷，杨超
(北京航空航天大学 航空科学与工程学院，北京 100191)

摘 要：变体飞行器能显著提升飞行器的气动性能，而变弯度柔性后缘是实现变体飞行器的重要方式之一。为探究柔性后缘动态偏转下机翼动响应特性及减缓效率，设计了一个变弯度柔性后缘机翼模型并开展了风洞试验。该机翼模型由承弯翼梁和6个3D打印的翼段组成。其中，2个翼段后缘分别设计2个变弯度柔性后缘舵面，这2个舵面分别用于动响应激励和动响应减缓控制。变弯度柔性后缘舵面由数字舵机、柔性索、波纹板结构和聚二甲基硅氧烷(PDMS)柔性蒙皮组成。对变弯度柔性后缘进行地面静态偏转试验和地面动态偏转试验测试，以研究后缘弯度变形规律及舵机动态时滞特性。在此基础上，在低速风洞试验中研究变弯度柔性后缘机翼动响应规律和基于变弯度柔性和闭环反馈控制的动响应减缓效率。风洞试验结果表明：机翼的翼尖加速度响应及翼根弯矩在频率为1.5～4 Hz时先增大后减小，并在接近机翼一弯频率时达到峰值。采用PID控制律和变弯度后缘进行闭环反馈控制后，在风速20 m/s、扰动频率2.2 Hz时翼尖加速度最大减缓效率达到70.18%，翼根弯矩的最大减缓效率为68.14%。此外，还提出了动响应减缓效率为正值的理论公式并分析了动响应减缓效率的影响机制和因素。

关 键 词：柔性后缘；变弯度；动响应；减缓效率；风洞试验
中图分类号：V211.5; TB553
文献标志码：A **文章编号**：1001-5965(2024)10-3239-11

变体飞行器是一种能实现实时、自适应的变形以应对不同的飞行环境的飞行器，这一概念最早由北大西洋公约组织(North Atlantic Treaty Organization, NATO)提出[1]。变体飞行器能显著提升飞行器的气动性能[2]，节约燃油，对航空运输业的发展有重要意义[3-6]。

机翼变形是实现变体飞行器的重要手段，变形形式[7]主要分为面内变形、面外变形[8-9]和翼型变形[10-12]。考虑到变形的有效性和实现的复杂程度，国内外研究主要集中在翼型变形方面。其中，变弯度后缘由于对气动力的影响较大，且仅需较小的驱动力就可以实现[13]而成为研究热点。因此，针对柔性变弯度后缘开展动响应减缓试验，以保障飞行器的飞行品质和飞行安全是十分必要的。

为开展动响应减缓试验，首先需要确定变弯度柔性后缘的结构形式。目前，变弯度柔性后缘有多种结构实现形式：Molinari等开发了分布式柔度结构以取代传统的副翼[14]；东京大学的Takahashi等提出了基于碳纤维复合材料波纹板结构的可变弯度后缘设计[15]；Woods等提出了一种基于鱼骨的主动变形机翼结构，通过钢丝拉伸使后缘弯曲等[16]。变弯度柔性后缘的结构形式要求有足够的刚度来承受气动载荷，同时又需要一定的柔性使之易于变形。本文综合试验中实现了后缘变形所需驱动力和结构所应承受的气动载荷等因素，并采用基于波纹板形式的变弯度柔性后缘结构。

收稿日期：2022-09-07；**录用日期**：2022-11-28；**网络出版时间**：2022-12-28 08:35
网络出版地址：link.cnki.net/urlid/11.2625.V.20221226.1412.002
基金项目：国家自然科学基金(11672018)
* **通信作者**. E-mail：yutingdai@buaa.edu.cn

引用格式：*张育鸣，戴玉婷，尉濡恺，等. 变弯度柔性后缘机翼动响应减缓试验[J]. 北京航空航天大学学报，2024，50(10)：3239-3249.*
ZHANG Y M，DAI Y T，WEI R K，et al. Experiment on dynamic response alleviation of a wing with variable-camber flexible trailing edge[J]. Journal of Beijing University of Aeronautics and Astronautics，2024，50(10)：3239-3249 (in Chinese).

同时，国内外还开展了一系列变弯度柔性后缘的地面试验和风洞试验。在地面试验方面，Zhao 等利用蜂窝结构的柔性后缘进行地面试验，探究了驱动力与后缘偏转角之间存在的非线性关系[17]；雷朝辉等基于柔顺翼肋设计并制作了一套可连续改变弯度的后缘柔性翼肋模块，该方案可使后缘偏转达到±15°，偏转速率达到 10 (°)/s[18]；东京大学的 Takahashi 等通过地面试验验证了后缘的偏转形式[15]。欧洲通过 AirGreen2 项目研究出一种新型多模态曲面变形襟翼的设计。通过地面试验和风洞试验验证的智能架构，使得在低速 (起飞/着陆) 和高速 (巡航) 条件下可以实现不同的变形模式[19]。在风洞试验研究方面，Takahashi 等进行了风洞试验评估了开发的变弯度柔性后缘机翼的气动性能[15]。结果表明，与传统的带铰链舵面的机翼相比，变弯度柔性后缘机翼的气动性能更加优越。Woods 等对开发的鱼骨主动变形机翼进行了风洞试验[16]。试验结果表明，该机翼的升阻比相较于传统机翼提高了 20%～25%。随后，研究人员在 2020 年进行了风洞试验，研究了鱼骨主动弯曲 (FishBAC) 变形装置的气动性能。结果表明，FishBAC 在所有迎角和升力系数上都比传统襟翼具有更高的气动效率[20-21]。2021 年；Marino 等[22] 也利用 FishBAC 进行仿真和风洞试验验证，最终得到变形机翼升阻比约为传统机翼的 2 倍[22]。波音公司和美国航空航天局共同研制了连续可变弯度后缘襟翼 (variable camber continuous trailing edge flap, VCCTEF) 系统。通过多段铰链连接的形式不断改变机翼后缘的弯度来改善机翼的升阻特性。初步试验表明，VCCTEF 减少了近 10% 的巡航阻力。2019 年，波音公司为 VCCTEF 设计了阵风载荷减缓控制系统。仿真结果表明，阵风响应的振幅减小了近 50%[23-24]。

目前，国际上关于变弯度柔性后缘的试验研究集中在可变形机翼的变形特性和气动力特性。基于变弯度柔性后缘的动响应减缓特性的试验研究亟需开展，以解决动态驱动、变形测量和载荷减缓策略的难题。本文设计了变弯度柔性后缘结构，提出了基于柔性后缘的风洞试验动态响应测量和基于主动控制的减缓方法。

1 变后缘机翼模型设计和地面试验

本文的研究对象是含有变弯度柔性后缘的平直机翼，如图 1 所示。依据风洞尺寸：机翼翼型为 NACA0015 翼型，弦长为 0.4 m，展长为 1.2 m。为减小蒙皮带来的附加刚度，将一弯频率控制在舵机动态特性的有效范围内，将机翼沿展向平分为 6 段。

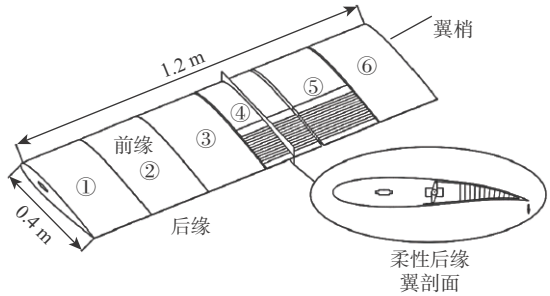


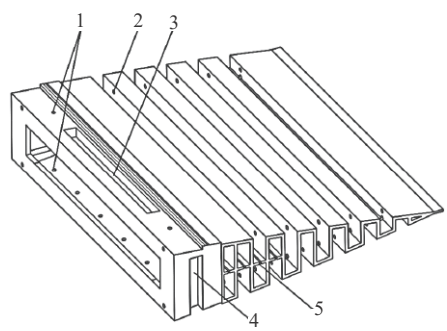
图 1 柔性变弯度后缘机翼设计

Fig. 1 Design of wing with variable-camber flexible trailing edge

翼段 4 和翼段 5 距前缘 65% 翼弦位置处设计了变弯度柔性结构。由于柔性后缘偏转可以改变机翼的气动力，其作用与传统铰链舵面相同，因此，在后文的风洞试验中，将翼段 4 的变弯度后缘称为给定扰动舵面，并将其用于扰动机翼产生动响应；翼段 5 的变弯度后缘称为响应控制舵面，用于减缓扰动舵面引起的机翼动响应。

该机翼变弯度主体结构为树脂 3D 打印的波纹板。由于波纹凸缘结构能承受较大法向气动载荷，且面内易于变弯，故选用并改进波纹结构为变弯度柔性后缘基本结构，如图 2 所示。其中，1 处螺纹孔及 3、4 处卡槽用于连接上、下翼段。同时为防止后缘偏转过程中波纹结构产生弦向压缩，确保蒙皮质量，还在前几段波纹结构中设计了闭室卡槽。

波纹区域从 65% 翼弦位置到 95% 翼弦位置。变弯度柔性后缘上下表面蒙皮选用 3M 公司出品的 VHB4910，也被称为丙烯酸酯蒙皮，弹性模量为 2.4 GPa，泊松比为 0.35。同时设计了预拉伸装置，对蒙皮施加预紧力后贴合到波纹板表面。为实现快速偏转响应，同时考虑到在动态偏转情况下驱动机构的可靠性，本文的连续变弯度机翼柔性后缘驱动机构采用航模舵机带动凯夫拉纤维驱动的形式。驱动装置选用 HS-7980 舵机，其扭矩为 38～46 kg·cm。舵机摇臂与变弯度后缘点通过凯夫拉柔



1. 螺纹孔；2. 螺纹孔；3. 卡槽；4. 卡槽；5. 闭室卡槽。

图 2 变弯度柔性后缘结构

Fig. 2 Structure of variable-camber flexible trailing edge

性索穿过图 2 中编号 2 处的小孔连接, 再通过摇臂的转动使后缘上偏或下偏。变弯度柔性后缘的结构设计及驱动形式如图 3 所示。

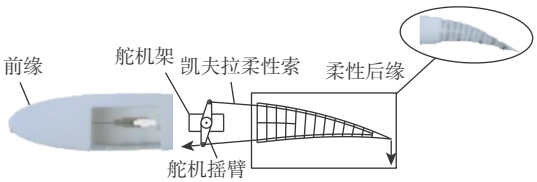


图 3　变弯度柔性后缘结构设计及驱动形式

Fig. 3　Structural design and driving form of variable-camber flexible trailing edge

当摇臂转动带动后缘偏转时, 对变弯度柔性后缘静变形的定义如图 4 所示。其中, 红色目标外形是通过在最后一块竖波纹板上施加一个沿柔性蒙皮切向方向的集中拉力, 再提取柔性后缘翼型中线与波纹板交点处的变形位移, 并采用二次多项式对波纹板中弧线变形曲线进行拟合得到的变形后翼型上下表面目标形状。

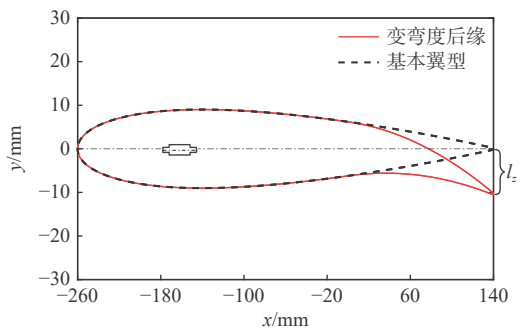


图 4　变弯度柔性后缘静变形定义

Fig. 4　Definition of static deformation of variable-camber flexible trailing edge

如图 4 所示, x为后缘弦向坐标, y为后缘运动后垂直翼型中线方向坐标, 以原点为后缘起始点, 定义后缘向下偏转为正。只考虑垂直翼型中线方向位移, 不考虑弦向变形, 定义后缘端点的变形量 l_z为后缘静变形。后续研究中所提舵面偏转位移均指后缘静变形。

通过地面振动试验得出了该结构的固有模态特性, 为后续风洞动响应试验提了供结构参考。振动试验现场照片如图 5 所示。连续变弯度机翼模型通过支架固支, 竖直放置。在模型上设置了 13 个锤击点, 每个点锤击 5 次取平均值。在翼尖的 1 号锤击点处布置一个加速度传感器, 采集连续变弯度机翼的加速度响应。通过分析激励信号和响应信号之间的关系, 可以得到连续变弯度机翼的模态振型和模态频率, 前 3 阶模态频率测量结果如表 1 所示。



图 5　地面振动试验

Fig. 5　Ground vibration test

表 1　地面试验测量的前 3 阶模态频率

Table 1　First three modal frequencies measured by ground tests

模态名称	频率/Hz
面外一弯	2.3
面外二弯	14.9
一扭	24.5

1.1　变弯度后缘静态偏转试验

该试验是通过 NI 数据采集系统给定舵机偏转信号, 旨在探究输入信号幅值与后缘变形的关系。同时试验采用 VIC-3D 非接触式视频测量系统来验证柔性后缘偏转形式。

试验中采用 NI 数据采集系统提供舵机偏转输入信号, 在输入信号幅值为 10°、15°、20°、25°、30° 时测定柔性变弯度后缘的变形量, 如图 6 所示。

由图中数据可知, 后缘静变形量与 NI 输入信号幅值呈正比关系。经数据拟合可得, NI 输入信号幅值与后缘静变形量的函数关系为

$$y = 0.532\,1x - 0.576\,4 \tag{1}$$

式中: x为 NI 输入信号幅值, (°); y为后缘变形量, mm。同时提取各个 NI 输入信号下后缘翼型表面标记的散斑坐标进行拟合处理。根据拟合的二次曲线得

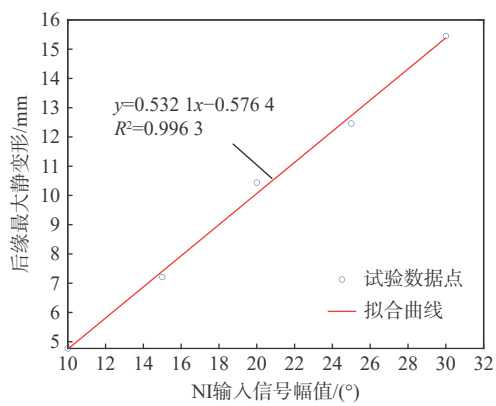


图 6　输入信号幅值与后缘静变形的关系

Fig. 6　Relationship between amplitude of input signal and static deformation of trailing edge

到输入信号幅值为 10°～30°时的变弯度柔性后缘偏转，如图 7 所示。

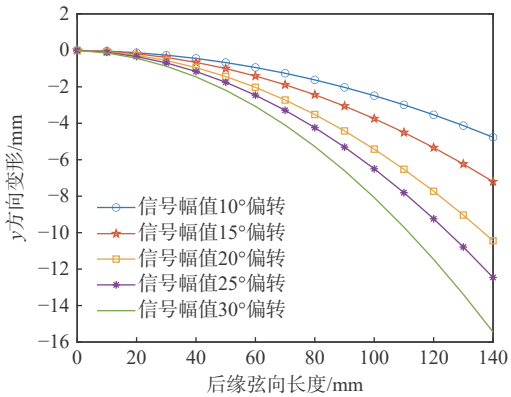


图 7 不同输入信号幅值下变弯度柔性后缘静态偏转

Fig. 7 Static deflection of variable-camber flexible trailing edge at different amplitudes of input signal

本文设计的变弯度柔性后缘偏转形式为二次曲线形式。取输入信号为 20°时的试验拟合结果与设计后缘的仿真结果进行对比，如图 8 所示。

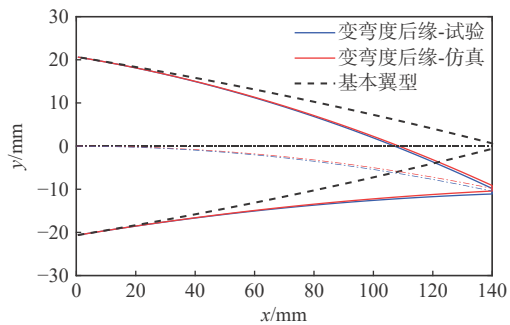


图 8 试验后缘变形拟合曲线与设计仿真结果对比（输入信号幅值 20°）

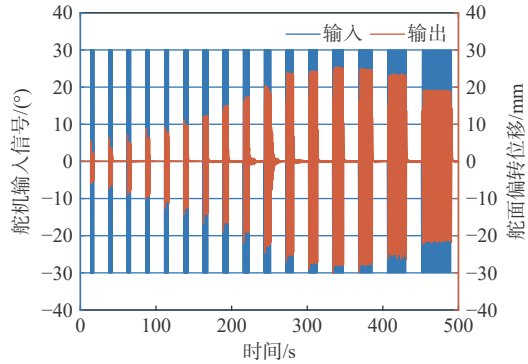
Fig. 8 Comparison between fitting curve of test trailing edge deformation and simulation results (amplitude of input signal of 20°)

由图 8 可知，变弯度柔性后缘的设计仿真偏转曲线与地面静态偏转试验拟合曲线基本吻合，从而验证了本文所设计柔性后缘的偏转形式为二次曲线，保证了后缘连续变弯度的特性。

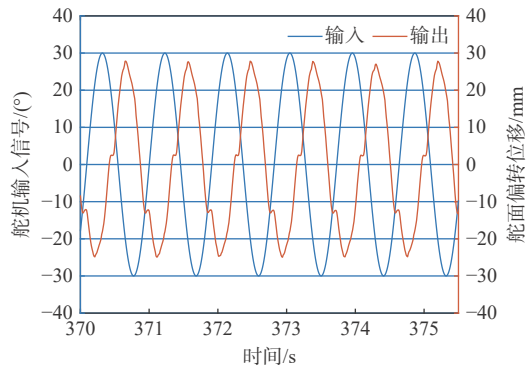
1.2 变弯度后缘动态偏转试验

通过后缘动态偏转特性试验，探究舵机输入与变弯度后缘动态偏转的幅值和相位关系。通过 NI 数据采集系统给定的舵机步进正弦扫频偏转信号，给定幅值为 30°，频率从 5 Hz 递减到 0.5 Hz，间隔为 0.3 Hz，输入的舵机信号与输出的后缘偏转位移信号如图 9 所示。

对分段对采集到的偏转位移信号进行正弦拟合，将输出的偏转位移信号的幅值、相位与 NI 输入信号的幅值、相位进行比较，得到后缘偏转位移信号的伯德图如图 10 所示。



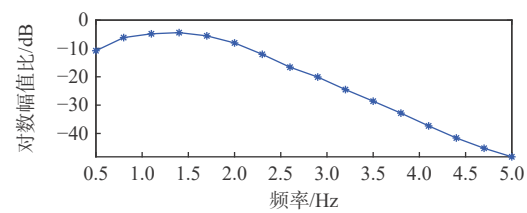
(a) 动态偏转试验位移输入输出



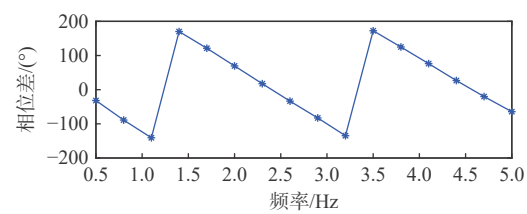
(b) 1.1 Hz输入信号局部放大

图 9 动态偏转试验位移输入输出

Fig. 9 Displacement input and output of dynamic deflection test



(a) 幅频特性曲线



(b) 相频特性曲线

图 10 后缘偏转位移信号伯德图

Fig. 10 Bode diagram of displacement signal of trailing edge deflection

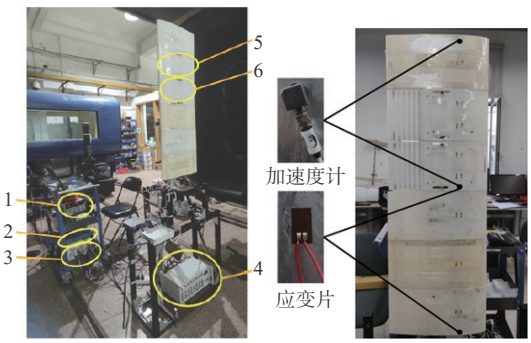
由图 10 可知，输入信号幅值为 30°时，后缘偏转位移随频率的增加先增大后减小。同时，输入信号与后缘偏转位移的相位差呈周期性变化趋势，且在周期内频率增加 0.3 Hz，相位滞后 50°。这对该驱动形式下的动响应减缓特性造成了很大影响。

2　变弯度柔性后缘动响应风洞试验

2.1　试验装置

将第 1 节中设计并制造的机翼模型用于研究在气流作用时，由扰动舵面引起的动响应特性，以及对控制舵面施加闭环主动控制律的动响应减缓特性。试验在北京航空航天大学的 D-4 低速风洞中进行。风洞长、宽、高均为 1.5 m，风速为 10～80 m/s。

为测量柔性变弯度后缘动态偏转时机翼的动响应信号，首先，采用激光位移计测量变弯度后缘动态偏转位移；其次在翼梁靠近机翼翼尖、翼段中部位置安装加速度计，并在靠近机翼中部、翼梁根部位置安装应变片，如图 11 所示。



1. 稳压电源；2. 信号调理仪；3. NI数采系统；4. 动态应变仪；5. 机翼模型控制舵面的翼段；6. 机翼模型扰动舵面的翼段。

图 11　风洞试验装置

Fig. 11　Wind tunnel test equipment

本文试验采用 NI 数据采集系统输入和采集信号，使用 Labview 编写数据输入采集程序，主要采集模型翼尖加速度信号、翼根部应变信号。采样频率为 100 Hz。风洞试验装置如图 11 所示。

该试验研究在典型风速、给定舵机驱动指令的条件下，在不同驱动频率下的变弯度后缘机翼的动响应和后缘偏转幅值。试验工况如下：试验风速为 12～20 m/s，间隔为 2 m/s；舵机正弦驱动幅值为 20°，驱动频率为 1.5 Hz、2.0 Hz、2.2 Hz、2.4 Hz、2.6 Hz、3.0 Hz、3.5 Hz、4.0 Hz。

2.2　试验结果

以风速为 20 m/s、后缘偏转频率为 2.2 Hz 的工况为例，采集到的翼尖加速度时域和频域响应如图 12 所示。

对于翼尖加速度，本文采用了巴特沃斯低通滤波器进行时域信号滤波，滤波器阶数为 10，采样频率为 100 Hz，截止频率为 5 Hz。以风速为 20 m/s、舵面激励频率为 2.2 Hz 为例，滤波前后的翼尖加速度响应对比如图 13 所示。

随后对滤波后的时域动响应信号用正弦函数进行拟合，得到不同工况下动响应的幅值。试验中

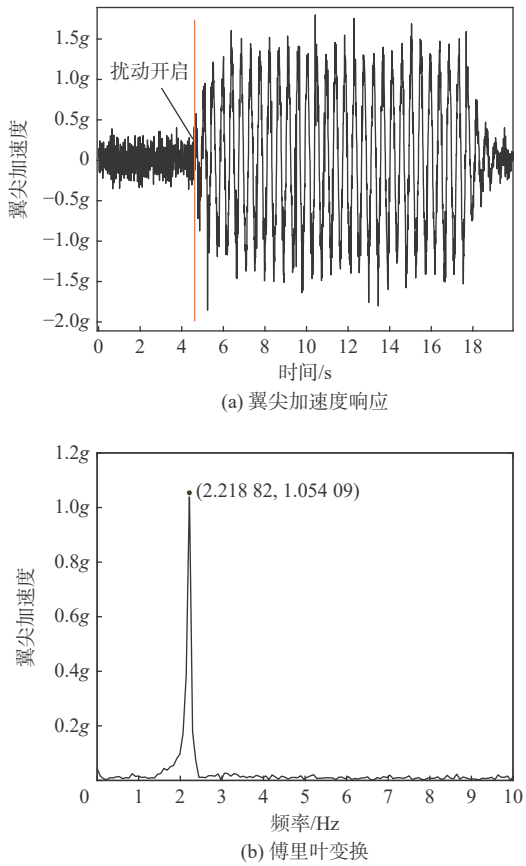


图 12　风速为 20 m/s、后缘偏转频率为 2.2 Hz 工况

Fig. 12　Wing tip acceleration response at wind speed of 20 m/s and trailing edge deflection frequency of 2.2 Hz

翼尖加速度和翼根弯矩幅值随舵面激励频率和风速变化关系如图 14 和图 15 所示。

由图 14 和图 15 可知，翼尖加速度与翼根弯矩变化趋势相似。翼尖加速度和翼根弯矩幅值随舵面激励频率的增加先增大后减小，并于 2.2 Hz 时达到峰值，这与机翼面外一弯（2.3 Hz）频率接近。一方面，由于舵面偏转频率逐渐靠近机翼固有模态频率，由舵面偏转引起的非定常气动力激励频率靠近机翼一弯模态频率，因而引起机翼受迫振动响应的幅值增大；另一方面，由于动压增加，舵面激励的气

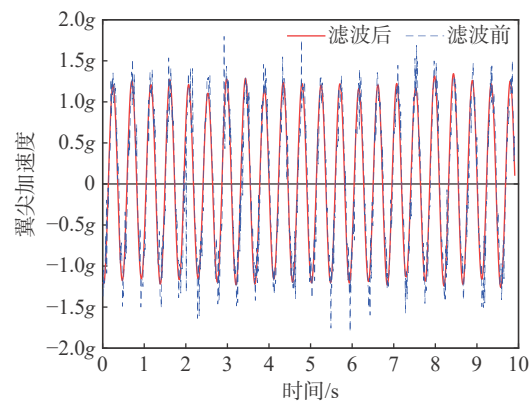


图 13　滤波前后的对比

Fig. 13　Comparison before and after filtering

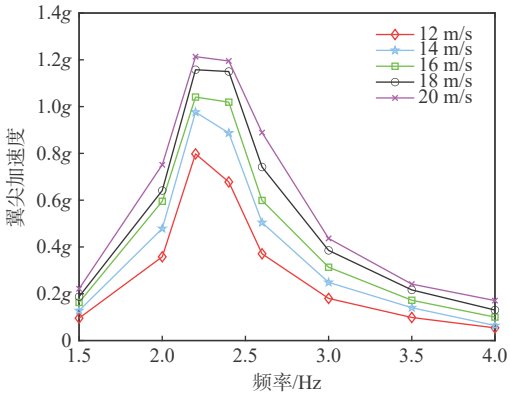


图 14 动响应试验翼尖加速度随频率变化关系

Fig. 14 Relationship between wing tip acceleration and frequency in dynamic response test

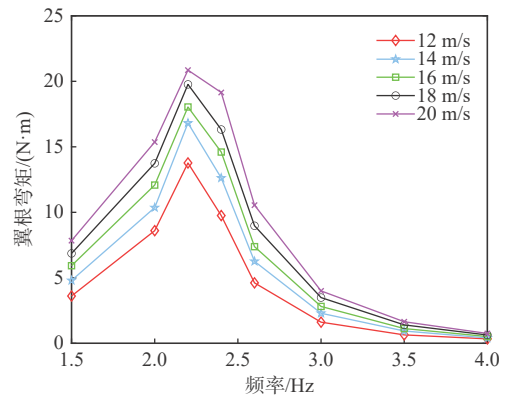


图 15 动响应试验翼根弯矩随频率变化关系

Fig. 15 Relationship between wing root bending moment with frequency in dynamic response test

动能量增加，所以翼尖加速度和翼根弯矩幅值随着风速的增加而逐渐增加。

3 动响应减缓试验

3.1 控制律设计

动响应减缓试验是通过给激励舵面正弦偏转信号，驱动激励舵面偏转产生非定常气动力，进而使弹性机翼产生动响应。待动响应稳定后，在主动控制律的作用下，驱动控制舵面偏转，从而减缓机翼动响应。

在试验中，给定扰动舵面正弦偏转信号的同时采集翼尖加速度和翼根应变信号，设计 PID 控制器参数为 k_p=−120、k_i=0、k_d=0，PID 反馈控制输入为翼尖加速度信号，控制输出为控制舵面舵机偏转信号。由于动响应试验中采集到的翼尖加速度信号存在较多噪声，经过 PID 控制器放大后会影响输出信号，所以考虑在控制回路中增加一个一阶 15 Hz 低通滤波器，滤波器传递函数如下：

$$T(s)=\frac{30\pi}{s+30\pi} \tag{2}$$

最终系统的控制框如图 16 所示。其中，$r(t)$ 为期望的动响应，$c(t)$ 为实际的动响应数据，$e(t)$ 为误差项，满足 $e(t)=r(t)-c(t)$，$u(t)$ 为控制舵面的舵偏输入信号。

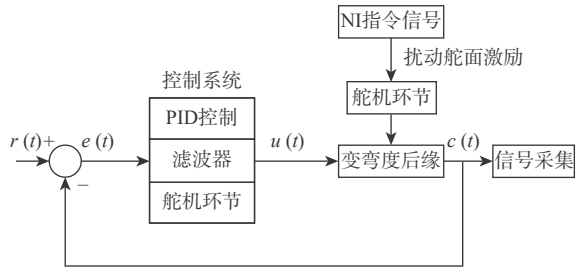


图 16 PID 控制框图

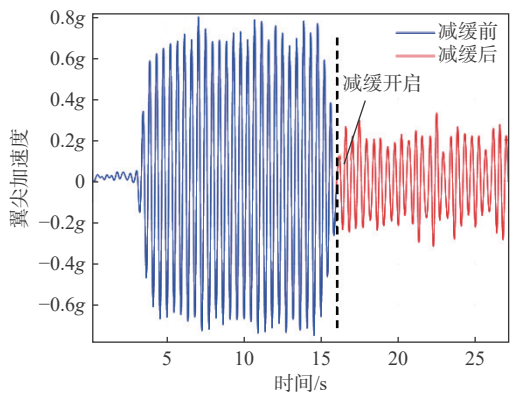
Fig. 16 PID control block diagram

3.2 试验结果

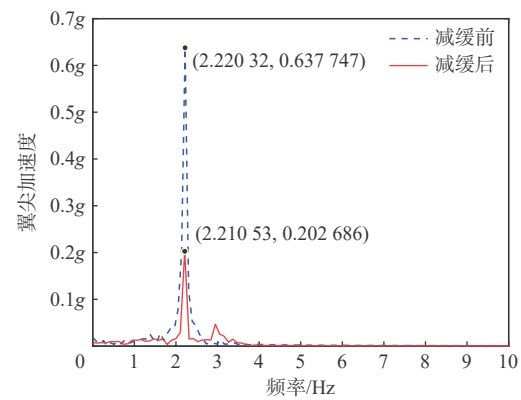
试验中采集到的部分数据如图 17 所示。

该减缓试验仍采用动响应试验所选的滤波器和滤波器参数进行滤波处理。控制舵面开启前后翼尖加速度响应幅值和翼根弯矩幅值随频率变化关系如图 18 所示。图 18(a)～图 18(h)为风速分别为 12 m/s、16 m/s、18 m/s、20 m/s 时舵面开启前和开启后翼尖加速度和翼根弯矩的对比。

由图 18 可知，翼尖加速度与翼根弯矩整体的变化趋势相似，对比控制舵面开启前、后两者的峰



(a) 翼尖加速度响应

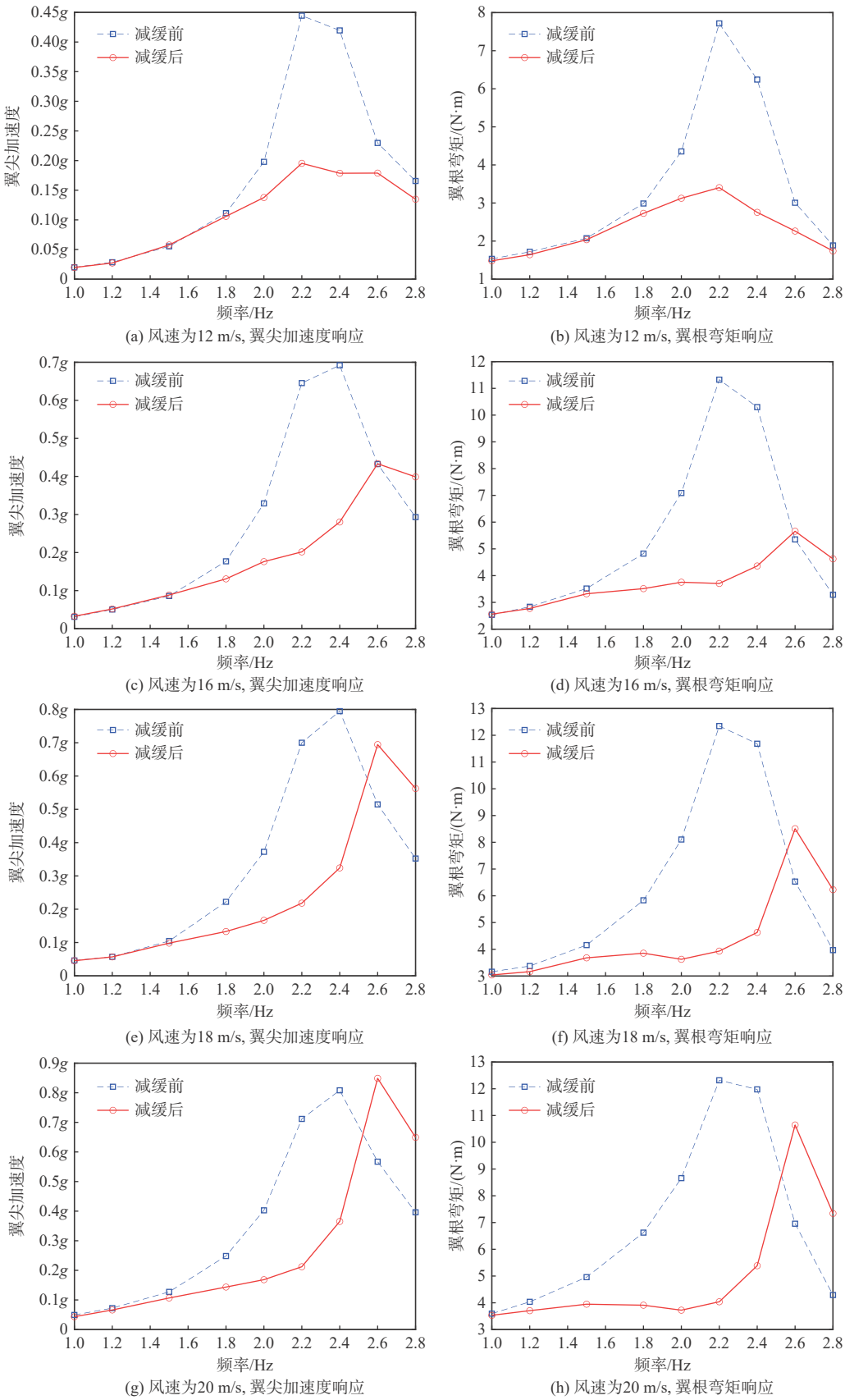


(b) 傅里叶变换

图 17 风速为 20 m/s，后缘偏转频率为 2 Hz 时动响应减缓

Fig. 17 Dynamic response alleviation at 20 m/s wind speed and 2.2 Hz trailing edge deflection frequency

值, 可知, 减缓效果较为明显。在控制舵面开启前翼尖加速度和翼根弯矩幅值随着频率的增加而先增大后减小, 并在靠近机翼一弯模态频率附近达到峰值, 这与第 2 节中的动响应试验结果的趋势基本一致; 在控制舵面开启后翼尖加速度和翼根弯矩幅值同样随着频率的增加而先增大后减小, 但其峰值



(a) 风速为12 m/s, 翼尖加速度响应　(b) 风速为12 m/s, 翼根弯矩响应

(c) 风速为16 m/s, 翼尖加速度响应　(d) 风速为16 m/s, 翼根弯矩响应

(e) 风速为18 m/s, 翼尖加速度响应　(f) 风速为18 m/s, 翼根弯矩响应

(g) 风速为20 m/s, 翼尖加速度响应　(h) 风速为20 m/s, 翼根弯矩响应

图 18　舵面开启前后动响应对比

Fig. 18　Comparison of dynamic responses before and after rudder surface opening

对应频率逐渐后移，且随风速的增大，峰值频率后移也增大。

在得到舵面开启前、后动响应之后，通过减缓效率的计算公式可以得出在一定风速下，各个频率对应的动响应减缓效率。减缓效率的计算公式如下：

$$\eta = (A_1 - A_2)/A_1 \times 100\% \tag{3}$$

式中：η为动响应减缓效率；A_1为舵面开启前动响应；A_2为舵面开启后动响应。翼尖加速度和翼根弯矩幅值减缓效率随频率的变化关系如图 19 和图 20 所示。

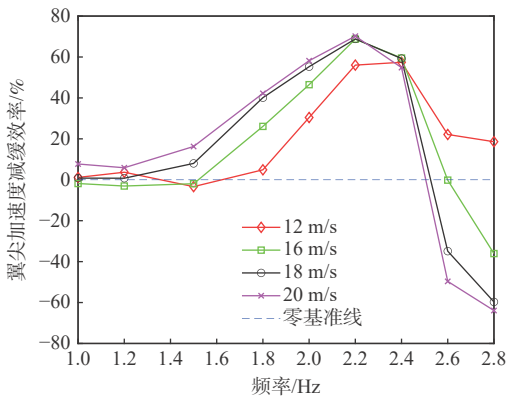


图 19 翼尖加速度减缓效率

Fig. 19 Wing tip acceleration alleviation efficiency

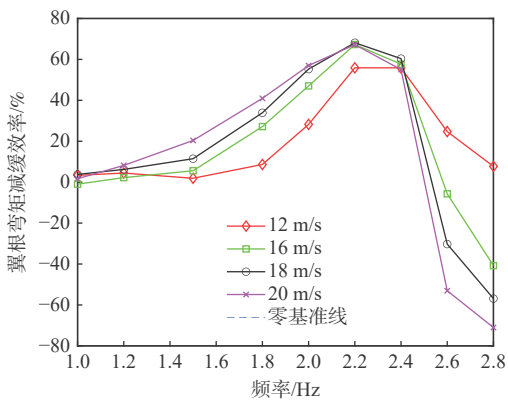


图 20 翼根弯矩减缓效率

Fig. 20 Wing root bending moment alleviation efficiency

3.3 动响应减缓效率分析

由图 19 和图 20 可知，翼尖加速度与翼根弯矩减缓效率整体的变化趋势相似。减缓效率随风速的增大而增大，这是由于在相同舵偏下，风速越大控制偏转力也越大，同时减缓效率随着频率的增加先增大后减小。当频率小于 1.6 Hz 时，减缓效率较低且随风速变化的规律不明显，这是由于此时机翼振动不强烈，反馈加速度信号较小且带有明显噪声；2～2.4 Hz 时的减缓效果较为明显，这是由于机翼一弯模态频率在此频率区间内，在相同激励的情况下，机翼一弯模态附近频率的机翼模态激励振动最为强烈如图 14 所示，而此时反馈的加速度信号也更大，控制舵面的偏转位移增加，如图 21 所示，减缓效果最为明显，在风速 20 m/s、舵面扰动频率 2.2 Hz 时翼尖加速度最大减缓效率达到 70.18%，翼根弯矩的最大减缓效率则为 68.14%。

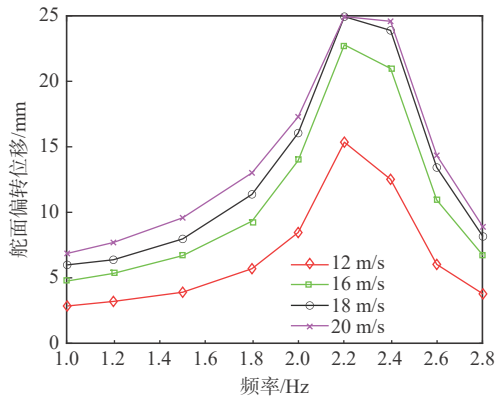


图 21 控制舵面偏转位移

Fig. 21 Deflection displacement of rudder surface

2.4 Hz 频率以上时动响应减缓效率不明显，且随着风速的增大减缓效率明显降低甚至达到负值。接下来对动响应减缓效率作进一步分析。开环系统的结构动力学方程为

$$\boldsymbol{M}\ddot{\boldsymbol{x}} + \boldsymbol{C}\dot{\boldsymbol{x}} + \boldsymbol{K}\boldsymbol{x} = \boldsymbol{f} \tag{4}$$

式中：$\boldsymbol{M}$为广义质量矩阵；$\boldsymbol{C}$为广义阻尼矩阵；$\boldsymbol{K}$为广义刚度矩阵；$\boldsymbol{x}$为广义模态坐标；$\boldsymbol{f}$为广义气动力。

当开启 PID 控制之后，扰动舵面和控制舵面引起的气动力可写为

$$\boldsymbol{F}_0 = \boldsymbol{\Phi}_0 \boldsymbol{f} \tag{5}$$

$$\boldsymbol{F}_1 = \boldsymbol{\Phi}_1 \boldsymbol{f} \tag{6}$$

式中：$\boldsymbol{\Phi}_0$为扰动舵面偏转模态；$\boldsymbol{\Phi}_1$为控制舵面偏转模态。由于舵面以正弦函数形式偏转，故扰动舵面偏转施加的激励气动力为$F_0\sin(\omega t + \varphi_0)$、控制舵面偏转施加的激励气动力为$F_1\sin(\omega t + \varphi_1)$。作差可得

$$F_0\sin(\omega t + \varphi_0) - F_1\sin(\omega t + \varphi_1) = \sqrt{F_0^2 - 2F_0F_1\cos(\varphi_0 - \varphi_1) + F_1^2}\sin(\omega t + \varphi) \tag{7}$$

其中，相位φ满足：

$$\sin\varphi = \frac{F_0\sin\varphi_0 - F_1\sin\varphi_1}{\sqrt{F_0^2 - 2F_0F_1\cos(\varphi_0 - \varphi_1) + F_1^2}} \tag{8}$$

$$\cos\varphi = \frac{F_0\cos\varphi_0 - F_1\cos\varphi_1}{\sqrt{F_0^2 - 2F_0F_1\cos(\varphi_0 - \varphi_1) + F_1^2}} \tag{9}$$

将式 (6) 右端项幅值与F_1对比可得出减缓效率为正的条件为

$$\cos(\varphi_0 - \varphi_1) > F_0/2F_1 \tag{10}$$

式中：F_0和φ_0分别为扰动舵面偏转产生激励力的幅

值和相位; F_1和φ_1分别为控制舵面偏转产生激励力的幅值和相位。则$\varphi_0-\varphi_1$为扰动舵面与控制舵面产生激励力的相位差。

假设扰动舵面与控制舵面的偏转位移与产生的激励力成正比, 则可以通过 NI 输入的指令信号和采集到的控制舵面偏转位移信号计算 2 种激励力的幅值比和相位差。扰动舵面偏转产生激励力的幅值F_0可由 NI 输入的指令信号对应 1.2 节中测得的舵机传函求出, 控制舵面偏转产生激励力的幅值F_1可由采集到的控制舵面偏转位移信号求出。

由图 16 可知, NI 输入信号还要经过舵机环节。由式 (7) 可知, 因为默认扰动舵面与控制舵面产生激励力有 180°的相位差, 所以满足:

$$\varphi_0-\varphi_1=\beta_0-\beta_1+180 \tag{11}$$

式中: β_0为 NI 输入的指令信号和采集到的控制舵面偏转位移信号相位差; β_1为舵机环节引起的相位滞后。

对频率为 2.6～2.8 Hz、风速为 12～20 m/s 的动响应减缓效率分析如表 2 和表 3 所示。

表 2　2.6 Hz 动响应减缓效率分析

Table 2　Dynamic response alleviation efficiency at 2.6 Hz

风速/(m·s^{-1})	β_0/(°)	β_1/(°)	$\cos(\varphi_0-\varphi_1)$	$F_1/2F_0$
12	−151	−33	0.469	0.465
16	−198	−33	0.967	1.186
18	−113	−33	−0.171	1.457
20	−94	−33	−0.481	1.705

表 3　2.8 Hz 动响应减缓效率分析

Table 3　Dynamic response alleviation efficiency at 2.8 Hz

风速/(m·s^{-1})	β_0/(°)	β_1/(°)	$\cos(\varphi_0-\varphi_1)$	$F_1/2F_0$
12	−205	−67	0.758	0.426
16	−277	−67	0.853	1.109
18	−288	−67	0.747	1.299
20	−318	−67	0.301	1.414

结合式 (9) 及表 2 和表 3 中的数据可知, 控制系统本身造成的相位滞后与舵机时滞效应产生的相位滞后对试验减缓效率有较大影响。动响应减缓是否有效取决于控制舵面和扰动舵面产生的激振力的相位差和幅值比。当频率为 2.6～2.8 Hz、风速达到 16 m/s 以上时, 控制舵面产生的激励力幅值增加, 无法满足不等式 (10), 则减缓效率将变为负值, 这便从理论上分析了图 19 和图 20 的减缓效率的机制。

4　结　论

1) 本文设计并制造了一种利用波纹板结构和由电机控制的凯夫拉绳无缝偏转的变弯度柔性后缘机翼模型, 并对其进行了地面试验, 探究了模型的模态特性及舵机特性。

2) 研究表明, 在扰动舵面激励频率为 1.5～4 Hz、风速为 12～20 m/s 的情况下, 翼尖加速度及翼根弯矩随舵面激励频率的增加先增大后减小, 且均在模型的一弯频率附近达到峰值。

3) 试验在风速为 20 m/s、扰动频率为 2.2 Hz 时取得对动响应的最大减缓效率。翼尖加速度最大减缓效率为 70.18%, 翼根弯矩的最大减缓效率为 68.14%。影响动响应减缓的主要因素为舵机时滞效应和控制系统本身造成的相位差, 以及控制舵面和扰动舵面偏转引起的激振力的比值。

参考文献 (References)

[1] MCGOWAN A M R, VICROY D D, BUSAN R C, et al. Perspectives on Highly adaptive or morphing aircraft[C]//NATO RTO AVT-168 Symposium. Lisbon: NTRS, 2009: RTO-MP-AVT-168.

[2] MAJID T, JO B W. Comparative aerodynamic performance analysis of camber morphing and conventional airfoils[J]. Applied Sciences, 2021, 11(22): 10663.

[3] 祝连庆, 孙广开, 李红, 等. 智能柔性变形机翼技术的应用与发展[J]. 机械工程学报, 2018, 54(14): 28-42.
ZHU L Q, SUN G K, LI H, et al. Intelligent and flexible morphing wing technology: A review[J]. Journal of Mechanical Engineering, 2018, 54(14): 28-42 (in Chinese).

[4] SOFLA A Y N, MEGUID S A, TAN K T, et al. Shape morphing of aircraft wing: Status and challenges[J]. Materials & Design, 2010, 31(3): 1284-1292.

[5] THILL C, ETCHES J, BOND I, et al. Morphing skins[J]. The Aeronautical Journal, 2008, 112(1129): 117-139.

[6] BARBARINO S, BILGEN O, AJAJ R M, et al. A review of morphing aircraft[J]. Journal of Intelligent Material Systems and Structures, 2011, 22(9): 823-877.

[7] 冷劲松, 孙健, 刘彦菊. 智能材料和结构在变体飞行器上的应用现状与前景展望[J]. 航空学报, 2014, 35(1): 29-45.
LENG J S, SUN J, LIU Y J. Application status and future prospect of smart materials and structures in morphing aircraft[J]. Acta Aeronautica et Astronautica Sinica, 2014, 35(1): 29-45 (in Chinese).

[8] ISMAIL N I, ZULKIFLI A H, ABDULLAH M Z, et al. Optimization of aerodynamic efficiency for twist morphing MAV wing[J]. Chinese Journal of Aeronautics, 2014, 27(3): 475-487.

[9] HUANG R, QIU Z P. Transient aeroelastic responses and flutter analysis of a variable-span wing during the morphing process[J]. Chinese Journal of Aeronautics, 2013, 26(6): 1430-1438.

[10] KOREANSCHI A, SUGAR GABOR O, ACOTTO J, et al. Optimization and design of an aircraft's morphing wing-tip demonstrator for drag reduction at low speed, Part I—Aerodynamic optimization using genetic, bee colony and gradient descent algorithms[J]. Chinese Journal of Aeronautics, 2017, 30(1): 149-163.

[11] KOREANSCHI A, GABOR O S, ACOTTO J, et al. Optimization

and design of an aircraft's morphing wing-tip demonstrator for drag reduction at low speeds, Part II - Experimental validation using Infra-Red transition measurement from Wind Tunnel tests[J]. Chinese Journal of Aeronautics, 2017, 30(1): 164-174.

[12] LI D C, LIU Q C, WU Y N, et al. Design and analysis of a morphing drag rudder on the aerodynamics, structural deformation, and the required actuating moment[J]. Journal of Intelligent Material Systems and Structures, 2018, 29(6): 1038-1049.

[13] AJAJ R M, PARANCHEERIVILAKKATHIL M S, AMOOZGAR M, et al. Recent developments in the aeroelasticity of morphing aircraft[J]. Progress in Aerospace Sciences, 2021, 120: 100682.

[14] MOLINARI G, QUACK M, ARRIETA A F, et al. Design, realization and structural testing of a compliant adaptable wing[J]. Smart Material Structures, 2015, 24(10): 105027.

[15] TAKAHASHI H, YOKOZEKI T, HIRANO Y. Development of variable camber wing with morphing leading and trailing sections using corrugated structures[J]. Journal of Intelligent Material Systems and Structures, 2016, 27(20): 2827-2836.

[16] WOODS B K S, FINCHAM J H S, FRISWELL M I. Aerodynamic modelling of the fish bone active camber morphing concept[C]// Proceedings of the RAeS Applied Aerodynamics Conference. Bristol: The Aeromatical Journal, 2014: 2224.

[17] ZHAO S W, LI D C, ZHOU J, et al. Numerical and experimental study of a flexible trailing edge driving by pneumatic muscle actuators[J]. Actuators, 2021, 10(7): 142.

[18] 雷朝辉, 宋晨, 张桢锴, 等. 后缘柔性可变形翼肋模块设计分析与试验[J]. 工程与试验, 2020, 60(4): 16-22.
LEI C H, SONG C, ZHANG Z K, et al. Design, analysis and test of morphing trailing edge flexible rib[J]. Engineering & Test, 2020, 60(4): 16-22 (in Chinese).

[19] PECORA R, AMOROSO F, SICIM M S. Design of a morphing test-article for large-scale, high-speed wind tunnel tests of an adaptive wing flap[C]//Active and Passive Smart Structures and Integrated Systems XV. San Diego: SPIE, 2021, 11588: 7-15.

[20] RIVERO A E, FOURNIER S, MANOLESOS M, et al. Wind tunnel comparison of flapped and FishBAC camber variation for lift control[C]// Proceedings of the AIAA Scitech 2020 Forum. Reston: AIAA, 2020: 1300.

[21] VARTIO E, SHAW E, VETTER T. Gust load alleviation flight control system design for a SensorCraft vehicle[C]// Proceedings of the 26th AIAA Applied Aerodynamics Conference. Reston: AIAA, 2008: 7192.

[22] WONG A, BIL C, MARINO M. Design and aerodynamic performance of a FishBAC morphing wing[C]// Proceedings of the AIAA SCITECH 2022 Forum. Reston: AIAA, 2022: 1298.

[23] LU C. Development of variable camber continuous trailing edge flap for performance adaptive aeroelastic wing[C]// Multidisciplinary Design Analysis and Optimization of Aerospace Composites. Piscataway: IEEE Press, 2020: 67-122.

[24] NGUYEN N T, CRAMER N B, HASHEMI K E, et al. Progress on gust load alleviation wind tunnel experiment and aeroservoelastic model validation for a flexible wing with variable camber continuous trailing edge flap system[C]// Proceedings of the AIAA Scitech 2020 Forum. Reston: AIAA, 2020: 0214.

Experiment on dynamic response alleviation of a wing with variable-camber flexible trailing edge

ZHANG Yuming，DAI Yuting*，WEI Rukai，HU Yating，YANG Chao

(School of Aeronautic Science and Engineering，Beihang University，Beijing 100191，China)

Abstract：Morphing aircraft can significantly enhance the aerodynamic performance of the aircraft, and variable camber flexible trailing edge is one of the important ways to achieve this. To investigate the dynamic response characteristics and alleviation efficiency of the wing under the dynamic deflection of flexible trailing edge, a variable camber flexible trailing edge wing model was designed and a wind tunnel tests was conducted. The wing model consisted of a bending wing beam and six 3D-printed wing panels. Two variable-camber flexible trailing edge rudder surfaces were installed at the trailing edge of two wing panels. These two rudder surfaces were used for dynamic response excitation and dynamic response alleviation control, respectively. The variable-camber flexible trailing edge rudder surface was composed of a digital actuator, flexible cables, a corrugated plate structure, and a flexible polydimethylsiloxane (PDMS) skin. Ground static and dynamic deflection tests were carried out for the variable-camber flexible trailing edge, so as to investigate the camber deformation law of the trailing edge and the dynamic time-delayed characteristics of the actuator. On this basis, a low-speed wind tunnel test was carried out to investigate the dynamic response law of the wing with variable-camber flexible trailing edge and the dynamic response alleviation efficiency based on variable-camber flexible trailing edge and closed-loop feedback control. The wind tunnel test results show that the wing tip acceleration response and the wing root bending moment increase first and then decrease in the frequency range of 1.5–4 Hz, and it reaches the peak value when approaching the first bending frequency of the wing. After closed-loop feedback control by proportional-integral-derivative (PID) control law and variable-camber flexible trailing edge, the maximum alleviation efficiency of the wing tip acceleration and wing root bending moment is 70.18% and 68.14%, respectively, at a wind speed of 20 m/s and disturbance frequency of 2.2 Hz. A theoretical formula with positive dynamic response alleviation efficiency was proposed, and the influencing mechanism and factors of dynamic response alleviation efficiency were analyzed.

Keywords：flexible trailing edge; variable camber; dynamic response; alleviation efficiency; wind tunnel test

Received：2022-09-07；**Accepted**：2022-11-28；**Published Online**：2022-12-28 08：35

URL：link.cnki.net/urlid/11.2625.V.20221226.1412.002

Foundation item：National Natural Science Foundation of China (11672018)

* **Corresponding author**. E-mail：yutingdai@buaa.edu.cn

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2022.0768

进口拐角波系对超燃冲压发动机尾喷管的性能影响

陈冲，葛建辉*，徐惊雷，苏鹏，俞凯凯
(南京航空航天大学 能源与动力学院，南京 210016)

摘　　要：为了探究进口拐角波系对超燃冲压发动机喷管的壁面压力分布作用机制，考虑进口拐角波系对超燃冲压发动机尾喷管性能的影响，开展不同拐角波及其强度对喷管气动参数影响的数值模拟分析。进行数值方法有效性和网格无关性验证，利用不同的前后台阶来形成进口拐角波系，研究了不同波系和波系强度对喷管气动性能的影响。结果表明，进口拐角波系造成的进口非均匀对喷管气动性能造成显著影响，其中推力系数在上下壁面存在后台阶时最大增加13.98%，升力在下壁面前台阶时提高94.32%，而俯仰力矩则在上壁面前台阶、下壁面后台阶时最多降低35.47%。当喷管进口存在单侧前台阶时，随台阶高度增加，升力与俯仰力矩的变化趋势一致；而喷管进口存在单侧后台阶时，结论相反。此外，壁面台阶对喷管性能的影响是相互独立的，且基本符合线性叠加原理。

关　键　词：超燃冲压发动机；喷管；膨胀波/激波；前/后台阶；气动性能
中图分类号：V231.3
文献标志码：A　　**文章编号**：1001-5965(2024)10-3250-12

高超声速飞行器因其广泛而不可替代的应用价值，得到了各航空航天大国的竞相研究[1-3]。超燃冲压发动机作为其主要的动力装置，具有显著的比冲优势[4-6]。超燃冲压发动机主要由进气道、隔离段、燃烧室以及喷管组成[7]。喷管是超燃冲压发动机的重要推力部件，相关研究表明，在高马赫数下其推力可以达到整个发动机推力的70%以上[8-9]。因此，对超燃冲压发动机尾喷管的性能研究具有重要意义。

超燃冲压发动机气流经过进气道、隔离段及燃烧室后会产生显著畸变，使喷管进口气流参数呈高度非均匀。由一维流理论可知：入口的非均匀气流会改变喷管进口动量和进口波系，显著影响喷管的性能。张晓源等[10]对基于煤油燃料超燃冲压发动机尾喷管的非均匀进口进行了研究，发现非均匀进口使得尾喷管净推力比均匀进口增加了1.2%左右。全志斌等[11]为了研究马赫数非均匀进口对超燃冲压发动机尾喷管性能的影响，进行了相关数值模拟研究和风洞试验。研究结果表明，非均匀进口会造成尾喷管推力下降2.92%～5.02%、负升力增加17.20%、俯仰力矩减小4.20%～6.70%。张璞等[12]针对基于最短长度理论设计的超燃冲压发动机单边膨胀喷管，在流量平均、总压不变的条件下，研究了由激波/膨胀波造成的马赫数非均匀进口对喷管性能的影响。结果表明，激波/膨胀波造成的马赫数非均匀入口使得喷管推力性能下降3%～6%，负升力达到均匀进口情况下升力的85.92%～117.20%，俯仰力矩减小2.5%～7.0%。陈以勒等[13]分别研究了进口气流角畸变、马赫数畸变以及两者耦合畸变对喷管性能的影响，发现了在气流角畸变影响下，

收稿日期：2022-09-14；**录用日期**：2022-11-01；**网络出版时间**：2022-11-15 14:18
网络出版地址：link.cnki.net/urlid/11.2625.V.20221114.1428.001
基金项目：国家自然科学基金(11802123)
*通信作者. E-mail：jianhuig@nuaa.edu.cn

引用格式：陈冲，葛建辉，徐惊雷，等. 进口拐角波系对超燃冲压发动机尾喷管的性能影响[J]. 北京航空航天大学学报，2024，50(10)：3250-3261. CHEN C，GE J H，XU J L，et al. Influence of inlet corner wave system on performance of scramjet nozzle[J]. Journal of Beijing University of Aeronautics and Astronautics，2024，50(10)：3250-3261 (in Chinese).

升力和俯仰力矩的增幅最高分别可达 51.84%和 12.11%。此外, 文献 [13] 也发现了气流角和马赫数畸变对喷管性能的影响是相互独立的。Mo 等[14]针对单边膨胀喷管在非均匀来流情况下的性能参数展开了研究, 研究表明, 在非均匀畸变模拟器产生的马赫数非均匀度为 150% 进口气流的情况下, 升力系数会降低 6%～8%。在此基础上, 为了降低进口非均匀对喷管的性能影响, 文献 [14] 还提出一种考虑非均匀来流的超燃冲压发动机喷管设计方法, 轴向推力系数在设计点约提高 1.75%, 升力和俯仰力矩系数分别增加约 6.51% 和 6.35%。此外 Gaffeng 等[15]提出一种应用于马赫数 Ma=15 飞行条件下基于有旋特征线法(Method of Characteristics, MOC) 的超燃冲压发动机喷管设计方法, 从而在来流不均匀的情况下可产生平行的出口气流和均匀的出口静压。

上述文献主要研究了喷管进口马赫数非均匀以及气流角畸变对喷管的性能影响, 但没有考虑喷管进口处拐角波系引起的非均匀对喷管性能的影响, 而喷管进口的拐角波系是广泛存在的, 其影响喷管流场分布, 显著影响喷管壁面的压力分布, 进而影响喷管的性能, 但目前公开文献中鲜有进口拐角波系对超燃冲压发动机喷管性能影响的研究, 因此有必要开展相关研究。本文首先介绍了采用的数值模拟方法, 并开展了数值方法有效性校验和网格无关性验证; 其次选择不同的前/后台阶来研究不同拐角波系及强度对喷管性能的影响, 并分析相应的影响规律。

1　数值模拟方法

1.1　数值方法有效性验证

本文采用 ANSYS Fluent 进行数值模拟。求解器选择密度基, 对流通量计算选择 Roe 平均通量差分 (Roe averaged flux difference splitting, Roe-FDS) 格式, 空间离散格式选择二阶迎风格式, 湍流模型使用剪切应力输运 k-ε 模型 (shear stress transport model, SST k-ε)。为获得壁面处准确的流动细节, 将控制壁面处网格 y^+ 约为 1。定义计算域进口边界条件为质量流量入口, 远场边界条件为压力远场, 远场出口为压力出口, 壁面均设置为绝热壁面。

为验证本文采用的数值方法的有效性, 选择了文献 [16] 中喷管 AW2 对本文的数值模拟方法进行验证, AW2 喷管构型示意如图 1 所示, 图中以喉道高度 h_t 进行无量纲化。由图 1 可知, 喷管出口面积与喉道面积比为 1.51。关于喷管具体构型和进出口流动参数详见文献 [16]。使用的计算网格示意

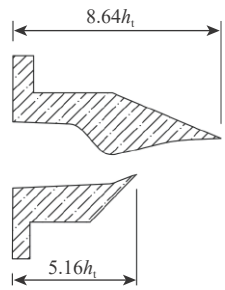


图 1　AW2 喷管构型示意

Fig. 1　AW2 nozzle configuration

如图 2 所示。数值模拟结果与实验得到的喷管上下壁面压力分布对比曲线如图 3 所示, 其中壁面压力分布以喷管进口总压 p^* 进行无量纲化, 由图 3 可知, 数值模拟得到的壁面压力分布与实验结果吻合良好, 喷管上壁面的数值结果存在明显波动, 这是由于文献 [16] 中给出的用于拟合喷管上壁面型线的点较少, 导致喷管上壁面光滑性差, 进而影响了喷管上壁面的数值结果。典型后台阶作用下的流场信息、后台阶的具体构型和来流流动参数详见文

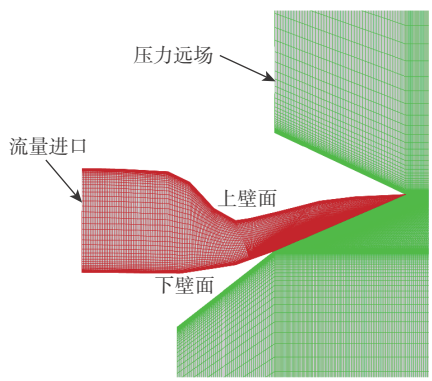


图 2　AW2 型喷管计算网格示意

Fig. 2　AW2 nozzle computing grid

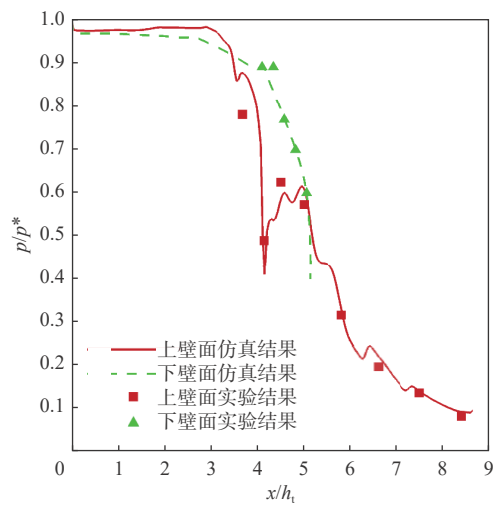


图 3　文献 [16] 中数值模拟和实验得到的壁面压力分布对比

Fig. 3　Comparison of wall pressure distribution between numerical simulation and experiment in reference [16]

献 [17], 数值模拟和实验获得的后台阶及其下游壁面压力分布对比曲线如图 4 所示, 以后台阶上游总压 p_0^*进行无量纲化。由图 4 可知, 数值模拟得到的台阶及其下游壁面压力分布与实验结果吻合良好。另外文献 [17] 还给出了沿流向距离后台阶结构 25 mm 处的竖直剖面上湍动能 k 的分布规律, 其中横坐标以当地边界层厚度δ进行无量纲化, 纵坐标以当地边界层外的主流速度的平方 U_e^2 进行无量纲化, 如图 5 所示。由图 5 可知, 虽然数值结果和实验结果略有偏差, 但二者整体分布趋势一致。由此证明了本文所采用的数值方法的有效性。

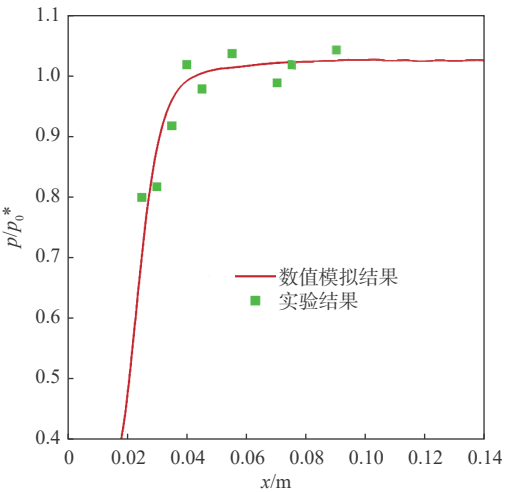


图 4 文献 [17] 中数值模拟和实验得到的壁面压力分布对比

Fig. 4 Comparison of wall pressure distribution between numerical simulation and experiment in reference [17]

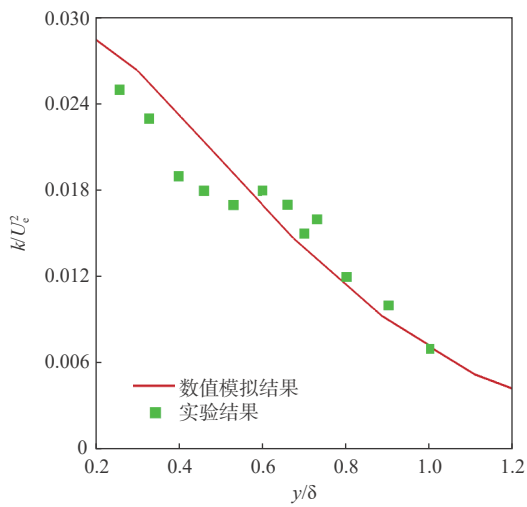


图 5 文献 [17] 中数值模拟和实验得到的湍动能分布对比

Fig. 5 Comparison of turbulent kinetic energy distribution between numerical simulation and experiment in reference [17]

1.2 网格无关性验证

本文所研究的喷管模型是基于 MOC 和最大推力理论设计出的超燃冲压发动机尾喷管[18], 并以此为基础进行进口拐角波系对喷管性能影响的研究。该喷管构型示意如图 6 所示, 以喷管进口高度 h_{in} 进行无量纲化, 出口与进口的面积比为 2.55。

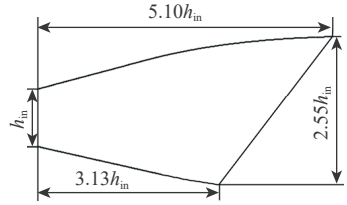


图 6 本文中喷管构型示意

Fig. 6 Nozzle configuration in this study

为减小网格分辨率对数值模拟结果的影响, 本节将进行网格无关性验证。网格划分工具使用 ICEM, 分别选择 7 万、14 万、28 万的网格量(定义为粗、中、细 3 种尺度网格)进行验证, 研究中通过偏置喷管上游的平直壁面来模拟前后台阶结构, 因此需在喷管入口前加上一等直管段作为实际计算域。其中, 中尺度网格示意如图 7 所示。壁面处第一层网格高度取 0.01 mm, 同时保证壁面处y^+约为 1。喷管计算域进口及出口气流参数如表 1 所示。在相同工况下, 利用不同尺度网格得到的下壁面压力对比曲线如图 8 所示, 其中壁面压力分布以计算域进口总压 p^*进行无量纲化。由图 8 可知, 在 x≈0.015 m 和 x≈0.260 m 附近, 对比中、粗尺度网格发现, 粗尺度网格的压力偏差比较大; 而中、细尺度网格计算得到的压力分布相互吻合。利用不同尺度网格得到的下壁面沿程摩擦系数对比曲线如图 9 所示, 由图 9 可知, 粗尺度网格的壁面摩擦系数偏差较大; 而中, 细尺度网格的壁面摩擦系数分布相互吻合。由此可知, 当网格量达 14 万后, 其数值结果不再受网格量的影响。因此, 为兼顾计算效率及精度, 在本文后续的数值模拟研究中将采用中等尺度网格。

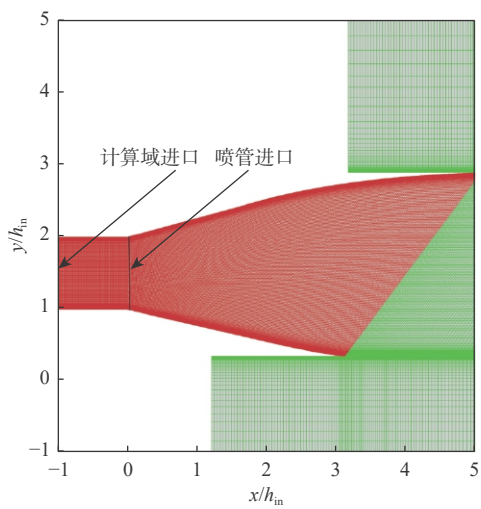


图 7 研究中使用的喷管计算网格示意

Fig. 7 Nozzle computing grid used in this study

表 1　计算域进出口流动参数

Table 1　Flow conditions of inlet and outlet in calculation domain

参数	数值
计算域进口马赫数Ma_{in}	1.76
计算域进口静压P_{in}/Pa	45 454.52
计算域进口静温T_{in}/K	1 821.97
环境压力P_e/Pa	3 466.0
环境温度T_e/K	219.57
巡航马赫数Ma_0	5.00
巡航高度H/km	23.00

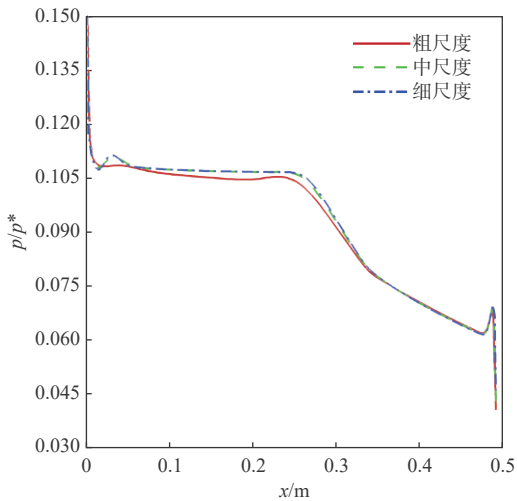


图 8　不同网格尺度获得的下壁面压力分布

Fig. 8　Pressure distributions on lower walls obtained from different grid scales

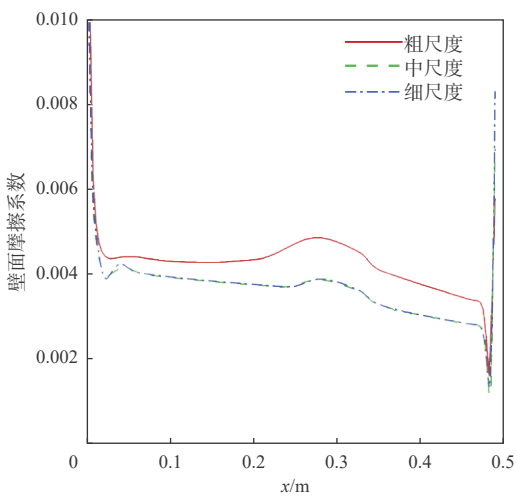


图 9　不同网格尺度获得的下壁面摩擦系数分布

Fig. 9　Friction coefficient distribution on lower walls obtained from different grid scales

2　拐角波系及相关气动参数定义

2.1　拐角波系定义

喷管流场的实验纹影如图 10 所示。由图 10 可知，喷管进口存在两道激波结构，并且在下游进一步反射。两道波系可能主要由以下两个原因导致的：1）从喷管的设计理念看，喷管的长度一般会受到严格限制，这使得喷管的初始膨胀段很短。加工精度的限制无法再现设计状态下极短的初始膨胀段，会导致高速气流快速膨胀进而形成激波结构；2）从喷管实际装配方面看，燃烧室出口与喷管进口由于装配限制容易形成前/后台阶，进而形成相应的波系结构。

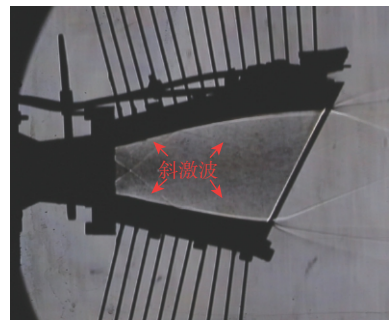


图 10　喷管流场的实验纹影

Fig. 10　Experimental schlieren diagram of nozzle flow field

本文采用上下偏移喷管上游平直壁面的方法形成前后台阶并利用不同的偏移距离代表台阶高度来形成不同强度的波系，以此来模拟实际喷管中由于初始膨胀段加工精度限制或者装配误差等引起的波系结构。超声速流场中前、后台阶形成波系的主要结构如图 11 所示。由图 11 可知，在前/后台阶作用下超声速流场中波系的结构非常复杂，存在激波、膨胀波以及波系与附面层的干扰等诸多结构。下文将主要针对单侧拐角波系影响和双侧拐角波系耦合影响展开研究。上壁面在前台阶作用下、下壁面在后台阶作用下的典型流场结构示意如图 12 所示。由图 12 可知，在上壁面存在前台阶时，超声速来流经过前台阶时结构会先形成激波，并在下壁面进行反射，随后气流经过激波在上壁面附近形成膨胀波系。下壁面存在后台阶时，超声速来流先向下壁面一侧偏转形成局部膨胀波扇。在膨胀波扇的影响下，当地气流方向会与下壁面产生夹角随后会形成一道斜激波，以实现气流方向与当地壁面切向的匹配。为了对问题进行更全面的研究，本文亦采用不同的台阶高度来研究不同波系强度对喷管性能的影响，由于未从公开文献中得知目前工程实际中超燃喷管装配偏差的具体范围，且本文侧重喷管内流动现象的理论研究，因此前后台阶的高度分别考虑了喷管进口高度的 0.5%、1%、3%、5%。前台阶高度用正值表示，后台阶高度用负值表示，台阶所处上、下壁面的位置用符号u和d表示，即$u_{+1\%}$表示上壁面存在相对于喷管进口高度为 1% 的前台阶。

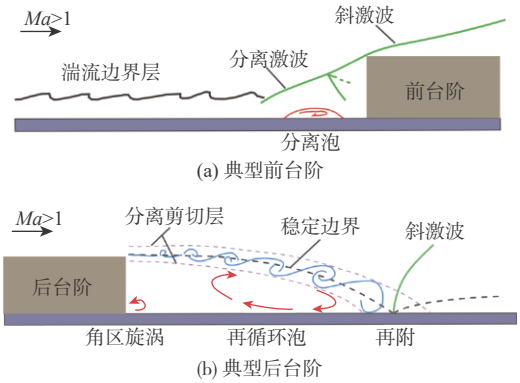


图 11 典型前后台阶下的流场结构示意

Fig. 11 Flow field structure under typical forward and backward step

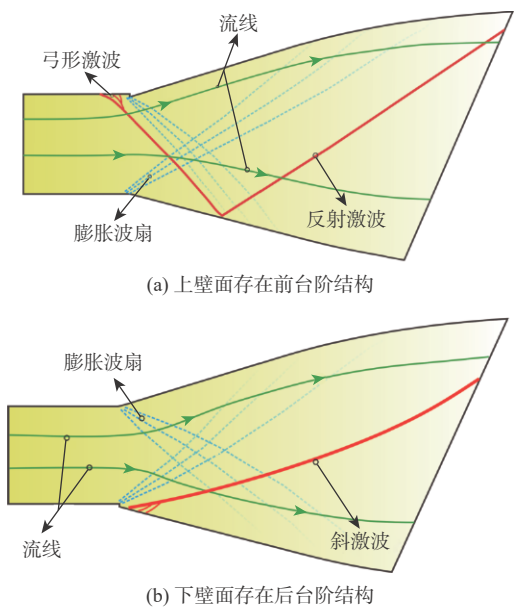


图 12 上壁面在前台阶作用下和下壁面在后台阶作用下的流场结构示意

Fig. 12 Flow field structure with forward and backward step on upper/lower wall

2.2 气动参数定义

本文主要研究喷管的推力系数 C_{fx}，其计算式为

$$C_{fx}=F/(F_{\mathrm{I}}-I_{\mathrm{in}}) \tag{1}$$

$$F_{\mathrm{I}}=\dot{m}\sqrt{\frac{2\gamma}{\gamma-1}RT^{*}\left[1-\left(\frac{P_{\mathrm{e}}}{p^{*}}\right)^{\frac{\gamma}{\gamma-1}}\right]} \tag{2}$$

$$\frac{p^{*}}{P_{\mathrm{in}}}=\left(\frac{T^{*}}{T_{\mathrm{in}}}\right)=\left(1+\frac{\gamma-1}{2}Ma_{\mathrm{in}}^{2}\right)^{\frac{\gamma}{\gamma-1}} \tag{3}$$

式中：F 为的喷管实际推力，其是由 fluent 读取的喷管上下壁面受力的水平分量之和；F_{I} 为喷管理想等熵推力；I_{in} 为喷管进口冲量；$\dot{m}$为流经喷管的质量流量；γ 为气体比热比（参考文献 [12-13] 中研究此类问题时采取了定比热比处理方法，因此本文认定 γ 为定值）；R 为气体常数；p^{*}为计算域进口总压；T^{*} 为计算域进口总温；俯仰力矩参考点的位置取在喷管进口中间处，本文规定低头力矩为正，反之为负。为了进行有效对比，本研究中保证不同状态下喷管的理想等熵推力与喷管进口冲量之差不变，即式 (1) 的分母不变。

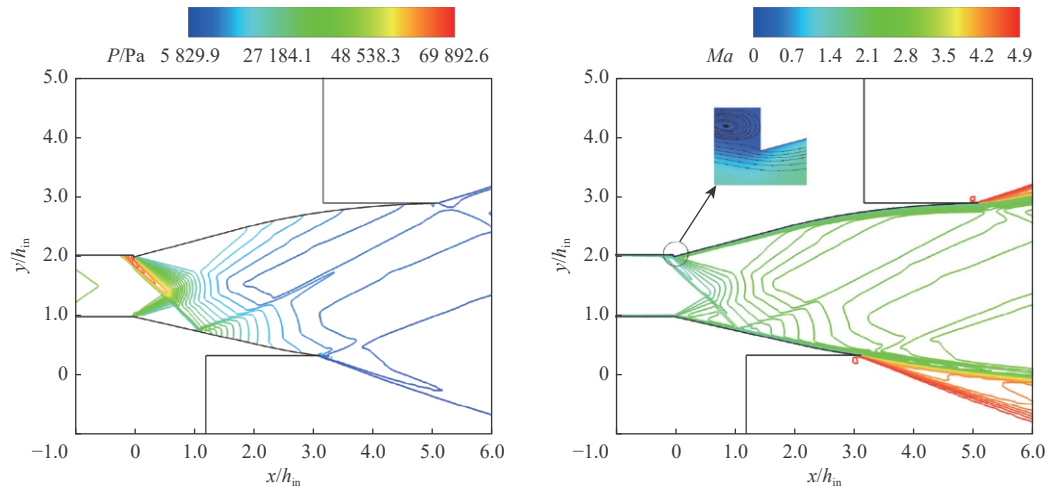
3 拐角波系对喷管气动性能的影响

3.1 单侧拐角波系对喷管气动性能的影响

3.1.1 上壁面拐角波系

本节首先考虑上壁面拐角波系对喷管气动性能的影响规律，在上壁面分别布置了 $u_{\pm0.5\%}$、$u_{\pm1\%}$、$u_{\pm3\%}$、$u_{\pm5\%}$共 8 种不同类型的台阶结构。$u_{+3\%}$台阶下喷管流场的压力和马赫数分布云图如图 13 所示，其中 P 为压力，Ma 为马赫数。由图 13 可知，当超声速气流经过上壁面的前台阶结构时会形成激波并在逆压梯度作用下产生微小分离区，随后在经过喷管的初始膨胀段时产生膨胀波系。该道激波的波角较大，因此其可在喷管下壁面进行反射，进而显著增加下壁面入射点后的压力分布。由于推力和升力分别是喷管壁面受力分布的水平和竖直分量，因此可以推断前台阶的加入可能会降低喷管的升力，并增加喷管推力。超声速气流经过上壁面后台阶时流场的压力和马赫数云图如图 14 所示。由图 14 可知，不同于前台阶下的喷管流场，超声速来流在后台阶的作用下，先向上壁面一侧偏转形成局部膨胀波扇。在膨胀波扇的影响下，当地气流方向会与上壁面产生夹角随后会形成一道斜激波，以实现气流方向与当地壁面切向的匹配。与同样台阶高度的$u_{+3\%}$情况相比，该道激波强度弱、波角小，因此从图 14 中也可以观察到其无法与喷管下壁面相交，但考虑上壁面后台阶结构引起的膨胀波和激波，喷管壁面的压力分布仍会受到影响，进而也会影响到喷管的气动性能。此外，图 13 和图 14 还显示喷管进口上游的等直段内存在因黏性干扰产生的相交波系，该波系强度较弱，因此其对气动性能的影响基本可忽略不计。

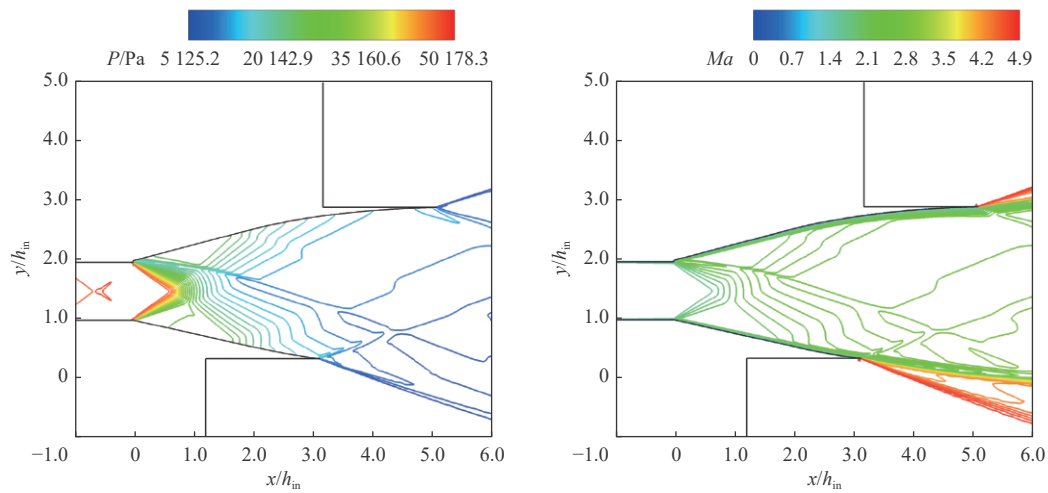
上壁面拐角波系作用下的喷管气动性能变化曲线如图 15 所示，图中 L 为升力，M 为俯仰力矩。由图 15 可知，前、后台阶结构均会使喷管推力增加，升力减小；对于俯仰力矩，二者导致的变化趋势正好相反：前台阶会减小俯仰力矩，后台阶反之。为进一步分析前、后台阶对喷管性能的影响机制，本节提取了$u_{\pm3\%}$、u_0（喷管进口无前台阶的情况）作用下的壁面压力分布，如图 16 所示。由图 16(c) 和图 16(d) 可知，前、后台阶结构均会引起喷管沿 y 方



(a) 压力分布　　(b) 马赫数分布

图 13　$u_{+3\%}$ 台阶下喷管流场压力和马赫数分布云图

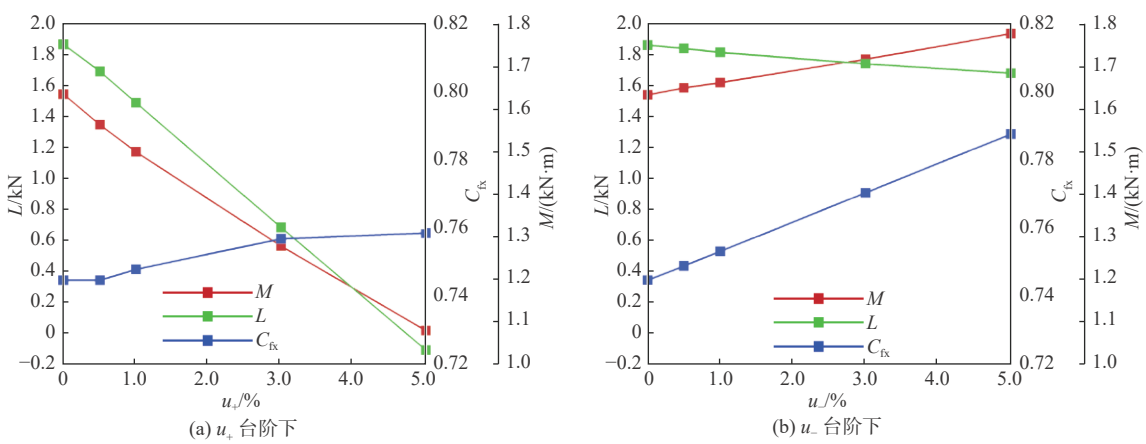
Fig. 13　Cloud charts of pressure and Mach number distributions in nozzle flow field under step of $u_{+3\%}$



(a) 压力分布　　(b) 马赫数分布

图 14　$u_{-3\%}$ 台阶下喷管流场压力和马赫数分布云图

Fig. 14　Cloud charts of pressure and Mach number distributions in nozzle flow field under step of $u_{-3\%}$

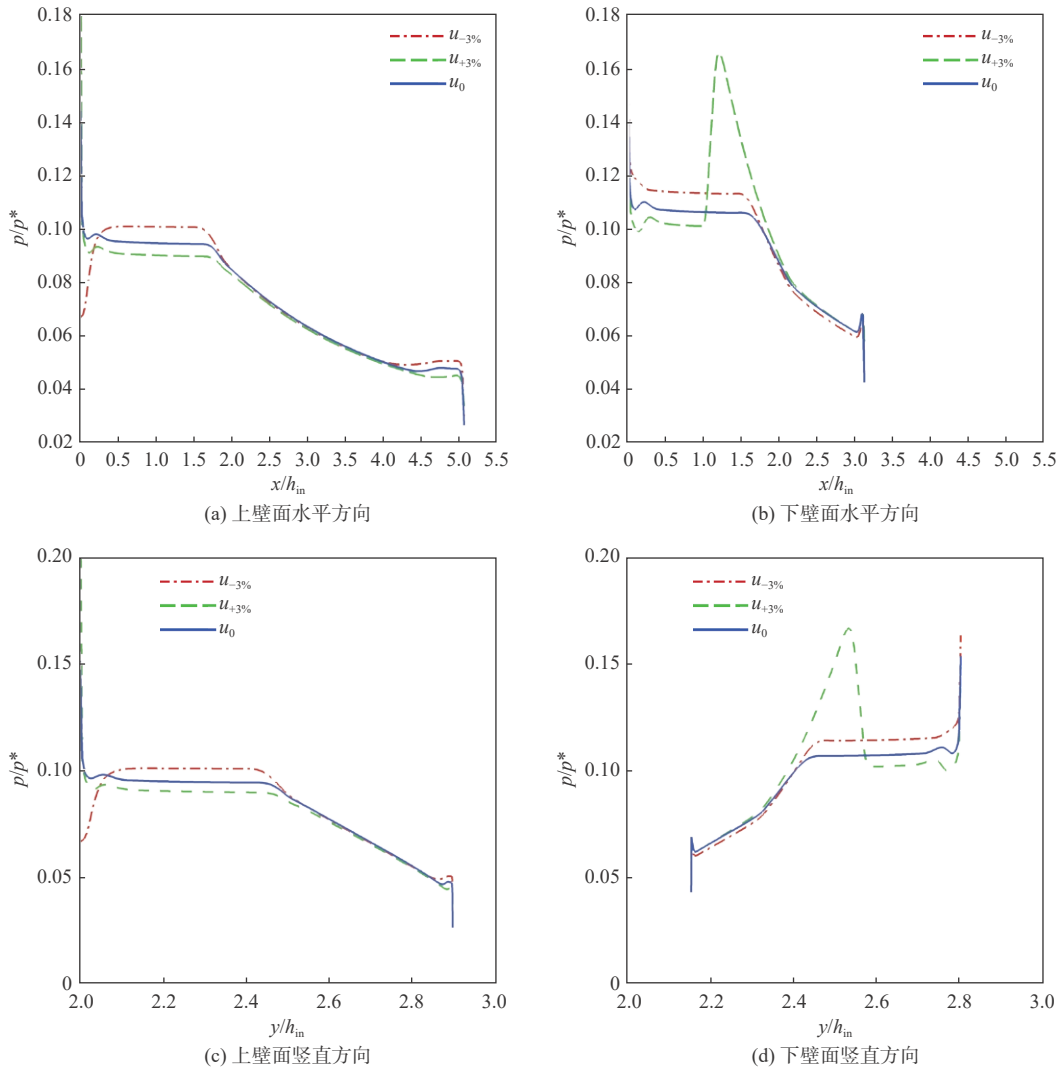


(a) u_+ 台阶下　　(b) u_- 台阶下

图 15　上壁面拐角波系下的喷管气动性能变化曲线

Fig. 15　Variation curves of aerodynamic performance of nozzle under corner wave system of upper wall

向壁面压力分布增加, 进而造成喷管推力的增加; 由于$u_{+3\%}$产生的激波强度更高且会在下壁面发生反射, 所以图 16(b) 中下壁面的$x/h_{\text{in}} \approx 1.3$处存在压力骤升区, 进而导致在$u_{+3\%}$情况下升力呈降低趋势, 甚至出现负升力, 而在$u_-$时, 喷管进口面积减小会让进口截面平均压力上升, 进而导致喷管壁面的沿



(a) 上壁面水平方向　(b) 下壁面水平方向

(c) 上壁面竖直方向　(d) 下壁面竖直方向

图 16　$u_{+3\%}$,$u_{-3\%}$和u_0作用下的壁面沿程压力分布

Fig. 16　Pressure distributions along walls under $u_{+3\%}$, $u_{-3\%}$, and u_0

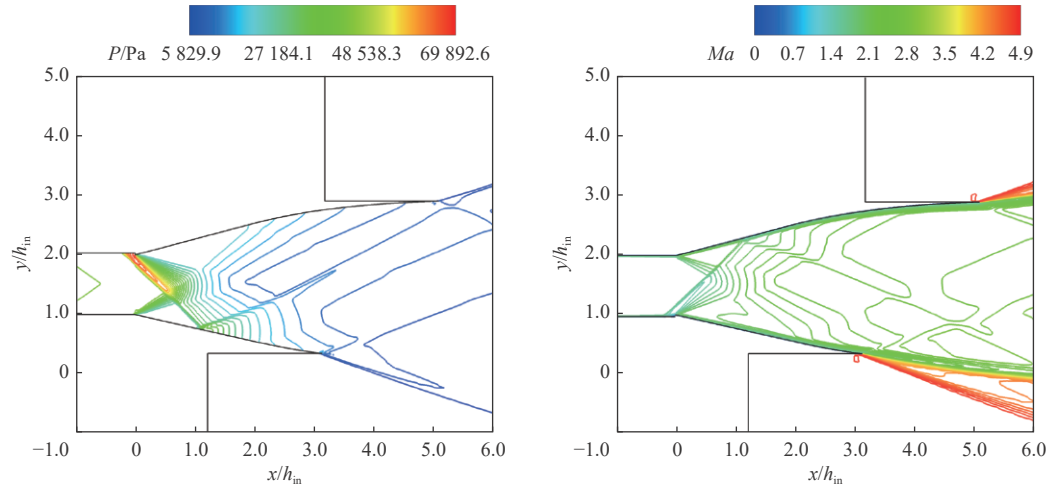
程压力分布相对于u_0呈整体上升趋势。尽管喷管上壁面进口拐角膨胀波会使上壁面进口附近形成低压区，但这个低压区较小，因此$u_{-3\%}$时喷管的推力会上升。

由数值模拟结果可知，在$u_{+3\%}$情况下，前台阶结构使喷管下壁面出现激波反射，导致喷管升力降低，这在很大程度上导致喷管俯仰力矩的下降；在$u_{-3\%}$时，尽管进口处仍有激波存在，但并没有在下壁面进行反射，这主要影响喷管上壁面的压力分布，在推升力的共同作用下，俯仰力矩呈上升趋势。

3.1.2　下壁面拐角波系

为研究下壁面拐角波系对喷管性能的影响规律，在下壁面设置了$d_{\pm0.5\%}$、$d_{\pm1\%}$、$d_{\pm3\%}$、$d_{\pm5\%}$共 8 台阶结构时的情况。在$d_{\pm3\%}$台阶结构时，喷管内部的压力和马赫数分布云图如图 17 和图 18 所示。由图 17 可知，来流经过下壁面的前台阶结构形成斜激波导致气流偏转，与$u_{+3\%}$类似，这道激波随后在喷管上壁面发生反射。在这道激波的下游区域，为消除气流与下壁面存在的夹角，超声速气流需在喷管下壁面进口处形成拐角膨胀波扇，以确保气流方向与当地壁面切线方向一致。结合 3.1.1 节的分析结果，喷管上壁面的激波反射区将显著改变上壁面的压力分布，进而使喷管推、升力均增加。由图 18 可知，下壁面进口处先形成膨胀波扇改变超声速气流方向，随后在膨胀波下游超声速流场产生激波使得下壁面附近气流方向与当地壁面切向平行。受进口拐角波系影响，此时喷管的气动性能也会改变。

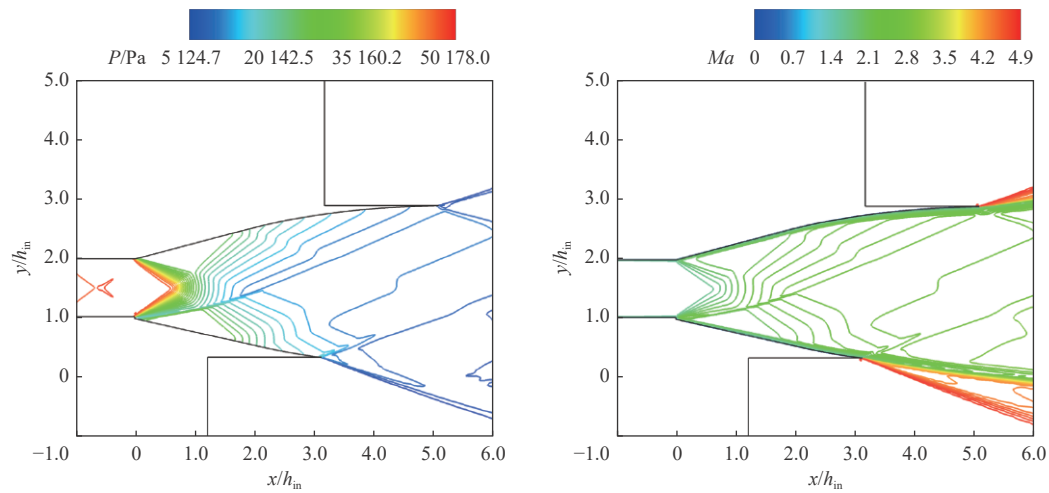
为进一步研究下壁面拐角波系对喷管性能的影响规律，提取了不同台阶类型和高度下的性能参数，如图 19 所示。由图 19 可知，下壁面台阶结构均会使喷管推力增加；在升力方面，喷管升力在$d_{+3\%}$时增加最大，而在$d_{-3\%}$下基本保持不变；此外，$d_{+3\%}$情况下俯仰力矩随着台阶高度的增加而增加，$d_{-3\%}$情况下俯仰力矩变化规律相反。为进一步分析喷管性能参数变化规律的成因，提取了喷管壁面的



(a) 压力分布　　(b) 马赫数分布

图 17　$d_{+3\%}$ 台阶下喷管流场压力和马赫数分布云图

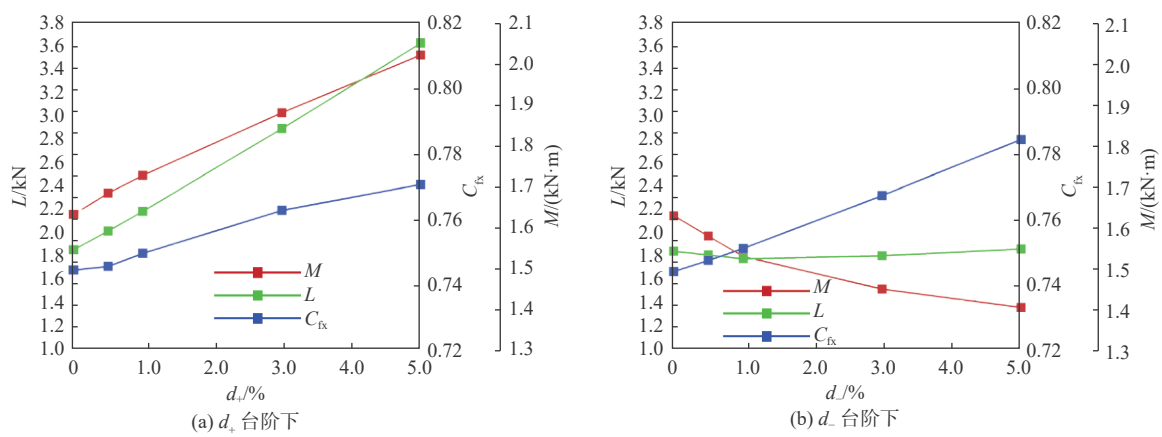
Fig. 17　Cloud charts of pressure and Mach number distributions in nozzle flow field under step of $d_{+3\%}$



(a) 压力分布　　(b) 马赫数分布

图 18　$d_{-3\%}$ 台阶下喷管流场压力和马赫数分布云图

Fig. 18　Cloud charts of pressure and Mach number distributions in nozzle flow field under step of $d_{-3\%}$

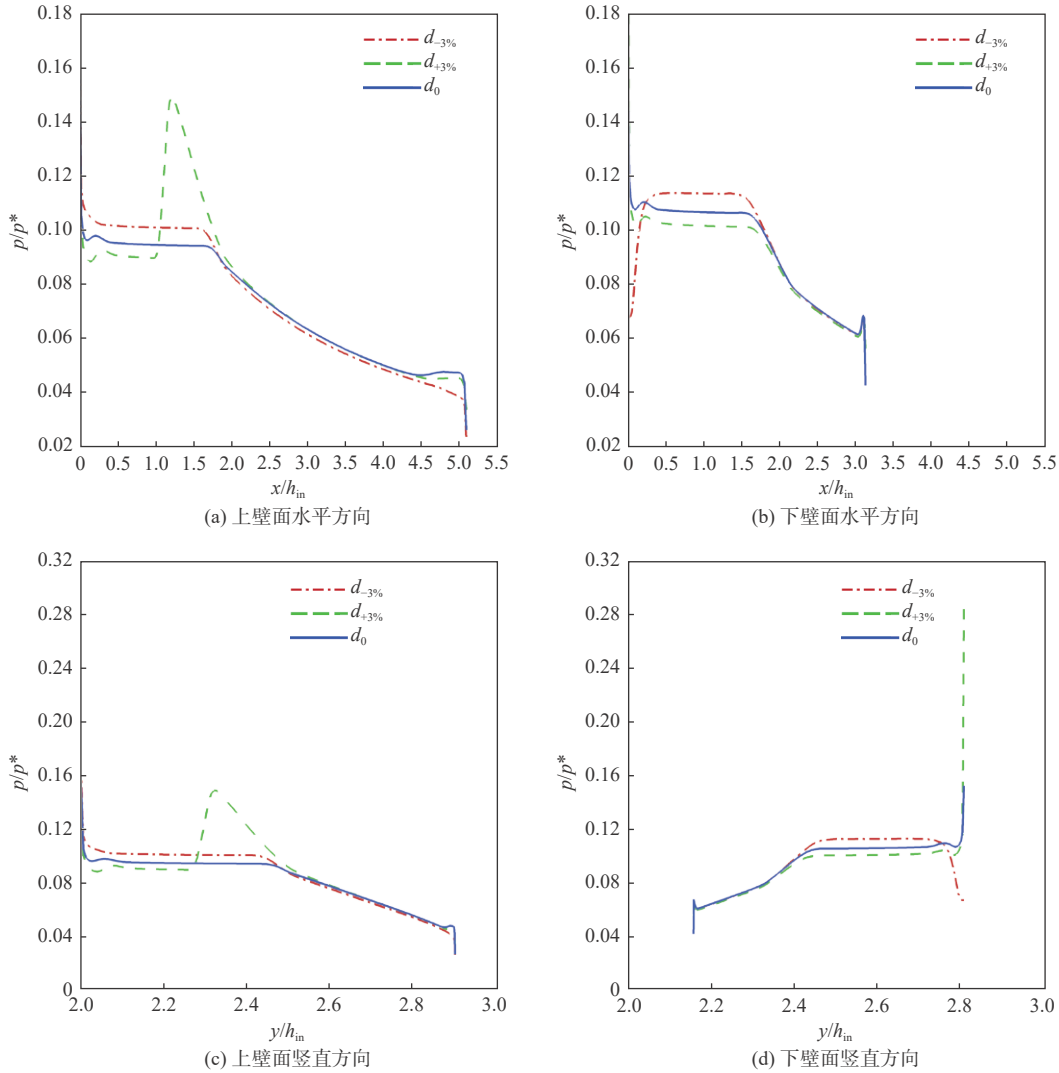


(a) d_+ 台阶下　　(b) d_- 台阶下

图 19　下壁面拐角波系下的喷管气动性能变化曲线

Fig. 19　Variation curves of aerodynamic performance of nozzle under corner wave system of lower wall

沿程压力分布, 如图 20 所示。由图 20 可知, 在 $d_{+3\%}$ 时, 喷管上壁面出现激波反射现象, 使得上壁面沿程压力出现骤升区, 从而增大了推力和升力, 也增加了俯仰力矩; 在 $d_{-3\%}$ 时, 喷管进口面积的减小会使得进口截面平均压力上升, 进而导致喷管壁面的沿程压力分布相对于 d_0(喷管进口无后台阶的情况)呈整体上升趋势。尽管喷管下壁面进口拐角膨胀波会使下壁面进口附近存在低压区, 但这个低



(a) 上壁面水平方向　(b) 下壁面水平方向

(c) 上壁面竖直方向　(d) 下壁面竖直方向

图 20　$d_{+3\%}$,$d_{-3\%}$、d_0作用下的壁面沿程压力分布

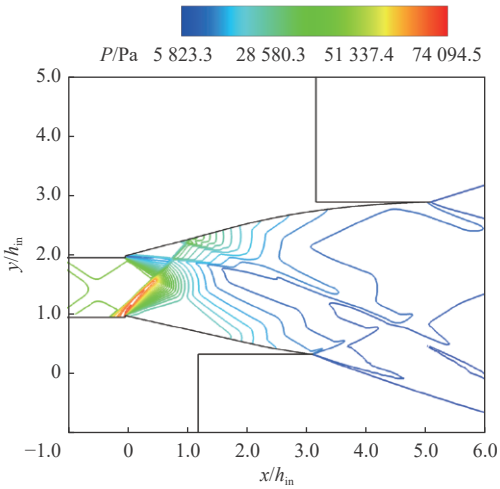
Fig. 20　Pressure distributions along walls under $d_{+3\%}$,$d_{-3\%}$, and d_0

压区较小，因此$d_{-3\%}$时喷管的推力上升。此外，由于$d_{-3\%}$时进口激波的波角较小，无法在喷管上壁面发生反射，所以其影响范围主要在下壁面，从而使俯仰力矩下降。

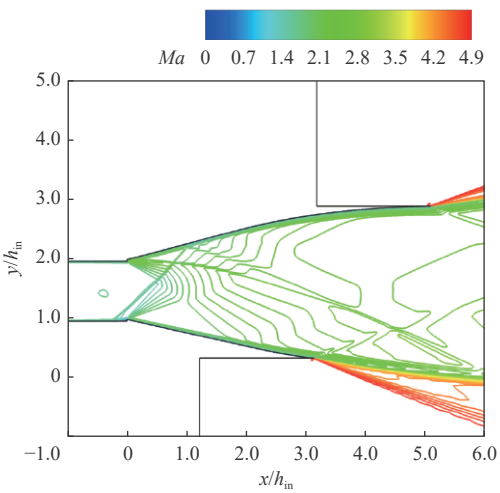
相比上壁面拐角波系作用下喷管的性能，台阶结构均使喷管推力增加。喷管进口激波，使在d_+作用下喷管的升力和俯仰力矩与u_+情况下的变化趋势相反；在u_-作用下喷管升力减小而d_-使喷管升力基本不变，u_-和d_-的俯仰力矩的变化趋势相反。综上所述，在单侧壁面的参数研究范围内，$u_{-5\%}$和$d_{-5\%}$情况下推力增幅接近且最大为5.40%，$d_{+5\%}$情况下升力最多可增加94.32%，$u_{+5\%}$情况下俯仰力矩最多可减小33.93%。

3.2　双侧拐角波系对喷管气动性能的耦合影响

为研究双侧壁面台阶结构对喷管性能的耦合影响规律，对$u_{\pm0.5\%}$、$u_{\pm1\%}$、$u_{\pm3\%}$、$u_{\pm5\%}$、$d_{\pm0.5\%}$、$d_{\pm1\%}$、$d_{\pm3\%}$、$d_{\pm5\%}$共16种构型进行了数值模拟研究。u_-d_+台阶构型时的喷管内部压力和马赫数云图如图21所示。由图21可知，超声速气流经过下壁面的前台阶结构时会先形成激波，该道激波的波角较大，因此可在喷管上壁面进行反射，随后在经过喷管的初始膨胀段时产生膨胀波扇。同时在上壁面进口处先形成膨胀波扇，在膨胀波扇下游超声速流场产生激波使得上壁面附近的气流方向与当地壁面切向平行。为了进一步分析双侧拐角波系对喷管气动性能的耦合影响，对单侧台阶和双侧台阶作用下的性能参数变化量进行比较。以升力为例，双侧u_-d_+与单独u_-和d_+作用下的升力变化曲线如图22所示。由图22可知，双侧u_-d_+时的升力变化量近似于单独u_-和d_+作用时升力变化量之和。这说明，壁面台阶结构对喷管性能的影响是相互独立的，基本遵循线性叠加原理。结合单侧台阶结构的结果，推测出双侧$u_{-5\%}d_{-5\%}$作用下喷管推力系数应增加最多，双侧$u_{-5\%}d_{+5\%}$时升力应增幅最大，双侧$u_{+5\%}d_{-5\%}$情况下俯仰力矩减小最多。u_+d_+,u_-d_-,u_+d_-,u_-d_+下的喷管气动参数随台阶高度的变化规律如



(a) 压力分布



(b) 马赫数分布

图 21　u_-d_+台阶下喷管流场压力和马赫数分布云图

Fig. 21　Cloud charts of pressure and Mach number distributions in nozzle flow field under step of u_-d_+

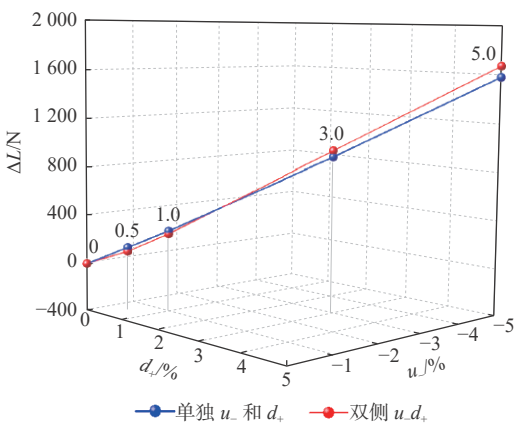
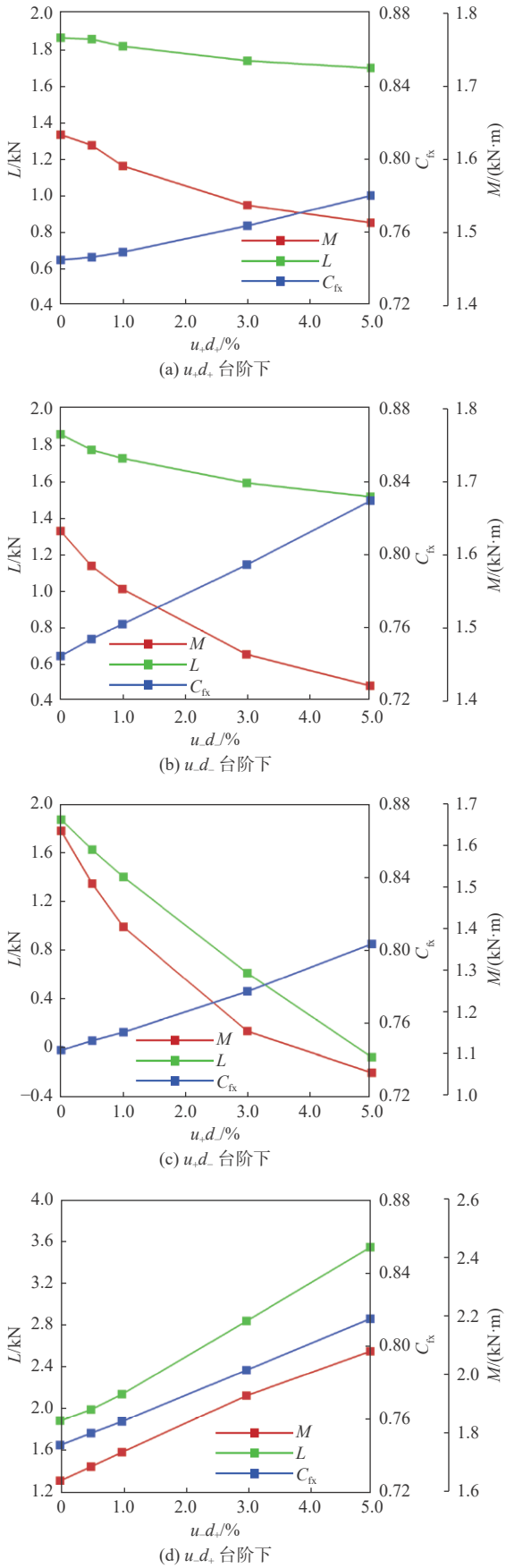


图 22　双侧u_-d_+与单独u_-和d_+作用下的升力变化曲线

Fig. 22　Comparison of nozzle lift variation under step of u_-d_+ and corresponding single-sided step type

图 23 所示。由图 23 可知，双侧$u_{-5\%}d_{-5\%}$情况下喷管推力系数增幅最大，在本文研究范围内的最大增幅可达到 13.98%，由于上壁面台阶在u_-情况下喷管



(a) u_+d_+台阶下

(b) u_-d_-台阶下

(c) u_+d_-台阶下

(d) u_-d_+台阶下

图 23　u_+d_+,u_-d_-,u_+d_-,u_-d_+下的喷管性能变化曲线

Fig. 23　Variation curves of aerodynamic performance of nozzle under u_+d_+,u_-d_-,u_+d_-, and u_-d_+

的升力降低最少，$u_{-5\%}d_{+5\%}$时喷管升力的增幅仅次于$d_{+5\%}$，由于下壁面台阶在$d_{-5\%}$时喷管俯仰力矩降低，所以喷管的俯仰力矩在u_+d_-时减小量略大于u_+时的减小量，减幅为35.47%。由此，验证了壁面台阶对喷管性能的影响是相互独立的，基本符合线性叠加规律。

4 结 论

考虑进口拐角波系对超燃冲压发动机尾喷管性能的影响，开展了不同拐角波及其强度对喷管气动参数影响的数值模拟分析，可得出以下结论：

1）超燃冲压发动机尾喷管的进口拐角波系引起喷管进口参数不均匀，喷管内产生的波系结构理论上可以使喷管气动性能发生明显变化，为喷管性能控制提供新思路。其中推力系数在上、下壁面均有后台阶时提升最多，在本文喷管构型的研究范围内其增幅最大可达13.98%，升力在下壁面前台阶时最多可提高94.32%，而俯仰力矩则在上壁面前台阶、下壁面后台阶作用下最多降低35.47%。

2）当本文研究的喷管构型进口的其中一侧壁面存在前台阶时，喷管另一侧壁面上将出现激波反射现象，随着台阶高度的增加，升力与俯仰力矩的变化渐趋一致；当喷管进口的其中一侧壁面存在后台阶时，由于该种情况下产生的激波波角较小，并不能与喷管另一侧壁面相交产生激波反射现象，所以此时随台阶高度增加，升力与俯仰力矩的趋势相反。

3）壁面台阶结构对喷管性能的影响基本是相互独立的，且符合线性叠加原理。在已知单侧台阶的影响规律的前提下，可直接通过单侧台阶对喷管的性能影响预估双侧台阶作用下喷管的性能变化规律。

参考文献（References）

[1] ORTON G F, SCUDERIT L F, SANGER P W, et al. Airbreathing hypersonic aircraft and transatmospheric vehicles[M]. Reston: American Institute of Aeronautics and Astronautics, Inc., 1997.

[2] FRY R S. A century of ramjet propulsion technology evolution[J]. Journal of Propulsion and Power, 2004, 20(1): 27-58.

[3] FERRI A. Review of scramjet propulsion technology[J]. Journal of Aircraft, 1968, 5(1): 3-10.

[4] FALEMPIN F H. Scramjet developments in France[M]. Reston: American Institute of Aeronautics and Astronautics, Inc., 2000.

[5] JOHNSTON P J , CUBBAGE J M , WEIDNER J P . Studies of engine-airframe integration on hypersonic aircraft[J]. Journal of Aircraft, 1970, 8(7): 495-501.

[6] MO J W, XU J L, GU R, et al. Design of an asymmetric scramjet nozzle with circular to rectangular shape transition[J]. Journal of Propulsion and Power, 2014, 30(3): 812-819.

[7] LANDSBERG W O, WHEATLEY V, SMART M K, et al. Performance of high Mach number scramjets - Tunnel vs flight[J]. Acta Astronautica, 2018, 146: 103-110.

[8] LU X, YUE L J, XIAO Y B, et al. Design of scramjet nozzle employing streamline tracing technique[C]// Proceedings of the 16th AIAA/DLR/DGLR International Space Planes and Hypersonic Systems and Technologies Conference. Reston: AIAA, 2009: AIAA2009-7248.

[9] HIRAIWA T, TOMIOKA S, UEDA S, et al. Performance variation of scramjet nozzle at various nozzle pressure ratios[J]. Journal of Propulsion and Power, 1995, 11(3): 403-408.

[10] 张晓源, 覃粒子, 刘宇. 非均匀入口与化学非平衡效应对尾喷管性能的影响研究[J]. 推进技术, 2014, 35(2): 166-171.
ZHANG X Y, QIN L Z, LIU Y. Non-uniform entrance and non-equilibrium chemical effects on scramjet nozzle performance[J]. Journal of Propulsion Technology, 2014, 35(2): 166-171 (in Chinese).

[11] 全志斌, 徐惊雷, 李斌, 等. 超燃冲压发动机尾喷管非均匀进口的冷流试验与数值模拟[J]. 航空学报, 2013, 34(10): 2308-2315.
QUAN Z B, XU J L, LI B, et al. Cold flow experiment and numerical simulation on nonuniform entrance flow of scramjet nozzle[J]. Acta Aeronautica et Astronautica Sinica, 2013, 34(10): 2308-2315 (in Chinese).

[12] 张璞, 徐惊雷, 全志斌. 激波/膨胀波造成的非均匀进口对单边膨胀喷管性能的影响及对比研究[C]//第七届全国高超声速科技学术会议. 北京: 中国力学 学会, 2014.
ZHANG P, XU J L, QUAN Z B. Effect of shock/expansion wave induced non-uniform inlet on performance of unilateral expansion nozzle and comparative study[C]//The 7th Hypersonic Science and Technology Conference. Beijing: Chinese Society of Mechanics, 2014(in Chinese).

[13] 陈以勒, 俞凯凯, 徐惊雷. 超燃冲压发动机尾喷管进口气流角畸变研究[J]. 推进技术, 2021, 42(12): 2694-2702.
CHEN Y L, YU K K, XU J L. Effects of inlet flow angle distortion on scramjet nozzle[J]. Journal of Propulsion Technology, 2021, 42(12): 2694-2702 (in Chinese).

[14] MO J W, XU J L, QUAN Z B, et al. Design and cold flow test of a scramjet nozzle with nonuniform inflow[J]. Acta Astronautica, 2015, 108: 92-105.

[15] GAFFNEY R L Jr.. Design of a Mach-15 total-enthalpy nozzle with non-uniform inflow using rotatioanl MOC[C]//Proceedings of the 43rd AIAA Aerospace Sciences Meeting and Exhibit. Reston: AIAA, 2005: AIAA2005-691.

[16] BARE E A, CAPONE F J. Static internal performance of convergent single-expansion-ramp nozzles with various combinations of internal geometric parameters[R]. Washington: National Aeronautics and Space Administration, 1989.

[17] ARAI T, SUGIYAMA H, HOMAREDA M, et al. Turbulence characteristics of supersonic boundary layer past a backward facing step[C]// Proceedings of the International Aerospace Planes and Hypersonics Technologies. Reston: AIAA, 1995.

[18] 陈以勒, 俞凯凯, 徐惊雷. 几何尺寸约束的超燃冲压发动机推力喷管设计[J]. 航空学报, 2021, 42(6): 124259.
CHEN Y L, YU K K, XU J L. New design method for scramjet nozzles with strong geometric constraints[J]. Acta Aeronautica et Astronautica Sinica, 2021, 42(6): 124259 (in Chinese).

Influence of inlet corner wave system on performance of scramjet nozzle

CHEN Chong，GE Jianhui*，XU Jinglei，SU Peng，YU Kaikai

(College of Energy and Power Engineering，Nanjing University of Aeronautics and Astronautics，Nanjing 210016，China)

Abstract: In order to explore the mechanism of the effect of inlet corner wave system on the wall pressure distribution of scramjet nozzle and investigate the influence of the inlet corner wave system on the performance of a scramjet nozzle, the numerical simulation and analysis of the influence of different corner waves and their intensity on the aerodynamic parameters of the nozzle were carried out. The effectiveness of the numerical method and the grid independence were verified. Different forward/backward steps were used to form the inlet corner wave system, and the influence of different wave systems and wave system intensities on the aerodynamic performance of the nozzle was studied. The results show that the inlet non-uniformity caused by the inlet corner wave system has a significant impact on the aerodynamic performance of the nozzle, in which the thrust coefficient can be increased by 13.98% at most when there is a backward step on the upper and lower walls, and the lift can be increased by 94.32% when there is a forward step on the lower wall. However, the pitching moment can be reduced by 35.47% at most when there is a forward step on the upper wall and a backward step on the lower wall. When there is a single-sided forward step at the nozzle inlet, the changing trend of lift and pitching moment is consistent with the increase in step height; when there is a single-sided backward step at the nozzle inlet, the conclusion is different. In addition, the influence of wall steps on nozzle performance is independent of each other and basically conforms to the linear superposition principle.

Keywords: scramjet; nozzle; expansion wave/shock wave; forward/backward steps; aerodynamic performance

Received：2022-09-14；**Accepted**：2022-11-01；**Published Online**：2022-11-15 14:18

URL：link.cnki.net/urlid/11.2625.V.20221114.1428.001

Foundation item：National Natural Science Foundation of China (11802123)

* **Corresponding author**. E-mail：jianhuig@nuaa.edu.cn

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2022.0784

链上、链下协同的智能合约即服务问题

刘浩宇[1]，马泽祁[1]，叶亚飞[2,*]，胡凯[1,2]

(1. 北京航空航天大学 软件开发国家重点实验室，北京 100191；2. 北京航空航天大学 云南研究院 昆明 650000)

摘 要：智能合约是区块链上的一种软件服务，具有价值传导、契约性和可信性等特点，当前各区块链系统上的智能合约数量越来越多，应用范围越来越广泛，传统智能合约的局限性也日益增大。隔离沙箱环境执行使智能合约与外界交互困难、资源受限、从而限制了智能合约的使用场景和性能。同时，传统的云服务等链下服务存在中心化风险，缺乏信任机制保障。提出智能合约即服务方案，提供链上合约与链下服务协同的方法，通过链下服务的链上认证，将服务生命周期纳入区块链管理，使之具有合约服务的可信化特点，以支持服务的注册、定制和组合；提出基于交易触发与事件驱动的通信与协同方法，在面对复杂业务场景或高资源需求任务时智能合约可与扩展服务组合构建工作流。经实验验证表明，智能合约即服务可以极大地拓展合约应用场景，通过将合约内执行的高资源耗费任务卸载到计算型扩展服务可提供更高的性能，性能优势与链下加速比正相关，如在执行1 200维矩阵乘法运算的任务时，性能可提升64%。

关 键 词：智能合约；智能合约即服务；链上、链下协同；服务认证；事件驱动

中图分类号：O213.2; V37; TP302.1

文献标志码：A **文章编号**：1001-5965(2024)10-3262-12

合约是一种具有法律效力的协议，它创建、定义和管理各方之间的相互权利和义务，规定了各方必须遵守的规则[1]。智能合约[2]是“智能化”的合约，其以数字代码的形式定义并可自动执行。区块链是以分布式存储和共识技术为基础，通过链式结构和隐私算法实现高可信、可追溯、不可篡改特性的技术。区块链的技术特性高度契合智能合约要求的执行环境，为其提供可靠的执行平台。区块链中智能合约模块被设计为一套独立的虚拟机制，以执行图灵完备的合约脚本。智能合约赋能后的区块链技术进入2.0时代，出现了以以太坊为代表的一系列支持智能合约执行的区块链系统，在金融场景的资产交易结算[3]、数据存证防伪[4]、供应链管理[5]等领域发挥着越来越重要的作用。各大区块链系统上已由众多参与方部署适用于各自领域，实现不同功能的海量智能合约。

然而智能合约应用依然面临着很多问题，场景上有很大的局限性，存在很多亟待解决的问题。为保障安全可信性，智能合约须运行在与外界隔绝的沙箱环境中，无法与外界交互。随着云计算、人工智能等新兴信息技术的发展，链下环境存在着丰富的数据、服务与计算资源。无法利用这些资源意味着智能合约技术无法协同主流的信息服务，限制其发挥技术的优越性。

同时上述常用的链下服务具有中心化风险，缺乏信任保障机制，缺少可追溯、高可信等特性，难以适用于强调安全性的场景。

为降低智能合约的使用门槛，业界提出了区块链即服务的思路，为特定场景提供区块链解决方案。然而不同场景智能合约之间相互独立且大部

收稿日期：2022-09-15；**录用日期**：2022-11-25；**网络出版时间**：2023-01-17 17:03

网络出版地址：link.cnki.net/urlid/11.2625.V.20230117.0909.001

基金项目：云南省重大科技专项 (202103AN080001-001,202002AA100007,202002AD080003)；软件开发国家重点实验室开放基金 (SKLSDE-2022ZX-11)

* **通信作者**. E-mail：yeyf@bhynii.com

引用格式：*刘浩宇，马泽祁，叶亚飞，等. 链上、链下协同的智能合约即服务问题 [J]. 北京航空航天大学学报，2024，50（10）：3262-3273. LIU H Y，MA Z Q，YE Y F，et al. Smart contracts-as-a-service with on-chain and off-chain collaboration [J]. Journal of Beijing University of Aeronautics and Astronautics，2024，50（10）：3262-3273 （in Chinese）.*

分逻辑复杂。在面对一个复杂业务场景时, 从零开发一个复杂合约的成本较高。

本文针对智能合约功能和性能受限、落地困难、传统服务存在中心化风险等问题, 提出智能合约即服务(smart contract as a service, SCaaS)的架构, 将链下环境中存在的丰富的数据与服务资源与链上的智能合约结合, 提出链下服务的合约化机制, 包括链下服务的注册、认证和管理, 将链下服务合约化, 变为扩展服务。采用事件驱动的通信模式, 设计链上智能合约与链下扩展服务的协同机制, 并通过非阻塞通信模型实现服务间的即时响应。以协同方案作为基础, 基于微服务的思路, 链上智能合约与链下扩展服务可组合成更灵活的工作流处理复杂的业务逻辑。

1　相关理论和技术

20 世纪 90 年代提出了智能合约, 被定义为计算机化的交易协议, 满足特定条件时自动触发[6]。智能合约推动区块链技术实现了阶跃式进步, 由最初的点对点金融系统, 扩展为可执行图灵完备脚本的分布式可信计算平台[7]。智能合约由触发条件、执行逻辑以及方法粒度的访问权限等组成。区块链上的智能合约以账户的形式存在, 各节点独立执行, 并共识确保正确性; 合约的部署与执行由交易触发, 交易执行结束后保存在区块内, 块间链式结构保证了交易的可追溯性和不可篡改性。

服务注册的现行标准是 UDDI[8](universal description, discovery, and integration), 用于提供服务的统一描述、发现和集成, 基于此标准出现了 Eureka[9]、Nacos[10] 等工具。数字证书是网络交互不可缺少的认证凭证, 已广泛应用于网络基础设施中, 提供用户合法身份的认证[11], 利用非对称加密、hash 签名等密码学方法认证网络实体。然而传统服务中的 UDDI 和 CA 认证均存在中心化风险, 无法在要求安全可信去中心化的区块链场景下使用。

当下区块链智能合约在 IoT[12]、分布式安全[13]、金融等场景[14] 获得广泛应用。然而, 相关技术未取得大范围的应用。学界和工业界提出了很多种方法来解决相关问题。

即服务是云计算的概念, 云计算提供方提供基础架构、应用平台、服务应用等资源, 解决传统软件存在的繁琐交付与部署的问题[15]。区块链即服务(blockchain as a service, BaaS)将区块链框架嵌入云计算平台, 利用云服务基础设施的部署和管理优势, 为开发者提供便捷高效的区块链生态系统和相关服务。通过这些基本的云服务, 开发者可以快速启动区块链网络来支持其应用程序, 而忽略底层架构的复杂性[16], 区块链即服务平台 BaaS 的总体架构图 1 所示[17]。目前 IBM、蚂蚁集团等企业已推出了功能完备的 BaaS 平台, 并提供成熟的即服务解决方案。

当下 BaaS 平台通过配套复杂冗长的智能合约解决特定场景的问题, 灵活性和扩展性较差。绝大部分智能合约可以抽象为一系列细粒度的合约组合, 比如资产转移、账户管理等。脱胎于 SoA 的微服务架构针对服务的扩展、灵活与复用性提出了成熟的解决方案, 是软件设计目前最常用的方式。智能合约微服务的思路也由此诞生。张富利等[18] 提出了一种智能合约微服务框架, 实现了智能合约在区块链网络上的“即插即用”; Tonelli 等[19] 提供了一个电子商务应用范例, 该应用运行于私有链提供的智能合约服务架构中, 证明了遵循微服务范式开

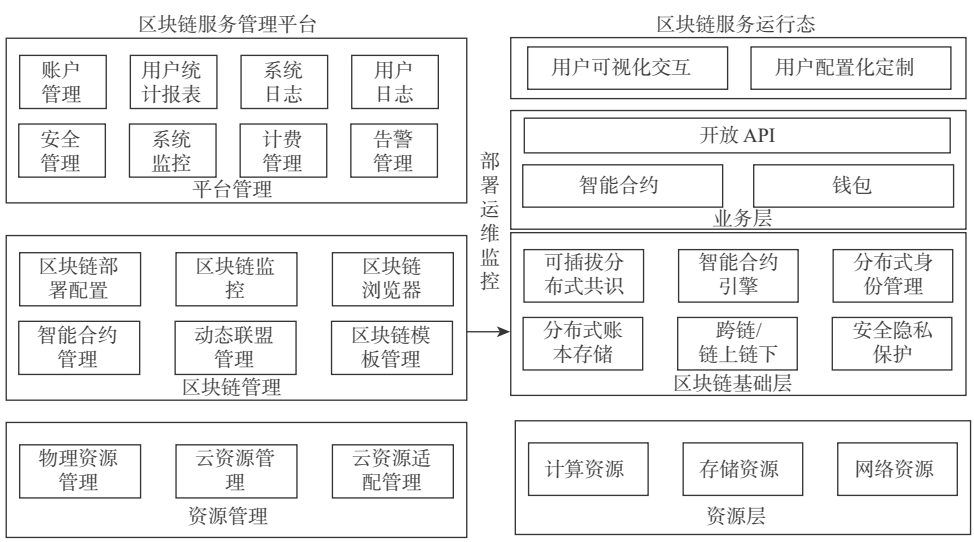


图 1　区块链即服务平台 BaaS 的总体架构

Fig. 1　Architecture of Platform BaaS

发智能合约服务系统的可行性。

智能合约微服务化的思路极大地增强了智能合约的灵活性与可扩展性。然而当前智能合约微服务化未考虑与链下丰富的服务协同，使用场景上依然存在局限性。

2 智能合约即服务架构

针对智能合约扩展性有限，无法有效利用链下资源的问题，提出智能合约即服务方案。作为智能合约服务化的手段，其主要目的是打通链上合约和链下传统服务。本节给出了智能合约即服务相关的定义架构、服务合约化机制以及链上链下协同机制。

定义1 智能合约即服务

智能合约即服务架构是指基于面向服务架构（service-oriented architecture, SOA），与BaaS平台配合，利用云服务、Web服务等基础设施和机制已有的资源和优势，为开发者提供便捷、可组合、可验证、可认证的各种粒度的智能合约服务和生态配套服务，使用户能方便地定制组合、构建和部署链上智能合约或链上认证的可信扩展服务，并且可以通过可信通道使链上、链下数据资源互通，使服务结果和数据上链存证，从而形成一种新的可信服务架构。

智能合约即服务架构如图2所示。

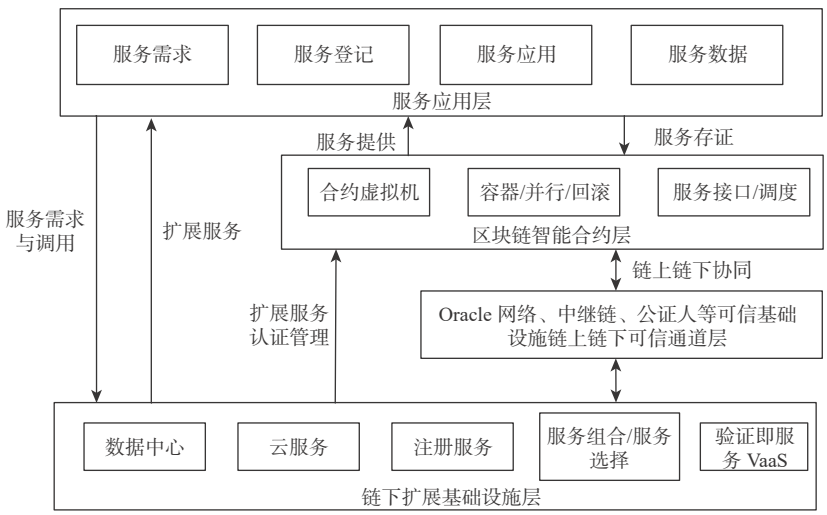


图2 SCaaS架构

Fig. 2 SCaaS Architecture Diagram

SCaaS方法的架构可以被分为4层。

1）链下扩展基础设施层：链下数据与服务资源的抽象层。在链上认证后利用智能合约对其生命周期进行管理。在面对复杂的业务场景时，提供服务选择与组合方法，并通过调用验证即服务VaaS[20]进行组合方法的合法性验证。

2）链上、链下可信通道层：基于预言机Oracle、中继链、公证人等机制实现。在保证合法性的基础上完成链上、链下的数据交互与协同操作，连通链下扩展基础设施层和区块链智能合约层。

3）区块链智能合约层：智能合约执行环境，即包括交易池、共识、通信、存储、合约虚拟机等在内的通用区块链系统。其支持容器化的合约执行，可进行交易执行的并行与回滚[21]。

4）服务应用层：与用户直接交互，接收服务需求，进行服务登记。区块链智能合约层与链下扩展基础设施层协同为服务应用层提供服务，并完成执行结果的链上存证。

链下扩展基础设施层包含了链上认证过的数据与服务资源，通过链上、链下可信通道层与区块链智能合约层协同。服务应用层作为用户界面接收用户需求并触发服务组合。

SCaaS根据服务执行环境分为2种服务形式。

定义2 扩展服务

链下普通服务经过注册认证后由智能合约进行SCaaS生命周期的管理，变为SCaaS可感知的扩展服务，与链上智能合约协同，包括平台自建服务和来自云平台的第三方服务等链下服务。具体描述、注册、认证及管理机制见第3节。

定义3 智能合约服务

合约C可被描述为一个三元组：合约方信息I（I_i为第i个合约方P_i的信息，i=1, ⋯, m, 共有m个人参与合约）、合约状态机M^*与合约方的执行状态机集合$\{M_1, M_2, \cdots, M_m\}$。

$$C=(I, M^*, \{M_1, M_2, \cdots, M_M\})$$

式中：合约状态机M^*可以表示为一个五元组$(Q, \Sigma, \delta^*, s^*, F^*)$，$Q=\{(q_1^*, q_2^*, \cdots, q_m^*, L), q_i^* \in q_{L_i}\}$。$Q$为

合约状态机所有状态的集合, L 为合约执行背景, q_i^*被包含于第 i 个合约方的状态集q_{l_i}中。Σ为输入事件的集合(一般是交易)。δ^*为转换函数的集合$(\delta^*: Q\times\Sigma\to Q)$。$s^*$为初始状态值$(s^*\in Q)$。$F^*$为终止状态的集合$(F^*\subseteq Q)$。每个区块链节点可以被理解为全量状态机的一个副本, 通过交易使得区块链保持一致性并进行流转。

合约的执行状态 M_i 表示第 i 个合约方 P_i 的执行状态机, 也可以表示为一个五元组$(q_{l_i},\Sigma,\delta_i,s_i,F_i)$, q_i 为合约方 P_i 的执行状态集合。Σ为输入事件的集合。δ_i为转换函数的集合$(\delta_i: q_{l_i}\times\Sigma\to q_{l_i})$。$s_i$为初始状态值$(s_i\in q_{l_i})$, F_i为终止状态的集合$(F_i\subseteq q_{l_i})$。

SCaaS 中的智能合约服务以标准合约定义为基础, 重点描述合约对外状态, 可以表示为三元组(CA,a,cm), 其中 CA 表征智能合约的链上属性, a 指合约地址, 通过 a 访问链上合约, cm 代表合约服务提供的参数列表, 通过 a,cm 完成对合约服务的请求。

2 种服务是根据业务逻辑原子化微粒度划分的。

3　扩展服务合约化方法设计

扩展服务合约化旨在将链下服务利用智能合约在链上进行认证, 基于合约对其生命周期进行管理, 基于区块链日志对其调用信息进行记录。认证后的服务被定义为扩展服务, 具有智能合约安全可信、不可篡改、可追溯等特性。

3.1　基于去中心化数字身份的认证机制

链下服务的认证机制包括数字身份设计与基于智能合约的数字证书生成。

SCaaS 管理多种类别、多个来源的扩展服务, 其需要为这些服务建立适当的数字身份, 统一整合及后续服务的调用。数字身份是指一组用于定义实体的特征数据, 包括三个维度: 身份标识、身份属性和身份交互[22]。身份标识是系统中能唯一指向对应实体的序列号或字符串; 身份属性描述与身份标识关联实体的相关信息; 身份交互是一套用于验证该身份信息的协议, 其定义了实体间交互时确认身份的方式。

SCaaS 中服务提供商提供链下服务, 而服务的有效性则由服务提供商和 SCaaS 中的管理合约共同保证。传统数字身份中的签发者和证明者由相同的组织或机构充当, 在依赖安全可信的“trustless”场景下, 这种方式会带来中心化风险。本文基于传统的数字身份, 将 SCaaS 中的数字身份划分为两部分应用于链下扩展服务: 身份标识和身份属性由云服务厂商提供; 身份交互由区块链系统通过另外发放的证书保证。服务提供商在服务注册时提交相关信息, 由平台整理后生成格式统一的数字身份, 数字身份模型如图 3 所示。

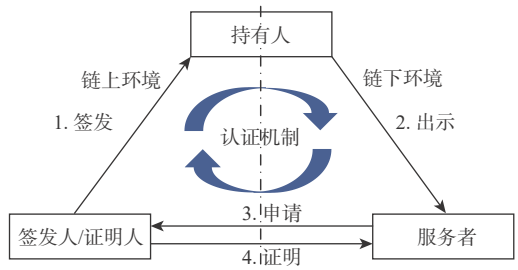


图 3　SCaaS 数字身份模型

Fig. 3　Digital identity model of SCaaS

扩展服务描述字段如表 1 所示。表 1 中, 统一资源定位符 URL 唯一描述服务, 用于服务发现与服务路由。服务状态哈希 codeBytes 是由用户提供字节数组形式的服务状态, 通过算法 1 生成, 可以用来表示当前状态下的服务快照、验证服务是否被篡改, 以保证服务可信。

表 1　链下扩展服务的数字身份字段

Table 1　Digital identity fields for off-chain scaling services

字段名	作用
服务名称name	
统一资源定位符URL	唯一指定服务
方法参数params	形参列表
服务提供商provider	
服务注册时间regi_time	
描述元信息meta_info	其他服务信息
codeBytes	服务状态快照

算法 1 第 4 步即 codeBytes 的位运算表示形式, 选取乘法基准实数为 31, 因为其等于 2^5-1(2 的幂次-1), 而且作为素数能够使生成的哈希值分布较为均匀。计算机中的位运算性能远优于加乘运算, 所以实际运算过程中一般采用位运算的形式。

算法 1　服务状态快照均匀哈希算法。

输入: 服务状态的字节数组 $[a_0,a_1,a_2,\cdots,a_{n-1},a_n]$。

输出: 服务状态快照哈希值 codeBytes。

```
1  codeBytes := a_0
2  for each a_i in [a_1,a_2,⋯,a_{n-1},a_n]:
3      do
4      codeBytes:= (codeBytes<<5) - codeBytes
5      codeBytes = codeBytes + a_i
6      done
```

获取如表 1 所示的扩展服务的相关信息之后, SCaaS 设计扩展服务认证合约完成扩展服务数字证书的生成, 主要过程如算法 2 所示。

算法 2 链下扩展服务的数字证书生成算法。

输入: 扩展服务数字身份 identity。

输出: 扩展服务数字证书 certificate。

```
certificate:=[]
certificate.add(version, serialNum, issuer)
Set validFrom & validTo according to regi_time
certificate.add(validFrom, validTo)
encry _fields=[codeBytes]
plain_fields:= identity expect codeBytes
certificate.add(encry _fields)
certificate.add(plain_fields)
Digist_certificate=digist(certificate)
Encry_certificate=encry(Digist_certificate,certificate)
```

初始化 Certificate,并依次添加:

1)version、serialNum、issuer 分别为认证合约预设的证书版本、当前扩展服务序列号、证书签发者。

2)根据服务注册时间生成扩展服务合法时间 validFrom、validTo。

3)服务状态快照字段 codeBytes。

4)链下扩展服务明文字段。

所有数据添加完毕后,使用摘要算法获取证书摘要 digist_certificate。最后使用非对称加密算法中的私钥对以上字段进行加密。

证书生成完毕后,将公钥以及签名作为证书字段下发到链下,同时在认证合约中保存一个证书副本。链下扩展服务获得该数字证书后,可在其他场景中使用,服务调用者可通过公钥和签名运行解密算法,并比较得到的明文与证书各字段,若一致则说明证书有效,从而确保服务可靠;否则,应当谨慎或拒绝调用对应服务。该合法性检查属于懒加载的过程,即调用前检查。

数字证书生成并下发后,扩展服务认证过程完成。

3.2 基于智能合约的管理机制

链下服务的管理机制主要指包括注册在内的生命周期管理与基于区块链日志的调用记录等。

服务注册是实现管理机制最关键的步骤,进行服务实例注册的同时,须对外提供获取服务实例的列表。为了避免扩展服务注册时的中心化风险,SCaaS 中的服务注册等生命周期管理交由扩展服务管理合约进行,其包含了若干合约方法,用于完成生命周期管理。SCaaS 中的管理合约基本架构如图 4 所示。

管理合约调用服务注册方法,认证扩展服务的合法性,生成数字证书后,创建并维护扩展服务实例。后续 SCaaS 的不同调用方进行服务调用时会自动触发日志上传模块。

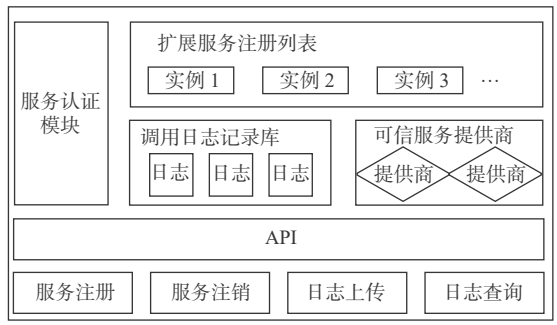


图 4 扩展服务管理合约基本架构

Fig. 4 Basic structure of extended service management contract

服务注销是注册的逆过程,管理合约通过 URL 检索并移除相关服务,删除认证证书副本。该过程也属于懒加载,在调用前认证时完成。

扩展服务的生命周期管理旨在利用区块链,使云服务具备智能合约特性。智能合约部署在区块链上,其静态的程序代码不允许人为修改;扩展服务的注册也是为达到类似目的而设计的机制,通过区块链颁发的认证证书及验证机制保证服务不可被非法篡改。云服务在该机制下具备更强的安全性,可预防篡改型恶意攻击,并发挥区块链的优势。

管理机制的另一重要部分是服务调用结果的存证与查询。传统链下服务的调用结果一般以日志形式存储到本地实例,具有极大的中心化风险,发生纠纷时难以保证日志的可靠可信,也会给其存储空间带来极大压力。

SCaaS 中的调用存证基于管理合约中的存证接口实现,设计了日志上传与查询的合约方法,调用记录将以区块链日志的形式存储在区块中。其基本机制如图 5 所示,图 5 中 k 为区块序号。

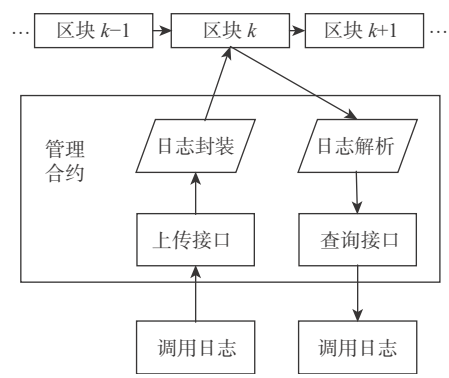


图 5 日志存证及查询机制

Fig. 5 Log storage and query mechanism

日志本身包括全局唯一的调用序列号、调用者、服务 URL、参数列表、调用时间、调用结果/失败(异常)原因等信息。扩展服务或智能合约被调

用时会自动触发日志上传动作。管理合约将接收到的调用相关信息封装为统一格式化日志后存入数据库中, 通过共识机制确保各节点对本次日志的内容达成一致, 从而避免了集中式存储调用日志时被篡改的可能性。

同时除了满足调用安全性保证与特定业务场景的查询需要外, SCaaS 中的日志查询接口还可被扩展服务或智能合约调用, 作为 RPC 使用。后序服务可通过调用日志查询接口获取前序服务的执行状态与执行结果, 作为本身的激活与输入。

扩展服务去中心化的身份认证机制与基于智能合约的管理机制使得链下普通服务具有了合约特性, 由区块链保证扩展服务的安全可信可追溯, 形成了链下服务与区块链系统交互的闭环, 真正意义上将云服务等链下服务融入智能合约服务平台中, 在保证服务可靠性、可追溯性的同时极大地拓展了平台提供的服务类型, 丰富了智能合约服务的应用场景。

4　链上链下协同方法设计

SCaaS 架构旨在提高服务的可扩展性, 将智能合约和扩展服务按照业务逻辑的最小单元进行划分。在面对复杂业务场景时, 各原子服务协同以工作流的方式共同完成任务。探究 SCaaS 架构下链上、链下协同的方法, 主要关注通信协同机制, 可信性由预言机等机制保证, 不在讨论范围内。

完成复杂任务的工作流由若干智能合约与扩展服务构成, 后序服务将会在前序服务执行结束并触发启动条件时开始运行, 因此需要一套通信机制来完成服务之间的启动状态触发的工作。传统微服务架构下服务间的调用方式有 2 种: 远程过程调用(remote procedure call, RPC)和事件驱动。SCaaS 架构下的服务工作流中, 假设相邻两个服务前序服务 A、后序服务 B, 对于 A、B 本身可以是智能合约或者扩展服务, 所以对于该场景下的通信也就存在 4 种形式, 如图 6 所示。

其中调用形式③通过区块链智能合约虚拟机

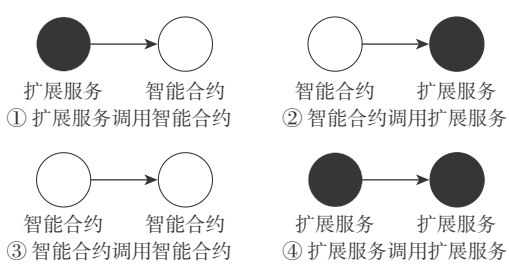


图 6　SCaaS 的 4 种通信场景

Fig. 6　Four communication scenarios of scaas

支持, 调用形式④为微服务调用的 RPC 方式, 可使用 RPC 或者事件驱动的方式。以下将着重说明调用形式①和调用形式②。

4.1　扩展服务与智能合约的协同方式

调用形式①有 2 种场景: 扩展服务链上认证和前序扩展服务调用后序智能合约。该过程可被拆解为为链下部分和链上部分, 如算法 3 和算法 4 所示, 算法 3 阐述链下服务调用管理合约完成注册的流程, 算法 4 阐述链下扩展服务调用链上智能合约的流程。

算法 3　扩展服务与智能合约的协同——链下。

输入:

认证: 服务元信息 metaInfo;

调用: 合约地址 *addr*, 方法调用 *callMethod*。

输出: 无, 状态为交易被缓存到交易池。

```
1  ori_transcation:={}
2  if authentication:
3      metaInfo_ compiled: = compile(metaInfo)
4      auth_transcation:=get_trans(metaInfo_ com-
piled)
5      ori_transcation= auth_transcation
6  eilf invoke:
7      invoke_tanscation:=get_trans(addr,call-
Method)
8      ori_transcation= invoke_tanscation
9      parsed_transcation:=rlp(ori_transcation)
10     encry_transcation:=encry(parsed_transca-
tion)
11   send encry_transcation to TX pool
```

如算法 3, 认证时使用注册交易, 将服务元信息编译成字节数组, 存放在注册交易的域中, 经过加密编码后发送到交易池; 调用时使用调用交易, 确定调用方合约地址、方法签名以及实参, 创建调用交易经过 RLP 编码和加密后发送到交易池(TX pool)。

算法 4　扩展服务与智能合约的协同——链上。

输入: 无。区块链定时从交易池捞取交易。

输出: 无。交易执行后更新链状态。

```
1  pre-action{select leader node, tx package,
consen-sus}
2  if authentication:
3      meta_info:=get_metaInfo(transcation)
4      identity:=create_identity(meta_info)
5      certificate(identity) // algo.2
6  eilf invoke:
7      contract_addr:=get_addr(transcation)
```

```
8    callMethod:=get_callMethod(transcation)
9    contract:=get_contract(contract_addr)
10   invoke(contract, callMethod)
11   store_log
12  create_repicet(transcation)
13  post-action{store block, update MPT}
```

链上部分如算法4所示，一轮做块经过选主、交易打包、共识之后进入交易执行阶段。认证时，区块链可从注册交易获取服务信息，按照算法2进行服务认证。调用时，首先从交易获取合约地址、方法签名和实参，获取合约实例后进行调用，并记录调用日志。交易执行后会生成交易回执，并进行全局状态更新与落块等操作。

4.2 智能合约与扩展服务的协同方式

SCaaS方法中，调用形式②基于非阻塞的事件驱动模型设计，由负责调用扩展服务的扩展服务调用合约（下称调用合约）负责完成各智能合约对扩展服务的调用，并记录调用日志。整体架构如图7所示。

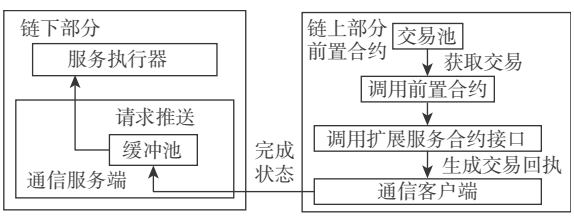


图7 链上调用链下通信架构

Fig. 7 On chain call off chain communication architecture

在区块链中对智能合约任何形式的调用只能通过交易的方式，前置合约pre-contract在执行过程中需要调用扩展服务，当交易执行器执行至扩展服务调用指令时，便会借助调用合法发起协同请求。调用合约与SCaaS链下部分在该场景的通信下被作为一个事件驱动模型的Client端和Server端，调用请求被发送至Buffer缓存等待处理，交由Selector分发，将请求交由注册在Selector上的对应扩展服务Handler处理。

为支持扩展服务的可扩展性，支持多种方法签名的服务调用，SCaaS中为调用形式②自定义了请求体编码结构，如表2所示，用于避免在buffer中可能出现的半包/粘包问题，以及在对扩展服务形参列表不做约束的情况下合约层对参数列表的泛化。

其中，消息头字段目前包含以下两个编号用于指明消息体内容：

#01#: 请求方智能合约地址；

#02#: 后序扩展服务方法签名与实参列表。

同时为了支持参数泛化，实参列表在调用合约中被编码为[datatype, value]的键值对形式，SCaaS链下通信Server侧完成请求数据的解码，以完成跨语言的扩展服务调用。

表2 自定义消息格式表

Table 2 Message format table

字段名	长度/字节	描述
Header	4	对消息体类型的标识
Body	Not-fixed	承载交易执行状态

4.3 安全性分析

安全性是SCaaS的根本要求。具体表现为，在扩展服务与链上合约协同过程不能引入外部风险。对于给定的场景，分别从组成工作流的链上合约与扩展服务上说明。

链上合约依然在隔离的沙箱环境执行，通过交易触发状态改变，让合约与扩展服务协同。通过密码学原理，共识机制等保障执行的一致可信，这与传统智能合约方案在安全性上没有区别。

对于扩展服务，智能合约颁发数字证书完成服务认证，调用前必须验证证书与签名，保证服务未被篡改。通过合约完成扩展服务的管理存证，调用日志被全量保存，保证可追溯性。

综上所述，在SCaaS对外提供服务的过程中，可以认为链上合约与扩展服务组成的协同工作流安全可信。

5 实验验证

以分布式AI场景为例，设计实验验证SCaaS在功能和性能上的的优势。实验一协同多个AI服务，通过认证及管理机制赋予其可信、可追溯等合约特性，证明SCaaS在功能上的优势，同时协同多个合约与扩展服务，验证其扩展性。实验二通过矩阵计算模拟模型训练的任务，突出说明SCaaS在分布式AI训练任务上的性能优势。将SCaaS方案与现有的区块链领域即服务相关工作进行对比，证明设计的先进性和创新性。

5.1 实验准备

5.1.1 平台环境

实验所使用的集群配置如下：硬件设备包括区块链集群、个人PC和交换机。4台戴尔R730服务器组成了最小的拜占庭集群运行区块链服务，每台服务器的CPU型号为Intel(R) Xeon(R) E5-2620, 主频2.40 GHz, 24cores, X86_64架构，内存64 GB。个人PC型号为Lenovo笔记本，CPU型号为Intel(R) Core(TM) i7-9750H主频2.60 GHz。内存为16 GB。NETGEAR G5308八口千兆交换机作为区块链节点相互、PC与区块链节点之间的网络通信设备。

区块链集群采用CentOS 7.0的操作系统，软件

依赖包括 MySQL 5.1.73, Redis 3.2.1, OpenJDK1.8.0。PC 采用 OpenJDK1.8.0 与 IntelliJ IDEA 2022.1。

5.1.2　方法介绍

机器学习包括模型的训练与使用, 其中涉及大量的矩阵运算。目前的机器学习主要依赖 GPU 设备来支持模型的训练与使用, 作为一种集中的训练结构, 相关数据集与资源均被中心化持有。集中式数据管理与存储极大地增加了数据被篡改的可能, 另外, 数据来源和生成数据的来源真实性无法得到保证。集中式架构下训练出来的模型缺少可信性。

去中心化 AI 的概念应运而生。作为人工智能与区块链的结合, 去中心化 AI 能够以分布式和去中心化的方式处理和执行可信、数字签名和安全的共享数据的分析或决策, 数据集在区块链上进行交易和存储, 无需可信第三方或中介, 基于智能合约的自治系统和机器可以随着时间的推移学习并适应变化, 从而做出可信和准确的决策结果, 这些结果经过区块链的所有挖矿节点的验证。这样的决定是无法反驳的, 并且可以被所有参与实体追踪、跟踪和验证[23]。

但链上资源毕竟有限, 使用普通集群运行的区块链节点无法完成模型训练的任务, 而使用高性能 GPU 集群构建的区块链集群的成本又过于高昂。目前去中心化 AI 常见的方式是联邦学习的方法, 本地完成训练后将阶段性的参数通过智能合约更新, 通过激励约束用户的作恶行为, 然而这依然存在节点作恶的可能。

5.2　实验结果

5.2.1　AI 协同任务

为验证 SCaaS 下应用场景的扩展性和通用性, 设计由多个扩展服务和智能合约组成的协同工作流, 如图 8 所示。

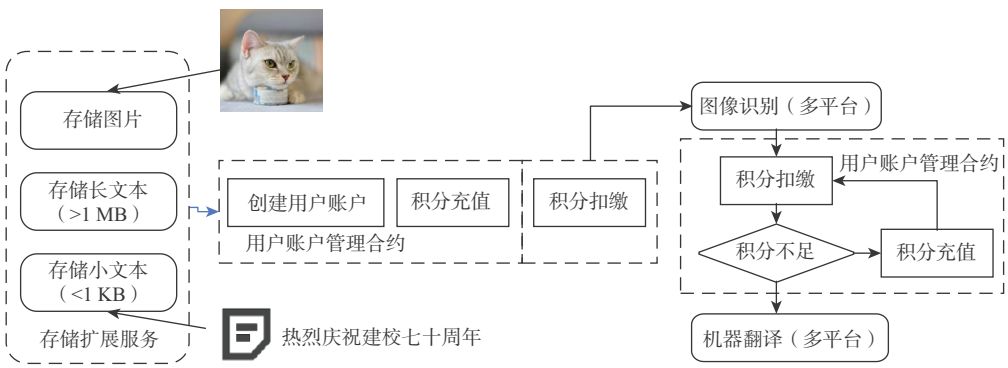


图 8　SCaaS 中的智能合约与人工智能扩展服务协同任务场景示例

Fig. 8　Examples of collaborative task scenarios between smart contracts and AI-extended services in SCaaS

存储扩展服务是基于 IPFS, 经过 SCaaS 认证的扩展服务, 其将图片、长文本、小文本等加密存储到 IPFS 中, 并将 cid 上传到智能合约管理。SCaaS 采取租户模式, 用户通过花费账户积分获取智能合约与扩展服务的调用权限。图像识别、机器翻译算法来自多个第三方云平台, 经过认证后成为扩展服务。

本次实验设计 SCaaS 协同流, 依次完成对图 8 中猫的图像识别任务和对“热烈庆祝建校七十周年”的文本机器翻译任务, 并完成存储和积分计算。扩展服务来源于华为云、腾讯云和阿里云。

单独调用图像识别和机器翻译扩展服务的结果如表 3 所示。

腾讯云的机器翻译扩展服务由 SCaaS 执行后的管理合约所存储的对应执行记录如图 9 所示。

通过构建的 SCaaS 协同工作流, 智能合约和扩展服务共同提供可信的分布式 AI 服务, 说明了 SCaaS 在功能场景上的优越性。

表 3　扩展服务执行结果

Table 3　Extended service implementation results

应用场景	图像识别	机器翻译
阿里云	猫(置信度: 70%)	warmly celebrate the 70th anniversary of the founding of the school
华为云	这是一只猫	celebrating the 70th anniversary of the School
腾讯云	猫为56%, 美国短毛猫为25%, 英国短毛猫为15%	warmly celebrate the 70th anniversary of the founding of the school

```
{"invokeId":1668590264121,"caller":"jack",
"serviceName":"TsTranslator","url":"http://127.0.0
.1:8082/api/SCaaS/CloudService/TsNlp/translateText",
"parameters":"热烈庆祝建校七十周年",
"invokeTime":1668590264121,"state":2,
"resultJson":"{result:Warmly celebrate the 70th
anniversary of the founding of the school}"}
```

图 9　腾讯云机器翻译业务场景的 SCaaS 调用日志

Fig. 9　SCaaS call logs for Tencent Cloud Machine translation business scenarios

5.2.2 模型训练任务

目前机器学习应用最广泛、最高效的方法是深度学习。模型训练的任务本质是通过包括大规模矩阵运算的正向传播和基于最小化损失函数的反向传播得到每个神经元的最优网络参数。

每个神经元的网络权重一般以矩阵的形式随机初始化，矩阵形式的输入数据与神经元的网络参数相乘得到预测值，基于预测值和真实值之间的差异设计损失函数，通过最小化损失函数完成网络参数的收敛。

矩阵乘法是模型训练过程中的核心运算，传统CPU架构无法支持大规模的矩阵相乘，所以目前常用的方式是GPU集群训练。

在分布式AI中，区块链作为分布式系统难以满足模型训练要求的硬件性能，这是限制分布式AI中区块链成为模型训练环境的最主要因素。

SCaaS能够充分利用已完成认证的链下扩展服务。在分布式AI的场景下，SCaaS可以将原先无法在链上完成的大规模矩阵计算任务卸载到链下的矩阵运算扩展服务中去，完成区块链上的机器学习模型训练。矩阵计算耗时与矩阵维度成正相关。

本实验注册华为云提供的高性能矩阵计算服务作为SCaaS中的扩展服务。通过对比单独链上计算矩阵乘法任务和卸载到扩展服务2种任务执行方式，验证SCaaS方法性能上的优势，考虑到网络波动性等影响因素，每个维数都进行了5次重复实验，得到的实验结果取均值，高维矩阵计算任务2种执行方式下执行时间比较如图10所示。在实验中，智能合约执行时间是指从发送矩阵运算交易到收到区块链系统计算结果的时间；SCaaS执行时间则包括SCaaS任务卸载、扩展服务中的矩阵计算以及扩展服务进行结果存证的时间。由图10可知，随着矩阵维度的上升，SCaaS执行矩阵计算任务的优势越来越明显，这充分说明了在处理高资源耗费问题时相比区块链独立执行的优越性。

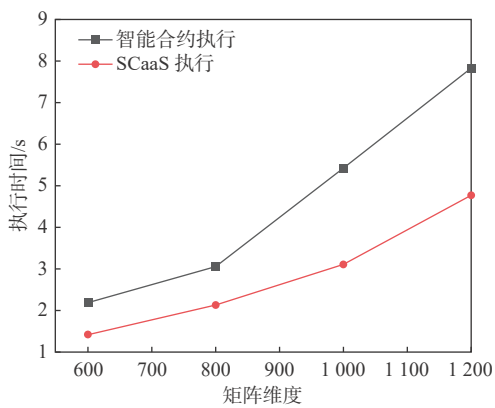


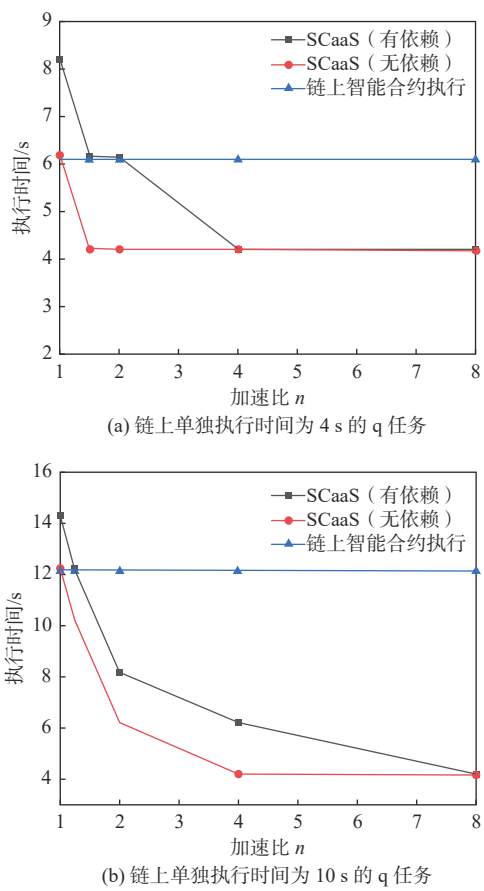
图10 高维矩阵计算任务2种执行方式下执行时间比较

Fig. 10 Comparison of execution time of high-dimensional matrix computing under two execution modes

SCaaS将任务卸载到扩展服务执行，相较于区块链合约执行任务，额外增加了链上与链下通信的时间，该开销使得SCaaS执行方法更适用于服务规模足够大或链下执行比链上足够快的情况。

对于任务A，定义区块链单独执行的时间为$t_{\mathrm{block,A}}$，SCaaS中扩展服务单独执行任务的时间为$t_{\mathrm{SCaaS,A}}$，加速比$n=\dfrac{t_{\mathrm{block,A}}}{t_{\mathrm{ScaaS,A}}}$。

为了寻找利用SCaaS协同智能合约与扩展服务，比单独的链上执行更具优势的加速比阈值，本实验设计了简单的三段式合约服务C_{t}，C_{t}可被拆分为3个子合约服务p、q、r。其中p、r为简单任务，执行耗时可忽略不计，因其始终运行在链上。而q为定时休眠任务，可通过设定不同休眠时间来模拟不同耗时的服务，分别在链上和链下扩展服务中运行。实验分别测试了链上单独执行耗时为4 s、10 s的q任务，不拆分情况下在链上执行完整合约的耗时，以及不同加速比下将q认证为扩展服务并在链下执行(p、r任务仍然在链上执行)的耗时，实验结果如图11所示。同时实验还考虑到服务间的依赖



(a) 链上单独执行时间为4 s的q任务

(b) 链上单独执行时间为10 s的q任务

图11 4 s、10 s的q任务链上智能合约执行和在SCaaS中的执行时间

Fig. 11 Execution time of 4 s and 10 s task q in on-chain smart contract and within the SCaaS

情况, 即 p 的后序服务 q 的激活是否依赖于 p 的执行完毕, 对于无依赖的情况, 在服务平台发起交易请求执行任务 p 的同时即可在链下并行执行任务 q, 而有依赖的情况则必须等待任务 p 在链上执行完成且链下收到反馈后才可执行 q, 2 种情况显然应在时间花销上有所区别。

在任务 q 对任务 p 有依赖的情况下, 对该实验模型进行算式推演, 可得出以下不等式:

$$2\times t_{\mathrm{c}}+\frac{t_{\mathrm{block,q}}}{n}<t_{\mathrm{c}}+t_{\mathrm{block,q}} \tag{1}$$

对不等式(1)进行化简可得

$$n>\frac{t_{\mathrm{block,q}}}{t_{\mathrm{block,q}}-t_{\mathrm{c}}} \tag{2}$$

式中: $t_{\mathrm{block,q}}$为 q 服务的链上执行时间, $t_{\mathrm{block,q}}=n\times t_{\mathrm{SCaaS,q}}$。 $t_{\mathrm{SCaaS,q}}$为 q 服务通过 SCaaS 注册为扩展服务后在链下的执行时间; t_{c}为从创建合约调用交易到区块链系统执行合约方法的通信时间, 其仅与实验环境有关, 故可看作常量。在本文的实验环境下, t_{c}为 2.1 s 左右, 带入式(2)可求得对应 4 s 的q任务, 智能合约即服务方法更具优势的临界加速比约为 2.1; 对于 10 s 的q任务, 智能合约即服务方法更具优势的临界加速比约为 1.27, 符合实验测量的结果。

在任务 q 对任务 p 有依赖时, 对于 4 s 的任务 q, 链下执行加速比 n>2 时认证为扩展服务更有优势; 对于 10 s 的任务 B, 加速比 n>1.25 时认证为扩展服务更有优势, 符合实验结果。同理其他规模的任务也可根据式 (2), 求出在链下扩展服务执行更优的加速比阈值, 进而选择合理的任务执行方法。

同理, 如果任务 p 与任务 q 之间不存在依赖, 仅任务 q 需要在任务 p 开始后即可开始执行, 则对加速比 n 的要求为

$$\max\left(t_{\mathrm{c}},\frac{t_{\mathrm{block,q}}}{n}\right)+t_{\mathrm{c}}<t_{\mathrm{block,q}} \tag{3}$$

根据实验结果, 对于高资源耗费型任务, 若为无依赖的合约任务, 卸载到链下能获得性能提升, 而存在依赖的合约任务, 则可通过式 (2) 计算出加速比阈值后综合选择执行方法。

5.3　与相关工作的对比分析

为了降低区块链落地的难度并从功能和性能上提升扩展性, 学界和工业界提出了一些方案, 其中包括区块链即服务平台, 如蚂蚁链 BaaS、Oracle BaaS; 依托于公链系统的预言机项目, 如 ChainLink; 学界在智能合约即服务领域的其他工作, 如文献 [19] 基于 DevOps 提出的 Mictract 平台。具体相关工作的特性对比如表 4 所示。

表 4　与业界其他相关工作成果的对比

Table 4　Comparison of characteristics with related work

解决方案	蚂蚁链 BaaS	Oracle BaaS	ChainLink	Mictract	SCaaS
适配链	Antchain和FabricBCOS	Fabric	Ethereum	Fabric	北航链
即时通信	否	否	否	否	是
合约扩展性	否	否	是	是	是
链下协同	否	否	是	否	是
协同方式			预言机共识网络		扩展服务认证管理
响应速度			慢		快

6　结　论

本文针对区块链智能合约方法在应用场景上的局限性, 提出基于链上、链下协同的智能合约即服务方案, 包括扩展服务认证和管理方法, 链上、链下协同方法, 这拓展了智能合约的应用场景并可以在一定任务规模下提升其运行性能。本文的主要工作可概括如下:

1) 通过合约化机制, 链下服务可认证为扩展服务。根据功能性实验的结果, 扩展服务可与智能合约高效协同, 对外提供可信服务。

2) 链上、链下通信协同机制能高效可信地完成链上、链下的协同交互, 这是能工作流处理复杂业务的基础。

3) 通过对比高资源耗费型任务在链上执行和 SCaaS 中利用扩展服务执行的时间成本, 验证 SCaaS 可以通过链下资源对链上任务进行加速, 从而提升了合约执行的性能。

基于目前的工作, 后续可以继续就智能合约服务化的业务完备性进行更深入的探究。在实际场景下, 针对一个问题会存在多种不同的组成方法, 如何在引入智能合约的基础上挑选出最优的组合方法是关键。同时工作流应符合长活事务模型, 为此 SCaaS 中的智能合约与扩展服务应该基于长活事务模型进行扩充。另外, 还可以从服务组合建模领域知识出发, 抽象智能合约与扩展服务的衡量指标, 基于启发式算法进行服务选择。

参考文献（References）

[1] VACCA A, DI SORBO A, VISAGGIO C A, et al. A systematic literature review of blockchain and smart contract development: techniques, tools, and open challenges[J]. Journal of Systems and Software, 2021, 174: 110891.

[2] MOHANTA B K, PANDA S S, JENA D. An overview of smart contract and use cases in blockchain technology[C]// 2018 9th International Conference on Computing, Communication and Networking Technologies (ICCCNT). Piscataway: IEEE Press, 2018: 1-4.

[3] 袁勇, 王飞跃. 区块链技术发展现状与展望[J]. 自动化学报, 2016, 42(4): 481-494.
YUAN Y, WANG F Y. Blockchain: the state of the art and future trends[J]. Acta Automatica Sinica, 2016, 42(4): 481-494 (in Chinese).

[4] 王鹏, 魏必, 王聪. 区块链技术在政务数据共享中的应用[J]. 大数据, 2020, 6(4): 105-114.
WANG P, WEI B, WANG C. Application of blockchain technology in government data sharing[J]. Big Data Research, 2020, 6(4): 105-114 (in Chinese).

[5] 周立群, 李智华. 区块链在供应链金融的应用[J]. 信息系统工程, 2016(7): 49-51.
ZHOU L Q, LI Z H. Application of blockchain in supply chain finance[J]. China CIO News, 2016(7): 49-51 (in Chinese).

[6] ZOU W Q, LO D, KOCHHAR P S, et al. Smart contract development: challenges and opportunities[J]. IEEE Transactions on Software Engineering, 2021, 47(10): 2084-2106.

[7] BUTERIN V. A next generation smart contract & decentralized application platform[EB/OL]. (2022-07-29)[2022-09-05]. https://ethereum.org/zh/whitepaper/.

[8] GANESHBABU S N, SRIRAM S, RANGARAJU K G. Smart cyber physical systems: advances, challenges and opportunities[M]. Abingdon: Taylor & Francis, 2020: 255-268.

[9] Spring Cloud. Service registration and discovery[EB/OL]. (2022-05-10)[2022-09-05]. https://spring.io/guides/gs/service-registration-and-discovery.

[10] Nacos. An easy-to-use dynamic service discovery, configuration and service management platform for building cloud native applications[EB/OL].(2022-06-12)[2022-09-14].https://nacos.io/en/docs/v1/what-is-nacos/.

[11] HARN L, REN J. Generalized digital certificate for user authentication and key establishment for secure communications[J]. IEEE Transactions on Wireless Communications, 2011, 10(7): 2372-2379.

[12] PARK J S, YOUN T Y, KIM H B, et al. Smart contract-based review system for an IoT data marketplace[J]. Sensors, 2018, 18(10): 3577.

[13] ABDELMABOUD A , AHMED A I A , ABAKER M, et al. Blockchain for IoT applications: taxonomy, platforms, recent advances, challenges and future research directions[J]. Electronics, 2022, 11(4): 11040630.

[14] ZHENG Z B, XIE S A, DAI H N, et al. An overview on smart contracts: challenges, advances and platforms[J]. Future Generation Computer Systems, 2020, 105(C): 475-491.

[15] SUNYAEV A, SUNYAEV A. Internet computing: principles of distributed systems and emerging internet-based technologies [M]. Switzerland: Springer Cham, 2020: 195-236.

[16] ZHENG W L, ZHENG Z B, CHEN X P, et al. NutBaaS: a blockchain-as-a-service platform[J]. IEEE Access, 2019, 7: 134422-134433.

[17] 魏凯, 卿苏德, 杨白雪, 等. 区块链即服务平台 BaaS 白皮书[R]. 北京: 中国信息通信研究院, 2019.
WEI K, QING S D, YANG B X, et al. Blockchain as a service platform baas white paper [R]. Beijing: China Institute of Information and Communications, 2019(in Chinese).

[18] 张富利, 侯培宇, 李杉杉, 等. 一种智能合约微服务化框架[J]. 软件学报, 2021, 32(11): 3423-3439.
ZHANG F L, HOU P Y, LI S S, et al. Framework for architecting smart contracts using microservices[J]. Journal of Software, 2021, 32(11): 3423-3439 (in Chinese).

[19] TONELLI R, PINNA A, BARALLA G, et al. Ethereum smart contracts as blockchain-oriented microservices[C]// Proceedings of the 19th International Conference on Agile Software Development: Companion. New York: ACM, 2018: 1-2.

[20] HU K, WAN J, LUO K, et al. Verification algebra for multi-tenant applications in VaaS architecture[J]. Software Testing, Verification and Reliability, 2021, 31(6): e1763.

[21] 王子凯, 朱健, 张伯钧, 等. 区块链与智能合约并行方法研究与实现[J]. 计算机科学, 2022, 49(9): 318-323.
WANG Z K, ZHU J, ZHANG B J, et al. Research and implementation of blockchain and smart contract parallel methods[J]. Computer Science, 2022, 49(9): 318-323(in Chinese).

[22] MIR U B, KAR A K, DWIVEDI Y K, et al. Realizing digital identity in government: Prioritizing design and implementation objectives for Aadhaar in India[J]. Government Information Quarterly, 2020, 37(2): 101442.

[23] SALAH K, REHMAN M H U, NIZAMUDDIN N, et al. Blockchain for AI: review and open research challenges[J]. IEEE Access, 2019, 7: 10127-10149.

Smart contracts-as-a-service with on-chain and off-chain collaboration

LIU Haoyu[1], MA Zeqi[1], YE Yafei[2,*], HU Kai[1,2]

(1. State Key Laboratory of software development, Beihang University, Beijing 100191, China;
2. Yunnan Research Institute of Beihang University, Kunming 650000, China)

Abstract: A smart contract is a software service on blockchain with the characteristics of value transfer, contractuality, and trustworthiness. The quantity of smart contracts on blockchain systems is currently rising, together with the application scope's expansion and the restrictions of conventional smart contracts' growing number. Isolated sandbox environment execution makes it difficult for smart contracts to interact with the outside world, and resources are limited, limiting the use scenarios and performance of smart contracts. At the same time, off-chain services have the risk of centralization and lack of trust mechanisms to protect them. We propose a Smart Contract-as-a-Service solution, which provides a method for on-chain contracts to collaborate with off-chain services. Through on-chain authentication of off-chain services, the service lifecycle is incorporated into blockchain management, so that it has the trustworthy characteristics of contract services and supports the registration, customization and combination of services. We proposes a communication and collaboration method based on transaction triggering and event-driven mechanisms, where smart contracts can be combined with extended services to construct workflows in the face of complex business scenarios or tasks with high resource demands. Empirical verification demonstrates that Smart Contracts-as-a-Service can significantly expand contract application scenarios and boost performance by shifting resource-intensive tasks from within the contract to computational extension services. The benefit of increased performance is positively correlated with the off-chain speedup ratio, for example, a 64% improvement in performance when performing multiplication of a 1 200-dimensional matrix.

Keywords: smart contract; smart contract as a service; collaboration on-chain and off-chain; service certification; event driven

Received: 2022-09-15; **Accepted**: 2022-11-25; **Published Online**: 2023-01-17 17:03

URL: link.cnki.net/urlid/11.2625.V.20230117.0909.001

Foundation items: Major Science and Technology Project of Yunnan Province (202103AN080001-001,202002AA100007,202002AD080003); Open fund of State Key Laboratory of software development (SKLSDE-2022ZX-11)

* **Corresponding author**. E-mail: yeyf@bhynii.com

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2022.0803

相变储能拓扑翅片的性能研究

殷健宝，邢玉明*，王仕淞，叶梦岩，王子贤，侯煦

(北京航空航天大学 航空科学与工程学院，北京 100191)

摘　　要：为强化翅片管式储能系统的传热速率和相变材料域的温度均匀性，基于带惩罚的固体各向同性材料方法，以两种分形结构和对应翅片体积分数为参数，采用月桂酸为相变材料，进行了拓扑优化设计，并对比研究了拓扑优化的强化导热效果和温度均匀性。研究表明：壁面温度 20 ℃时，拓扑优化有更好的导热效果，比分形优化分别减少 21.18% 和 12.68% 的总凝固时间，相变材料温度最高降低了 7.33 ℃和 4.30 ℃，平均降低了 0.98 ℃和 3.85 ℃；拓扑优化也具备更好的温度均匀性，相变材料平均方差分别为对应分形的 33.38% 和 72.13%；壁温偏离拓扑的设计参数时，其热性能也基本未发生改变。以上结果传导为翅片设计提供了一定参考。

关　键　词：相变材料；拓扑优化；分形优化；数值模拟；翅片优化

中图分类号：V221$^+$.8; V221$^+$.92; V228.3; V245.3$^+$43

文献标志码：A　　**文章编号**：1001-5965(2024)10-3274-09

空间热管理技术的进步对于处理高热流系统和降低总体质量至关重要[1]。相变材料(phase change materials, PCM)可以利用相变材料的融化和凝固过程吸收和释放大量热量，是一种理想的热管理材料，有相变工作温度变化小和储能密度高的优点[2]。换热设备中集成相变材料有助于提供额外的热管理能力，在航空热管理[3-4]和航天热管理[1,5]有广泛的应用前景。特别是在空载高能激光设备的热管理上，相变材料有助于提高空载冷却设备的功率和降低尺寸[6]。Shanmugasundaram 等人[7]在空载激光设备热管理平台引入相变材料部件，最高可以节约 36% 的系统质量。

但是除金属相变材料外，大多数相变材料均存在导热系数较低的缺点，阻碍了储能系统的储热/放热过程，制约了其进一步的发展。主要有两种强化相变材料导热的方法：添加高导热系数材料和添加翅片。高导热系数材料主要可以分为宏观高导热材料[8-9]、多孔材料[10-11]和高导热纳米粒子[12]。另一种方法是在相变材料侧添加翅片，以此增加换热面积。传统翅片主要有环肋[13]、直肋[14]和针肋[15]3 种类型。与其他方法相比，添加翅片具有施工成本低，制造简便、易于维护和鲁棒性好[16]的优点。但是传统翅片的材料存在分布不合理的问题，以直肋为例，距离管壁较远的 PCM 需要更大的翅片传热面积，普通矩形翅片较难实现这种翅片体积的合理分配。

近年来，许多研究人员将注意力转向了对不同结构翅片的研究。分形翅片仿照自然界的分形结构，可以实现较强的导热强化作用。Sheikholeslami 等[17]提出了一种新型的雪花分形翅片，通过有限元分析发现，优化后的分形翅片比纯 PCM 和同体积直翅的凝固速率分别提高了 7.8 倍和 1.7 倍，而且相比纳米粒子强化，增强效果更好，但没有较大幅度降低最大储能总量。Yu[18]等建立了树状分形翅片，通过数值模拟和响应面法对强化结构进行了优化，优化的结构比同体积直翅总融化时间减少了

收稿日期：2022-09-23；**录用日期**：2023-02-16；**网络出版时间**：2023-03-08 09:43
网络出版地址：link.cnki.net/urlid/11.2625.V.20230307.1527.003
基金项目：航空科学基金 (20172851018)
* **通信作者**. E-mail：xym505@126.com

引用格式：*殷健宝，邢玉明，王仕淞，等. 相变储能拓扑翅片的性能研究 [J]. 北京航空航天大学学报，2024，50（10）：3274-3282.*
YIN J B，XING Y M，WANG S S，et al. Study of performance of topological fin for phase change energy storage [J]. Journal of Beijing University of Aeronautics and Astronautics，2024，50（10）：3274-3282（in Chinese）.

26.7%。

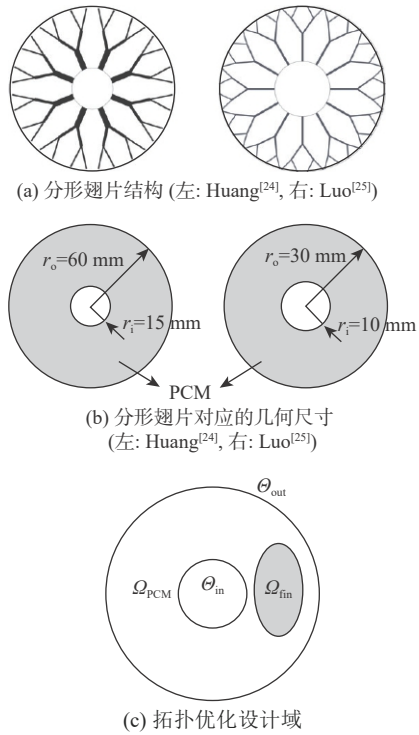
拓扑优化技术本质上是设计域材料的一种分布方法, 有着约束条件少和自由度高的优点。这一技术最初用于解决结构强度问题[19], 2007 年, Bruns[20]等在热力学中应用了这项技术, 2017 年, Pizzolato[21]等对管式储能换热器进行了拓扑优化研究, 以储罐能量最小化和时间最小化为优化目标, 得到了符合要求的翅片结构, 由此发现拓扑优化存在其他设计难以形成的设计特点和趋势; 而后研究了重力条件下的拓扑结构优化, 与纵向翅片相比, 拓扑翅片速率在蓄热和放热阶段分别提高了 37% 和 15%[22]。Tian[23]等采用带惩罚的固体各向同性材料 (solid isotropic materials with penalization parameterization, SIMP) 方法进行拓扑优化, 分析拓扑结构时, 通过分形维数计算和误差分析, 发现拓扑结构与实际叶片分形结构类似, 接近自然界的优化结构。在热领域, 拓扑优化相对于分形优化研究起步晚、研究内容少, 具体性能研究也比较少。分形优化主要集中在尺寸优化和形状优化上, 因此确定最佳结构往往需要一定数量的数值模拟结果, 并利用一些优化方法, 比如响应面法。变量越多, 所需算例越多, 这些限制意味着分形优化的自由度比较低, 结构确定所需的工作量比较大。拓扑优化自由度高, 工作量远少于分形优化, 可是优化设置要求更高, 而且拓扑翅片能否达到优异分形翅片的强化传热效果也未得到证实。所以有必要对这两种新型翅片的性能进行对比, 提供一种优化研究的参考。

本文针对相变储能系统翅片优化, 以两种优秀分形结构几何和翅片体积分数作为设计参数, 以最小温度方差为优化目标, 采用 SIMP 法对月桂酸/铝的二维翅片管进行了传热的拓扑优化, 并与对应分形进行了数值模拟对比。同时对拓扑优化的传热速率和温度均匀性进行了研究, 并改变了壁温, 研究了偏离设计参数对热性能的影响。

1　拓扑优化

1.1　几何尺寸

为了和较好的分形翅片进行对比, 选取了 Huang[24] 和 Luo[25] 的分形结构, 如图 1(a) 所示, 左边为 Huang 的分形结构示意图, 右方为 Luo 的分形结构示意图。图 1(b) 则为对应分形的几何尺寸, 也是拓扑优化的基本几何尺寸, 其中 r_i 为翅片管内管半径; r_o 为翅片管外管半径。图 1(c) 则为拓扑优化设计域示意, 其内外径与图 1(b) 相同。设计域 Ω 包括两个部分: 相变材料区 Ω_{PCM} 和高导热的翅片区 Ω_{fin}。边界条件 Θ_{in} 为恒温边界, Θ_{out} 为绝热边界。



(a) 分形翅片结构 (左: Huang[24], 右: Luo[25])

(b) 分形翅片对应的几何尺寸 (左: Huang[24], 右: Luo[25])

(c) 拓扑优化设计域

图 1　拓扑优化的几何尺寸和设计域确定示意

Fig. 1　Schematic diagram of geometry size and design domain determination for topology optimization

根据不同的目标函数, 将高导热材料填充其中。图 2 是 Huang 和 Luo 的基本树状分形单元各级尺寸示意, 其中, L_n 代表第 n 级分叉的长度; δ_n 代表第 n 级分叉的厚度; 下标代表分支的级数; L_0、L_1 和 L_2 分别代表第 0 级、第 1 级和第 2 级的分支长度; δ_0、δ_1 和 δ_2 分别代表第 0 级、第 1 级和第 2 级的分支厚度。对应文献各级的尺寸如表 1 所示。使用具体尺寸绘图后, 可以确定 Huang 和 Luo 对应分形

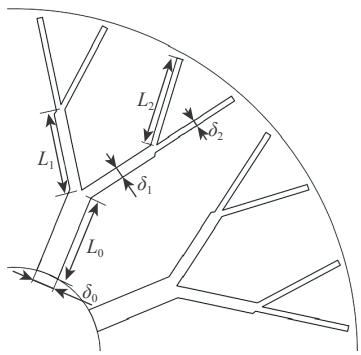


图 2　树状分形的各级尺寸示意

Fig. 2　Schematic diagram of dimensions at each level of a tree fractal

表 1　树状分形的尺寸

Table 1　Dimensions of a tree fractal

文献	L_0/mm	L_1/mm	L_2/mm	δ_1/mm	δ_2/mm	δ_3/mm
文献[24]	14.00	15.40	16.90	4.00	2.00	1.00
文献[25]	9.78	6.91	4.88	0.71	0.50	0.36

翅片的体积分数分别为0.138, 1.099 5。二者相变材料均为月桂酸,翅片为铝材料,计算所用的热物性[24]如表2所示。ρ、c_p、k、L和T_m分别代表密度、比热容、导热系数、相变潜热值和相变温度。

表2 材料热物性参数

Table 2 Thermal properties of material

ρ/(kg·m^{-3})		c_p/(kJ·kg^{-1}·℃$^{-1}$)		k/(W·m^{-1}·℃$^{-1}$)		L/(kJ·kg^{-1})(月桂酸)	T_m/℃(月桂酸)
月桂酸	铝	月桂酸	铝	月桂酸	铝		
1 007(固)/862(液)	2 730	1.7(固)/2.4(液)	0.896	0.22(固)/0.147(液)	167	178	42～48

1.2 数学模型

1.2.1 控制方程

相变传热采用焓法,其控制方程为

$$\rho\frac{\partial h}{\partial t}=\nabla\cdot(k\nabla T) \tag{1}$$

焓h表达式为

$$h=c_p(T-T_{ref})+L \tag{2}$$

式中,T为温度;T_{ref}为参考温度。

1.2.2 材料物性插值

SIMP方法[26]在优化设计方面有着广泛的应用。通过插值对物理量进行重新定义,如下所示:

$$k(x)=k_{PCM}+\gamma^{p_{factor}}(k_{fin}-k_{PCM}) \tag{3}$$

$$\rho(x)=\rho_{PCM}+\gamma(\rho_{fin}-\rho_{PCM}) \tag{4}$$

$$c_p(x)=c_{p,PCM}+\gamma\left(c_{p,fin}-c_{p,PCM}\right) \tag{5}$$

由式(1)和式(2)可以得到相变潜热带来的热源项,其插值形式为

$$Q(x)=-(1-\gamma)\rho\frac{\partial L}{\partial t}=-(1-\gamma)\rho L\frac{\partial\xi}{\partial t} \tag{6}$$

式中,下标PCM和fin分别代表相变材料和高导热翅片材料;ξ为相变材料的液相分数;γ表示材料的伪密度,导热强化区域γ=1,相变材料区域γ=0;p_{factor}为惩罚因子;t为时间;x表示材料单元位置,通过式(1)在设计域有限元节点上对两种材料热物性进行插值。

1.2.3 目标函数

拓扑优化目标为提高翅片的导热强化能力,同时考虑到温度均匀性这一重要指标,故以总设计域Ω的最小方差为优化目标,如式(7)所示,并受翅片体积分数约束,如式(8)所示:

$$\min F(\gamma)=\frac{1}{V_\Omega}\int_\Omega(T-T_{av})^2\mathrm{d}V \tag{7}$$

$$\text{s.t.}\quad C(\gamma)=\frac{\int_\Omega\gamma\mathrm{d}V}{\int_\Omega\mathrm{d}V}-\varphi\leqslant 0 \tag{8}$$

式中,V为体积;V_Ω为总设计域体积;T_{av}为域内平均温度;F和C分别为目标函数和约束函数;φ为高导热翅片的体积分数。

1.2.4 过滤法和投影法

求解Helmholtz型的偏微分方程可以实现对结果的过滤[27]。这样可以有效解决相邻节点γ值相差较大引起解的震荡。相关方程如下所示:

$$\nabla\cdot(-r_f\nabla\rho_f)+\nabla\rho_f=\gamma \tag{9}$$

式中,r_f为过滤半径;ρ_f为过滤后的伪密度。通过定义过滤半径,将半径内的平均值作为节点变量的值。r_f针对Huang和Luo分别取0.018 mm和0.002 6 mm。

在过滤之后,仍然会出现大量的中间值,通过投影函数对中间值进行处理[28]、几何重构后,结果更接近0～1分布,投影函数如下所示:

$$\rho_p=\frac{\tanh\left(\beta\left(\rho_f-\theta_\beta\right)\right)+\tanh\left(\beta\theta_\beta\right)}{\tanh\left(\beta\left(1-\theta_\beta\right)\right)+\tanh\left(\beta\theta_\beta\right)} \tag{10}$$

式中,ρ_p为投影后的伪密度;β为投影斜率;θ_β为投影点。这样可以保证翅片的体积分数不会发生改变,而且有利于得到明显的边界。投影过程β和θ_β分别取8和0.5。

1.2.5 优化算法

利用全局收敛的移动渐进[29](globally convergent method of moving asymptotes, GCMMA)算法对伪密度进行优化更新。GCMMA法属于一阶单调凸近似方法,是典型的结构优化求解方法之一。拓扑优化时的数学模型可以简化为式(11),计算流程如图3所示。其中灵敏度分析采用离散伴随方法[22]等计算目标和约束的梯度。

$$\begin{cases}\text{find}\ \ \gamma=\left[\gamma_1,\gamma_2,\cdots,\gamma_n\right]^{\mathrm{T}}\in\mathrm{R}^n\\ \text{Min}\ \ F(\gamma)\\ \text{s.t.}\ \ C(\gamma)\leqslant 0\\ \qquad 0.001\leqslant\gamma_i\leqslant 1,\ i=0,\cdots,n\end{cases} \tag{11}$$

1.2.6 边界条件和初始条件

储热翅片管内管壁为恒温热源,表达式如下:

$$T_{in}=20\ ^\circ\mathrm{C} \tag{12}$$

外管壁为绝热边界条件,表达式如下:

$$q_{out}=0 \tag{13}$$

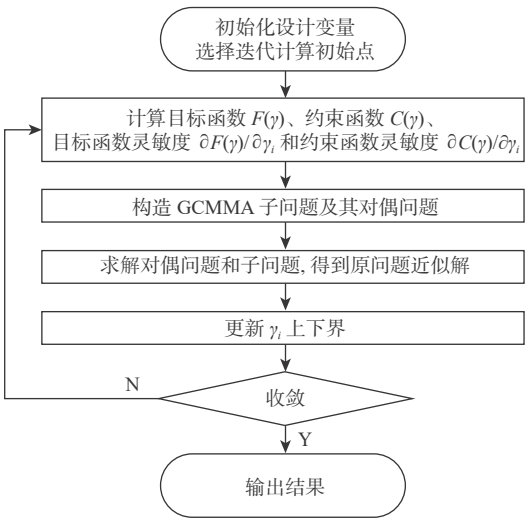


图 3　GCMMA 算法流程

Fig. 3　Flow chart of GCMMA algorithm

单元管初始温度均匀一致, 表达式如下:

$$T_0 = 49\ ^\circ\mathrm{C} \tag{14}$$

其中, T_{in} 为内壁面温度; q 为热流; q_{out} 为外壁面热流; T_0 为系统初始温度。

2　数值模拟

2.1　物理模型

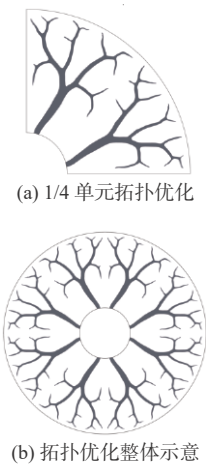
优化过程在 COMSOL Multiphysics 中进行, Huang 和 Luo 对应分形翅片的体积分数分别为 0.138 和 0.099 5, 初始材料的伪密度 γ 以及惩罚因子 p 取 0.5 和 5。采用 GCMMA 进行优化更新计算材料伪密度, 令最大总迭代步数为 2 000, 优化终止残差为 10^{-8}, 采取双曲正切投影得到清晰的结构边界, 同时利用 1.2.2 节的物性插值方法和 1.1.5 节的优化边界条件。采取 1/4 模型作为优化目标, 在 Huang 和 Luo 对应几何下, 分别划分 20 873 和 20 118 的网格量, 其内管半径分别为 15 mm 和 10 mm, 外管半径分别为 60 mm 和 30 mm, 对应的优化结果分别如图 4 和图 5 所示。拓扑优化结果与分形结构有一定的相似度, 但从优化理论上看, 二者却存在很大不同, 因此本文优化结果也不是严格意义上的分形结构。

2.2　焓-多孔介质法

采用焓-多孔介质法模拟相变热沉吸热凝固的过程。为降低模拟的复杂性, 在数值模型中作以下合理假设:

1) 将液态相变材料视为不可压缩的牛顿流体, 且流动为层流, 忽略相变过程中的密度变化;

2) 材料初始温度恒定均匀, 且均为各向同性。



(a) 1/4 单元拓扑优化

(b) 拓扑优化整体示意

图 4　拓扑优化结果(r_i=15 mm, r_o=60 mm)

Fig. 4　Topology optimization results (r_i=15 mm, r_o=60 mm)



(a) 1/4 单元拓扑优化

(b) 拓扑优化整体示意

图 5　拓扑优化结果(r_i=10 mm, r_o=30 mm)

Fig. 5　Topology optimization results (r_i=10 mm, r_o=30 mm)

2.2.1　控制方程

1) 翅片的导热模型

能量方程:

$$\rho_{Al} c_{p,Al} \frac{\partial T_{Al}}{\partial t} = \nabla^2 (k_{Al} T_{Al}) \tag{15}$$

2) 固液相变模型

连续性方程:

$$\nabla \cdot \boldsymbol{V} = 0 \tag{16}$$

动量方程:

$$\rho_P \frac{\partial \boldsymbol{V}}{\partial t} + \rho_P \boldsymbol{V}(\nabla \cdot \boldsymbol{V}) = -\nabla P_P + \mu_P \nabla^2 \boldsymbol{V} + \boldsymbol{S} \tag{17}$$

其中:

$$\boldsymbol{S} = -A_m \frac{(1-\xi)^2}{\xi^3 + 0.001} \boldsymbol{V} \tag{18}$$

$$\xi = \begin{cases} 0 & T \leqslant T_{solidus} \\ \dfrac{T - T_{solidus}}{T_{liquidus} - T_{solidus}} & T_{solidus} < T < T_{liquidus} \\ 1 & T \geqslant T_{liquidus} \end{cases} \tag{19}$$

能量方程：

$$\frac{\partial(\rho_{\mathrm{P}}h_{\mathrm{P}})}{\partial t}+\nabla\cdot(\rho_{\mathrm{P}}\boldsymbol{V}h_{\mathrm{P}})=\nabla\cdot(k_{\mathrm{P}}\nabla T_{\mathrm{P}})-\rho_{\mathrm{P}}L\frac{\partial\rho_{\mathrm{P}}}{\partial t} \tag{20}$$

式中，$\boldsymbol{V}$为速度矢量；P为静压；$\boldsymbol{S}$为动量源项；h为总焓；下标Al和P分别代表材料铝和PCM。在焓-多孔介质法中，相变材料存在3种阶段：固体阶段、液体阶段和糊状阶段。A_{m}是糊状区常数，受相变材料的热物理性质的影响，糊状区常数一般为10^5到10^8之间[30]，本文取10^5。

2.3 无关性验证

以图4(a)所示的结构进行讨论，计算了不同网格数量(48 538、82 640和117 508)和时间步长(1 s、0.5 s和0.1 s)下换热单元液体质量分数的变化情况，如图6所示。可以看出，当网格数量由48 538提升到82 640和117 508时，计算结果基本不变，时间步长对结果的影响也比较小。为了减少计算量和提高准确性，选用48 538的网格量和1 s的计算步长。

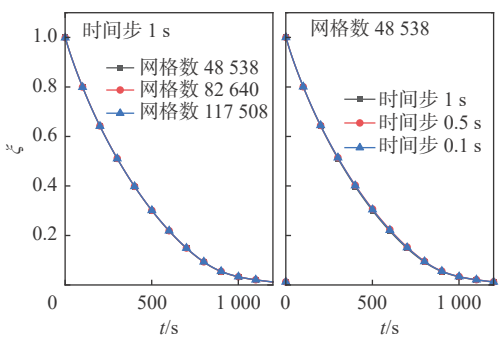
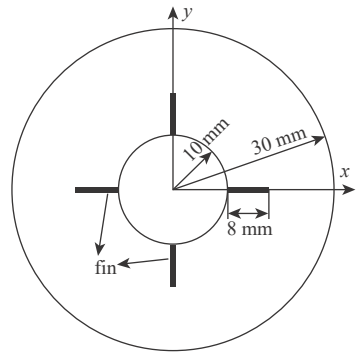


图6 网格无关性验证和时间步长无关性验证

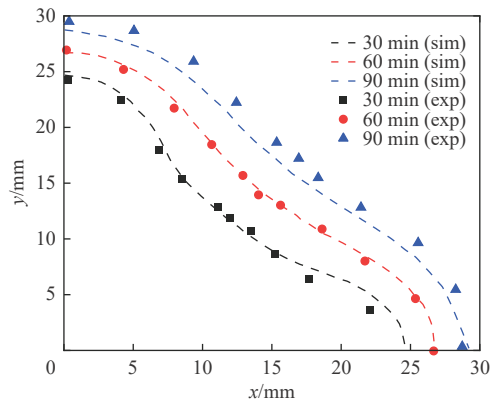
Fig. 6 Grid independent verification and time step independent verification

2.4 有效性验证

为了验证现有模型的可靠性，与文献[31]中的实验数据进行比较。将数值模拟得到的固-液相界面位置与文献[31]中的实验界面进行了比较，模型示意分别如图7(a)所示，其中x和y分别代表横、纵坐标值。储热单元外管半径为30 mm，内管半径为10 mm，内管表面焊有4个长8 mm、厚1 mm的翅片。管厚1.4 mm，管壳和翅片均为铝材料，由于铝的导热系数较高且管壁较薄，所以忽略管厚。PCM采用熔点55～57 ℃的石蜡，其密度为771.2 kg/m^3，比热容为2.176 kJ/(kg·℃)，导热系数为0.089 W/(m·℃)，潜热值为232.4 kJ/kg。PCM初始温度为65 ℃，内管为冷却边界，温度为20 ℃，外壳为绝热边界条件。图7(b)为结果对比，结果表明，焓-多孔介质法对凝固界面的判断基本与实验一致。



(a) Ismail[30] 实验单元示意



(b) 数值模拟相变界面与实验的对比

图7 模型有效性验证

Fig. 7 Validation of model

3 结果与讨论

为了更全面地研究拓扑优化的性能，除了比较PCM液态分数的下降情况，还引入PCM的平均温度$\overline{T}_{\mathrm{PCM}}$和平均温度方差$\overline{\sigma}_{\mathrm{PCM}}{}^2$作为评价指标，其中方差可以用于评价温度的均匀性，如下式所示：

$$T_{\mathrm{PCM}}=\frac{1}{V_{\mathrm{PCM}}}\int_{\Omega_{\mathrm{PCM}}}T\mathrm{d}V \tag{21}$$

$$\overline{T}_{\mathrm{PCM}}=\frac{\int_0^{t_{\mathrm{solid}}}T_{\mathrm{PCM}}\mathrm{d}t}{t_{\mathrm{solid}}} \tag{22}$$

$$\sigma_{\mathrm{PCM}}{}^2=\frac{1}{V_{\mathrm{PCM}}}\int_{\Omega_{\mathrm{PCM}}}(T-T_{\mathrm{PCM}})^2\mathrm{d}V \tag{23}$$

$$\overline{\sigma}_{\mathrm{PCM}}=\frac{\int_0^{t_{\mathrm{solid}}}\sigma_{\mathrm{PCM}}{}^2\mathrm{d}t}{t_{\mathrm{solid}}} \tag{24}$$

式中：T_{PCM}为PCM域的温度，与时间有关，代表某一时刻PCM域的平均温度；$\overline{T}_{\mathrm{PCM}}$为PCM域的平均温度，代表凝固阶段$T_{\mathrm{PCM}}$的平均值；$t_{\mathrm{solid}}$为凝固的总时长；$\sigma_{\mathrm{PCM}}{}^2$为PCM域的瞬时温度方差，与时间有关；$\overline{\sigma}_{\mathrm{PCM}}{}^2$为PCM域的平均温度方差，代表凝固阶段$\sigma_{\mathrm{PCM}}{}^2$的平均值。

3.1　拓扑优化与分形优化的性能对比

先将 Huang[24] 的分形和对应拓扑优化进行对比。如图 8 所示, 左侧为拓扑优化云图, 右侧为分形优化云图, 每张云图左半测代表液相百分数, 右半侧代表温度。由图可以看出, 凝固进行到 400 s 时, 拓扑优化的凝固速率略快一些, 温度热点也比较分散, 总体差距并不明显。过程进行到 800 s 时, 二者出现明显差异, 拓扑翅片外缘的相变材料已经接近完全凝固, 温度分布也比较均匀。分形翅片此时还有大块相变材料未凝固, 离翅片距离也较远, 可以预见之后的凝固速度会进一步减慢, 而且分形翅片边缘出现了低温区, 这说明传热均匀性并不好, 会进一步抑制导热。800 s 后, 拓扑翅片相较分形翅片仍能保持较强的导热强化作用, 温度均匀性也更好。图中拓扑仍存在局部热点, 主要是受翅片体积分数限制, 分支无法进一步生长, 如果将翅片初始体积分数提高, 再通过后处理适当缩减翅片体积, 就可以进一步提高导热性能和温度均匀性。

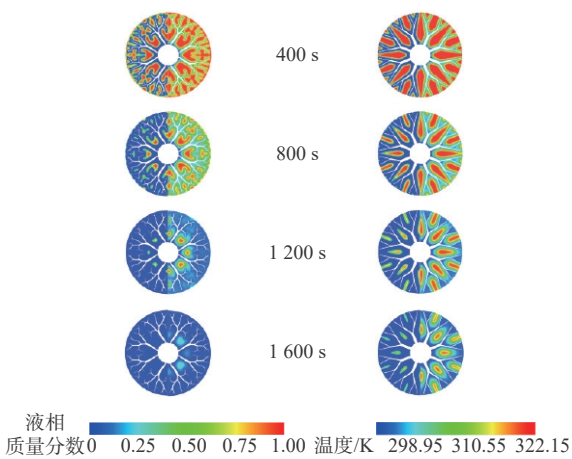


图 8　拓扑优化和 Huang 的分形优化液相质量分数云图 (左半) 和温度云图 (右半)

Fig. 8　Liquid fraction contour (left half) and Temperature contour (right half) of topology optimization and Huang's fractal optimization

进一步对二者液相分数的变化情况进行对比, 如图 9 所示, 可以看到, 初始阶段分形翅片的传热速率略快一些, 这是因为其 0 级分支厚度大、热阻小, 利于热量的传递。随着过程的进行, 拓扑翅片的传热速率超过分形翅片, 差距逐步拉大。由图 8 可以看到在 800 s 时, 分形翅片与液态相变材料距离较远, 近翅片区域温度的降低也减小了传热温差, 所以凝固速率下降加快。拓扑优化完全凝固需 1 496 s, 比分形优化减少了 21.18% 的总凝固时间。

如图 10 所示, 拓扑优化 $\overline{T}_{PCM}$ 为 33.76 ℃, 比分形低了 0.98 ℃, 而瞬时 T_{PCM} 的最大差值达到了 7.33 ℃。在温度均匀性方面, 拓扑优化相变材料内

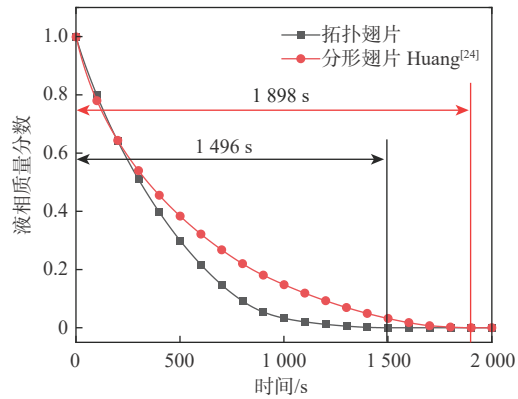


图 9　拓扑优化和 Huang 分形优化的液相分数对比

Fig. 9　Comparison diagram of liquid fraction between topology optimization and Huang's fractal optimization

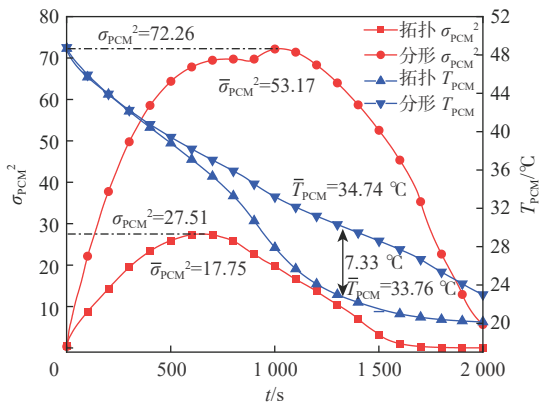


图 10　拓扑优化和 Huang 分形优化的 PCM 温度方差与温度对比

Fig. 10　Comparison diagram of σ_{PCM}^2 and T_{PCM} between topology optimization and Huang's fractal optimization

部温度方差较小, 最大 σ_{PCM}^2 为 27.51, 是分形的 38.07%, $\overline{\sigma}_{PCM}^2$ 为 17.75, 是分形的 33.38%。分形优化在 879 s 时方差突然上升, 同时 T_{PCM} 的下降速率略微增加, 这都是翅片结构分布不合理导致的局部温度较低造成的。拓扑优化可以有效改善分形在径向的温度均匀性, 在实际应用中, 这不仅可以提高系统效率, 还能延长设备的使用寿命。

对 Luo[25] 分形和对应拓扑优化进行对比。如图 11 所示, 拓扑结构仍然具备较好的凝固速率和均匀性, 但是没有很明显的差异, 主要两方面的原因: 一方面, 是结构本身造成的, Huang 的圆管尺寸大 (r_i=15 mm, r_o=60 mm), 总共 8 个分支, Luo(r_i=10 mm, r_o=30 mm) 的圆管内外距离小, 总共 12 个分支, 所以 Luo 翅片与相变材料的体积小, 传热距离更短, 传热速率更快; 另一方面, 由于翅片体积占比更小 (9.95%), 拓扑翅片生长受到一定限制, 因此优化效果相对差一些。

对二者液相分数进行对比, 如图 12 所示, 拓扑优化凝固总时长 427 s, 比分形缩短了 12.68% 的融

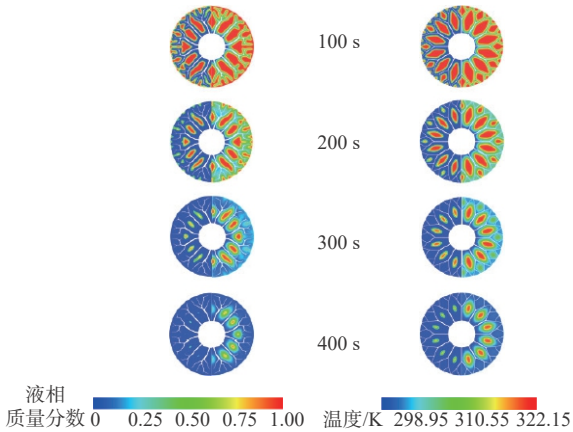


图 11 拓扑优化和 Luo 的分形优化液相质量分数云图 (左半) 和温度云图 (右半)

Fig. 11 Liquid fraction contour (left half) and temperature contour (right half) of topology optimization and Luo's fractal optimization

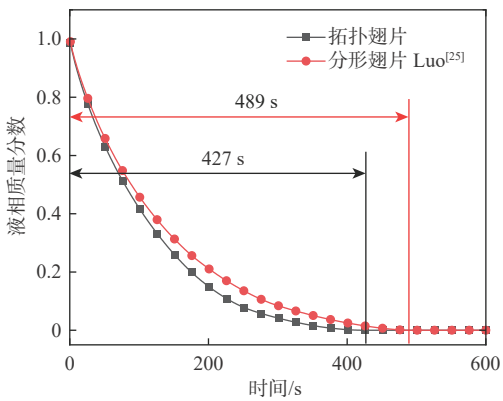


图 12 拓扑优化和 Luo 分形优化的液相分数对比

Fig. 12 Comparison diagram of liquid fraction between topology optimization and Luo's fractal optimization

化时间。在较小内外管径差和较低翅片体积分数下的情况，拓扑优化仍然有优秀的强化导热性能。

由图 13 可以看出，拓扑优化内部温度方差仍然较小，最大σ^2_{PCM}为 46.14，是分形的 70.89%；$\overline{\sigma}^2_{\mathrm{PCM}}$为 33.23，是分形的 72.13%；$\overline{T}_{\mathrm{PCM}}$为 29.39 ℃，比分形低了 3.85 ℃，而瞬时 T_{PCM} 的最大差值达到了 4.30 ℃。由于拓扑的第一个分支有较大的厚度，所以凝固速率和温度下降速率一直领先于分形。结合两种分形结构的模拟结果，零级分支的厚度相对最终凝固时间可能影响并不大，在满足强度的基础上，薄且多分支的分形结构也许会有更好的传热效率和温度均匀性。

拓扑优化自由度较高，可是其对网格划分、边界参数、投影参数、翅片体积分数和初始值等都比较敏感，所以结果不一定优于分形优化。而分形优化对几何规律比较敏感，如果更改分形的几何方程或者增加级数，则有可能会得到更好的优化结果，

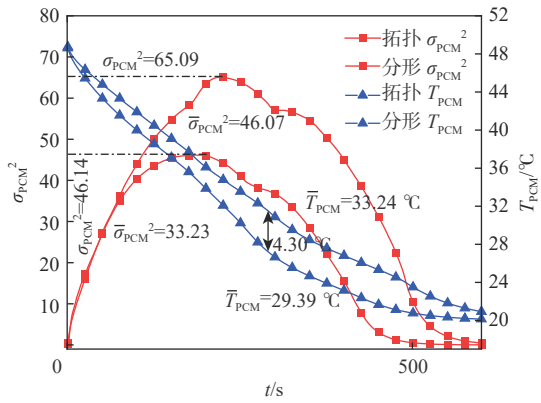


图 13 拓扑优化和 Luo 分形优化的 PCM 温度方差与温度对比

Fig. 13 Comparison diagram of σ_{PCM}^2 and T_{PCM} between topology optimization and Luo's fractal optimization

比如减小初始级数的翅片厚度，增加分支数量和分数级数。所以上述对比只说明拓扑优化结果可以优于分形优化结果。

3.2 壁温对拓扑优化结构的影响

拓扑翅片结构是基于特定温度边界条件生成的，以 Huang 分形和对应拓扑为研究对象，改变内壁面温度，研究其性能变化，如图 14 所示。壁面温度降低 10 ℃，拓扑凝固总时间降低到 1 142 s，比分形提高了 20.42%；相反地，壁面增加 10 ℃，拓扑凝固总时间增加到 2 953 s，比分形提高了 21.98%。随着壁温的降低，拓扑优化相对分形凝固速率的提高幅度有较小增加。由此基本可以认为，冷源偏离拓扑优化的设计温度后，其结构也保持一定的导热性能。对不同壁面条件下的相变域温度方差进行研究，如图 15 所示。壁温越低，翅片与相变材料温差越大，σ^2_{PCM}越大。壁温 10 ℃ 时，拓扑优化σ^2_{PCM}最大值和$\overline{\sigma}^2_{\mathrm{PCM}}$分别为 33.17 和 52.30，分别为对应分形的 39.03% 和 33.11%。壁温变化后拓扑优化的温度均匀性也较好。

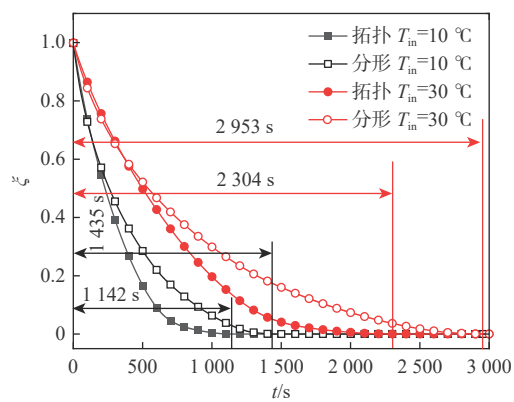


图 14 拓扑和分形优化在不同壁面温度下的液相质量分数

Fig. 14 Liquid fraction of topological and fractal optimization at different wall temperatures

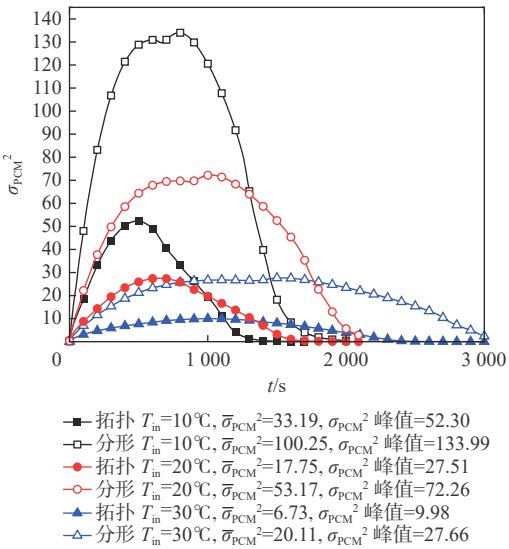


图 15　拓扑和分形优化在不同壁面温度下的 PCM 温度方差

Fig. 15　σ_{PCM}^2 of topological and fractal optimization at different wall temperatures

4　结　论

拓扑优化为相变储能系统的设计优化提供了新的方法, 其可以实现优于分形优化的设计, 而且所需计算资源更少, 设计自由度更高, 为翅片优化提供了参考。以两种优秀的分形几何结构和翅片体积为参数, 设计了对应的拓扑结构, 并进行了传热速率和温度均匀性的对比, 得到以下主要结论:

1) 拓扑优化具备更快的传热速率, 对比两种分形优化结果, 分别减少了 21.18% 和 12.68% 的凝固总时间, T_{PCM} 最大降低了 7.33 ℃ 和 4.30 ℃, $\overline{T}_{PCM}$降低了 0.98 ℃ 和 3.85 ℃。

2) 拓扑优化温度均匀性较好, $\overline{\sigma}_{PCM}^2$分别为两种分形的 33.38% 和 72.13%。

3) 在壁面温度偏离设计参数时, 相比对应分形, 拓扑优化的导热强化效果和温度均匀性基本未发生变化。

参考文献（References）

[1] KABIR M, GEMEDA T, PRELLER E, et al. Design and development of a PCM-based two-phase heat exchanger manufactured additively for spacecraft thermal management systems[J]. International Journal of Heat and Mass Transfer, 2021, 180: 121782.

[2] TATSIDJODOUNG P, LE PIERRÈS N, LUO L G. A review of potential materials for thermal energy storage in building applications [J]. Renewable and Sustainable Energy Reviews, 2013, 18: 327-349.

[3] WANG S S, XING Y M, HAO Z L, et al. Experimental study on the thermal performance of PCMs based heat sink using higher alcohol/graphite foam[J]. Applied Thermal Engineering, 2021, 198: 117452.

[4] HOU X, XING Y M, HAO Z L. Multi-objective optimization of a composite phase change material-based heat sink under non-uniform discrete heating[J]. Applied Thermal Engineering, 2021, 197: 117435.

[5] JÄCKEL R, TAPIA F, GUTIÉRREZ-URUETA G, et al. Design of an aeronautic pitot probe with a redundant heating system incorporating phase change materials[J]. Flow Measurement and Instrumentation, 2020, 76: 101817.

[6] SHANMUGASUNDARAM V, RAMALINGAM M, DONOVAN B, et al. Aircraft based pulsed power system thermal management options with energy storage[C]//Proceedings of the 4th International Energy Conversion Engineering Conference and Exhibit (IECEC). Reston: AIAA, 2006: AIAA2006-4025.

[7] SHANMUGASUNDARAM V, RAMALINGAM M, DONOVAN B. Thermal management system with energy storage for an airborne laser power system application[C]//Proceedings of the 5th International Energy Conversion Engineering Conference and Exhibit (IECEC). Reston: AIAA, 2007: AIAA2007-4817.

[8] VELRAJ R, SEENIRAJ R, HAFNER B, et al. Heat transfer enhancement in a latent heat storage system[J]. Solar Energy, 1999, 65: 171-180.

[9] BABAPOOR A, AZIZI M, KARIMI G. Thermal management of a Li-ion battery using carbon fiber-PCM composites[J]. Applied Thermal Engineering, 2015, 82(2): 281-290.

[10] OPOLOT M, ZHAO C R, LIU M, et al. Influence of cascaded graphite foams on thermal performance of high temperature phase change material storage systems[J]. Applied Thermal Engineering, 2020, 180: 115618.

[11] ZHANG H, LI X Y, LIU L Q, et al. Experimental investigation on paraffin melting in high porosity copper foam under centrifugal accelerations[J]. Applied Thermal Engineering, 2020, 178: 115504.

[12] ALGARNI S, MELLOULI S, ALQAHTANI T, et al. Experimental investigation of an evacuated tube solar collector incorporating nano-enhanced PCM as a thermal booster[J]. Applied Thermal Engineering, 2020, 180: 115831.

[13] XU H T, WANG N, ZHANG C Y, et al. Optimization on the melting performance of triplex-layer PCMs in a horizontal finned shell and tube thermal energy storage unit[J]. Applied Thermal Engineering, 2020, 176: 115409.

[14] MAHDI J M, LOHRASBI S, GANJI D D, et al. Accelerated melting of PCM in energy storage systems via novel configuration of fins in the triplex-tube heat exchanger[J]. International Journal of Heat and Mass Transfer, 2018, 124: 663-676.

[15] KATESHIA J, LAKHERA V J. Analysis of solar still integrated with phase change material and pin fins as absorbing material[J]. Journal of Energy Storage, 2021, 35: 102292.

[16] AGYENIM F, HEWITT N, EAMES P, et al. A review of materials, heat transfer and phase change problem formulation for latent heat thermal energy storage systems (LHTESS)[J]. Renewable and Sustainable Energy Reviews, 2010, 14(2): 615-628.

[17] SHEIKHOLESLAMI M, LOHRASBI S, GANJI D D. Numerical analysis of discharging process acceleration in LHTESS by immersing innovative fin configuration using finite element method[J]. Applied Thermal Engineering, 2016, 107: 154-166.

[18] YU C, WU S C, HUANG Y P, et al. Charging performance optim-

ization of a latent heat storage unit with fractal tree-like fins[J]. Journal of Energy Storage, 2020, 30: 101498.

[19] BENDSØE M P. Optimal shape design as a material distribution problem[J]. Structural Optimization, 1989, 1(4): 193-202.

[20] BRUNS T E. Topology optimization of convection-dominated, steady-state heat transfer problems[J]. International Journal of Heat and Mass Transfer, 2007, 50(15-16): 2859-2873.

[21] PIZZOLATO A, SHARMA A, MAUTE K, et al. Topology optimization for heat transfer enhancement in Latent Heat Thermal Energy Storage[J]. International Journal of Heat and Mass Transfer, 2017, 113: 875-888.

[22] PIZZOLATO A, SHARMA A, MAUTE K, et al. Design of effective fins for fast PCM melting and solidification in shell-and-tube latent heat thermal energy storage through topology optimization[J]. Applied Energy, 2017, 208: 210-227.

[23] TIAN Y, LIU X L, XU Q, et al. Bionic topology optimization of fins for rapid latent heat thermal energy storage[J]. Applied Thermal Engineering, 2021, 194: 117104.

[24] HUANG Y P, LIU X D. Charging and discharging enhancement of a vertical latent heat storage unit by fractal tree-shaped fins[J]. Renewable Energy, 2021, 174: 199-217.

[25] LUO X M, GU J A, MA H Q, et al. Numerical study on enhanced melting heat transfer of PCM by the combined fractal fins[J]. Journal of Energy Storage, 2022, 45: 103780.

[26] ZHOU M, ROZVANY G. The COC algorithm, Part II: Topological, geometrical and generalized shape optimization[J]. Computer Methods in Applied Mechanics and Engineering, 1991, 89(1-3): 309-336.

[27] LAZAROV B S, SIGMUND O. Filters in topology optimization based on Helmholtz-type differential equations[J]. International Journal for Numerical Methods in Engineering, 2011, 86(6): 765-781.

[28] WANG F W, LAZAROV B S, SIGMUND O. On projection methods, convergence and robust formulations in topology optimization [J]. Structural and Multidisciplinary Optimization, 2011, 43(6): 767-784.

[29] Guo L H, Tang W C. Deep Study and Improvement of the Globally Convergent Version of MMA[J]. Applied Mechanics and Materials, 2014, 635: 1666-1670.

[30] SHMUELI H, ZISKIND G, LETAN R. Melting in a vertical cylindrical tube: numerical investigation and comparison with experiments[J]. International Journal of Heat and Mass Transfer, 2010, 53(19-20): 4082-4091.

[31] ISMAIL K A R, ALVES C L F, MODESTO M S. Numerical and experimental study on the solidification of PCM around a vertical axially finned isothermal cylinder[J]. Applied Thermal Engineering, 2001, 21(1): 53-77.

Study of performance of topological fin for phase change energy storage

YIN Jianbao, XING Yuming*, WANG Shisong, YE Mengyan, WANG Zixian, HOU Xu

(School of Aeronautic Science and Engineering, Beijing University of Aeronautics and Astronautics, Beijing 100191, China)

Abstract: In order to improve the temperature uniformity of the phase change material domain and the heat transfer rate of the finned tube energy storage system, two types of fractal structures and their corresponding fin volume fraction were used as design parameters, and lauric acid was used as the phase change material to carry out topology optimization design. This was done using the Solid Isotropic Material with Penalty method. The thermal conductivity enhancement and temperature uniformity of topology optimization are compared. The findings indicate that the topology optimization has a better thermal conductivity impact and can shorten the overall solidification time by 21.18% and 12.68%, respectively, when the wall temperature is 20 ℃. The phase change material temperature is reduced by 7.33 ℃ and 4.30 ℃, 0.98 ℃ and 3.85 ℃ on average. At the same time, the topology optimization also has better temperature uniformity, and the average variance of phase change material is 33.38% and 72.13% of the corresponding fractal, respectively. When the wall temperature deviates from the design parameters of the topology, the thermal performance does not change. This study provides some reference for fin design.

Keywords: phase change material; topology optimization; fractal optimization; numerical simulation; fin optimization

Received: 2022-09-23; **Accepted**: 2023-02-16; **Published Online**: 2023-03-08 09:43

URL: link.cnki.net/urlid/11.2625.V.20230307.1527.003

Foundation item: Aeronautical Science Foundation of China (20172851018)

* **Corresponding author**. E-mail: xym505@163.com

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2022.0814

一种基于多链协同的联盟链改进模型

赵鲜鲜[1,2]，谭海波[1,2]，赵赫[1,*]，周桐[3]，程昊天[1,2]，李金泽[1,2]

(1. 中国科学院合肥物质科学研究院，合肥 230031；2. 中国科学技术大学，合肥 230026；
3. 安徽中科晶格技术有限公司，合肥 230088)

摘 要：监管友好、性能优异的联盟链是我国政务和商用区块链的首选。然而联盟链通常因为有效节点少、节点独立性低和发展能力弱的原因，降低系统的可靠性。因此提出了同盟链：一种通过协同多个联盟链来提升系统可靠性的模型。同盟链通过联盟链间相互层叠存储链上数据降低联盟链历史数据被篡改的可能性，通过协同存储区块链加密数据保障账本数据恢复能力；使用聚合签名、数据完整性验证等方法解决数据传输完整性、来源真实性问题；设计双向数据存在性验证方法检验传输数据的连续性，确保协同数据相互背书的真实性；通过与现有方案进行通信、存储和计算消耗上的对比来验证本方案的有效性。

关 键 词：区块链；联盟链；可靠性；去中心化；数据完整性；数据真实性
中图分类号：TP315
文献标志码：A **文章编号**：1001-5965(2024)10-3283-14

中央政治局第十八次集体学习以来，区块链技术已上升为国家科技战略。我国区块链产业发展迅速，为多个行业的新发展增信赋能。2021年我国启动了“区块链创新应用试点行动”，在国务院、央行、最高法等18个重要部门启动区块链创新试点工作。同年5月，工信部联合中央网信办发布了《关于加快推动区块链技术应用和产业发展的指导意见》，指出到2025年我国区块链产业综合实力达到世界先进水平，产业初具规模，区块链应用渗透经济社会多个领域，形成场景化示范应用。

区块链按节点准入机制分为公有链、联盟链和私有链[1]。联盟链节点因严格的准入机制具有较高的信任度，故大多数联盟链采用性能较高的传统BFT[2]类分布式共识算法加快达成共识。因此，监管友好、性能优异的联盟链成为区块链落地商用的首选[3]。我国以联盟链为主的区块链产业规模至2021年已突破65亿元[4]。

系统可靠性更加关注数据在系统中的长期存在性[5]，区块链系统的可靠性主要体现在历史数据长期存在和难以篡改两方面。联盟链应用的主要参与方多为权威机构、第三方平台、企业等，仍然存在一定的中心化特征，相较于公有链，联盟链通常在合谋等一些攻击上更易实现[6]，因而区块链的历史数据能在一定时间内被篡改，影响其可靠性。因此，承载了大量重要行业应用的联盟链亟须获得更强的可靠性保障。本文将现有联盟链可靠性低的原因总结如下：

1）有效节点数少。有效节点参与维护联盟链，参与联盟链的有效节点数越少，合谋等攻击越有可能发生[6]。联盟链参与节点数较少的原因是区块链使用者受提供解决方案企业的规模、建链成本以及使用机构的组织结构制约。

2）节点独立性低。通常来说，维护联盟链的节点行为相互独立程度越高，区块链的去中心化程度

收稿日期：2022-09-29；**录用日期**：2022-11-25；**网络出版时间**：2023-03-09 16:07
网络出版地址：link.cnki.net/urlid/11.2625.V.20230309.1406.007
基金项目：国家重点研发计划 (2021YFB2700800)
*通信作者. E-mail：zhaoh@hfcas.ac.cn

引用格式：赵鲜鲜，谭海波，赵赫，等. 一种基于多链协同的联盟链改进模型 [J]. 北京航空航天大学学报，2024，50（10）：3283-3296.
ZHAO X X，TAN H B，ZHAO H，et al. A consortium chain improvement model based on multi-chain collaboration[J]. Journal of Beijing University of Aeronautics and Astronautics，2024，50（10）：3283-3296（in Chinese）.

越强，篡改难度越大。节点独立性受区块链本身承载业务的特性所限，多数维护区块链的节点基本隶属同一行业，甚至是同一机构[4]，且一般无法接受其他机构或与行业无关的主体。

3）可持续发展能力弱。区块链系统的可迁移性和恢复能力越强，历史数据长期存在性越好。可持续能力主要体现在技术选型等原因产生的数据需要进行数据迁移[7]和系统数据丢失、崩溃或停止运行情况下的容灾备份。现有大部分联盟链一旦发生数据损毁或停止运行等事件后就无法恢复数据以及证明数据存在。

针对以上原因，本文提出同盟链模型，以期在不干扰区块链原有事务和不显著增加区块链的维护负担下，降低链上历史数据的篡改概率和设计账本恢复机制，以提升区块链的可靠性。通过不同联盟链间持续存储对方每个阶段发送的区块链数据，形成区块链数据的相互层叠存储，使同盟链模型中的每个节点都能间接参与所有联盟链数据的维护；通过对加密区块链数据的协同存储，提升链上数据存储的去中心化程度，在有相关证明的情况下能够对损毁或缺失的数据进行反向恢复或佐证链上数据的真实性。

1 相关工作

1.1 现有联盟链可靠性提升方法

现有提高联盟链可靠性的方法主要为单链改造和多链协同两种。

单链改造分为直接扩展区块链节点数量和改进共识节点可靠性。直接扩展节点数量是最简单的可靠性提升方法，理论上能够降低系统被攻击的风险，但该方法增加的节点并非一定为有效节点或独立性高的节点，一定程度上可能具有中心化特征。改进共识节点可靠性通过提高共识节点的服务可靠性或随机性来提升系统抵御风险的能力[8-9]，但该方法在节点数较少或大部分共识节点信任度不高的情况下不能提高系统可靠性，且难以在已运行的联盟链上进行改造。

多链协同指借助其他区块链来提高原有链可靠性，比较典型的是锚定模式[10-11]和跨链协同[12-15]存储模式。锚定模式是指将原有区块链的数据定期存储至公有链或主链中，通过公有链或主链所有节点存储该段区块链数据进行背书，2013年PoE[10](proof of existence)提出将数据定期锚定至比特币区块链中，通过返回的时间戳等数据反向验证其在原有区块链中的存在性。跨链协同模式指原有区块链数据流通在多个其他区块链中，通过其他链该流转数据的相关交易证明原有链上该数据的存在性。Optimistic rollup[13]是以太坊首主导的扩展性技术，在数据流转的基础上保证了数据可用性。其主要思想：节点缴纳一定保证金成为聚合者，聚合者将一段时间内原有链的交易数据压缩并聚合，再通过智能合约将其存储在另一条链上，并进入较长时间的等待期。若等待期内没有节点发出欺诈证明则交易会被确认，聚合者将得到奖励，反之聚合者将受到惩罚。

1.2 多链协同机制面临的问题

单链改造中除直接提升节点数量方案外，大概率还需要暂停区块链的运行，这对于正在使用区块链处理业务的机构来说是难以接受的，且暂停期间数据容易被篡改，而直接增加节点数量可能会带来中心化的风险和较高的成本（特别是在实际操作节点用户较少的联盟链中），故本文采用多链协同方式来提高区块链的可靠性，而采用多链协同机制则需要对数据来源真实性、数据存在性、完整性进行保障[16]。

数据来源真实性，主要对链间数据来源进行认证，现阶段大多采用数字证书、数字签名等数据认证方案[17]。本文默认每条联盟链由一个真实且唯一的数字身份进行初始化并通信。联盟链的准入机制使得第三方通常无法查看链上数据，数据来源的真实性判断就转变为如何证明节点向外传输的数据是真实的区块链数据。因为联盟链能够追溯到每一个节点的身份，且大多数共识节点是诚实的，因此，我们可以采用多数节点投票签名和多个不同共识节点确认消息的方式解决上述问题，同时运用聚合签名压缩多个节点签名以减少通信数据量的大小。

数据存在性，本文主要针对区块链交易，只有数据发送给其他区块链后验证是否被真实存储，附加到区块链账本中的数据才具有比较高的防篡改能力。区块链的安全性由所有节点进行维护，交易是将数据添加到账本后的承载实体。本文通过对比交易数据并使用Merkle树[18]验证该交易在其他区块链中的存在性。

数据完整性，一方面体现在传输数据所指向区块链的连续性，防止发送方伪造区块链；另一方面体现在传输的加密数据在保存方中的完整性，防止保存方篡改、删除甚至重新外包加密数据。对于发送方，本文通过密码学累加器保证数据链的连续性。密码累加器通过具有准交换属性和单向性的哈希函数开发出一种可用于为时间戳和成员资格测试的单向累加器加密协议，于1994年首次被Benaloh[19]提出，分为对称累加器和非对称累加器，其能够高效地证明元素是否存在于集合中。以每

次传输的数据为累加器元素构造累加器, 若根据之前的累加器根和本次数据能够构造和发送方相同的新累加器值, 即隐式地确认了之前传输数据的存在性, 进而保证多个节点确认后发送方区块链的完整性; 对于保存方, 本文采用数据可拥有证明(provable data possession, PDP)防止加密数据被篡改, 并通过使用不同密钥加密数据防止数据外包。PDP 是一种用于远程完整性的方案, 基于该方案可以检测外包数据中数据的损坏。Ateniese[20] 在 2007 年定义了通用的 PDP 协议架构, 通常由数据发送方随机挑选存储的元数据发起挑战, 数据存储方根据挑战访问小部分文件块生成证据证明保存数据的完整性。

由于本文的重点不在于密码学工具的改进, 简单起见, 所用工具均采用较简单的实现方式, 其他诸如改进或具有相似作用的密码学工具都可以替代本文所述工具, 以提升安全性或效率。

2　同盟链模型

2.1　系统架构

本文的系统架构如图 1 所示, 主要由联盟链、信使节点和普通节点 3 种角色, 以及成员通信配置(alliance intercommunication configuration, AIC)、通信服务、同盟通信协议 (alliance communication protocol, ACP) 组成。相互背书的数据为协同数据, 区块链上确认一定区块数量后由该联盟链的信使节点调用通信服务按照协议 ACP 发送。接受方联盟链的信使节点通过通信服务监听并解析协同数据, 将符合条件的数据构造交易上链, 并将需要保存的加密数据存储到相关介质中, 接受发送方的完整性挑战并给出挑战应答。该存储介质可以是云存储, 也可以是硬盘等。为了保证信使节点的服务质量, 信使节点由联盟链中的共识节点产生。

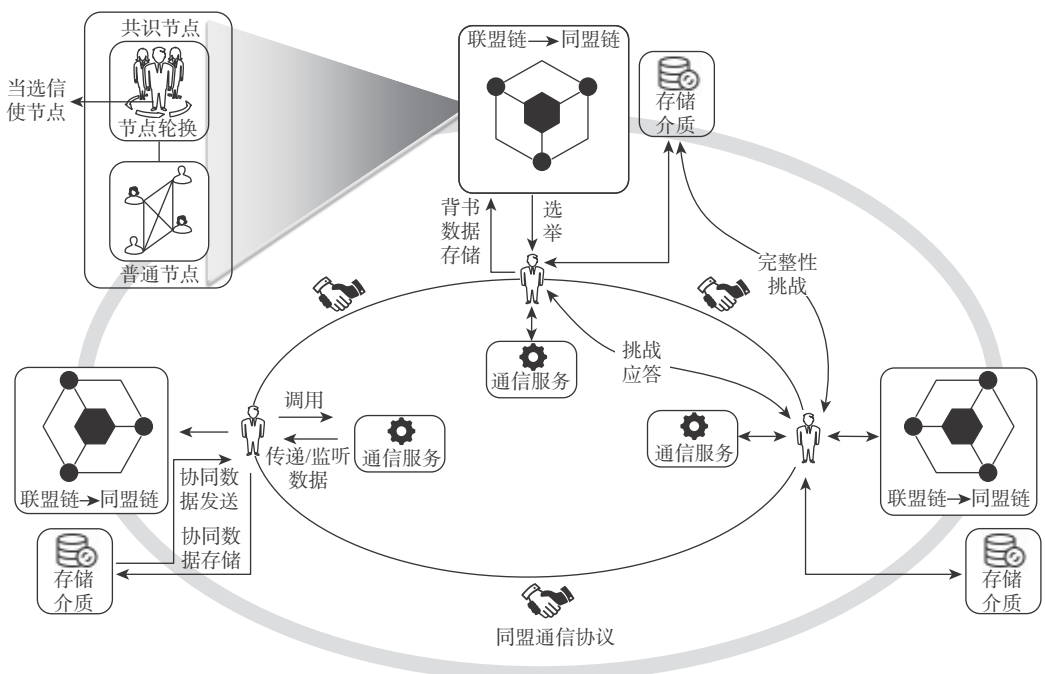


图 1　同盟链系统框架

Fig. 1　System framework of alliance chain

在同盟链模型中, 多个联盟链在自己的账本中不断附加其他结盟区块链协同消息的标识信息, 从而形成区块链数据间的层叠, 若此时篡改一个已协同区块, 难度将增大。如图 2 所示, 联盟链 A 中第一个轮次 Epoch1 的协同数据 A_1, 其唯一标识满足一定要求后在 Epoch2 中被附加在联盟链 B 和 C 中的某个区块里, 随着联盟链之间不断相互存储不同轮次的唯一标识, 联盟链之间的数据层叠存储, 一个已协同的数据还需要篡改其他存储该数据标识的区块链中对应的数据, 联盟链之间形成同盟关系。

2.2　系统组件

同盟链: 需要提升可靠性的联盟链, 相互协同存储后组成同盟链, 符合 BFT 类协议的基本假设。联盟链中包含普通节点、共识节点和信使节点。根据数据流向联盟链可以分为数据发送方和数据协同方。

信使节点: 信使节点是同盟链模型中数据的收发方, 通过通信服务接收验证其他联盟链传输的消息、发起对传输数据的签名、聚合所有消息签名并通过通信服务发送数据至其他联盟链。

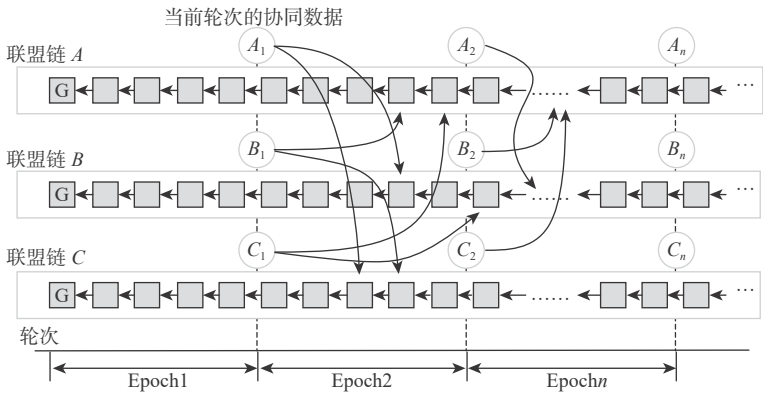


图 2 链间数据存储层叠示意

Fig. 2 Schematic of data storage overlapping between blockchains

普通节点：普通节点负责对同盟链中的事务进行验证和签名。

协同数据：协同数据包括链上背书数据 EndorseData 和链下存储数据 ArchiveData。EndorseData 包括当前轮次发送区块的区块头和必要的验证数据，能够唯一标识该数据的信息通过验证后存储在数据协同方的帐本中，即多个联盟链数据存储形成层叠关系。ArchiveData 包括数据发送方加密的区块链数据和其他需要存储的数据如运行日志等，其不仅可以恢复数据，也能帮助数据发送方恢复到当时的运行状态。ArchiveData 存储在数据协同方的存储介质中，只有当选的信使节点才能向其中追加数据，其他信使节点仅被允许查看，因此该数据不是必需的传输数据。本文使用开源库 Zlib[21] 对 ArchiveData 进行压缩来减少数据大小，后文所述 ArchiveData 均为已被压缩的数据。

通信服务：通信服务和链上节点绑定，具体功能包括监听、路由与数据转换。监听功能分为链上数据监听和协同数据监听，链上数据监听指该服务能够获取链上的状态和数据，协同数据监听指能够获取其他联盟链发送的消息；路由功能根据 AIC 中的路由信息找到传输的对象；数据转换功能将链上监听的数据按照 ACP 的格式进行包装进并请求信使节点对包装后数据的签名。信使节点收到数据后，根据 ACP 对数据展开解析；使用双向数据存在性验证方法（verifiable bidirectional method, VBM）检验区块链的连续性和历史协同数据交易的存在性；校验数据是否符合 AIC 中的各项规则，通过后数据协同方对该数据进行存储。普通节点不具备监听和路由的功能。简便起见，本文后面仅描述节点操作，不再对节点和服务间的交互细节进行阐述。

2.3 成员通信配置 AIC

联盟链间通信时，需要一些配置信息帮助路由找到其他信使节点，并检测联盟链、节点以及传输信息是否合格，关键配置如表 1 所示。

表 1 AIC 关键配置项

Table 1 Key configuration items of AIC

配置项	作用
Ecert	现任信使节点当选证明
AEnvInfo	其他联盟链信使节点证书列表和路由信息
AConfigList	其他联盟链配置信息
EDataQuene	收发双方待确认、已确认数据次数映射列表
ECerList	记录每个成员近期传输证明映射列表
BlackNode	记录禁止通信的信使节点名单
BlackChain	记录作恶联盟链唯一标识列表
Miniconfig	所有联盟链确认加入同盟链模型的最低配置

表中，AConfigList 记录其他联盟链的配置，该配置为成员加入时发送的节点总数、链类型、公钥列表等，可以检验联盟链数据合法性；EDataQuene 能防止其他联盟链的单个信使节点作恶，存储当前轮次发送区块数量、最新区块的高度和对应的确认次数，前一次数据确认后，列表对应字段加 1；ECerList 能够防止由于异常导致的累加器根值不同无法检验或丢弃有用数据的问题；BlackNode 记录作恶节点地址和公钥；BlackChain 记录作恶区块链的 ChainID；Miniconfig 能够防止新成员加入导致的同盟分裂。

2.4 同盟通信协议 ACP

联盟链间的通信依赖于同盟通信协议 ACP，是为了实现联盟链间交换信息而设计的规范，其详细结构如图 3 所示。其中，图 3 左侧为数据传输的详细格式，发送链和接收链有唯一标识，统称为 ChainID，from 和 to 表示数据收发方；消息类型 MType 包括同盟成员加入 MemAdd、协同传输 MemTran、同盟配置更新 MemUpdate、信使节点更新、同盟成员退出 MemQuit 和数据使用 DataUse；Content 是根据 MType 构造的消息内容；Extra 是自

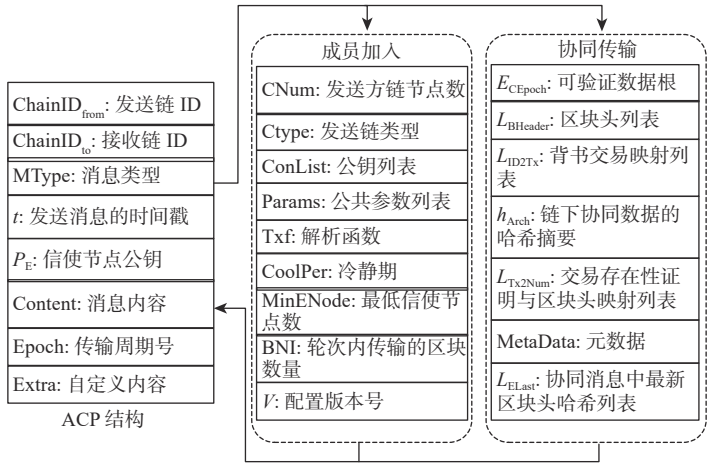


图 3　ACP 结构

Fig. 3　Structure of ACP

定义内容, 用于存放成员间相互协商的其他参数, 以及信使节点当选证明和警告信息。图 3 右侧为 MemAdd 和 MemTran 消息的构造。其中, L_{ID2Tx} 包括<ChainID, T_{x_i}>, T_{x_i}是序号为 i 的交易, 该交易需要对应区块链的信使节点加密; L_{ELast} 包括<ChainID, h_{BLast}, EC>, h_{BLast} 为一个轮次中最新的区块头哈希, EC 为当前链的链上背书证明; L_{Tx2Num}=<T_{x_i}, MerkleProof, N_B>, Merkle-Proof 为交易 T_{x_i} 的证明路径, N_B 为区块号; 成员退出时只存放最后发送消息的可验证数据和返回的数据索引; 同盟配置更新仅构造更新的参数; 信使节点轮换用新信使节点的当选证明 Ecert 和其他消息类型一起发送。

2.5　双向数据存在性验证方法 VBM

联盟链之间不仅需要数据发送方证明传输数据为真实的区块链数据, 而且需要证明数据协同方之前传输的数据经过验证后已经存储在数据发送方的区块链中。VBM 基于累加器构造, 累加器成员为 EndorseData 的唯一标识, 累加器根为背书数据证明 (endorse certification, EC)。

所有联盟链维护一个 EC, 不同链可能会由于网络延迟等原因异步接收消息从而使下一步发送协同消息中的 EC 值不同, 但由于 AIC 中存储了 ECerList, 因此能够借助近期历史数据对不同链进行异步验证。

数据发送方第i个发送轮次发送的区块头列表如图 4 所示, T_{x_3}为轮次 i 中数据协同方被确认的历史协同数据所构造的交易, 序号为 3。EC 构造算法 ConEC() 将发送区块列表哈希值 h_{Blist}、列表最新区块头哈希 h_{BLast}、列表最新区块高度 H_{BLast}、链下存储数据摘要 h_{Arch} 和 Epoch 构造成最新的累加元素添加进累加器中。数据协同方对 EC 的合法性验证算法 VerifyEC() 分为两步, 第一步验证协同方已达确认阈值的协同数据是否在发送方区块链数据中, 首先验证T_{x_3}的内容是否为已达阈值的确认数据, 通

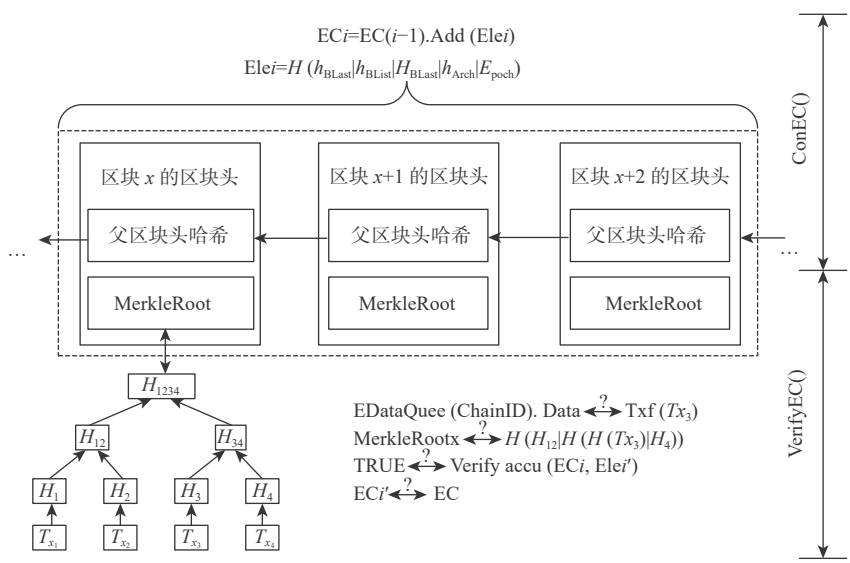


图 4　EC 构造和验证过程

Fig. 4　Process of EC construction and verification

过T_{x_3}的路径证明生成MerkleRootx, 对比是否与区块x中的根值MerkleRoot相等, 即判断其是否在区块链中; 第二步判断传输数据是否连续, 数据协同方根据ConEC()重新计算得到累加元素Ele$i$', 并根据前一次的证明EC(i−1)构造ECi', 与数据发送方发送的ECi进行对比。由于所有联盟链共用一个累加器, 故需要ECerList保存每个联盟链近几次的EC, 否则协同方只能证明该段数据的存在性, 而不能证明单个联盟链数据的连续性。

3 同盟链通信过程

本文所设计方法可以支持多个联盟链, 为了更加简单直观地描述同盟链的通信过程, 在此以联盟链A和B的行为阐述主要流程, 并将A作为行为发起方。

3.1 初始化阶段

由于联盟链不对外公开, 初始化时难以保障参数的正确性, 本文假设ChainID、信使节点公钥与IP、ACP协议已通过可信方式被其他联盟链所知。

1) 参数初始化

选定单个联盟链的安全参数γ, 基数为p的循环群G_1和G_2, 双线性映射$e:G_1\times G_2\to G_T$, 选择两个抗碰撞的安全散列函数H_0、$H_1:\{0,1\}^*\to G_1$, 私钥k和公钥P作为联盟链的聚合公私钥对, 根据每个联盟链的ChainID构造ArchiveData加密的公私钥对$\{k_{\mathrm{ID}},P_{\mathrm{ID}}\}$, 向所有联盟链公开$P$和$P_{\mathrm{ID}}$, 联盟链内部公开私钥$k$和$k_{\mathrm{ID}}$, 为各节点保存获取的密钥。系统参数为

$$\mathrm{Params}=\{\gamma,G_1,G_2,H_0,H_1,p,e,P,P_{\mathrm{ID}}\}$$

2) 信使节点选取

由于不同的区块链共识节点选取方式的差异, 难以使用统一的选举方式选出信使节点, 本文使用原区块链内已被认可的共识算法在共识节点集合中选举信使节点, 对此不再赘述。

A选出任期i内的信使节点$E_A^{i_0}$和j个候选节点$\mathrm{NodeListE}=\{E_A^{i_0},\cdots,E_A^{i_j}\}$, 当信使节点$E_A^{i_0}$作恶或者宕机时, 候选节点会广播当选信息并从已确认的位置开始执行。为了使链中原有的共识过程尽可能不受其他联盟链的影响, 当选的信使节点在任期内将不进行链内事务的共识。

E_A为当选节点, 当选后其在A中广播当选信息Ecert=$\{T_{\mathrm{S}},\mathrm{SigPE},P_{\mathrm{E}}\}$, $\{k_{\mathrm{E}},P_{\mathrm{E}}\}$为节点公私钥, T_{S}是信使节点竞选时公布的时间戳, σ_i表示其他节点对当前信使节点T_{S}和P_{E}的签名, Sig表示聚合签名, SigPE是E_A对Sig的签名。

3.2 构造阶段

E_A遵循ACP协议并根据传输消息的类型构造不同的消息M。

当构造MemAdd时, E_A首先向其他联盟成员请求同盟最低配置Miniconfig和其他联盟链的配置信息, 验证后向所在链节点广播并存入AIC相应字段中。E_A根据协议构造链初始配置, 广播该配置并收集区块链中其他节点对该配置的签名。MemAdd未被确认前不能进行协同数据的发送, 只能接受其他联盟链的信息。当被确认后从创世区块开始每个轮次都发送BNI数量的区块。

当构造MemTran时, E_A根据传输的区块头列表L_{BHeader}调用函数ConEC()计算EC, 检查当次传输的区块中是否有其他联盟链之前发送的协同数据, 若有, 则构造L_{ID2Tx}和L_{Tx2Num}。若发送ArchiveData还需要使用B对应的密钥对原始数据加密, 则会生成同态标签并发起历史协同数据完整性的挑战。同态标签生成时, 信使节点首先会查找ChainID对应的公私钥对$\{k_{\mathrm{ID}},P_{\mathrm{ID}}\}$。将ArchiveData进行分块处理后得到$F=\{m_1,\cdots,m_n\}$, 选取分块的唯一文件标识$v$和辅助变量$u$, 生成认证元$\theta_i=(H(m_i)\cdot u^{m_i})^{k_{\mathrm{ID}}}$, 集合$\Phi=\{\theta_1,\cdots,\theta_n\}$, 对认证元集合构造Merkle认证哈希树[18]M_{Tree}, M_{Tree}存放入字段Metadata中。在标签生成后, E_A可以在任意时间发起对该数据的挑战。E_A从之前的F中随机选出c块, 并从每个块索引中选取一个随机数v_i, 生成挑战$C_{\mathrm{ID}}=\{i,v_i\}$。

应答方需要在相应时间内进行回应。E_B接收到该消息时检索ArchiveData, 并生成应答。E_B根据挑战方发送的挑战数据, 计算证明$\{\theta,\mu\}$、块索引i在M_{Tree}中的路径证明Ω_i和$H(m_i)$, 一起发送给挑战发起的信使节点, 公式如下:

$$\theta=\prod_{i=1}^{c}\theta_i^{v_i} \tag{1}$$

$$\mu=\sum_{i=1}^{c}v_i m_i \tag{2}$$

当构造MemUpdate时, E_A根据所在联盟链或Miniconfig进行参数调整, 在新配置被确认前, 仍需要采用旧配置进行通信。

当构造MemQuit时, 若联盟链正常退出, E_A则将所在区块链最新的EC放入Content中, 等待其他联盟链返回A中已被确认数据的索引信息和对该EC的签名作为归档信息, 等待Coolper时间后可退出。

警告消息将发送错误消息的节点公钥、节点IP、错误消息描述、错误消息摘要、错误消息的聚合签名放入字段Extra中。

3.3　同步阶段

同步流程如图 5 所示, 流程详细叙述如下:

1) 在数据发送方中, 若 E_A 为第一次进行协同数据传输, 则需要将 Ecert 放入 Extra 字段中, 再根据传输消息类型 MType 构造消息 M。

2) E_A 将消息 M 在 A 中发起共识。

3) A 中其他节点对消息 M 验证并签名。

4) 签名的节点将签名 σ 发送给 E_A。

5) 当收集到多数签名后, E_A 将签名进行聚合生成 Sig。签名聚合的具体流程如下:

a. 联盟链内节点验证 E_A 对消息的签名是否真实, 真实则对 M 进行签名, 签名为 $\sigma = kH_0(M)$。

b. 信使节点收到签名后验证签名是否合法, 合法则接受该签名, 并记录签名的公钥至公钥列表 $\text{ConList} = \{P_1, \cdots, P_n\}$。当收集到超过半数的签名后开始进行聚合。

c. 计算 n 个用户对消息 M 的聚合签名 $\text{Sig} = \sum_{i=1}^{n} \sigma_i$。

6) E_A 构造 ACP 并签名后发送给其他联盟链。

7) B 中的信使节点 E_B 解析数据并检查消息的合法性、是否有警告消息, 不合法则构造警告消息, 若有警告消息则查看消息类型并采取相应措施。若为 MemTran, 则可以调用 VBM 方法检验历史协同数据的存在性和确认对发送方之前数据的确认, 并对完整性挑战作出应答, 该应答可以实时传输给 E_A。若消息类型为 DataUse, 则验证后查找相应的数据并构造证明传输给 E_A。

8) 信使节点 E_B 将验证后的数据 M 签名后广播给所在区块链中的其他节点。

9) 普通节点将 M 存放在 EDataQuene 中, 当信使节点发起上链交易后将相应数据从 EDataQuene 中删去, 上链数据构造交易为 $T_x = \{\text{ChainID}, N_B, h_{\text{BLast}}, h_{\text{Arch}}, \text{Epoch}, \text{BNI}\}$, 被确认的数据将会接受完整性挑战。

10) E_A 对 E_B 返回的挑战应答数据进行验证。

11) E_B 将构造的数据传输给 E_A。

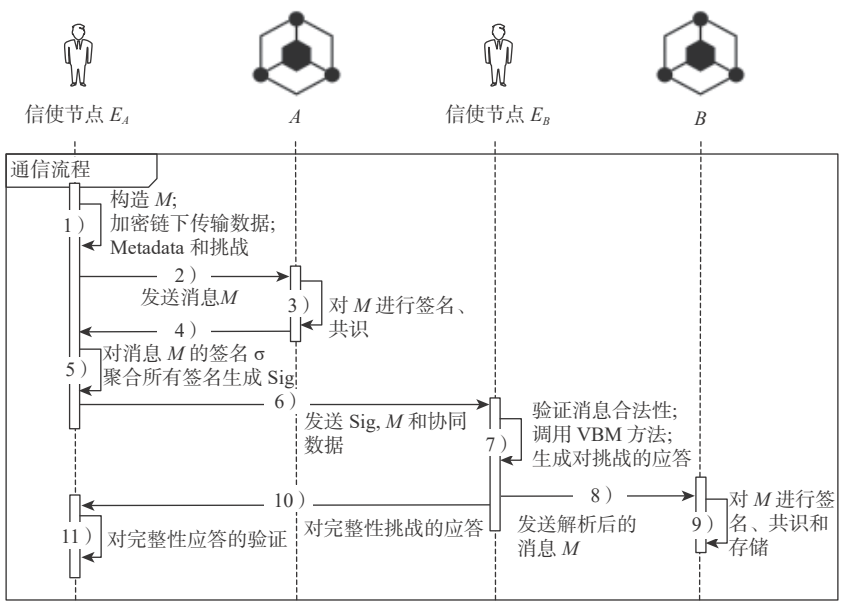


图 5　同步流程

Fig. 5　Synchronization process

3.4　验证阶段

为防止数据发送方中某信使节点伪造数据, 传输数据需经过多个信使节点确认才可保障正确性, 本文在联盟链结盟初期设置了信使节点的最低数量 MinENode。信使节点在接收到数据时需要先对消息进行验证, 包括 t、ChainID、信使节点的 Ecert 以及消息的摘要, 通过验证等式 $e(\text{Sig}, p) = e\left(H(M), \sum_{i=1}^{n} P_i\right)$ 是否相等判断聚合签名是否有效, 然后根据不同的消息类型进行不同的验证。若验证不通过, 将构造警告消息。

当验证 MemAdd 时, E_B 对消息 M 校验流程如图 6 所示。

1) E_B 检查签名消息的节点是否为新的信使节点, 若是, 执行步骤 2; 若否, 执行步骤 4。

2) E_B 检查 ChainID 下是否有归档信息, 即 EC 的存在性证明, 若有, 则证明该联盟链之前参与过同盟链的协同过程, 执行步骤 3; 若没有归档数据, 则执行步骤 4。

3) 查看消息 M 中归档信息对应的相应索引,

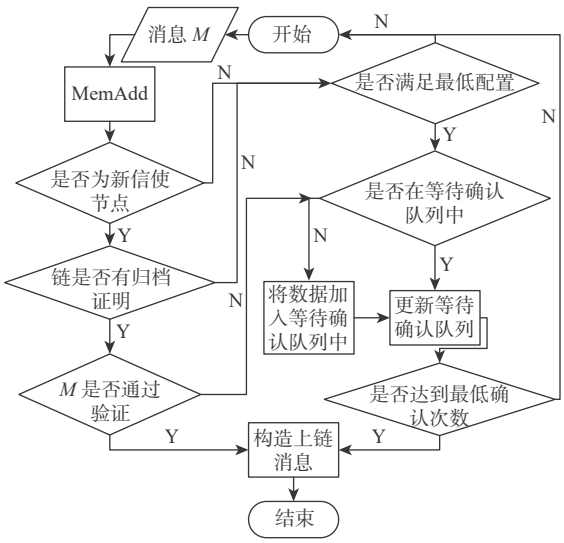


图 6 MemAdd 类型消息验证流程

Fig. 6 Validation flowchart for MemAdd type message

若找到，则执行步骤 5，若找不到则说明该归档信息有误，执行步骤 4。

4）E_B 查看 AIC 中关于联盟的最低配置，若消息 M 满足，则执行步骤 5，若不满足则丢弃 M，构造警告消息，同时继续监听新的消息。

5）E_B 查看消息 M 是否已经在等待确认的数据队列 EDataQuene 中，若无，则将消息 M 加入 EDataQuene 中并将确认次数置为 1，等待满足最低确认次数的阈值 (最低信使节点数)，若 EDataQuene 中已存在，则增加 M 的确认次数，执行步骤 6。

6）若未达到最低确认次数的阈值，等待新的信使节点传递对消息 M 的确认，重复执行步骤 1～5；若已达到最低消息阈值，则执行步骤 7。

7）E_B 对该消息 M 构造上链消息。

当验证 MemTran 类型数据时，E_B 根据 VBM 验证发送的数据是否合法、真实且连续。

当验证 MemUpdate 时，E_B 检查配置项中是否符合满足 MiniConfig，不满足，则丢舍该配置，满足，则将该配置放在 EDataQuene 中，等待满足最低确认次数。在确认最新配置之前，使用最新已被确认的配置进行通信。

当验证 MemQuit 类型数据时，E_B 收到退出消息后，对 A 保存在链上的配置信息和状态信息交易哈希做归档，并将链上有关的交易哈希索引等信息构造进 Content 中，并等待链内共识完整后传输给 E_A。E_A 在接收到其他成员传输的信息后，存储相应数据，等待 Coolper 后退出。正常退出的联盟链在退出后可以根据索引信息在相应区块链中查找历史协同数据，根据归档证明能够快速加入同盟链并证明历史协同数据的存在性。

除此之外，还需要对链下协同数据的完整性挑战应答进行验证，即验证交易路径并计算根节点的哈希值，验证通过则判断等式 3) 是否相同：

$$e(\theta,p)\stackrel{?}{=}e\left(\prod_{i=1}^{c}H(m_i)^{v_i}u^{\mu},P_{\mathrm{ID}}\right) \tag{3}$$

若流程中有不正确的数据，包括假冒信使节点、错误传输消息、错误的完整性证明、严重超时的数据确认和证明以及故意的数据不确认等，信使节点除了针对可能丢失的传输消息进行重传外，严重超时和确认的信使节点将会被放入黑名单中，超过一定次数后便不再与该节点进行通信，而假冒信使节点会被直接禁止通信。异常信息会被构造入警告信息，重传信息会被构造进 ACP 的 Content 里，并短时间内进行传输，不占用正常协同消息的时间。警告消息由 Coolper 时间来发送情况说明撤销处罚。

3.5 数据使用阶段

该阶段在 A 需要账本数据恢复或需要其他联盟链提供链上数据背书证明时进行，分为数据索取和返回两个阶段。

数据索取：E_A 构造 DataUse 类型数据 M_1。E_A 将需要请求的 A 链的数据编号发送给其他联盟链，即区块号区间<N_{Bs}、N_{Be}>和链下数据同步序号区间<N_{Es}, N_{Ee}>、N_{Bs} 和 N_{Be} 表示开始和结束的区块号、N_{Es} 和 N_{Ee} 表示开始和结束的轮次号。

数据返回：E_B 验证 M_1 后按照其中的编号将相关数据构造成 M_2 发送给 E_A。E_B 查找包含 N_{Be} 的链上背书交易，根据交易中的 BNI、N_B 进行查找包含 N_{Bs} 的背书交易 T_x，将<N_{Es}、N_{Ee}>期间所有的交易数据内容生成背书交易列表 ETxList。若还需要 ArchiveData，则根据<N_{Es}、N_{Ee}>查找相应发送轮次内的链下加密数据 ArchiveData，生成链下协同数据列表 ArcList。

E_A 验证 M_2 的合法性后，便开始进行数据的验证和比对。将 A 链的相关数据和 ETxList 进行对比，包括计算数据的完整性、验证数据合法性，若有多个联盟链则均需要进行对比，验证成功后根据 Epoch 将链下数据进行排列，最后通过对应的私钥 k_{ID} 对数据进行解密，从而完成账本数据的恢复。若无 ArchiveData，直接使用证明即可。

4 安全性分析

4.1 合谋攻击

当大部分节点进行作恶时，能够很快篡改区块链数据。总节点数 N_{node}，完整存储数据节点数为 N_{copy}，当网络中作恶节点数超过 N_{copy} 的一半时，攻击节点可以在有限时间内共同篡改数据。成功攻

击的概率 P_{Fault} 如式(4)所示:

$$P_{\text{Fault}}=\begin{bmatrix}N_1\\N_2\end{bmatrix}\times\begin{bmatrix}N_{\text{nodeFault}}\\N_2\end{bmatrix}\Big/\begin{bmatrix}N_{\text{node}}\\N_2\end{bmatrix}\tag{4}$$

式中, N_1=N_{node}-$N_{\text{nodeFault}}$, $N_{\text{nodeFault}}$ 表示作恶节点数, N_2=N_{copy}/2。在总结点数 N_{node} 为 30, N_{copy} 为 20 时, 由式(4)得出图 7, 即得出账本被篡改的概率。其中横坐标为作恶节点占总节点数的百分比, 纵坐标为篡改数据的成功概率。由图 7 可以看出, 当作恶节点数在 40% 以下时, 被篡改的概率小于 10%。节点数较少时, 作恶节点数超过 40% 的概率会很大。

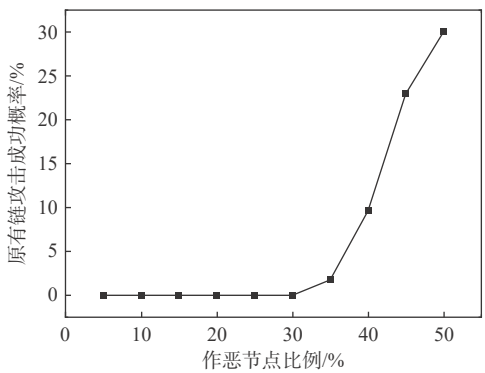


图 7　数据篡改成功概率

Fig. 7　Probability of successful data tampering

采用本文方案时, 该数据在其他联盟链上背书后, 恶意节点篡改数据需要攻破其他联盟链, 设所有联盟链配置相同, 则恶意节点成功更改的概率为

$$P_{\text{Fault}}=\prod_{i=1}^{C\text{num}}\begin{bmatrix}N_1\\N_2\end{bmatrix}\times\begin{bmatrix}N_{\text{nodeFault}}\\N_2\end{bmatrix}\Big/\begin{bmatrix}N_{\text{node}}\\N_2\end{bmatrix}\tag{5}$$

根据该公式作图 8, 得出联盟链数量影响下单个联盟链被成功攻击的概率, 其中横坐标为联盟链节点个数, 纵坐标为攻击成功概率。由此看到, 本文的方案可以有效抵御单个联盟链阶段对历史确认数据的合谋攻击, 增加一个同等配置的联盟链就可以使占比 50% 的攻击节点篡改成功概率下降至 10% 以下, 随着联盟链数量的增多, 该概率将进一步下降。

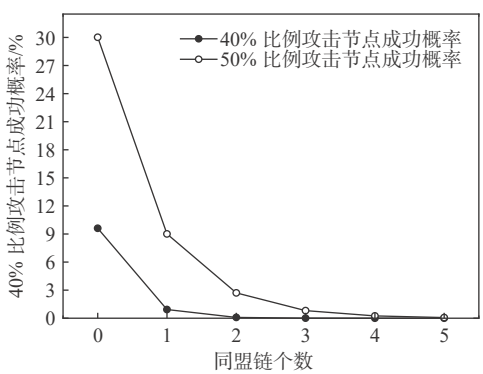


图 8　使用本方案的数据篡改成功概率

Fig. 8　Probability of successful data tampering with our scheme

4.2　中间人攻击

本方案使用信使节点作为传输数据的媒介, 为了保证双方通信时数据的安全性, 本文针对单个信使节点作恶的情况提出在加入初期联盟链之间协商最低信使节点确认次数来降低作恶的可能。通过在共识节点中选取信使节点来保证服务的可靠性。初始化时,单个节点有概率对区块链的参数进行造假。假设单个节点不诚实的概率为 δ, 那么在最低确认次数为 3 的情况下, 作恶的概率降低为 $\delta^3(0<\delta<1)$。

同时, 为了防止传输期间信使节点进行合谋, 本文要求联盟链中大多数节点对传输的消息进行签名, 并采用聚合签名对所有消息签名进行聚合以减少数据大小。即使恶意节点可以伪装所有其他签名者, 并可以对诚实签名者发起选择消息攻击, 它在任何概率多项式时间内也不可能伪造诚实签名者的有效聚合签名。因此, 信使节点单次发送消息不可伪造, 即单次消息可以代表整个联盟链。在接收到数据时, 聚合签名正确性验证过程如下:

$$e(\text{Sig},p)=e\left(H(M)\sum_{i=1}^{n}k_i,p\right)=e\left(H(M),\sum_{i=1}^{n}P_i\right)\tag{6}$$

ACP 在传输时会附带其他联盟链该轮次发送并已被确认的 EC 值, 该 EC 值可以增强其他联盟链对传输数据正确性的验证。

4.3　链下存储数据完整性欺骗

单次数据损坏或未存储的比例为 f_0, 挑选 c 块, 使用完整性验证方法时有 $1-(1-f_0)^c$ 的概率识别出协同方未完整保存数据。同时, 由于其他联盟链可能会合谋, 导致数据协同方的 ArchiveData 备份数量少于联盟链数量, 可通过使用对应 ChainID 生成的密钥对 ArchiveData 进行加密, 生成不同的数据, 使得所有联盟链存储的数据不一致, 以避免该种情况的发生。

数据协同方若不能提供正确的证据完整性挑战, 则不能通过以下等式:

$$\begin{aligned}e(\theta,p)&=e\left(\prod_{i=1}^{c}(H(m_i)u^{m_i})^{v_ik_{\text{ID}}},P\right)=\\&e\left(\prod_{i=1}^{c}H(m_i)^{v_i}\prod_{i=1}^{c}u^{v_im_i},P\right)=\\&e\left(\prod_{i=1}^{c}H(m_i)^{v_i}u^{\mu},P\right)\end{aligned}\tag{7}$$

4.4 上链背书欺骗

同盟链中联盟链之间由于不一定能够看到其他链的数据，因此在数据协同方接收数据后可能会出现不上链存储数据的情况。本文规定联盟链间需要在一定轮次内传输其他联盟链发送的历史协同数据所在的交易进行验证。通过解析交易中的内容并计算路径 MerkleRoot，该 MerkleRoot 必须存在于传输的区块头列表中的某个区块头中，因为数据传输方不太可能会伪造自己需要背书的区块列表，故通过验证后数据协同方会相信自己之前传输的数据已经在对方的区块链中进行存储。

4.5 背书数据不一致

数据发送方可能会由于节点作恶、轮换等原因造成发送数据不一致。本文采取多信使节点确认的措施，并将每一次传输的数据构造成累加器[22]，当传输新数据时，由于累加器元素能够唯一确定单次传输的数据，因此若信使节点传输的新累加器不是基于之前的累加器和当前轮次传输数据构造的，则协同方验证不会通过。本文中，由于所有联盟链都对一个累加器进行更新，采用 RSA 累加器[22]不用担心陷门泄露问题，故本文采用通用累加器来保证传输数据的真实性。

4.6 发送方链分叉导致的数据不一致

当发送方的数据出现分叉后，数据在协同方中存在待确认和已确认两种情况，已确认指协同数据已在协同方中进行上链操作。

待确认数据可以重新发送该分叉所在轮次的相关数据，发送方会对协同方的警告消息做出数据重传的回应，在 Coolper 时间后，未得到最低信使节点数量确认的数据会被协同方删除，此时链上存储的是分叉后的数据。

已确认的数据不能被更改，若发送方链分叉后原先发送的数据不在主链中，则需要发送方根据前一次的协同数据构造 EC，此时的协同方存储了主链数据和分叉数据。

5 实验与评估

本节基于以太坊实现了同盟链的原型系统，并与其他方案在存储、通信和计算开销方面进行对比。

在实验中，对以太坊近一年内的数据进行随机采样，得出平均区块大小为 S_b=83 KB，平均每个区块 124 笔交易。同时设置区块头大小为 0.5 KB，单次发送区块数量为 100，单次协同数据分块为 100，挑战块数量 c 为 20，BLS 公钥长度为 256 bit，聚合签名长度为 256 bit，节点签名长度为 256 bit，哈希摘要为 256 bit，p 为 160 bit，累加器大小为 512 bit，RSA 模长为 1 024 bit，Merkle 树路径长度约为 7 层，此时路径大小为 256 Byte。本文使用基于 BLS 的聚合签名、基于 RSA 的累加器以及其他常规密码学工具。不同场景下可变参数不同，为了减少对实验结果的干扰，本次测量将解析的函数、自定义内容项、信使节点黑名单这些可变长度项设为空或范围内的最小值。在原型系统中选取区块编号为 1 261 到 1 460 的数据进行对比和测量。

5.1 链类型对比

引言提到现有区块链类型根据准入机制分为公有链、私有链和联盟链。随着区块链商用化的普及，现有链暴露出了各自的缺点：公有链对监管不友好，私有链仅适合机构内部使用，联盟链有效节点少、节点独立性低、链可持续发展能力差。本文提出的同盟链模型介于公有链和联盟链之间，每个联盟链作为一个整体成为同盟链模型的维护者。本文通过联盟链之间的相互存储背书使所有参与结盟的区块链节点能间接维护其他联盟链账本，这样可以较好地整合联盟链和公有链的优点，更符合现有商用化落地的需求。不同链类型对比如表 2 所示。

表 2 不同区块链类型对比

Table 2 Comparison of different blockchain types

对比维度	公有链	私有链	联盟链	同盟链
准入机制	无	有	有	有
去中心化程度	高	低	中	较高
交易速度	低	高	高	高
监管难易	难	易	易	易

5.2 方案对比

提高共识节点可靠性不能从本质上提升整个区块链的可靠性，故本文不对其作对比分析。本文主要从账本篡改难度、有效节点数量增幅、节点独立程度、有无账本恢复能力、适用范围和对原有链交易性能的影响程度 6 个方面同其他方案进行对比，结果如表 3 所示。

表 3 不同方案可靠性指标对比

Table 3 Comparison of reliability indicators of different schemes

方案	账本篡改难度	有效节点数量增幅	节点独立程度	账本恢复能力	适用范围	交易体验影响程度
方案1	中	中	低	无	广	中
方案2	高	高	高	无	较广	小
方案3	高	高	高	有	窄	大
本方案	高	高	高	有	广	小

方案 1：直接增加节点数量，能够增大账本篡改难度，若增加的节点之间存在一定行为关联，即不

完全独立, 则有效节点数低于新增节点数。一般而言, 节点数目增多, 区块链的维护难度上升、交易速度下降。该方案在遭受攻击或意外时主要信赖于使用区块链机构自身恢复措施的强度, 本身并不能主动恢复账本数据。

方案 2[11]: 锚定模式, 通过在公有链中发送上链交易存储数据, 借助公有链所有的维护节点共同抵御账本篡改类型的攻击。因该方案将数据存储在公有链中需要手续费, 成本随着时间呈线性增长, 所以有可能存在交易拥堵等情况导致系统可用性显著下降。该方法需要节点加入公有链发起上链交易, 故上链数据真实性依赖于该节点, 该节点是否作恶直接影响原有链的去中心化程度和安全性。该方案只能判断公有链中锚定的数据是否被更改, 无法恢复数据, 后面将不对其进行定量对比分析。

方案 3[13]: 多链协同模式, 利用传统跨链思想进行数据交互。当前跨链技术主要面向不同区块链资产之间的交易兑换, 需要借助可信第三方监听原有链来完成数据交换。Optimistic rollup 考虑了数据可用性, 但需要维护其他链, 且区块链间能够相互访问, 适用范围有限。同时, 该方案等待期较长, 对交易体验影响较大, 若等待期内无验证者提交欺诈证明, 则该数据就会被确认存在安全风险。

本方案: 通过设计较为完备的通信机制和验证措施使参与结盟的区块链存储数据形成层叠存储, 此时若想改动已经协同过的区块链需要所有联盟链的同意, 从而区块链中的各节点能够间接参与到其他区块链数据的维护工作中, 形成区块链间有效节点数的共享。同时通过协同存储进行数据灾备, 最终提升原有链的可靠性。相比方案 1, 本方案能够更好地保证增强节点数的独立性, 不大幅降低原有区块链的交易速度; 相比方案 2, 本方案对监管更友好, 节约成本; 相比方案 3, 本方案不需要强制要求区块链之间的业务关联性和资产透明性, 数据交互时也不需要锁定资产, 对联盟链来说, 没有第三方节点监听和介入原有区块链, 适用范围更广。

5.3　存储消耗

本文为了增强数据的可用性, 采用了链下存储的方式, 但方案 1 与方案 3 的方式都是链上存储, 为了较为全面地衡量不同方案的存储消耗, 将从节点存储开销(上链数据)和区块链存储总量(总共增加的存储量)两个方面进行对比。

1) 节点存储开销

方案 1 仅增加节点, 不增加单节点存储, 故本文仅和方案 3 进行比较。一个轮次中使用不同方案单节点的存储量增幅如表 4 所示, 能够看出, 在每个联盟链有效个数为 6、不同有效节点个数的情况下, 本方案单节点存储开销增幅较小, 优于方案 3, 但劣于方案 1, 相对存储总量和其他成本来说可以忽略不计。

表 4　单节点存储量增幅

Table 4　Increment of single node storage　%

有效节点个数	方案3	本方案
6	0	0
12	53.413 6	0.002 4
18	112.137 2	0.004 8
36	275.985 1	0.120 5
60	495.254 3	0.216 9

2) 区块链存储总量

本方案中, 节点在链下需要存储的数据主要为 AIC 和 ArchiveData。此时计算 3 个联盟链联合情况下的存储, 每条联盟链成员为 6, 最小信使节点数量为 2, 时间戳、成员链 ID 等数字占用 32 bit, ECerList 存储个数为最低信使节点数量加 1, AIC 所占大小为 1.48 KB。N_{Num} 为联盟链数量, 第 i 个联盟链一个轮次内区块数量为 $N_{\mathrm{BNI},i}$, 则压缩后加密数据传输大小总量 S_{c} 约为

$$S_{\mathrm{c}}=\sum_{i=1}^{N_{\mathrm{Num}}}N_{\mathrm{BNI},i}\times S_{\mathrm{b}}\approx 8.92\ \mathrm{MB} \tag{8}$$

在叶子节点数为 100 时, M_{Tree} 大小为 $S_{M_{\mathrm{Tree}}}=\mathrm{Total}(m)\times S_{\mathrm{h}}=(m^0+\cdots+m^{\log_m n})S_{\mathrm{h}}=7.10$ KB, 元数据的存储量为 $S_{\mathrm{mdata}}=n\times S_{\mathrm{h}}=3.13$ KB。

方案 1 中新增节点全量存储区块链数据, 故随节点数量的增加区块链存储总量增多。方案 3 没有增加原有链节点数量, 但所有节点需要存储其他链压缩后的数据, 在增加同等有效节点个数的情况下其存储总量低于方案 1。本方案上链数据最少, 仅为协同数据的唯一标识及其定位数据, 但该协同数据中的 ArchiveData 需要进行链下存储。如图 9 所示, 本方案的存储总量在 3 个方案中最低。

5.4　通信消耗

本节量化本方案联盟链间的通信消耗并和其他方案进行对比。本方案中主要的消耗为不同类型下的 ACP 和 ArchiveData 的传输、辅助信息如对 M 的聚合签名、哈希摘要等。其中, 辅助信息数据通信大小为 I_1=0.09 KB, MemAdd 和 MemUpdate 类型的 ACP 大小为 I_2=0.32 KB。MemTran 类型的 ACP 大小为 I_3=50.34 KB。MemTran 类型的 ArchiveData 数据仅传输加密区块链数据和验证数据,即 I_4=117.42 KB 和 I_5=0.35 KB。DataUse 类型是 MemTran

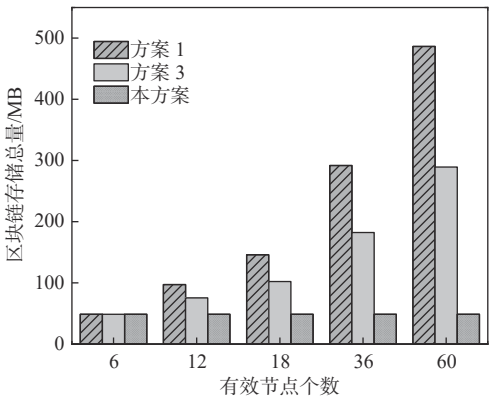


图 9 区块链一个轮次的存储总量

Fig. 9 The total storage capacity of the blockchain in single interval

类型的逆过程，数据返回的过程和 MemTran 通信消耗相同，即 I_3 或 I_3+I_4，数据索取时额外的索引通信消耗为 I_6=0.07 KB。MemQuit 类型消息在 Content 中仅有 EC 消息，但需要发送方和其他联盟链进行交互，此处次数为 3，通信量为 I_7=0.34 KB。

通信消耗和节点数量、共识信息大小、共识信息数量呈正比。一般来说，每个区块内的每笔交易都需要进行全网广播，P2P 网络通信复杂度随着节点数量的增加而上升。方案 1 中增加的节点需要参与每一笔交易的共识，在 100 个区块，每个区块 124 笔交易下共识信息数量显著增多，其额外产生的通信消耗远高于其他方案，故本文仅与方案 3 进行对比。方案 3 不增加原有链节点数量，聚合压缩多笔交易后在其他链上发起交易并进行存储，通信消耗主要表现为原有链和存储方所有节点对交易的共识过程。实验结果如图 10 所示，可以看出，本方案增加的通信消耗在 3 个方案中最少。

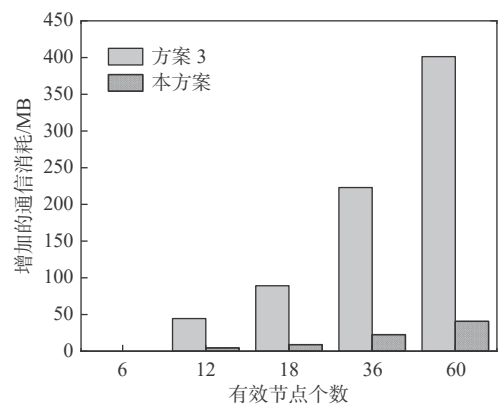


图 10 一个轮次内增加的通信消耗

Fig. 10 Increased communication consumption in single interval

5.5 计算消耗

传输数据的构造为本方案中最耗时的操作，主要分为链下协同数据的加密和数据完整性验证操作中辅助数据的构造。为了保障安全性，本文采用 RSA 算法进行加密，并通过多线程异步加密的方式将加密时间缩短至 327 ms。从单节点和整个区块链两个维度对比不同方案所增加的计算消耗。对单节点来说，方案 1 不增加耗时；方案 3 中，聚合节点压缩数据需耗时 0.2 s 左右，其他节点验证该聚合交易需额外耗时 0.008 s 左右；本方案中，信使节点进行同盟链各项事务需耗时 1.4 s 左右的时间，其他节点签名和验证需花费 0.03 s 左右，信使节点操作和原有链事务异步进行，不停顿原有链。

方案 1 中所有新增节点都要参与交易和区块的共识，在本文所设参数下会产生非常大的计算消耗。方案 3 的计算消耗来源于数据的压缩和共识；本方案的计算消耗主要为传输数据的构造和验证。一个轮次内不同方案给整个区块链带来的计算时间消耗总量如表 5 所示。可以看出，方案 3 的计算消耗总量最少，本方案的计算消耗总量略高于方案 3，但远低于方案 1。

表 5 计算开销额外总耗时

Table 5 Additional total elapsed time spent on computing overhead s

有效节点个数	方案1额外总耗时	方案3额外总耗时	本方案额外总耗时
6	0	0	0
12	117.541 1	0.246 6	1.531 6
18	209.253 1	0.254 4	1.566 0
36	486.230 0	0.277 8	2.835 7
60	855.912 7	0.309 0	3.857 1

6 结 论

联盟链是我国区块链落地的主要形态，但存在有效节点数少、节点独立性低以及可持续发展能力弱的问题，导致联盟链的去中心化程度低、可靠性不足，在运行发展过程中可能出现历史数据被合谋篡改、数据丢失后难以恢复的风险和隐患。对于这些问题，本文贡献如下：

1）提出联盟链之间相互存储其他结盟区块链数据的唯一标识形成数据之间的层叠关系，并协同存储加密区块链数据保证账本的可恢复性，增强系统的可靠性和去中心化程度。

2）设计通信机制使得不同联盟链之间有统一的传输和验证方式，保障不同区块链之间的通信，对方案运行之中会出现的消息类型给出相应构造和验证方案。

3）分析同盟链的安全性问题，证明同盟链能够抵御中间人的攻击，降低区块链对历史数据合谋攻

击的概率, 同时通过完整性验证、密码累加器和延迟确认解决方法中存在的其他问题。

4) 通过分析比对其他方案得出, 同盟链能够提高原有链去中心化程度、有效节点数量、节点间独立性、账本篡改难度, 同时对交易性能影响程度低, 适用范围比较广, 在存储链下协同数据时能够反向恢复数据对应的区块链。

5) 通过实验对比证明本方案在区块链存储总量和通信开销上有优势, 在计算消耗上略有劣势, 但同盟链模型不会阻塞原有链业务, 也不会停顿原有链。

本文所提方法仍有不足之处: ①需要多数节点诚实且可信为前提; ②对信使节点要求较高; ③不同链类型需要开发不同的交易解析工具。同时, 本文仅对同构区块链进行了实验, 所提方法仅对单个同盟链模型进行了分析, 对于多个同盟链模型交叠情况没有过多考虑。下一步将针对以上问题继续开展研究。

参考文献 (References)

[1] ZHENG Z B, XIE S A, DAI H N, et al. An overview of blockchain technology: architecture, consensus, and future trends[C]// 2017 IEEE International Congress on Big Data (BigData Congress). Piscataway: IEEE Press, 2017: 557-564.

[2] CASTRO M, LISKOV B. Practical Byzantine fault tolerance and proactive recovery[J]. ACM Transactions on Computer Systems, 2002, 20(4): 398-461.

[3] ZHU X Y, BADR Y. Identity management systems for the Internet of Things: a survey towards blockchain solutions[J]. Sensors, 2018, 18(12): 4215.

[4] 中国电子信息产业发展研究院. 2021 年中国区块链年度发展白皮书[R/OL]. (2022-06-14)[2022-07-24]. https://dsj.guizhou.gov.cn/xwzx/gnyw/202206/t20220614_74881576.html.

[5] 武腾, 薛磊, 郑东, 等. P2P 持久存储系统可靠性分析与数据维护优化[J]. 信息安全与通信保密, 2009, 7(8): 149-153.
WU T, XUE L, ZHENG D, et al. Reliability analysis and data maintenance optimization of P2P durable storage system[J]. Information Security and Communications Privacy, 2009, 7(8): 149-153 (in Chinese).

[6] GUO H Q, YU X J. A survey on blockchain technology and its security[J]. Blockchain: Research and Applications, 2022, 3(2): 100067.

[7] BANDARA H D, XU X W, WEBER I. Patterns for blockchain data migration[C]// Proceedings of the European Conference on Pattern Languages of Programs 2020. New York: ACM, 2020: 1-19.

[8] LEI K, ZHANG Q C, XU L M, et al. Reputation-based Byzantine fault-tolerance for consortium blockchain[C]// 2018 IEEE 24th International Conference on Parallel and Distributed Systems (ICPADS). Piscataway: IEEE Press, 2018: 604-611.

[9] YAN C Y, ZHANG C, LU Z G, et al. Blockchain abnormal behavior awareness methods: a survey[J]. Cybersecurity, 2022, 5(1): 5.

[10] ARÁOZ M. Proof of existence[R/OL]. [2022-07-24]. http://docs.proofofexistence.com.

[11] 谭海波, 周桐, 赵赫, 等. 基于区块链的档案数据保护与共享方法[J]. 软件学报, 2019, 30(9): 2620-2635.
TAN H B, ZHOU T, ZHAO H, et al. Archival data protection and sharing method based on blockchain[J]. Journal of Software, 2019, 30(9): 2620-2635 (in Chinese).

[12] HOPE-BAILIE A, THOMAS S. Interledger: creating a standard for payments[C]//Proceedings of the 25th International Conference Companion on World Wide Web-WWW '16 Companion. New York: ACM, 2016: 281-282.

[13] 叶少杰, 汪小益, 徐才巢, 等. BitXHub: 基于侧链中继的异构区块链互操作平台[J]. 计算机科学, 2020, 47(6): 294-302.
YE S J, WANG X Y, XU C C, et al. BitXHub: side-relay chain based heterogeneous blockchain interoperable platform[J]. Computer Science, 2020, 47(6): 294-302 (in Chinese).

[14] FLOERSCH K. Ethereum smart contracts in L2: optimistic rollup[R/OL]. (2019-08-28)[2022-11-30]. https://medium.com/plasma-group/ethereum-smart-contracts-in-l2-optimistic-rollup-2c1cef2ec537.

[15] BUTERIN V. Chain interoperability[R/OL]. R3 research paper, 2016 (2016-09-09)[2022-07-24]. https://allquantor.at/blockchainbib/pdf/vitalik2016chain.pdf.

[16] VIGIL M, BUCHMANN J, CABARCAS D, et al. Integrity, authenticity, non-repudiation, and proof of existence for long-term archiving: a survey[J]. Computers & Security, 2015, 50: 16-32.

[17] LAKHWANI K, KAUR R, KUMAR P, et al. An extensive survey on data authentication schemes in cloud computing[C]// 2018 4th International Conference on Computing Sciences (ICCS). Piscataway: IEEE Press, 2018: 59-66.

[18] MERKLE R C. A certified digital signature[M]// Advances in Cryptology — CRYPTO' 89 Proceedings. New York: Springer New York, 2007: 218-238.

[19] BENALOH J, DE MARE M. One-way accumulators: a decentralized alternative to digital signatures[M]// Advances in cryptology — EUROCRYPT '93. Berlin: Springer Berlin Heidelberg, 1994: 274-285.

[20] ATENIESE G, DI PIETRO R, MANCINI L V, et al. Scalable and efficient provable data possession[C]// Proceedings of the 4th International Conference on Security and Privacy in Communication Netowrks. New York: ACM, 2008: 1-10.

[21] Zlib home site[EB/OL]. (1995-05-01)[2022-03-28]. http://www.zlib.net.

[22] 祁健. 累加器在区块链中的应用研究[D]. 南京: 南京信息工程大学, 2020: 22-33.
QI J. Research on the application of accumulator in blockchain[D]. Nanjing: Nanjing University of Information Science & Technology, 2020: 22-33 (in Chinese).

A consortium chain improvement model based on multi-chain collaboration

ZHAO Xianxian[1, 2], TAN Haibo[1, 2], ZHAO He[1, *], ZHOU Tong[3], CHENG Haotian[1, 2], LI Jinze[1, 2]

(1. Hefei Institutes of Physical Science, Chinese Academy of Sciences, Hefei 230031, China;
2. University of Science & Technology of China, Hefei 230026, China;
3. Anhui ZhongKeJingGe Technology Co., Ltd., Hefei 230088, China)

Abstract: In recent years, blockchain technology has been developing rapidly, and It plays a critical role in fintech, supply chain, medical health, data security and other fields. Due to friendly supervision and excellent performance, the consortium chain has become China's first choice for government and commercial blockchain. However, it is usually faced with problems such as the small number of effective nodes, low node independence, and weak sustainable development ability, which reduces the system's reliability. We suggest the alliance model, a blockchain approach that coordinates several consortium chains to increase system stability, in light of the aforementioned issues. The model stores data by overlapping each other between chains to reduce the possibility of data tampering and ensure data recovery. In addition, we use the methods of aggregated signature and data integrity verification to solve the problems of transmission data integrity and source authenticity. In order to verify the legitimacy of reciprocal endorsement of collaborative data and to check the continuation of the sent data, we also develop an existence verification method. To confirm its efficacy, this scheme's communication, storage, and computational usage are compared to those of the current techniques.

Keywords: blockchain; consortium chain; reliability; decentralization; data integrity; data authenticity

Received: 2022-09-29; **Accepted**: 2022-11-25; **Published Online**: 2023-03-09 16:07
URL: link.cnki.net/urlid/11.2625.V.20230309.1406.007
Foundation item: National Key Research and Development Program of China (2021YFB2700800)
* **Corresponding author**. E-mail: zhaoh@hfcas.ac.cn

《北京航空航天大学学报》征稿简则

《北京航空航天大学学报》是北京航空航天大学主办的以航空航天科学技术为特色的综合性自然科学学术期刊(月刊)。本刊以反映航空航天领域研究成果与动态、促进学术交流、培养科技人才和推动科技成果向社会生产力转化为办刊宗旨。本刊为中国自然科学技术核心期刊,并被 Ei Compendex 等国内外权威文献检索系统收录。本刊向国内外公开发行,为进一步提高办刊质量和所刊出文章的学术水平,特制定本简则。

1 论文作者及内容

1.1 本刊面向海内外所有学者。

1.2 主要刊载与航空航天科学技术有关的材料科学及工程、飞行器设计与制造、宇航科学与工程、信息与电子技术、控制技术和自动化工程、流体力学和动力工程、计算机科学及应用技术、可靠性工程与失效分析等领域的研究文章。航空航天科学技术民用方面以及具有航空航天工程背景的应用数学、应用物理、应用力学和工程管理等方面的文章也在本刊优先考虑之列。

2 来稿要求

2.1 论文应具有创新性、科学性、学术性和可读性。

2.2 论文为原创作品,尚未公开发表过,并且不涉及泄密问题。若发生侵权或泄密问题,一切责任由作者承担。

2.3 主题明确,数据可靠,图表清晰,逻辑严谨,文字精练,公式符号标准规范,标点符号正确。

2.4 文稿撰写顺序:中文题名(一般不超过 20 个汉字),作者中文姓名、单位、所在城市、邮政编码(作者顺序始终不得更改),中文摘要(包括目的、方法、结果及结论),中文关键词(5 ~ 8 个),中图分类号,引言,正文,参考文献,英文题名,作者英文姓名、单位、所在城市、邮政编码、国别,英文摘要,英文关键词。首、末页下角注明基金项目名称及编号,作者信息。

2.5 作者请登录本刊网页进行在线投稿。

3 稿件的审核、录用与版权

3.1 来稿须经专家两审和主编、编委讨论后决定刊用与否。

3.2 若来稿经审查后认定不宜在本刊发表,将及时告知作者。如果在投稿满 3 个月后仍未收到本刊任何通知,作者有权改投其他刊。在此之前,请勿一稿多投,否则一切后果自负。

3.3 来稿一经作者签字并在本刊刊出,即表明所有作者都已经认可其版权转至本刊编辑部。本刊在与国内外文献数据库或检索系统进行交流及合作时,不再征询作者意见。

邮寄地址:100191 北京市海淀区学院路 37 号 《北京航空航天大学学报》编辑部

办公地点:北京航空航天大学办公楼 405,407 房间

电　　话:(010)82316698,82316981,82315499,82317448

E-mail: jbuaa@ buaa. edu. cn

http://bhxb. buaa. edu. cn

http://www. buaa. edu. cn

《北京航空航天大学学报》
第五届编辑委员会

主　任（主　编）：赵沁平

（以下按姓氏笔画为序）

副主任（副主编）：丁希仑　王少萍　孙志梅　李秋实　李焕喜　杨嘉陵
苗俊刚　相　艳　徐立军　钱德沛　曹晋滨

编　委：马殿富　王　琪　王　聪　邓小燕　王青云　王荣明　刘　宇
刘　红　江　洁　刘　强　闫　鹏　朱天乐　刘铁钢　齐铂金
陈万春　邹正平　苏东林　杨世春　沈成平　邱志平　宋知人
杨树斌　张晓林　杨晓奕　杨继萍　李惠峰　吴新开　张瑞丰
杨照华　宋凝芳　周　锐　林宇震　林贵平　战　强　姚仰平
胡庆雷　赵秋红　段海滨　赵巍胜　席　平　郭　宏　徐　洁
徐世杰　郭洪波　康　锐　翟　锦　熊华钢

北京航空航天大学学报

Beijing Hangkong Hangtian Daxue Xuebao

（原《北京航空学院学报》）

（月刊　1956年创刊）

第50卷 第10期 2024年10月

主管单位　中华人民共和国工业和信息化部
主办单位　北京航空航天大学
主　　编　赵沁平
编辑出版　《北京航空航天大学学报》编辑部
邮　　编　100191
地　　址　北京市海淀区学院路37号
印　　刷　北京科信印刷有限公司
发　　行　《北京航空航天大学学报》编辑部
发行范围　国内外发行
联系电话　(010) 82316698 82316981
82315499 82317448
电子信箱　jbuaa@ buaa.edu.cn

JOURNAL OF BEIJING UNIVERSITY OF AERONAUTICS AND ASTRONAUTICS (JBUAA)

(Monthly, Started in 1956)

Vol.50 No.10 October 2024

Administrated by　Ministry of Industry and Information Technology of the People's Republic of China
Sponsored by　Beijing University of Aeronautics and Astronautics (BUAA)
(Beijing 100191, P. R. China)
Chief Editor　ZHAO Qinping
Edited and Published by　Editorial Board of JBUAA
Printed by　Beijing Kexin Printing Co., Ltd.
Distributed by　Editorial Board of JBUAA
Telephone　(010) 82316698 82316981
82315499 82317448
E-mail　jbuaa@buaa.edu.cn
http://bhxb.buaa.edu.cn

中国标准连续出版物号：ISSN 1001-5965 / CN 11-2625/V

ISSN 1001-5965

9 771001 596243 10>

国内定价：50.00元／期

微信公众号　超星电子书　QQ群