

DOI:10.16356/j.2097-6771.2026.03.015

## 考虑维修干扰的涡轴发动机气路性能预测

邱朝阳<sup>1</sup>, 蔡景<sup>1</sup>, 陈颖<sup>2</sup>

(1.南京航空航天大学民航学院,南京 211106;2.中国人民解放军 93145 部队,上海 200231)

**摘要:**针对涡轴发动机性能建模过程中存在的工况复杂、稳态与瞬态数据难以有效区分、维修活动干扰性能退化建模等关键技术难题,提出一种基于试车数据的涡轴发动机气路性能预测方法。首先,采用 Pearson 相关性分析对试车数据进行降维处理,剔除冗余信息以提升建模效率;其次,通过构建滑动窗口结合变异系数阈值判定准则,实现试车数据中稳态数据的精准提取,解决稳态与瞬态数据混淆的问题;然后,基于相对退化量分析识别数据中维修活动的发生时间点,构建性能退化分段建模策略,有效排除维修活动对退化建模的干扰;最后,利用随机森林算法拟合发动机气路参数间的非线性多耦合关系,完成涡轴发动机气路性能的精准预测。为验证所提方法的有效性,采用 3 段试车数据进行重复验证,结果表明,各预测指标的平均相对误差均小于 1.09%。该方法可有效实现涡轴发动机气路性能退化趋势的精准预测,为发动机故障诊断、健康状态监测及性能预测等工程实践提供可靠的技术支撑。

**关键词:**涡轴发动机;气路性能;数据驱动;稳态识别;随机森林

中图分类号:V263

文献标志码:A

文章编号:1005-2615(2026)03-0615-12

## Prediction for Gas Path Performance of Turboshaft Engines Considering Maintenance Interference

QIU Zhaoyang<sup>1</sup>, CAI Jing<sup>1</sup>, CHEN Ying<sup>2</sup>

(College of Civil Aviation, Nanjing University of Aeronautics & Astronautics, Nanjing 211106, China; 2. Unit 93145 of the Chinese People's Liberation Army, Shanghai 200231, China)

**Abstract:** To address the key technical challenges in turboshaft engine performance modeling, such as complex operating conditions, difficulty in effectively distinguishing steady-state and transient data, and interference of maintenance activities on degradation modeling, a gas path performance prediction method for turboshaft engines based on test run data is proposed. First, Pearson correlation analysis is adopted to reduce the dimensionality of test run data, eliminating redundant information to improve modeling efficiency. Second, a sliding window combined with a coefficient of variation threshold criterion is constructed to accurately extract steady-state data from test run data, solving the problem of confusion between steady-state and transient data. Third, based on relative degradation analysis, the occurrence time of maintenance activities in the data is identified, and a segmented modeling strategy for performance degradation is established to effectively eliminate the interference of maintenance activities on degradation modeling. Finally, the random forest algorithm is used to fit the nonlinear multi-coupling relationship between engine gas path parameters, so as to achieve accurate prediction of turboshaft engine gas path performance. To verify the effectiveness of the proposed method, three segments of test run data are used for repeated verification.

**基金项目:**航空科学基金(2024L042052001)。

**收稿日期:**2025-12-16;**修订日期:**2026-03-20

**通信作者:**蔡景,男,副教授,E-mail:caijing@nuaa.edu.cn。

**引用格式:**邱朝阳,蔡景,陈颖.考虑维修干扰的涡轴发动机气路性能预测[J].南京航空航天大学学报(自然科学版),2026,58(3):615-626. QIU Zhaoyang, CAI Jing, CHEN Ying. Prediction for gas path performance of turboshaft engines considering maintenance interference[J]. Journal of Nanjing University of Aeronautics & Astronautics(Natural Science Edition),2026,58(3):615-626.

The results show that the average relative error of each prediction index is less than 1.09%. This method can effectively realize the accurate prediction of the degradation trend of turboshaft engine gas path performance, and provide reliable technical support for engineering practices such as engine fault diagnosis, health condition monitoring and performance prediction.

**Key words:** turboshaft engine; gas path performance; data-driven; steady-state identification; random forest (RF)

涡轴航空发动机作为直升机等旋转翼飞行器的核心动力源,其气路性能对整个发动机的性能有着至关重要的影响。气路性能是指发动机进气、压缩、燃烧、膨胀和排气的整个过程中,气流的流动状态、压比、温度、流量等一系列参数的综合表现。这些参数直接影响发动机的推力、效率、燃油消耗率、寿命等关键性能指标,进而关乎飞行器的飞行性能与运行安全。然而,航空发动机是结构复杂的气动热力学系统,通常工作在高温、高压、高速、强振动等恶劣的环境下<sup>[1]</sup>。随着飞行循环数的增加,发动机部件磨损和外来撞击等因素会不可避免地造成发动机零部件和性能的退化<sup>[2]</sup>,从而对飞行安全产生威胁。此外,气路故障是造成空中停车类事故的主要原因,其产生的维护费用也最高<sup>[3]</sup>。因此,开展航空发动机的气路性能研究具有重要的工程意义与实用价值。

对航空发动机的气路性能进行评估的方法主要分为两类,分别是基于模型的方法和基于数据的方法<sup>[4]</sup>。基于模型的评估方法是通过建立系统的物理模型或者数学模型对其状态进行评估的一种方法。该方法基于发动机的热力学原理,将发动机的工作过程通过一系列非线性方程来描述。这些方程反映了发动机内部各个部件(如压气机、燃烧室、涡轮等)的物理和热力学特性。基于模型的评估方法的评估结果可信度高,可解释性强,但建模过程比较困难,尤其面对航空发动机这类复杂系统,模型的验证也比较困难<sup>[5]</sup>。

与基于模型的方法不同,基于数据的评估方法是一种基于大量发动机试验数据的建模方法。通过收集并分析发动机在不同工作条件下的实际运行数据,采用数据拟合技术来建立一个“黑箱”模型<sup>[6]</sup>。这种模型不依赖于发动机内部的物理原理,而是通过数据间的关系来反映发动机的性能特征<sup>[7]</sup>。该方法的优势在于,它能够准确反映发动机在复杂、非线性环境下的实际表现,特别是在无法获取详细物理模型的情况下非常有效<sup>[8]</sup>。通过对发动机试验数据的训练和建模,该方法能够有效捕捉发动机的气路性能变化,为故障诊断、健康监测和性能预测等任务提供支持。

基于数据的气路性能评估是当前研究最为广泛、应用最为成熟的性能评估方法。近年来人工神经网络(Artificial neural network, ANN)<sup>[9]</sup>、卷积神

经网络(Convolutional neural network, CNN)<sup>[10-12]</sup>、深度置信网络(Deep belief network, DBN)<sup>[13-14]</sup>、支持向量机(Support vector machine, SVM)<sup>[15]</sup>等被广泛用于发动机气路系统的性能评估,并取得良好的效果<sup>[16]</sup>。

然而,目前涡轴发动机性能建模仍面临以下问题:(1)GJB 242A—2018第3.1.1节定义了数十种发动机工作状态,试车数据也显示单次试车包含多达10个以上的工况切换,数据工况复杂度显著高于单工况测试场景;(2)工况切换期间瞬态数据与稳态数据混杂,影响建模精度;(3)发动机清洗等维修活动能显著恢复气路性能,若忽略维修活动而将全段数据视为单一退化过程,会导致退化模型出现明显偏差。

针对上述问题,提出一种基于试车数据的涡轴发动机气路性能预测方法:首先通过Pearson相关性分析对试车数据进行数据降维;其次通过构建滑动窗口和变异系数阈值判定对试车数据进行稳态提取;然后通过相对退化量分析识别数据中存在的维修活动时间点,从而实现性能退化分段建模,排除维修活动干扰;最后利用随机森林算法对发动机气路参数之间的非线性和多耦合关系进行建模,最终实现对涡轴发动机气路性能的预测。该方法能够对发动机气路性能的退化情况进行有效预测,为故障诊断、健康监测和性能预测等任务提供支持。

## 1 数据分析

### 1.1 试车数据结构

使用的数据集是某型涡轴发动机地面试车台试车数据,试车依据GJB 242A—2018标准进行试验,时间跨度从2024-06-18~2025-05-24,共计进行了540次试车任务,产生了540个试车数据文件。执行试车任务时,在试验台架上加装了高精度传感器,采样频率为1 000 Hz,用于实时监测发动机的各种气路参数,收集到的有效气路性能参数如表1所示。

### 1.2 试车数据参数Pearson相关性分析

如表1所示,试车程序采集的气路参数种类众多,需要从中挑选合适参数表征发动机气路性能。根据GJB 242A—2018文件描述,表征发动机气路性能参数主要包含:功率/当量功率、空气流量、耗油率、发动机转速、 $t_{4.5}$ 温度以及其他参数,具体如表2所示。

表 1 试车数据包含参数

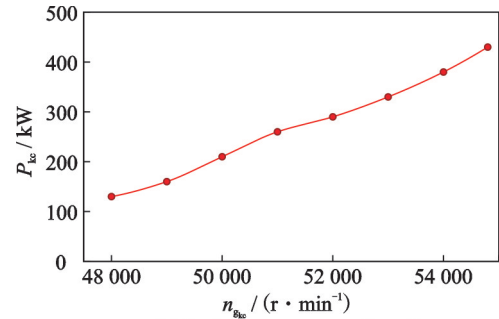
符号	中文名称	单位
$t_1$	进气温度	°C
$p_1$	进气压力	Pa
$n_g$	高压轴转速	r/min
$n_p$	低压轴转速	r/min
$n_s$	输出轴转速	r/min
$t_{3.4}$	燃气涡轮前温度	°C
$t_{4.5}$	自由涡轮进口总温	°C
$t_9$	排气温度	°C
$T_{dn}$	输出扭矩	N·m
$P$	轴功率	kW
$W_f$	燃油流量	kg/h
$s_{fc}$	燃油消耗率	g/(kW·h)
$\Pi_{c.1}$	压比	—
$W_a$	空气流量	kg/s
$P_L$	滑油压力	MPa
$t_{1.1}$	滑油进油温度	°C
$t_{1.2}$	滑油回油温度	°C

表 2 影响发动机气路性能的关键参数

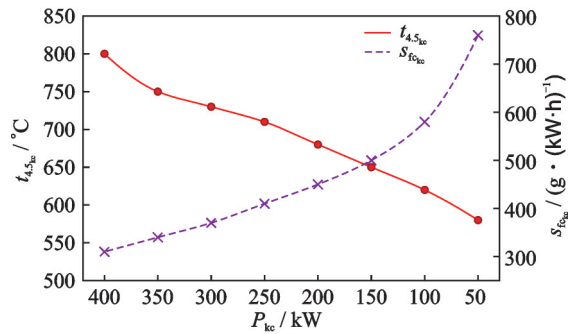
Table 2 Key parameters affecting engine gas path performance

关键参数	参数作用
高压轴转速、 低压轴转速	反映发动机的转速和工作状态,高速运转代表高负荷工作状态
排气温度、燃气涡轮 出口温度、燃气涡轮 进口温度	反映了燃烧室和涡轮的工作效率和功率输出,温度高低对应不同的工况
输出扭矩与输出转速	直接与发动机的输出功率相关,扭矩与转速越高,表示发动机处于高负荷状态
轴功率与耗油量	耗油量和功率直接反映发动机在不同状态下的燃油效率和功率输出
空气流量	进气流量影响压气机的工作效率,在应急状态和全功率状态下流量增大
振动参数(高压转子 振动、低压转子振动、 振动总量)	在高负荷状态下,振动可能增大,反映了机械结构的工作状态
功率/当量功率	反映发动机的实际输出功率与预期设计功率的匹配程度
$t_{4.5}$ 温度	反映涡轮出口温度,表示燃烧室燃烧效率和发动机运行状态

而在工程实践上,通常采用两种模式来全面表征和评价发动机气路性能,分别是在固定转速下评估功率及相关参数,以及在固定输出功率下评估燃油消耗率与涡轮后温度。这两种模式被广泛应用于发动机研制、试验和使用过程中的性能评估。例如如图 1 为试车试验中定期开展的性能评估试验结果,其采用的评估方法即为上述两种模式。因此,结合 GJB 标准要求和工程实践,初步选择高压轴转速  $n_g$ 、轴功率  $P$ 、自由涡轮进口总温  $t_{4.5}$  和燃油消耗率  $s_{fc}$  四个参数来表征涡轴发动机气路性能。



(a) Performance curve of  $n_{gk} - P_{kc}$



(b) Performance curves of  $P_{kc} - t_{4.5}$  and  $s_{fc}$

图 1 某型涡轴发动机性能评估曲线

Fig.1 Performance evaluation curves of a certain turboshaft engine

为了验证上述参数选择的合理性,对表 1 所示的 17 个参数进行 Pearson 相关性分析。

分析结果如图 2 所示。从热力图来看,  $P$  与输出扭矩  $T_{dn}$  的相关系数高达 1.000, 与燃油流量  $W_f$ 、压比  $\Pi_{c.1}$ 、空气流量  $W_a$ 、高压轴转速  $n_g$  的相关系数分别为 0.997、0.993、0.991、0.981。上述参数分别对应发动机扭矩输出、燃油供给、压缩做功、进气流量等关键气动热力学环节,全面表征发动机功率输出状态;  $P$  与各参数的高度相关性,决定其可作为发动机输出能力的核心代表指标,无需额外引入  $T_{dn}$ 、 $W_f$  等冗余参数,简化建模与分析过程。

$t_{4.5}$  与燃气涡轮前温度  $t_{3.4}$  的相关系数为 0.992, 与排气温度  $t_9$  为 0.956, 与  $T_{dn}$ 、 $P$  均为 0.991。这说明  $t_{4.5}$  能够综合反映燃烧室出口热状态与涡轮级间热力过程,是表征发动机燃气热状态的最直接指标,且其与  $t_{3.4}$ 、 $t_9$  的高相关性证明以  $t_{4.5}$  代替多个温度参数具有充分的代表性。

$s_{fc}$  与  $n_g$  的相关系数为 -0.922, 与  $P$  为 -0.847, 与  $W_f$  为 -0.853, 与  $t_{4.5}$  为 -0.857。  $s_{fc}$  与其他参数均呈显著负相关,是唯一从经济性维度表征发动机气路效率的指标。同时  $s_{fc}$  与  $P$ 、 $t_{4.5}$  并非高度共线,三者能从输出能力、热状态、经济效率 3 个相互补充的维度共同表征发动机气路综合性能,避免了信息冗余。

综上,  $P$ 、 $t_{4.5}$ 、 $s_{fc}$  三个参数各自与多个气路关键参数高度相关,同时三者之间不完全共线,从物理

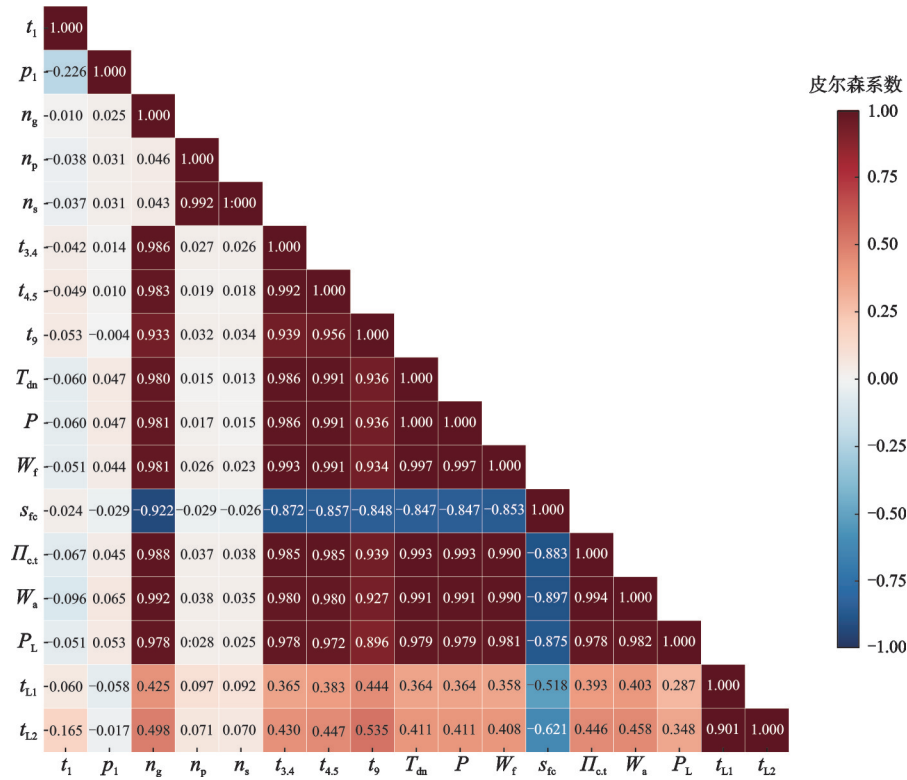


图2 气路参数相关性分析热力图

Fig.2 Thermodynamic diagram of gas path parameter correlation analysis

含义上覆盖了发动机气路性能的3个核心维度,选取这3个参数作为气路性能表征指标具有充分的统计依据和工程合理性。

## 2 数据预处理

### 2.1 不同工况的数据转换

在发动机性能监测与评估之前,需要将实测工作参数转化至标准大气状况,消除可变环境因素对发动机性能参数造成的干扰。具体而言,发动机在实际运行中的输出功率、燃油消耗率、排气温度以及进气流量等关键性能指标,均受到当时环境大气压力、环境温度及相对湿度的显著影响。换算能够消除环境温度、压力、湿度波动对性能参数的干扰,确保不同时间、不同环境下采集的试车数据具有可比性,避免将环境因素导致的参数波动误判为发动机性能退化。因此,为了对发动机的性能进行精准评估,依据 GJB 242A—2018 标准规定,将提取的性能数据换算至海平面标准大气状态,换算公式具体为

$$n_{g_{kc}} = n_g \times \left( \frac{288.15}{t_1 + 273.15} \right)^{0.5} \quad (1)$$

$$P_{kc} = P \times \frac{101\,325}{p_{t_1}} \times \left( \frac{288.15}{t_1 + 273.15} \right)^{0.5} \quad (2)$$

$$t_{4.5_{kc}} = (t_{4.5} + 273.15) \times \left( \frac{288.15}{t_1 + 273.15} \right)^{0.85} - 273.15$$

$$W_{f_{kc}} = W_f \times \frac{101\,325}{p_{t_1}} \times \left( \frac{288.15}{t_1 + 273.15} \right)^{0.5} \times \frac{H_u}{43\,136} \quad (3)$$

$$s_{f_{kc}} = \frac{W_{f_{kc}}}{P_{kc}} \quad (4)$$

式中: $n_{g_{kc}}$ 、 $P_{kc}$ 、 $t_{4.5_{kc}}$ 、 $W_{f_{kc}}$ 、 $s_{f_{kc}}$ 为海平面、标准大气、基准燃油热值条件下的高压轴转速、轴功率、自由涡轮进口总温、燃油流量和燃油消耗率; $P_{t_1}$ 为发动机进口总压; $H_u$ 为燃油低热值。

该换算公式适用于带独立燃气发生器和自由涡轮的分轴式燃气轮机,仅在换算转速相等的相似工况下有效,且要求不存在进气压力损失、排气背压、燃料热值偏差、高湿度高海拔低雷诺数等偏离标准大气的工况。

### 2.2 稳态数据提取

从540个发动机试车数据文件中选取1个典型文件对1.2节选定的4个核心参数进行可视化展示,如图3所示。可以看出,在一次试车程序中发动机经历了多种工况的测试。转速参数表现出明显的多阶段、多工况运行特点,符合典型发动机试车中“慢车-中间状态-最大状态”的测试剖面设计;功率参数与温度参数、转速参数呈现出高度的正相关关系。温度参数的梯度变化完整反映了发动机从启动、过渡状态到额定功率运行的全过程。高温

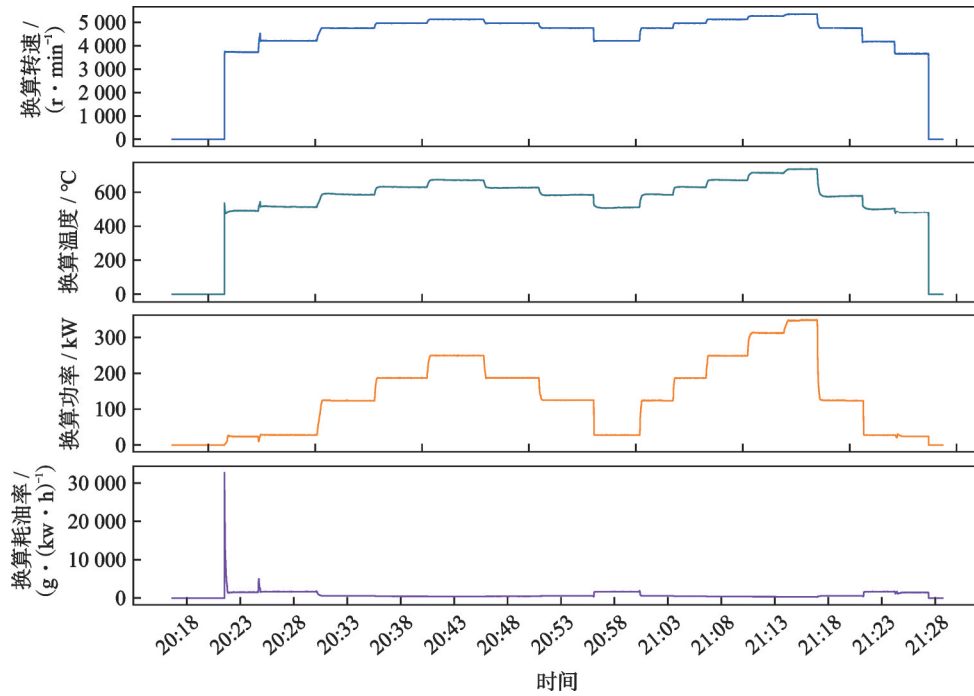


图3 试车数据文件核心参数

Fig.3 Core parameters of test run data file

区域集中出现在额定功率附近,符合燃气涡轮发动机热力循环中温度与功率输出正相关的基本规律;而燃油消耗率则与功率呈现负相关关系,在额定功率运行时试车数据耗油率最低。

由于发动机试车程序存在多阶段、多工况的运行特点,测试数据既包含于特定工况下运行的稳态段,也包含启动、加速、减速等过程的瞬态段,稳态段数据和瞬态数据混杂其中。若直接使用原始数据会因波动干扰导致分析结果失真。只有稳态段数据能够反映发动机在特定工况下的稳定输出特性。用稳态段中多个点数据的平均值代替单个点的数据,也能够一定程度上减小因测量导致的误差。因此,为避免瞬态数据中的参数剧烈变化掩盖发动机真实性能水平并减小测量误差,需要对原始数据进行稳态提取。

### 2.2.1 数据重采样

发动机试车原始数据采样频率较高,为1000 Hz,若逐行判断每个时刻是否处于稳态,会产生巨大的计算量,进而导致稳态提取效率低下、耗时过长。因此,首先对原始数据按200 ms的时间间隔进行重采样,即原始数据中每200 ms的所有数据重采样为新数据的一行数据,新数据取原所有数据的中位数作为新数据的重采样值。这样就将采样频率为1000 Hz的原始数据压缩为了采样频率为5 Hz的重采样数据,大大降低计算量的同时用中位数平滑掉了块内的尖峰噪声,为后续提取分析提供了平稳、等间隔的高质量数据。

### 2.2.2 逐点状态识别

从重采样数据的第1行数据开始建立滑动窗口,窗口大小为8 s。对于采样频率为5 Hz的重采样数据,每个窗口包含40行数据,即窗口 $D_k = \{x_k, x_{k+1}, \dots, x_{k+40}\}$ ,其中 $x_k \in \mathbb{R}^m$ 是第 $k$ 行数据, $m$ 是数据维度。对窗口中的数据做以下操作。

(1)对窗口中40行数据的 $n_{g_{kc}}$ 、 $P_{kc}$ 、 $s_{fc_{kc}}$ 三个参数,分别计算每个参数的变异系数(Coefficient of variation, CV)

$$CV = \frac{\sigma}{\mu} \times 100\% \quad (6)$$

式中: $\sigma$ 为标准差, $\mu$ 为均值。变异系数是衡量数据离散程度的相对指标。

(2)对窗口中40行数据的 $t_{4.5_{kc}}$ 参数,计算标准差 $\sigma$ 。

(3)阈值判定:当窗口 $D_k$ 同时满足以下4个条件时,将 $x_k$ 标记为稳态。

$$CV(n_{g_{kc}}) \leq 0.3\%$$

$$CV(P_{kc}) \leq 1\%$$

$$\sigma(t_{4.5_{kc}}) \leq 5^\circ\text{C}$$

$$CV(s_{fc_{kc}}) \leq 2\%$$

式中 $t_{4.5_{kc}}$ 采用滚动标准差,而 $n_{g_{kc}}$ 、 $P_{kc}$ 、 $s_{fc_{kc}}$ 采用变异系数,主要基于它们量纲和工程意义的差异; $n_{g_{kc}}$ 、 $P_{kc}$ 、 $s_{fc_{kc}}$ 三个参数在不同工况下绝对数值跨度悬殊,若直接用标准差进行阈值判定,低转速段的标准差天然较小,会导致在低工况下更容易被误判为稳态,采用变异系数可归一化各工况的波动程度,实现跨工

况统一判断;而  $t_{4.5_{kc}}$  温度的工况间变化相对平缓,且工程上对温度的绝对波动有明确的物理意义和工程标准,故采用标准差与绝对阈值比较更直观。

4个参数稳态阈值的确定方法为:从540个试车数据文件中随机抽取20个包含多种工况的典型文件,人工标注其中明确的稳态段;统计所有人工标注稳态段内各参数的CV和标准差分布;取分布的95百分位数初步作为阈值进行稳态提取。后续根据提取效果适当调整,得到最终的筛选阈值。

当窗口  $D_k$  执行完上述操作后,继续对窗口  $D_{k+1}$  执行上述操作,直到没有足够的窗口  $x_{k+40}$  使得窗口  $D_k$  满足窗口大小为8s。

### 2.2.3 连续稳态段提取

对上述逐点判断是否处于稳态后的数据,寻找连续的稳态段  $X_i = \{x_n, x_{n+1}, \dots, x_{n+m}\}$ ,其中  $x_n$  为

稳态段的起始点,即  $x_n$  处于稳态而  $x_{n-1}$  不处于稳态; $m$  是使  $X_i$  中所有点均为稳态点的最大值。对识别到的稳态段进行以下操作:

(1)修剪:去掉每段首尾1s的部分,避免边缘波动;

(2)过滤:仅保留修剪后持续时间大于等于6s的稳态段,排除短暂稳定的工况;

(3)汇总:记录每个稳态段的起止时间、持续时间,并计算段内  $n_{g_{kc}}$ 、 $P_{kc}$ 、 $t_{4.5_{kc}}$  和  $s_{fc_{kc}}$  参数的均值,以此作为该稳态段的参数值。

提取效果如图4所示,可以看出在2024-6-18的一次试车任务中共提取出11个稳态段,并且图4中4个参数的稳态段上下——对齐,说明每个稳态段的4个参数均处于稳态,符合上述须同时满足4个阈值判定条件的提取规则。

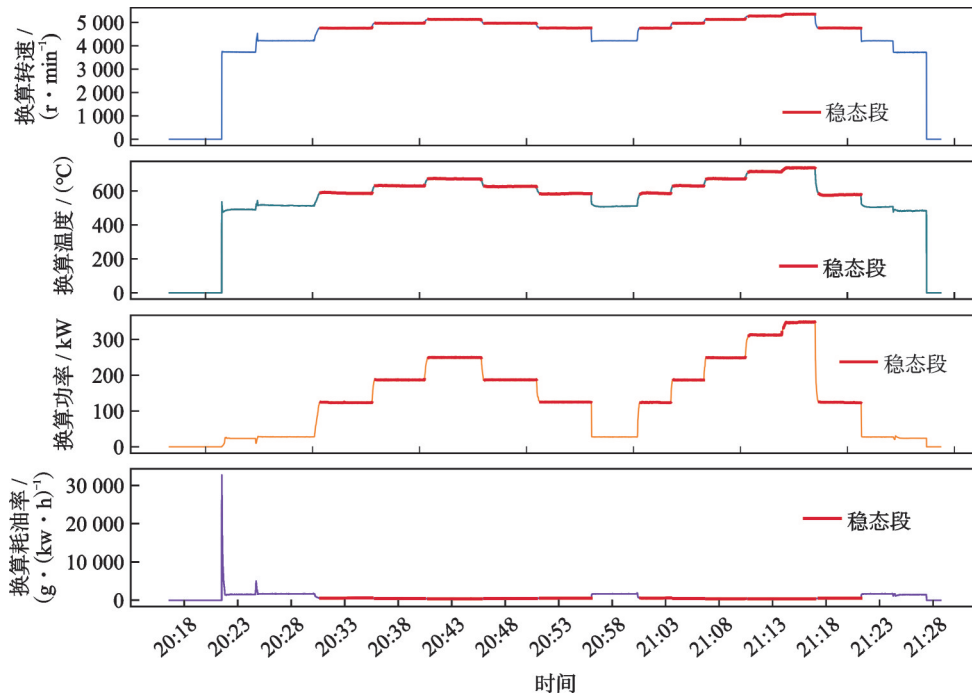


图4 稳态提取效果示意

Fig.4 Schematic diagram of steady-state extraction effect

### 2.2.4 提取结果

按照上述的提取步骤,从540个试车数据文件中共计提取出1955段稳态数据,涵盖5018454个数据点。每段稳态数据包含的参数有起止时间,持续时间,累计工作时间,稳态段平均  $n_{g_{kc}}$ 、 $P_{kc}$ 、 $t_{4.5_{kc}}$  和  $s_{fc_{kc}}$ 。

稳态数据的部分统计结果如图5所示。在转速参数( $n_{g_{kc}}$ )方面,直方图集中在  $4.8 \times 10^4 \sim 5.3 \times 10^4$  r/min,均值约  $4.99 \times 10^4$  r/min、中位数  $5.11 \times 10^4$  r/min,说明稳态段多发生在中高功率区域,符合发动机在试车、调试阶段长时间保持高压转子额定转速的工况特点。占比较小的低速稳态段通常对应启停或特定调试点。

稳态数据占比方面,约65%的采样点被归类为稳态数据,表明试验以稳态台架工况为主;剩余35%为加速、减速或参数波动阶段,体现发动机在工况切换或调节控制时的非稳态状态。

持续时间方面,第1四分位数  $Q_1$  为39.6s,中位数为77.4s,第3四分位数  $Q_3$  为5.6min,四分位距(Interquartile range, IQR)为4.9min,第99百分位数  $P_{99}$  约355s。上述统计结果表明,多数稳态段持续时间较短,主要用于验证单一工况下的瞬态性能;而长尾段则对应少数持续数百秒的耐久性或稳态性能评估工况,体现了多工况、短驻留与少量长时间稳态验证相结合的典型试验策略。

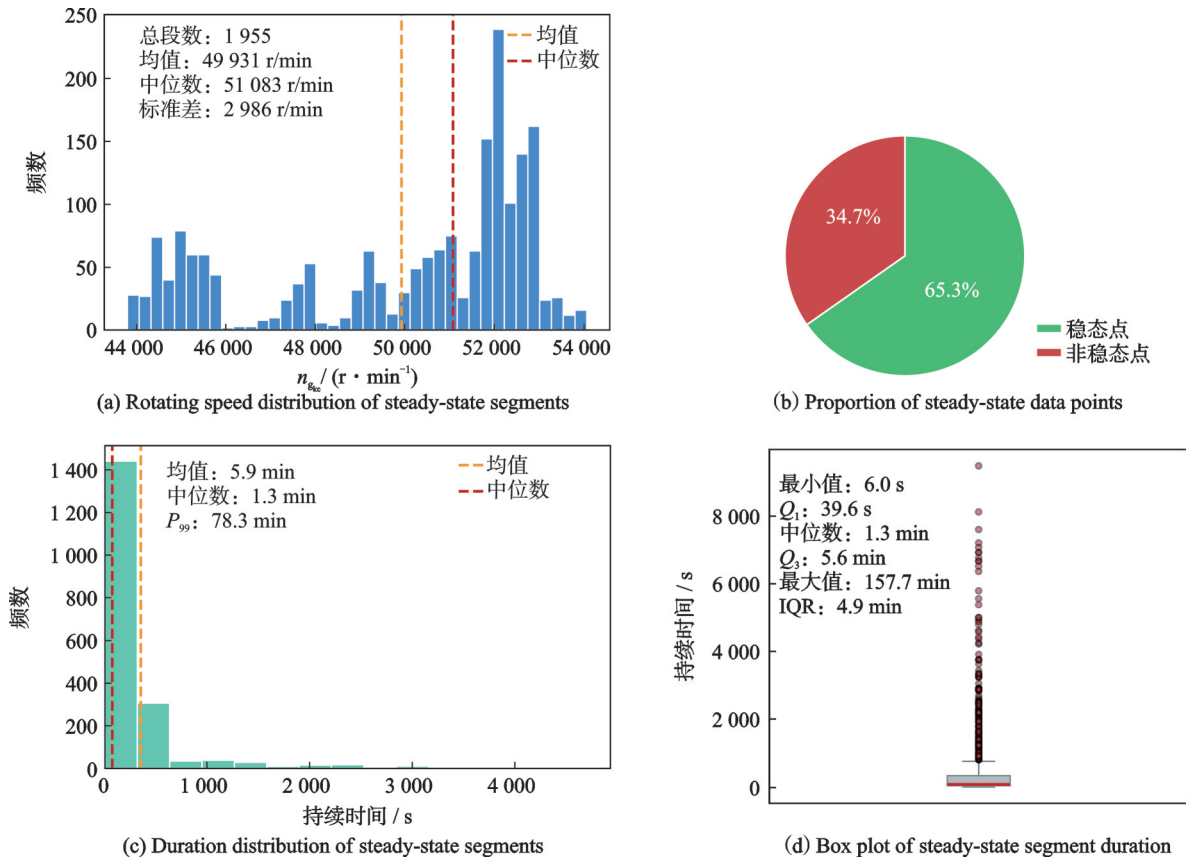


图 5 稳态数据统计结果

Fig.5 Statistical results of steady-state data

### 2.3 维修活动识别

发动机清洗是降低发动机腐蚀风险、恢复发动机动力性能的一种经济而又有效的维修活动,现已成为发动机维护规程上的一项必要程序。通过发动机清洗,能够使因积垢等因素导致的发动机气流通道部件效率下降、整机推力降低、耗油率升高等性能退化得到极大程度的恢复<sup>[17]</sup>。

另外,涡轴发动机的维修活动还包括叶片修复更换、燃油喷嘴清洁、密封件更换等,这些操作均可能使性能参数出现类似清洗的恢复。这些维修活动能够使发动机性能得到恢复,使性能退化量清零或降低,进而影响性能预测。因此,本文提出一种维修活动识别方法,无需依赖外部维修记录,通过捕捉维修活动所导致的发动机性能显著恢复现象,对维修前后数据进行分段建模,从而剥离发动机性能突变对预测的干扰。

#### 2.3.1 计算相对退化量

识别维修事件的核心思路是识别发动机性能突然得到改善的时间点。但各稳态段的转速各不相同,其功率等气路性能参数无法直接比较。为此,假设当发动机处于同一退化程度时,各转速下的气路性能参数的相对退化量相同,而相对退化量

会随着运行时间而逐渐增大,在维修事件时骤然减小。基于这个假设,通过构建发动机气路性能基线模型,计算各稳态段功率的相对退化量,识别相对退化量的突变点,从而得到维修事件时间点。

由于该型发动机的设计曲线无法直接获取,利用该型发动机耐久试验性能数据,通过拟合得到各转速下的功率值,以此作为该发动机的性能基线模型。

得到基线模型后,对上文提取得到的每一行稳态数据,计算该稳态转速在基线模型中的功率基线值  $P_{bl}$ ,并进一步计算稳态数据真实值  $P_{kc}$  相对于基线值  $P_{bl}$  的相对偏差  $\Delta P = \frac{P_{kc} - P_{bl}}{P_{kc}}$ 。

#### 2.3.2 稳健线性拟合

由于数据采集误差、基线模型误差等因素,数据中不可避免地会有异常值的存在。若不进行处理,这些异常值后续可能会被错误地识别为维修事件。为排除异常值对识别的干扰,对  $\Delta P$  进行稳健线性拟合(Robust line fit),通过迭代修正机制降低异常值的干扰,具体步骤如下。

(1) 初始拟合:对所有点做一次线性拟合  $\hat{y} = kx + b$ ,得到趋势线;

(2) 计算残差:计算每个点与趋势线的残差

$$\sigma_i = y_i - \hat{y}_i;$$

(3) 异常值判断:计算所有点的残差的中位数绝对偏差(Median absolute deviation, MAD)为  $\text{median}(|\delta - \text{median}(\delta)|)$ , 然后对每一个点分别进行判断,若  $|\delta_i - \text{median}(\delta)| > 3\text{MAD}$ , 则将该点视为异常值剔除;

(4) 迭代:剔除异常值后重复步骤(1)~(3), 直到不再有异常值被剔除。

### 2.3.3 跳变点识别

对剔除异常值的数据,通过分析  $\Delta P$  的正向跳变特征识别潜在维修事件。对每一行稳态数据的  $\Delta P$ , 分别计算其后 10 行数据和前 10 行数据的  $\Delta P$  平均值,通过前后均值的对比,判断当前时间点是否为可能的维修事件。判断该时间点为可能的维

修事件需同时满足两个条件。

(1) 改善幅度达标:前窗口均值 - 后窗口均值  $\geq$  最小改善幅度,即性能有明显提升;

(2) 改善前性能较差:前窗口均值  $\leq$  性能阈值,即改善前的性能处于需要维修的状态。

具体到本文方法中,最小改善幅度为基线性能的 3.5%, 该数值基于对所有点改善幅度的统计结果确定;而性能阈值指的是本文中的基线性能。文中建立基线模型的数据取自该发动机耐久试验报告,其代表的性能应为该型号发动机在合规使用、标准维护下,寿命末期的最低保证性能。因此当发动机性能接近或低于基线模型时,可以认为此时发动机性能较差。

通过上述规则,对剔除异常值后的数据进行识别,识别结果如图 6 所示。

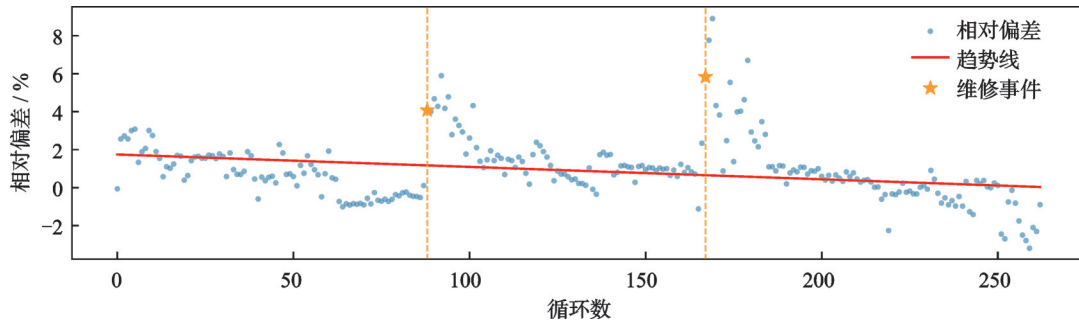


图 6 维修点识别结果

Fig.6 Maintenance point identification results

### 2.3.4 分段结果

所使用的试车数据时间跨度为 2024-06-18~2025-05-24, 按上节识别结果所示分别于 2024-10-11 和 2024-11-11 进行了两次发动机维修。将试车数据按上述时间点分为 3 段, 每一段分别重置累计运行时间, 以备后续使用。各段基本信息如表 3 所示, 可见 3 个时段的稳态段数量与总运行时间相差不大, 均具备足够的样本量用于后续模型训练。

表 3 分段后各段情况

Table 3 Information of each segment after segmentation

数据段	日期范围	稳态段数	总运行时间/s
第 1 段	2024-06-18~2024-10-10	670	461 286
第 2 段	2024-10-11~2024-11-11	638	366 568
第 3 段	2024-11-12~2025-05-24	647	288 689

## 3 性能预测结果

本节对上文确定的用于表征发动机气路性能的关键参数, 即  $n_{g_{kc}}$ 、 $P_{kc}$ 、 $t_{4.5_{kc}}$  和  $s_{fc_{kc}}$  进行预测分析。

根据 1.2 节参数分析的结果, 选定“根据  $n_g$  预测  $P$ ”与“根据  $P$  预测  $t_{4.5}$  和  $s_{fc}$ ”两种模式进行  $P_{kc}$ 、 $t_{4.5_{kc}}$  和  $s_{fc_{kc}}$  的预测。同时, 为考虑发动机运行过程中的缓慢退化, 将累计运行时间也作为性能参数加入模型训练, 从而实现了对发动机退化性能的预测。

针对发动机转速、功率、温度、燃油消耗率、累计运行时间等参数之间存在的高阶非线性耦合关系, 使用随机森林算法实现发动机性能的预测。随机森林(Random forest, RF)算法是一种基于决策树的集成学习算法。它利用 bootstrap 重抽样法生成多个互不相关的分类回归树(Classification and regression tree, CART), 该算法对单棵 CART 的分类能力要求不高, 但通过独立训练和投票机制避免了决策树算法中的过拟合问题, 提高了对待评估样本的预测正确率<sup>[18-19]</sup>。随机森林算法能够自动捕捉发动机特征之间复杂的非线性关系, 并通过多棵树的集成提升稳定性与抗噪能力, 既无需复杂特征工程, 也不容易过拟合, 在精度、可解释性和工程实用性之间取得了良好平衡。

### 3.1 根据 $n_g$ 预测 $P$

通过第 2 节的试车数据预处理,从原始的 540 个试车数据文件中提取得到了 3 个时段的稳态数据文件。每个时段的稳态数据均按时间顺序将前 80% 的稳态段作为训练集,后 20% 的稳态段作为测试集,以  $n_{gc}$  和累计运行时间为输入特征,采用随机森林算法构建预测模型,通过训练集数据完成模型训练,实现通过输入  $n_{gc}$  和累计运行时间预测  $P_{kc}$ 。同时在训练时确保 20% 的测试集完全不参与模型训练,防止数据泄露。

主要超参数设置如下:决策树数量  $n\_estimators = 100$ ,最大深度  $max\_depth = 15$ ,最小分割样本数  $min\_samples\_split = 5$ ,最小叶节点样本数  $min\_samples\_leaf = 2$ ,随机种子  $random\_state = 42$ 。其中,限制最大深度(15层)与最小叶节点样本数(2个)共同起到正则化作用,防止单棵决策树过度生长导致过拟合;100 棵树的集成规模在保证预测稳定性的同时兼顾了计算效率。上述参数基于经验设置并经初步验证,在本文数据集上取得了良好的预测效果。

利用训练好的模型对测试集数据进行预测,采用决定系数  $R^2$ 、均方根误差(Root mean square error, RMSE)和平均绝对误差(Mean absolute error, MAE)作为模型评价指标,3 段稳态数据的预测值与实际值对比如图 7~9 所示,预测误差统计如表 4 所示。可以看出,各段的最大相对误差都在 5% 以内,平均误差能够控制在 1% 左右,误差处于工程可接受范围。

观察各段的误差统计数据,可以发现第 3 段数据的平均相对误差略大于前两段,推测原因为第 3 段是第 2 次清洗后的数据,发动机累计运行时间最长,退化程度最深,并且第 3 段数据的时间跨度也是 3 段数据中最长的,但数据量却与前两段相当,导致预测精度略有下降。

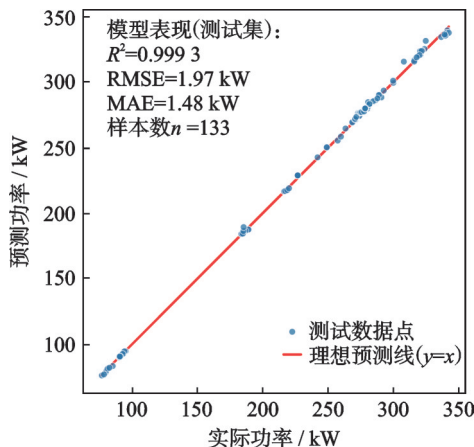


图 7 第 1 段数据预测结果

Fig.7 Prediction results of the first segment data

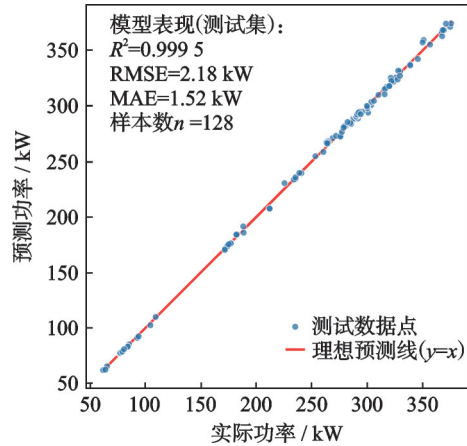


图 8 第 2 段数据预测结果

Fig.8 Prediction results of the second segment data

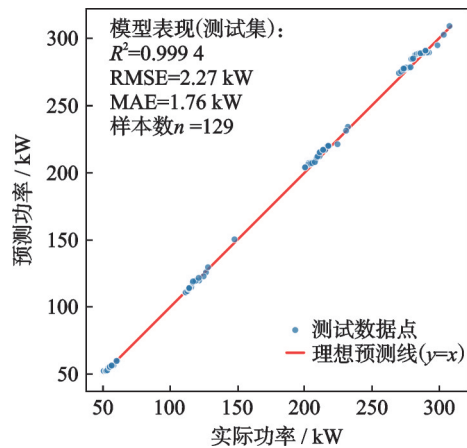


图 9 第 3 段数据预测结果

Fig.9 Prediction results of the third segment data

表 4 功率预测误差统计

Table 4 Statistics of power prediction error

数据段	最高绝对误差/kW	MAE/kW	最高相对误差/%	平均相对误差/%
第 1 段	7.732 6	1.483 5	2.51	0.60
第 2 段	8.587 5	1.521 6	2.45	0.66
第 3 段	5.663 7	1.764 9	4.11	1.09

### 3.2 根据 $P$ 预测 $t_{4.5}$ 和 $s_{fc}$

将 3 段数据以  $P_{kc}$  和累计运行时间为特征,利用与 3.1 节相同的方法进行  $t_{4.5}$  和  $s_{fc}$  的预测,预测结果如图 10 所示。预测误差统计如表 5 所示。可以看出,各段两个预测目标的最高相对误差都在 7% 以内,平均相对误差都控制在 1% 以内,处于工程可接受范围。

观察误差分布,在平均相对误差相近的同时,最高相对误差明显大于功率的预测结果,说明相比于功率,温度与燃油消耗率的瞬时波动更大。其中最大误差主要集中在高功率工况区,该区域对应发动机高负荷运行状态,说明在高功率工况下,发动机的参数波动更剧烈。

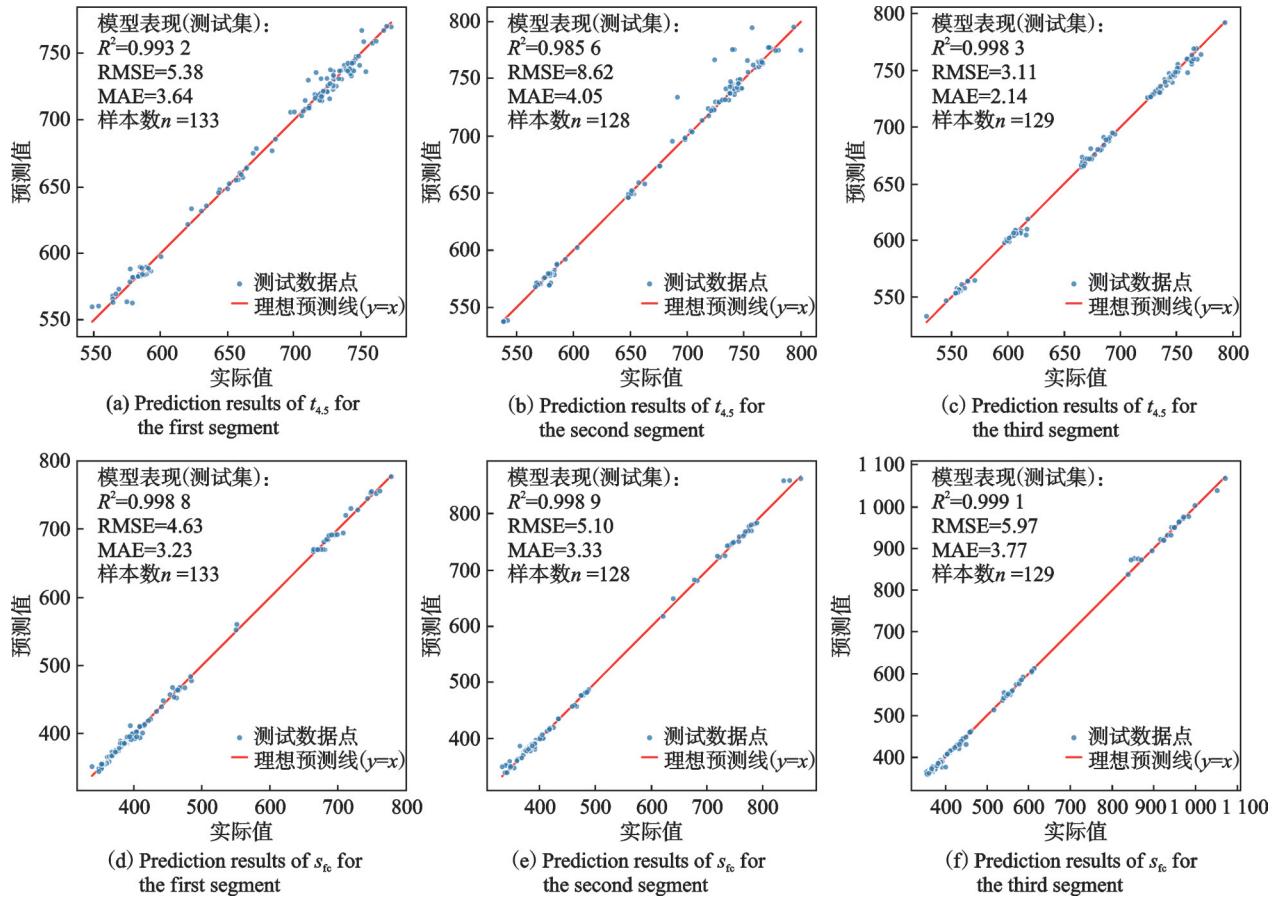


图 10 各段  $t_{4.5kc}$  与  $s_{fckc}$  预测结果

Fig.10 Prediction results of  $t_{4.5kc}$  and  $s_{fckc}$  for each segment

表 5  $t_{4.5kc}$  与  $s_{fckc}$  误差统计

Table 5 Error statistics of  $t_{4.5kc}$  and  $s_{fckc}$

数据段	预测目标	最高绝对误差	MAE	最高相对误差/%	平均相对误差/%
第 1 段	$t_{4.5kc}$	19.056 3 °C	3.640 1 °C	2.793 8	0.538 4
	$s_{fckc}$	17.171 0 g/(kw·h)	3.227 9 g/(kw·h)	4.350 4	0.706 5
第 2 段	$t_{4.5kc}$	42.424 6 °C	4.046 5 °C	6.130 8	0.571 2
	$s_{fckc}$	22.336 1 g/(kw·h)	3.330 4 g/(kw·h)	5.984 8	0.736 3
第 3 段	$t_{4.5kc}$	11.579 4 °C	2.135 1 °C	1.870 9	0.318 0
	$s_{fckc}$	26.833 2 g/(kw·h)	3.772 1 g/(kw·h)	6.124 2	0.753 7

3.3 对比实验

为验证本文所提方法的有效性与优势,以相同的训练集、测试集划分和相同的输入特征,将本文方法与以下 3 种对比算法在功率  $P$  预测任务上进行比较。各对比算法均按照文献中推荐的常用参数。

(1)不分段 RF:不对试车数据按维修活动时间点进行分段处理,直接将全段 1 955 个稳态数据作为一个整体训练 RF 模型,其余超参数设置与本文方法完全相同。

(2)支持向量回归(Support vector regression, SVR):以径向基函数(Radial basis function, RBF)为核函数,正则化参数  $C=100$ ,容错率  $\epsilon=0.1$ ,核参数 gamma 采用 sklearn 默认的 scale 策略,使用与

本文方法相同的分段数据训练。

(3)  $K$  近邻回归 ( $K$ -nearest neighbors regression, KNN):采用欧氏距离度量,邻居数  $K=5$ ,权重策略为均匀权重(weights=uniform),并行计算利用全部 CPU 核心,使用与本文方法相同的分段数据训练。

各算法预测误差统计如表 6 所示。可以看出,本文分段 RF 方法在所有指标上均取得最优结果。

综合 3.1 节与 3.2 节的预测结果可知,基于 RF 算法构建的预测模型在两种预测模式下均表现出优异的性能:以  $n_{fc}$  和累计运行时间为特征预测  $P_{kc}$  时,各数据段平均相对误差控制在 0.60%~1.09% 之间;以  $P_{kc}$  和累计运行时间为特征预测  $t_{4.5kc}$  与  $s_{fckc}$  时,平均相对误差均不超过 0.76%,最高相对误差

表 6 各算法功率预测误差对比

Table 6 Comparison of power prediction errors of each algorithm

算法	$R^2$	RMSE/kW	MAE/kW	最高相对误差/%	平均相对误差/%
分段 RF(本文)	0.999 3	1.97	1.48	2.51	0.60
不分段 RF	0.989 4	7.51	6.25	14.41	3.51
SVR	0.841 2	30.17	19.63	35.09	7.88
KNN	0.941 6	18.29	12.03	14.89	5.06

未突破 7%,所有核心指标的预测精度均满足工程应用要求。累计运行时间的引入有效捕捉了发动机性能退化规律,而 3 段独立数据的重复验证则规避了单一数据集的偶然性误差。

由本节对比实验可知,本文提出的分段 RF 方法在涡轴发动机气路性能预测任务中具有显著优势。维修活动识别与分段建模有效消除了维修干扰,使数据退化趋势连续可学习;RF 算法凭借对多工况非线性特征的自适应建模能力,相比 SVR 和 KNN 具有更强的工程适用性,能够满足涡轴发动机气路性能预测的精度要求。

整体来看,该预测方法能够精准拟合气路参数间的非线性多耦合关系,可以通过输出发动机气路关键性能指标的预测结果,为发动机全生命周期健康管理提供数据支撑。

## 4 结 论

(1)提出了一种从发动机试车数据中提取稳态数据的方法,通过构建滑动窗口和变异系数阈值判断提取多工况数据中的稳态数据,以稳态段的气路性能参数代替瞬态数据进行建模,有效提升了模型预测精度与运行可靠性。

(2)提出了一种通过相对退化量分析识别数据中存在的维修活动时间点的方法。通过稳健线性拟合迭代修正机制降低异常值的干扰,提升识别的准确性,从而实现性能退化的分段建模,成功排除了维修活动对建模的干扰。

(3)提出了一种基于 RF 算法的涡轴发动机气路性能预测方法,将累计运行时间作为特征加入模型训练从而考虑退化对发动机气路性能的影响,进一步提高了预测结果的准确性。各段内 3 个预测指标平均误差均控制在 1% 左右,验证了该方法的有效性 with 准确性。

### 参考文献:

- [1] WANG H F, JIANG W, DENG X Y, et al. A new method for fault detection of aero-engine based on isolation forest[J]. *Measurement*, 2021, 185: 110064.
- [2] 曹明,黄金泉,周健,等.民用航空发动机故障诊断与健康管理系统现状、挑战与机遇 I:气路、机械和

FADEC 系统故障诊断与预测[J]. *航空学报*, 2022, 43(9): 1-34.

CAO Ming, HUANG Jinquan, ZHOU Jian, et al. Current status, challenges and opportunities of civil aero-engine diagnostics & health management I: Diagnosis and prognosis of engine gas path, mechanical and FADEC[J]. *Acta Aeronautica et Astronautica Sinica*, 2022, 43(9): 1-34.

- [3] 林京,张博瑶,张大义,等.航空燃气涡轮发动机故障诊断研究现状与展望[J]. *航空学报*, 2022, 43(8): 7-20.

LIN Jing, ZHANG Boyao, ZHANG Dayi, et al. Research status and prospect of fault diagnosis for gas turbine aeroengine[J]. *Acta Aeronautica et Astronautica Sinica*, 2022, 43(8): 7-20.

- [4] ZHANG C Q, CUI L F, ZHANG Q Z, et al. Online anomaly detection for aeroengine gas path based on piecewise linear representation and support vector data description[J]. *IEEE Sensors Journal*, 2022, 22(23): 22808-22816.

- [5] 周良.基于数据驱动的航空发动机健康评估[D].南京:南京航空航天大学,2022.

ZHOU Liang. Aero-engine health assessment based on data-driven[D]. Nanjing: Nanjing University of Aeronautics and Astronautics, 2022.

- [6] JOLY R B, OGAJI S O T, SINGH R, et al. Gas-turbine diagnostics using artificial neural-networks for a high bypass ratio military turbofan engine[J]. *Applied Energy*, 2004, 78(4): 397-418.

- [7] 刘恩宏,王富宁,张敏,等.基于试车和总体性能仿真数据驱动的燃气涡轮发动机性能预测[J/OL]. *航空动力学报*: 1-14 [2026-01-23]. <https://doi.org/10.13224/j.cnki.jasp.20250308>.

LIU Enhong, WANG Funing, ZHANG Min, et al. Performance prediction of gas turbine engines driven by test and overallperformance simulation data[J/OL]. *Journal of Aerospace Power*: 1-14 [2026-01-23]. <https://doi.org/10.13224/j.cnki.jasp.20250308>.

- [8] 寇容海,李文博,党庆庆,等.数据与模型驱动的航天器姿态控制系统故障诊断[J]. *航空学报*, 2025, 46(S1): 145-160.

KOU Ronghai, LI Wenbo, DANG Qingqing, et al. Fault diagnosis of spacecraft attitude control system driven by data and model[J]. *Acta Aeronautica et As-*

- tronautica Sinica, 2025, 46(S1): 145-160.
- [9] 潘鹏飞. 基于试飞数据的航空发动机状态监测与故障诊断[J]. 推进技术, 2021, 42(12): 2826-2837.  
PAN Pengfei. Flight data based condition monitoring and fault diagnosis of aero-engine[J]. Journal of Propulsion Technology, 2021, 42(12): 2826-2837.
- [10] 王奕惟, 莫李平, 王奕首, 等. 基于全航段 QAR 数据和卷积神经网络的航空发动机状态辨识[J]. 航空动力学报, 2021, 36(7): 1556-1563.  
WANG Yiwei, MO Liping, WANG Yishou, et al. Aero-engine status identification based on full-segment QAR data and convolutional neural network[J]. Journal of Aerospace Power, 2021, 36(7): 1556-1563.
- [11] 韩淞宇, 邵海东, 姜洪开, 等. 基于提升卷积神经网络的航空发动机高速轴承智能故障诊断[J]. 航空学报, 2022, 43(9): 150-163.  
HAN Songyu, SHAO Haidong, JIANG Hongkai, et al. Intelligent fault diagnosis of aero-engine high-speed bearings using enhanced CNN[J]. Acta Aeronautica et Astronautica Sinica, 2022, 43(9): 150-163.
- [12] LIAO Z B, WANG J, LIU J X, et al. Uncertainties in gas-path diagnosis of gas turbines: Representation and impact analysis[J]. Aerospace Science and Technology, 2021, 113: 106724.
- [13] 钟诗胜, 李旭, 张永健. 基于 DBN 的不均衡样本驱动民航发动机故障诊断[J]. 航空动力学报, 2019, 34(3): 708-716.  
ZHONG Shisheng, LI Xu, ZHANG Yongjian. Fault diagnosis of civil aero-engine driven by unbalanced samples based on DBN[J]. Journal of Aerospace Power, 2019, 34(3): 708-716.
- [14] 张鑫, 郭顺生, 李益兵, 等. 基于拉普拉斯特征映射和深度置信网络的半监督故障识别[J]. 机械工程学报, 2020, 56(1): 69-81.  
ZHANG Xin, GUO Shunsheng, LI Yibing, et al. Semi-supervised fault identification based on Laplacian eigenmap and deep belief networks[J]. Journal of Mechanical Engineering, 2020, 56(1): 69-81.
- [15] 王修岩, 李萃芳, 高铭阳, 等. 基于 SVM 和 SNN 的航空发动机气路故障诊断[J]. 航空动力学报, 2014, 29(10): 2493-2498.  
WANG Xiuyan, LI Cuifang, GAO Mingyang, et al. Fault diagnosis of aero-engine gas path based on SVM and SNN[J]. Journal of Aerospace Power, 2014, 29(10): 2493-2498.
- [16] 张世杰. 数据驱动的民航发动机气路系统故障诊断研究[D]. 成都: 电子科技大学, 2025.  
ZHANG Shijie. Research on data-driven fault diagnosis of civil aviation engine gas path system[D]. Chengdu: University of Electronic Science and Technology of China, 2025.
- [17] 石硕. 民用航空发动机性能维持清洗方法研究[J]. 科技信息, 2013(15): 84-85.  
SHI Shuo. On-wing compressor washing investigation for civil aero-engine performance enhancement[J]. Science & Technology Information, 2013(15): 84-85.
- [18] 刘兴, 王艳, 纪志成. 基于随机森林的风电功率短期预测方法[J]. 系统仿真学报, 2021, 33(11): 2606-2614.  
LIU Xing, WANG Yan, JI Zhicheng. Short-term wind power prediction method based on random forest[J]. Journal of System Simulation, 2021, 33(11): 2606-2614.
- [19] 王奕森, 夏树涛. 集成学习之随机森林算法综述[J]. 信息通信技术, 2018, 12(1): 49-55.  
WANG Yisen, XIA Shutao. A survey of random forests algorithms[J]. Information and Communications Technologies, 2018, 12(1): 49-55.

(编辑:张蓓)