运载火箭推进系统的时间序列异常分类：提升长短期记忆网络准确率和数据质量的快速统计检测器

Sean P. Engelstad,[1] Samuel R. Darr,[2] Matthew Taliaferro[3], 和 Vinay K. Goyal[4]
The Aerospace Corporation，美国加利福尼亚州埃尔塞贡多 90245

在发射前支持执行/不执行决策，需要根据设计鉴定阶段制定的红线限值评估实时遥测数据。通常会使用来自地面测试或以往飞行的同族数据来检测初始故障模式及其发生时间；然而，这种方法高度依赖工程判断，且用于新型运载火箭时更容易出错。为解决这些局限，我们采用长短期记忆网络（LSTM）对时间序列异常进行监督分类。不过，由模拟异常数据得出的初始训练标签可能并非最优，因为异常强度、异常稳定时间及其他因素存在差异。在本研究中，我们提出一种基于马氏距离和正向、反向检测比例的新型统计检测器，用于调整监督训练标签。我们在地面级推进系统的数字孪生仿真上验证了该方法，每次试验运行 20.8 分钟，并包含$\boldsymbol{O(10^8)}$个训练时间步。统计数据重新标注使 LSTM 分类器的精确率和召回率分别提高了 7% 和 22%。

I. 引言

L运载火箭需要经过大量测试、分析和检查，以确保具备高可靠性，从而成功将有效载荷送入预定轨道。在新型运载火箭的研发中，推进系统往往是最具挑战性的领域之一 [1-4]。在地面测试和运行期间发现的异常，可以为纠正措施提供指导；实施这些措施后，发射可靠性会有所提高。任务运营中心的工程师实时监控数千个遥测参数。将传感器数据绘制成图，并与地面测试结果、预测、红线限值或历史发射数据进行对比，以评估系统状态。红线限值依据鉴定试验或分析确定 [5-7]。在项目早期，同族数据有限；对于相对较新的运载火箭，红线限值可能不足以检测细微或未预见的异常

[1] MTS研究生实习生，结构部，AIAA会员。（通讯作者，邮箱：sengelstad312@gatech.edu）[2] 工程专家，流体力学部，AIAA会员。[3] 工程专家，流体力学部，AIAA会员。[4] 首席主任，飞行器系统部，AIAA副院士。

批准公开发布。OTR 2025-00908

运载火箭。在作者此前的研究中 [8, 9], , 研究人员开发了一种机器学习方法，用于实时检测和识别异常，以协助工程师实时评估遥测数据。该机器学习模型基于模拟异常进行训练，能够根据遥测数据实时提供异常预测。

Wu 等人讨论了时间序列异常分类领域的最新进展[10], , 而文献 [11, 12]讨论了时间序列异常检测。常见的机器学习（ML）策略包括重构、预测和直接分类方法。已有数百种机器学习模型得到应用，包括长短期记忆网络 [13], 、Transformer [14], 、自编码器 [15], 、变分自编码器[16], , 以及采用小波变换或傅里叶变换的模型[17]。一种新策略 [18] 采用小波变换进行预处理，并使用LSTM和Transformer架构，在基准检测数据集上取得了更优结果。

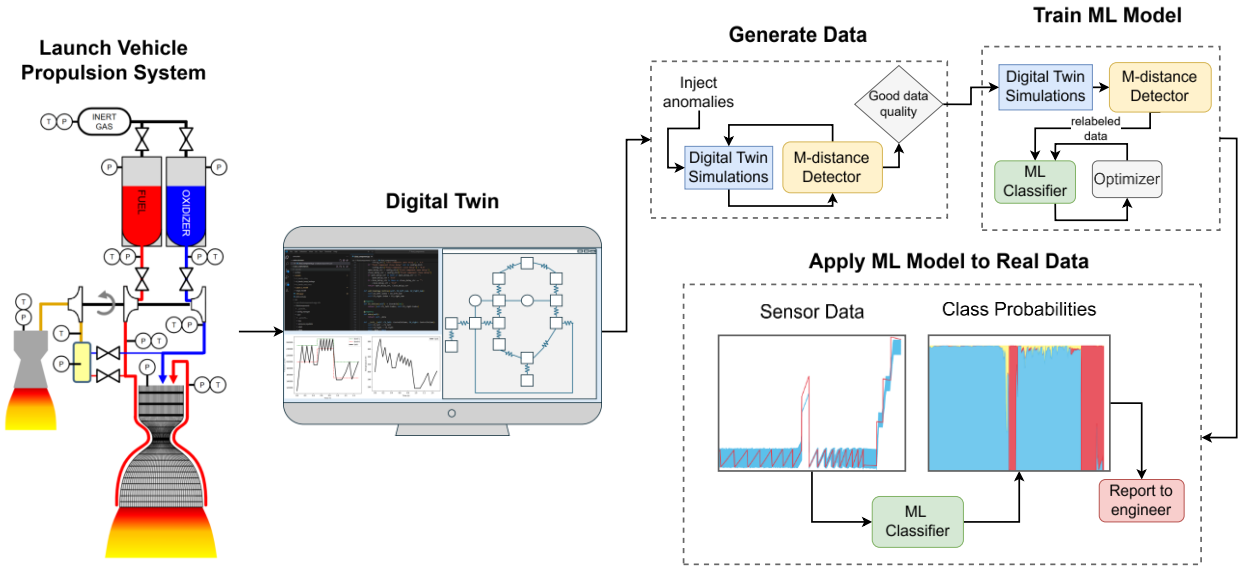


图 1. 论文的主要组成部分：利用来自数字孪生的模拟异常数据训练机器学习分类器。

本文在参考文献 [8, 9], 中所述方法论的基础上，采用定制的（长短期记忆）LSTM模型 [13, 26]，对时间序列数据进行监督式异常分类。本研究中的时间序列分类问题涉及超过 4 亿个时间步、7 个输入参数，以及非周期性的阀门动作；这些动作会导致压力和温度传感器数据中出现斜率不连续性。不规则振荡源于压力带阀门逻辑，即通过阀门循环动作将容器压力控制在特定目标范围或压力带内。在先前研究 [9], 使用的较小数据集上，基于 Transformer 和自编码器的重构与预测策略难以应对非周期性阀门动作造成的斜率不连续性。无监督方法的精确率也往往较低，并且在多类别检测或分类方面可能不如监督式方法 [32]。此外，连续小波变换 [17] 大幅增加了数据集大小，导致这个大型问题超出了内存限制。因此，有必要采用先前研究 [9], 中的定制 LSTM 模型；该模型在小规模推进系统数据集上取得了 0.967 的 F1分数。

本研究旨在展示一种模拟异常与机器学习方法，并将其应用于数据更复杂、时间步更多且更贴近实际的推进系统。与先前研究 [9], 相比，输入参数数量从 4 个增加到 7 个，异常类别数量从 21 个增加到 25 个，时间步数量从 100 个增加到 2720 个。此外，研究还对三个离散的发射阶段进行了建模。研究同时考虑了起始时间可变、持续时间涵盖短、中、长多种时间尺度的异常。

与 [9], 中概述的工作不同，分类标签通过基于时间序列的异常检测器进行校正。与作者之前的测试案例相比，异常生成设置从 20 种大幅增加至 101 种，因此异常并非总能被检测到，具体取决于发射阶段、阀门指令、异常强度或持续时间。这导致监督分类数据的质量较差。我们没有改用在阀门切换数据上表现不佳的无监督分类器，而是使用无监督时间序列异常检测器校正监督分类标签[9], 。

一种采用小波变换预处理以及 LSTM 和 Transformer 架构的新策略 [18] ，在基准检测数据集上表现更优。与其他方法相比，我们更倾向于使用一种轻量级的、基于统计的检测器，该检测器采用马氏距离（M距离），以便在数据集生成和质量改进的多个迭代过程中调整异常生成设置。以往的 M-距离检测方法依赖机器学习模型 [19, 20, 21] 或频率提取 [22]，以获得给定时间步的异常预测。时间窗口和前向 - 后向策略可将随时间变化的 M 距离转换为每个时间步的异常预测。这使我们能够在每个时间步获得准确的检测预测。

本研究的主要组成部分详见图 1，其中包括用于生成模拟异常数据的运载火箭推进系统数字孪生模型。研究采用无监督检测器来调整异常生成设置，并重新标注监督分类数据，以提高数据质量。最后，训练长短期记忆网络监督分类器模型，并将其应用于真实传感器数据。

II. 方法论

A. 流体系统异常仿真

图3所示的地面级推进系统采用 GiNA [23], 进行仿真。GiNA 是一维瞬态CFD求解器，适用于具有 10 至 100 个控制容积的运载火箭推进系统。GiNA 使用有限体积法求解耦合的质量-能量方程。通过基于瑞利数的相关式预测各气相空间中的传热传质，避免了典型 CFD 分析中的空间离散化。采用二阶隐式龙格-库塔格式和自适应时间步，对控制方程进行数值求解。由于自由度较低$O(10^2)$，GiNA 的运行时间主要由物性评估决定，而非线性求解器。GiNA 仿真通常需要约 10 分钟来模拟 2 小时的实际时间，因此能够生成大量数据集用于机器学习训练。

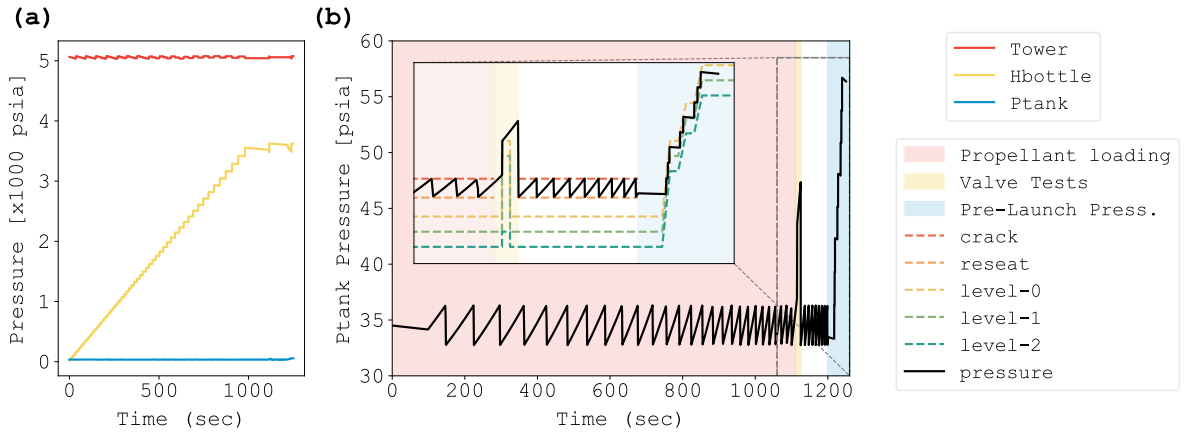


图2。系统的标称响应，包括（a）各控制容积中的压力，以及（b）每个发射增压阶段的推进剂箱压力和阀门压力水平。

GiNA 配备了流体组件（FC）模块，其中包含 4 种主要的流体组件：压力调节器、关断阀（或“开关式”阀门）、可变面积阀和泄压阀。图2展示了本研究推进系统中使用的流体组件，包括四个关断阀 SOV-TOW、SOV-HB、SOV-PT1 和 SOV-PT2，以及一个锁闭式泄压阀 LRV。基于压力的关断阀具有由压力带逻辑决定的 0 或 1 阀门状态。基于压力的指令逻辑可通过循环开启和关闭阀门，使其在预设范围或压力带内控制罐绝对压力或罐间压差。锁闭式泄压阀循环开启和关闭，使罐压保持在开启压力与回座压力之间；除非收到锁闭指令，迫使阀门保持关闭。

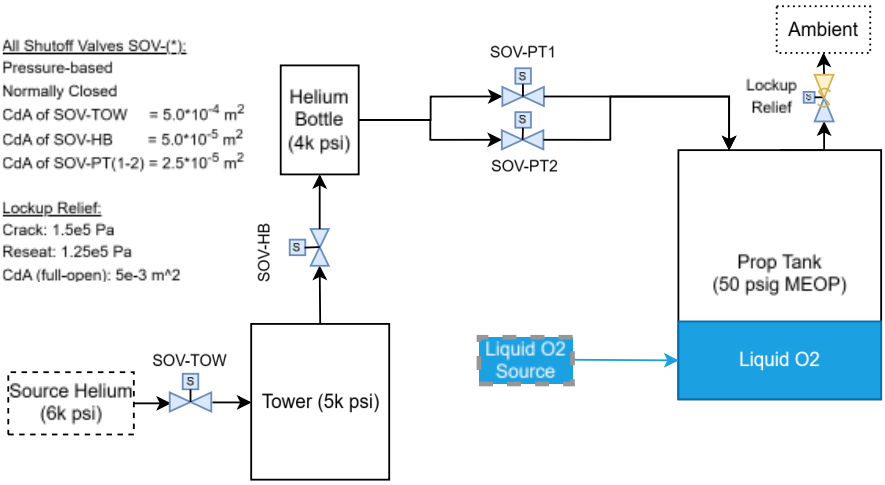


图3。分组 级推进系统，使用气态氦和液氧为推进剂贮箱增压 **k.**

图3中的测试案例旨在模拟一个简化的地面至运载器增压系统。氦气由氦源边界提供，用于给塔架、氦气瓶和推进剂罐加压。关断阀用于控制流向各自下游容积的氦气流。锁闭式泄压阀（LRV）用于排放罐气枕中的气体，以

使压力保持在预先设定的开启压力与回座压力之间。仿真持续1250秒，包含三个运行阶段，如图2左图中的阴影区域所示：推进剂装载、阀门测试和发射前加压。

在 0 至 1100 秒的推进剂装载阶段，氦气瓶加压，同时将液氧（LO_2）装入推进剂贮箱。随着 LO_2 进入推进剂贮箱，气枕容积减小，压力升高；多余的压力通过 LRV 排放至环境压力。在标称运行情况下，推进剂装载期间 LRV 约开启并重新落座 50 次；随着气枕容积缩小，每次开启与重新落座之间的时间间隔逐渐缩短。

1100 秒后，将进行阀门测试。LRV 被锁定，SOV-PT1 和 SOV-PT2 开启，以对推进剂贮箱加压。在 1125 秒时，LRV 解锁，推进剂贮箱减压，直至恢复到 LRV 的标称开启—重新落座压力范围；这一状态持续至下一发射阶段的 1200 秒。

1200 秒时，发射前加压阶段开始。再次锁定 LRV 并开启关断阀，对推进剂贮箱加压；压力分多个区间逐步提升，直至 1250 秒。完整的标称仿真耗时 1.52 分钟，对应 20.8 分钟的实际时间。

B. 异常数据生成

我们使用代码中的异常数据生成（ADG）模块，生成了大量模拟异常数据集。通过蒙特卡洛模拟，从单个部件异常库中随机选取异常进行仿真。为使数据更贴近实际，每次试验和每种故障模式都包含范围广泛的随机输入。例如，每个控制容积的初始压力和温度都是随机设定的，泄漏面积、阀门缓慢开启延迟、开始时间和异常持续时间等异常设置也同样如此（部分示例见附录中的表A.1）。对于许多设置，使用了以 3σ 为界的正态分布。蒙特卡洛异常仿真使用 60 个CPU处理器，在 11.4 小时内完成了 15000 次试验。每次试验均保存了压力和温度传感器的七个输入参数，包括用于监督训练的异常发生时间。

图3所示推进系统的可用异常见附录中的图A.1，共包括 1 个标称类别和 24 个异常类别。其中有八种短持续时间异常，持续时间约为一秒，包括失效开启和失效关闭故障。内部泄漏属于中等持续时间异常，持续约 40 至 100 秒。其余异常，例如缓慢关闭、缓慢开启和高开启压力等，会持续整个仿真过程。尽管有些长持续时间异常可能只会在阀门动作的短暂时段内显现。例如，只有在阀门循环关闭时，缓慢关闭异常才会显现，而整个仿真过程中阀门循环关闭可能仅发生少数几次。

最初根据工程专业知识选取了101种异常生成设置，表A.1中列出了示例。随后，使用下一节讨论的M-距离检测器识别不可检测异常，并调整异常生成

设置。这些“不可检测”异常无法通过统计方法或人工观察识别。例如，最初约有40%的3类和8类内部泄漏无法检测。因此，我们开展了图4所示的泄漏研究，以确定每种泄漏故障的最低可检测泄漏率，并调整异常生成设置，从而改善数据质量。具体而言，随后根据异常起始时间设定泄漏率边界（见附录中的表A.1）。

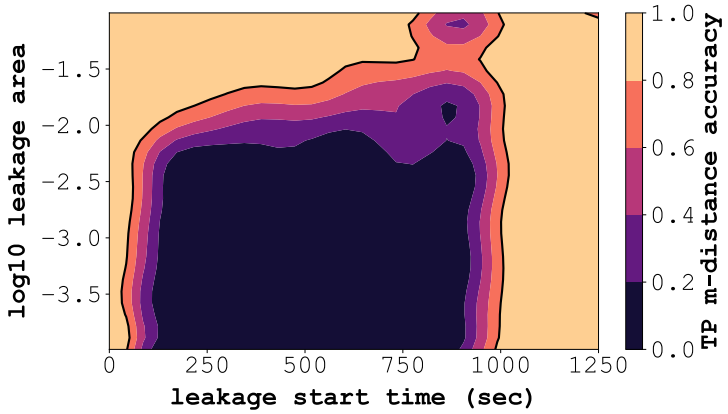


图 4. 8类 SOV-HB阀门的内部泄漏研究。

C. 马氏距离异常检测

在基线模拟异常数据集中，有若干示例中的异常过于微弱，难以检测（即与标称离散无法区分）；或者异常恢复到标称状态所需的时间与基线真实标签中的时间不同。为提高数据质量，无监督时间序列异常检测器会调整异常生成设置，并对监督分类器数据重新标注。

M距离用于衡量多元正态分布中某点到均值的标准化距离[28, 29], ，其作用类似于单变量正态分布中的 z 分数。考虑区间 $t \in [a, b]$ 内的 T 时间步，使得 x、$\mu \in R^T$ 分别表示单个参数的输入状态和均值，$\Sigma \in R^{T \times T}$ 表示协方差矩阵。均值 μ 和协方差 Σ 根据数据集中标称试验的样本指标计算得出。因此，时间窗口 $[a, b]$ 上的 M距离为:

$$d_{[a,b]} = \sqrt{(x-\mu)^T(\Sigma + \theta^2 I)^{-1}(x-\mu)|_{t\in[a,b]}} \quad (1)$$

其中 $\theta = 10^{-4}$ fo r 对于我们的数据集，并注意到，当 x 位于协方差特征基之外时，M距离按 $1/\theta$ 缩放 is.

该数据集在每个时间步有七个输入参数。作为近似，假设各参数相互独立，并取每个参数的最大M距离 p, $d_{[a,b]} = \max_p \{d^p_{[a,b]}\}$。这样可降低协方差矩阵求逆的计算成本，将矩阵尺寸从$7T \times 7T$ 减至 $T \times T$。由于协方差矩阵跨越整个时间范围，因此不仅可以检测点异常，还可以识别上下文异常或曲线形状的变化 [31] 。

由于数据具有高度非线性和斜率不连续性，对于给定概率水平$P(d < d^*) = P^*$，临界M距离并不服从高斯分布。事实上，临界M距离通常比高斯值高出30%或40%。因此，使用标称集合数据计算自适应M-距离截断值 $d^{nom}_{[a,b]}(P^*)$，概率水平为 $P^* = 0.99$。为方便起见，定义检测比例 $F_{[a,b]}$，

$$F_{[a,b]} = \frac{d_{[a,b]}}{d^{\text{nom}}_{[a,b]}(P^*)};\ ([a,b] \text{ is anom}) \leftrightarrow \left(F_{[a,b]} > 1\right) \tag{2}$$

因此，大于 1 的值为异常，小于 1 的值为标称值。

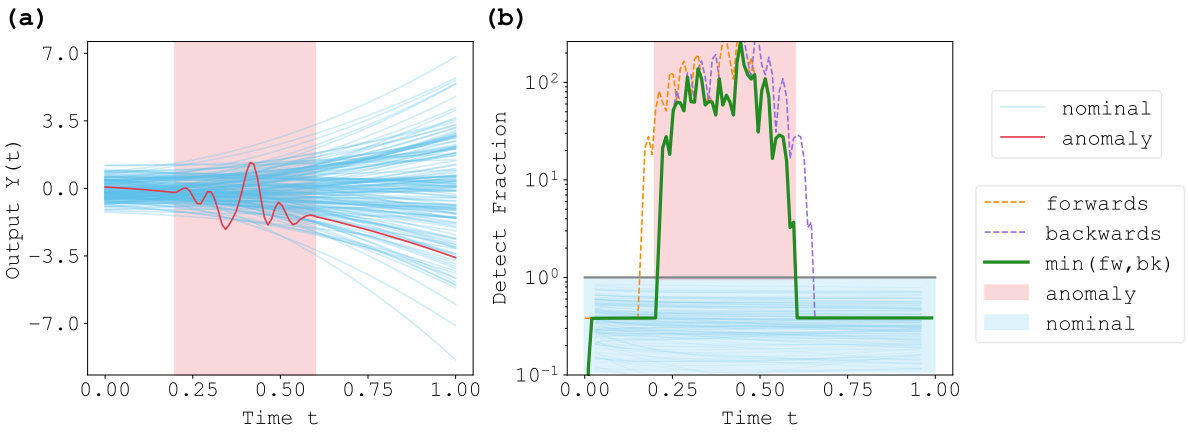


图 5. M-距离方法：(a) 参考文献**[26, 27]** 中的单个零均值高斯过程的标称数据和异常数据，如表 A.2 所示；(b) 异常数据的检测比例。

然而，检测比例仅表明时间窗口 $[a,b]$ 中是否存在异常，而不指出异常发生在哪个时间步。在向前的时间窗口（t, $t + T\Delta t$]中，可能会在实际异常事件发生前产生假阳性。类似地，在向后的时间窗口 $[t - T\Delta t, t$中，可能会在异常事件发生后产生假阳性。为减少假阳性，采用向前和向后时间窗口中的最小检测比例，

$$F^{(T)}_{\text{FB}}(t) = \min\left\{F_{[t-T\Delta t,t]}, F_{[t,t+T\Delta t]}\right\} \tag{3}$$

如图 5 所示。请注意，向前和向后的时间窗口对异常事件的起始和结束位置并不一致，因此采用可能的最短异常时间跨度。这样可以更准确地预测异常事件的开始和结束时间。

附录中图5和表 A.2的数据是根据参考文献 [26, 27] 中所述的高斯过程核函数生成的。这里选用了特殊的时间窗口核函数（见附录中的表 A.2），因此异常的精确起止时间是已知的。图 A.2 还表明，较小的检测窗口大小 T 效果更好，因为它们能提高对异常起止时间的预测准确度。在图2所示的推进系统数据集中，短

持续时间异常使用 20 个时间步窗口，长持续时间异常使用 100 个时间步窗口，点异常使用 1 个时间步窗口。因此，总体检测比例为，

$$F_{\text{ovr}}(t) = \max\left\{F_{FB}^{(1)}(t), F_{\text{FB}}^{(20)}(t), F_{\text{FB}}^{(100)}(t)\right\} \quad (4)$$

该方法在考虑多种时间尺度的同时，给出了保守的异常预测。使用公式 (4) 中的总体检测比例，M-距离方法能够识别全部 24 种异常类型。为提高检测器的运行速度，针对 100 个时间步窗口采用了 5 的步长，因此每个协方差矩阵的大小仅为 20×20。此外，对全部 15000 次试验进行多线程处理也提升了运行速度。

D. 机器学习分类工具

机器学习分类器模型的完整描述见 [9]。在 TensorFlow [24]中，使用了一个带外层 EinsumDense 层的自定义长短期记忆网络 [13, 30] 模型。模型使用长度为 100 个时间步的时间窗口，以减少长短期记忆网络中的训练权重数量。模型输入是一个由归一化输入状态构成的 3 张量 $\hat{X} \sim (N_b, N_t, N_p)$，其中 $N_b = 15000$ 表示试验次数，N_t =2721 表示时间步数，$N_p = 7$ 表示输入参数数量。模型输出是一个分类概率 3 张量，形状为 $Y \sim (N_b, N_t, N_c)$，其中 $N_c = 25$ 表示类别数。最终数据集包含 15000 次试验，具有 2.85×10^8 个训练时间步。

TensorFlow 长短期记忆网络模型采用分类交叉熵损失函数进行训练，并通过计算多分类 F1 分数进行评估 [25], , 相关方程见 [9]。此外，在参考文献 [9], 中，我们的长短期记忆网络模型已被证明对丢弃率、层数、学习率和时间窗口大小等超参数具有鲁棒性。由于训练成本较高，无法进行超参数调优，因此沿用先前研究中的相同最优超参数。具体而言，该模型包含 3 个长短期记忆网络层，每层含有[100, 75, 50] 个神经元，学习率为 3×10^{-4}，并采用长度为 100 个时间步且交错排列的时间窗口，因此仅使用最后 50 个时间步进行预测。

III. 结果

A. 机器学习模型训练

长短期记忆网络 TensorFlow 模型使用完整的地面级系统数据集进行训练，该数据集包含 15000 次试验和 25 个数据类别。每种情况下均训练 750 轮，并采用 80%-10%-10% 的训练-验证-测试集划分。训练使用两块 NVIDIA Titan RTX TU102 GPU，每块的算力为 16.31 万亿次浮点运算，精度为 32 位浮点精度。训练结果和指标如表 I 所示。未重新标记的数据集和 M-距离重新标记的数据集均分别进行了训练，每个机器学习模型约需 2.5 天完成训练。M距离重新标记通过调整每次试验的真实类别概率标签，使其与检测器相匹配。得益于

多线程、Python 即时编译和协方差步长，{v1}({v2}) 个时间步的 M距离重新标记仅用 1.486 小时便完成。表 I 显示，重新标记后训练损失和测试损失均显著降低，测试 F1 分数也提高了 12%。

表 1 - LSTM模型的训练指标

数据集	重新标记 时间，小时	训练 时间，小时	训练（测试） Loss	测试 F1 分数
未重新标记	n/a	62.849	0.512 (0.449)	0.710
已重新标记	1.486	61.075	0.165 (0.326)	0.831

M距离重新标记会调整监督分类器的真实标签，使其与检测器对异常和标称状态的预测相匹配。图6展示了一个示例，其中类别2和3类故障的基线真实标签与实际异常持续时间不符。例如，在2类故障开启异常中，阀门仅失效开启了1.3秒（时间非常短），但较高的系统残差在仿真剩余时间内一直持续。尽管阀门已不再处于失效状态，这些较高的系统残差仍处于标称分布之外，因此被M-距离检测器标记为异常。类别2和3类故障在阀门故障期间可以区分，但故障发生后却非常相似，此时压力塔架和温度塔架残差缓慢下降。因此，对于具有相似高系统残差特征的异常试验，手动调整了长短期记忆网络模型的F1分数，使阀门故障后产生相似的类别预测。通常，同一部件的故障开启、内部泄漏和缓慢开启异常在故障发生后的残差特征也相似。

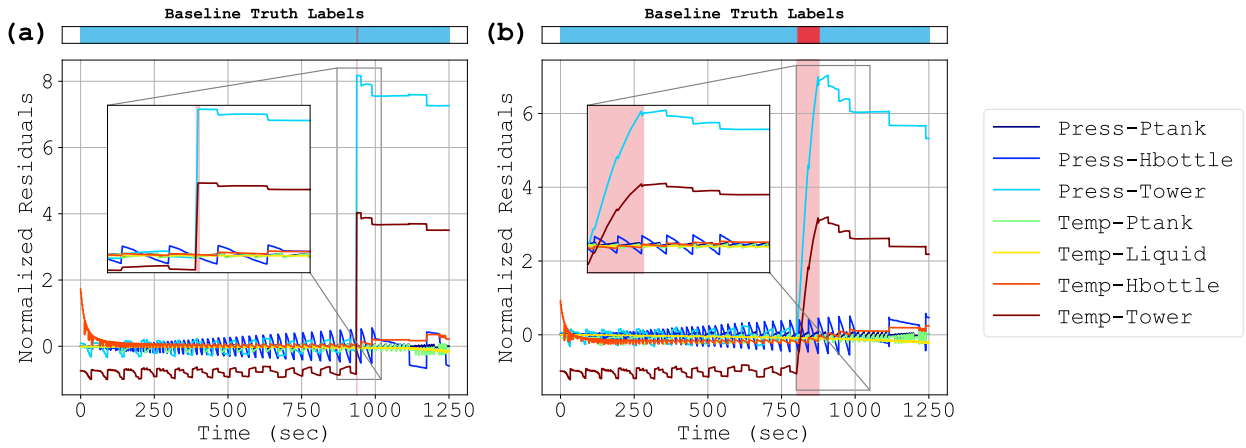


图6.（a）类别2失效开启和（b）3类内部泄漏的残差状态。阀门故障期间的异常特征有所不同，但故障发生后，残差较高，表现相似。

表II展示了“未重新标记”“重新标记”和“重新标记 + 已修正”数据集在测试数据上的总体指标。这里，“重新标记”表示已执行 M距离重新标记，而“已修正”表示针对故障后

类别预测所做的校正。结果表明，重新标记和故障后校正显著提高了各类别的准确率。精确率提高了 7%，表明假阳性减少；召回率提高了 22%，表明假阴性减少。因此，重新标记的最大作用似乎是通过扩展异常来减少假阴性。此外，对于已重新标记 + 并修正的情况，随着数据集大小增加，测试 F1 分数提高了 9%。

据推测，随着试验次数增加，F1分数也会提高，因为7500次试验可能无法完全覆盖异常空间——考虑到我们的数据中有101种异常生成设置和3个发射阶段。

表 2 - 按数据集划分的 LSTM模型测试数据指标

试验次数	数据类型	精确率	召回率	F1
7500	未重新标记	0.741	0.522	0.613
	已重新标记	0.815	0.733	0.772
	已重新标记+已更正	0.828	0.759	0.792
15000	未重新标记	0.850	0.665	0.703
	已重新标记	0.879	0.820	0.833
	已重新标记+已校正	**0.916**	**0.858**	**0.876**

经过 M距离重新标记和故障后校正后，最终的混淆矩阵如图7所示，所用测试数据集包含 1500 次试验（占数据的 10%）。该混淆矩阵对应的 F1分数为 88%，精确率为 92%，召回率为 86%。较高的精确率表明假阳性极少。较低的召回率表明假阴性多于假阳性。通过使用 M距离调整更多异常生成设置并增加试验次数，很可能大幅减少第 0 列中剩余的假阴性。

B. 测试数据示例应用

接下来，考虑三个示例，以展示 M距离重新标记对类别概率预测的影响：(1) 图8展示了一次试验，其系统残差较高，并且在阀门故障后仍持续存在；(2) 图9展示了一个永久故障的阀门，其仅在阀门动作时间内对系统产生影响；(3) 图10展示了另一次试验，其故障后系统残差较高，因此需要对图6和表II中的分类错误进行人工修正。

首先，在图8中展示了LRV的21类慢开异常。阀门失效持续1.47秒，导致推进剂贮箱压力升高，并经过31秒后恢复至标称值。M-距离方法准确预测了压力恢复至标称值所需的时间。此外，与未重新标记的情况相比，已重新标记情况下的LSTM模型分类特征具有更少的假阳性和假阴性，使人工观察者更容易识别该异常。

图9展示了 SOV-PT 1和2 的14类慢开异常。在整个仿真过程中，阀门都被施加了该故障模式；然而，只有在阀门测试和发射前加压发射阶段的阀门动作时间内，才会表现出偏离标称行为

。当异常不属于标称族数据时，M距离对偏离标称行为的预测结果一致。在此示例中，无论是未重新标记的情况（图9c）还是已重新标记的情况（图9d），都正确预测出了异常。不过，已重新标记的机器学习模型（图9d）假阳性更少，对工程师而言也更可靠。此外，如果没有进行重新标记，一些处于标称分布内的训练时间步就会被错误地标记为异常，从而污染数据集，导致机器学习预测效果变差。

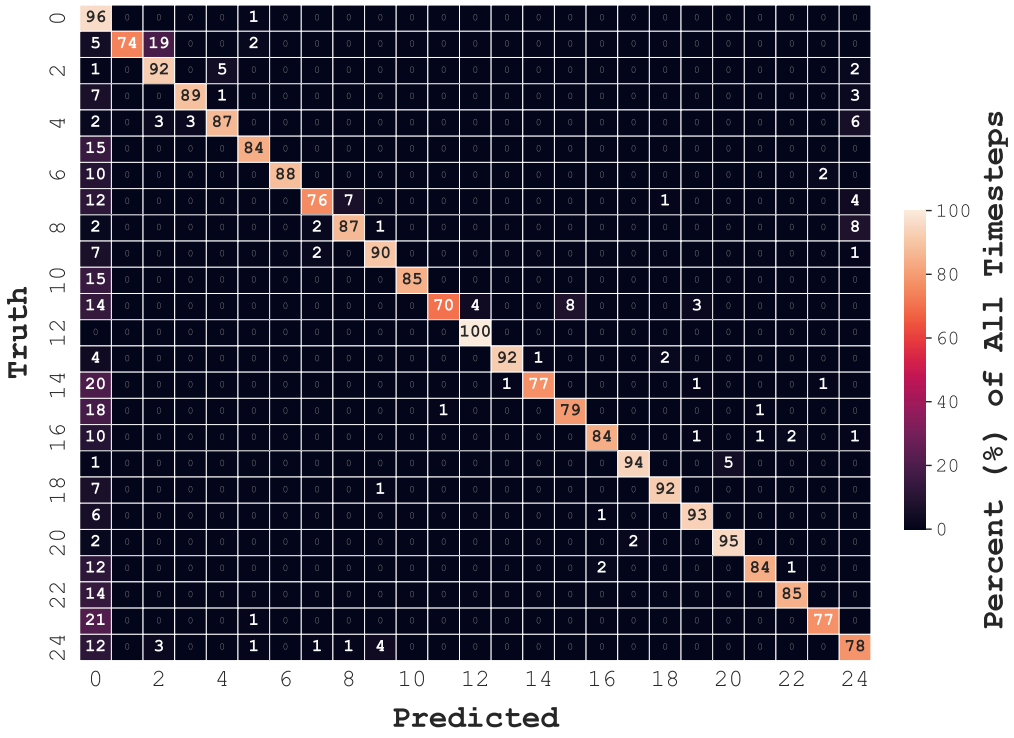


图7. 经M距离重新标记和故障后人工校正后的最终混淆矩阵。

图10展示了一个2类故障开启异常SOV-TOW。阀门仅故障开启了1.23秒，但这导致塔压升高，并且在仿真剩余时间内未恢复至标称值。如图6所示，阀门故障开启后，类别2和3类故障具有相似的异常特征，且系统残差较高。因此，在这次试验中对分类器错误进行了人工修正，使其在阀门故障后接受3类预测。在图8至图10的三次试验中，M距离重新标记准确预测了偏离标称行为，并降低了LSTM模型的假阳性率和假阴性率。这表明，提升数据质量对于开发实用的运载火箭推进系统自动异常识别工具至关重要。

IV. 结论

结合机器学习与物理学的　　我们在先前工作中开发的工具 [8, 9] 被应用于以下对象的异常分类　　a

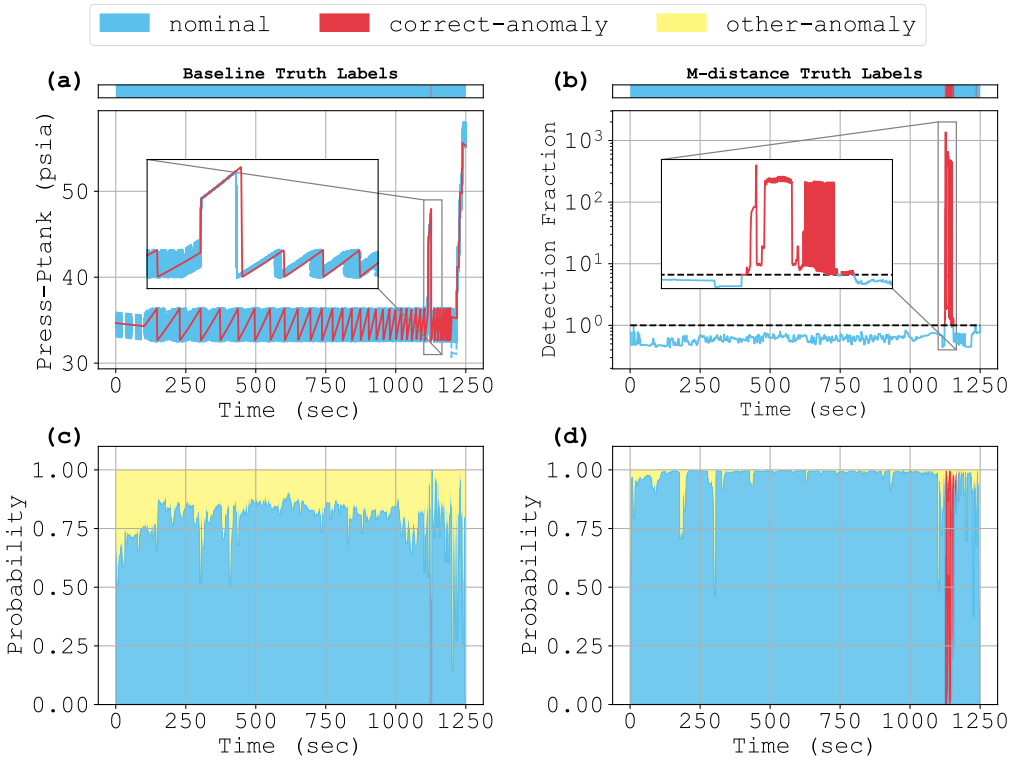


图8。LRV阀的缓慢开启异常，其中（a）为原始传感器数据，（b）为M-距离检测，（c）和（d）分别为重新标记前后的机器学习预测。

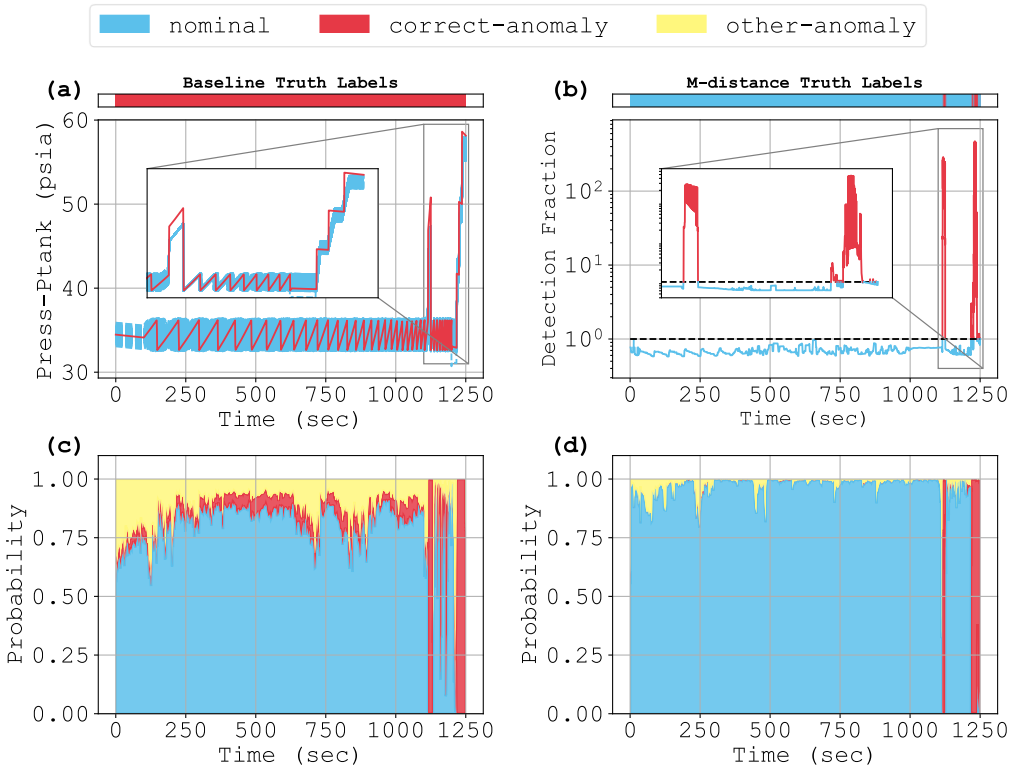


图9。SOV-PT1,2阀的缓慢开启，其中（a）为原始传感器数据，（b）为M-距离检测，（c）和（d）分别为重新标记前后的机器学习预测。

大规模地面至运载器增压系统，其压力带阀门更贴近现实，仿真时间也比之前的测试案例 [9] 更长。训练数据集包含超过 4 亿个时间步、24 种失效模式、每次试验 20.8 分钟的实时仿真，以及各种短时间尺度和长时间尺度的模拟异常。共采用 101 种异常生成设置，为每个模拟异常生成不同强度、持续时间和起始时间的异常。发现机器学习模型在基线标签上的表现不佳后，我们开发了一种 M-距离无监督时间序列检测方法，用于修正监督训练标签。我们还使用 M-距离方法改进异常生成设置，确定最小可检测泄漏面积。结果表明，M-距离方法使我们的长短期记忆网络模型的测试 F1 分数提高了 12%；若对可接受的类别混淆进行调整，则提高了 16%。在若干示例试验中，M-距离方法正确识别出异常在哪些发射前阶段可被检测到，还能够延长具有较高系统残差的异常，并找到最小可检测异常强度。

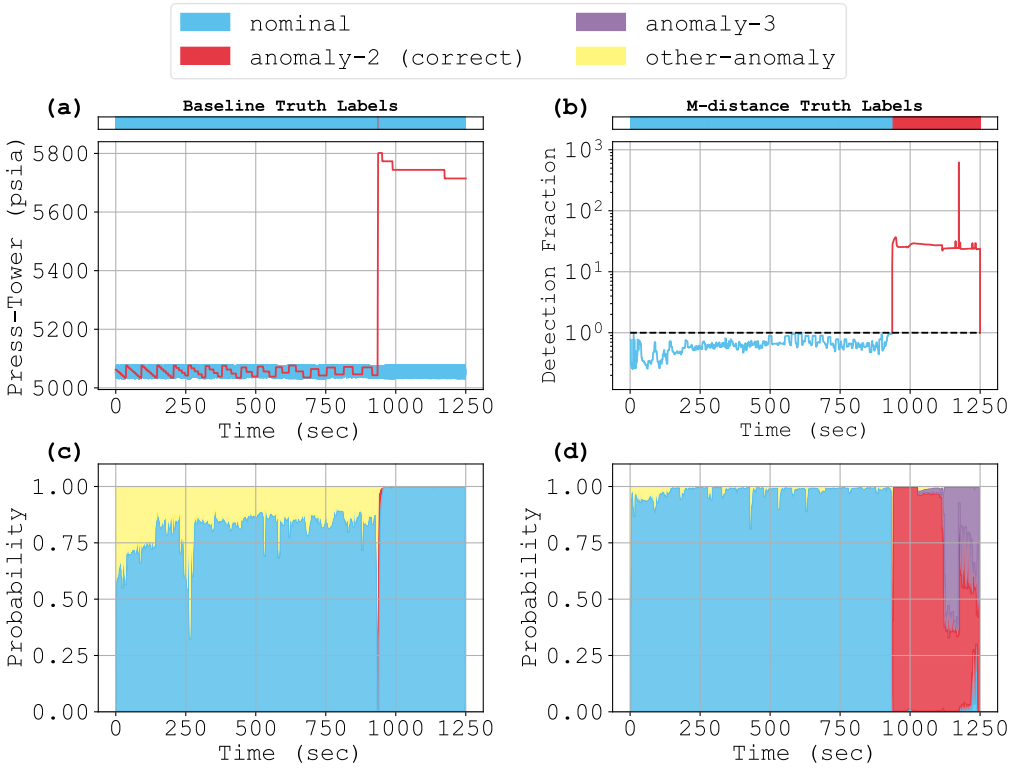


图 10. SH-TOW 阀门的故障开启异常：(a) 原始输入数据，(b) M-距离检测，(c) 和 (d) 重新标记前后的机器学习预测。

M-距离方法显著提高了监督训练数据的质量，降低了长短期记忆网络模型的假阳性率。此外，在使用重新标注并经过人工校正的数据训练最终长短期记忆网络模型后，F1 分数达到 88%，表明该分类器模型具有良好的效果。这说明，提高数据质量对于开发适用于运载火箭推进系统的实用自动异常识别工具至关重要。

A. 实用经验总结

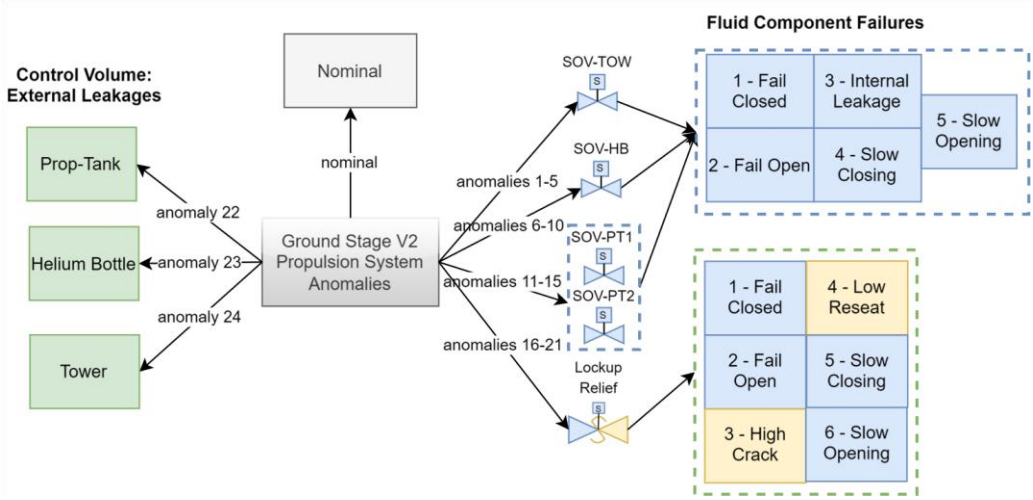
以下是我们从这项工作中总结出的实践经验：

1. 应使用 M-距离方法，根据时间确定最小可检测泄漏面积。泄漏面积两端的压差会发生变化；如果泄漏面积保持恒定，在某些阶段，压差可能过小，无法产生可检测的泄漏率。这样可以确保在训练机器学习模型时纳入适当的可检测范围。

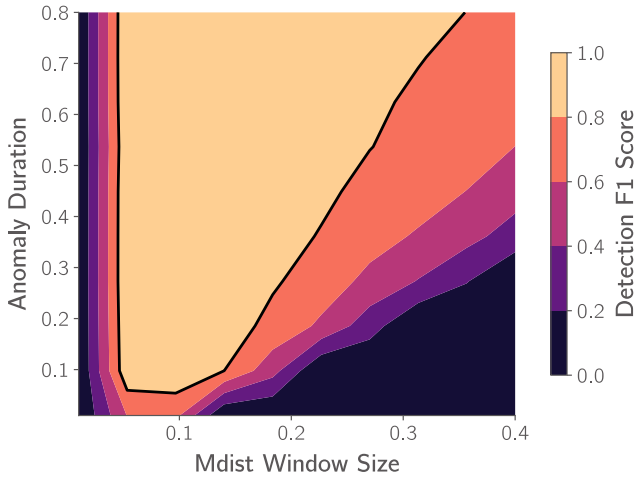
2. 数据集大小对于获得良好的 M-距离和机器学习模型训练结果非常重要。研究发现，每种异常采用 300 次试验是数据集规模的良好起点，但在本数据集中，每种异常需要 600 次试验才能获得更高的准确率。M-距离误报可能表明存在欠采样，需要增加试验次数。3. 如果某个输入参数的标称变化过小，阀门脱离故障状态后，M-距离方法仍可能报告弱异常；而在发射过程中，这些情况不会被视为异常。这些弱异常会在重新标记后降低机器学习模型的性能。因此，为避免此类错误，控制容积的初始压力和温度，以及裂纹压力和回座压力，均选取了足够大的标准差。绘制标称集合数据并检查图示区域的覆盖是否足够密集，会很有帮助。

4. M-距离方法可以通过延长那些在阀门不再处于故障状态后仍产生持续偏离标称状态特征的异常，来修正异常类别标签。然而，阀门不再处于故障状态后，几种会产生较大残差的异常看起来完全相同，机器学习模型会将它们混淆；事实上，即使是人也无法区分这些异常。例如，同一部件上的失效开启、内部泄漏和缓慢关闭异常，在阀门故障后具有相同的特征。实际应用中，人工监视员可以依据最先报告的异常类别，因为它最有可能对应异常处于活动状态的时段。人工专业知识也有助于判断哪些异常会产生相似的异常后特征，从而辨别故障的真实原因。

附录



图A.1. 图3中地面级推进系统可用的异常。



图A.2. GP核数据的M距离F1分数随检测窗口大小和异常持续时间的变化。

表A.1 - 随机异常生成设置，其中 $\boldsymbol{H(t)}$ 是 t>**1000** 秒时的赫维赛德函数

	SOV-TOW	SOV-HB	SOV-PT1,2	LRV
失效开启分数	[0.1, 0.3]	[0.1, 0.3]	[0.1, 0.3]	[0.1, 0.5]
缓慢开启延迟（秒）	[0.1, 0.25]	[0.1, 0.25]	[0.025, 0.1]	[1.0, 3.0]
缓慢关闭延迟（秒）	[5, 10]	[1.0, 2.5]	[1.5, 3.0]	[5.5, 7.5]
内部泄漏分数	$\begin{bmatrix} -2.8 - 1.2H(t), .8 - 1. \\ -1 \\ 2H(t) \end{bmatrix}$ 以10为底的对数	$\begin{bmatrix} -2.3 - .7H(t) .7 - . \\ -1 \\ 3H(t) \end{bmatrix}$	[-2.52, -2.0]	---

表 A.2 – 零均值高斯过程的单一输出，例如 $\boldsymbol{Y_{nom}(t) \sim N(0, k_{nom}(t, t'))}$，使用参考文献 **[26, 27]** 中的时间窗口核。

核	表达式
$k_{nom}(t, t')$	$0.3 + 0.4tt' + 3(tt')^2 + t(t')^2 + t'(t)^2$
$W_{[a,b]}(t, t')$	$relu(1 - \left\|\frac{2t - (a+b)}{b-a}\right\|) \cdot relu(1 - \left\|\frac{2t' - (a+b)}{b-a}\right\|)$
$k_{anom}(t, t')$	$k_{nom}(t, t') + 0.3 \cdot W_{[a,b]}(t, t') \cdot \exp[-2 \cdot sin^2(10\pi\|t - t'\|)]$

致谢

本研究由 The Aerospace Corporation 的内部研究与开发经费资助。推进这项技术的工作未使用其他资金来源。

参考文献

[1] Guarro, S.,《运载火箭可靠性与风险度量指标的定义和估算及其与需求的关系》，《概率安全评估与管理》PSAM 14，洛杉矶（加利福尼亚州），2018年。[2] Williams, C. H., Ross, L. J., and Nieberding, J. J.,《基于人为因果因素对概念系统进行运载火箭故障概率预测评估的方法论》，2018年AIAA SPACE与航天论坛及展览会，奥兰多（佛罗里达州），2018年。https://doi.org/10.2514/6.2018-5413[3] Rising, J. M., and Leveson, N. G.,《航天运载火箭的系统理论过程分析》，《空间安全工程杂志》，第5卷，第3-4期，2018年，第153–183页。https://doi.org/10.1016/j.jsse.2018.06.004[4] Morse, E. L., Putney, B. F., and Fragola, J. R.,《可靠性增长与平均值的注意事项：半人马座案例研究》，《2011年可靠性与可维护性年度研讨会论文集》，IEEE，布埃纳维斯塔湖（佛罗里达州），2011年。https://doi.org/10.1109/RAMS.2011.5754433[5] Paté-Cornell, M. E., and Guikema, S. D.,《航天运载火箭“初期问题”的概率分析》，《概率安全评估与管理》，施普林格，柏林（德国），2004年，第2193–2198页。https://doi.org/10.1007/978-0-85729-410-4_353[6] Arnold, T. M.,《航天飞机主发动机（SSME）红线限值管理史》，AIAA SPACE 2011会议及展览会，长滩（加利福尼亚州），2011年。

[7] Huang, Z., Aguilar, R., Larkin, J., Swider, J., and Rocketdyne, A.,《基于可靠性的健康监测红线投票逻辑设计》，2014年可靠性与可维护性研讨会，IEEE，科罗拉多斯普林斯（科罗拉多州），2014年。https://doi.org/10.1109/RAMS.2014.6798507[8] Darr, S., Engelstad, S., Knighton, T., Taliaferro, M., and Goyal, V.,《用于推进系统异常识别的机器学习与物理相结合工具》，AIAA 2023-2375。2023年1月。https://doi.org/10.2514/6.2023-2375[9] Engelstad, S., Darr, S., Knighton, T., Taliaferro, M., Goyal, V., and Kennedy, G.,《用于推进系统异常识别的机器学习与物理相结合工具》，《航天器与火箭》，第61卷，第5期，2024年，第1219-1232页。https://doi.org/10.2514/1.A35897[10] Wu, Y., Sicard, B., and Gadsden, S.,《物理信息机器学习：异常检测与状态监测应用综述》，《应用专家系统》，第255卷，2024年。https://doi.org/10.1016/j.eswa.2024.124678[11] Wang, F., Jiang, Y., Zhang, R., Wei, A., Xie, J., and Pang, X.,《多变量时间序列深度异常检测综述：分类、应用与发展方向》，《传感器》（巴塞尔），第25卷，第1期，2025年。https://doi.org/10.3390/s25010190[12] Boniol, P., Liu, Q., Huang, M., Palpanas, T., and Paparrizos, J.,《深入探究时间序列异常检测：十年回顾》，arXiv:2412.20512 [cs.LG], 2024.[13] Hochreiter, S., and Schmidhuber, J.,《长短期记忆》，《神经计算》，第9卷，第8期，1997年12月，第1735-1780页。DOI: 10.1162/neco.1997.9.8.1735。[14] Vaswani, A., Shazeer, N., Parmar, [illegible] 第31届神经信息处理系统国际会议论文集，2017年。DOI: 10.5555/3295222.3295349。[15] Chao, M., Kulkarni, C., Goebel, K, and Fink, O.,《混合式深度故障检测与隔离：结合深度神经网络和系统性能模型》，[illegible] 第4期，2023年。https://doi.org/10.36001/ijphm.2019.v10i4.2621[16] Chen, Y., Liu, J., Peng, J., Wu, Y., Xu, Y., and Zhang, Z.,《自编码变分贝叶斯》，《剑桥艺术与科学探索》，第2卷，第1期，2024年。

[17] Guc, F., and Chen, Y., “基于物理-信息迁移学习框架的传感器故障诊断”, 《传感器》, 第 22 卷, 第 8 期, 第 2913 页, 2022 年。https://doi.org/10.3390/s22082913[18] Huang, Y., Liu, P., Han, R., Yin, T., Dou, Q., and Song, Y., “FreqWave-TranDuD：一种基于小波和傅里叶变换的多变量时间序列异常检测方法”,

Stemocer. A.. and Thoemel. I.. “一种在尊重立方体卫星机载资源限制的前提下使用深度学习进行异常检测的方法”, 《航空航天信息系统期刊》, 第 20 卷, 第 12 期, 第 859-872 页, 2023 年。

[20] Pang, J., Liu, D., Peng, Y. 和 Peng, X., 《用于多变量航天器遥测序列异常检测的时间依赖马氏距离》《ISA Transactions》, 第 140 卷, 354–367 页, 2023 年。https://doi.org/10.1016/j.isatra.2023.06.002

[21] Hou, Y., Chen, Z., Wu, M., Foo, C.S., Li, X. 和 Shubair, R., 《基于马氏距离的用于异常检测的对抗网络》IEEE 国际声学、语音与信号处理会议, 2020 年。https://doi.org/10.1109/ICASSP40776.2020.9053206

[22] Hu, Z., Xiao, M., Zhang, L., Liu, S. 和 Ge, Y., 《利用频率特征和 Parzen 窗密度估计的基于马氏距离的模拟滤波器异常检测方法》, 《Journal of Electronic Testing》, 第 32 卷, 681–693 页, 2016 年。https://doi.org/10.1007/s10836-016-5623-z

[23] Taliaferro, M. 和 Darr, S., 《运载火箭内部流体流动与热力学建模》, AIAA Scitech论坛2024, https://doi.org/10.2514/6.2024-2293。[24] Abadi, M., Barham, P., Chen, J., Chen, Z., Davis, A., Dean, J., Devin, M., Ghemawat, S., Irving, G., Isard, M., Kudlur, M., Levenberg, J., Monga, R., Moore, S., Murray, D., Steiner, B., Tucker, P., Vasudevan, V., Warden, P. Wicke, M., Yu, Y. 和 Zheng, X., 《TensorFlow：一种用于大规模机器学习的系统》, 第12届USENIX操作系统设计与实现研讨会（OSDI 16）, 美国佐治亚州萨凡纳, 2016年, 第265–283页。DOI：10.5555/3026877.3026899[25] Opitz, J. 和 Burst, S., 《宏平均F1和宏平均F1》, 2021年。https://doi.org/10.48550/arXiv.1911.03347[26] Engelstad, S., Burke, B. 和 Kennedy, G., 《利用机器学习改进飞机机翼的屈曲预测，以实现高效结构优化》, AIAA Scitech 2024, 2024年。https://doi.org/10.2514/6.2024-3981

[27] Rasmussen, C. 和 Williams, C.，《机器学习中的高斯过程》，麻省理工学院出版社，2006年。ISB N 0- 262-18253-X。[28] Garthwaite, P., Elfadaly, F. 和 Crawford, J.，《马氏距离的修正置信区间》，《统计与概率快报》，第127卷，第131-137页。https://doi.org/10.1016/j.spl.2017.03.029[29] Reiser, B.，《马氏距离的置信区间》，《统计通讯》–仿真与计算，第30卷，第1期，第37-45页，DOI: 10.1081/SAC-100001856。[30] Gers, F. A., Schmidhuber, J. 和 Cummins, F.，《学习遗忘：使用长短期记忆网络进行持续预测》，《神经计算》，第12卷，~~第10期，2000年，第2451–2471页。~~ https://doi.org/10.1162/089976600300015015[31] Foorthuis, R.（2021）。异常的本质与类型：数据偏差综述。《国际数据科学与分析期刊》，12(4)，297-331。https://doi.org/10.1007/s41060-021-00265-1 [32] Memarzadeh M., Matthews B. 和 Weckler D.，《利用深度学习检测飞行运行数据中的异常》，NASA技术报告，2023年7月18日。