arXiv:2601.01357v1 [cs.LG] 4 Jan 2026

迈向大语言模型赋能的自主燃烧研究：用于自我纠正建模工作流的文献感知智能体

KeXiao[a,b], Haoze Zhang[a],RunzeMao[a],HanLi[a,b,*],ZhiX.Chen[a,b,*]

[a]北京大学力学与工程科学学院湍流与复杂系统国家重点实验室，中国北京 *100871* [b]科学智能研究院（*AISI*），中国北京 *100080*

摘要

大语言模型（LLM）的迅速发展正推动人工智能转变为自主研究伙伴，但在燃烧建模等复杂科学领域，关键的能力缺口依然存在。在这些领域，要提供切实可行的人工智能辅助，必须将领域文献知识与稳健的执行能力无缝结合，以支持计算流体力学（CFD）代码等需要专业知识的工具。为弥补这一缺口，我们推出 FlamePilot，这是一种旨在通过自动化、自我纠正的计算流体力学工作流赋能燃烧建模研究的大语言模型智能体。FlamePilot 的独特之处在于采用原子工具架构，确保能够在 OpenFOAM 及 DeepFlame 等扩展框架中稳健地设置并执行复杂数值模拟。该系统还能够从科学文献中学习，提取关键信息，指导数值模拟从初始设置推进到结果优化。在一项公开基准测试上的验证表明，FlamePilot 的可执行性得分达到满分 1.0，成功率为 0.438，超过了此前报告的智能体最佳得分，分别为 0.625 和 0.250。此外，针对中度或强烈低氧稀释（MILD）燃烧数值模拟开展的详细案例研究，展示了该系统作为协作研究副驾驶的效能：FlamePilot 自主将一篇研究论文转化为配置完成的数值模拟，执行模拟、对结果进行后处理、提出基于证据的改进，并管理多步参数研究直至收敛，期间只需极少的人工干预。通过采用透明且可解释的范式，FlamePilot 为人工智能赋能的燃烧建模奠定了基础框架，促进形成协作伙伴关系：智能体负责工作流编排，使研究人员能够专注于高层次分析。

关键词：大语言模型智能体；用于研究的大语言模型；燃烧建模；计算流体力学自动化

*通讯作者。

1. 引言

大语言模型（LLM）的快速发展正显著影响科学研究，推动人工智能（AI）从计算辅助工具转变为自主研究伙伴[1, 2]。通过自动化假设生成、实验设计、执行和分析等认知任务，由大语言模型驱动的人工智能智能体系统能够加速研究进程，使研究人员得以投入更多精力进行创造性和高层次思考。

尽管针对通用研究背景 [3, 4] 以及数学[5], 、化学 [6], 和生命科学 [7], 等专业领域开发此类代理已取得显著进展，但燃烧研究界至今仍缺乏一个实用的且易于使用的AI代理系统。这样的系统必须将领域特定知识与稳健的执行能力相结合，才能有效处理 OpenFOAM [8]等专业工具（用于计算流体力学（CFD）应用）以及Cantera[9]（用于化学动力学）。

虽然 MetaOpen-FOAM [10–12] 和其他框架 [13–19] 已推动了 OpenFOAM 工作流自动化的发展，但它们仍不足以实现广泛应用。基准测试研究表明，这些框架在复杂问题上的成功率较低 [20], ，说明其执行可靠性不足。此外，它们依赖特定版本的 OpenFOAM，限制了实用性，也妨碍了与定制研究求解器的兼容。更重要的是，当前大多数智能体无法自主整合科学文献中的知识

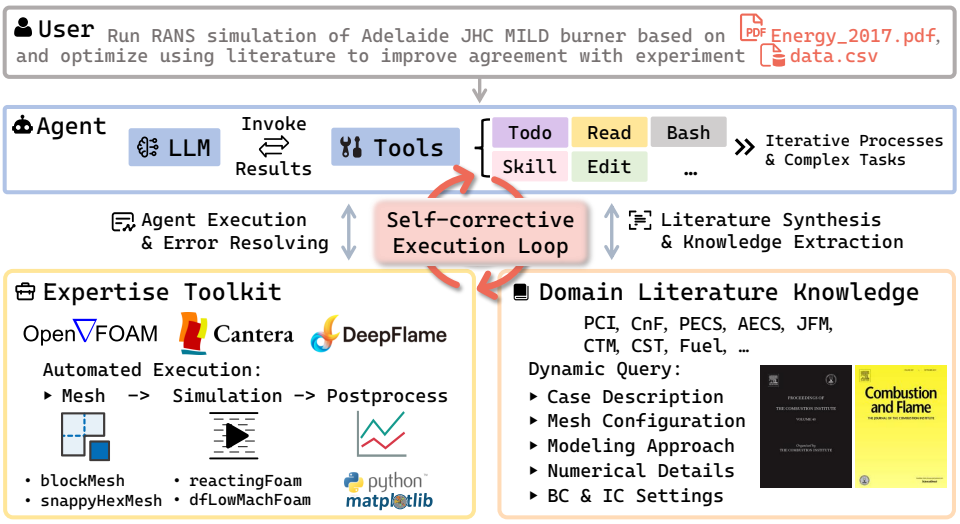

图 1：FlamePilot 智能体架构示意图，展示了大语言模型推理、基于工具的执行以及动态领域知识理解如何集成，以实现自动化计算流体力学工作流。

，而是依赖用户提示来注入关键领域背景。这种无法利用基础知识来源的能力，阻碍了它们在实际研究工作流中的有效应用。

本研究开发了 FlamePilot，这是一款用于燃烧建模、由大语言模型驱动的研究助手，可将领域知识与稳健的 OpenFOAM 执行能力相结合。该系统作为由人类监督的协作伙伴，利用执行反馈和科学文献采取自我纠正措施，以确保结果在物理上准确且可复现。这一架构优先考虑可验证性和可解释性，以奠定可信赖的 AI 辅助所需的科学严谨性。受编码智能体启发，FlamePilot 通过文件操作和命令行执行处理 OpenFOAM 工作流，并支持 DeepFlame 等标准及扩展框架 [21–23]。在一个中度或强烈低氧稀释（MILD）燃烧案例上的实证验证证实了它作为研究副驾驶的有效性，并在自动化计算流体力学工作流中始终保持可验证性和可解释性。对可验证性和可解释性的重视，确立了建立人们对 AI 助手的信任并在计算科学中实现可复现结果所必需的科学严谨性。

2. 智能体系统与架构

FlamePilot 将成熟的命令行界面（CLI）编码智能体范式 [24–27]应用于 OpenFOAM 工作流，从而无缝融入工作流程，充当“研究副驾驶”。如图 1 所示，其架构包含三个核心组件：负责工作流编排的中央大语言模型、用于仿真控制的专业工具包，以及用于提供领域知识的文献综合模块。这种统一设计既支持自动化执行，也能提供基于知识的辅助。

不同于复杂的多智能体系统，FlamePilot务实的单智能体架构优先考虑可执行性和透明度。它运用原子工具和专用技能，实现自主错误纠正，同时确保研究人员全程监督。通过让所有决策都可查阅，并依据动态提取的文献开展数值模拟，这种方法确保了不透明自动化所缺乏的科学严谨性和可重复性。

2.1. 工具集成

科学计算中的大语言模型智能体要有效运行，需要灵活多样的工具集成；现有计算流体力学智能体受限于专业化的单体功能，无法满足这一需求。FlamePilot 通过一套遵循成熟编码智能体范式的原子工具来解决这一问题，其中包括文件读写、目录列表、grep 搜索和 bash 执行。这套细粒度工具集支持动态组合，以处理复杂的计算流体力学任务：配置字典、执行求解器，并通过日志分析诊断错误。

为从 OpenFOAM 教程算例中检索知识，FlamePilot 采用直接文件系统搜索，而非静态检索增强生成，并利用 OpenFOAM 的键值对结构，在教程目录中高效进行模式匹配。工作流编排由任务管理系统负责，提供任务分解、规划和进度跟踪，从而实现对多阶段模拟的自主管理。这种受先进编码智能体启发的推理机制 [24, 25], 使 FlamePilot 能够自主管理复杂的多阶段模拟，同时保留透明且可审计的执行路径，便于研究人员进行必要的监督。

2.2. 领域知识集成

有效的科学推理需要将领域知识融入基础工作流执行之外。FlamePilot 通过一个结构化的

“技能”系统实现这一点，该系统受到 Claude 的 Agent Skills[24], 启发，能够动态增强智能体的能力，赋予其专业知识。

系统将知识组织为模块化单元，其中包含指令、脚本和资源，智能体会按需发现并加载这些内容。我们为 OpenFOAM 及其衍生框架 DeepFlame 实现了专用技能。DeepFlame 技能专门向智能体提供有关该框架如何从 OpenFOAM 扩展而来的知识，包括其独特的配置规范和求解器要求。这种模块化方法使系统能够使用相同的基础工具集在不同的仿真环境中运行，无需针对特定工作流程开发专用工具。

除了操作知识，FlamePilot 还具备研究分析能力，使其决策以公认的科学发现为依据。专门的论文分析技能会指导智能体将学术 PDF 转换为 Markdown 格式，并提取关键的计算流体力学参数。这些参数包括几何结构和网格细节、模型选择，以及以往研究中记载的具体调优参数。通过系统地将原始文献转换为结构化、可查询的格式，这项技能弥合了已发表研究与实际配置之间的差距，使智能体能够依据科学文献库中的循证背景信息做出计算流体力学决策。

3. 结果与讨论

3.1. 基础：*FlamePilot* 在计算流体力学任务中的核心能力

我们使用 CFDLLM-Bench 套件中的 FoamBench 高级部分，评估 FlamePilot 配置和执行复杂 OpenFOAM 案例的能力 [20]。该基准测试包含 16 个由专家设计的 OpenFOAM 案例，要求将自然语言描述转化为复杂操作，包括湍流模型选择和网格生成。

我们仅向 FlamePilot 提供测试查询，并允许其访问本地 OpenFOAM 教程目录，未给予任何额外的人工指导。我们在相同基准测试上将其性能与其他计算流体力学智能体进行了比较，结果总结于图 2。有关评估指标的详细说明，请参见 [20]。

FlamePilot 获得了满分的可执行性得分（M_{exec}），在全部 16 项测试中均成功生成了可执行的 OpenFOAM 案例。这一成绩显著优于此前报告的最佳基准测试得分 0.625，并直接验证了我们的架构设计；原子工具集成和单智能体执行模型的稳健性使灵活的错误纠正和快速迭代成为可能，从而满足了实现这一性能所需的要求。

相比之下，虽然成功率为 0.438，基于NMSE的准确度（M_{NMSE}）为 0.469，也超过了此前最先进水平记录的 0.250 和

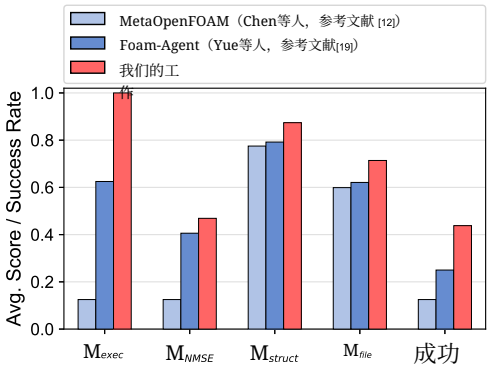

图2：智能体在FoamBench-Advanced [20] 案例中的性能。FlamePilot（我们的工作）实现了完美的可执行性（M_{exec} = 1.0），展现出卓越的操作稳健性。MetaOpenFOAM [12] 和 Foam-Agent [19] 的性能指标基于 [20]中报告的数值。

0.406，但绝对值仍然偏低。这种差异源于 M_{NMSE} 指标会逐点比较标量场，即使解在物理上合理，也可能因网格拓扑、边界条件设置位置或收敛标准略有差异而产生较高误差。我们逐个案例进行的人工分析验证了 FlamePilot 所生成模拟流场的物理合理性，并发现了若干具体案例：在这些案例中，FlamePilot 使用 snappyHexMesh 处理复杂几何体所生成的网格优于基准测试的参考网格。

这些结果确立了FlamePilot配置和执行多种计算流体力学模拟的核心能力。尽管FoamBench主要关注非反应流，但其在基础OpenFOAM操作中展现出的能力为其应用于更复杂的MILD燃烧建模领域奠定了必要基础，下一节将对此进行展示。

3.2. 案例研究：*FlamePilot*作为*MILD*燃烧的交互式研究助理

本案例研究通过将 FlamePilot 应用于 MILD燃烧，展示了它作为交互式研究助理所发挥的变革性作用。这项研究展现了 FlamePilot 如何超越简单的自动化，成为能够管理复杂迭代工作流程的协作伙伴，并以科学文献为依据。

研究工作始于一位研究人员提供了一篇关于 MILD燃烧 的论文 [30]，并要求复现其基准案例。计算域如图3a所示。FlamePilot 自动处理 PDF，提取几何、边界条件、湍流-化学相互作用模型、化学动力学和求解器规格等关键参数，随后将这些信息转化为一套完整配置且可执行的

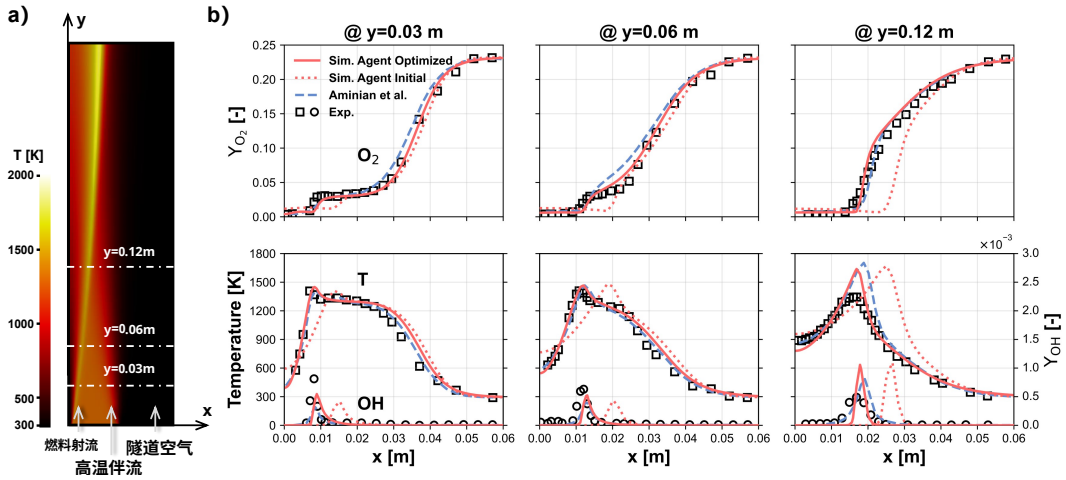

图 3：阿德莱德 JHC MILD 燃烧器的数值模拟结果对比：(a) 二维计算域示意图（未按比例绘制）；(b) 标量分布展示了从初始数值模拟到经过文献指导优化后的最终结果的改进情况，表明其与实验数据 [28]的吻合度显著提高。从初始设置到最终优化的整个过程仅需数小时，其中仿真运行时间占据主要部分。优化后的结果准确度与文献中已有的成熟数值模拟结果相当 [29]。

DeepFlame 案例。此流程建立了从已发表研究到实际运行模拟的直接路径，大幅缩短了从文献综述到获得初步计算结果所需的时间。

当初始数值模拟结果与实验数据 [28] 存在显著差异时（图 3b），FlamePilot 会分析所提供的文献 [29, 31, 32]，据此生成基于证据的改进建议。研究人员选择调整 k-ε模型常数（$C_{1\varepsilon}$），并针对入口湍动能开展参数研究，同时通过修改预配置 blockMeshDict 中的两个参数，手动改进网格质量。随后，FlamePilot 自主执行完整工作流程：实施模型更改、运行参数研究、处理数据并生成可视化结果。

从初始设置到最终优化的完整过程仅需数小时，其中数值模拟运行时间占据了主要部分。优化后的结果与实验测量值的吻合度显著提高，准确性达到了文献中已有数值模拟的相当水平 [29]。这一成功的改进验证了 FlamePilot 基于证据进行优化的能力，同时凸显了其作为计算研究效能倍增器的作用。在整个过程中，研究人员通过提供背景指导、分析结果以确定科学方向，并就模型调整和网格细化作出关键决策来保持监督。FlamePilot 则负责处理工作流编排、后处理以及受文献启发的假设生成等繁重工作。

4. 结论

本文介绍了 FlamePilot，这是一款大语言模型智能体，通过将深厚的领域知识与可靠的执行能力相结合，弥合了燃烧建模中的关键缺口，适用于 OpenFOAM 等复杂工具。FlamePilot 的架构展示了实现自动化、自我纠正的计算流体力学工作流的一条切实可行的路径。

实证验证证实了 FlamePilot 的能力：它在 FoamBench-Advanced基准测试上取得了满分的可执行性得分，证明其原子工具集具备研究所需的可靠性。此外，一项 MILD燃烧案例研究表明，FlamePilot 能够作为交互式研究副驾驶，吸收文献知识来配置数值模拟、提出基于证据的改进，并自主管理参数研究，从而打通已发表知识与经验证结果之间的闭环。

通过采用单智能体、人类参与闭环范式，FlamePilot 确保了透明度并为研究人员提供指导，建立起协作伙伴关系，将工作流编排交由系统处理，从而加快科学发现。本研究为人工智能赋能的燃烧科学奠定了基础框架。其设计天然具备扩展性，可适配多种计算流体力学代码库，并将随着基础大语言模型的进步而持续改进，使智能体人工智能成为可持续的效能倍增器，同时恪守科学严谨性与可复现性。

利益冲突声明

作者声明不存在已知的利益冲突。

致谢

本研究得到国家自然科学基金委员会资助（项目编号：92270203、52276096和52441603）。

参考文献

[1] J. Wei, Y. Yang, X. Zhang, Y. Chen, X. Zhuang, Z. Gao, D. Zhou, G. Wang, Z. Gao, J. Cao, Z. Qiu, X. He, Q. Zhang, C. You, S. Zheng, N. Ding, W. Ouyang, N. Dong, Y. Cheng, S. Sun, L. Bai, B. Zhou, 《从科学人工智能到智能体科学：自主科学发现综述》，arXiv:2508.14111 [cs]（2025年8月）。[2] X. Wang, J. Xu, A. H. Feng, Y. Chen, H. Guo, F. Zhu, Y. Shao, M. Ren, H. Yi, S. Lian, H. Yang, T. Wu, H. Hu, S. Xiang, X.-Y. Zhang, C.-L. Liu, 《自主研究指南：科学智能体综述》（2025年8月）。[3] J. Chai, S. Tang, R. Ye, Y. Du, X. Zhu, M. Zhou, Y. Wang, W. E, Y. Zhang, L. Zhang, S. Chen, 《SciMaster：迈向通用科学人工智能体，第一部分。X-Master作为基础：我们能在人类最后的考试中领先吗？》，arXiv:2507.05241 [cs]（2025年7月）。[4] L. Mitchener, A. Yiu, B. Chang等, 《Kosmos：用于自主发现的人工智能科学家》，arXiv:2511.02824 [cs]（2025年11月）。[5] A. Novikov, N. V˜u, M. Eisenberger, E. Dupont, P.- S. Huang, A. Z. Wagner, S. Shirobokov, B. Ko- zlovskii, F. J. R. Ruiz, A. Mehrabian, M. P. Ku- mar, A. See, S. Chaudhuri, G. Holland, A. Davies, S. Nowozin, P. Kohli, M. Balog, 《AlphaEvolve：用于科学与算法发现的编码智能体》，arXiv:2506.13131 [cs]（2025年6月）。[6] T. D. Pham, A. Tanikanti, M. Kec.eli, 《ChemGraph：用于计算化学工作流的智能体框架》，arXiv:2506.06363 [physics]（2025年6月）。[7] K. Huang, S. Zhang, H. Wang, Y. Qu, Y. Lu, Y. Roohani, R. Li, L. Qiu, G. Li, J. Zhang, D. Yin, S. Marwaha, J. N. Carter, X. Zhou, M. Wheeler, J. A. Bernstein, M. Wang, P. He, J. Zhou, M. Snyder, L. Cong, A. Regev, J. Leskovec, 《Biomni：通用生物医学人工智能体》（2025年6月）。[8] H. Jasak, 《OpenFOAM：科研与工业中的开源计算流体力学》，《国际海军建筑与海洋工程杂志》1（2）（2009）89–94。[9] D. G. Goodwin, H. K. Moffat, I. Schoegl, R. L. Speth, B. W. Weber, 《Cantera：用于化学动力学、热力学和传输过程的面向对象软件工具包》，https://www.cantera.org，3.1.0版（2024年）。[10] Y. Chen, X. Zhu, H. Zhou, Z. Ren, 《MetaOpenFOAM：基于大语言模型的计算流体力学多智能体框架》，版本号：2（2024年）。[11] Y. Chen, L. Zhang, X. Zhu, H. Zhou, Z. Ren, 《OptMetaOpenFOAM：基于计算流体力学的敏感性分析与参数优化的大语言模型驱动思维链》，版本号：1（2025年）。[12] Y. Chen, X. Zhu, H. Zhou, Z. Ren, 《MetaOpenFOAM 2.0：用于自动化计算流体力学模拟与后处理的大语言模型驱动思维链》，版本号：1（2025年）。[13] L. Xu, D. Mohaddes, Y. Wang, 《用于火灾动力学模拟的大语言模型智能体》，版本号：1（2024年）。[14] Z. Dong, Z. Lu, Y. Yang, 《微调用于自动化计算流体力学模拟的大语言模型》，《理论与应用力学快报》15（3）（2025年）100594。[15] E. Fan, K. Hu, Z. Wu, J. Ge, J. Miao, Y. Zhang, H. Sun, W. Wang, T. Zhang, 《ChatCFD：用于端到端计算流体力学自动化的由大语言模型驱动的智能体，结合领域特定结构化推理，版本号：2（2025年）。[16] S. Pandey, R. Xu, W. Wang, X. Chu，OpenFOAMGPT：一种用于基于 OpenFOAM 的计算流体力学的检索增强型大语言模型（大语言模型）智能体，《流体物理》37（3）（2025年）035120。[17] J. Feng, R. Xu, X. Chu, OpenFOAMGPT 2.0：面向计算流体力学的端到端可信自动化，arXiv:2504.19338 [physics]（2025年4月）。[18] Z. Xu, L. Wang, C. Wang, Y. Chen, Q. Luo, H.-D. Yao, S. Wang, G. He，CFDagent：用于复杂流动模拟的语言引导型零样本多智能体系统，版本号：1（2025年）。[19] L. Yue, N. Somasekharan, Y. Cao, S. Pan, Foam-Agent：迈向自动化智能计算流体力学工作流，版本号：1（2025年）。[20] N. Somasekharan, L. Yue, Y. Cao, W. Li, P. Emami, P. S. Bhargav, A. Acharya, X. Xie, S. Pan，CFDLLM-Bench：用于评估大语言模型在计算流体力学领域表现的基准测试套件，arXiv:2509.20374 [cs]（2025年10月）。[21] R. Mao, M. Lin, Y. Zhang, T. Zhang, Z.-Q. J. Xu, Z. X. Chen，DeepFlame：深度学习赋能的反应流模拟开源平台，《计算物理通讯》291（2023年）108842。[22] R. Mao, M. Zhang, Y. Wang, H. Li, J. Xu, X. Dong, Y. Zhang, Z. X. Chen，利用图形处理器和机器学习模型加速反应流模拟的集成框架，《燃烧学会论文集》40（1-4）（2024年）105512，出版商：Elsevier BV。[23] R. Mao, X. Dong, X. Bai, Z. Wu, G. Dang, H. Li, Z. X. Chen，DeepFlame 2.0：面向低马赫数条件下原生GPU机器学习加速反应流模拟的新版本，《计算物理通讯》312（2025年）109595。[24] Claude Code。https://www.anthropic.com/claude-code（2025年）。[25] OpenCode。https://github.com/sst/opencode（2025年）。[26] Gemini Code CLI。https://github.com/google-gemini/gemini-cli（2025年）。[27] Codex。https://openai.com/codex/（2025年）。[28] B. Dally, A. Karpetis, R. Barlow，稀释热伴流中湍流非预混射流火焰的结构，《燃烧学会论文集》29（1）（2002年）1147–1154。[29] J. Aminian, C. Galletti, S. Shahhosseini, L. Tognotti，MILD燃烧器的数值研究：混合场、化学动力学及湍流-化学相互作用分析，《流动、湍流与燃烧》88（4）（2012年）597–623。[30] Z. Li, A. Cuoci, A. Sadiki, A. Parente，采用雷诺平均纳维-斯托克斯方法和详细化学反应机理对阿德莱德热伴流射流燃烧器进行全面数值研究，《能源》139（2017年）555–570。[31] Z. Zhao, T. Zhang, X. Li, L. Zhang, Z. Zhang, Y. Chen, F. Wu, C. Luo, C. Zheng，MILD富氧条件下高温伴流中CH4/NH3射流火焰的NO生成机理：伴流CO2和H2O的影响，Fuel 313（2022年）123030。[32] M. Ferrarotti, Z. Li, A. Parente，采用有限速率化学反应燃烧模型模拟MILD燃烧时混合模型的作用，《燃烧学会论文集》37（4）（2019年）4531–4538。