

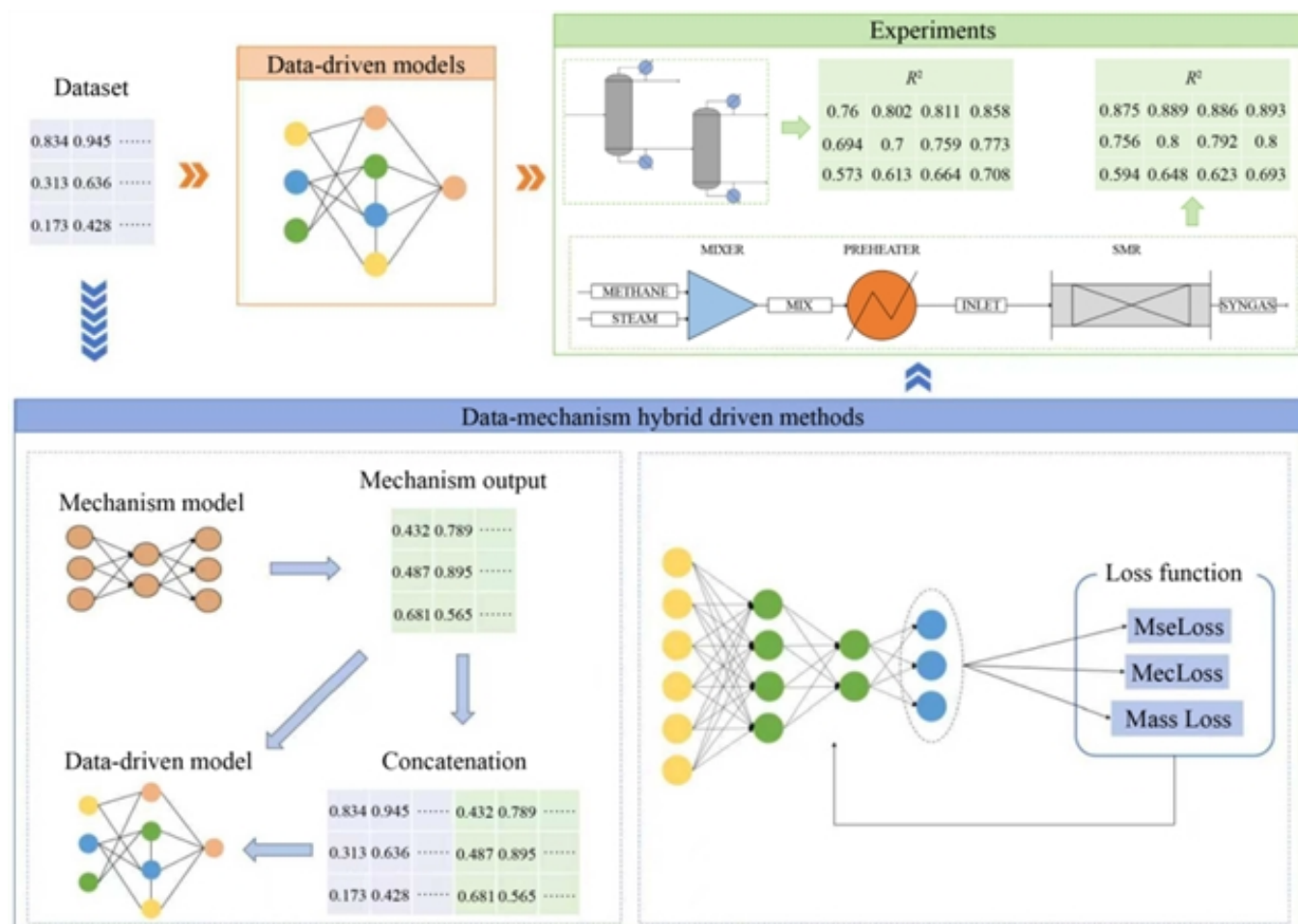
小样本高噪声解决方案：流程工业混合驱动方法

作者：writer 来源：科学网

本文原地址：<https://www.iikx.com/news/progress/41615.html>

本文仅供学习交流之用，版权归原作者所有，请勿用于商业用途！

小样本高噪声解决方案：流程工业混合驱动方法。流程工业中，关键质量变量离线测量延迟问题突出，软测量技术成为实时预测的核心手段，但纯机理模型计算成本高，纯数据驱动模型在小样本及含噪声场景下精度骤降，成为行业痛点。近期，大连理工大学赵珺教授团队提出多类数据-机理混合驱动方法，在两大典型化工过程中验证有效，为流程工业软测量优化提供了新路径。



研究方法

研究以随机森林（RF）、极端梯度提升（XGBoost）、人工神经网络（ANN）三类经典数据驱动模型为基础，设计了五类混合驱动方法（a-机理输出驱动，b-输入拼接增强，c-

机理约束损失（仅 ANN）， d-单机理-数据融合（仅 ANN）， e-双机理-数据融合（仅 ANN）），采用苯?甲苯?二甲苯（BTX）精馏与蒸汽甲烷重整（SMR）两大化工案例验证。本文整体研究框架如图1所示，单数据?机理混合驱动方法如图2所示，双机理-数据融合方法如图3所示。

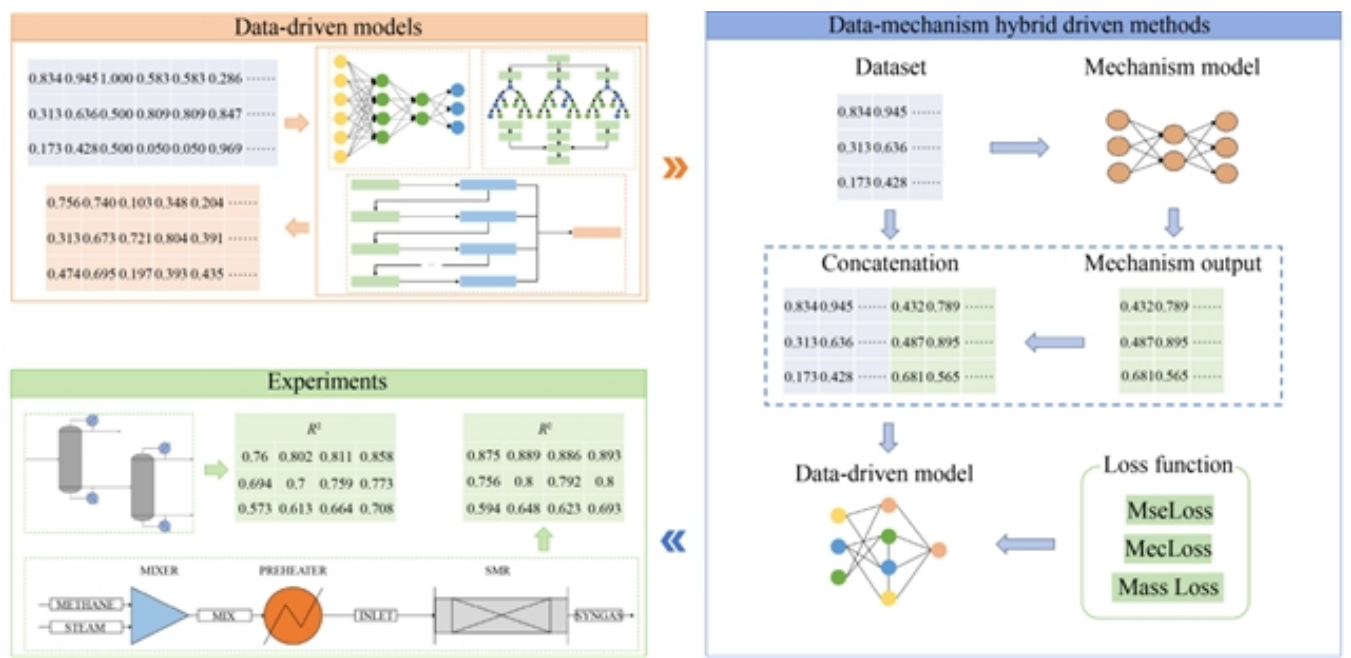


图1.面向不同样本量与噪声强度数据集的化工过程关键变量回归预测方法框架：数据驱动模型 → 数据?机理混合驱动方法 → 工业案例验证。

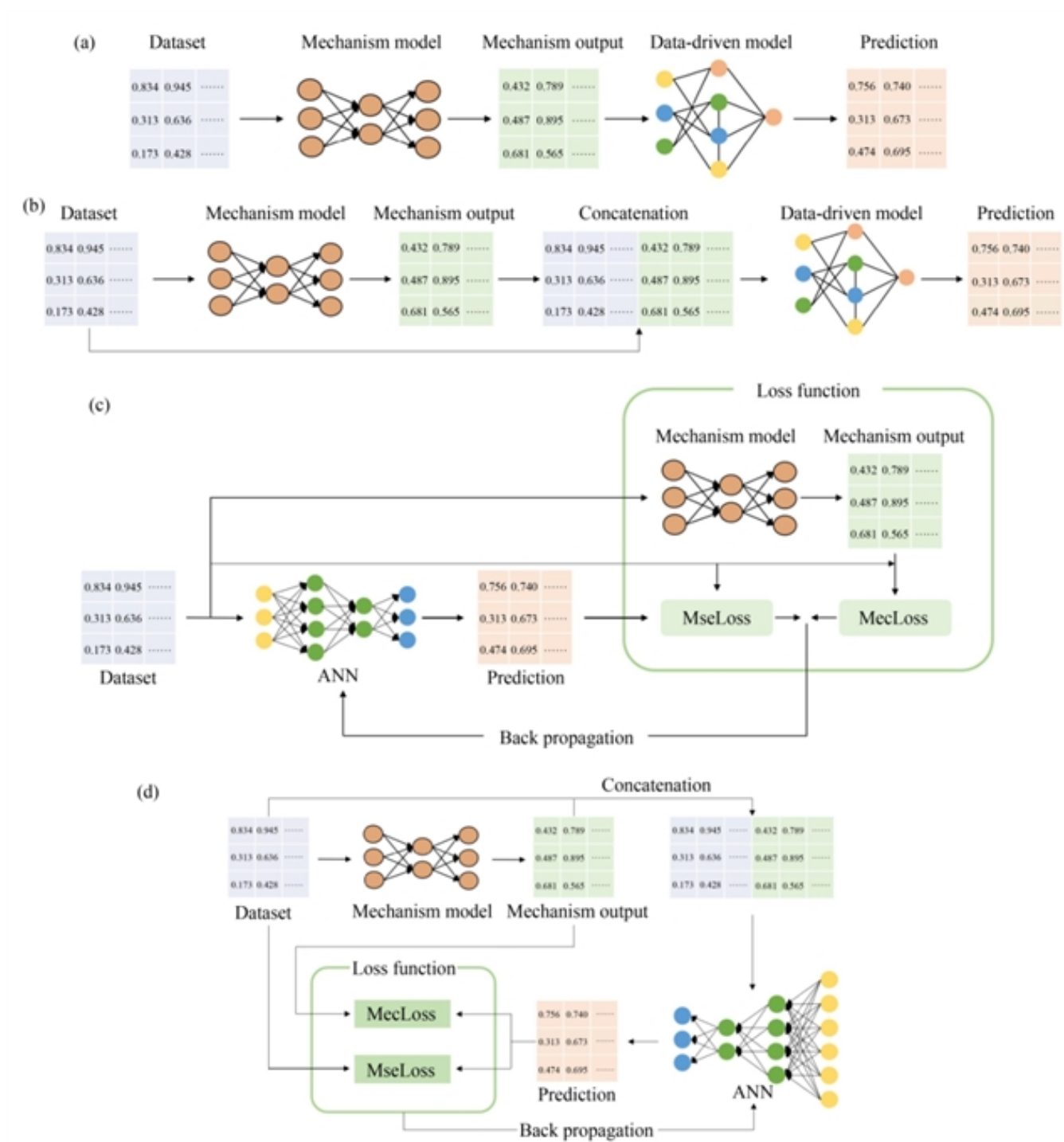


图2.不同数据-机理混合方法逻辑框架：(a) 机理输出驱动；(b) 输入拼接增强；(c) 机理约束损失；(d) 单机理-数据融合。

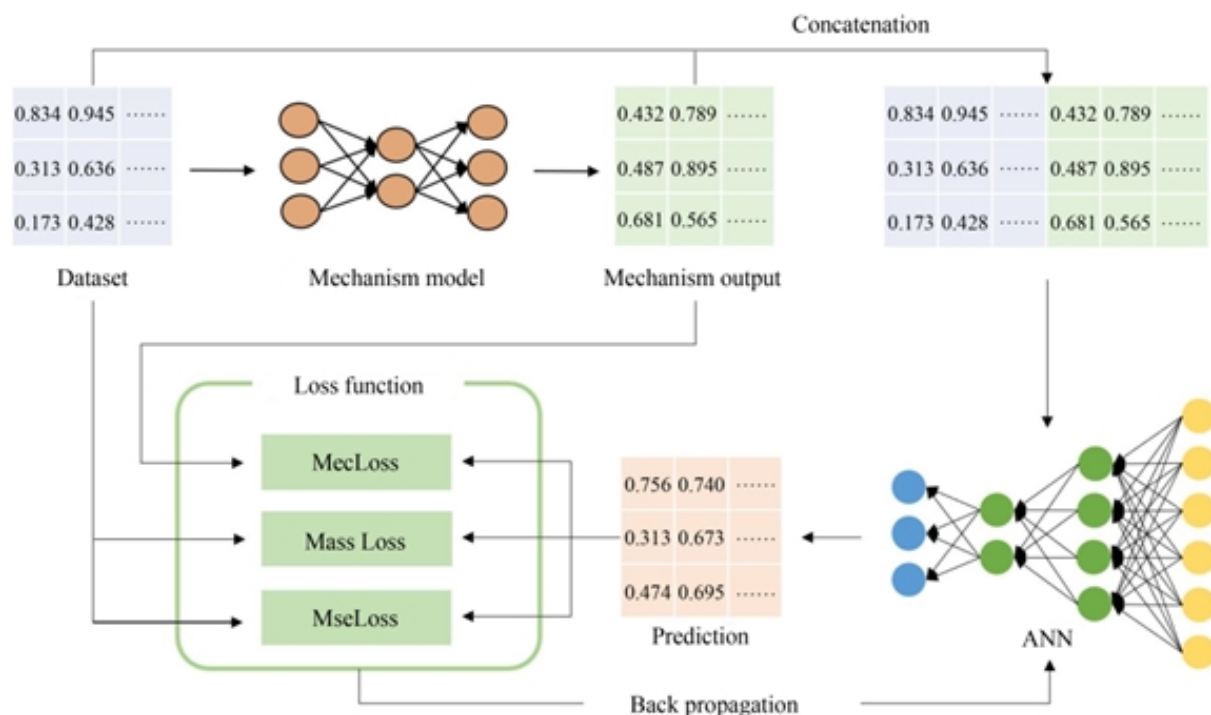


图3.在单机理-数据融合驱动方法的基础上，提出了双机理-数据混合方法逻辑框架。

核心结果

研究通过对比纯数据驱动与混合驱动模型的决定系数 R^2 值，明确各类方法性能，核心结果如下：

基础模型性能对比

XGBoost整体最优，抗噪声能力强；RF在100样本小样本场景表现突出；ANN对噪声最敏感，样本量越小，噪声越高，性能下降越明显。

混合方法有效性

所有混合方法均能提升预测精度，且小样本、高噪声场景下提升效果更显著。在10%~20%噪声、100~400样本范围内，RF、XGBoost、ANN的 R^2 最高分别提升5.2%、17.7%、36.2%，其中ANN提升潜力最大。

各类混合方法性能差异

方法-a性能极不稳定，多数场景降低精度，不推荐使用。方法-b是RF和XGBoost的最优适配方法，稳定提升精度， R^2 提升0.003~0.081。方法-c和方法-d是ANN的最优适配方法， R^2 提升0.003~0.159，稳定性最优；噪声强度升高后，基于ANN的方法-d效果更佳。在方法-d基础上进一步提升ANN精度， R^2 额外提升0.001~0.022，高值预测区间更稳定。双机理-数据混合方法-e能够进一步提高单机理-数据混合方法的预测精度。

案例适配性

BTX精馏：低噪声下XGBoost+方法-b、ANN+方法-d最优；高噪声下XGBoost+方法-b最稳定。SMR重整：绝大多数场景下XGBoost+方法-b表现最优，ANN+方法-d/-e低噪声优势明显。

结论与展望

数据-机理混合驱动方法显著优于纯数据驱动模型，能有效解决小样本、含噪声场景下的预测精度不足问题；预测精度提升幅度受噪声强度、样本量、基础模型影响，高噪声下XGBoost结合混合方法更稳定；RF、XGBoost适配方法-b，ANN适配方法-d，引入多重机理约束方法-e可进一步提升性能。当前所用虚拟机理模型结构较简单，限制了性能上限，后续将采用更严谨及贴合工业实际的机理模型，进一步提升混合驱动方法的精度与工业适用性，推动其在更多流程工业场景落地应用。（来源：EngineeringJournals微信公众号）

相关论文信息：<https://journal.hep.com.cn/fcse/EN/10.1007/s11705-026-2632-z>

特别声明：本文转载仅仅是出于传播信息的需要，并不意味着代表本网站观点或证实其内容的真实性；如其他媒体、网站或个人从本网站转载使用，须保留本网站注明的“来源”，并自负版权等法律责任；作者如果不希望被转载或者联系转载稿费事宜，请与我们联系。

作者：王超等 来源：《工程·化学工程》

更多 科学进展 请访问 <https://www.iikx.com/news/progress/>

本文版权归原作者所有，请勿用于商业用途，[爱科学iikx.com](https://www.iikx.com)转发