DOI: 10.19666/j.rlfd.202609026

基于物理引导长短期记忆-物理信息神经网络的磨煤机出口温度预测方法

邱天，王兴宇，张得海

（华北电力大学控制与计算机工程学院，北京 102206）

［摘 要］【目的】针对磨煤机出口温度预测中数据驱动模型难以同时兼顾预测精度与能量平衡物理一致性的问题，提出一种融合磨煤机物理先验的出口温度预测方法，以提高预测结果的准确性和物理合理性。【方法】以某 350 MW 燃煤机组 MPS 型中速磨煤机运行数据为基础，构建物理引导长短期记忆-物理信息神经网络（Physics-Guided Long Short-Term Memory-Physics-Informed Neural Network，PG-LSTM-PINN）。模型构造风煤比和热风占比以表征风煤匹配及冷热风混合状态，利用工况差异修正时间注意力，在长短期记忆网络（Long Short-Term Memory，LSTM）遗忘门中引入热惯性衰减因子，并依据能量平衡关系构造物理损失。通过模型对比、物理一致性评价、消融、小样本和物理损失权重敏感性实验评价模型性能。【结果】PG-LSTM-PINN 的 R^2为 0.981 1，均方根误差（Root Mean Square Error，δ_{RMSE}）和平均绝对误差（Mean Absolute Error，δ_{MAE}）分别为 0.078 2 ℃和 0.061 1 ℃，相比 LSTM 分别降低 50.69%和 52.41%；物理残差均方根误差 $\delta_{RMSE,p}$和物理残差平均绝对误差 $\delta_{MAE,p}$分别为 21.35 kJ/s 和 16.23 kJ/s。消融实验中，移除任一物理引导模块后，预测误差和物理残差均增大；小样本实验中，PG-LSTM-PINN 在各训练比例下均保持较低预测误差；物理损失权重敏感性实验中，η=1.301 3×10^{-9}时 δ_{RMSE}最低。【结论】所提方法能够提高磨煤机出口温度预测精度，减小能量平衡物理残差，并改善训练数据较少条件下的预测表现。

［关 键 词］中速磨煤机；长短期记忆网络；物理信息神经网络；注意力机制；能量平衡

［引用本文格式］（本段作者勿动）作者姓名. 中文标题[J]. 热力发电, 2026, 55(1): 起始页码-终止页码. 作者姓名. 英文标题[J]. Thermal Power Generation, 2026, 55(1): 起始页码-终止页码.

Coal Mill Outlet Temperature Prediction Method Based on Physics-Guided Long Short-Term Memory-Physics-Informed Neural Network

Qiu Tian, Wang Xingyu, Zhang Dehai

(School of Control and Computer Engineering, North China Electric Power University, Beijing 102206, China)

Abstract: [Objective] Data-driven models for coal mill outlet temperature prediction often have difficulty maintaining high prediction accuracy while also satisfying the physical requirement of energy balance. This study proposes an outlet temperature prediction method that incorporates physical priors of the coal mill, with the aim of improving both the numerical accuracy and the physical consistency of the prediction results. [Methods] Operating data from an MPS medium-speed coal mill in a 350 MW coal-fired generating unit were used to construct a Physics-Guided Long Short-Term Memory-Physics-Informed Neural Network (PG-LSTM-PINN). The method introduces the air-coal ratio and hot-air ratio to represent the matching relationship between primary air and coal feed and the mixing state of hot and cold air. Differences between operating conditions are then used

收 稿 日 期：XXXX-XX-XX 修 回 日 期：XXXX-XX-XX 接 受 日 期：XXXX-XX-XX 网络首发日期：XXXX-XX-XX
基 金 项 目：国家科技重大专项“全工况全流程智能控制系统及自治运行控制技术”（2025ZD1701004）
Supported by: National Science and Technology Major Project of the Ministry of Science and Technology of China (CoalMajor Project, 2025ZD1701004).
第一作者简介：邱天（1976），男，博士，副教授，硕士生导师，主要研究方向为发电过程状态监测与故障诊断，E-mail：qiutian@ncepu.edu.cn。
通信作者简介：王兴宇（2001），男，硕士研究生，主要研究方向为深度学习与火电过程时序预测，E-mail：1033327542@qq.com。

to correct the temporal attention mechanism so that historical information is weighted according to its relevance to the current condition. A thermal-inertia decay factor is embedded in the forget gate of the Long Short-Term Memory (LSTM) network to represent the dynamic influence of the mill thermal process on state updating. In addition, a physical loss is constructed from the energy-balance relationship and incorporated into model training. Model comparison, physical-consistency evaluation, ablation experiments, reduced-training-data experiments, and physical-loss-weight sensitivity experiments were conducted to evaluate the contribution and effectiveness of the proposed method. Together, these experiments evaluate prediction accuracy, energy-balance physical consistency, and performance with limited training data. [Results] PG-LSTM-PINN achieved a coefficient of determination (R^2) of 0.981 1, a root mean square error (δ_{RMSE}) of 0.078 2 ℃, and a mean absolute error (δ_{MAE}) of 0.061 1 ℃. Compared with LSTM, its δ_{RMSE} and δ_{MAE} were reduced by 50.69% and 52.41%, respectively. The physical-residual root mean square error ($\delta_{RMSE,p}$) and physical-residual mean absolute error ($\delta_{MAE,p}$) were 21.35 kJ/s and 16.23 kJ/s, respectively. In the ablation experiments, removing any physical-guidance module increased both the prediction error and the physical residual, indicating that each module contributed to the overall performance of the complete model. In the reduced-training-data experiments, PG-LSTM-PINN maintained relatively low prediction errors at all tested training proportions. The sensitivity analysis of the physical-loss weight showed that the lowest δ_{RMSE} was obtained when η was $1.301\ 3\times10^{-9}$. [Conclusion] The proposed PG-LSTM-PINN integrates operating-condition representation, temporal information weighting, thermal-inertia modeling, and energy-balance constraints within a unified prediction framework. The results show that it can improve the accuracy of coal mill outlet temperature prediction, reduce the energy-balance physical residual, and provide better prediction performance when the amount of training data is limited.

Key words: medium-speed coal mill; long short-term memory network; physics-informed neural network; attention mechanism; energy balance

磨煤机是燃煤机组制粉系统的关键辅机，其运行状态直接影响锅炉燃烧稳定性和制粉系统安全。对于直吹式制粉系统，出口温度是表征磨内热工状态的重要参数：温度过高会增大煤粉着火和自燃风险，过低则可能造成原煤干燥不足，影响煤粉输送和燃烧效率。准确预测出口温度有助于保障制粉系统稳定运行[1]。

现有研究包括机理建模、数据驱动和物理信息融合。机理建模方面，Gao 等[1]和 Li 等[2]分别建立含质量、能量守恒及传热、水分过程的磨煤机模型；王朝阳等[3]研究制粉系统扰动与变负荷控制；景峥等[4]构建灰箱模型分析关键运行因素。随着分散控制系统（Distributed Control System，DCS）和历史数据库完善，Pullanikkattil 等[5]验证了现场数据用于出口温度预测的可行性；牛玉广等[6]表明深度时空模型能够利用磨煤机历史多变量信息。物理信息融合方面，物理信息神经网络（Physics-Informed Neural Network，PINN）将控制方程或物理先验引入训练。吴婷婷等[7]以质量和能量守恒约束磨煤机机理模型参数及中间变量辨识；邓志平等[8]、胡淑昂等[9]分别将物理约束用于生物质气化产物分布和锅炉燃烧 NO_x 浓度预测；Zhang 等[10]将换热、热惯性和能量平衡先验引入双向长短期记忆网络（Bidirectional Long Short-Term Memory，BiLSTM）进行供热负荷预测。

基于上述分析，本文提出物理引导长短期记忆-物理信息神经网络（Physics-Guided Long Short-Term Memory-Physics-Informed Neural Network，PG-LSTM-PINN）磨煤机出口温度预测方法，以长短期记忆网络（Long Short-Term Memory，LSTM）为预测主体。与已有研究相比，吴婷婷等[7]主要利用质量与能量守恒约束开展磨煤机动态机理模型参数及中间变量辨识，Zhang 等[10]面向供热负荷预测构建融合热工先验的 BiLSTM 模型；经典 PINN[11]主要面向连续控制方程的正、反问题，通常利用自动微分计算方程导数。本文针对 MPS 型中速磨煤机离散运行序列的出口温度预测，将风煤匹配、冷热风混合、热惯性及能量平衡等物理先验分别融入输入表征、历史信息加权、循环状态更新和训练约束，并采用相邻预测值前向差分构造离散时序物理残差。采用某 350 MW 燃煤机组现场运行数据开展模型对比、物理一致性、消融、小样本及物理损失权重敏感性实验，验证所提方法的有效性。

1 基于 PG-LSTM-PINN 的出口温度预测模型

1.1 模型总体结构

磨煤机出口温度受风煤匹配、冷热风混合、热惯性和能量平衡等过程影响。为利用先验，构建以 Hochreiter 和 Schmidhuber[12]提出的 LSTM 为主体的

PG-LSTM-PINN，参考 Raissi 等[11]的 PINN 思想，将物理知识嵌入输入表征、历史信息加权、状态更新和训练目标，架构见图 1。

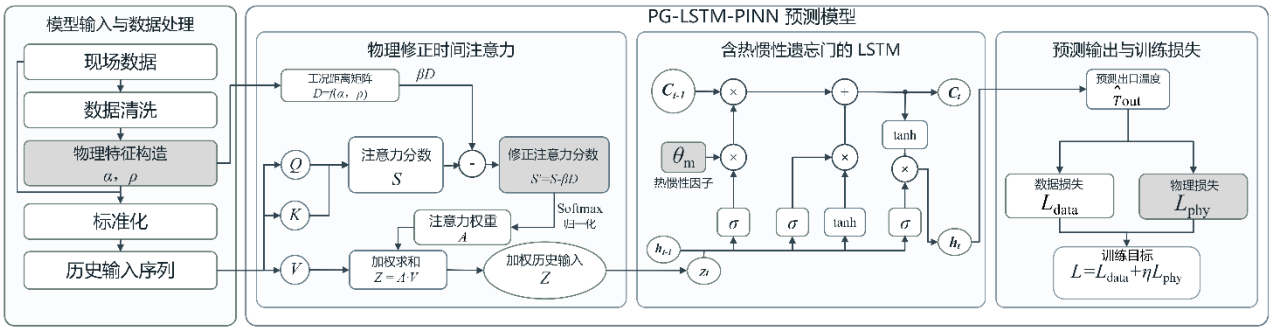


图 1 PG-LSTM-PINN 磨煤机出口温度预测模型整体架构

Fig.1 Overall Framework of the PG-LSTM-PINN Model for Coal Mill Outlet Temperature Prediction

模型以 6 个运行变量及历史出口温度构成 7 个基础输入通道，加入风煤比 α 和热风占比 ρ 后为 9 通道输入；根据两项物理特征的工况差异修正时间注意力，加权后输入 LSTM，在遗忘门引入等效热惯性衰减因子；最后将预测出口温度代入能量平衡关系构造物理残差，与数据损失组成训练目标。与仅在训练目标中附加物理约束不同，本文将磨煤机物理先验贯穿输入特征构造、历史信息加权、LSTM 状态更新和损失函数设计。

1.2 物理特征构造

原始变量难以直接表征一次风与给煤量的匹配，故构造风煤比 α：

$$\alpha = \frac{W_{\mathrm{air}}}{W_{\mathrm{c}}} \tag{1}$$

式中：α 为风煤比，无量纲；W_{air} 为一次风质量流量，kg/s；W_{c} 为给煤量，kg/s。

α 表示单位给煤量对应的一次风供给水平，用于表征一次风与给煤负荷的匹配状态。

一次风总量相近时，冷热风比例仍影响入口携热能力，故结合通道最大流量构造热风占比 ρ：

$$\rho = \frac{U_{\mathrm{H}} W_{\mathrm{H,max}}}{U_{\mathrm{H}} W_{\mathrm{H,max}} + U_{\mathrm{L}} W_{\mathrm{L,max}}} \tag{2}$$

式中：ρ 为热风占比，无量纲；U_{H} 和 U_{L} 分别为热风门和冷风门开度，%；$W_{\mathrm{H,max}}$ 和 $W_{\mathrm{L,max}}$ 分别为热风通道和冷风通道最大质量流量，kg/s。$W_{\mathrm{H,max}}$ 和 $W_{\mathrm{L,max}}$ 属于机理模型关键动态参数，其辨识结果在算例分析中给出。

α 和 ρ 分别表征风煤匹配和冷热风混合，由基础变量构造，既作为增强输入，也用于度量历史工况差异。

1.3 物理修正时间注意力机制

注意力机制[13]按历史特征相关性分配权重，但相似特征未必对应相同工况。为减少工况混合，引入由风煤比 α 和热风占比 ρ 构造的工况距离。

设历史窗口长度为 l，增强输入线性映射为查询 Q、键 K 和值 V，时间步 i 对 j 的普通注意力分数为：

$$S_{ij} = \frac{q_i^{\mathrm{T}} k_j}{\sqrt{d_{\mathrm{k}}}} \tag{3}$$

式中：S_{ij} 为普通注意力分数；q_i 和 k_j 分别为时间步 i 的查询向量和时间步 j 的键向量；d_{k} 为键向量维数。

为消除量纲和尺度影响，对 α 和 ρ 作 Z-Score 标准化，标准化参数仅由训练集计算；标准化后仍记为 α 和 ρ。时间步 i 和 j 的物理工况距离为：

$$D_{ij} = \left|\alpha_i - \alpha_j\right| + \gamma\left|\rho_i - \rho_j\right| \tag{4}$$

式中：D_{ij} 为物理工况距离；α_i、α_j 和 ρ_i、ρ_j 分别为两个时间步的风煤比和热风占比；γ 为热风占比差异调节系数。

以工况距离修正注意力分数：

$$S_{ij}^{'} = S_{ij} - \beta D_{ij} \tag{5}$$

式中：$S_{ij}^{'}$ 为物理修正后的注意力分数；β 为物理修正强度系数。

经 Softmax 归一化得到注意力权重：

$$A_{ij} = \frac{\exp(S_{ij}^{'})}{\sum_{m=1}^{l} \exp(S_{im}^{'})} \tag{6}$$

式中：A_{ij} 为时间步 i 对时间步 j 的注意力权重；m 为归一化求和索引。

加权值向量得到输出特征：

$$z_i = \sum_{j=1}^{l} A_{ij} v_j \tag{7}$$

式中：z_i为时间步 i 经物理修正时间注意力融合后的特征向量；v_j为时间步 j 对应的值向量。

各时间步输出组成特征序列：

$$Z = [z_1, z_2, \ldots, z_l] \tag{8}$$

式中：Z 为历史特征序列。注意力分数、物理工况距离、调节系数和注意力权重均无量纲；查询、键、值、特征表示、网络权重和偏置为网络内部量，不对应具体物理量。

由式（4）和式（5）可知，工况越近，距离惩罚越弱，模型越倾向利用相近工况的历史信息；输出序列送入 LSTM。

1.4 热惯性遗忘门

LSTM 通过门控学习长期依赖[14]，但标准遗忘门未显式反映磨内物料和金属蓄热引起的热状态衰减。为此引入等效热惯性衰减因子，时间步 t 的标准遗忘门为：

$$f_{t,\mathrm{std}} = \sigma\left(W_{\mathrm{f}}\left[z_t, h_{t\text{-}1}\right] + b_{\mathrm{f}}\right) \tag{9}$$

式中：$f_{t,\mathrm{std}}$为无量纲标准遗忘门；$\sigma(\cdot)$为 Sigmoid 函数；z_t和 $h_{t\text{-}1}$分别为当前输入特征和上一时步隐藏状态；W_{f}、b_{f}为权重矩阵和偏置。

等效热惯性衰减因子定义为：

$$\theta_{\mathrm{m}} = \exp\left(-\frac{\Delta t}{\tau_{\mathrm{m}}}\right) \tag{10}$$

式中：θ_{m}为等效热惯性衰减因子，无量纲；Δt 为采样间隔，取 1 s；τ_{m}为可学习等效热惯性时间尺度参数，s。

训练过程中采用正值参数化保证 $\tau_{\mathrm{m}}>0$。

热惯性遗忘门定义为：

$$f_{t,\mathrm{phy}} = f_{t,\mathrm{std}} \odot \theta_{\mathrm{m}} \tag{11}$$

式中：$f_{t,\mathrm{phy}}$ 为无量纲热惯性遗忘门；⊙为逐元素乘法，即标准遗忘门各维元素与热惯性衰减因子相乘。

$f_{t,\mathrm{phy}}$在 LSTM 单元内部替代标准遗忘门参与细胞状态更新，其余门控结构保持不变。

1.5 基于能量平衡的物理损失函数

在数据损失基础上引入物理损失，以约束预测结果满足能量平衡。集总模型的磨内蓄热变化率等于输入与输出总热流率之差：

$$c_{\mathrm{mix}}\left(M_{\mathrm{c}} + M_{\mathrm{pf}} + M_{\mathrm{metal}}\right)\frac{\mathrm{d}T_{\mathrm{out}}}{\mathrm{d}t} = Q_{\mathrm{in}} - Q_{\mathrm{out}} \tag{12}$$

式中：c_{mix} 为磨内煤与金属混合物等效比热容，kJ/(kg·℃)；M_{c}、M_{pf}、M_{metal}分别为磨内原煤、煤粉和参与换热金属质量，kg，其中 M_{metal}为机理模型关键动态参数，辨识结果见算例分析；T_{out}、$\frac{\mathrm{d}T_{\mathrm{out}}}{\mathrm{d}t}$ 分别为出口温度及变化率，单位为℃、℃/s；Q_{in}、Q_{out}为输入、输出总热流率，kJ/s。

输入总热流率为一次风、原煤、磨煤机做功、密封风和漏入冷风热流率之和：

$$Q_{\mathrm{in}} = Q_{\mathrm{air}} + Q_{\mathrm{rc}} + Q_{\mathrm{mac}} + Q_{\mathrm{seal}} + Q_{\mathrm{le}} \tag{13}$$

式中：Q_{air}、Q_{rc}、Q_{mac}、Q_{seal}、Q_{le} 分别为一次风、原煤、磨煤机做功转化、密封风和漏入冷风热流率，均为 kJ/s。输出热流包括出口一次风与密封风带出、设备散热、水分蒸发和煤粉升温。将预测 $\hat{T}_{\mathrm{out}}$ 代入相关项，得到 $\hat{Q}_{\mathrm{out}}$：

$$\begin{aligned}\hat{Q}_{\mathrm{out}} &= \hat{c}_{\mathrm{out}}\hat{T}_{\mathrm{out}}(1+K_{\mathrm{seal}})W_{\mathrm{air}} + K_{\mathrm{loss}}Q_{\mathrm{in}} \\ &+\Delta\hat{M}W_{\mathrm{c}}(2\,500 + \hat{c}_{\mathrm{p,v}}\hat{T}_{\mathrm{out}} - c_{\mathrm{w}}T_{\mathrm{rc}}) \\ &+\frac{100 - M_{\mathrm{ar}}}{100}W_{\mathrm{c}}\hat{c}_{\mathrm{pf}}(T_{\mathrm{out}} - T_{\mathrm{rc}})\end{aligned} \tag{14}$$

式中：$\hat{Q}_{\mathrm{out}}$ 为预测输出总热流率；$\hat{c}_{\mathrm{out}}$ 为出口湿空气比热容；$\hat{T}_{\mathrm{out}}$ 为预测出口温度；K_{seal}、W_{air}、K_{loss} 和 Q_{in} 分别为密封风系数、一次风流量、设备散热系数和输入总热流率；$\Delta\hat{M}$ 为煤中蒸发水分质量分数；W_{c} 为给煤量；$\hat{c}_{\mathrm{p,v}}$、c_{w}、T_{rc}、M_{ar} 和 $\hat{c}_{\mathrm{pf}}$ 分别为水蒸气定压平均比热容、水比热容、原煤温度、收到基煤全水分和煤粉比热容。$\hat{Q}_{\mathrm{out}}$ 单位为 kJ/s；$\hat{c}_{\mathrm{out}}$、$\hat{c}_{\mathrm{p,v}}$、c_{w} 和 $\hat{c}_{\mathrm{pf}}$ 单位均为 kJ/(kg·℃)；$\hat{T}_{\mathrm{out}}$ 和 T_{rc} 单位均为℃；W_{air} 和 W_{c} 单位均为 kg/s；K_{seal} 和 K_{loss} 无量纲；M_{ar} 单位为%；$\Delta\hat{M}$ 为 0～1 形式的无量纲质量分数；常数 2 500 为水的近似汽化潜热，单位为 kJ/kg。带帽变量表示预测值或由预测出口温度计算的温度相关量。式（13）和式（14）的热流率分项参考 Gao 等[1]，关键动态参数依据现场数据重新辨识。

同一连续运行段内，对相邻预测起点的出口温度预测值作前向差分，近似变化率：

$$\frac{\mathrm{d}\hat{T}_{\mathrm{out}}}{\mathrm{d}t} \approx \frac{\hat{T}_{\mathrm{out}}(t+\Delta t) - \hat{T}_{\mathrm{out}}(t)}{\Delta t} \tag{15}$$

式中：t 为时间，s；Δt 为采样间隔，取 1 s；两预测值均为℃，差分变化率为℃/s。经典 PINN 通常利用自动微分（automatic differentiation，AD）计算连续控制方程导数[11]；本文针对 LSTM 离散序列，以式（15）一阶前向差分近似式（12）的时间导数。差分仅在同一连续运行段内构造，使用相邻预测起点的模型预测值，不使用预测时刻之后的实测出口

温度。

将式（14）的预测输出总热流率、式（15）的温度变化率及同一目标时刻的输入总热流率和辅助状态量代入式（12），构造物理残差：

$$r_{\text{phy}} = c_{\text{mix}}(M_{\text{c}} + M_{\text{pf}} + M_{\text{metal}})\frac{\mathrm{d}\hat{T}_{\text{out}}}{\mathrm{d}t} - Q_{\text{in}} + \hat{Q}_{\text{out}} \tag{16}$$

式中：r_{phy}为能量平衡物理残差，单位为 kJ/s，其余符号含义同前。

数据损失为：

$$L_{\text{data}} = \frac{1}{N}\sum_{n=1}^{N}\left(T_{\text{out},n}^{*} - \hat{T}_{\text{out},n}^{*}\right)^2 \tag{17}$$

物理损失为：

$$L_{\text{phy}} = \frac{1}{N}\sum_{n=1}^{N} r_{\text{phy},n}^2 \tag{18}$$

总损失为：

$$L = L_{\text{data}} + \eta L_{\text{phy}} \tag{19}$$

式中：N为式（17）和式（18）的有效样本数；$T_{\text{out},n}^{*}$和$\hat{T}_{\text{out},n}^{*}$分别为第 n 个样本标准化后的实测标签和预测值，均为无量纲量；$r_{\text{phy},n}$为能量平衡物理残差；η为物理损失权重。

训练时，L_{data}由标准化出口温度计算，为无量纲量；计算 L_{phy}前先将预测温度反标准化，再代入能量平衡方程。r_{phy}单位为 kJ/s 且未额外归一化，故 L_{phy}和 η 的单位分别为$(\text{kJ/s})^2$和$(\text{kJ/s})^{-2}$；η用于量纲转换和两项损失的数值尺度平衡。

上述设计从特征、结构和损失三层嵌入物理先验，符合物理信息机器学习的基本思路[15-17]。

2 算例分析

2.1 数据来源与实验设置

研究对象为某 350 MW 燃煤机组 MPS 型中速磨煤机，数据覆盖 2017 年 11 月 1—15 日，包括 A1—A13 共 13 个独立运行段，采样间隔为 1 s。输入为冷、热风阀门开度 U_{L}、U_{H}，给煤量 W_{c}、一次风流量 W_{air}、磨煤机电流 I、入口一次风温度 T_{in}及历史出口温度，预测下一采样时刻实测出口温度。

机理模型参考 Gao 等[1]的 MPS 动态模型。通用参数依据文献和设备资料，关键动态参数由现场数据重新辨识；采用 MATLAB 参数估计器，以一次风流量、磨煤机电流和出口温度三项输出误差平方和为目标，辨识后模型结构与参数固定，结果见表 1。

表 1　磨煤机机理模型参数辨识结果

Tab.1 Parameter Identification Results of the Coal Mill Mechanistic Model

物理含义及符号	辨识值	单位
冷一次风通道最大流量 $W_{\text{L,max}}$	40	kg·s^{-1}
热一次风通道最大流量 $W_{\text{H,max}}$	90.656	kg·s^{-1}
一次风流量动态环节惯性时间常数 T_1	10.6	s
入口一次风温度动态环节惯性时间常数 T_2	4.098	s
原煤向煤粉转化系数 K_{conv}	0.6	s^{-1}
煤粉流出系数 K_{pf}	0.081	—
煤粉质量电流比例系数 K_1	0.18	—
原煤质量电流比例系数 K_2	0.3	—
电流常数项 K_3	19.859	—
参与换热金属质量 M_{metal}	5 099.8	kg

注：“—”表示原始文献未给出明确单位，本文沿用其参数定义，不表示无量纲。

机理模型参数仅用 A1 运行段辨识；A3 仅用于独立动态校核，不参与辨识、调整或校正，并固定使用 A1 参数。A3 时间范围为 2017 年 11 月 2 日 19:20:00—21:39:59，采样间隔 1 s，共 8 400 个采样点。一次风流量 W_{air}、磨煤机电流 I 和出口温度 T_{out} 的 R^2分别为 0.827 3、0.790 1和0.851 2，δ_{RMSE} 分别为 1.532 1 kg/s、1.798 1 A和1.111 4 ℃，δ_{MAE} 分别为 1.023 1 kg/s、1.344 4 A和0.872 3 ℃。机理模型能够反映主要可测输出的动态变化，但对局部波动仍存在偏差。

各运行段独立驱动机理模型，磨内原煤量 M_{c} 和煤粉量 M_{pf} 初值均为 0，段间不传递状态。各段采用相同长度的预设段首数据预热，以减小零初值瞬态影响；预热区间及异常值、缺失值均不用于样本构造。预热后，辅助状态沿段内时间顺序前向递推。A12 和 A13 固定采用 A1 辨识参数，不重新辨识、调整或校正；测试段辅助状态和物理项不使用预测目标时刻及其后的实测出口温度。变量及单位见表 2。

表 2　磨煤机出口温度预测数据集主要变量及单位

Tab.2 Main Variables and Units of the Coal Mill Outlet Temperature Prediction Dataset

变量及符号	单位	变量及符号	单位
热风阀门开度 U_{H}	%	风煤比 α	无量纲
冷风阀门开度 U_{L}	%	热风占比 ρ	无量纲
一次风流量 W_{air}	kg·s^{-1}	磨内原煤量 M_{c}	kg
给煤量 W_{c}	kg·s^{-1}	磨内煤粉量 M_{pf}	kg
磨煤机电流 I	A		
入口一次风温度 T_{in}	℃	磨煤机出口温度 T_{out}	℃

加入风煤比 α 和热风占比 ρ 后，输入由 7 通道

扩展为 9 通道；M_c 和 M_{pf} 用于计算物理损失和物理一致性指标。

模型基于 Python 3.10.13、PyTorch 2.6.0 和 CUDA 12.4 实现，平台配置 16 GB 内存和 NVIDIA GeForce GTX 1650 显卡。采用自适应矩估计（Adaptive Moment Estimation，Adam）优化，最多训练 100 轮并早停。采用 Optuna[18]和树结构 Parzen 估计（Tree-structured Parzen Estimator，TPE）[19]寻优；η 在 10^{-11}～10^{-6}范围按对数尺度采样，与其他超参数共同进行 30 次 Optuna-TPE 试验，以 A12 验证集出口温度 δ_{RMSE} 为目标，A13 测试集不参与超参数选择，仅用于最终评价。除无需训练的持久性预测外，各学习模型及实验配置均独立运行 5 次，评价指标为 5 次实验结果的平均值。主要参数见表 3。

表 3 模型训练参数与超参数设置

Tab.3 Training Parameters and Hyperparameter Settings

参数名称	取值
最大训练轮数	100
批量大小	256
学习率	$1.747\ 2\times10^{-3}$
LSTM 隐藏层维度	128
LSTM 层数	1
热惯性时间尺度初值	10.703 0
物理损失权重 η	$1.301\ 3\times10^{-9}$
热风占比差异调节系数 γ	3.407 8
物理修正强度系数 β	4.845 1

2.2 数据预处理与评价指标

检查各运行段的缺失值、时间戳和测量异常，清洗后得到 215 280 个有效时间点，不跨段插值或构造样本。

按式（1）和式（2）构造风煤比 α 和热风占比 ρ，并将输入、辅助状态及出口温度按时间对齐。A1—A11、A12 和 A13 分别为训练、验证和测试集，有效时间点数依次为 164 340、19 440 和 31 500，分别用于参数学习、选择与早停及性能评价。

输入和预测目标采用 Z-Score 标准化，均值与标准差仅由训练集计算并用于验证集和测试集。设 x 为待标准化变量：

$$x^* = \frac{x-\mu_x}{\sigma_x} \tag{20}$$

式中：x 和 x^*分别为标准化前、后变量，x^*无量纲；μ_x 和 σ_x 分别为训练集对应变量的均值、标准差。

预测结果反标准化为具有实际单位的出口温度：

$$\hat{T}_{out} = \hat{T}^*_{out}\sigma_T + \mu_T \tag{21}$$

式中：$\hat{T}^*_{out}$ 和 $\hat{T}_{out}$ 分别为标准化空间和反标准化后的出口温度预测值，Tout 单位为℃；μ_T 和 σ_T 分别为训练集出口温度标签的均值、标准差。

各独立数据段内以 60 s 历史窗口构造样本，训练、验证和测试集分别得到 163 680、19 380 和 31 440 个监督样本，共 214 500 个，见表 4。表 4 时间范围为原始记录范围。

表 4 数据集划分

Tab.4 Dataset Partition

数据集	包含数据段	样本数	时间范围
训练集	A1—A11	163 680	2017-11-01 12:00:00—2017-11-12 04:09:59
验证集	A12	19 380	2017-11-12 06:00:00—2017-11-12 11:24:59
测试集	A13	31 440	2017-11-15 05:20:00—2017-11-15 14:09:00

模型从预测精度和能量平衡物理一致性两方面评价。预测精度采用决定系数 R^2、均方根误差（Root Mean Square Error，δ_{RMSE}）和平均绝对误差（Mean Absolute Error，δ_{MAE}）[20]：

$$R^2 = 1-\frac{\sum_{i=1}^{N}(y_i-\hat{y}_i)^2}{\sum_{i=1}^{N}(y_i-\bar{y})^2} \tag{22}$$

$$\delta_{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i-\hat{y}_i)^2} \tag{23}$$

$$\delta_{MAE} = \frac{1}{N}\sum_{i=1}^{N}|y_i-\hat{y}_i| \tag{24}$$

式中：y_i 和 $\hat{y}_i$ 分别为测试集实测和预测出口温度，℃；$\bar{y}$ 为测试集标签均值，℃；N 为有效预测点数。R^2 无量纲，δ_{RMSE} 和 δ_{MAE} 的单位均为℃。

物理一致性采用物理残差均方根误差 $\delta_{RMSE,p}$ 和物理残差平均绝对误差 $\delta_{MAE,p}$ 评价。$\delta_{RMSE,p}$ 对较大残差更敏感，$\delta_{MAE,p}$ 反映绝对残差平均水平；指标越小，能量平衡满足程度越高。

$$\delta_{RMSE,p} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}r_{phy,i}^2} \tag{25}$$

$$\delta_{MAE,p} = \frac{1}{N}\sum_{i=1}^{N}|r_{phy,i}| \tag{26}$$

式中：N 为物理残差数量；$r_{phy,i}$ 为第 i 个预测点的能量平衡物理残差，kJ/s；$\delta_{MAE,p}$ 和 $\delta_{RMSE,p}$ 的单位均为 kJ/s。评价不跨运行段边界，仅使用相邻预测值均

存在的位置。

2.3 预测性能分析

选取支持向量回归（Support Vector Regression，SVR）、随机森林（Random Forest，RF）、极端梯度提升（Extreme Gradient Boosting，XGBoost）、多层感知机（Multilayer Perceptron，MLP）、LSTM（7 通道）、LSTM（9 通道）、BiLSTM 和长短期记忆-物理信息神经网络（Long Short-Term Memory-Physics-Informed Neural Network，LSTM-PINN）进行对比，各模型采用相同的数据划分、历史窗口、下一采样时刻预测目标和预处理流程。LSTM（7 通道）使用 7 个基础输入通道，LSTM（9 通道）在相同网络结构和训练设置下增加风煤比 α 和热风占比 ρ。另设持久性预测，以当前时刻实测出口温度预测下一采样时刻，表示为 $\hat{T}_{out,t+1}=T_{out,t}$，间隔 1 s；结果见表 5。

表 5 不同模型的预测性能对比

Tab.5 Prediction Performance Comparison of Different Models

模型	R^2	δ_{RMSE}/℃	δ_{MAE}/℃
持久性预测	0.923 3	0.157 5	0.125 3
SVR	0.775 6	0.268 4	0.186 2
RF	0.812 0	0.245 6	0.184 8
XGBoost	0.858 9	0.213 7	0.164 6
MLP	0.901 6	0.178 4	0.148 6
LSTM（7 通道）	0.922 6	0.158 6	0.128 4
LSTM（9 通道）	0.931 2	0.149 5	0.121 5
BiLSTM	0.930 4	0.150 4	0.122 4
LSTM-PINN	0.948 1	0.129 5	0.106 1
PG-LSTM-PINN	**0.981 1**	**0.078 2**	**0.061 1**

持久性预测的 R^2、δ_{RMSE} 和 δ_{MAE} 分别为 0.923 3、0.157 5 ℃和 0.125 3 ℃，略优于 LSTM（7 通道），表明 1 s 单步预测中出口温度具有较强的短时连续性。由表 5 可知，深度学习模型总体优于传统机器学习模型，说明历史时序信息有助于提高预测精度。LSTM（9 通道）较 LSTM（7 通道）的 δ_{RMSE} 和 δ_{MAE} 分别降低 5.74%和 5.37%，说明风煤比 α 和热风占比 ρ 提供了有效的工况信息；BiLSTM 的 δ_{RMSE} 和 δ_{MAE} 较 LSTM（7 通道）有所降低。进一步引入物理先验后，LSTM-PINN 的 δ_{RMSE} 和 δ_{MAE} 降至 0.129 5 ℃和 0.106 1 ℃，PG-LSTM-PINN 进一步降至 0.078 2 ℃和 0.061 1 ℃，R^2达到 0.981 1；较 LSTM（7 通道）、LSTM-PINN 和持久性预测，其 δ_{RMSE} 和 δ_{MAE} 分别降低 50.69%/52.41%、39.61%/42.41%和 50.35%/51.24%。

图 2 和图 3 选取 A13 测试集同一连续 25 min 典型片段，共 1 500 个采样点，按照同时覆盖相对平稳和明显动态变化过程的原则非随机选取。前 800 点为相对平稳区间，PG-LSTM-PINN 的 δ_{RMSE} 和 δ_{MAE} 分别为 0.063 0 ℃和 0.045 0 ℃；后 700 点为明显动态变化区间，分别为 0.092 0 ℃和 0.071 0 ℃。由图 2 和图 3 可知，传统机器学习模型能够反映总体变化趋势，但预测曲线相对平滑；LSTM、BiLSTM 和 LSTM-PINN 对时序变化的跟踪能力有所提高，但在低温谷值和高温峰值附近仍存在偏差。PG-LSTM-PINN 在低温谷值及快速下降、回升阶段与实测值更为接近，在两类工况下均保持较好的跟踪能力。

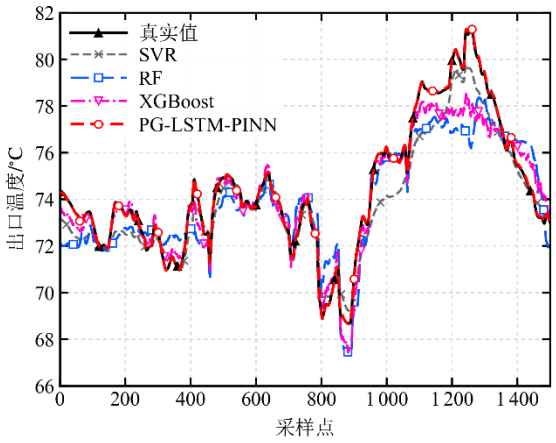


图 2 传统机器学习模型与 PG-LSTM-PINN 的出口温度预测对比

Fig.2 Outlet Temperature Prediction Comparison of Traditional Machine Learning Models with PG-LSTM-PINN

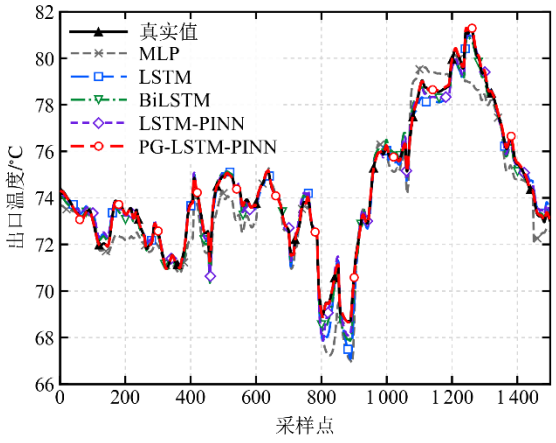


图 3 深度学习模型与 PG-LSTM-PINN 的出口温度预测对比

Fig.3 Outlet Temperature Prediction Comparison of Deep Learning Models with PG-LSTM-PINN

2.4 物理一致性分析

将各模型预测出口温度代入能量平衡方程，结合辨识机理模型辅助状态，计算完整测试集物理残差，$\delta_{RMSE,p}$ 和 $\delta_{MAE,p}$ 见图 4。

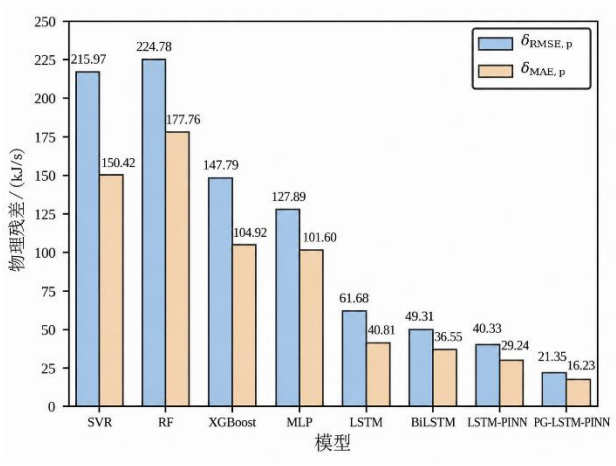


图 4 不同模型的物理一致性评价结果

Fig.4 Comparison of Physical Consistency Metrics of Different Models

PG-LSTM-PINN 的 $\delta_{RMSE,p}$ 和 $\delta_{MAE,p}$ 分别为 21.35 kJ/s 和 16.23 kJ/s；较 LSTM 分别降低 65.39% 和 60.23%，较 LSTM-PINN 分别降低 47.06%和 44.49%，物理残差进一步减小。

2.5 消融实验

设置基础 LSTM、完整模型及 4 个单模块移除变体进行单因素消融，配置见表 6。

表 6 消融实验模型变体配置

Tab.6 Model Variant Configurations for Ablation Experiments

模型变体	物理特征输入	物理修正时间注意力	热惯性遗忘门	物理损失
基础 LSTM	否	否	否	否
移除物理特征输入	否	是	是	是
移除物理修正时间注意力	是	否	是	是
移除热惯性约束	是	是	否	是
移除物理损失	是	是	是	否
完整模型	是	是	是	是

六组模型按表 6 顺序的 R^2分别为 0.922 6、0.952 3、0.968 0、0.969 3、0.951 2 和 0.981 1，完整模型最高，各单模块移除后 R^2均有所下降。由图 5 可知，完整模型的 δ_{RMSE} 和 δ_{MAE} 均最低；移除物理特征输入或物理损失后误差增幅相对较大，移除物理修正时间注意力或热惯性约束后误差也有所增加。

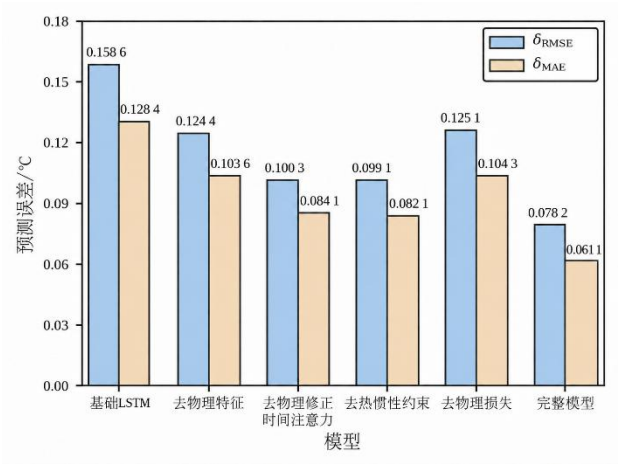


图 5 不同模型变体的预测误差对比

Fig.5 Comparison of Prediction Errors of Different Model Variants

由图 6 可知，完整模型的 $\delta_{RMSE,p}$ 和 $\delta_{MAE,p}$ 分别为 21.35 kJ/s 和 16.23 kJ/s，低于基础 LSTM 的 61.68 kJ/s 和 40.81 kJ/s。各模块移除后物理残差均增大；结合图 5 可见，预测误差也均增加，说明各模块对预测精度和物理一致性均有贡献。

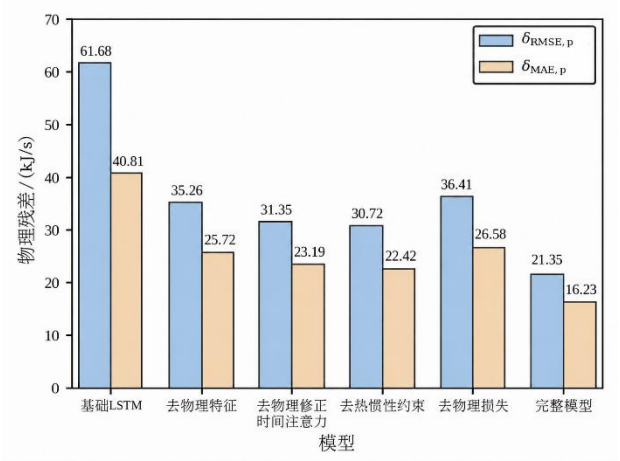


图 6 不同模型变体的物理残差指标对比

Fig.6 Comparison of Physical Residual Metrics of Different Model Variants

2.6 小样本预测实验

固定 A12 验证集和 A13 测试集，从 A1—A11 各段 60 s 历史窗口构造的训练样本中随机抽取 20%、40%、60%、80%和 100%；其余设置不变，对比 LSTM、LSTM-PINN 与 PG-LSTM-PINN，见图 7。

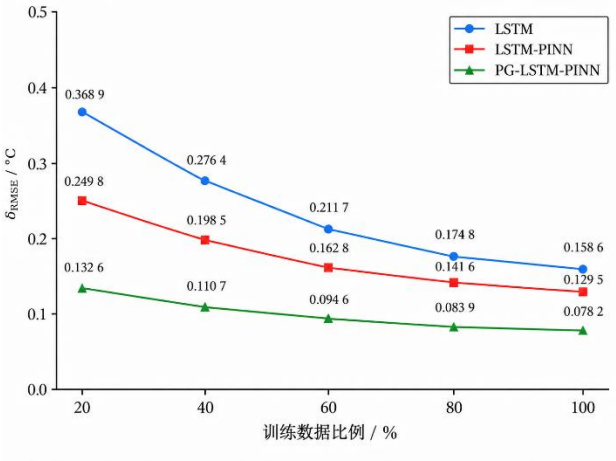


图 7　不同训练数据比例下的预测 δ_{RMSE}

Fig.7 Prediction δ_{RMSE} at Different Training Data Proportions

PG-LSTM-PINN 在各比例下的 δ_{RMSE} 最低。20% 时，LSTM、LSTM-PINN 和 PG-LSTM-PINN 的 δ_{RMSE} 为 0.368 9、0.249 8 和 0.132 6 ℃；100%时为 0.158 6、0.129 5 和 0.078 2 ℃。完整模型用 20%样本的误差低于基础 LSTM 用全部样本，表明物理先验有助于提高小样本预测准确性。

2.7　物理损失权重敏感性分析

物理损失权重 η 用于平衡数据损失与能量平衡物理损失。保持其余模型结构和训练参数不变，将 η 设为 10^{-11}、10^{-10}、$1.301\ 3\times10^{-9}$、10^{-8}、10^{-7}和 10^{-6}，比较预测 δ_{RMSE}，见图 8。

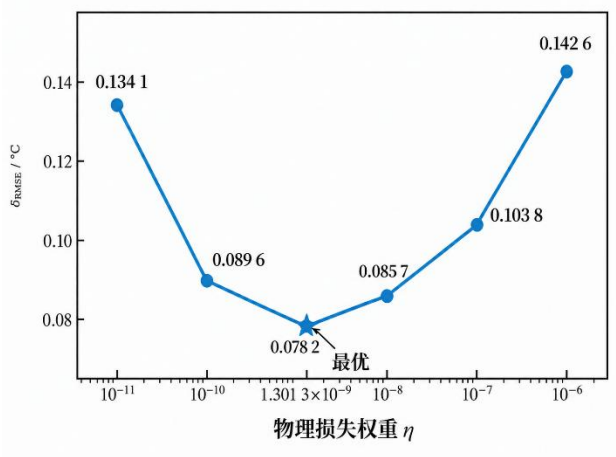


图 8　物理损失权重敏感性分析

Fig.8 Sensitivity Analysis of the Physical Loss Weight

η 从 10^{-11}增至 $1.301\ 3\times10^{-9}$时，δ_{RMSE} 由 0.134 1 ℃降至最低 0.078 2 ℃；增至 10^{-6}时升为 0.142 6 ℃。说明权重过小时物理约束较弱，过大时数据损失相对作用减弱，均不利于预测精度。

3　结论

针对磨煤机出口温度预测难以兼顾精度与物理一致性的问题，本文提出 PG-LSTM-PINN，并以某 350 MW 燃煤机组数据验证，结论如下：

1）与已有研究相比，本文面向 MPS 型中速磨煤机离散序列，将风煤比和热风占比、物理修正时间注意力、热惯性遗忘门及能量平衡损失分别融入输入表征、历史加权、状态更新和训练约束，并以前向差分构造离散物理残差，体现物理对象、网络结构和方程实现的差异。

2）A13 测试集上，PG-LSTM-PINN 的 R^2、δ_{RMSE} 和 δ_{MAE} 为 0.981 1、0.078 2 ℃和 0.061 1 ℃；较持久性预测，δ_{RMSE} 和 δ_{MAE} 降低 50.35%和 51.24%；较 LSTM（7 通道）降低 50.69%和 52.41%。模型能较好跟踪典型工况中的快速下降、回升及局部波动。

3）PG-LSTM-PINN 的 $\delta_{RMSE,p}$ 和 $\delta_{MAE,p}$ 为 21.35 kJ/s 和 16.23 kJ/s，均低于对比模型；移除任一物理引导模块后，预测误差和物理残差均增大，小样本下模型仍保持较低预测误差。

本文结果基于单台 350 MW 燃煤机组的现有运行数据，模型在其他机组、煤种及更广泛变负荷工况下的适用性仍需通过扩大数据来源与工况覆盖范围进一步验证。

［参 考 文 献］

[1] Gao Y K, Zeng D L, Liu J Z. Modeling of a medium speed coal mill[J]. Powder Technology, 2017, 318: 214-223.

[2] Li X W, Wu Y C, Chen H F, et al. Coal mill model considering heat transfer effect on mass equations with estimation of moisture[J]. Journal of Process Control, 2021, 104: 178-188.

[3] 王朝阳，石慧，王文毓，等. MPS 型磨煤机制粉系统动态模型与瞬态运行优化控制研究[J]. 中国电机工程学报, 2024, 44(16): 6518-6528.
Wang C Y, Shi H, Wang W Y, et al. Study on dynamic model and transient operation optimization control of the MPS coal mill pulverizing system[J]. Proceedings of the CSEE, 2024, 44(16): 6518-6528.

[4] 景峥，伍克宁，刘百慧，等. 燃煤机组中速磨煤机制粉系统建模及运行特性研究[J]. 动力工程学报，2026, 46(1): 32-41.
Jing Z, Wu K N, Liu B H, et al. Study on Modeling and Operation Characteristics of Pulverized-coal Preparation System with Medium Speed Coal Mill for Coal-fired Units[J]. Journal of Chinese Society of Power Engineering, 2026, 46(1): 32-41.

[5] Pullanikkattil S, Yerolla R, Vilanova R, et al. Interpretable machine learning model for temperature prediction in coal pulverizer of thermal power plants[J]. International Journal of Coal Preparation and Utilization, 2026, 46(3): 679-703.

[6] 牛玉广, 徐健, 杜鸣, 等. 基于时空图注意网络的磨煤机故障预警[J]. 动力工程学报, 2025, 45(9): 1519-1526.
Niu Y G, Xu J, Du M, et al. Spatio-temporal Graph Attention Network-based Fault Warning of Coal Mill[J]. Journal of Chinese Society of Power Engineering, 2025, 45(9): 1519-1526.

[7] 吴婷婷, 任少君, 朱保宇, 等. 基于物理信息神经网络的中速磨煤机动态模型参数辨识方法[J]. 热能动力工程, 2025, 40(7): 154-163.
Wu T T, Ren S J, Zhu B Y, et al. Dynamic Parameter Identification Method for Medium Speed Coal Mill based on Physics-Informed Neural Networks[J]. Journal of Engineering for Thermal Energy and Power, 2025, 40(7): 154-163.

[8] 邓志平, 任少君, 翁琪航, 等. 基于物理信息神经网络的生物质气化产物分布预测方法[J]. 动力工程学报, 2024, 44(5): 719-726.
Deng Z P, Ren S J, Weng Q H, et al. Prediction Method for Biomass Gasification Product Distribution Based on Physics-informed Neural Network[J]. Journal of Chinese Society of Power Engineering, 2024, 44(5): 719-726.

[9] 胡溆昂, 郭文康, 李益国. 基于物理单调最小二乘支持向量机的锅炉燃烧 NOx 生成浓度预测方法[J]. 动力工程学报, 2026, 46(5): 137-145.
Hu X, Guo W K, Li Y G. Prediction method of boiler combustion NOx generation concentration based on physics-constrained monotonic least squares support vector machine[J]. Journal of Chinese Society of Power Engineering, 2026, 46(5): 137-145.

[10] Zhang Z Y, Zhang L, Gao X Y, et al. A physics-informed BiLSTM for heat load forecasting at district heating substations[J]. Journal of Building Engineering, 2026, 125: 116119.

[11] Raissi M, Perdikaris P, Karniadakis G E. Physics-informed neural networks: a deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations[J]. Journal of Computational Physics, 2019, 378: 686-707.

[12] Hochreiter S, Schmidhuber J. Long short-term memory[J]. Neural Computation, 1997, 9(8): 1735-1780.

[13] Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need[C]//Advances in Neural Information Processing Systems 30. Red Hook: Curran Associates, Inc., 2017: 5998-6008.

[14] Torres J F, Hadjout D, Sebaa A, et al. Deep learning for time series forecasting: a survey[J]. Big Data, 2021, 9(1): 3-21.

[15] Willard J, Jia X W, Xu S M, et al. Integrating scientific knowledge with machine learning for engineering and environmental systems[J]. ACM Computing Surveys, 2023, 55(4): 66.

[16] Karniadakis G E, Kevrekidis I G, Lu L, et al. Physics-informed machine learning[J]. Nature Reviews Physics, 2021, 3(6): 422-440.

[17] Cuomo S, Schiano Di Cola V, Giampaolo F, et al. Scientific machine learning through physics-informed neural networks: where we are and what's next[J]. Journal of Scientific Computing, 2022, 92: 88.

[18] Akiba T, Sano S, Yanase T, et al. Optuna: a next-generation hyperparameter optimization framework[C]//Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. New York: ACM, 2019: 2623-2631.

[19] Watanabe S. Tree-structured Parzen estimator: understanding its algorithm components and their roles for better empirical performance[EB/OL]. arXiv:2304.11127, 2023[2026-08-30]. https://arxiv.org/abs/2304.11127.

[20] Chicco D, Warrens M J, Jurman G. The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation[J]. PeerJ Computer Science, 2021, 7: e623.