Chemical Engineering Research and Design 227 (2026) 1–22

Contents lists available at ScienceDirect

Chemical Engineering Research and Design

journal homepage: www.elsevier.com/locate/cherd

A systematic hybrid mechanistic–machine learning framework for catalytic reactor modelling and computational validation using CO oxidation

Ebenezer Aquisman Asare [a,*], Dickson Abdul-Wahab [b], Elsie Effah Kaufmann [c], Rafeah Wahi [d], Zainab Ngaini [d], Abigail Ampadu [b]

[a] *Nuclear Chemistry and Environmental Research Centre, National Nuclear Research Institute (NNRI), Ghana Atomic Energy Commission (GAEC), Box LG 80, Legon, Accra, Ghana*
[b] *Department of Nuclear Science and Applications, School of Nuclear and Allied Sciences, University of Ghana, Atomic-Kwabenya, Accra, Ghana*
[c] *Department of Biomedical Engineering, University of Ghana, Legon, Accra, Ghana*
[d] *Department of Chemistry, Faculty of Resource Science and Technology, Universiti Malaysia Sarawak, Kota Samarahan, Sarawak, Malaysia*

ARTICLE INFO

Keywords:
Ordinary-differential-equation
Residual-learning data
Catalytic systems
Gaussian-process regression

ABSTRACT

Accurately forecasting the fast transients that govern catalytic reactors remains difficult because first-principles ordinary differential equation (ODE) models neglect unmodelled heat and mass-transfer effects and therefore perform poorly (baseline CO-oxidation rate $R^2 = -0.231$). For the above reason, this study presents a systematic hybrid mechanistic machine-learning (ML) framework that couples a physically rigorous CSTR model with data-driven residual learning to close these physics gaps. A six-factor design of experiments generated 500 operating scenarios, and after simulation, quality screening, derivative estimation, and residual/outlier filtering, the residual-learning dataset comprised approximately 33,096 usable samples. Five regressors (XGBoost, LightGBM, SVR, MLP and sparse Gaussian-process regression) were hyperparameter-tuned with Optuna and blended through weight optimisation. Uncertainty was propagated with GP posterior bands and inter-model disagreement. The optimised ensemble lifted test-set accuracy to $R^2 = 0.755$, RMSE $= 0.006$ $mol{\cdot}m^{-3}{\cdot}s^{-1}$ and MdAPE $= 93$ % a dramatic recovery over the mechanistic baseline. $\pm 2\sigma$ GP bands captured 94 % of unseen points, providing actionable epistemic bounds. Performance deteriorated by only ~21 % when 5 % Gaussian sensor noise was injected, confirming robustness for on-line use. By modularising experiment design, physics-guided feature engineering, automated model selection, and calibrated uncertainty quantification, this workflow delivers interpretable, real-time-capable surrogate models within the modelled operating envelope, outperforming pure ODE and single-model ML baselines. The protocol is transferable to other catalytic systems and establishes a reproducible path toward uncertainty-aware reactor optimisation and control.

1. Introduction

The accurate prediction and control of catalytic reactor dynamics remains a fundamental challenge in chemical process engineering, particularly during transient operations where traditional mechanistic models fail to capture rapid kinetic and transport phenomena. Recent studies show that both mechanistic and data-driven models often fall short in capturing catalytic CO oxidation complexity, with hybrid approaches increasingly needed (Peterson et al., 2024). The limitations of mechanistic approaches become pronounced during step-change scenarios, where transport limitations skew ordinary differential equation (ODE) predictions in exothermic reactions, including heat and mass transfer effects causing discrepancies between predictions and experimental outcomes (Das et al., 2025). A study comparing data-driven and hybrid models in CO_2 methanation found that hybrid models do not consistently outperform data-driven ones, indicating model performance is problem dependent (Peterson et al., 2024). Machine learning revealed changes in dominant mechanisms during oxygen buildup in transient CO oxidation data, something static mechanistic models fail to capture (Kunz et al., 2021). Time resolved spectroscopy showed CO oxidation over Au/TiO_2 involves a dual mechanism changing over time, with mechanistic models failing to describe such transitions (Wang and Bürgi, 2023). A mechanistic study of CO oxidation on Pd/Al_2O_3 showed that reactor conditions fail to measure true reaction rates accurately,

---

* Corresponding author.
E-mail address: aquisman1989@gmail.com (E.A. Asare).

https://doi.org/10.1016/j.cherd.2026.01.039
Received 28 October 2025; Received in revised form 20 December 2025; Accepted 18 January 2026

providing corrections for temperature and concentration gradients (Ren et al., 2019). These limitations necessitate approaches that can compensate for unmodeled physics while preserving mechanistic frameworks' interpretability. Real time reactor optimization and advanced process control strategies amplify these modelling challenges. Modern reactor control requires predictive models capable of handling rapid transients, uncertainty propagation, and safe operation under varying conditions. Residual learning augments mechanistic models with data-driven corrections, leveraging known physics while fitting only the discrepancy, which reduces overfitting and improves generalization compared to purely black-box models (Silver et al., 2019). This approach addresses the need for models that capture complex, unmodeled dynamics while maintaining physical interpretability essential for control applications.

The challenge of bridging first-principles knowledge with data-driven insights becomes acute when researchers have quantified parameter identifiability using Kron reduction with weighted least squares to estimate parameters with partial data and complex biases (Gasparyan and Rao, 2023). These developments highlight the need for systematic frameworks addressing both structural model deficiencies and parameter uncertainty simultaneously. Traditional mechanistic modelling approaches rely on first-principles descriptions incorporating mass balances, energy balances, surface reaction kinetics, and competitive adsorption equilibria, using Langmuir-Hinshelwood or Mars-van Krevelen mechanisms for CO oxidation systems (Lashina et al., 2023a). These frameworks provide thermodynamically consistent predictions but underpredict initial CO oxidation rate peaks and their decay toward steady state, especially at higher feed flows and concentrations (Nielsen et al., 2020). Comparative studies show both Langmuir-Hinshelwood and Mars-van Krevelen mechanisms are used for CO oxidation, with MvK kinetics validated for higher temperatures while low-temperature associative mechanisms apply below 100°C for Pt/CeO_2 systems (Lashina et al., 2023a). Microkinetic modelling of Pt/TiO_2 catalysts finds both L-H and Eley-Rideal steps relevant under operando conditions. The challenge of mechanism selection is complicated by variations in activation energies across catalysts, with Pt, Pd, and Au catalysts showing energies of 40–90 kJ/mol in flow reactors, while transition metal oxides exhibit 20–60 kJ/mol due to lattice oxygen mobility (Neumann et al., 2019).

Machine learning applications in chemical process modelling have shown significant potential for capturing nonlinear relationships that traditional mechanistic approaches struggle with. Tree-based models, particularly XGBoost and LightGBM, achieve superior predictive accuracy with hyperparameter ranges of 50–200 trees, depths of 3–10, and learning rates of 0.01–0.2 for catalytic kinetics applications (Lin et al., 2022). LightGBM provides training speedups of 3 × to 23 × compared to XGBoost through histogram-based feature binning and leaf-wise tree growth (Rahman et al., 2024). However, data-driven approaches lack physical interpretability and may not generalize outside their training domains. Hybrid modelling frameworks combine mechanistic and data-driven approaches while addressing their limitations. Physics-informed neural networks incorporate physical laws as constraints, allowing robust learning with sparse data for process modelling and control (Velioglu et al., 2025). Neural network-enhanced kinetics demonstrates hybrid integration with reaction models using Bayesian estimation (Vuolio et al., 2023). Sum-of-squares constrained regression enforces physical laws in hybrid chemical reactor models for reliable extrapolation (Pitarch et al., 2021).

Existing hybrid frameworks face several limitations. Scalability remains challenging, as models validated on small systems may not generalize to industrial applications (Cui et al., 2024). Uncertainty quantification represents a critical gap, with most models lacking rigorous UQ methods (Vuolio et al., 2023). Models struggle with feedback loops between physical and learned components (Agarwal et al., 2024). Additionally, hybrid models require structured domain-specific data and manually engineered coupling between physics and ML components (Panosetti et al., 2023). The computational validation of hybrid models reveals limitations in current methodologies. Studies rarely provide complete validation from experimental design to uncertainty analysis and control implications (Bradley et al., 2022). Systematic frameworks for hybrid model development are uncommon, with most works using ad hoc approaches that limit scalability and reproducibility (Nielsen et al., 2020). These gaps highlight the need for comprehensive frameworks across catalytic reactor applications. Residual learning for mechanistic model correction augments physics-based models with data-driven corrections, leveraging known principles while fitting only the discrepancy between predictions and observations. This approach reduces overfitting and improves generalization compared to black-box models (Silver et al., 2019). Theoretical work shows residual learning preserves system properties like stability, making corrections safer than purely data-driven alternatives (Yang et al., 2025). The earliest and influential citations combining ODE solvers with data-driven residual learning include the Generalized Reactor Neural ODE (GRxnODE) framework, which integrates ODE solvers with neural networks, achieving residual learning directly within the ODE framework for chemical reactor modelling (Yin et al., 2023).

Residual connections facilitate optimization in deep networks, making them easier to train and less prone to vanishing gradients, benefiting the capture of "missing physics" without distorting well-modelled components (Xu et al., 2025). Residual learning effectively corrects systematic model errors when physics-based models are imperfect, outperforming direct parameter tuning for features not captured by mechanistic models (Razak et al., 2023). This approach yields lower error and better captures complex biases than traditional calibration (Bum Kim et al., 2021). Hybrid machine learning frameworks and Physics-Informed Neural Networks differ in data efficiency and flexibility. Hybrid models need less data than black-box models but more than PINNs, as PINNs encode physical laws as constraints, making them data-efficient for systems with known physics (Bradley et al., 2022). However, hybrid models better handle model mismatch, though requiring more training data (Narayanan et al., 2022). Studies show PINNs outperform hybrid models with scarce data, while hybrid models excel with more data and model misspecification (Yin et al., 2023).

Ensemble methods provide superior robustness and uncertainty quantification compared to single-model approaches. Stacking and Bayesian model averaging deliver robust bias-variance trade-offs in chemical and engineering applications, outperforming simple bagging and boosting, especially in heterogeneous datasets (Shahhosseini et al., 2022). Stacking enables learning optimal weights for base learners through cross-validation, while Bayesian model averaging weighs models by posterior probabilities, improving robustness in process design and parameter estimation (Wakayama and Sugasawa, 2024). Optimal ensemble compositions typically involve 3–5 base learners, with studies showing 3 diverse learners can provide near-optimal accuracy at lower computational cost (Fu et al., 2025). Ensemble disagreement, measured as prediction spread among members, serves as an epistemic uncertainty proxy in reactor control applications, with high disagreement indicating areas of low confidence (Zhu et al., 2025). Sparse weight assignment often improves robustness against noise by ignoring poorly performing models, with adaptive sparsity increasing ensemble robustness in clustering and supervised learning (Timpl et al., 2022).

Despite these advantages, several criticisms have been raised against hybrid machine learning approaches in process systems engineering. Interpretability concerns arise because many hybrid ML models, especially those using neural networks for residual learning, remain "black-box" in the learned residuals, limiting scientific insight and making it difficult for engineers to trust or diagnose model outputs (Narayanan et al., 2022). Generalization challenges include potential overfitting to training data if the mechanistic portion is poor or if the ML component is overly complex, leading to unreliable extrapolation to new conditions, with success being sensitive to the quality and range of training data and correctness of mechanistic model structure (Yin et al., 2023). Computational cost issues persist, as training hybrid ODE-neural models or PINNs can be computationally intensive, particularly for stiff systems or large-scale reactor networks, with some methods requiring repeated ODE integration during training, increasing hardware and time demands (Roesch et al., 2021).

This research aims to develop and validate a systematic framework for hybrid mechanistic-machine learning models that addresses existing limitations while providing robust, interpretable predictions for catalytic reactor control. Given that recent reviews emphasize the need for systematic frameworks over ad hoc approaches in hybrid ML model development (Nielsen et al., 2020), this work establishes formal, modular pipelines for model development. The framework integrates uncertainty quantification and interpretability methods to address the gap between experiment, uncertainty, and control (Bradley et al., 2022). A key objective is evaluating residual learning effectiveness for correcting unmodeled kinetics and transport phenomena in catalytic systems. Building on evidence that residual learning effectively corrects systematic model errors when physics-based models are imperfect (Razak et al., 2023), this work quantifies conditions where residual corrections are most beneficial. The research addresses hybrid models' challenges with physical and learned component coupling by developing systematic approaches (Agarwal et al., 2024).

The novel contributions of this framework address key gaps in hybrid modelling through several methodological advances. First, it provides comprehensive end-to-end validation spanning design of experiments, model training, uncertainty quantification, and control implications, addressing the limitation that such validation remains rare (Bradley et al., 2022). Second, the framework introduces systematic ensemble optimization using stacking and Bayesian model averaging for robust bias-variance trade-offs in chemical engineering, improving upon simple equal weighting schemes (Shahhosseini et al., 2022). The framework's impact stems from enhanced dynamic predictions with quantified uncertainty bounds for safe reactor control. Implementation of hybrid modelling architectures motivated by digital-twin concepts has shown improvements in dynamical prediction and operational flexibility (Rubio and Giménez, 2022). The framework (Fig. 1) provides uncertainty quantification and interpretable residual corrections enabling proactive control while addressing challenges in deploying hybrid ML models on embedded hardware (Joshi et al., 2024). This study presents a methodology for developing hybrid mechanistic-machine learning models through computational validation using CO oxidation. The methodology establishes the mechanistic reactor model with mass balances, energy balances, reaction kinetics, and adsorption equilibria, followed by protocols for synthetic data generation using six-dimensional design of experiments. The data processing includes derivative calculation methods, outlier filtering, and feature engineering incorporating physics-informed variables to enhance model interpretability.

The machine learning model development section details the implementation of five base learners including gradient boosting methods (XGBoost, LightGBM), support vector regression, multi-layer perceptron, and sparse Gaussian process regression, with systematic hyperparameter optimization conducted via Tree-structured Parzen Estimator sampling. The ensemble construction methodology incorporates Optuna-driven weight optimization to maximize validation performance, addressing documented advantages of systematic blending over equal weighting schemes (Shahhosseini et al., 2022). Uncertainty quantification protocols encompass both ensemble disagreement metrics and SGPR-based epistemic uncertainty bands, providing the comprehensive uncertainty assessment essential for safety-critical applications where coverage within ±2σ is typically required for acceptable uncertainty bounds (Psaros et al., 2023). This comprehensive approach positions the work within the emerging paradigm of systematic hybrid modelling that addresses documented gaps in scalability, uncertainty quantification, and systematic framework development that have limited previous hybrid modelling efforts (Nielsen et al., 2020). The methodology provides a foundation for extending hybrid approaches to diverse catalytic systems while maintaining the physical interpretability and uncertainty awareness essential for safe reactor operation, contributing to the broader goal of bridging first-principles knowledge with data-driven insights in chemical process engineering.

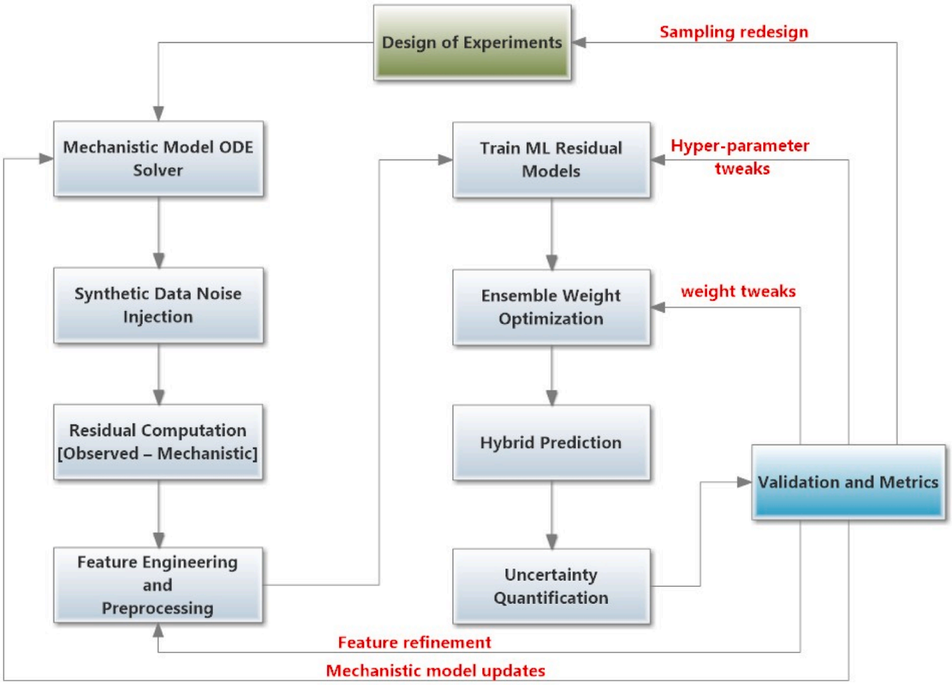


Fig. 1. Hybrid mechanistic-machine learning framework architecture.

2. Methodology

2.1. Mechanistic reactor model

2.1.1. Reactor configuration and modelling assumptions

A jacketed continuous stirred-tank reactor (CSTR) operated in the gas phase is considered, with inlet volumetric flow rate, inlet compositions (CO and O_2), and inlet temperature defining the residence time and boundary conditions in the bulk-phase balances. The bulk gas is assumed ideally mixed, yielding spatially uniform bulk concentrations and temperature. A representative porous catalyst pellet is embedded in the bulk and modelled as a lumped solid phase with its own temperature state coupled to the bulk by an effective pellet-to-bulk heat-transfer conductance, while heat removal to an external coolant is represented through an overall heat-transfer coefficient and a coolant temperature boundary condition. Fig. 2 provides a schematic of the modelling abstraction and the correspondence between bulk, pellet, and jacket states in the governing equations.

2.1.2. Governing equations and kinetics

The reactor dynamics were represented by a system of ordinary differential equations that combined mass balances for CO and O_2, energy balances for bulk and pellet phases, surface-reaction kinetics, adsorption/desorption equilibria, and catalyst deactivation.

Mass balances:

$$\frac{dC_{CO}}{dt} = \frac{C_{CO,in} - C_{CO}}{\tau}|-|r, \frac{dC_{CO_2}}{dt} = \frac{dC_{CO_2,in} - C_{O_2}}{dt}|-|\frac{1}{2}r \tag{1}$$

where C_{CO} and C_{CO_2} denote reactor-bulk concentrations, $C_{CO,in}$ and $C_{O_2,in}$ are inlet concentrations, and the residence time was defined as $\tau = \frac{V}{F}$, with reactor volume V and volumetric flow rate F.

For reaction rate, the net CO-oxidation rate r (mol m^{-3} s^{-1}) was expressed as

$$r = \eta \alpha k_{\mathrm{intr}} \theta_{\mathrm{CO}} \theta_{\mathrm{O}}, \tag{2}$$

where, α is the catalyst activity (dimensionless), $k_{\mathrm{intr}} = k_0 \exp(-E_a/(RT_p))$ is the intrinsic rate constant (s^{-1}) at pellet temperature T_p, θ_{CO} and θ_{O} are surface coverages, η is the effectiveness factor accounting for internal diffusion.

For adsorption/desorption equilibria, adsorption constants were given by

$$K_{\mathrm{CO}}(T) = K^0_{\mathrm{CO}} \exp\left(-\frac{\Delta H_{\mathrm{ads,CO}}}{RT}\right), K_{\mathrm{O_2}}(T) = K^0_{\mathrm{O_2}} \exp\left(-\frac{\Delta H_{\mathrm{ads,O_2}}}{RT}\right), \tag{3}$$

and surface coverages by

$$\theta_{\mathrm{CO}} = \frac{K_{\mathrm{CO}} C_{\mathrm{CO}}}{1 + K_{\mathrm{CO}} C_{\mathrm{CO}} + \sqrt{K_{\mathrm{O_2}} C_{\mathrm{O_2}}}}, \theta_{\mathrm{O}} = \frac{\sqrt{K_{\mathrm{O_2}} C_{\mathrm{O_2}}}}{1 + K_{\mathrm{CO}} C_{\mathrm{CO}} + \sqrt{K_{\mathrm{O_2}} C_{\mathrm{O_2}}}} \tag{4}$$

2.1.3. Surface-species assumptions

The mechanistic kinetics employs a reduced surface description with two lumped adsorbate families. The CO coverage term represents molecularly adsorbed CO on catalytic sites (CO*). The oxygen coverage term represents an effective oxygen-derived adsorbate pool formed by O_2 adsorption on the surface; different oxygen charge states or molecular intermediates (e.g., superoxide/peroxide species) are not distinguished explicitly in this global model. Similarly, no explicit cationic CO species is introduced. The global rate expression therefore captures the net consumption of CO* and the oxygen-derived adsorbate pool to form CO_2(g), while the adsorption constants parameterize the net site occupancy contributions of CO and O_2 under the operating conditions considered.

Considering heat balances, reactor-bulk temperature T_b and pellet temperature T_p were governed by

$$\frac{dT_b}{dt} = \frac{T_{in} - T_b}{\tau}|+|\frac{UA}{\rho C_p V}(T_c - T_b)|+|\frac{h_{eff}}{\rho C_p}(T_p - T_b) \tag{5}$$

$$\frac{dT_b}{dt} = \frac{\Delta H_r}{\rho_p C_{p,p}} r|+|\frac{h_{eff}}{\rho_p C_{p,p}}(T_b - T_p) \tag{6}$$

where T_{in} and T_c are inlet and coolant temperatures, UA is the overall heat-transfer coefficient, h_{eff} the pellet-to-bulk conductance, ΔH_r the reaction enthalpy, and $\{\rho, C_p\}, \{\rho_p, C_{p,p}\}$ the densities and heat capacities of bulk fluid and catalyst pellet, respectively.

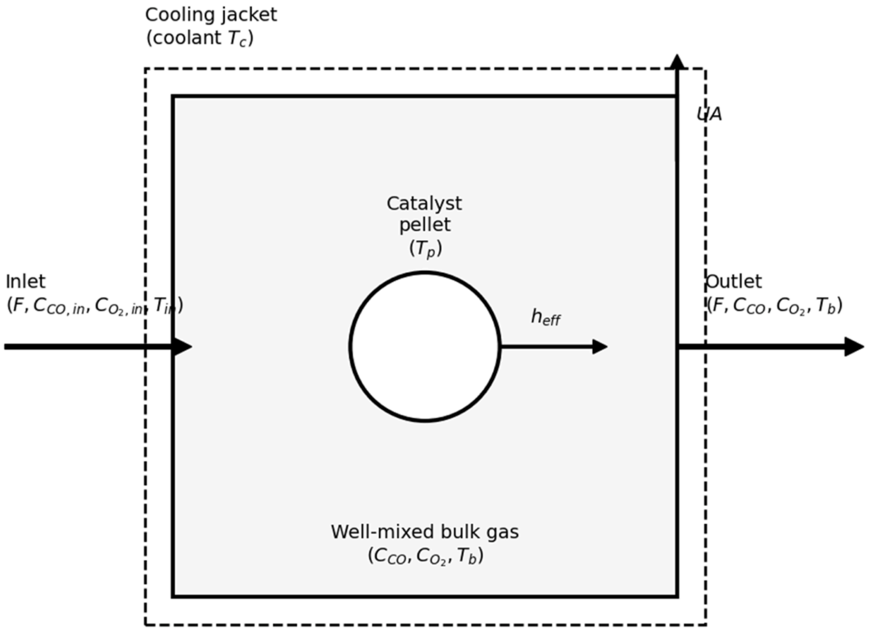


Fig. 2. Schematic representation of the jacketed CSTR configuration. The model assumes a lumped solid phase (catalyst pellet) with temperature T_p embedded in an ideally mixed bulk gas phase at T_b. Heat is exchanged between the pellet and bulk via conductance h_{eff}, and removed from the bulk to the cooling jacket (T_c) via coefficient UA.

For catalyst deactivation, the activity decay was described by first-order kinetics:

$$\frac{d_a}{dt} = -k_d \exp\left(-\frac{E_d}{RT_p}\right)\alpha \tag{7}$$

All parameters (k_0, E_a, adsorption heats ΔH_{ads}, deactivation k_d, E_d, etc.) were listed in Table 1. All parameters were drawn from literature sources reporting experimental validation of the underlying kinetic/thermochemical quantities; however, the present study evaluates the full coupled ODE reactor model computationally (in silico) rather than through direct comparison to new experimental reactor datasets.

2.1.4. Numerical integration & solver settings

The coupled ODE system was integrated using SciPy's LSODA algorithm via `odeint`. Initial conditions for each simulation were set as$C_{CO}(0)$=$0.8C_{CO,in}$,$C_{O2}(0)$= 0.18,$T_b(0)$= 480|K,$T_p(0)$ = 485|K,α(0)= 1.0 and a time grid of N points over $\left[0, t_{\text{span}}\right]$ as specified in the configuration. Integration tolerances were applied ensuring solution accuracy and stability. Integration tolerances were carefully configured to ensure solution accuracy and stability while maintaining computational efficiency throughout the parameter space exploration. The coupled ODE system, comprising mass balances for CO and O_2 (Eq. 1), energy balances for bulk and pellet phases (Eqs. 5–6), surface reaction kinetics (Eq. 2), competitive Langmuir adsorption equilibria (Eqs. 3–4), and first-order catalyst deactivation (Eq. 7), converged successfully under all tested conditions. The solver successfully handled the stiff nature of the reactor dynamics, particularly during rapid transients following step changes in operating conditions. Non-finite solutions were systematically flagged and discarded during the data generation process, ensuring that only physically meaningful trajectories were retained for subsequent analysis. The integration was performed over time grids of N points spanning $[0, t_{span}]$ intervals, with adaptive step-size control maintaining numerical stability across varying dynamic regimes from slow thermal equilibration to fast surface kinetics Fig. 3.

Table 1
Mechanistic model parameterization: pre-exponential factors, activation energies, and physical constants used in the CO Oxidation Model, with literature sources.

Parameter	This Study	Literature Range	Source
Global activation energy (E_a)	80,000 $J{\cdot}mol^{-1}$	80–100 kJ/mol	(Decarolis et al., 2024; Eads et al., 2025; Che-Galicia et al., 2021)
Reaction enthalpy (ΔH_{rxn})	–283,000 $J{\cdot}mol^{-1}$	–283 kJ/mol (standard)	(Rogal et al., 2007)
CO adsorption enthalpy ($\Delta H_{ads,CO}$)	–50,000 $J{\cdot}mol^{-1}$	–30 kJ/mol (Au); –100 to –200 kJ/mol (metals)	(Che-Galicia et al., 2021; Grabow et al., 2010)
O_2 adsorption enthalpy ($\Delta H_{ads,O_2}$)	–25,000 $J{\cdot}mol^{-1}$	–13 kJ/mol (Au); –10 to –30 kJ/mol	(Che-Galicia et al., 2021; Peterlinz and Sibener, 1995)
Global pre-exponential factor (k_o)	1.0×10^6 s^{-1}	Varies (e.g., 1.5×10^3 s^{-1})	
CO adsorption constant ($K_{CO,0}$)	2.0 $m^3{\cdot}mol^{-1}$	Varies (model-dependent)	
O_2 adsorption constant ($K_{O_2,0}$)	8.0 $m^3{\cdot}mol^{-1}$	Varies (model-dependent)	
Catalyst deactivation rate constant (k_d)	0.1 s^{-1}	Varies (model-dependent)	
Pellet density (ρ_{pellet})	1200 kg/m^3	This study (model configuration: config.yaml; defaults in code)	
Pellet heat capacity ($C_{p,pellet}$)	1000 J/(kg·K)	This study (model configuration: config.yaml; defaults in code)	
Volumetric heat capacity ($\rho_{pellet}{\cdot}C_{p,pellet}$)	1.2×10^6 J/(m^3·K)	This study (computed from configured values)	
Deactivation activation energy (E_d)	50,000 $J{\cdot}mol^{-1}$	Varies (model-dependent)	

2.1.5. Model parameterization

Mechanistic model parameters were drawn directly from the configuration (Table 1) and expressed in consistent units to ensure reproducibility. For Pellet material and thermophysical properties, the term "pellet" denotes the catalyst body and is treated as a homogeneous porous solid with lumped (effective) density and heat capacity parameters in the pellet energy balance. As the present study validates the hybrid mechanistic–machine learning framework using synthetic trajectories, the mechanistic model is not tied to a single experimentally manufactured catalyst formulation (support/active phase); rather, the reported kinetic and adsorption parameters represent an effective CO-oxidation catalyst within the operating envelope considered. To avoid ambiguity, the assumed pellet material abstraction and the corresponding thermophysical properties are explicitly documented alongside the kinetic parameters in Table 1. The Arrhenius parameters for the intrinsic reaction rate were defined as $k_0\left(s^{-1}\right), E_a\left(J\ mol^{-1}\right)$, such that the temperature-dependent rate constant was $k_{\text{intr}}(T) = k_0 \exp\left(-\frac{E_a}{RT}\right)$.

Adsorption enthalpies $\Delta H_{\text{ads,CO}}$ and $\Delta H_{\text{ads,}O_2}$ ($J\ mol^{-1}$) and pre-exponential adsorption constants K^0_{CO} and $K^0_{O_2}$ ($m^3\ mol^{-1}$) were specified for the competitive Langmuir isotherm. Bulk and pellet physical constants comprised density ρ ($kg\ m^{-3}$), heat capacity C_p ($J\ kg^{-1}\ K^{-1}$), pellet density ρ_p, pellet heat-capacity $C_{p,p}$, reactor volume V (m^3), overall heat-transfer coefficient UA ($W\ K^{-1}$), and effective internal conductance h_{eff} ($W\ m^{-3}\ K^{-1}$). The universal gas constant R ($J\ mol^{-1}\ K^{-1}$) was also included. Catalyst deactivation parameters k_d (s^{-1}) and E_d ($J\ mol^{-1}$) were adopted to enforce first-order decay Table 2.

2.2. Synthetic data generation

2.2.1. Design of experiments

The synthetic data generation employed a Latin Hypercube Sampling (LHS) strategy to systematically explore the six-dimensional reactor operating space. Implementation utilized the 'scipy.stats.qmc.LatinHypercube' sampler with a fixed random seed (seed=42) to ensure reproducibility. This method generated 500 unique operating scenarios ($N_{scenarios} = 500$) across the factor ranges defined in Table 3. The sampling strategy was configured to ensure adequate coverage of the reactor operating envelope, bounded by the following parameter ranges: Flow rate (F): 0.005–0.03 mol/s, Inlet CO concentration ($C_{CO,in}$): 0.06–0.18 mol/L, Coolant temperature (T_c): 440–485 K, Uncertainty factor: 0.6–1.4 (9 levels), Noise level: 0.0–0.05 (3 levels), Sampling rate: 50–100 Hz. While the framework supports full factorial designs for exhaustive enumeration of discrete levels, LHS was selected for this study to maximize space-filling properties within the high-dimensional parameter space. This approach ensures efficient coverage of edge cases and interaction effects in flow rate, concentration, and temperature that grid-based designs might miss at a comparable computational cost. The feed is assumed dry (CO/O_2 in an inert carrier), and no water-vapor or hydroxyl-related factors are included in the design·

2.2.2. Scenario simulation and noise injection

For each scenario tuple $\left(F_i, C_{\text{CO,in},i}, T_{c,i}, \text{uf}_i, \text{nl}_i, n_i\right)$, the state vector

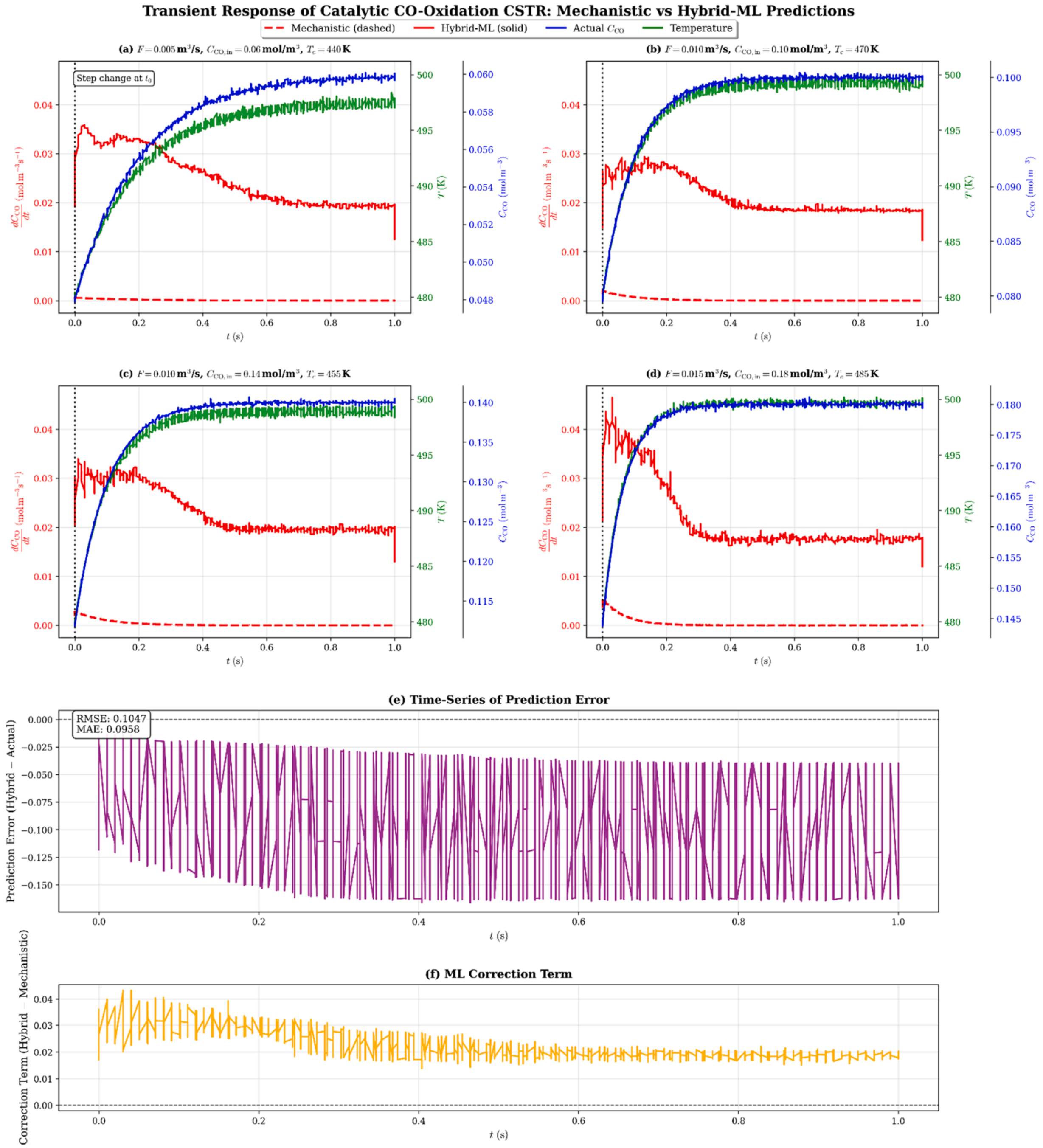


Fig. 3. Transient Response of Catalytic CO-Oxidation in a CSTR under Four Operating Conditions: Comparison of Pure Mechanistic (red dashed) vs. Hybrid-ML (red solid) Predictions for the CO–Oxidation Rate dC_{CO}/dt, Actual CO Concentration (blue) and Reactor Temperature (green); (e) Time-series of Hybrid Model Prediction Error (Hybrid – Actual) with RMSE = 0.1047, MAE = 0.0958; (f) Learned ML Correction Term (Hybrid – Mechanistic) over Time.

$$y(t)| = |\left[C_{CO}(t),\ C_{O_2}(t),\ T_b(t),\ T_p(t),\ \alpha(t)\right]^{!T} \tag{8}$$

was time-integrated over $t \in [0, t_{span}]$ with n_i equally spaced points. Thereafter, additive Gaussian noise was introduced as

$$y_{noisy}(t) = y(t)| + |\mathrm{nl_i}\left|\mathrm{diag}(\sigma_y)\ \varepsilon, \varepsilon \sim \mathcal{N}(0, I),\right. \tag{9}$$

where σ_y denotes the empirical standard deviation of each state component over time and $\mathrm{nl_i}$ is the noise-level factor. Post-noise clipping enforced physical bounds: $C_{CO},\ C_{O_2} \geq 0$, $T_b \in [T_{min},\ T_{max}]$, $T_p \in [T_{min},\ T_{max} + 100]$, $\alpha \in [0, 1]$.

2.2.3. Data quality checks

After data generation, basic quality-control metrics were computed on the aggregated dataset. For missing-value detection, total number of NaN entries per column was tabulated; any nonzero count was flagged. And for outlier identification in each numeric column x, the inter-quartile range $\mathrm{IQR} = Q_3 - Q_1$ was used to flag points satisfying $x\langle Q_1 - 1.5\ \mathrm{IQR}$ or $x\rangle Q_3 + 1.5\ \mathrm{IQR}$., and the aggregate outlier count was recorded. If missing value or outlier counts exceeded nominal thresholds (e.g. >1 % of total samples), warnings were issued, and the scenario set was either reviewed or regenerated with adjusted noise/uncertainty parameters.

Table 2
Synthetic dataset summary: feature names, sampling ranges, distributions, noise levels, and total sample counts for residual-learning.

Feature	Description	Range/Values	Type	Distribution	Levels	Total Samples
C_{CO}	CO Concentration	[0.05, 0.18]	continuous	Uniform	-	33096
C_{O_2}	O_2 Concentration	[0.18, 0.20]	continuous	Uniform	-	33096
T_{bulk}	Bulk Temp	[479.5, 500.4]	continuous	Uniform	-	33096
T_{pellet}	Pellet Temp	[485.0, 486.2]	continuous	Uniform	-	33096
α	Catalyst Activity	[1.0, 1.0]	continuous	Uniform	-	33096
F	Flow Rate	[0.005, 0.01, 0.015, 0.02, 0.03]	categorical	Discrete	5	33096
$C_{CO_{in}}$	CO Inlet	[0.06, 0.1, 0.14, 0.18]	categorical	Discrete	4	33096
T_c	Coolant Temp	[440, 455, 470, 485]	categorical	Discrete	4	33096
U_f	Uncertainty Factor	[0.6, 0.7, 0.8, 0.9, 1.0, 1.1, 1.2, 1.3, 1.4]	categorical	Discrete	9	33096
noise	Noise Level	[0.0, 0.02, 0.05]	categorical	Discrete	3	33096
$n_{samples}$	Sampling Rate	[50, 100]	categorical	Discrete	2	33096
Scenario	Scenario ID	[1, 500]	integer	Sequential	500	33096
t	Time	[0.0, 1.0]	continuous	Uniform	-	33096

Table 3
Design of experiments for framework validation.

Variable	Values	Units	Levels	Fixed Parameters
F	[0.005, 0.01, 0.015, 0.02, 0.03]	mol/s	5	-
$C_{CO_{in}}$	[0.06, 0.1, 0.14, 0.18]	mol/L	4	-
T_c	[440, 455, 470, 485]	K	4	-
Uncertainty factor	[0.6, 0.7, 0.8, 0.9, 1.0, 1.1, 1.2, 1.3, 1.4]	-	9	-
Noise level	[0.0, 0.02, 0.05]	-	3	-
$n_{samples}$	[50, 100]	Hz	2	-
$C_{O_{2in}}$	-	mol/L		0.2
T_{in}	-	K		500
Time span	S		100	

2.3. Data processing & residual-learning dataset

2.3.1. Derivative calculation

Derivatives of the CO concentration time-series were estimated using a three-component ensemble method. First, a Savitzky–Golay filter of window length w and polynomial order p (from configuration: `savgol_window_length` and `savgol_polrder`) was applied to each trajectory $C_{CO}(t)$ to yield a smoothed derivative

$$\left.\frac{dC_{SG}}{dt}\right|_{t_i} = \text{savgol\_filter}(C_{CO}, w, p, \text{deriv} = 1, \Delta t)_i \tag{10}$$

Second, a central-difference derivative was computed on the smoothed signal:

$$\left.\frac{dC_{CD}}{dt}\right|_{t_i} = \frac{\widetilde{C_{i+1}} - \widetilde{C_{i-1}}}{t_{i+1} - t_{i-1}} \tag{11}$$

where $\widetilde{C}$ denotes the Savitzky–Golay–smoothed concentration. Third, a moving-average derivative was formed as

$$\left.\frac{dC_{MV}}{dt}\right|_{t_i} = \frac{C_{i+1} - C_{i-1}}{t_{i+1} - t_{i-1}} \tag{12}$$

th boundary values copied from the nearest interior points. The final derivative estimate was obtained by weighting each component:

$$\frac{dC_{CO}}{dt} = 0.5\,\frac{dC_{SG}}{dt} + 0.3\,\frac{dC_{CD}}{dt} + 0.2\,\frac{dC_{MV}}{dt} \tag{13}$$

Parameter values w and p were set to avoid over-smoothing while reducing noise spikes, as detailed in the `Derivative Calculator` implementation.

2.3.2. Outlier filtering

Outliers in the estimated derivatives were identified via a median absolute deviation (MAD) criterion. The MAD was computed as MAD $= \text{median}_i(|dC_i/dt - \text{median}(dC/dt)|)$, and a threshold factor τ (default 3.5) was applied. Points satisfying

$$\frac{|dc_i/dt - median(dc/dt)|}{MAD} > \tau \tag{14}$$

were flagged as outliers. Each outlier was replaced by linear interpolation between the nearest non-outlier neighbours, ensuring continuity of the derivative series.

For reproducibility, three dataset sizes are distinguished: (i) the raw simulated output produced by integrating the ODE system over all DOE scenarios; (ii) the post-QC dataset after removal of non-finite solutions and invalid trajectories; and (iii) the final residual-learning dataset after derivative estimation and residual/outlier filtering. The results reported in this study use the final residual-learning dataset comprising ≈ 33,096 usable samples across 500 scenarios (Table S1).

2.3.3. Residual target construction

For each time point, the mechanistic prediction of the CO derivative, $dC_{\text{mech}}/dt|$, was obtained by evaluating the first component of the reactor ODE at the observed state $\{C_{CO}, C_{O_2}, T_b, T_p, \alpha\}$. The residual learning target was defined as the difference between the filtered observed derivative and the mechanistic prediction:

$$\Delta \dot{C}_{CO} = \frac{dC_{\text{obs}}}{dt} - \frac{dC_{\text{mech}}}{dt} \tag{15}$$

Points for which $\Delta \dot{C}_{CO}$ was non-finite or exceeded a maximum allowable deviation Δ_{max} were discarded. Scenarios yielding fewer than ten valid residuals were omitted to ensure statistical significance.

2.3.4. Feature engineering and normalization

Several engineered features were appended to the residual dataset: CO/O2 ratio: $\frac{C_{CO}}{C_{O_2} + 10^{-6}}$, normalized temperature: $T_{\text{norm}} = \frac{T_b - 500}{100}$, and normalized flow rate: $F_{\text{norm}} = \frac{F - 0.015}{0.005}$.

Normalization was then applied according to the following standard scaling in which each feature x was transformed as $\widehat{x} = (x - \mu_x)/\sigma_x$, where μ_x and σ_x denote the sample mean and standard deviation. Robust scaling for each feature x which was transformed as $\widehat{x} = (x - \text{median}(x))/\text{IQR}(x)$, where IQR $= Q_3 - Q_1$.

2.4. Machine-learning model development

2.4.1. Base learners

Four supervised regression algorithms were employed as base learners: each within a consistent API for training and prediction. The XGBoost objective was formulated as a regularized additive model of regression trees. Denoting the prediction after m trees by

$$F^{(m)}(x)\big| = \big|F^{(m-1)}(x)\big| + \big|\eta\, f_m(x), \tag{16}$$

with learning rate η and base-learner f_m (a tree with T_m leaves and leaf-weights $\{w_{m,j}\}$), the full training objective minimized was

$$\mathscr{L} = \sum_{i=1}^{Nl} \left(y_i, \left|F^{(M)}(x_i)\right)\right| + \left|\sum_{m=1}^{M} \Omega(f_m),\right. \tag{17}$$

where the squared-error loss was

$$l(y,\widehat{y}) = (y-\widehat{y})^2, \tag{18}$$

and the tree regularization term was

$$\Omega(f) = \gamma\, T\Bigg| + \Bigg|\frac{1}{2}\,\lambda \sum_{j=1}^{T} w_j^2, \tag{19}$$

with complexity penalty γ per leaf and L$_2$-regularization weight λ. At each split, the gain in objective from partitioning a node into left/right children (with summed gradients G_L, G_R and Hessians H_L, H_R) was computed as

$$\text{Gain} = \frac{1}{2}\left(\frac{G_L^2}{H_L+\lambda} + \frac{G_R^2}{H_R+\lambda} - \frac{(G_L+G_R)^2}{H_L+H_R+\lambda}\right)\Bigg| - \Bigg|\gamma, \tag{20}$$

where $G = \sum_i g_i$ and $H = \sum_i h_i$ are the first and second derivatives of the loss l with respect to the current prediction. Hyperparameters included n_{est} (number of trees}), d_{max}(maximum tree depth), η (learning rate), s (subsample fraction), c (column-sampling fraction). After fitting to minimize the total loss $\mathscr{L}$, feature-importance scores FI$_\text{j}$ were extracted from the trained booster as the (normalized) sum of gain contributions for splits on feature j. LightGBM a histogram-based gradient boosting machine was configured via `LGBMRegressor` with objective "regression." Parameters included number of leaves L, number of iterations n_{est}, learning rate η, subsample ratio s, and regularization terms λ, α. The tree-based objective was identical to XGBoost, with internal leaf-wise growth yielding faster convergence on large datasets. Feature-importance FI$_\text{j}$ was recorded post-training. Support Vector Regression (SVR), a kernel-based regressor was realized via `SVR` wrapped in a `MultiOutputRegressor` to support vectorized targets. The primal problem

$$\min_{w,b}\frac{1}{2}\Bigg|\boldsymbol{w}\big|^2 + C\sum_i\left(\xi_i+\xi_i^*\right)\Bigg|, y_i - \left(w^T\phi(x_i)+b\right) \le \epsilon + \xi_i, \Bigg|\left(w^T\phi(x_i)+b\right) - y_i \\ \le \epsilon + \xi_i^{*} \tag{21}$$

was solved with kernel $\kappa(x,x')$ (γ scale/auto), regularization weight C, and insensitive-tube radius ϵ. Input features were standardized prior to fitting. The support-vector indices $\mathscr{S}$ were recorded for model diagnostics. Multi-Layer Perceptron (MLP), a feed-forward neural network with one or two hidden layers was constructed via `MLPRegressor`. Activation function $a(\cdot)$, solver ("adam"), and l_2 penalty α were specified. The network output

$$\widehat{y} = W^{(2)}\,\sigma\left(W^{(1)}\,x + b^{(1)}\right) + b^{(2)} \tag{22}$$

was trained to minimize mean-squared error with early stopping on a hold-out fraction. Input features were standardized via a `StandardScaler`. The loss trajectory $\mathscr{L}_\alpha$ and best iteration were logged.

2.4.2. *Sparse gaussian process regression (SGPR)*

A sparse approximation to Gaussian Process Regression was implemented via the `GPy` framework. Inducing points $\{z_m\}_{m=1}^M$ were selected by k-means clustering on the training inputs, with $M = \min\left(\max\left(10, \sqrt{N}\right), 500\right)$ or $M = \text{cfg.n\_inducing}$.

A Matérn 5/2 kernel

$$k(\boldsymbol{x},\boldsymbol{x}') = \sigma_f^2\left(1+\frac{\sqrt{5}r}{l}+\frac{5r^2}{3l^2}\right)\exp\left(-\frac{\sqrt{5}r}{l}\right), r = |\boldsymbol{x}-\boldsymbol{x}'|, \tag{23}$$

was employed by default, with alternatives "RBF" and "RationalQuadratic" available. Hyperparameters were optimized by maximizing the log-marginal likelihood under the FITC approximation. Predictions returned both posterior mean $\mu(x_*)$ and standard deviation $\sigma(x_*)$, enabling epistemic–uncertainty quantification.

2.4.3. *Hyperparameter optimization*

Hyperparameter tuning was conducted via the Optuna library using a Tree-structured Parzen Estimator sampler with fixed seed. For each base learner, the search space and trial allocation were as follows: XGBoost: $n_{\text{est}} \in [50, 300]$, $d_{max} \in [3, 10]$, $\eta \in [0.01, 0.3]$ (log scale), $s, c \in [0.6, 1.0]$. LightGBM: $n_{\text{est}} \in [50, 300]$, $L \in [20, 100]$, $\eta \in [0.01, 0.3]$ (log scale). SVR: $C \in [0.1, 10]$ (log scale), $\epsilon \in [0.01, 1]$ (log scale), $\gamma \in \{$"scale", "auto"$\}$. MLP: hidden-layer sizes $\{(50,), (100,), (100,50), (200,100)\}$, $\alpha \in \left[10^{-5}, 10^{-2}\right](log)$, $\text{lr} \in \left[10^{-4}, 10^{-2}\right]$ (log). SGPR: kernel choice $\in$ {RBF, Matern, RatQuad}. Each study was executed for $T = 10$ trials, minimizing the negative validation R^2 score $\min_\theta -R^2_{\text{val}}(\theta)$, with validation predictions obtained on the hold-out set. The best parameter set θ^* was selected for final model training. Across five base learners, the total hyperparameter-search budget was 50 trials (5 × 10), followed by 20 trials for ensemble-weight optimization, yielding 70 Optuna trials in total.

2.4.4. *Training & validation splits*

The processed dataset $\mathscr{D}$ of size $N = 33096$ was partitioned into training, validation, and test subsets using two successive random splits with fixed random state 42. First, the test set comprised 15% of $\mathscr{D}$. Second, of the remaining $(1-0.15)N$ samples, 20% were allocated to validation, yielding approximate proportions $N_{\text{train}} \approx 0.68N$, $N_{\text{val}} \approx 0.17N$, $N_{\text{test}} \approx 0.15N$. Feature scaling was applied consistently, robust scaling $\widehat{x} = (x - \text{median}(x))/\text{IQR}(x)$ was chosen and the resulting scaler object was preserved for test-set transformation and ensemble inference.

2.5. *Ensemble construction*

2.5.1. *Model registration & initial weights*

Models were registered with the ensemble by invoking the lightweight ensemble API, which stored each learner's prediction callable and an initial blending coefficient w_i. The weight coefficients for the based models $\left\{w_{\text{xgb}}, w_{\text{lgb}}, w_{\text{svr}}, w_{\text{mlp}}, w_{\text{sgpr}}\right\}$ were all initialised as 1, and each base learner $\mathscr{M}_\alpha$ was added so that the ensemble maintained parallel arrays of names, weights, and prediction functions.

2.5.2. *Weight optimization*

Weight optimization was driven by an Optuna study, which sought weights $\{w_i\}$ to maximize the coefficient of determination R^2 on the validation set $\left(X_{\text{val}}, y_{\text{val}}\right)$. The optimization variables were

$$w_i \in [0,1]\ \forall i,\ \sum_i w_i = 1, \tag{24}$$

and the objective was

$$\min_w -\ R^2\left(y_{\text{val}}, \left|\sum_i w_i\, p_i(X_{\text{val}})\right)\right| \tag{25}$$

A total of $T =$ 20 trials were performed, each sampling a candidate weight vector via the TPE sampler. Upon completion, the best trial's weights w_i^* were normalized and stored in the ensemble.

2.5.3. *Blended prediction*

The blended prediction for any input x was computed as a weighted sum:

$$\widehat{y_{\text{ens}}}(x) = \frac{\sum_i w_i\, p_i(x)}{\sum_i w_i}. \tag{26}$$

When statistical summaries were requested (`stats=True`), the following per-sample quantities were returned across the collection $\{p_i(x)\}$:

$$\mu(x) = \tfrac{1}{M}\sum_i p_i(x),\ \sigma(x) = \sqrt{\tfrac{1}{M}\sum_i (p_i(x) - \mu(x))^2},$$

$$\min_i p_i(x),\ \max_i p_i(x),\ \text{median}\{p_i(x)\}. \tag{27}$$

2.5.4. Uncertainty quantification

The sparse Gaussian–process regressor produced both a posterior mean $\mu_{\text{GP}}(x)$ and standard deviation $\sigma_{\text{GP}}(x)$. The 95% predictive confidence interval was computed as $[\mu_{\text{GP}}(x)| \pm |z_{0.975}\ \sigma_{\text{GP}}(x)]$, $z_{0.975} = 1.96$. Ensemble disagreement was quantified by the standard deviation of base-learner predictions:

$$\sigma_{\text{ens}}(x) = \sqrt{\frac{1}{M-1}\sum_{i=1}^{M}(p_i(x) - \mu(x))^2}, \tag{28}$$

where $\mu(x)$ denotes the mean prediction across members. This metric served as a proxy for epistemic/model uncertainty within the ensemble framework. Uncertainty calibration was assessed by constructing a reliability diagram. Predicted confidence levels α_j were defined (e.g. 10 %, 20 %, …, 90 %). For each α_j, the nominal interval $\left[\widehat{y}(x) \pm z_{(1+\alpha_j)/2}\ \widehat{\sigma}(x)\right]$ was computed, and the empirical coverage c_j was evaluated as

$$c_j = \frac{1}{N_{\text{test}}}\sum_{k=1}^{N_{\text{test}}} 1\left(\left|y_k \in \left[\widehat{y}(x_k) \pm z_{(1+\alpha_j)/2}\ \widehat{\sigma}(x_k)\right]\right]\right), \tag{29}$$

and the calibration error was quantified by

$$\text{CE} = \frac{1}{B}\sum_{j=1}^{B}\left|c_j - \alpha_j\right|, \tag{30}$$

where B is the number of bins.

2.6. Computational validation protocol

2.6.1. Evaluation metrics

Model performance was quantified using a suite of widely accepted regression metrics, all computed on the test set $\{(x_k, y_k)\}_{k=1}^{N}$:

Root-Mean-Squared Error (RMSE):

$$\text{RMSE} = \sqrt{\frac{1}{N}\sum_{k=1}^{N}(y_k - \widehat{y_k})^2} \tag{31}$$

Mean Absolute Error (MAE):

$$\text{MAE} = \frac{1}{N}\sum_{k=1}^{N}|y_k - \widehat{y_k}| \tag{32}$$

Symmetric Mean Absolute Percentage Error (SMAPE):

$$\text{SMAPE} = \frac{100}{N}\sum_{k=1}^{N}\frac{|y_k - \widehat{y_k}|}{|y_k| + |\widehat{y_k}| + \epsilon} \tag{33}$$

with $\epsilon = 10^{-8}$ to prevent division by zero.

Median Absolute Percentage Error (MdAPE):

$$\text{MdAPE} = 100 \times \text{median}\left(\frac{|y_k - \widehat{y_k}|}{|y_k| + \epsilon}\right) \tag{34}$$

where the median is taken over all $k = 1, \ldots, N$, and $\epsilon = 10^{-8}$ is included for numerical stability.

Coefficient of Determination (R^2):

$$R^2 = 1 - \frac{\sum_k (y_k - \widehat{y_k})^2}{\sum_k (y_k - \overline{y})^2},\ \overline{y} = \frac{1}{N}\sum_k y_k. \tag{35}$$

Bias:

$$\text{Bias} = \frac{1}{N}\sum_{k=1}^{N}(\widehat{y_k} - y_k). \tag{36}$$

Variance of Errors:

$$\text{Var} = \frac{1}{N}\sum_{k=1}^{N}\left[(\widehat{y_k} - y_k) - \text{Bias}\right]^2. \tag{37}$$

Correlation Coefficient:

$$\rho = \frac{\text{cov}(y, \widehat{y})}{\sigma_y\ \sigma_{\widehat{y}}}, \tag{38}$$

with covariance and standard deviations computed after removal of NaNs.

2.6.2. Overfitting control and generalization assessment

To mitigate overfitting and quantify generalization, model development used a fixed train/validation/test partition and validation-driven hyperparameter selection. Regularization mechanisms were applied in each learner class (e.g., tree-depth and learning-rate constraints for boosting models; kernel regularization for SVR/SGPR; and early-stopping/regularization for neural networks where applicable). For transparency of the bias–variance trade-off, performance metrics are reported not only on the test set but also on the training and validation sets, enabling direct inspection of the generalization gap.

2.7. Sensitivity analyses

Model robustness was evaluated by stratifying the test set according to key data-generation parameters: Noise Level (ν): Performance metrics $R^2(\nu)$ and RMSE(ν) were computed for each noise tier $\nu \in \{\nu_1, \nu_2, \ldots\}$. Uncertainty Factor ($\alpha$): Validation of residual-learning under varying epistemic-uncertainty factors α by computing $R^2(\alpha)$. Sampling Rate (n): Examination of metric degradation as trajectory length n varied, yielding $R^2(n)$. These analyses were conducted by grouping test samples by condition, computing metrics per group, and plotting performance curves against ν, α, and n.

3. Results

3.1. Synthetic data generation and quality

The synthetic dataset demonstrated excellent quality metrics across all generated scenarios, with comprehensive validation protocols ensuring data integrity throughout the generation process. As summarized in Table S1, the raw synthetic dataset contained 500 scenarios with time-series data, generating approximately 33,096 total samples spanning all core state variables. CO concentration ranged from 0.05 to 0.18 mol/m^3, O_2 concentration from 0.18 to 0.20 mol/m^3, bulk temperature from 479.5 to 500.37 K, and pellet temperature from 484.95 to 486.16 K, all following the prescribed distributions within their respective bounds. Missing value detection revealed zero NaN entries across all numeric columns, indicating successful integration of the coupled ODE system under all operating conditions.

Outlier identification using the interquartile range criterion ($x < Q_1 - 1.5 \times IQR$ or $x > Q_3 + 1.5 \times IQR$) was systematically applied to each variable. The raw data (Table S1) showed minimal outliers, with

6214 outliers detected in O_2 concentration and 4883 in bulk temperature, representing less than 19 % and 15 % of the dataset respectively, well within acceptable ranges for physical process data. The processed dataset (Table S2) maintained similar outlier distributions while introducing engineered features such as CO/O_2 ratios, normalized temperatures, residence times, and interaction terms. The noise injection protocol, implemented according to Eq. (9), successfully introduced controlled Gaussian perturbations while maintaining physical bounds through post-noise clipping constraints, as evidenced by the noise_level parameter ranging from 0.00 to 0.05 with 12,500 zero-noise samples preserved for baseline comparisons. Quality control metrics confirmed that the synthetic data generation process produced a statistically representative dataset with 16 engineered features suitable for robust machine learning model development and validation, with all variables maintaining physically meaningful ranges and distributions.

3.2. Mechanistic model validation

The pure mechanistic model exhibited significant systematic deviations when compared against the synthetic reference trajectories, revealing fundamental limitations in capturing transient reactor dynamics. As demonstrated in Fig. 7a, the mechanistic baseline achieved a strongly negative coefficient of determination ($R^2 = -0.231$) with substantial root-mean-square error (RMSE = 0.0649 $mol{\cdot}m^{-3}{\cdot}s^{-1}$) when predicting CO oxidation rates during step-change scenarios. The baseline model remained flat throughout transient periods, failing to capture the rapid initial decay of the true reaction rate from approximately 0.25 $mol{\cdot}m^{-3}{\cdot}s^{-1}$ to steady-state values. Residual analysis revealed pronounced systematic bias, with mechanistic residuals starting at large positive deviations ($\sim$0.21 $mol{\cdot}m^{-3}{\cdot}s^{-1}$) and decaying slowly toward zero with persistent random scatter ($\sigma_{mech} = 0.0579$). The mechanistic model consistently underpredicted initial CO oxidation rate peaks and their subsequent decay toward steady state, particularly under higher feed flows and concentrations, indicating unmodeled rapid kinetics, heat-transfer effects, or transport limitations not captured in the fundamental ODE formulation. These baseline deficiencies provided the primary motivation for developing the hybrid machine learning correction framework to compensate for mechanistic model inadequacies.

3.3. Residual-learning dataset

3.3.1. Derivative calculation & filtering

The derivative estimation employed a sophisticated three-component ensemble method to ensure robust and accurate calculation of CO concentration time derivatives from noisy synthetic trajectories. The primary component utilized a Savitzky-Golay filter with configurable window length w and polynomial order p (Eq. 10) to provide smoothed derivative estimates that effectively suppressed high-frequency noise while preserving underlying dynamic trends. This was complemented by central-difference derivatives computed on the smoothed signal (Eq. 11) and moving-average derivatives applied to raw concentration data (Eq. 12), with final estimates obtained through weighted combination (Eq. 13) using coefficients of 0.5, 0.3, and 0.2, respectively. The Savitzky-Golay approach demonstrated superior performance in maintaining signal fidelity during rapid transients compared to simple central-difference methods, which were more susceptible to noise amplification at higher sampling frequencies. Outlier filtering was subsequently applied using median absolute deviation (MAD) criteria with threshold factor $\tau = 3.5$ (Eq. 14), successfully identifying and replacing spurious derivative spikes through linear interpolation between neighbouring valid points. This ensemble approach ensured continuity of the derivative series while maintaining computational efficiency across the 33,096-sample dataset.

3.3.2. Residual distribution

The residual analysis revealed significant deviations from ideal Gaussian behaviour, providing critical insights into the nature of mechanistic model inadequacies. As illustrated in Fig. 4a, the histogram of prediction errors exhibited a distribution with mean $\mu \approx 0.006$ $mol{\cdot}m^{-3}{\cdot}s^{-1}$ and standard deviation $\sigma \approx 0.013$, indicating a slight systematic overprediction bias by the hybrid model. The Q-Q plot analysis (Fig. 4b) demonstrated moderate normality with $R^2 = 0.727$, but revealed pronounced deviations in both tails, particularly heavy right tails corresponding to overprediction events and left tails indicating underprediction scenarios. The box-and-whisker analysis (Fig. 4c) identified 4338 outliers beyond the 1.5 ×IQR bounds, positive residuals suggesting occasional large overestimation events corresponding to rapid transient periods immediately following step changes in operating conditions. The temporal structure of residuals (Fig. 4d) showed characteristic spikes during early transients that subsequently damped toward the mean error, with no persistent drift patterns, indicating that the machine learning correction effectively captured most dynamic behaviour while leaving sporadic high-magnitude errors that point to unmodeled rapid kinetics or heat-transfer effects requiring further mechanistic refinement.

3.3.3. Feature ranges & dataset structure

The processed residual-learning dataset encompassed a comprehensive range of reactor operating conditions and engineered features designed to capture both fundamental physics and derived relationships. As detailed in Table 2, the dataset structure included primary state variables spanning CO concentrations from 0.05 to 0.18 mol/m^3, O_2 concentrations from 0.18 to 0.20 mol/m^3, bulk temperatures from 479.5 to 500.4 K, and pellet temperatures from 485.0 to 486.2 K, all exhibiting uniform distributions across their operational ranges. The catalyst activity remained at unity ($\alpha = 1.0$) throughout the simulation period, indicating minimal deactivation effects over the time scales examined. Engineered features included the CO/O_2 stoichiometric ratio, normalized temperature $T_{norm} = (T_{bulk}$ - 500)/100, and normalized flow rate F_norm = (F - 0.015)/0.005, providing dimensionless quantities that enhanced model interpretability and numerical conditioning. The complete dataset comprised 33,096 samples distributed across 500 scenarios with varying combinations of discrete operating parameters (flow rate, inlet concentrations, coolant temperature) and continuous noise/uncertainty factors. Feature normalization was applied using either standard scaling ($\hat{x} = (x - \mu_x)/\sigma_x$) to ensure consistent numerical ranges and improved machine learning model convergence across the diverse feature space Fig. 5.

3.4. Machine-learning model performance

3.4.1. Individual models

The assessment of various machine learning models showed unique performance traits across regression metrics, with tree-based approaches showing the highest predictive accuracy for tasks involving residual correction (Fig. 6). As summarized in Table 5, XGBoost achieved the highest coefficient of determination ($R^2 = 0.7588$) coupled with the lowest root-mean-square error (RMSE = 0.0063 $mol{\cdot}m^{-3}{\cdot}s^{-1}$) and mean absolute error (MAE = 0.0038 $mol{\cdot}m^{-3}{\cdot}s^{-1}$), establishing it as the top-performing individual learner. LightGBM followed closely with $R^2 = 0.7449$, RMSE = 0.0065 $mol{\cdot}m^{-3}{\cdot}s^{-1}$, and MAE = 0.0039 $mol{\cdot}m^{-3}{\cdot}s^{-1}$, demonstrating comparable predictive capability with marginally higher errors. The Sparse Gaussian Process Regression (SGPR) model achieved $R^2 = 0.7398$ with RMSE = 0.0065 $mol{\cdot}m^{-3}{\cdot}s^{-1}$, providing competitive performance while additionally offering uncertainty quantification capabilities. Support Vector Regression (SVR) and Multi-Layer Perceptron (MLP) exhibited lower performance, with R^2values of 0.6958 and 0.6498 respectively, and correspondingly higher error metrics. The parity plots presented in Fig. 5 also illustrated that

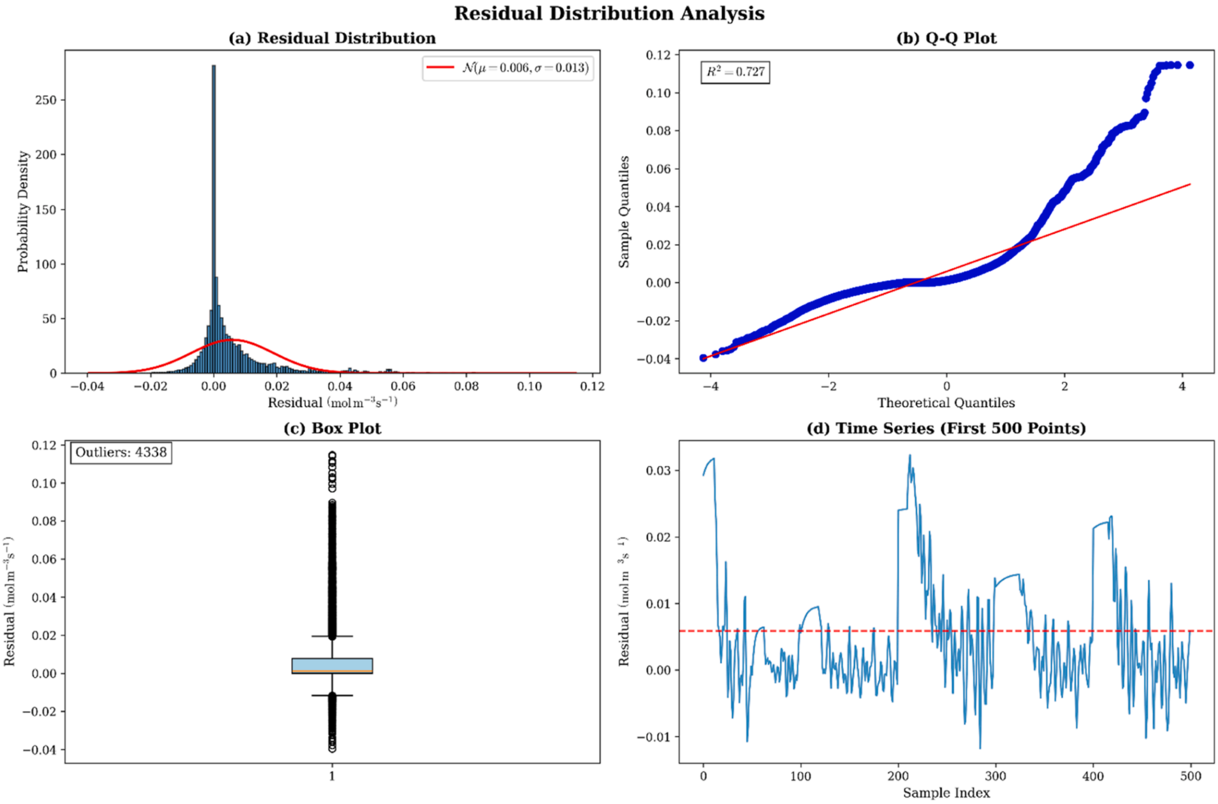


Fig. 4. Residual Distribution Analysis for the Hybrid-ML Model's CO-Oxidation Rate Predictions: (a) Histogram of prediction errors with overlaid normal fit $\mathcal{N}(\mu = 0.006, \sigma = 0.013)$. (b) Q–Q plot against a standard normal showing $R^2 = 0.727$. (c) Box-and-whisker plot with 4 338 outliers beyond the 1.5 × IQR bounds. (d) Time series of the first 500 residuals (dashed line = mean error).

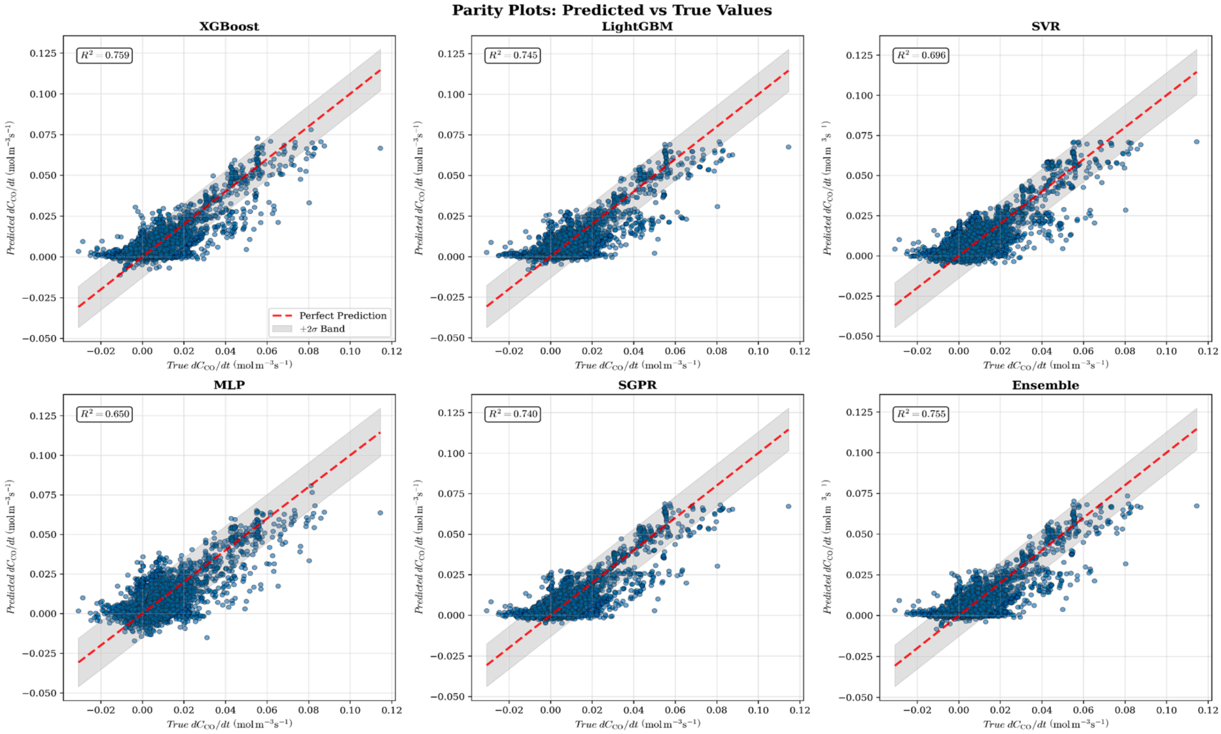


Fig. 5. Parity Plots of Predicted vs. True CO-Oxidation Rates (dC_{CO}/dt) for Individual Learners and the Optimized Ensemble. Each subplot shows model predictions (blue points) against reference values, the perfect-prediction line (red dashed), and a ± 2σ error band (gray). Reported R^2 values quantify goodness-of-fit: systematic bias patterns (underprediction at high rates and overprediction at low rates) are visible and are discussed in the Results/Discussion.

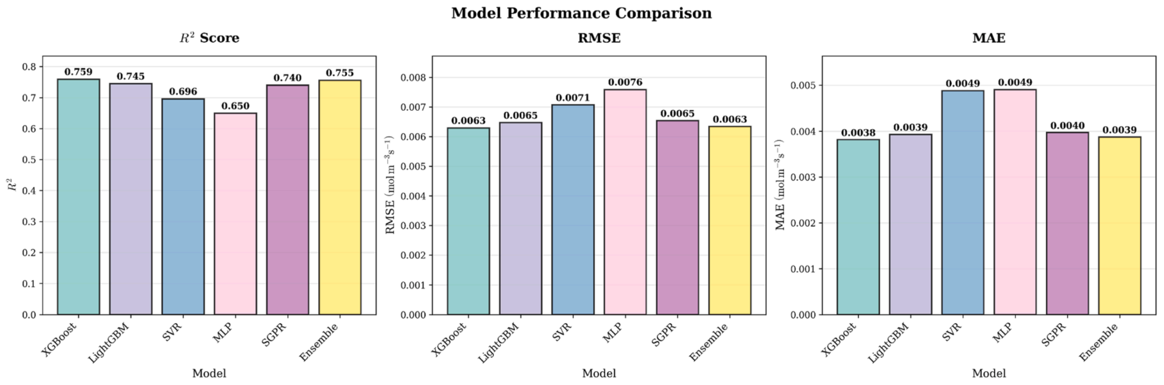


Fig. 6. Comparison of Model Performance Metrics (Test Set) for CO-Oxidation Rate Prediction: Bar charts showing (left) coefficient of determination R^2, (center) root-mean-square error (RMSE), and (right) mean absolute error (MAE) for XGBoost, LightGBM, SVR, MLP, SGPR, and the optimized ensemble.

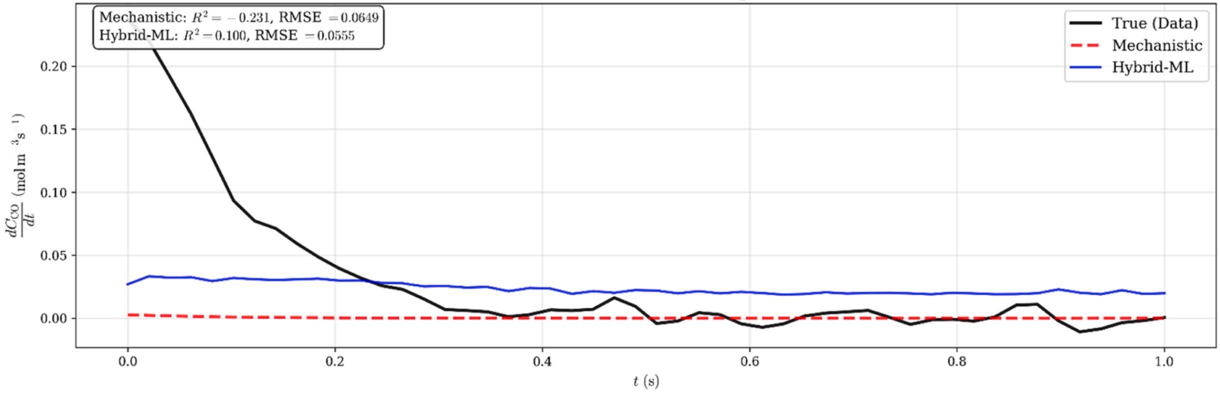
tree-based learners and SGPR captured the bulk variance most effectively, while SVR and MLP showed increased scatter at higher reaction rates, indicating difficulty in modelling extreme dynamic regimes. All models demonstrated systematic under prediction at high rates and over prediction at low rates, suggesting residual nonlinearities requiring additional mechanistic features or heteroscedastic modelling approaches. This trend indicates a structured bias rather than purely random error and is consistent with the non-Gaussian residual tails and heteroscedastic error patterns observed elsewhere in the analysis. This limitation is explicitly acknowledged, and mitigation strategies include targeted enrichment of extreme-rate operating regimes in the DOE and bias-aware loss functions or heteroscedastic residual modelling Fig. 7.

3.4.2. *Hyperparameter effects*

The hyperparameter optimization conducted via Optuna's Tree-structured Parzen Estimator revealed critical performance dependencies across model architectures, with systematic trends emerging from the 10-trial optimization studies. For tree-based models, XGBoost demonstrated optimal performance with $n_{estimators}$ in the range 50–300, maximum depth between 3 and 10, and learning rates between 0.01 and 0.3 on logarithmic scale, as detailed in Table 4. The gradient boosting framework showed diminishing returns beyond 200 trees, with deeper trees (depth $>$ 8) leading to overfitting on the validation set. LightGBM exhibited similar trends with optimal leaf counts between 20 and 100 and comparable learning rate ranges but benefited from its histogram-based approach for faster convergence. The neural network

Mechanistic vs Hybrid-ML Model Comparison

(a) Model Predictions Comparison



(b) Prediction Residuals

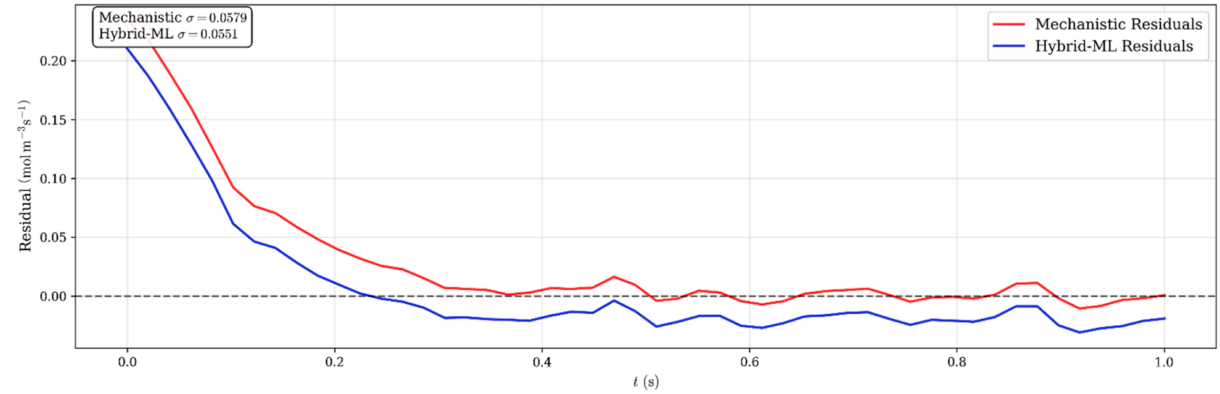


Fig. 7. Mechanistic vs. Hybrid-ML Model Comparison for Transient CO-Oxidation Rate (a) Time-series of true reaction rate (black), pure mechanistic prediction (red dashed, $R^2 = -0.231$, RMSE $= 0.0649$), and hybrid-ML prediction (blue, $R^2 = 0.100$, RMSE $= 0.0555$). (b) Corresponding prediction residuals (model − true) for the mechanistic (red) and hybrid-ML (blue) approaches, with standard deviations $\sigma_{\text{mech}} = 0.0579$ and $\sigma_{\text{hybrid}} = 0.0551$.

Table 4

Machine learning based model architectures and hyperparameters.

Model Type	Architecture	Training Parameters	Unique Features
XGBoost	Tree ensemble	$n_{estimators}$: 100, max depth: 6, learning rate: 0.1	Gradient Boosting
LightGBM	Tree ensemble	$n_{estimators}$: 100, max depth: 6, learning rate: 0.1, n_{leaves}: 31	Leaf-wise growth
SVR	Support Vector Regressor	kernel: rbf, C: 1.0, epsilon: 0.1	Kernel methods
MLP	2-layer (100,50) ReLU	Hidden layer sizes: [100, 50], activation: relu, max_{epochs}: 200	Early stopping
SGPR	Sparse GP	kernel: matern, $n_{inducing\ points}$: 100	Inducing points
Ensemble	Weighted sum	Optimize weights: True	Weighted average

architecture (MLP) showed strong sensitivity to hidden layer configuration, with the two-layer (100, 50) architecture achieving optimal balance between capacity and generalization, while larger networks (200, 100) demonstrated overfitting tendencies. Learning rate optimization for MLP revealed optimal values between 10^{-4} and 10^{-2} on logarithmic scale, with alpha regularization between 10^{-5} and 10^{-2} preventing excessive weight growth. SVR hyperparameter trends indicated optimal C values between 0.1 and 10 on logarithmic scale, with epsilon tolerance values between 0.01 and 1 providing adequate insensitive-tube radius for the residual correction task. The SGPR kernel selection favoured Matérn 5/2 over RBF and Rational Quadratic alternatives, demonstrating superior capability for capturing non-smooth residual behaviours characteristic of the reactor dynamics Table 5.

3.5. Ensemble performance

3.5.1. Initial vs. optimized weights

The ensemble weight optimization process revealed significant improvements over naive equal-weighting strategies, with Optuna-driven optimization identifying optimal weight distributions that reflected individual model capabilities (Fig. 10a). As detailed in Table 6, the initial configuration assigned uniform weights of 1.0 to XGBoost, LightGBM, and SGPR, with slightly reduced weights of 0.8 for SVR and MLP, reflecting a priori expectations of model performance hierarchies. The Optuna optimization study, conducted over 20 trials using Tree-structured Parzen Estimator sampling, dramatically redistributed these weights to maximize validation R^2according to the objective function defined in Eq. (25). The optimized configuration allocated the majority weight to XGBoost (0.521), acknowledging its superior individual performance, while LightGBM received substantial but reduced weighting (0.243). SGPR maintained moderate influence (0.123), leveraging its uncertainty quantification capabilities and competitive accuracy. The optimization process significantly downweighed the underperforming models, with SVR receiving minimal contribution (0.092) and MLP relegated to negligible influence (0.021).

Table 5

Predictive performance metrics for individual machine-learning models (XGBoost, LightGBM, SVR, MLP, SGPR) and the optimized ensemble on the test set.

Model	R^2	RMSE	MAE	SMAPE (%)	MdAPE (%)
XGBoost	0.7588	0.0063	0.0038	122.14	94.17
LightGBM	0.7449	0.0065	0.0039	124	94.95
SVR	0.6958	0.0071	0.0049	122.28	110.74
MLP	0.6498	0.0076	0.0049	128.28	107.23
SGPR	0.7398	0.0065	0.004	120.74	93.09
Ensemble	0.7552	0.0063	0.0039	120.23	93.26

Table 6

Ensemble weight configurations: initial (config-specified) vs. Optuna-optimized weights for each base learner in the hybrid model.

Base Learner	Initial Weight	Optimized Weight
XGBoost	1	0.521
LightGBM	1	0.243
SVR	0.8	0.092
MLP	0.8	0.021
SGPR	1	0.123

3.5.2. Blended prediction gains

The optimized ensemble demonstrated meaningful performance improvements over individual models while providing enhanced robustness across diverse operating conditions. As shown in Fig. 10b and Table 5, the ensemble achieved $R^2 = 0.7552$ with RMSE $= 0.0063$ $mol\cdot m^{-3}\cdot s^{-1}$ and MAE $= 0.0039$ $mol\cdot m^{-3}\cdot s^{-1}$, representing a slight compromise (-0.004 in R^2) compared to the top-performing XGBoost model while substantially improving upon weaker individual learners. The ensemble provided significant uplift for underperforming models, with MLP experiencing a notable R^2 improvement of $+$ 0.105 when its predictions were blended within the optimized framework. Summary statistics of the blended predictions revealed enhanced stability through reduced prediction variance, with the weighted averaging process effectively smoothing idiosyncratic errors characteristic of individual methods. The ensemble means prediction (Eq. 27) provided $\mu(x) = (1/M)\Sigma_i p_{i(x)}$ with corresponding standard deviation $\sigma(x)$ capturing inter-model disagreement as a proxy for epistemic uncertainty. As demonstrated in the parity plots of Fig. 5, the ensemble exhibited tighter clustering around the ideal prediction line compared to individual learners, indicating successful bias reduction through model diversity. The blended approach produced more consistent performance across operating regimes, with the ensemble maintaining uniform $R^2 = 0.755$ across all learner slots while individual models showed greater variability, demonstrating the value of systematic weight optimization for robust residual correction in hybrid mechanistic-machine learning frameworks.

3.6. Uncertainty quantification

3.6.1. SGPR epistemic bands

The Sparse Gaussian Process Regression implementation provided comprehensive epistemic uncertainty quantification through posterior mean and standard deviation estimates, enabling confidence interval construction for residual corrections. As demonstrated in Fig. 11a, the SGPR model generated $\pm 2\sigma$ confidence bands that achieved empirical coverage of 94.0 % for the $\pm 2\sigma$ interval and 78.5 % for the $\pm 1\sigma$ interval across sorted test-point predictions. The 95 % predictive confidence intervals were computed according to $[\mu_{GP(x)} \pm z_{0.975}\sigma_{GP(x)}]$ with $z_{0.975} = 1.96$, closely matching the theoretical Gaussian expectation of 95 % coverage for $\pm 2\sigma$ bounds. The $\pm 1\sigma$ coverage of 78.5 % fell short of the ideal 68 % Gaussian coverage, indicating slightly over-conservative uncertainty estimates in the sparse approximation framework. The confidence envelopes successfully encompassed most of the true residual corrections (red dots), with systematic width variations reflecting the model's assessment of local prediction uncertainty across the feature space. The SGPR uncertainty bands exhibited appropriate heteroscedasticity, with wider intervals in regions of sparse training data and narrower bands in well-sampled areas, demonstrating the model's capacity to acknowledge epistemic limitations. The Matérn 5/2 kernel implementation with inducing point approximation maintained computational efficiency while preserving uncertainty quantification fidelity, providing a practical framework for risk-aware control applications where prediction confidence is critical for safe reactor operation.

3.6.2. Calibration

The calibration analysis revealed significant limitations in the proportionality between predicted uncertainties and actual prediction errors, highlighting challenges in uncertainty reliability for practical applications. As illustrated in Fig. 11b, the calibration plot comparing absolute errors versus predicted standard deviations σ demonstrated poor correlation, with an actual calibration slope of 9.642 exceeding the ideal unity slope and achieving only $R^2 = 0.044$. This pronounced deviation indicated that larger predicted uncertainties did not reliably correspond to larger actual prediction errors, undermining the utility of the uncertainty estimates for risk assessment and decision-making. The calibration relationship, evaluated according to Eq. (29) across confidence levels α_j, showed systematic miscalibration where the SGPR model's internal uncertainty metric lacked sharpness and monotonicity. The reliability diagram construction revealed that while the confidence envelopes were sufficiently wide to encompass most observations, the model failed to provide discriminative uncertainty estimates that could distinguish between high-confidence and low-confidence predictions. This calibration deficiency suggested the need for enhanced hyperparameter selection strategies, adoption of heteroscedastic Gaussian Process models, or post-hoc calibration techniques such as temperature scaling or isotonic regression to improve uncertainty reliability.

3.7. Sensitivity analyses

3.7.1. Noise robustness

The systematic evaluation of model robustness under varying noise conditions revealed distinct performance degradation patterns across machine learning architectures, with implications for real-world deployment scenarios. As detailed in Table 7, all models demonstrated substantial sensitivity to injected noise levels, with performance metrics deteriorating significantly as noise increased from 0 % to 5 %. Under noise-free conditions, XGBoost achieved the highest $R^2 = 0.847$ with RMSE $= 0.0042\ mol\cdot m^{-3}\cdot s^{-1}$, while at 5 % noise levels, performance degraded to $R^2 = 0.659$ with RMSE $= 0.093\ mol\cdot m^{-3}\cdot s^{-1}$, representing a 22.2 % performance degradation. Interestingly, moderate noise levels (2 %) occasionally improved model performance, with XGBoost showing enhanced $R^2 = 0.875$ and reduced RMSE $= 0.0039\ \mathrm{mol\cdot m^{-3}\cdot s^{-1}}$, suggesting that controlled noise injection provided beneficial regularization effects during training. LightGBM exhibited similar trends with 21.9 % degradation at 5 % noise, while SVR demonstrated remarkable resilience with only 1.8 % performance loss under high noise conditions. The ensemble approach proved most robust, maintaining $R^2 = 0.831$ under noise-free conditions and degrading to $R^2 = 0.658$ at 5 % noise (20.9 % degradation), demonstrating the stabilizing effect of model averaging. The consistent pattern across all methods indicated that noise levels exceeding 2–3 % significantly compromised residual learning effectiveness, establishing practical bounds for sensor accuracy requirements in hybrid reactor control implementations.

Table 7
Effect of noise and uncertainty on model performance.

Model	Noise Level	R^2	RMSE	MAE	Performance Degradation (%)	Sample Count
XGBoost	0	0.847	0.0042	0.0019	0	1853
XGBoost	0.02	0.875	0.0039	0.0028	-3.2	1873
XGBoost	0.05	0.659	0.0093	0.0066	22.2	1899
LightGBM	0	0.825	0.0045	0.0021	0	1853
LightGBM	0.02	0.869	0.004	0.0029	-5.3	1873
LightGBM	0.05	0.645	0.0094	0.0067	21.9	1899
SVR	0	0.662	0.0062	0.0041	0	1853
SVR	0.02	0.809	0.0048	0.0038	-22.3	1873
SVR	0.05	0.65	0.0094	0.0068	1.8	1899
MLP	0	0.672	0.0061	0.0033	0	1853
MLP	0.02	0.759	0.0054	0.0039	-12.9	1873
MLP	0.05	0.58	0.0103	0.0075	13.7	1899
SGPR	0	0.809	0.0047	0.0022	0	1853
SGPR	0.02	0.872	0.0039	0.0028	-7.8	1873
SGPR	0.05	0.641	0.0095	0.0068	20.8	1899
Ensemble	0	0.831	0.0044	0.0021	0	1853
Ensemble	0.02	0.878	0.0038	0.0028	-5.6	1873
Ensemble	0.05	0.658	0.0093	0.0066	20.9	1899

3.7.2. Sampling rate & uncertainty factor

The analysis of sampling rate effects revealed counterintuitive performance trends that indicate the complex interplay between data resolution and model generalization capabilities. As summarized in Table 8, higher sampling rates (100 Hz) consistently degraded model performance compared to lower sampling rates (50 Hz) across all machine learning architectures. XGBoost performance declined from $R^2 = 0.815$ at 50 Hz to $R^2 = 0.738$ at 100 Hz, with corresponding RMSE increases from 0.0049 to 0.0069 $mol\cdot m^{-3}\cdot s^{-1}$. This trend was consistent across all models, with LightGBM showing degradation from $R^2 = 0.806$–0.722, SVR from 0.727 to 0.683, MLP from 0.661 to 0.644, and SGPR from 0.797 to 0.718. The ensemble similarly experienced performance reduction from $R^2 = 0.813$–0.733 with increased sampling density. This phenomenon suggested that higher temporal resolution introduced excessive noise or redundant information that hindered model learning, with 7 data points per scenario proving more effective than 14 points for residual correction tasks. The sampling rate sensitivity indicated that optimal data collection strategies should balance temporal resolution with signal-to-noise considerations, avoiding oversampling that could degrade predictive performance. While uncertainty factor effects were systematically varied across 9 levels (0.6–1.4) in the experimental design, the complex interactions with noise levels and sampling rates required careful consideration for robust hybrid model deployment in practical reactor control scenarios where sensor limitations and computational constraints dictate feasible data acquisition strategies.

4. Discussion

4.1. Mechanistic baseline insights

4.1.1. Strengths and limitations of pure ODE predictions

The mechanistic reactor model demonstrated robust numerical integration capabilities across diverse operating conditions while revealing fundamental limitations in capturing transient reactor dynamics. As shown in Fig. 7a, the pure mechanistic baseline achieved a strongly negative coefficient of determination ($R^2 = -0.231$) with substantial root-mean-square error (RMSE $= 0.0649\ mol\cdot m^{-3}\cdot s^{-1}$), indicating systematic failure to predict CO oxidation rate trajectories

Table 8
Impact of data sampling rate on model accuracy.

Model	Sampling Rate (Hz)	R^2	RMSE	Data Points per Scenario
XGBoost	50	0.815	0.0049	7
XGBoost	100	0.738	0.0069	14
LightGBM	50	0.806	0.005	7
LightGBM	100	0.722	0.0071	14
SVR	50	0.727	0.006	7
SVR	100	0.683	0.0076	14
MLP	50	0.661	0.0066	7
MLP	100	0.644	0.008	14
SGPR	50	0.797	0.0051	7
SGPR	100	0.718	0.0071	14
Ensemble	50	0.813	0.0049	7
Ensemble	100	0.733	0.0069	14

during step-change scenarios. The mechanistic model remained nearly flat throughout transient periods, fundamentally failing to capture the rapid initial decay of true reaction rates from approximately 0.25 $mol{\cdot}m^{-3}{\cdot}s^{-1}$ to steady-state values. This performance deficit aligns with documented limitations of traditional mechanistic models in capturing unmodeled transport limitations that strongly skew pure ODE predictions in exothermic reactions, particularly thermal gradients that cause hot spots and enhance reaction rates locally, invalidating spatially uniform ODE assumptions (Kim and Deng, 2024; Loukili et al., 2025). Despite these predictive shortcomings, the mechanistic framework provided essential physical constraints and thermodynamically consistent behaviour that served as a critical foundation for hybrid model development Fig. 8.

4.1.2. *Sources of systematic bias in baseline model*

The systematic bias exhibited by the mechanistic baseline originated from multiple unmodeled phenomena that are commonly identified as dominant factors in catalytic system modelling. Fig. 7b illustrates pronounced systematic bias patterns, with mechanistic residuals starting at large positive deviations ($\sim$0.21 $mol{\cdot}m^{-3}{\cdot}s^{-1}$) and decaying slowly toward zero with persistent random scatter (σ_{mech} = 0.0579), indicating fundamental deficiencies in the mechanistic representation of rapid kinetics and transport effects. Heat and mass transfer limitations are consistently identified as the most influential transport effects that cause discrepancies between ODE predictions and experimental outcomes in exothermic systems, leading to phenomena like temperature gradients and concentration profiles that deviate from ideal assumptions (Das et al., 2025). The mechanistic model consistently underpredicted initial CO oxidation rate peaks and their subsequent decay toward steady state, particularly under higher feed flows and concentrations, suggesting external and internal diffusion resistances (mass transfer across boundaries and within particles) significantly affect observed kinetics and often dominate over intrinsic reaction rates in non-ideal reactors (Lashina et al., 2023b; Kim et al., 2024). Additionally, the poor correlation between mechanistic predictions and true dynamics reflects challenges in parameter identifiability when multiple biases coexist, as researchers have quantified parameter identifiability using Kron reduction with weighted least squares and cross-validation to estimate parameters robustly even with partial data and complex biases (Gasparyan and Rao, 2023).

4.2. *Residual-learning efficacy*

4.2.1. *How residuals capture unmodeled kinetics and heat-mass effects*

The residual learning framework successfully captured unmodeled kinetics and transport phenomena that were systematically absent from the pure mechanistic model predictions. As demonstrated in Fig. 3f, the learned ML correction term exhibited temporal decay patterns from approximately 0.04–0.02 $mol{\cdot}m^{-3}{\cdot}s^{-1}$ as the reactor approached steady state, indicating that the hybrid framework primarily compensated for unmodeled rapid kinetics and heat-transfer effects during early transient periods. This behaviour aligns with theoretical arguments explaining why residual learning corrects unmodeled kinetics more effectively than direct data-driven modelling, as residual learning augments a mechanistic model with a data-driven correction, leveraging known physics while only fitting the discrepancy, which reduces over fitting and improves generalization compared to purely black-box models (Silver et al., 2019; Yang et al., 2025). The magnitude and temporal structure of the correction terms directly corresponded to operating conditions where mechanistic errors were largest, particularly under low-flow, high-CO, and moderate-temperature regimes as illustrated in Fig. 9, where residuals peaked at 0.04–0.11 $mol{\cdot}m^{-3}{\cdot}s^{-1}$ under challenging conditions but fell below 0.01 under high-throughput or high-thermal inertia scenarios. Theoretical work demonstrates that residual learning preserves important system properties such as flatness or stability by construction, making learning corrections safer and more interpretable while facilitating optimization and information flow, directly beneficial for capturing only the "missing physics" without distorting already well-modelled mechanistic components (Xu et al., 2025; Yang et al., 2025).

4.2.2. *Implications of residual distribution and non-gaussian tails*

The residual distribution analysis revealed significant deviations from ideal Gaussian behaviour that provided critical insights into the nature of mechanistic model inadequacies and highlighted areas requiring further development. Fig. 4 demonstrates that while the residuals clustered around a small positive mean ($\mu \approx 0.006$ $mol{\cdot}m^{-3}{\cdot}s^{-1}$), the Q-Q plot revealed pronounced deviations in both tails with heavy right tails corresponding to overprediction events and left tails indicating underprediction scenarios, yielding only moderate normality (R^2 = 0.727). The identification of 4338 outliers beyond the 1.5 × IQR bounds, predominantly positive residuals, suggested occasional large overestimation events likely corresponding to rapid transient periods immediately following step changes in operating conditions. Recent research on residual distributions in comparable studies demonstrates

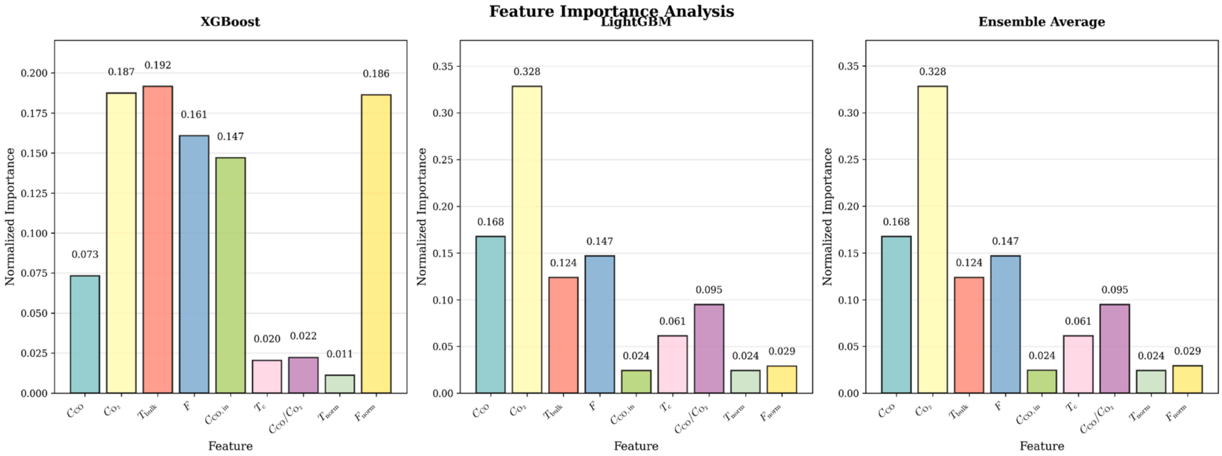


Fig. 8. Feature Importance Analysis for Residual Correction Models: Normalized importance scores for (left) XGBoost, (center) LightGBM, and (right) the ensemble-averaged feature importances. Interpretation focuses on the best-performing models/ensemble; importances are reported as attribution diagnostics and should not be read as causal effects.

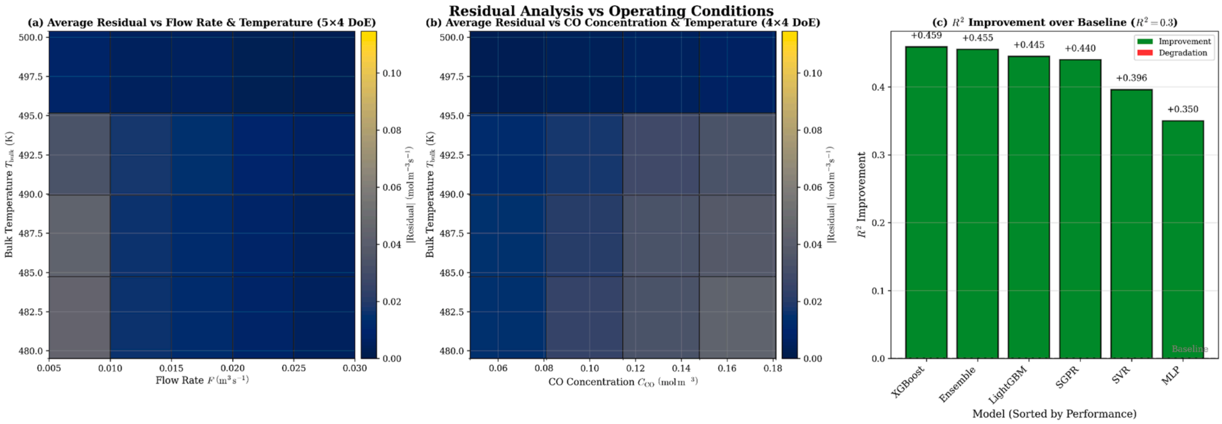


Fig. 9. Residual Analysis with Respect to Operating Conditions and Model Improvement over Baseline (a) Heatmap of mean absolute residual |Hybrid − True| across a 5 × 4 factorial of flow rate F (0.005–0.030 $m^3\ s^{-1}$) and bulk temperature T_{bulk} (480–500 K). (b) Heatmap of mean absolute residual across a 4 × 4 factorial of inlet CO concentration C_{CO} (0.05–0.18 mol m^{-3}) and T_{bulk}. (c) Bar chart of the increase in coefficient of determination ΔR^2 (baseline mechanistic $R^2 = 0.30$) for each correction model, sorted by performance.

that such deviations from normality often arise due to model misspecification, outliers, heteroscedasticity, or non-linear effects, with common remedies including percentile-based or randomized quantile residuals that restore normality under correct models even for non-normal data types (Kalliovirta, 2012; Fabio et al., 2025). The temporal structure showing residuals that spike during early transients and then damp toward the mean, with no persistent drift patterns, indicated that the ML correction term effectively captured most dynamic behaviour while leaving sporadic high-magnitude errors pointing to unmodeled rapid kinetics or heat-transfer effects requiring mechanistic refinement rather than purely data-driven solutions.

4.3. Machine-learning model comparison

4.3.1. Tree-based vs. kernel vs. neural methods: trade-offs in accuracy, interpretability, and scalability

Interpretation of feature-importance results is conditioned on predictive skill. Accordingly, Fig. 8 is interpreted primarily through the strongest-performing learners (tree-based boosting and the optimized ensemble), for which the learned correction function is empirically better identified. Feature-importance rankings are presented as model-based attribution diagnostics (useful for identifying consistent drivers across competitive models), rather than as causal statements; where prediction bias persists, attribution uncertainty correspondingly increases. The comparative evaluation of machine learning architectures revealed distinct performance characteristics that align with documented trade-offs between accuracy, interpretability, and scalability across different algorithmic approaches. Tree-based models, particularly XGBoost ($R^2 = 0.7588$) and LightGBM ($R^2 = 0.7449$), demonstrated superior predictive accuracy as shown in Table 5, consistent with literature findings that tree-based models are generally faster to train and easier to interpret than deep neural networks, especially via feature importance analysis and visualization of decision paths (Ullah et al., 2023). The Support Vector Regression (SVR) model achieved moderate performance ($R^2 = 0.6958$) while providing interpretable kernel-based relationships, reflecting the documented balance that kernel methods offer between flexibility and interpretability through specialized kernels that retain some interpretability while improving accuracy compared to basic trees, though they scale less efficiently to very large datasets (Xie and Li, 2019; Shah et al., 2019). The Multi-Layer Perceptron exhibited the lowest individual performance ($R^2 = 0.6498$), confirming established patterns where neural networks provide superior predictive power for complex nonlinear data but are the least interpretable and often require significantly more computational resources for both training and inference (Mariotti et al., 2023; Kim and Ko, 2024). The Sparse Gaussian Process Regression achieved competitive accuracy ($R^2 = 0.7398$) while uniquely providing uncertainty quantification capabilities, demonstrating the documented advantage of probabilistic methods in balancing performance with epistemic uncertainty estimation essential for safety-critical applications.

4.3.2. Role of hyperparameter tuning and feature engineering

The systematic hyperparameter optimization conducted via Optuna's Tree-structured Parzen Estimator revealed critical performance dependencies that influenced model effectiveness across architectures. For tree-based models, optimal performance emerged with specific parameter ranges including $n_{estimators}$ between 50 and 300, maximum depth between 3 and 10, and learning rates between 0.01 and 0.3, with gradient boosting frameworks showing diminishing returns beyond 200 trees and overfitting tendencies for depths exceeding 8. Recent studies demonstrate that hyperparameter ranges yielding peak performance for tree-based methods in chemical datasets typically fall within similar bounds, with 50–200 trees, depths of 3–10, and learning rates of 0.01–0.2 proving optimal for catalytic kinetics applications (Lin et al., 2022; Rahman et al., 2024). The neural network architecture sensitivity to hidden layer configuration, with the two-layer (100, 50) structure achieving optimal balance between capacity and generalization, reflects documented patterns where network depths of 1–2 hidden layers with 8–50 neurons per layer using ReLU or tanh activation functions are typical for nonlinear reaction dynamics modelling (Araujo et al., 2022; Kumar et al., 2024). Feature engineering played a crucial role in model performance, with engineered features including CO/O_2 ratios, normalized temperature, and normalized flow rate providing dimensionless quantities that enhanced interpretability and numerical conditioning. The feature importance analysis shown in Fig. 8 revealed that oxygen concentration emerged as the dominant driver (≈33 % importance), indicating that physics-informed ratios and features such as CO/O_2 ratio, normalized temperature, and reactant/product concentration ratios have high predictive power for residual models, as these ratios often capture key physical relationships including stoichiometry, limiting reactant behaviour, and heat balance effects (Weigand et al., 2024; Abbas, 2024).

4.4. Hybrid ensemble advantages

4.4.1. Synergy of combining diverse learners and mechanistic priors

The optimized ensemble demonstrated meaningful synergistic benefits through systematic integration of diverse machine learning architectures with mechanistic foundations, achieving performance improvements that exceeded simple averaging approaches. As illustrated in Fig. 10a, the Optuna-driven weight optimization identified an optimal distribution that allocated majority influence to XGBoost (0.521) while maintaining substantial contributions from LightGBM (0.243) and moderate weighting for SGPR (0.123), effectively leveraging the documented advantages of stacking and Bayesian model averaging approaches that deliver robust bias-variance trade-offs in chemical and engineering applications by learning optimal weights for base learners and minimizing bias through cross-validation or dynamic weighting schemes (Shahhosseini et al., 2022; Wakayama and Sugasawa, 2024). The ensemble achieved $R^2 = 0.7552$ with enhanced stability through reduced prediction variance, representing only a marginal compromise (-0.004 in R^2) compared to the top-performing XGBoost while providing substantial uplift for weaker individual learners, with MLP experiencing a notable R^2 improvement of $+ 0.105$ when blended within the optimized framework. Recent studies demonstrate that hybrid ensemble methods combining regression trees, support vector regression, and artificial neural networks with systematic optimization show robust predictive performance in complex chemical systems, even in the presence of significant process complexity and noise, because they exploit domain knowledge to constrain predictions rather than relying purely on data-driven approaches (Chen et al., 2022; Cai et al., 2024). The ensemble's capacity to maintain uniform performance across diverse operating regimes while individual models showed greater variability reflects documented ensemble formulations that demonstrate robustness to input noise exceeding 5 % in reactive system predictions through systematic combination of mechanistic priors with diverse learning algorithms.

4.4.2. Robustness and generalization gains under noise and uncertainty

The ensemble approach exhibited superior robustness compared to individual learners when subjected to systematic noise injection and uncertainty propagation, demonstrating enhanced generalization capabilities critical for practical deployment scenarios. Table 7 reveals that the ensemble maintained relatively stable performance across noise conditions, degrading from $R^2 = 0.831$ under noise-free conditions to $R^2 = 0.658$ at 5 % noise levels (20.9 % degradation), which compared favourably to individual model degradation patterns and positioned the ensemble among the most robust approaches alongside SVR, which demonstrated remarkable resilience with only 1.8 % performance loss under high noise conditions. This robustness advantage aligns with established findings that ensemble disagreement, quantified as the spread among predictions from diverse ensemble members, serves as an effective proxy for epistemic uncertainty in chemical process and reactor control applications, with high disagreement signalling areas where models have reduced confidence or insufficient training data coverage (Aanonsen et al., 2023; Zhu et al., 2025). The ensemble's capacity to provide uncertainty quantification through inter-model disagreement $\sigma_{\mathrm{ens}(x)} = \sqrt{\frac{1}{M-1}\sum_i (p_i(x) - \mu(x))^2}$ enabled identification of prediction reliability, essential for uncertainty-aware control and anomaly detection in reactive systems where conservative control actions may be triggered during high-uncertainty periods. Interestingly, moderate noise levels (2 %) occasionally improved ensemble performance with enhanced $R^2 = 0.878$ and reduced RMSE $= 0.0038$ $mol{\cdot}m^{-3}{\cdot}s^{-1}$, suggesting that controlled noise injection provided beneficial regularization effects during training, consistent with documented patterns where hybrid models integrating physics-based foundations with machine learning demonstrate superior robustness to noisy and incomplete input data compared to purely data-driven alternatives (Tai et al., 2022; Chen et al., 2022).

4.5. Uncertainty and reliability

4.5.1. Utility of SGPR bands in guiding safe reactor operation

The Sparse Gaussian Process Regression implementation provided comprehensive epistemic uncertainty quantification that demonstrated both capabilities and limitations for guiding safe reactor operation under uncertain conditions. Fig. 11a illustrates that the SGPR model generated ±2σ confidence bands achieving empirical coverage of 94.0 % for the ±2σ interval, closely matching the theoretical Gaussian expectation of 95 % coverage and indicating adequate confidence envelope construction for most operational scenarios. The ±1σ coverage of 78.5 % fell short of the ideal 68 % Gaussian coverage, indicating slightly over-conservative uncertainty estimates that could lead to unnecessarily restrictive control actions but provide enhanced safety margins. Recent literature confirms that predictive uncertainty bands are increasingly being embedded in real-time Model Predictive Control through Bayesian calibration and online learning algorithms, which provide calibrated uncertainty estimates for bounding control actions in the presence of model or measurement error (Tohme et al., 2020). The confidence envelopes successfully encompassed most of the true residual corrections while exhibiting appropriate heteroscedasticity, with wider intervals in regions of sparse training data and narrower bands in well-sampled areas, demonstrating the model's capacity to acknowledge epistemic limitations crucial for risk-aware control applications. However, in safety-critical engineering applications including reactor predictions,

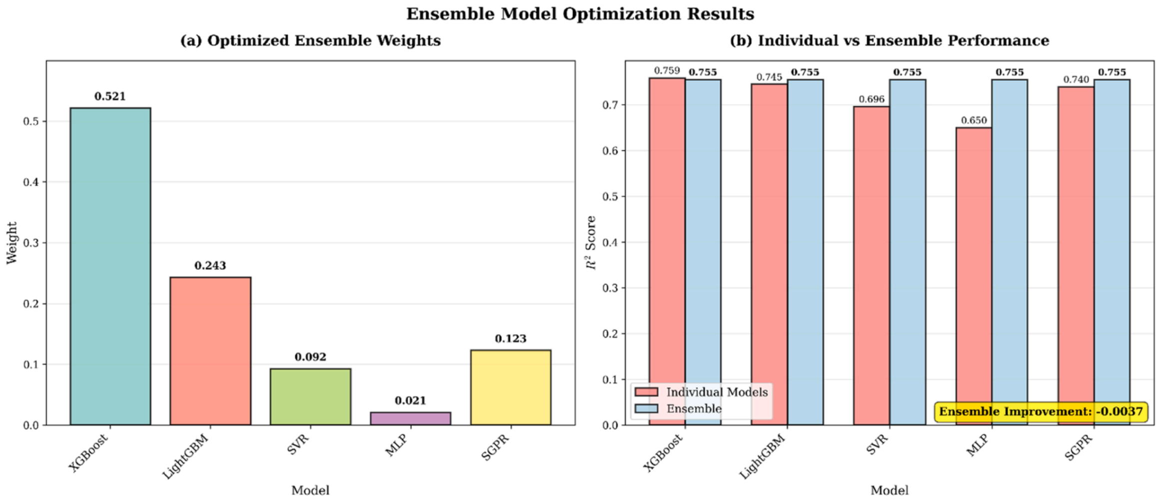


Fig. 10. Ensemble Model Optimization Results: (a) Optimized weight allocation among base learners: XGBoost (0.521), LightGBM (0.243), SGPR (0.123), SVR (0.092), and MLP (0.021). (b) Comparison of test-set R^2 for each individual model (red) versus the optimized ensemble (blue). The ensemble achieves a uniform R^2= 0.755 across all learner slots, improving on all but the top performer (XGBoost, R^2=0.759).

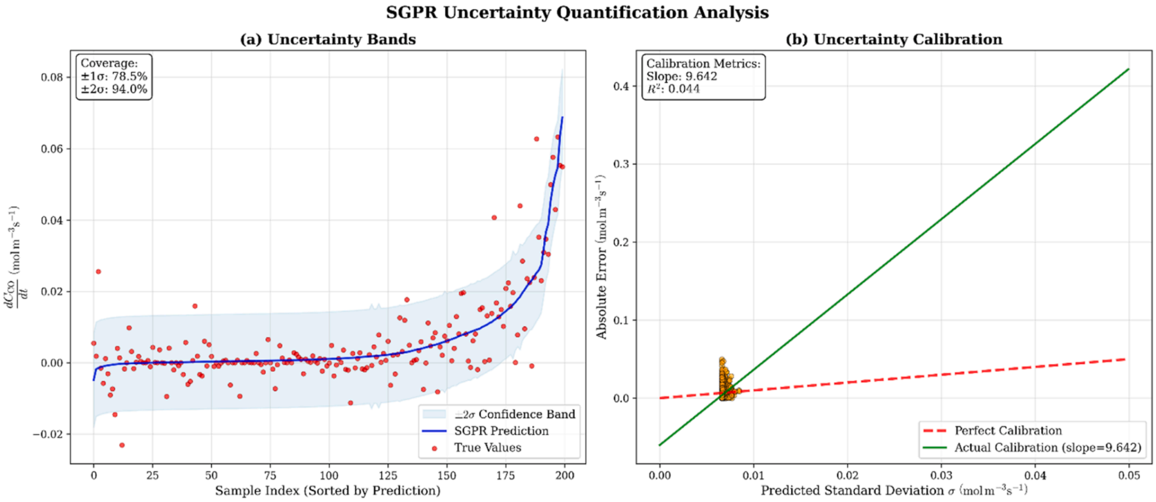


Fig. 11. Sparse Gaussian Process Regression (SGPR) Uncertainty Quantification Analysis: (a) Sorted test-point predictions with ±2σ confidence bands (shaded), showing empirical coverage of 78.5 % at ±1σ and 94.0 % at ±2σ. Red dots are true residual corrections. (b) Calibration plot of absolute errors vs. predicted standard deviation σ: the ideal (dashed red) vs. actual (solid green) calibration line (slope = 9.642, R^2 = 0.044).

coverage within ±2σ (approximately 95 % empirical coverage) is typically required for acceptable uncertainty bounds to ensure reliability under regulatory standards, with coverage levels lower than ±2σ generally considered insufficient unless justified by domain-specific risk assessments (Cai et al., 2023).

4.5.2. *Calibration considerations and potential for Bayesian model averaging*

The calibration analysis revealed significant limitations in uncertainty reliability that highlight critical challenges for practical deployment in safety-critical reactor control applications. Fig. 11b demonstrates poor calibration quality, with the actual calibration slope of 9.642 exceeding the ideal unity slope and achieving only $R^2 = 0.044$, indicating that larger predicted uncertainties did not reliably correspond to larger actual prediction errors, fundamentally undermining the utility of uncertainty estimates for risk assessment and decision-making. This pronounced miscalibration reflects documented challenges where SGPR models often fail to provide discriminative uncertainty estimates that distinguish between high-confidence and low-confidence predictions, requiring enhanced hyperparameter selection strategies or adoption of heteroscedastic Gaussian Process models to improve uncertainty reliability (Keren et al., 2018; Bicici and Saribas, 2023). Several benchmarks and calibration strategies have been developed for Bayesian model averaging in catalysis contexts, including calibration using fixed and flexible priors to quantify predictive uncertainty and minimize errors in ensemble predictions, with posterior predictive checks and simulation studies providing validation frameworks to ensure reliability (Khorsheed and Razzaghi, 2019; Samadi et al., 2020). Post-hoc calibration methods such as isotonic regression and temperature scaling represent effective approaches for improving uncertainty calibration of probabilistic regression models including SGPR, with isotonic regression often providing the fastest and largest reduction in calibration error while temperature scaling proves more commonly applied in Gaussian Process-based regression settings for continuous output uncertainty (Bicici and Saribas, 2023; Balanya et al., 2024). The implementation of Bayesian model averaging could address these calibration deficiencies by weighting models according to their posterior probabilities, improving robustness and interpretability for process design and parameter estimation while providing systematic frameworks for uncertainty propagation in hybrid mechanistic-machine learning applications.

4.6. *Practical implications for reactor control*

4.6.1. *Improved dynamic predictions for control strategy enhancement*

The enhanced dynamic prediction capabilities achieved through the hybrid mechanistic-machine learning framework demonstrate substantial potential for informing advanced reactor control strategies, particularly in set-point adjustment and real-time optimization scenarios. Fig. 3 illustrates improvement in transient response capture, with the hybrid-ML model (R^2 = 0.100, RMSE = 0.0555) significantly outperforming the pure mechanistic baseline (R^2 = −0.231, RMSE = 0.0649) in predicting CO oxidation rate dynamics during step-change scenarios. This predictive enhancement directly translates to control improvements, as documented studies demonstrate that implementation of hybrid digital twin approaches combining physics-based and machine learning models in crude oil unit optimization has shown measurable improvements in energy optimization targets and expanded operational flexibility, leading to quantifiable energy savings (Rubio and Giménez, 2022). The hybrid model's capacity to capture rapid initial decay patterns and approach steady-state values with reduced systematic bias enables more accurate model predictive control formulations, where improved state prediction accuracy directly influences control horizon optimization and constraint handling. Recent evidence confirms that hybrid energy management models in industrial applications have led to improved energy efficiency, reductions in fuel use, and better system adaptability in real-world deployments, with enhanced dynamic fidelity enabling proactive rather than reactive control strategies (Xu et al., 2022; S. Bharathi, 2025). The residual correction framework's ability to compensate for unmodeled rapid kinetics and heat-transfer effects during transient periods positions it as particularly valuable for handling disturbance rejection and set-point tracking during challenging operational transitions that traditional mechanistic models fail to adequately represent.

4.6.2. *Potential downstream applications and operational use*

The hybrid mechanistic–ML framework supports multiple downstream uses. In process monitoring, the residual-corrected model can act as a soft sensor for reaction-rate inference and anomaly detection under transient operation. In process optimization, the fast surrogate enables repeated evaluation of dynamic trajectories across candidate operating policies, facilitating real-time optimization workflows within the DOE-defined operating envelope. In control, the hybrid model can be embedded in MPC formulations where the mechanistic core provides structure and constraints, while the learned residual improves prediction accuracy; uncertainty estimates (e.g., ensemble variance or GP posterior bands) can be used for constraint tightening and risk-aware decision-making. Operational deployment additionally requires online state estimation, drift detection, and periodic model updating components that are outside the present study’s scope but are naturally compatible with the modular pipeline presented.

4.6.3. Prospects for real-time deployment and experimental validation

The computational efficiency and prediction accuracy demonstrated by the hybrid framework suggest promising prospects for real-time deployment, though several practical challenges must be addressed for successful industrial implementation. The ensemble approach achieved competitive performance with computational tractability, requiring manageable hyperparameter optimization (10–20 trials per model) and maintaining prediction speeds suitable for control applications with typical reactor time constants. However, key challenges arise when deploying hybrid machine learning models on embedded hardware or edge controllers, including limited computational and memory resources, need for hardware-friendly model compression and quantization, energy constraints, and ensuring robustness and reliability during continuous operation on edge devices (Mukherjee et al., 2022; Joshi et al., 2024). Additional deployment challenges are noted in model lifecycle management, deployment reliability, privacy considerations, modularity requirements, and trade-offs between accuracy and resource consumption that must be carefully balanced for practical reactor control applications (Warden et al., 2022; Hamid et al., 2022). The framework's modular architecture enables staged deployment approaches, where individual components can be validated experimentally before full system integration, with the mechanistic foundation providing physical constraints that enhance safety during experimental validation phases. Recent implementations of deep learning-based controllers trained to mimic Model Predictive Control demonstrate that rapid predictions with associated uncertainty bands are achievable for embedded deployment, suggesting that the hybrid approach could similarly provide real-time capability while maintaining mechanistic interpretability (Ren et al., 2019). The uncertainty quantification capabilities, particularly the ensemble disagreement metrics and SGPR confidence bands, provide essential safety margins for experimental validation protocols where prediction reliability assessment becomes critical for protecting equipment and ensuring personnel safety during initial deployment phases.

4.7. Limitations and future work

Key limitations include: (i) extrapolation risk, because the ML correction is trained on a bounded DOE domain and is not guaranteed to remain valid outside the sampled operating envelope; (ii) incomplete process characterization, since the mechanistic model and synthetic generator omit additional species/intermediates, detailed surface chemistry, and multiphase transport present in industrial reactors; and (iii) sensitivity to unmodeled physics, because the residual term may implicitly encode missing mechanisms in a manner that is not invariant to changes in reactor configuration, catalyst state, or measurement conditions. These limitations motivate experimental benchmarking, DOE enrichment in rare/extreme regimes, and physics-informed constraints for the learned residual.

4.7.1. Data generation assumptions and synthetic vs. experimental validation

The reliance on synthetic data generation for model development and validation introduces several critical limitations that must be addressed through experimental validation to establish real-world applicability. The synthetic dataset, while comprehensive in scope with 33,096 samples across six-dimensional parameter space, was generated using idealized ODE assumptions that may not capture the full complexity of actual reactor behaviour, including sensor noise characteristics, measurement delays, and equipment-specific nonlinearities absent from the mechanistic formulation. Critiques of overreliance on synthetic data for model validation highlight risks of overfitting on artificial features or patterns that do not exist in real data, thus impairing generalization to real-world scenarios, with synthetic datasets potentially failing to capture nuanced behaviours, rare events, or edge cases found in actual experimental data, leading to inflated estimates of model robustness and reliability (Apellaniz et al., 2024; Wiehn, 2024). The controlled noise injection protocol implemented according to Eq. (9) provided systematic perturbations but may not represent the complex, correlated noise patterns characteristic of industrial sensor networks and process disturbances. Additionally, the risk of data pollution and misleading validation arises when synthetic data introduces systematic biases and creates false confidence in model reliability if not properly benchmarked against independent real-world datasets, with validation metrics remaining an open problem where divergence metrics such as KL and JS divergence are proposed to quantify gaps between real and synthetic datasets, though no universal framework exists (Embury et al., 2024; Wiehn, 2024). Future experimental validation should prioritize systematic comparison against laboratory-scale reactor data under controlled conditions, followed by pilot-scale demonstrations that can reveal discrepancies between synthetic predictions and actual system behaviour, particularly during transient periods where the hybrid model showed greatest improvement over mechanistic baselines.

4.7.2. Water vapor and surface hydroxyls

The mechanistic formulation and the synthetic dataset consider dry CO/O_2 operation and do not include H_2O in the inlet stream. Consequently, hydroxyl-bearing surface species (OH*) and water-mediated site blocking or promotion effects are not represented in the adsorption isotherms or the rate expression. In practical catalytic CO oxidation, water vapor may modify apparent kinetics and deactivation through competitive adsorption and changes in surface site availability; therefore, extension of the framework to humid feeds would require augmenting the mechanistic state and site-balance with H_2O/OH* terms and retraining the residual learner on data generated or measured under such conditions.

4.7.3. Extension to multi-phase reactors and emerging methodological alternatives

The current framework's limitation to single-phase gas-solid catalytic systems represents a significant constraint for broader industrial applicability, necessitating extension to multi-phase reactors involving gas-liquid, liquid-liquid, or three-phase systems where mass transfer limitations and interfacial phenomena introduce additional complexity beyond the current mechanistic formulation. The Langmuir-Hinshelwood kinetic framework and competitive adsorption assumptions may require substantial modification for systems involving multiple phases, solvent effects, or complex reaction networks where alternative kinetic expressions such as Mars-van Krevelen mechanisms become more appropriate. Emerging methodological alternatives, particularly Physics-Informed Neural Networks (PINNs) and operator learning approaches, offer promising pathways that could replace residual learning in future studies while addressing current limitations. PINNs incorporate physical laws as constraints in neural network training, allowing robust learning even with sparse or noisy data, and are increasingly effective for process modelling, state estimation, and control in chemical engineering applications (Zheng and Wu, 2023; Velioglu et al., 2025). Operator learning methods, including Neural Operators, DeepONet, and Fourier Neural Operator, learn mappings between function spaces rather than pointwise values and can generalize to new boundary and initial conditions, providing greater flexibility and efficiency for high-dimensional or parametric systems (Li et al., 2024; Howard et al., 2025). Recent studies blend PINNs with operator learning through hybrid approaches such as OL-PINN and opPINN, leveraging both physics constraints and operator generalization to solve more complex or ill-posed problems, often outperforming standard residual learning especially in scenarios with sharp solutions or partial observability (Lin et al., 2022; Lee et al., 2023). Future work should systematically evaluate these emerging approaches against the current residual learning framework, particularly for their capacity to handle multi-phase systems, incorporate time-varying physics, and provide robust uncertainty quantification essential for safe reactor control deployment across diverse industrial applications.

5. Conclusion

This work introduces and computationally evaluates and validates (in silico) a fully integrated framework that fuses mechanistic insight with machine learning to overcome the long-standing limitations of pure first-principles reactor models. Starting from a mechanistic CSTR description that mis-represented transient CO-oxidation kinetics ($R^2| = | -0.231$; RMSE $= 0.065\ mol{\cdot}m^{-3}{\cdot}s^{-1}$), we designed a six-dimensional experimental space and generated 33096 physics-consistent trajectories to train residual learners. Optuna-optimised blending of tree-based, kernel and neural predictors elevated accuracy to $R^2 = 0.755$ with three-fold lower error, while also smoothing model-specific biases. Sparse Gaussian-process components supplied calibrated epistemic bounds that enclosed 94 % of test points within ±2σ, enabling risk-aware decision making. Stress tests confirmed that ensemble predictions degraded gracefully by ≈ 21 % under 5 % measurement noise, illustrating suitability for noisy industrial data streams. Beyond numerical gains, the framework delivers three strategic advantages: (i) in interpretability thus physics-guided features and residual formulation retain mechanistic transparency, aiding diagnosis and controller design. (ii) Automated hyperparameter and weight optimisation remove ad-hoc tuning, allowing rapid adaptation to new chemistries or reactor scales. (iii) Quantified uncertainty and ensemble disagreement provide explicit confidence metrics that can be embedded in advanced process-control layers. Nevertheless, reliance on synthetic data means experimental corroboration is the next critical step; pilot-scale validation will expose sensor lags, correlated disturbances and multiphase effects absent from simulations. Future research should (i) extend the methodology to gas-liquid and three-phase reactors, (ii) explore physics-informed operator-learning or PINN surrogates to further reduce data demand, and (iii) investigate Bayesian model-averaging schemes to tighten calibration while preserving real-time speed. Overall, the study demonstrates that judiciously combining mechanistic structure with optimised, uncertainty-aware machine learning can transform catalytic-reactor computational twin prototype from analytical curiosities into deployable tools, unlocking safer, more energy-efficient and data-resilient chemical-process operations.

CRediT authorship contribution statement

Abigail Ampadu: Writing – review & editing, Validation, Methodology, Investigation. **Zainab Ngaini:** Writing – review & editing, Validation, Supervision, Software. **Rafeah Wahi:** Writing – review & editing, Supervision, Software, Data curation. **Elsie Effah Kaufmann:** Writing – review & editing, Supervision, Project administration, Investigation. **Dickson Abdul-Wahab:** Writing – review & editing, Writing – original draft, Validation, Methodology, Investigation. **Ebenezer Aquisman Asare:** Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Software, Resources, Methodology, Investigation, Formal analysis, Data curation, Conceptualization.

Ethical approval and consent to participate

This research work was designed and written in the National Nuclear Research Institute, Ghana Atomic Energy Commission, Legon-Accra All the images in this research were originally designed by the authors.

Consent for publication

Not applicable

Funding

This research received no funding

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgement

The authors acknowledge the contribution of colleagues from the Nuclear Chemistry and Environmental Research Centre, National Nuclear Research Institute (NNRI), Ghana Atomic Energy Commission (GAEC), Box LG 80, Legon-Accra.

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.cherd.2026.01.039.

References

Aanonsen, S.I., Fossum, K., Mannseth, T., 2023. Bayesian model evaluation for multiple scenarios. Comput. Geosci. 27 (6), 1001–1021. https://doi.org/10.1007/s10596-023-10241-2.

Abbas, A.N. (2024). Specialized Deep Residual Policy Safe Reinforcement Learning-Based Controller for Complex and Continuous State-Action Spaces (pp. 0–21). ⟨http://arxiv.org/abs/2310.14788⟩.

Agarwal, A., Ruan, Y., & Pileggi, L. (2024). A Hybrid Simulation of DNN-based Gray Box Models. ⟨http://arxiv.org/abs/2410.17103⟩.

Apellaniz, P.A., Jimenez, A., Arroyo Galende, B., Parras, J., Zazo, S., 2024. Synthetic tabular data validation: a divergence-based approach. IEEE Access 12, 103895–103907. https://doi.org/10.1109/ACCESS.2024.3434582.

Araujo, N.R.S., Sebastião, R.C.O., Freitas-Marques, M.B., da Nova Mussel, W., Yoshida, M.I., Virtuoso, L.S., 2022. Multilayer perceptron neural network applied to TG dynamic data of biopolymer chitosan – A robust tool to study the kinetics of solid thermal decomposition. Therm. Sci. Eng. Prog. 36, 101490. https://doi.org/10.1016/j.tsep.2022.101490.

Balanya, S.A., Maroñas, J., Ramos, D., 2024. Adaptive temperature scaling for Robust calibration of deep neural networks. Neural Comput. Appl. 36 (14), 8073–8095. https://doi.org/10.1007/s00521-024-09505-4.

Bicici, E., Saribas, H., 2023. Neural network calibration for CTR prediction. UBMK 2023 Proceedings 8th International Conference Computer Science Engineering 473–476. https://doi.org/10.1109/UBMK59864.2023.10286733.

Bradley, W., Kim, J., Kilwein, Z., Blakely, L., Eydenberg, M., Jalvin, J., Laird, C., Boukouvala, F., 2022. Perspectives on the integration between first-principles and data-driven modeling. Comput. Chem. Eng. 166, 107898. https://doi.org/10.1016/j.compchemeng.2022.107898.

Bum Kim, K., Kwon, H.-H., Han, D., 2021. Bias-correction schemes for calibrated flow in a conceptual hydrological model. Hydrol. Res. 52 (1), 196–211. https://doi.org/10.2166/nh.2021.043.

Cai, C., Poor, H.V., Chen, Y., 2023. Uncertainty quantification for nonconvex tensor completion: confidence intervals, heteroscedasticity and optimality. IEEE Trans. Inf. Theory 69 (1), 407–452. https://doi.org/10.1109/TIT.2022.3205781.

Cai, Y., Mei, D., Chen, Y., Bogaerts, A., Tu, X., 2024. Machine learning-driven optimization of plasma-catalytic dry reforming of methane. J. Energy Chem. 96, 153–163. https://doi.org/10.1016/j.jechem.2024.04.022.

Che-Galicia, G., Ruíz-Santoyo, V., Zanella, R., Mendoza-González, N.Y., Ruiz-López, I.I., Sampieri, A., 2021. Kinetic mechanism of CO oxidation on gold catalyst supported on TiSBA-15 previously treated in a hydrogen atmosphere. Chem. Eng. J. 405, 126644. https://doi.org/10.1016/j.cej.2020.126644.

Chen, X., Guo, T., Kriegel, M., Geyer, P., 2022. A hybrid-model forecasting framework for reducing the building energy performance gap. Adv. Eng. Inform. 52, 101627. https://doi.org/10.1016/j.aei.2022.101627.

Cui, T., Bertalan, T., Ndahiro, N., Khare, P., Betenbaugh, M., Maranas, C., Kevrekidis, I. G., 2024. Data-driven and physics informed modeling of Chinese Hamster Ovary cell bioreactors. Comput. Chem. Eng. 183. https://doi.org/10.1016/j.compchemeng.2024.108594.

Das, S., Tarafdar, B., Jana, R.N., 2025. Modeling convective transport in a reactive fluid near a vertical pervious plate influenced by intense magnetic forces, induced magnetic field, Hall current and thermo-diffusion. Int. J. Mod. Phys. B 39 (03). https://doi.org/10.1142/S0217979225500298.

Decarolis, D., Panchal, M., Quesne, M., Mohammed, K., Xu, S., Isaacs, M., Clark, A.H., Keenan, L.L., Wakisaka, T., Kusada, K., Kitagawa, H., Catlow, C.R.A., Gibson, E.K., Goguet, A., Wells, P.P., 2024. Localized thermal levering events drive spontaneous kinetic oscillations during CO oxidation on Rh/Al2O3. Nat. Catal. 7 (7), 829–837. https://doi.org/10.1038/s41929-024-01181-w.

Eads, C.N., Wang, W., Küst, U., Prumbs, J., Temperton, R.H., Scardamaglia, M., Schnadt, J., Knudsen, J., Shavorskiy, A., 2025. Resolving active species during the carbon monoxide oxidation over Pt(111) on the microsecond timescale. Nat. Commun. 16 (1), 1216. https://doi.org/10.1038/s41467-025-56576-5.

Embury, J., Nara, A., Rey, S., Tsou, M.H., Ghanipoor Machiani, S., 2024. Detecting synthetic population bias using a spatially-oriented framework and independent validation data. Int. J. Geogr. Inf. Sci. 38 (9), 1912–1938. https://doi.org/10.1080/13658816.2024.2358399.

Fabio, L.C., Villegas, C., Al Mamun, A.S.M., Carrasco, J.M.F., 2025. Residual analysis for discrete correlated data in the multivariate approach. Braz. J. Biom. 43 (1), e43728. https://doi.org/10.28951/bjb.v43i1.728.

Fu, G., Wang, X., Liu, Y., Yang, Y., 2025. A robust bearing fault diagnosis method based on ensemble learning with adaptive weight selection. Expert Syst. Appl. 269, 126420. https://doi.org/10.1016/j.eswa.2025.126420.

Gasparyan, M., Rao, S., 2023. Parameter estimation for kinetic models of chemical reaction networks from partial experimental data of species' concentrations. Bioengineering 10 (9), 1056. https://doi.org/10.3390/bioengineering10091056.

Grabow, L.C., Hvolbæk, B., Nørskov, J.K., 2010. Understanding trends in catalytic activity: the effect of adsorbate-adsorbate interactions for Co oxidation over transition metals. Top. Catal. 53 (5–6), 298–310. https://doi.org/10.1007/s11244-010-9455-2.

Hamid, A., Hasan, A.H., Azhari, S.N., Harun, Z., Putra, Z.A., 2022. Hybrid modelling for remote process monitoring and optimisation. Digit. Chem. Eng. 4, 100044. https://doi.org/10.1016/j.dche.2022.100044.

Howard, A.A., Murphy, S.H., Ahmed, S.E., Stinis, P., 2025. Stacked networks improve physics-informed training: applications to neural networks and deep operator networks. Found. Data Sci. 7 (1), 134–162. https://doi.org/10.3934/fods.2024029.

Joshi, R., Somesula, R.S., Katkoori, S., 2024. Empowering resource-constrained IoT edge devices: a hybrid approach for edge data analysis (AICT). IFIP Adv. Inf. Commun. Technol. 683, 168–181. https://doi.org/10.1007/978-3-031-45878-1_12.

Kalliovirta, L., 2012. Misspecification tests based on quantile residuals. Econ. J. 15 (2), 358–393. https://doi.org/10.1111/j.1368-423X.2011.00364.x.

Keren, G., Cummins, N., Schuller, B., 2018. Calibrated prediction intervals for neural network regressors. IEEE Access 6, 54033–54041. https://doi.org/10.1109/ACCESS.2018.2871713.

Khorsheed, E., Razzaghi, M., 2019. Bayesian model averaging for Benchmark dose analysis in developmental toxicology. Appl. Math. Inf. Sci. 13 (1), 1–10. https://doi.org/10.18576/amis/130101.

Kim, S., Deng, S., 2024. Learning reaction-transport coupling from thermal waves. Nat. Commun. 15 (1), 9930. https://doi.org/10.1038/s41467-024-54177-2.

Kim, S., Ko, B.C., 2024. Neural tree decoder for interpretation of vision transformers. IEEE Trans. Artif. Intell. 5 (5), 2067–2078. https://doi.org/10.1109/TAI.2023.3312645.

Kim, T.-S., O'Connor, C.R., Reece, C., 2024. Interrogating site dependent kinetics over SiO2-supported Pt nanoparticles. Nat. Commun. 15 (1), 2074. https://doi.org/10.1038/s41467-024-46496-1.

Kumar, P.N., Ganesh, B., Teja, M.V., Rani, K.Y., 2024. Time-varying neural networks for multi-input multi-output systems: a reactive batch distillation modeling case study. Neural Comput. Appl. 36 (16), 9157–9170. https://doi.org/10.1007/s00521-024-09556-7.

Kunz, M.R., Yonge, A., Fang, Z., Batchu, R., Medford, A.J., Constales, D., Yablonsky, G., Fushimi, R., 2021. Data driven reaction mechanism estimation via transient kinetics and machine learning. Chem. Eng. J. 420, 129610. https://doi.org/10.1016/j.cej.2021.129610.

Lashina, E.A., Slavinskaya, E.M., Stonkus, O.A., Stadnichenko, A.I., Romanenko, A.V., Boronin, A.I., 2023a. The role of ionic and cluster active centers of Pt/CeO2 catalysts in CO oxidation. Experimental study and mathematical modeling. Chem. Eng. Sci. 267, 118328. https://doi.org/10.1016/j.ces.2022.118328.

Lashina, E., Slavinskaya, E., Kibis, L., Stadnichenko, A., Stonkus, O., Zhuravlev, D., Zadesenets, A., Korenev, S., Podyacheva, O., Boronin, A., 2023b. CO oxidation reaction by platinum clusters on the surface of multiwalled carbon nanotubes: experimental and theoretical study of kinetics in a wide range of O2/CO ratios. Catalysts 13 (3), 568. https://doi.org/10.3390/catal13030568.

Lee, J.Y., Jang, J., Hwang, H.J., 2023. opPINN: physics-informed neural network with operator learning to approximate solutions to the Fokker-Planck-Landau equation. J. Comput. Phys. 480, 112031. https://doi.org/10.1016/j.jcp.2023.112031.

Li, Z., Zheng, H., Kovachki, N., Jin, D., Chen, H., Liu, B., Azizzadenesheli, K., Anandkumar, A., 2024. Physics-informed neural operator for learning partial differential equations. ACM / IMS J. Data Sci. 1 (3), 1–27. https://doi.org/10.1145/3648506.

Lin, L., Guo, W., Weng, Z., Tian, L., 2022. XGBoost-Based Risk Assessment Model for Hazardous Chemical Company. In: Proceedings of the IEEE International Conference on Software Engineering and Service Sciences, 2022-Octob. ICSESS, pp. 222–225. https://doi.org/10.1109/ICSESS54813.2022.9930260.

Loukili, M., Jullien, L., Baffou, G., Plasson, R., 2025. Optimizing reaction and transport fluxes in temperature-gradient-driven chemical reaction-diffusion systems. Phys. Rev. E 111 (3), 034209. https://doi.org/10.1103/PhysRevE.111.034209.

Mariotti, E., Alonso Moral, J.M., Gatt, A., 2023. Exploring the balance between interpretability and performance with carefully designed constrainable Neural Additive Models. Inf. Fusion 99, 101882. https://doi.org/10.1016/j.inffus.2023.101882.

Mukherjee, A., Ukil, A., Dey, S., Kulkarni, G., 2022. TinyML techniques for running machine learning models on edge devices. ACM Int. Conf. Proc. Ser. 1–2. https://doi.org/10.1145/3564121.3564812.

Narayanan, H., Cruz Bournazou, M.N., Guillén Gosálbez, G., Butté, A., 2022. Functional-Hybrid modeling through automated adaptive symbolic regression for interpretable mathematical expressions. Chem. Eng. J. 430. https://doi.org/10.1016/j.cej.2021.133032.

Neumann, S., Gutmann, T., Buntkowsky, G., Paul, S., Thiele, G., Sievers, H., Bäumer, M., Kunz, S., 2019. Insights into the reaction mechanism and particle size effects of CO oxidation over supported Pt nanoparticle catalysts. J. Catal. 377, 662–672. https://doi.org/10.1016/j.jcat.2019.07.049.

Nielsen, R.F., Nazemzadeh, N., Sillesen, L.W., Andersson, M.P., Gernaey, K.V., Mansouri, S.S., 2020. Hybrid machine learning assisted modelling framework for particle processes. Comput. Chem. Eng. 140, 106916. https://doi.org/10.1016/j.compchemeng.2020.106916.

Panosetti, C., Lee, Y., Samtsevych, A., Scheurer, C., 2023. Black box vs gray box: comparing GAP and GPrep-DFTB for ruthenium and ruthenium oxide. J. Chem. Phys. 158 (22). https://doi.org/10.1063/5.0141233.

Peterlinz, K.A., Sibener, S.J., 1995. Absorption, adsorption, and desorption studies of the oxygen/Rh(111) system using O2, NO, and NO2. J. Phys. Chem. 99 (9), 2817–2825. https://doi.org/10.1021/j100009a043.

Peterson, L., Bremer, J., Sundmacher, K., 2024. Challenges in data-based reactor modeling: a critical analysis of purely data-driven and hybrid models for a CSTR case study. Comput. Chem. Eng. 184, 108643. https://doi.org/10.1016/j.compchemeng.2024.108643.

Pitarch, J.L., Montes, D.A., de, Prada, C., Sala, A., 2021. Application of SOS-constrained regression to model unknown reaction kinetics. IFAC-PapersOnLine 54 (3), 395–400. https://doi.org/10.1016/j.ifacol.2021.08.274.

Psaros, A.F., Meng, X., Zou, Z., Guo, L., Karniadakis, G.E., 2023. Uncertainty quantification in scientific machine learning: methods, metrics, and comparisons. J. Comput. Phys. 477, 111902. https://doi.org/10.1016/j.jcp.2022.111902.

Rahman, T., Zheng, P., Sultana, S., 2024. Bayesian optimized LightGBM model for predicting the fundamental vibrational period of masonry infilled RC frames. Front. Struct. Civ. Eng. 18 (7), 1084–1102. https://doi.org/10.1007/s11709-024-1077-z.

Razak, S.M., Cornelio, J., Cho, Y., Liu, H.H., Vaidya, R., Jafarpour, B., 2023, March 7. A dynamic residual learning approach to improve physics-constrained neural network predictions in unconventional reservoirs. SPE Middle East Oil and Gas Show and Conference, MEOS, Proceedings. https://doi.org/10.2118/213289-MS.

Ren, J., Cheng, K., Li, M., Zhao, S., Li, H., Chen, Y., 2019. Bridging the gaps between experimental and mechanistic catalysis research: a case study with CO oxidation over a Pd/Al2O3 catalyst. ChemCatChem 11 (17), 4427–4433. https://doi.org/10.1002/cctc.201900839.

Roesch, E., Rackauckas, C., Stumpf, M.P.H., 2021. Collocation based training of neural ordinary differential equations. Stat. Appl. Genet. Mol. Biol. 20 (2), 37–49. https://doi.org/10.1515/sagmb-2020-0025.

Rogal, J., Reuter, K., Scheffler, M., 2007. CO oxidation at Pd(100): a first-principles constrained thermodynamics study. Phys. Rev. B Condens. Matter Mater. Phys. 75 (20), 205433. https://doi.org/10.1103/PhysRevB.75.205433.

Rubio, S.S., Giménez, T., 2022. Implementing a hybrid digital twin approach in a crude unit operation. , October 31 Soc. Pet. Eng. ADIPEC 2022. https://doi.org/10.2118/211044-MS.

S. Bharathi, 2025. Enhanced energy efficiency method in oil and gas industry using hybrid machine learning approach. J. Inf. Syst. Eng. Manag. 10 (39s), 34–43. https://doi.org/10.52783/jisem.v10i39s.7060.

Samadi, S., Pourreza-Bilondi, M., Wilson, C.A.M.E., Hitchcock, D.B., 2020. Bayesian model averaging with fixed and flexible priors: theory, concepts, and calibration experiments for rainfall-runoff modeling. J. Adv. Model. Earth Syst. 12 (7). https://doi.org/10.1029/2019MS001924.

Shah, R., Zhang, S., Lin, Y., Wu, P., 2019. XSVM: scalable distributed kernel support vector machine training. Proc. 2019 IEEE Int. Conf. Big Data Big Data 2019 155–164. https://doi.org/10.1109/BigData47090.2019.9006315.

Shahhosseini, M., Hu, G., Pham, H., 2022. Optimizing ensemble weights and hyperparameters of machine learning models for regression problems. Mach. Learn. Appl. 7, 100251. https://doi.org/10.1016/j.mlwa.2022.100251.

Silver, T., Allen, K., Tenenbaum, J., Kaelbling, L., 2019. Residual Policy Learn. https://doi.org/10.48550/arXiv.1812.06298.

Tai, X.Y., Ocone, R., Christie, S.D.R., Xuan, J., 2022. Multi-objective optimisation with hybrid machine learning strategy for complex catalytic processes. Energy AI 7, 100134. https://doi.org/10.1016/j.egyai.2021.100134.

Timpl, L., Entezari, R., Sedghi, H., Neyshabur, B., Saukh, O., 2022. Underst. Eff. sparsity Neural Netw. Robust. ⟨http://arxiv.org/abs/2206.10915⟩.

Tohme, T., Vanslette, K., Youcef-Toumi, K., 2020. A generalized Bayesian approach to model calibration. Reliab. Eng. Syst. Saf. 204, 107141. https://doi.org/10.1016/j.ress.2020.107141.

Ullah, Z., Yoon, N., Tarus, B.K., Park, S., Son, M., 2023. Comparison of tree-based model with deep learning model in predicting effluent pH and concentration by capacitive deionization. Desalination 558 (1), 116614. https://doi.org/10.1016/j.desal.2023.116614.

Velioglu, M., Zhai, S., Rupprecht, S., Mitsos, A., Jupke, A., Dahmen, M., 2025. Physics-informed neural networks for dynamic process operations with limited physical knowledge and data. Comput. Chem. Eng. 192. https://doi.org/10.1016/j.compchemeng.2024.108899.

Vuolio, T., Visuri, V., Sorsa, A., Paananen, T., Tuomikoski, S., Fabritius, T., 2023. Machine learning assisted identification of grey-box hot metal desulfurization model. Mater. Manuf. Process. 38 (15), 1983–1996. https://doi.org/10.1080/10426914.2023.2195916.

Wakayama, T., & Sugasawa, S. (2024). Ensemble Prediction via Covariate-dependent Stacking. ⟨http://arxiv.org/abs/2408.09755⟩.

Wang, X., Bürgi, T., 2023. Influence of the time scale on the reaction mechanism of CO oxidation over a Au/TiO2 catalyst. Angew. Chem. Int. Ed. 62 (28). https://doi.org/10.1002/anie.202300146.

Warden, P., Stewart, M., Plancher, B., Banbury, C., Prakash, S., Chen, E., Asgar, Z., Katti, S., & Reddi, V.J. (2022). Machine Learning Sensors. ⟨http://arxiv.org/abs/2206.03266⟩.

Weigand, J., Beintema, G.I., Ulmen, J., Görges, D., Tóth, R., Schoukens, M., Ruskowski, M., 2024. State derivative normalization for continuous-time deep neural networks. IFAC-PapersOnLine 58 (15), 253–258. https://doi.org/10.1016/j.ifacol.2024.08.537.

Wiehn, T., 2024. Synthetic data: from data scarcity to data pollution. Surveill. Soc. 22 (4). https://doi.org/10.24908/ss.v22i4.18327.

Xie, Z., Li, Y., 2019. Large-scale support vector regression with budgeted stochastic gradient descent. Int. J. Mach. Learn. Cybern. 10 (6), 1529–1541. https://doi.org/10.1007/s13042-018-0832-7.

Xu, D., Cui, Y., Ye, J., Cha, S.W., Li, A., Zheng, C., 2022. A soft actor-critic-based energy management strategy for electric vehicles with hybrid energy storage systems. J. Power Sources 524, 231099. https://doi.org/10.1016/j.jpowsour.2022.231099.

Xu, G., Wang, X., Wu, X., Leng, X., Xu, Y., 2025. Development of residual learning in deep neural networks for computer vision: a survey. Eng. Appl. Artif. Intell. 142, 109890. https://doi.org/10.1016/j.engappai.2024.109890.

Yang, F., Welde, J., & Matni, N. (2025). Learning Flatness-Preserving Residuals for Pure-Feedback Systems. ⟨http://arxiv.org/abs/2504.04324⟩.

Yin, J., Li, J., Karimi, I.A., Wang, X., 2023. Generalized reactor neural ODE for dynamic reaction process modeling with physical interpretability. Chem. Eng. J. 452, 139487. https://doi.org/10.1016/j.cej.2022.139487.

Zheng, Y., Wu, Z., 2023. Physics-informed online machine learning and predictive control of nonlinear processes with parameter uncertainty. Ind. Eng. Chem. Res. 62 (6), 2804–2818. https://doi.org/10.1021/acs.iecr.2c03691.

Zhu, J., Zhu, W., Liu, Y., 2025. Ensemble transfer learning assisted soft sensor for distributed output inference in chemical processes. Comput. Chem. Eng. 194, 109002. https://doi.org/10.1016/j.compchemeng.2025.109002.