Chinese Journal of Chemical Engineering 95 (2026) 313–327

Contents lists available at ScienceDirect

Chinese Journal of Chemical Engineering

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

Full Length Article

Combining multiscale decomposition with time-series modeling: An approach for petrochemical processes

Mengjie Luo [1], Mengxuan Zhang [2], Xiaogang Shi [1], Xingying Lan [1,*], Yongdong Liu [2,*], Wei Li [2]

[1] State Key Laboratory of Heavy Oil Processing, China University of Petroleum-Beijing, Beijing 102249, China
[2] China Industrial Control Systems Cyber Emergency Response Team, Beijing 100040, China

ARTICLE INFO

Article history:
Received 15 April 2025
Received in revised form
1 March 2026
Accepted 3 March 2026
Available online 1 April 2026

Keywords:
Chemical processes
Multiscale
Prediction
Variational mode decomposition
Neural networks
Attention

ABSTRACT

Accurately predicting key product features is essential for process optimization and safety in petrochemical processes. However, conventional data-driven AI models often struggle to adequately capture the strong nonlinearity and multi-scale characteristics of these systems. To address this challenge, this paper proposes a multi-scale prediction framework that integrates variational mode decomposition (VMD), attention mechanisms, and temporal convolutional networks (TCN), referred to as DATCN. This framework facilitates the extraction of multi-scale features across temporal, spatial, and frequency domains, enabling cross-scale interaction modeling in complex processes. Comparative experiments on the Tennessee Eastman (TE) benchmark and industrial fluid catalytic cracking (FCC) long-term time-series datasets demonstrate that the proposed model achieves superior prediction accuracy (RMSE reduced by 24.75% to 37.26% *versus* other deep-learning models). The results demonstrate the framework's capability to address both nonlinear dynamics and spatiotemporal heterogeneity, offering a promising solution for intelligent optimization in petrochemical production.
© 2026 The Chemical Industry and Engineering Society of China, and Chemical Industry Press Co., Ltd. All rights are reserved, including those for text and data mining, AI training, and similar technologies.

1. Introduction

The petrochemical industry lies at the heart of the energy and materials supply chain and plays a vital role in the global economy [1]. As regulations on new energy tighten and demand for high-value materials grows, there is a pressing need to optimize and control petrochemical processes. This calls for industrial models that are both accurate and broadly applicable. However, petrochemical processes typically involve multiphase flow, coupled heat and mass transfer, and complex chemical reactions [2,3]. They exhibit clear multi-scale behavior in time and space [4–6] and respond strongly to feedstock variability. Developing mechanistic unit-operation models that capture these multi-scale dynamics with high accuracy and generality remains a major challenge.

Artificial intelligence models have become catalysts for unlocking value from industrial data [7,8] and are widely applied in petrochemistry [9–14]. Feedforward neural networks excel at predicting yields and optimizing process outputs [15–18]. Convolutional neural networks (CNN) and long short-term memory networks (LSTM) attract attention for time-series modeling, thanks to their powerful feature extraction and long-dependency learning abilities [19–21].

However, complex petrochemical processes, such as fluid catalytic cracking, hydrocracking, and reforming, require coordinated operation of multiple units, where key features are tightly interrelated [22,23]. Process features (*e.g.*, temperature, pressure, flow rates) and the underlying chemical reactions interact strongly [24,25]. These interactions drive nonlinear, multi-scale, and non-steady behavior [23,24,26–28]. Therefore, data-driven models constructed directly from raw multivariate time-series data often exhibit limited predictive performance [29,30]. Some researchers have built hybrid frameworks that combine multiple predictive models to reveal hidden patterns in nonlinear systems [31–34]. However, the high dimensionality and multi-scale nature of multi-unit processes still pose severe challenges for achieving both high accuracy and strong generalization [23]. This limitation arises from two main factors. First, dynamic features at different time scales in the raw data are mixed, preventing the model from clearly

* Corresponding authors.
E-mail addresses: lanxy@cup.edu.cn (X. Lan), liuyongdong@cics-cert.org.cn (Y. Liu).

https://doi.org/10.1016/j.cjche.2026.03.001
1004-9541/© 2026 The Chemical Industry and Engineering Society of China, and Chemical Industry Press Co., Ltd. All rights are reserved, including those for text and data mining, AI training, and similar technologies.

distinguishing long-term trends and short-term fluctuations. Second, redundant and irrelevant variables accumulate in the high-dimensional space during modeling, which obscures the key dynamic features driving the prediction task. Therefore, separating multi-scale features and performing targeted screening and reconstruction of high-dimensional variables before time-series modeling are essential for improving the predictive performance of petrochemical process models.

VMD is an adaptive, non-recursive signal processing method that effectively addresses mode mixing issues in Empirical Mode Decomposition (EMD) and its variants [35–39]. It decomposes a time series into a set of smooth intrinsic mode functions (IMFs) across multiple frequency scales, reducing non-stationarity and mode mixing [23]. By reducing nonlinearities and redundancy, VMD improves dynamic feature extraction and has been widely used in complex industrial data analysis [23,40–43]. However, capturing long-term dependencies in these multi-scale features is challenging for traditional temporal models, which suffer from vanishing or exploding gradients, limited parallelism, and weak long-term dependency modeling. The TCN was developed to overcome these issues [44]. TCN integrates the local pattern extraction capabilities of CNN with the temporal modeling strengths of recurrent architecture [45,46]. Without recurrence, TCN avoids gradient issues and stably learns long-term dependencies. Causal convolutions further preserve temporal order, ensuring valid predictions [47,48].

In multi-unit petrochemical processes, different variables and their multi-scale modal components have varying importance for prediction. VMD can decouple multi-scale dynamic features, but it also increases the feature dimension. Without proper feature selection, feeding all VMD components equally into TCN may introduce redundant information and reduce TCN's ability to capture key features.

Based on the analysis above, this paper tackles the modeling challenges of non-stationarity, high-dimensional coupling, and multi-scale dynamics in petrochemical processes. We propose DATCN, a multi-scale attention-based temporal prediction framework that integrates VMD decomposition, feature recalibration, and TCN modeling. The main innovations and contributions are summarized in three points:

(1) Multi-scale signal decoupling and structured feature representation

Enhanced VMD decomposes non-stationary petrochemical signals into distinct frequency components, effectively resolving feature overlapping and providing a structured input for the model.

(2) Multi-variable multi-scale feature recalibration

The Squeeze-and-Excitation (SE) attention mechanism is integrated before temporal modeling to adaptively weight and filter structured features, suppressing redundant noise while highlighting key dynamic information critical for process monitoring.

(3) Long-term sequence modeling with causal convolution

TCN captures long-term dependencies through dilated causal convolutions, improving stability and generalization across complex operating conditions.

The rest of this paper is structured as follows: Section 2 introduces the mathematical and physical models used, while Section 3 outlines the modeling process and results of industrial experiments, both TE process and industrial FCC unit are analyzed and discussed.

2. Proposed Methodology

As shown in Fig. 1, this paper proposes DATCN, a multi-scale modeling framework integrating temporal-spatial-frequency for complex petrochemical processes modeling. First, operational data are collected from petrochemical units and preprocessed through cleaning, outlier removal, and normalization to construct a high-dimensional structured time-series dataset. To reduce the influence of redundant features, a hybrid Filter-Wrapper feature selection method developed by our group is applied [31], enhancing data quality for subsequent modeling. At the frequency scale, a Whale Optimization Algorithm (WOA)-based strategy is employed to adaptively optimize the parameters of VMD. This enables progressive decomposition of the signal into IMFs with distinct frequency characteristics, effectively denoising the data while revealing latent dynamics across different frequency bands. Subsequently, the resulting multi-frequency components are processed by a spatial attention mechanism, which learns dynamic weights to highlight salient features in the spatial dimension. These spatially enhanced multi-scale frequency features are subsequently fed into TCN for time-scale modeling. It leverages causal convolution to preserve temporal causality and dilated convolution to expand the receptive field for the integrated modeling of space-time and frequency scales. The TCN efficiently captures long-term dependencies and complex temporal patterns. Each IMF is modeled individually, and the final prediction is obtained *via* signal reconstruction. By combining time-aware learning, spatial feature enhancement, and frequency-domain decomposition, the proposed framework enables comprehensive multi-scale dynamic modeling, offering a robust solution for intelligent prediction in industrial petrochemical systems.

2.1. Adaptive signal decomposition module

2.1.1. VMD framework

The primary process of the VMD algorithm involves iteratively searching for the optimal solution of the variational model to determine the center frequency and bandwidth of each decomposed component. Assuming that the original signal z is decomposed into k intrinsic mode components, a constrained variational model is constructed, as shown in Eq. (1):

$$\begin{cases} \min_{\{u_k,\omega_k\}} \left\{ \sum_k \left\| \partial_t [(\delta(t) + j/(\pi t)) \times u_k(t)] e^{-j\omega_k t} \right\|_2^2 \right\} \\ \text{s.t.} \ \sum_{i=1}^{k} u_k(t) = z(t) \end{cases} \tag{1}$$

where $u_k = \{u_1, u_2, \ldots, u_k\}$ represents the modal functions, $\omega_k = \{\omega_1, \omega_2, \ldots, \omega_k\}$ denotes the center frequency of each mode, t represents time. To solve the constrained optimization problem, the VMD algorithm introduces penalty terms and Lagrange multipliers, converting the constrained variational model into an unconstrained variational problem, as shown in Eq. (2):

$$\begin{aligned} L(\{u_k\}, \{\omega_k\}, \lambda) = \alpha \sum_k \left\| \partial_t [(\partial(t) + j/(\pi t)) \times u_k(t)] e^{-j\omega_k t} \right\|_2^2 \\ + \left\| z(t) - \sum_k u_k(t) \right\|_2^2 + \langle \lambda(t), z(t) - \sum_k u_k(t) \rangle \end{aligned} \tag{2}$$

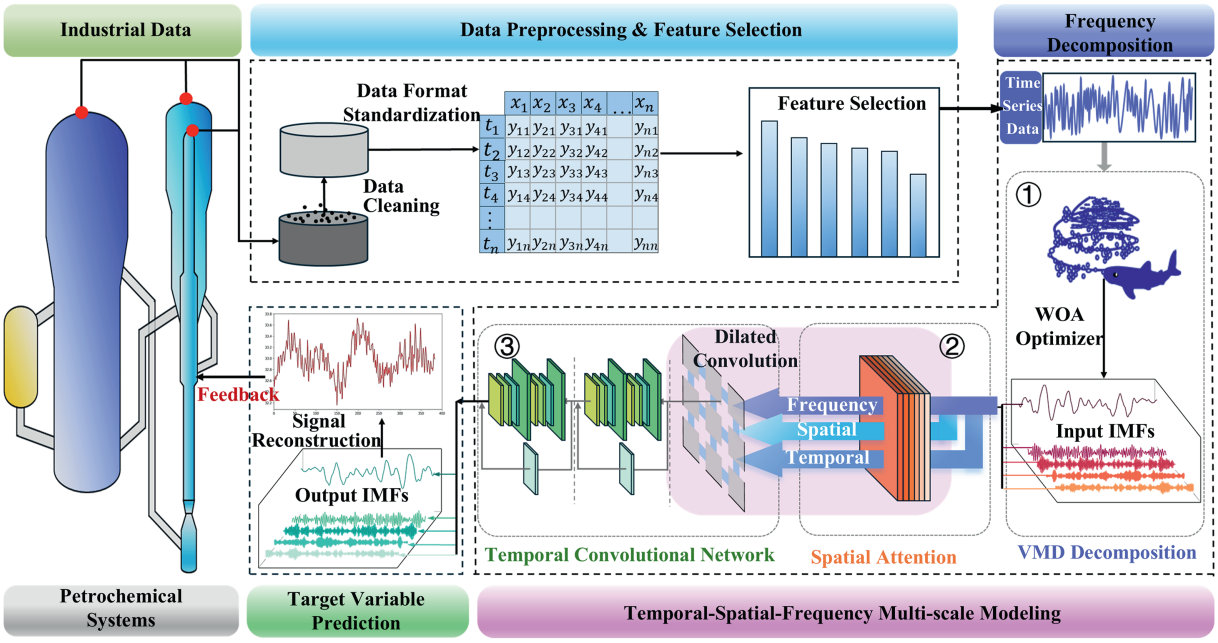


Fig. 1. The overall framework of the proposed DATCN model.

where α is the second-order penalty factor, and λ is the Lagrange multiplier. Using the alternating direction multiplier method to solve, we get Eq. (3):

$$\hat{u}_k^{n+1}(\omega)=\frac{\hat{f}(\omega)-\sum_{i\neq k}\hat{u}_i(\omega)+\hat{\lambda}(\omega)/2}{1+2\alpha(\omega-\omega_k)^2} \tag{3}$$

where ω is the frequency, n denotes the number of iterations, $\hat{u}_k^{n+1}(\omega)$, $\hat{z}(\omega)$, $\hat{\lambda}(\omega)$ are the Fourier transforms of $u_k^{n+1}(t)$, $z(\omega)$ and $\lambda(t)$. The formulas for updating $\hat{u}_k$, ω_k and $\hat{\lambda}$ are given in Eqs. (4)–(6):

$$\hat{u}_k^{n+1}(\omega)=\frac{z(\omega)-\sum_{i=1,i<k}^{k}\hat{u}_k^{n+1}(\omega)-\sum_{i=1,i>k}^{k}\hat{u}_i^{n}(\omega)+\hat{\lambda}^{n}(\omega)\Big/2}{1+2\alpha\left(\omega-\omega_k^b\right)^2} \tag{4}$$

$$\omega_k^{n+1}=\frac{\int_0^\infty \omega\left|\hat{u}_k^{n+1}(\omega)\right|^2 \mathrm{d}\omega}{\int_0^\infty \left|\hat{u}_k^{n+1}(\omega)\right|^2 \mathrm{d}\omega} \tag{5}$$

$$\hat{\lambda}^{b+1}(\omega)=\hat{\lambda}^{b}(\omega)+\lambda\left[\hat{z}(\omega)-\sum_k\hat{u}_k^{b+1}(\omega)\right] \tag{6}$$

The iteration stop condition is shown in Eq. (7), where ε is a threshold value:

$$\sum_k\left\|\hat{u}_k^{n+1}-\hat{u}_k^{n}\right\|_2^2\Big/\left\|\hat{u}_k^{n}\right\|_2^2<\varepsilon \tag{7}$$

The detailed execution flow of the VMD algorithm is illustrated in Fig. 2. First, u_k, ω_k, and λ are initialized. Then, $\hat{u}_k$ and ω_k are updated according to Eqs. (4) and (5). Next, the algorithm checks if the condition in Eq. (7) is met. If not, the iteration continues with updates to n and k. If the condition is satisfied, the Lagrange multiplier λ is updated. Finally, the algorithm evaluates whether the stopping condition for the iteration is met. If so, the iteration process concludes; otherwise, it continues to update.

2.1.2. Parameter optimization strategy

VMD performs well in signal decomposition, but its effectiveness depends on selecting appropriate hyperparameters, especially the number of modes (k) and the penalty factor (α), which significantly affect the results [49]. If k is too small, the signal may not be fully decomposed, leading to incomplete separation of key frequency components and potential mode aliasing. If k is too large, over-decomposition can occur, generating spurious components. The value of α affects the stationarity of the decomposed modes and the retention of detailed information. A smaller α may produce finer decomposition but risks overfitting by keeping redundant frequencies, while a larger α may smooth the result but lose important frequency information. Thus, choosing the right values for k and α is crucial for effective signal decomposition.

The WOA simulates the hunting behavior of humpback whales. It achieves efficient global search and local exploitation through the shrinking encirclement and spiral bubble net mechanisms [50]. Compared to Genetic Algorithm (GA) and Particle Swarm Optimization (PSO), WOA converges faster. It also shows greater stability in high-dimensional parameter spaces. In this study, WOA optimizes the key parameters of VMD (k and α) to improve signal decomposition accuracy.

In VMD decomposition, each IMF component ideally corresponds to a narrowband mode with a distinct center frequency and a concentrated spectrum. To assess decomposition quality, this paper uses minimum envelope entropy as a fitness function to optimize the decomposition parameters.

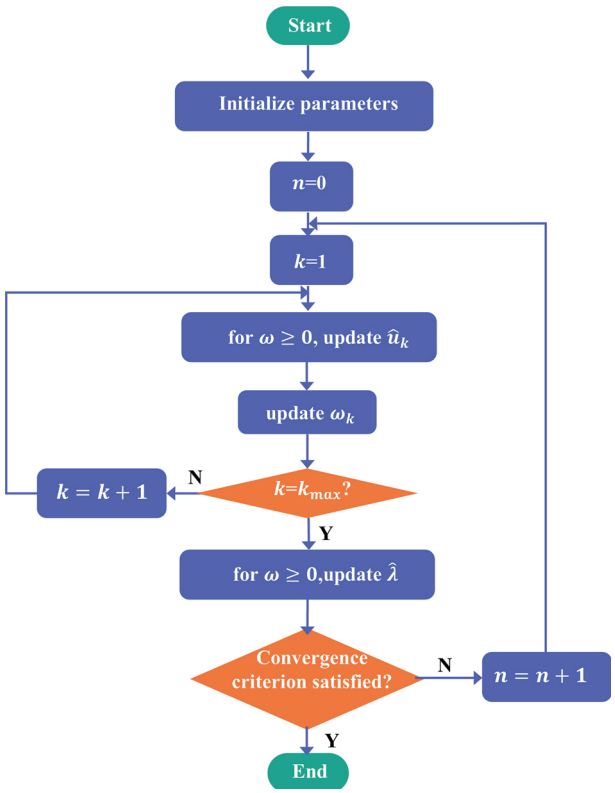


Fig. 2. Flow chart of the VMD algorithm.

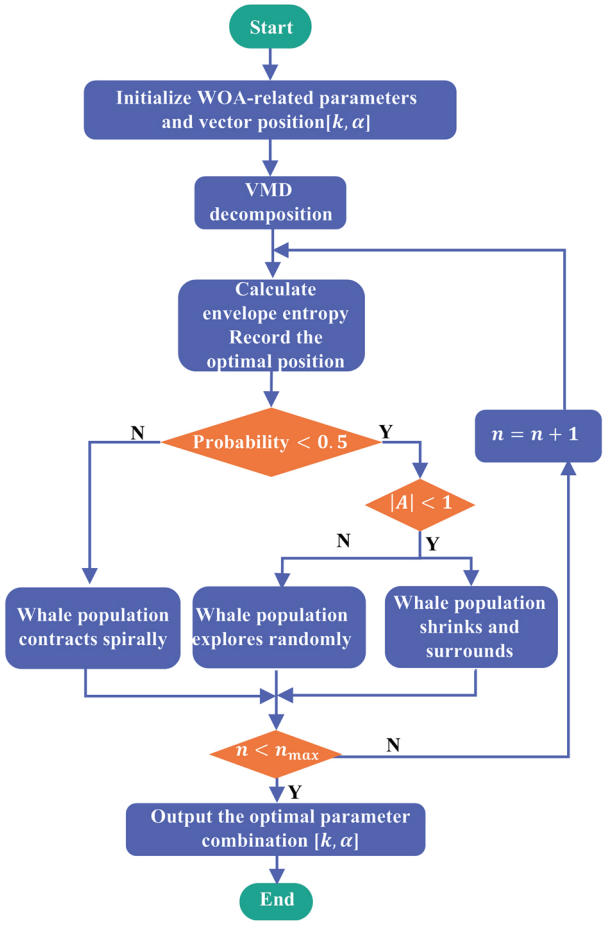


Fig. 3. Flowchart of the WOA-based VMD parameter optimization process.

The envelope entropy E_p is calculated from the envelope signal $a(t)$, which is obtained from the original signal using the Hilbert transform. First, the envelope sequence is normalized:

$$g_j = a(j) \Big/ \sum_{i=1}^{N} a(i) \tag{8}$$

Envelope entropy is defined as:

$$E_{\mathrm{p}} = -\sum_{j=1}^{N} g_j \log_2 g_j \tag{9}$$

Envelope entropy serves as a measure of signal complexity. Lower envelope entropy indicates a more structured and orderly signal, reflecting higher information content and less noise [51]. Minimizing envelope entropy guides the optimization of VMD parameters (k, α), enabling the generation of IMF components that are more regular and stable. This improves both the physical interpretability of the decomposition results and the effectiveness of subsequent time-series modeling. The flowchart illustrating the WOA-based optimization of VMD parameters is shown in Fig. 3.

2.2. Squeeze-and-excitation spatial attention module

The importance of different VMD-derived modes and variables varies significantly across prediction tasks. Therefore, a spatial attention module based on the Squeeze-and-Excitation (SE) structure is integrated to recalibrate spatial feature representation. As shown in Fig. 4, it adaptively weights features through lightweight channels to emphasize informative features and suppress irrelevant ones with low computational cost [52,53]. This mechanism improves the ability to learn complex physicochemical processes in petrochemical production, significantly enhancing prediction accuracy and generalization.

The implementation of the SE module consists of the following three steps:

2.2.1. Global feature squeeze

Global average pooling is applied along the temporal dimension of each channel to obtain global feature statistics. The compressed statistics of the i-th feature are given by:

$$z_i = F_{\mathrm{sq}}(x_i) = \frac{1}{T}\sum_{i=1}^{T} x_i(t) \tag{10}$$

T denotes the time step. The compressed vector z_i represents the average response of the feature over the current time window.

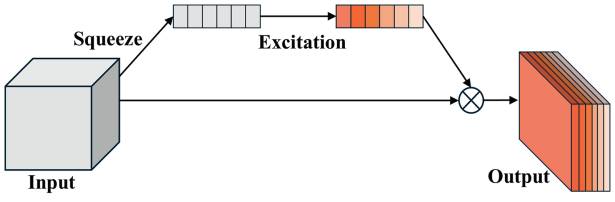


Fig. 4. Structure of the SE attention module.

2.2.2. Adaptive weight excitation

To capture nonlinear dependencies among features, a gating mechanism using two fully connected layers generates the channel weight vector s:

$$s_i = \sigma(W_2\delta(W_1 z_i)) \tag{11}$$

here, σ denotes the Sigmoid activation function, δ denotes the ReLU activation function, and W_1 and W_2 are the weight matrices of the two fully connected layers, respectively.

This step allows the model to automatically identify key and noisy modes and assign appropriate weights to each.

2.2.3. Feature recalibration

Finally, the weight vector s is multiplied element-wise with the original feature matrix to obtain the weighted feature matrix $\tilde{x}$:

$$\tilde{x}_i = s_i \cdot x_i \tag{12}$$

After recalibration, features strongly related to the prediction task are enhanced, while irrelevant or noise-dominated features are suppressed. This yields optimized inputs for subsequent TCN modeling, improving the model's ability to capture and learn critical modes.

2.3. Temporal convolutional network model

TCN is a convolutional neural network designed for time series modeling. It combines causal convolution, dilated convolution, and residual connections to capture long-term dependencies while preserving temporal structures. Compared to traditional RNN, TCN supports parallel and efficient computation, making it well-suited for modeling complex dynamics in petrochemical processes.

2.3.1. Causal convolution mechanism

Causal convolution is a fundamental operation in TCN designed to preserve temporal consistency by preventing information leakage from future time steps. At any time t, the output depends only on the current and past inputs within the kernel's receptive field. This design ensures that the model learns true causal relationships, as illustrated in Fig. 5.

Assume that the input time series is $x(t) = \{x_0, x_1, x_2, \ldots, x_t\}$ and the convolution kernel is j, then the causal convolution can be defined as Eq. (13):

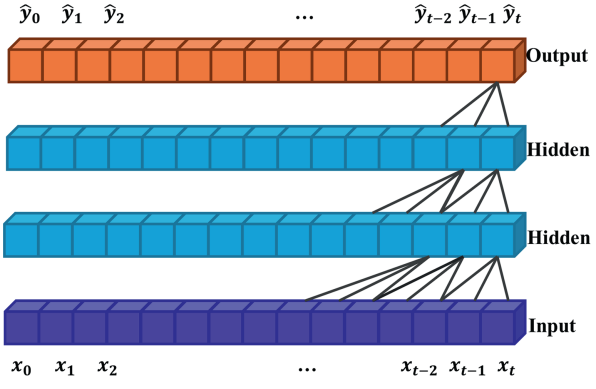


Fig. 5. Schematic diagram of causal convolutional network.

$$\hat{y}(t) = \sum_{i=0}^{m-1} p(i)\cdot x_{t-i} \tag{13}$$

where $p(i)$ represents the i-th weight of the convolution kernel, m denotes the size of the kernel, and $\hat{y}(t)$ is the output at time t.

2.3.2. Dilated convolution structure

Traditional convolutional neural networks often struggle with time series data due to a limited receptive field. The fixed kernel size constrains their ability to capture long-term dependencies-an essential requirement in industrial applications. While pooling layers can enlarge the receptive field, they often lead to information loss. In contrast, dilated convolution in TCN introduces a dilation factor d, which inserts gaps between kernel elements. This design enlarges the receptive field without reducing resolution, enabling the model to learn long-time dependencies while preserving data integrity, as shown in Fig. 6.

The dilated convolution is defined as Eq. (14):

$$\hat{y}(t) = \sum_{i=0}^{m-1} p(i)\cdot x_{t-d\cdot i} \tag{14}$$

where $t-d\cdot i$ indicates the direction of the historical data. When $d = 1$, the dilated convolution reduces to standard convolution. As d increases, the kernel can cover a wider temporal range, allowing the model to capture longer-term dependencies. The receptive field of dilated convolution grows exponentially with the number of layers, approximately $O(m\cdot 2^l)$, allowing TCN to effectively capture long-term dependencies from multi-scale temporal patterns in industrial data.

2.3.3. Residual learning block

To address gradient vanishing and explosion during deep network training, TCN incorporates residual blocks into its architecture, as shown in Fig. 7. These residual connections alleviate information loss during training by directly passing input signals to deeper layers through shortcut paths [54]. Each residual block contains two dilated convolution layers followed by ReLU activation. Weight regularization is applied to prevent gradient explosion during training.

As shown in Fig. 8, the TCN architecture comprises multiple stacked residual blocks. Each block captures temporal dependencies at different scales using causal and dilated convolutions, followed by a fully connected layer for final prediction.

3. Results and Discussion

In order to examine the performance and application potential of the DATCN model on complex chemical processes, we fully verify the performance of the model through the TE simulation process and the industrial FCC process.

3.1. Case study I: Tennessee Eastman process

The TE process is a simulation platform developed by Eastman Chemical Company, reflecting actual chemical reaction processes. Its process chart is shown in Fig. 9. The system includes five main units: reactor, condenser, compressor, separation tower, and stripping tower. As a standard test platform, the TE process simulates the dynamic complexities of actual production environments [55]. The generated data is free from external interference

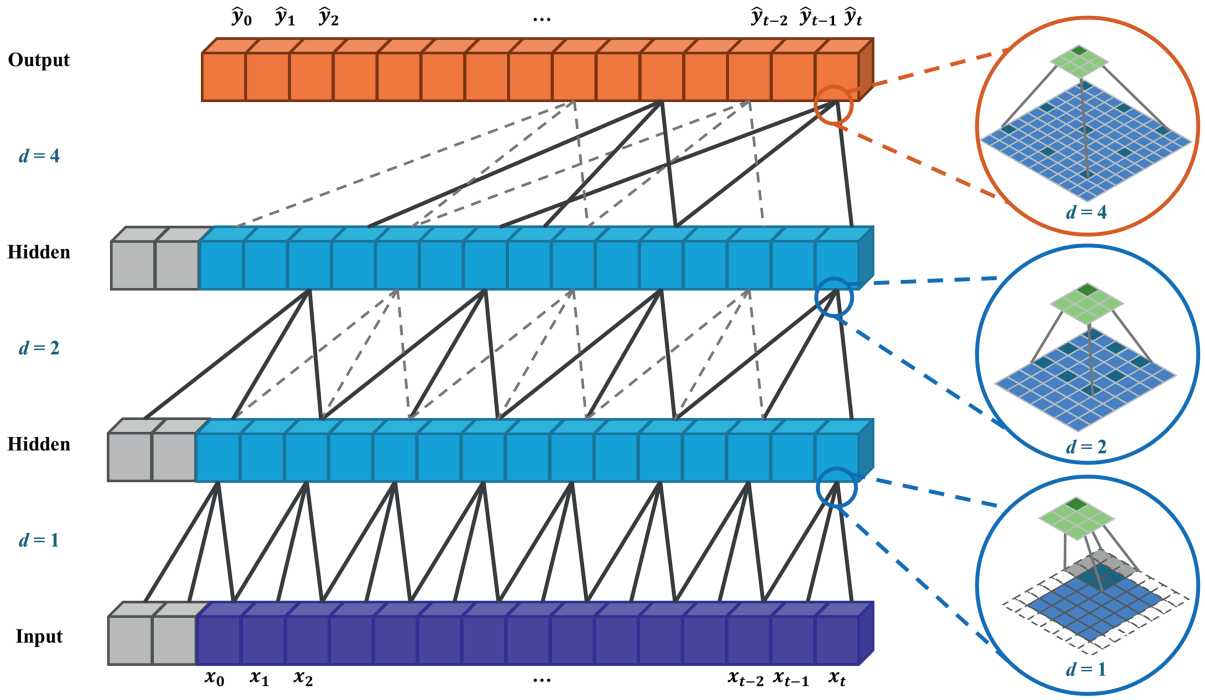


Fig. 6. Structure of dilated convolution with different dilation factors.

and displays dynamic variations, nonlinearity, and multi-scale behaviors. Therefore, this study uses TE process data to validate the proposed model's performance.

The TE process has 52 features: 12 control, 19 analytical, and 21 direct measurement features. The verification dataset includes 48 h of simulation data under normal conditions, with samples taken every 3 min, totaling 960 samples. Due to the steady-state nature of the TE process under normal conditions, the dataset is limited in size and highly regular. It is split 6:4 into training and test sets, with the relatively large test set used to rigorously evaluate the model's prediction accuracy and generalization in small-sample conditions. The model's performance is validated using the mole content of component A in the purge as the target feature.

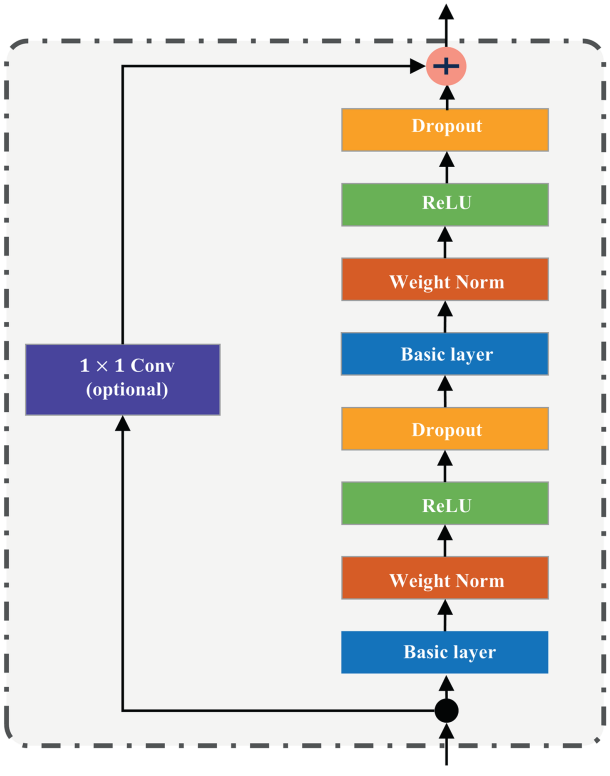


Fig. 7. Schematic diagram of residual block structure.

To reduce model complexity and improve prediction accuracy, we adopt the Filter-Wrapper model proposed by our group for feature selection [31]. The selected optimal feature subset is listed in Table 1.

Based on the feature selection results, VMD is first applied to decompose the TE data. Given the differences in physical mechanisms and frequency characteristics among process variables such as temperature and flow rate, each variable is independently decomposed in sequence with a unified mode number k to maintain a consistent input structure. The parameters $k \in [2,8]$ and $\alpha \in [200, 3000]$ are optimized using WOA, yielding optimal values of $k = 5$ and $\alpha = 1721$. Under these parameters, VMD decomposes the signal into five IMF components, as shown in Fig. 10(a). For n variables, decomposition yields $n \times k$ IMF components, which are concatenated along the channel dimension to form a high-dimensional input matrix $x_{in} \in R^{(t\times(n\times k))}$. Each IMF is then used as input to the TCN model for individual prediction, and the outputs are reconstructed to obtain the final prediction of the target variable, as shown in Fig. 10(b). The prediction results demonstrate that the DATCN model not only captures the overall data distribution accurately but also mitigates the influence of non-stationarity on predictions, yielding more stable and consistent results.

To further evaluate the model's prediction accuracy, Fig. 11 presents the distributions of the true and predicted values, along with their relative errors. The results show strong consistency between the two distributions, with most relative errors below

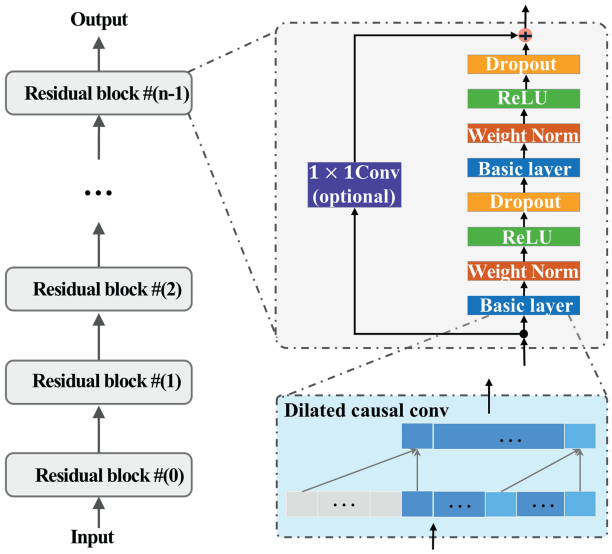


Fig. 8. Schematic diagram of TCN network structure.

0.01, confirming the high accuracy of the DATCN model. This performance is primarily attributed to DATCN's multiscale signal processing and modeling strategy, which allows the model to capture the overall dynamic characteristics of the data and suppress noise.

To comprehensively evaluate model performance, DATCN is compared with standard TCN, several hybrid models (TCN-LSTM, TCN-GRU, and CNN-LSTM), as well as Transformer and graph neural network (GNN) models. The predicted results of different models are presented in Fig. 12.

As shown in Fig. 12, the TCN model (Fig. 12(a)) can reasonably predict trend changes in the data, but its prediction curve exhibits significant sawtooth fluctuations. This indicates that the single TCN structure is highly sensitive to high-frequency noise, leading to unstable prediction results and relatively scattered point cloud distributions. The hybrid models TCN-LSTM (Fig. 12(b)) and TCN-GRU (Fig. 12(c)) leverage the memory capabilities of RNN units to mitigate the non-stationarity. However, they fail to effectively capture the global trend and show lower prediction accuracy. The prediction curves of CNN-LSTM (Fig. 12(d)), Transformer (Fig. 12(e)), and GNN models (Fig. 12(f)) are overly smooth, almost eliminating local details and transient fluctuations. This is due to their global or strong feature extraction mechanisms, such as CNN pooling, Transformer global attention, and GNN neighborhood aggregation, which can excessively smooth signals rich in high-frequency dynamics.

To evaluate model performance, Root Mean Square Error (RMSE), Mean Square Error (MSE), Mean Absolute Error (MAE), and Relative Error (RE) are used to assess prediction accuracy and stability, as shown in Fig. 13. The results show that DATCN consistently outperforms all baseline models. Its prediction error is reduced by at least 23.43%, with a maximum improvement of 52.94% over the other models. This performance improvement is primarily attributed to DATCN's multi-scale collaborative modeling mechanism. VMD decomposes non-stationary signals into multiple scales, reducing signal volatility and enhancing stationarity. The spatial attention mechanism adaptively emphasizes key modal features while suppressing redundant information, improving the effectiveness of input features. TCN efficiently captures long-range temporal dependencies through causal and dilated convolutions, fully leveraging historical dynamic information. The synergy of these

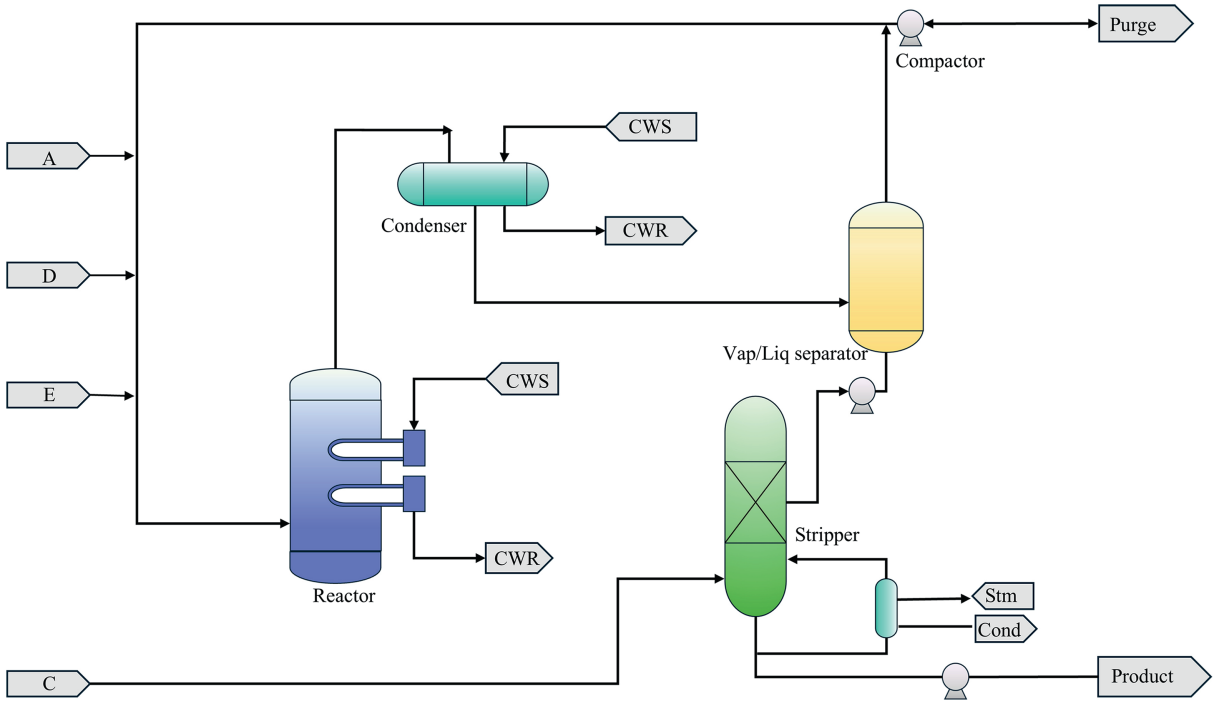


Fig. 9. Process flow diagram of the TE process.

Table 1
Optimal feature subset for the TE process.

Feature ID	Description	Feature ID	Description
XMEAS7	The pressure of reactor	XMEAS18	The temperature of stripper
XMEAS10	Venting rate	XMEAS19	The steam flow of stripper
XMEAS11	The temperature of separator	XMEAS20	Compressor power
XMEAS13	Liquid level of separator	XMV1	D feed flow
XMEAS16	Liquid level of stripper	XMV3	A feed flow

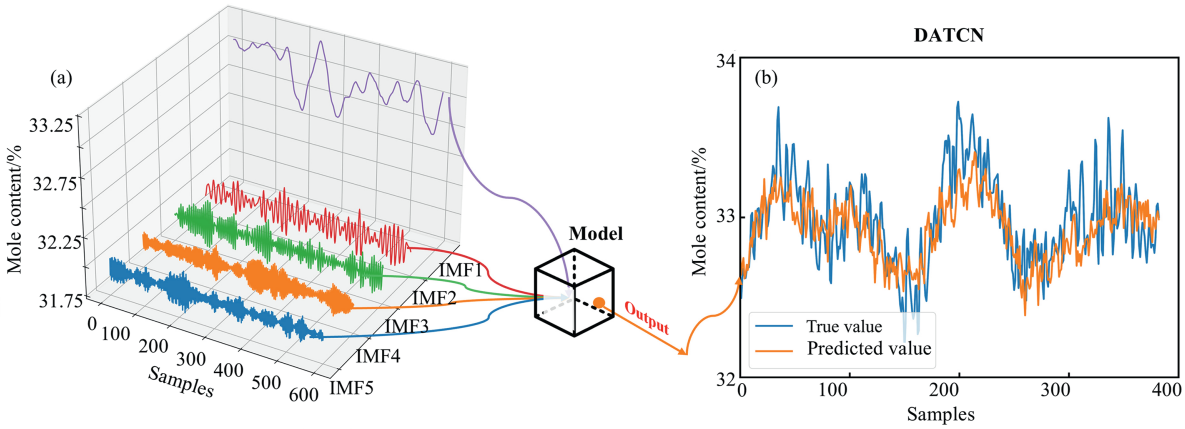


Fig. 10. Results of component A in the TE process: (a) VMD decomposition; (b) DATCN prediction comparison.

three components across temporal, spatial, and frequency domains not only suppresses noise effectively but also preserves high-frequency physical details. This approach avoids jagged fluctuations and excessive smoothing commonly seen in other models, achieving superior prediction accuracy and stability on TE data.

To further validate DATCN's practical performance, a case study from an industrial petrochemical unit is presented next to demonstrate the model's effectiveness under highly volatile and noisy conditions.

3.2. Case study II: industrial FCC process

FCC process is a key unit in deep petroleum refining, primarily used to convert heavy oil into fuel oil and light olefins [56]. It consists of three main subsystems: the reaction-regeneration system, the fractionation system, and the absorption–stabilization system. Among these, the reaction-regeneration system is central to the process and includes the riser reactor, disengager, and regenerator, forming a raw material-product-catalyst circulation system [57]. This system is characterized by tightly coupled flow, reaction, heat transfer, and mass transfer processes, exhibiting strong nonlinearity, multi-scale behavior, and nonstationary dynamics [8,58]. In practice, even minor variations in reaction conditions can significantly impact system performance, with disturbances propagating throughout the riser-disengager-regenerator loop due to strong causality. Consequently, the development of a high-precision prediction model is crucial for optimizing FCC operations and guiding process parameter adjustments.

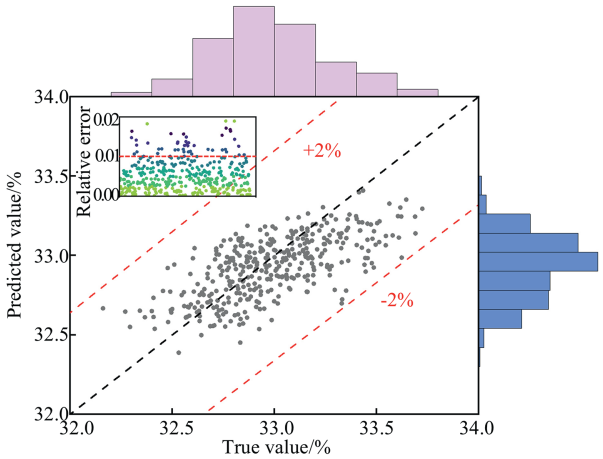


Fig. 11. Relative error distribution of DATCN for component A in the TE process.

This study evaluates the performance of DATCN by predicting the gasoline yield from an FCC unit at a refinery in Northwest China. The schematic of the reaction-regeneration system is shown in Fig. 14.

This study utilizes 97 DCS feature bit numbers of industrial data, spanning 90 days with a sampling interval of 20 min, resulting in a total of 6480 entries. To enhance data completeness and accuracy, preprocessing is required. First, missing data is filled using polynomial interpolation based on neighboring cycles. Then,

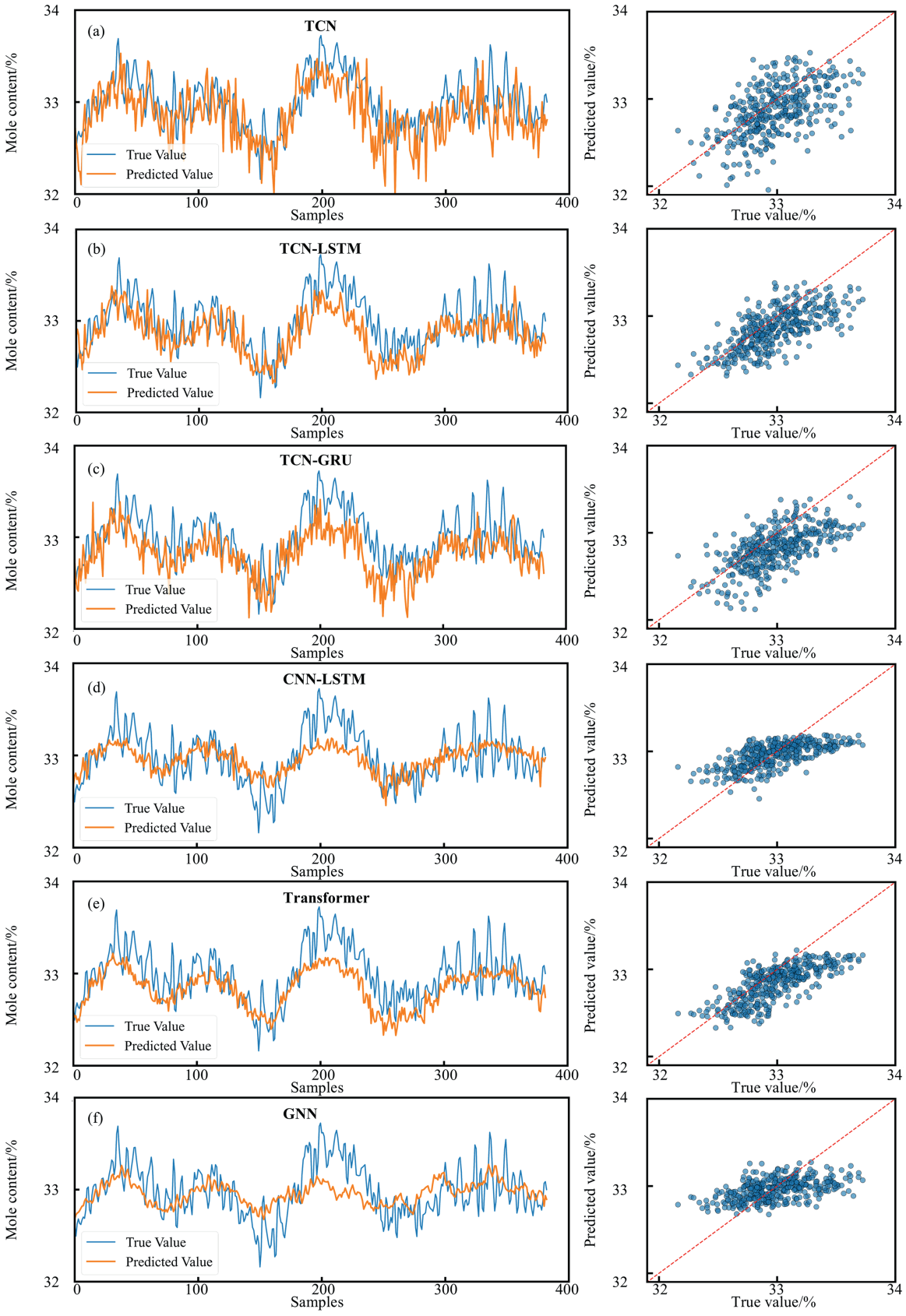


Fig. 12. Prediction comparison of component A in the TE process using different models: (a) TCN; (b) TCN-LSTM; (c) TCN-GRU; (d) CNN-LSTM; (e) Transformer; (f) GNN.

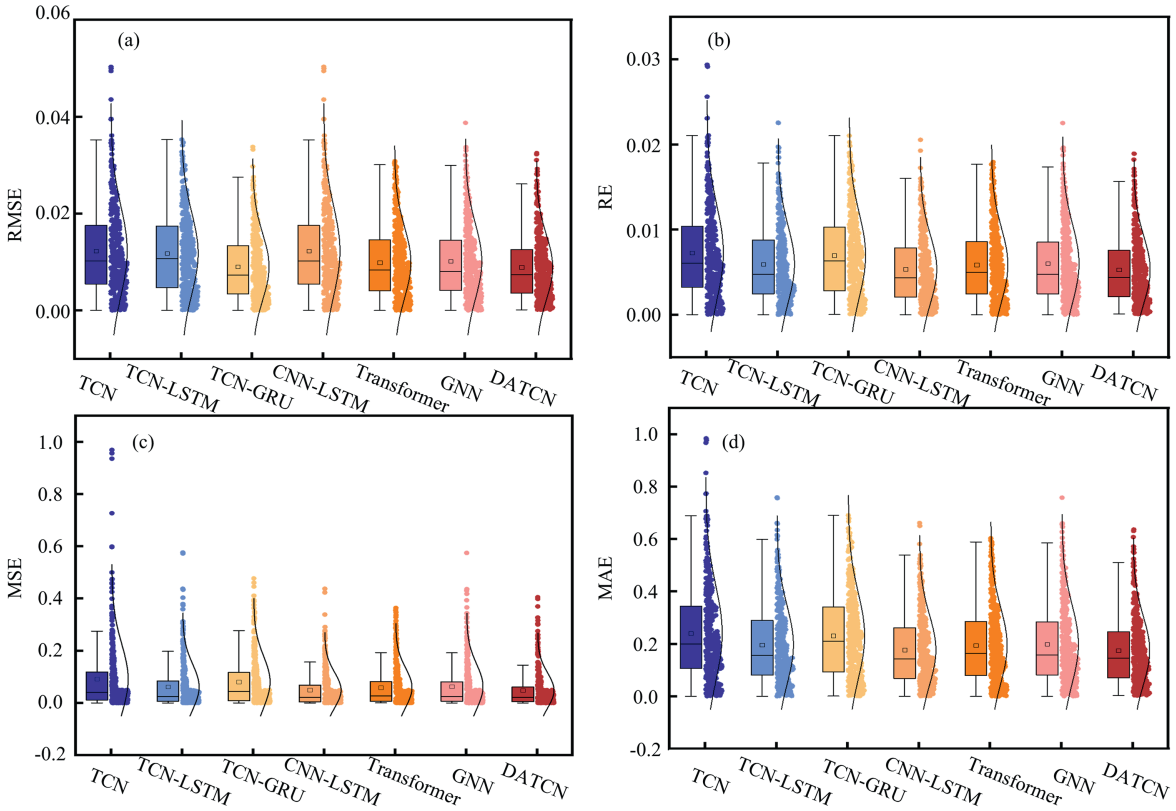


Fig. 13. Comparison of performance metrics for component A in the TE process: (a) RMSE; (b) RE; (c) MSE; (d) MAE.

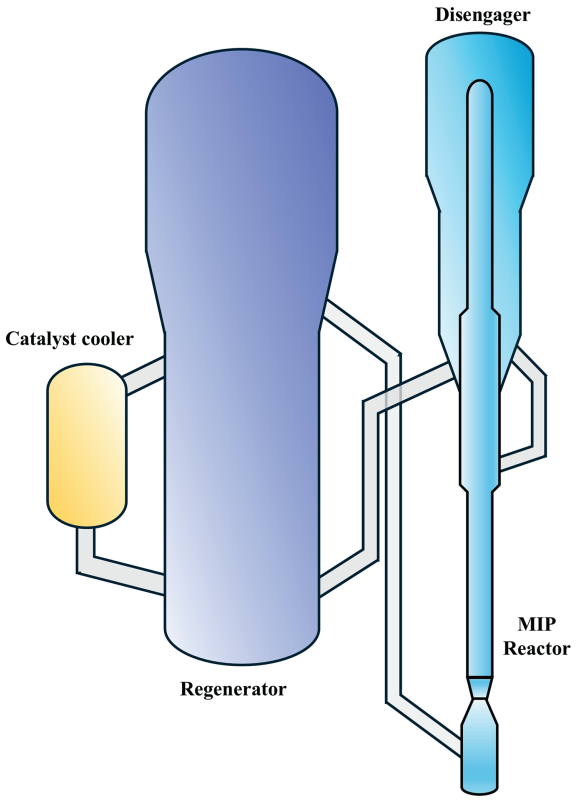


Fig. 14. Process flow diagram of the reaction-regeneration system in the FCC process.

outliers are removed using the standard deviation of the data and the PauTa criterion.

Given the large number of features, feature selection is performed prior to modeling and validation. The hybrid Filter-Wrapper feature selection method, proposed by Lan's group, is employed, and its effectiveness has been validated in several complex petrochemical processes. This method identifies 34 features most relevant to the target features. Table 2 lists the descriptions and corresponding DCS bit numbers of the optimal feature subsets. For data security reasons, the actual bit numbers are omitted.

The optimal feature subset selected includes features such as the flow rate of raw material atomizing vapor, the flow of pre-lifting dry gas and vapor, the steam flow rate in the stripping section, fluidized and conveying air flow rates in the catalyst cooler, as well as temperatures and pressures in key system components such as the riser, disengager, regenerator, and catalyst cooler. To verify the relevance of these features, a mechanism analysis is conducted. The flow rates of raw material atomizing vapor, pre-lifting dry gas, pre-lifting vapor, and steam in the stripping section directly influence the flow of oil, gas, and catalyst, which in turn affect product yield [59]. The temperatures in the riser, disengager, regenerator, and of the catalyst directly impact reaction selectivity and catalyst activity [60]. Similarly, the pressures in the riser, disengager, and cyclone affect catalyst flow and distribution, influencing both reaction selectivity and yield. The pressure drop across the spent catalyst slide valve determines the catalyst circulation and fluidization, influencing its contact with the oil. Additionally, the FCC unit may experience excess heat, affecting reaction depth. The catalyst cooler serves to cool the catalyst,

Table 2
Optimal feature subset for the industrial FCC unit.

Feature ID	Description	Feature ID	Description
FIC001	The flow rate of raw material atomizing vapor	TIA003	The temperature of catalyst in the small-scale feeding line
FIC002	The flow of pre-lifting dry gas in riser	TI006	The temperature at the lower part of the dilute phase in the regenerator R-102
FIC003	The flow of pre-lifting vapor in riser	TICA002	The temperature of the dense phase in the regenerator R-102
FIC004	The steam flow rate in the stripping section	TI007	The temperature at the main air inlet of the regenerator coke burning drum
FIC005	The flow of combustion oil to regenerator R-102	TI008	The temperature of the catalyst in the cold regenerator feed line
FI001	The fluidized air flow rate in catalyst cooler R-103A	TIA004	The temperature of the gas collection chamber at the top of the regenerator R-102
FI002	The fluidized air flow rate in catalyst cooler R-103B	TI009	The temperature at the upper part of the catalyst cooler R-103A
FI003	The conveying air flow rate in catalyst cooler R-103B	TI010	The temperature at the upper part of the catalyst cooler R-103B
TIA001	The outlet temperature of oil and gas from the top of disengager R-101	TI011	The temperature at the middle section of the catalyst cooler R-103A
TI001	The bottom temperature of the first reaction region in riser	TI012	The temperature at the bottom section of the catalyst cooler R-103A
TI002	The inlet temperature of the single-stage cyclone in disengager R-101	PI001	The feedstock pressure of the riser reactor R-101A
TICA001	The temperature of the pre-lifting section in the riser reactor R-101A	PDRCA001	The pressure drop across the spent catalyst slide valve
TI002A	The temperature of gas collection chamber in the top of the R-101	PDR001	The pressure drop across the single-stage in the disengager R-101
TI003	The upper temperature of the dilute phase in the disengager R-101	PDR002	The pressure drop of the primary cyclone in the disengager R-101
TI004	The upper temperature of the dilute phase in the disengager R-101	PDIA001	The balance pressure difference in the riser reactor
TI005	The bottom temperature of the dilute phase in the disengager R-101	PDRA001	The pressure drop across the catalyst cooler slide valve
TIA002	The catalyst temperature in the large-scale feeding line	WR001	The catalyst level of catalyst cooler

removing excess heat and regulating the regenerator temperature [61]. Therefore, the fluidized and conveying air flow, along with the temperature and catalyst level in the catalyst cooler, play a crucial role in product yield. By combining hybrid Filter-Wrapper feature selection with mechanism analysis, this study identifies 34 key features as inputs for the DATCN model, thereby improving the prediction accuracy for gasoline yield.

It is worth noting that this study specifically focuses on predicting gasoline yield over a continuous future period, which differs from traditional regression tasks and presents significant challenges in industrial systems. Industrial FCC data exhibits strong nonlinearity, non-stationarity, and high noise, making models with limited historical samples prone to underfitting. Therefore, unlike the 6:4 training-test split used for the TE process, this study adopts an 8:2 split based on chronological order. The first 80% of the data forms the training set, enabling the model to fully capture long-term dependencies and latent dynamic trends. The remaining 20% serves as the test set to evaluate the model's generalization and stability over future periods. This time-sequence-based partitioning not only reflects real industrial scenarios where historical data are leveraged to predict future outcomes but also effectively prevents temporal information leakage from random shuffling.

To evaluate the performance of the DATCN model proposed in this paper, Figs. 15 and 16 present its application results in FCC gasoline yield prediction using the optimal feature subset from Table 2, along with a comparative analysis against multiple models. As illustrated in the figures, the predicted values of DATCN closely track the dynamic trajectory of the actual data, while the tight clustering of scatter points along the diagonal line confirms the model's high accuracy and robust stability. While other models perform well in regions with minor fluctuations (*e.g.*, after 500 sample points), their accuracy declines in areas with larger fluctuations (0-500 sample points). This indicates their difficulty in handling complex industrial time-series data and capturing its dynamic characteristics. In contrast, the DATCN model maintains its predictive accuracy even in regions with significant fluctuations, highlighting its robustness for industrial data.

The performance metrics of the different models are presented in Fig. 16(h). The DATCN model demonstrates significant advantages across four metrics: RMSE, MSE, RE, and MAE, with all values achieving optimal results. Specifically, compared to other models, the DATCN model reduces MSE by 43.38% to 60.64%, RMSE by 24.75% to 37.26%, MAE by 22.91% to 36.88%, and RE by 23.88% to 37.71%. These results validate the DATCN model's superior accuracy and reliability in predicting complex petrochemical processes.

This performance enhancement is attributed to the framework's effective decoupling of multi-scale information and its hierarchical modeling capability. Specifically, VMD decomposes complex industrial signals into physically meaningful intrinsic modes. Subsequently, the SE attention mechanism adaptively highlights the dominant modes of critical variables, such as riser reaction temperature and regeneration pressure, thereby filtering out noise and ensuring feature clarity. Finally, TCN utilizes causal dilated convolutions to capture

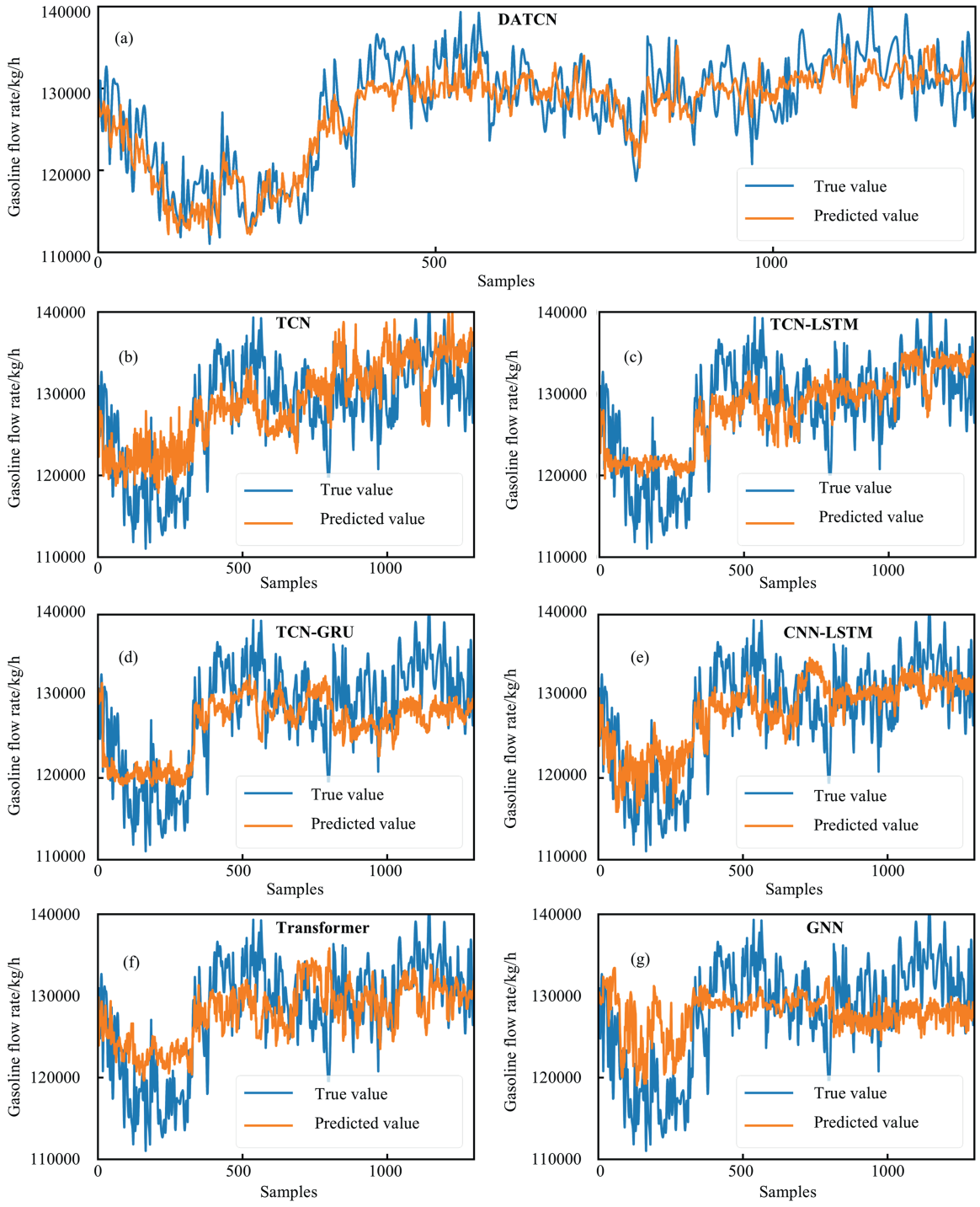


Fig. 15. Prediction comparison of gasoline flow rate in the FCC process using different models: (a) DATCN; (b) TCN; (c) TCN-LSTM; (d) TCN-GRU; (e) CNN-LSTM; (f) Transformer; (g) GNN.

multi-scale dynamic features. Its expanded receptive field covers extended time windows, enabling the precise capture of nonlinear cumulative effects from historical disturbances on current yield.

In summary, the performance enhancement of DATCN transcends the mere accumulation of model complexity. Instead, it stems from the synergistic integration of frequency-domain decoupling, feature recalibration, and causal sequence modeling. This synergy effectively reconciles the trade-off between noise suppression and dynamic pattern capture, ensuring enhanced accuracy and stability in predicting complex petrochemical processes.

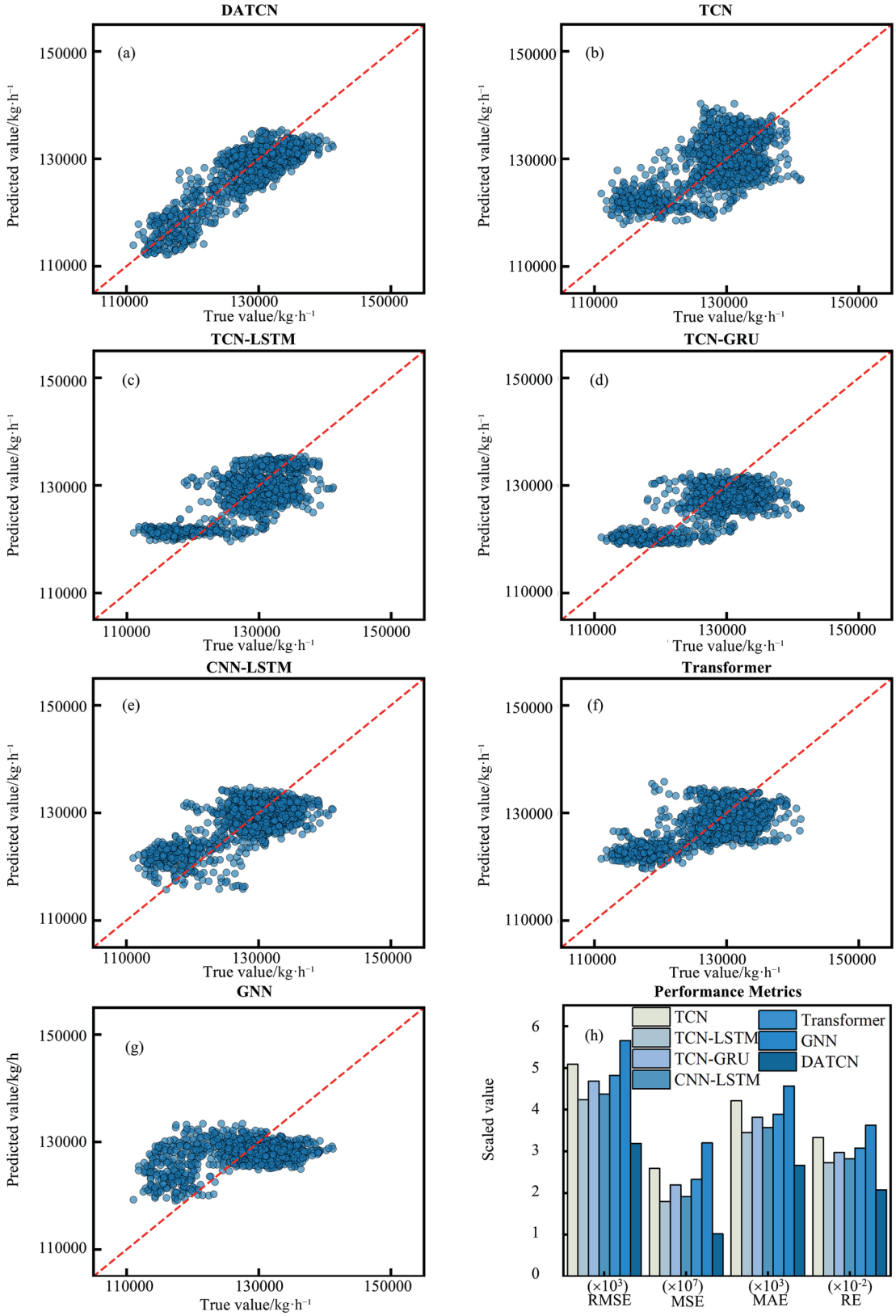


Fig. 16. Prediction scatter plots and metrics for gasoline flow rate: (a)-(g) Scatter plots of DATCN and benchmark models; (h) Performance metrics.

4. Conclusions

To address the strong nonlinearity, multiscale characteristics, and non-stationarity of petrochemical process data, this paper proposes the DATCN prediction framework, which integrates mode decomposition, attention mechanisms, and TCN in a synergistic manner. VMD decomposes raw signals into modes at multiple scales. The attention mechanism adaptively recalibrates these features to emphasize key dynamics while suppressing noise. TCN captures long-term dependencies through causal dilated convolutions, enabling robust prediction of complex process dynamics.

Comparative experiments on the TE process and real industrial FCC unit data show that DATCN consistently outperforms several benchmark models in both prediction accuracy and stability across various process variables, demonstrating its robustness and practical applicability in industrial processes.

Despite its strong performance, deploying DATCN faces challenges related to computational complexity and sensitivity to changes in operating conditions. Future research will focus on two main directions: (1) developing lightweight feature decomposition and parameter adaptation techniques for online deployment, and (2) incorporating domain knowledge through physics-informed learning to improve interpretability and maintain predictive consistency under prolonged and complex operating conditions.

CRediT Authorship Contribution Statement

Mengjie Luo: Writing – original draft, Visualization, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. Mengxuan Zhang: Methodology. Xiaogang Shi: Writing – review & editing, Supervision, Resources. Xingying Lan: Writing – review & editing, Validation, Supervision, Resources, Project administration, Methodology, Funding acquisition, Conceptualization. Yongdong Liu: Validation, Resources. Wei Li: Validation, Resources.

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.

Acknowledgements

The present work is financially supported by the National Natural Science Foundation of China (22021004).

References

[1] Y.J. Wang, C. Qian, S.J. Qin, Attention-mechanism based DiPLS-LSTM and its application in industrial process time series big data prediction, *Comput. Chem. Eng.* 176 (2023) 108296.
[2] Y.F. Chen, A.L. Li, X.Y. Li, D. Xue, J. Long, Efficient JITL framework for nonlinear industrial chemical engineering soft sensing based on adaptive multi-branch variable scale integrated convolutional neural networks, *Adv. Eng. Inform.* 58 (2023) 102199.
[3] S.B. Yang, J. Moreira, Z.K. Li, Bioinspired encoder–decoder recurrent neural network with attention for hydroprocessing unit modeling, *Ind. Eng. Chem. Res.* 62 (44) (2023) 18526–18540.
[4] J.M. Lu, Y.F. Wang, Z.Y. Zhou, J.X. Wu, Y.H. Zhang, L.Z. Zhang, Q. Shi, S.B. He, C. M. Xu, Molecular transformation of heavy oil during slurry phase hydrocracking process: influences of operational conditions, *Pet. Sci.* 22 (2) (2025) 884–893.
[5] Z. Li, Y.H. Ying, M.L. Yang, L. Zhao, L. Zhao, W.L. Du, Monitoring and path optimization of catalytic reformer in a refinery: principal component analysis and A* algorithm application, *Expert Syst. Appl.* 209 (2022) 118358.
[6] M.X. Zhang, Z. Yang, Y.P. Zhao, M.Z. Lv, X.Y. Lan, X.G. Shi, J.S. Gao, C.K. Li, Z. Yuan, Y. Lin, A hybrid safety monitoring framework for industrial FCC disengager coking rate based on FPM, CFD, and ML, *Process. Saf. Environ. Prot.* 175 (2023) 17–33.
[7] T. Dias, R. Oliveira, P. Saraiva, M.S. Reis, Predictive analytics in the petrochemical industry: research Octane Number (RON) forecasting and analysis in an industrial catalytic reforming unit, *Comput. Chem. Eng.* 139 (2020) 106912.
[8] C.W. Tang, J.L. Huang, M. Xu, X. Liu, F. Yang, W.T. Feng, Z.N. He, J.C. Lv, Attention-based early warning framework for abnormal operating conditions in fluid catalytic cracking units, *Appl. Soft Comput.* 153 (2024) 111275.
[9] J. Abdi, G. Mazloom, F. Hadavimoghaddam, A. Hemmati-Sarapardeh, S.H. Esmaeili-Faraj, A. Bolhasani, S. Karamian, S. Hosseini, Estimation of the flow rate of pyrolysis gasoline, ethylene, and propylene in an industrial olefin plant using machine learning approaches, *Sci. Rep.* 13 (2023) 14081.
[10] C.S. He, C.W. Zhang, T.F. Bian, K.X. Jiao, W.K. Su, K.J. Wu, A. Su, A review on artificial intelligence enabled design, synthesis, and process optimization of chemical products for industry 4.0, *Processes* 11 (2) (2023) 330.
[11] M. Misra, H.H. Yue, S.J. Qin, C. Ling, Multivariate process monitoring and fault diagnosis by multi-scale PCA, *Comput. Chem. Eng.* 26 (9) (2002) 1281–1293.
[12] J.H. Cho, J.M. Lee, S. Wook Choi, D. Lee, I.B. Lee, Fault identification for process monitoring using kernel principal component analysis, *Chem. Eng. Sci.* 60 (1) (2005) 279–288.
[13] Q.M. Zhao, K.X. Bi, T. Qiu, Data-driven intelligent modeling framework for the steam cracking process, *Chin. J. Chem. Eng.* 61 (2023) 237–247.
[14] Y.M. Bai, S.Y. Xiang, F.F. Cheng, J.S. Zhao, A dynamic-inner LSTM prediction method for key alarm variables forecasting in chemical process, *Chin. J. Chem. Eng.* 55 (2023) 266–276.
[15] M.K. Lim, J.S. Yun, K.H. Cho, J.W. Yoon, U.H. Lee, A. Ferreira, A. Mafalda Ribeiro, I.B.R. Nogueira, J. Park, J.K. Kim, K. Kim, Multi-objective optimization of ANN-based vacuum pressure swing adsorption process for ethane and ethylene separation, *J. Ind. Eng. Chem.* 143 (2025) 221–239.
[16] I. Alhajree, G. Zahedi, Z.A. Manan, S.M. Zadeh, Modeling and optimization of an industrial hydrocracker plant, *J. Petrol. Sci. Eng.* 78 (3–4) (2011) 627–636.
[17] W.J. Zhu, X.W. Liu, X. Hou, J.Y. Hu, Z.H. Diao, Application of machine learning to process simulation of n-pentane cracking to produce ethylene and propene, *Chin. J. Chem. Eng.* 28 (7) (2020) 1832–1839.
[18] L. Iapteff, J. Jacques, M. Rolland, B. Celse, Reducing the number of experiments required for modelling the hydrocracking process with Kriging through Bayesian transfer learning, *J. R. Stat. Soc. Ser. C* 70 (5) (2021) 1344–1364.
[19] Y.J. Wang, Y.C. Zhang, Z. Wu, H.G. Li, P.D. Christofides, Operational trend prediction and classification for chemical processes: a novel convolutional neural network method based on symbolic hierarchical clustering, *Chem. Eng. Sci.* 225 (2020) 115796.
[20] Y.M. Han, C.Y. Fan, M. Xu, Z.Q. Geng, Y.H. Zhong, Production capacity analysis and energy saving of complex chemical processes using LSTM based on attention mechanism, *Appl. Therm. Eng.* 160 (2019) 114072.
[21] F. Yang, Y.S. Sang, J.C. Lv, J. Cao, Prediction of gasoline yield in fluid catalytic cracking based on multiple level LSTM, *Chem. Eng. Res. Des.* 185 (2022) 119–129.
[22] T.W. Zhang, F.F. Shen, Z. Li, X. Peng, W.M. Zhong, Synergizing low-carbon planning and operation for sustainable integrated refinery-petrochemical processes under arrival time uncertainty: a large-scale hierarchical energy-efficiency optimization perspective, *Appl. Energy* 377 (2025) 124497.
[23] J. Long, C. Huang, K. Deng, L. Wan, G.H. Hu, F. Zhang, Novel hybrid data-driven modeling integrating variational modal decomposition and dual-stage self-attention model: applied to industrial petrochemical process, *Energy* 304 (2024) 131895.
[24] M.K. Khaldi, M. Al-Dhaifallah, O. Taha, Artificial intelligence perspectives: a systematic literature review on modeling, control, and optimization of fluid catalytic cracking, *Alex. Eng. J.* 80 (2023) 294–314.
[25] G. He, C.L. Zhou, T. Luo, L. Zhou, Y.Y. Dai, Y.G. Dang, X. Ji, Online optimization of fluid catalytic cracking process *via* a hybrid model based on simplified structure-oriented lumping and case-based reasoning, *Ind. Eng. Chem. Res.* 60 (1) (2021) 412–424.
[26] S. Santhakumaran, Y.A.W. Shardt, Data-driven nonlinear system identification and SIR particle filtering for chemical process monitoring and prediction, *IFAC-PapersOnLine* 58 (14) (2024) 434–439.
[27] S.W. Xiong, L. Zhou, Y.Y. Dai, X. Ji, Attention-based long short-term memory fully convolutional network for chemical process fault diagnosis, *Chin. J. Chem. Eng.* 56 (2023) 1–14.
[28] J. Huang, X.Y. Sun, X. Yang, K.X. Peng, Fault detection for chemical processes based on non-stationarity sensitive cointegration analysis, *ISA Trans.* 129 (Pt B) (2022) 321–333.
[29] J. Long, L. Ye, H.F. Peng, Z. Tian, Efficient prediction framework for large-scale nonlinear petrochemical process based on feature selection and temporal-attention LSTM: applied to fluid catalytic cracking, *Chem. Eng. Sci.* 301 (2025) 120733.
[30] S.Z. Liu, Q.L. Liu, S. Ahmed, J.J. Wang, F.Y. Lei, D.F. Zhao, A method for the early prediction of abnormal conditions in chemical processes combined with physical knowledge and the data-driven model, *J. Loss Prev. Process. Ind.* 86 (2023) 105185.

[31] Q. Li, M.X. Zhang, X.G. Shi, X.Y. Lan, X.Q. Guo, Y.L. Guan, An intelligent hybrid feature subset selection and production pattern recognition method for modeling ethylene plant, *J. Anal. Appl. Pyrolysis* 160 (2021) 105352.
[32] N. Liu, C.M. Zhu, M.X. Zhang, X.Y. Lan, A multiscale adaptive framework based on convolutional neural network: application to fluid catalytic cracking product yield prediction, *Pet. Sci.* 21 (4) (2024) 2849–2869.
[33] M.X. Zhang, D.F. Cao, X.Y. Lan, X.G. Shi, J.S. Gao, An ensemble-learning approach to predict the coke yield of commercial FCC unit, *Ind. Eng. Chem. Res.* 61 (24) (2022) 8422–8431.
[34] W.J. Song, V. Mahalec, J. Long, M.L. Yang, F. Qian, Modeling the hydrocracking process with deep neural networks, *Ind. Eng. Chem. Res.* 59 (7) (2020) 3077–3090.
[35] K. Dragomiretskiy, D. Zosso, Variational mode decomposition, *IEEE Trans. Signal Process.* 62 (3) (2014) 531–544.
[36] H.Y. Chen, H. Huang, Y. Zheng, B. Yang, A load forecasting approach for integrated energy systems based on aggregation hybrid modal decomposition and combined model, *Appl. Energy* 375 (2024) 124166.
[37] N.E. Huang, Z. Shen, S.R. Long, M.C. Wu, H.H. Shih, Q.A. Zheng, N.C. Yen, C.C. Tung, H.H. Liu, The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis, *Proc. Math. Phys. Eng. Sci.* 454 (1971) (1998) 903–995.
[38] Z.H. Wu, N.E. Huang, Ensemble empirical mode decomposition: a noise-assisted data analysis method, *Adv. Adapt. Data Anal.* 1 (1) (2009) 1–41.
[39] J.R. Yeh, J.S. Shieh, N.E. Huang, Complementary ensemble empirical mode decomposition: a novel noise enhanced data analysis method, *Adv. Adapt. Data Anal.* 2 (2) (2010) 135–156.
[40] Y.H. Zhang, K.Z. Zhong, X.P. Xie, Y.Z. Huang, S. Han, G.Z. Liu, Z.Y. Chen, VMD-ConvTSMixer: spatiotemporal channel mixing model for non-stationary time series forecasting, *Expert Syst. Appl.* 271 (2025) 126535.
[41] H.Y. Zhang, W.S. Wu, X.Y. Song, Z.X. Chen, Improving density logging resolution by VMD-CEEMDAN-ICWT method and its application in thin layer identification, *Geoenergy Sci. Eng.* 239 (2024) 212993.
[42] Q.H. Wang, L.H. Li, Research on semi-supervised soft sensor modeling method for sulfur recovery unit based on ISSA-VMD-ESN, *Chem. Eng. Sci.* 298 (2024) 120397.
[43] X.L. Li, X.F. Zou, J.F. Cheng, M. Tang, P.F. Hu, FMM-VMD-Transformer: a hybrid deep learning model for predicting natural gas consumption, *Digit. Eng.* 2 (2024) 100005.
[44] S.J. Bai, J.Z. Kolter, V. Koltun, An empirical evaluation of generic convolutional and recurrent networks for sequence modeling, arXiv: 1803.01271 (2018), https://doi.org/10.48550/arXiv.1803.01271.
[45] Q. Li, X.Y. Ren, F. Zhang, L. Gao, B. Hao, A novel ultra-short-term wind power forecasting method based on TCN and Informer models, *Comput. Electr. Eng.* 120 (2024) 109632.
[46] L.S. Ke, L.L. Fang, J.H. Meng, J.C. Peng, J. Wu, M.Q. Lin, D.I. Stroe, Identification of the aging state of lithium-ion batteries *via* temporal convolution network and self-attention mechanism, *J. Energy Storage* 84 (2024) 110999.
[47] J.J. Zhu, L.C. Su, Y.W. Li, Wind power forecasting based on new hybrid model with TCN residual modification, *Energy AI* 10 (2022) 100199.
[48] S.J. Qian, T. Peng, Z.H. Tao, X. Li, M.S. Nazir, C. Zhang, An evolutionary deep learning model based on XGBoost feature selection and Gaussian data augmentation for AQI prediction, *Process. Saf. Environ. Prot.* 191 (2024) 836–851.
[49] Y.H. Zhao, X.H. Peng, T. Tu, Z. Li, P.Y. Yan, C. Li, WOA-VMD-SCINet: hybrid model for accurate prediction of ultra-short-term Photovoltaic generation power considering seasonal variations, *Energy Rep.* 12 (2024) 3470–3487.
[50] M.K. Fang, F.F. Zhang, Y. Yang, R. Tao, R.F. Xiao, D. Zhu, The influence of optimization algorithm on the signal prediction accuracy of VMD-LSTM for the pumped storage hydropower unit, *J. Energy Storage* 78 (2024) 110187.
[51] K. Zhang, H.N. Ting, Y.M. Choo, Baby cry recognition based on WOA-VMD and an improved Dempster–Shafer evidence theory, *Comput. Methods Progr. Biomed.* 245 (2024) 108043.
[52] N. Saleem, H. Elmannai, S. Bourouis, A. Trigui, Squeeze-and-excitation 3D convolutional attention recurrent network for end-to-end speech emotion recognition, *Appl. Soft Comput.* 161 (2024) 111735.
[53] C.Q. Xu, J.L. Wang, J. Zhang, Forecasting the power consumption of a rotor spinning machine by using an adaptive squeeze and excitation convolutional neural network with imbalanced data, *J. Clean. Prod.* 275 (2020) 122864.
[54] B.L. Li, J. Li, X.S. Wu, H.Q. Tang, Gated recurrent unit and temporal convolutional network with soft thresholding and attention mechanism for tool wear prediction, *Measurement* 240 (2025) 115546.
[55] Y.M. Li, T. Peng, W. Sun, C.L. Ji, Y.H. Wang, Z.H. Tao, C. Zhang, M. Shahzad Nazir, A soft sensor model based on CNN-BiLSTM and IHHO algorithm for Tennessee Eastman process, *Measurement* 218 (2023) 113195.
[56] M. Huang, Y. Zheng, S.Y. Li, Distributed economic model predictive control for an industrial fluid catalytic cracking unit ensuring safe operation, *Control Eng. Pract.* 126 (2022) 105263.
[57] Q. Yang, A.S. Berrouk, Y.P. Du, H. Zhao, C.H. Yang, M.A. Rakib, A. Mohamed, A. Taher, CFD investigation of hydrodynamics, heat transfer and cracking reactions in a large-scale fluidized catalytic cracking riser, *Appl. Math. Model.* 40 (21–22) (2016) 9378–9397.
[58] I. Al Aslani, G. Lezcano, J.M. Colom, A. Alahmadi, T. Shoinkhorova, A. Dikhtiarenko, M.M. Cui, L. Alfilfil, I. Morales Osorio, K. Almajnouni, P. Castaño, J. Gascon, Reaction network and kinetic modeling for the direct catalytic cracking of Arabian light crude oil to chemicals, *Chem. Eng. J.* 498 (2024) 154981.
[59] Y.M. John, R. Patel, I.M. Mujtaba, Modelling and simulation of an industrial riser in fluid catalytic cracking process, *Comput. Chem. Eng.* 106 (2017) 730–743.
[60] A. Gupta, D. Subba Rao, Effect of feed atomization on FCC performance: simulation of entire unit, *Chem. Eng. Sci.* 58 (20) (2003) 4567–4579.
[61] J. Chang, F.D. Meng, L.Y. Wang, K. Zhang, H.G. Chen, Y.P. Yang, CFD investigation of hydrodynamics, heat transfer and cracking reaction in a heavy oil riser with bottom airlift loop mixer, *Chem. Eng. Sci.* 78 (2012) 128–143.