Systems & Control Transactions

Research Article - Peer Reviewed Conference Proceeding
ESCAPE 35 - European Symposium on Computer Aided Process Engineering
Ghent, Belgium. 6-9 July 2025
Jan F.M. Van Impe, Grégoire Léonard, Satyajeet S. Bhonsale,
Monika E. Polańska, Filip Logist (Eds.)

PSE PRESS

Integration of Yield Gradient Information in Numerical Modeling of the Fluid Catalytic Cracking Process

Wenle Xu[a,b], Baohua Chen[a,c], and Tong Qiu[a,b]*

[a] Tsinghua University, Department of Chemical Engineering, Beijing, China
[b] Tsinghua University, State Key Laboratory of Chemical Engineering, Beijing, China
[c] PetroChina Guangxi Petrochemical Company, Qinzhou, Guangxi, China
* Corresponding Author: qiutong@tsinghua.edu.cn.

ABSTRACT

Fluid catalytic cracking is a crucial process in the refining industry, capable of converting lower-quality feedstocks into higher-value products. Due to the variability in feedstock properties and fluctuations in product market prices, timely adjustment and optimization of the FCC unit are essential. In this context, data-driven models have garnered increasing attention for their capacity to handle the complex, nonlinear reactions involved in the FCC process. However, on account of the limited operating range of the plants and the black-box nature of data-driven models, relying solely on these models for optimization may lead to contradictory decisions in optimization processes. To address these challenges, we integrate gradient information of product yields with respect to key variables derived from the mechanistic model Petro-SIM, into the training process of data-driven models. To mitigate the high computational demands of the Petro-SIM model, we propose the use of active learning methods for efficient sampling and thereby constructing a surrogate model. The results demonstrate that the active learning approach reduces the required sampling size by 25%. More importantly, the data-driven model trained with gradient information improves the accuracy of trend direction prediction by 34.6%, significantly enhancing its effectiveness in supporting the optimization process. The code will be available at https://github.com/xwl514/fcc-hybrid-loss.

Keywords: Fluid Catalytic Cracking, Machine Learning, Data-Driven Model, Active Learning, Gradient Information

1. INTRODUCTION

Fluid catalytic cracking (FCC) is a crucial process in petroleum refining, converting heavy oils into lighter fuels, such as gasoline and diesel. This process enhances the utilization efficiency of crude oil, reduces resource waste, and meets market demand for light fuels, making it indispensable in both petroleum refining and the chemical industry. Fluctuations in upstream processes and market prices of FCC products necessitate timely adjustments and optimization of the catalytic cracking process conditions to maximize efficiency and profitability. Achieving such optimization depends on accurate modeling of the process, which enables the prediction of key variables directly linked to profitability in response to changes in operating conditions [1].

Modeling of the FCC process can be categorized into mechanism-based and data-driven approaches. Mechanistic models are founded on fundamental physical laws, such as mass conservation, energy conservation, and chemical reaction kinetics. However, due to the complexity of feedstock composition and reaction pathways in FCC, establishing a strict model is challenging. One common simplification is the lumped kinetic model, which consolidates numerous individual compounds into several virtual components based on their similar properties. Reaction pathways are then established for these virtual lumps, and the rate constants are calibrated using plant data. The earliest application in the FCC domain is a three-lump model, which predicts the conversion rate and product distribution under varying reaction conditions, thereby enabling the selection of an optimal operating mode based on production demands [2]. Subsequently, more detailed lumped models have been

developed, offering a more refined description of feedstock and product compositions [3, 4]. However, these traditional lumped models do not account for the molecular structure of the components and are not based on actual reaction mechanisms. In contrast to traditional lumping methods, the structure-oriented lumping (SOL) approach decomposes complex hydrocarbon molecules into molecular fragments or structural groups [5]. This method facilitates the modeling of reactant and product compositions at the molecular level and more closely aligns with the true reaction mechanisms of catalytic cracking, offering improved extrapolation capabilities. However, mechanistic models often require simplifications, which can lead to a loss of accuracy. Furthermore, the development and calibration of these models involve large-scale parameter estimation, a process that is time-consuming and demands high-quality data.

With the advancement of machine learning technologies, data-driven modeling methods have increasingly been applied to the modeling of the FCC process. Machine learning methods, especially deep learning techniques, possess strong nonlinear modeling capabilities and can demonstrate higher accuracy in FCC modeling. However, due to the typically limited operating range, the extrapolation capability of these models is constrained. Although several hybrid modeling approaches have been proposed, these methods often focus primarily on the interpretability and extrapolation of the predicted yield values [6, 7]. However, they do not directly account for the gradient information of product yield with respect to operational variables. Factors such as noise in the factory data can cause the model to unintentionally learn distorted gradient information, potentially leading to incorrect optimization directions.

To address these issues, this study aims to directly integrate the gradient information into the data-driven modeling process, thereby enhancing the model's effectiveness in optimizing the FCC process. The primary innovations of this study are as follows:

1. Active learning for mechanistic sampling: employ active learning methods to sample data from the mechanistic model, Petro-SIM. This approach significantly reduces the time and computational costs associated with data sampling.
2. Gradient calculation via surrogate model: utilize samples from the mechanistic model to construct a surrogate model. The gradient information of yield variables with respect to

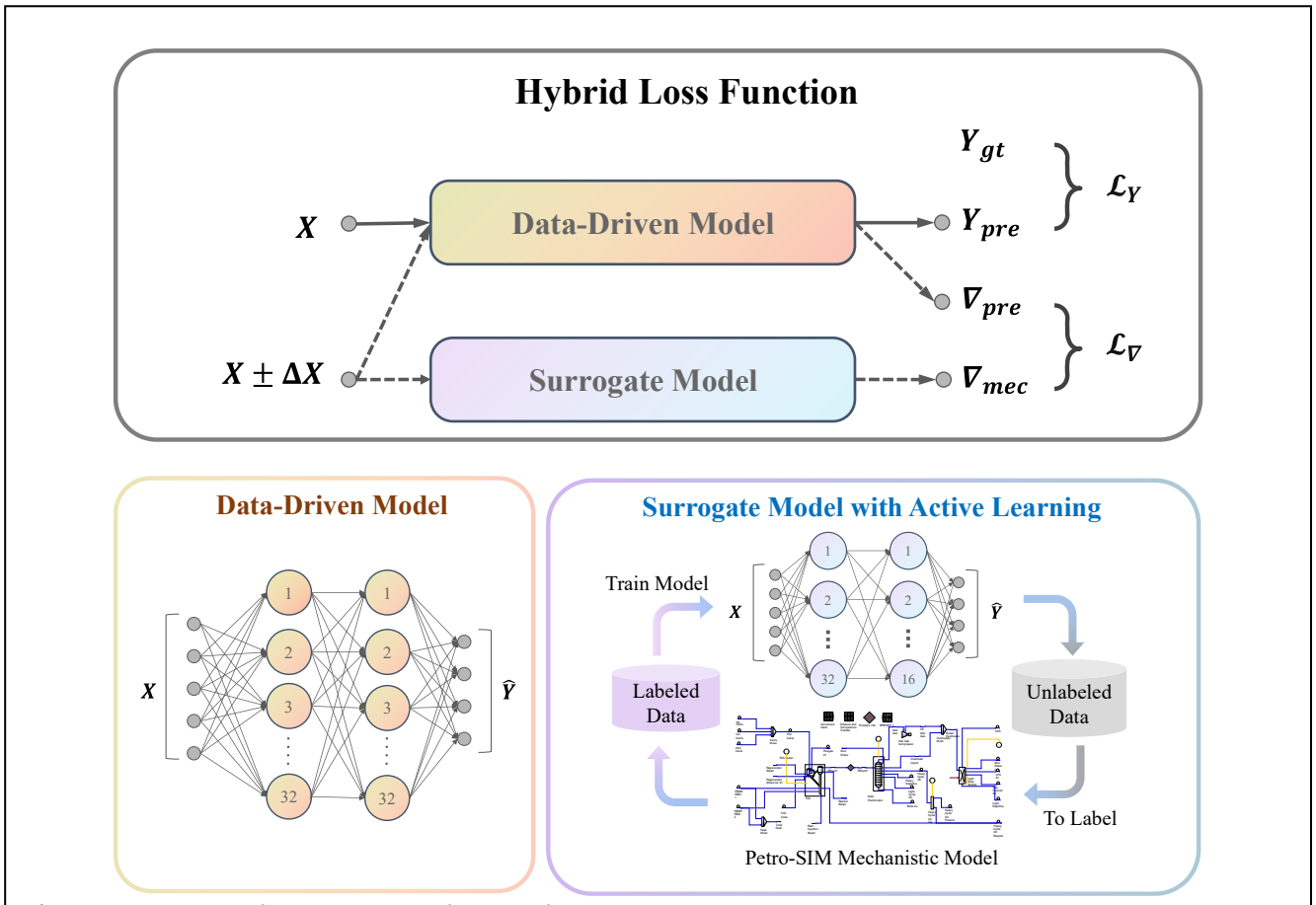

Figure 1: Overview of our hybrid loss function framework.

operational variables at plant sample points is computed using the central difference method.

3. Gradient loss function: introduce a novel product yield gradient loss function tailored for the FCC process. This function enables the training of data-driven models and enhances their reliability in optimization tasks.

2. METHODS

An overview of our method is shown in Figure 1. First, we apply active learning techniques to sample data from the mechanistic model and use the labeled samples to construct a surrogate model. Next, we develop a data-driven model, and during its training, we not only minimize the deviation between the model's output and the true values, but also minimize the gradient differences between the data-driven model and the surrogate model. A detailed explanation of these components is provided in the following sections.

2.1 Surrogate Model with Active Learning

In our hybrid loss function framework, we incorporate the gradient information of target variables with respect to operating variables. However, conducting response tests on operating variables over a wide range during plant operation is impractical, making it challenging to obtain accurate gradient information directly from factory data. To address this limitation, we leverage the mechanistic model, which is based on real reaction kinetics and provides reliable gradient information that reflects actual reaction kinetics.

However, computing the gradient information for each data point using the mechanistic model during every training epoch is computationally prohibitive, especially for the large plant data. When the training set consists of 100,000 points, training a single epoch could take hundreds of hours, making such a workload impractical. To overcome this challenge, we adopt a two-step strategy. First, we utilize active learning methods to sample data from the mechanistic model. By selectively sampling the most informative data points, we significantly reduce the number of required samples. Then, we construct a surrogate model to calculate the gradient information. This integrated approach ensures computational feasibility while maintaining high accuracy in gradient estimation.

We select the key operational variables as input variables and those that significantly influence economic performance as output variables, as shown in Table 1. These variables serve as inputs and outputs for the mechanistic model, surrogate model, and data-driven model.

Table 1: Key process input and output variables.

Variable	Description	Unit
x_1	Feed flowrate	℃
x_2	Feed preheat temperature	℃
x_3	Reflux flowrate	℃
x_4	Reactor temperature	t/h
x_5	Regenerator temperature	t/h
y_1	Dry gas yield	wt%
y_2	Liquefied gas yield	wt%
y_3	Diesel yield	wt%
y_4	Gasoline yield	wt%

2.1.1 Mechanistic Model

The mechanistic model we use is the FCC-SIM model from Petro-SIM (KBC Advanced Technologies Ltd., Walton-on-Thames, UK), which is widely used for the design and optimization of FCC units [7]. We develop an FCC model and calibrate it using factory data to ensure better alignment with the plant's operational conditions. The mechanistic model simulates the FCC process based on input variables and outputs the yields of key products, such as gasoline and diesel, which are the primary variables of interest.

2.1.2 Active Learning

Building a surrogate model requires sampling from the mechanistic model. To reduce the time consumption associated with this sampling, we propose using active learning methods. By selecting the most valuable data for labeling based on criteria such as uncertainty or diversity, this approach minimizes the labeling effort, thereby improving both learning efficiency and model performance. The active learning sampling framework is illustrated in Figure 2.

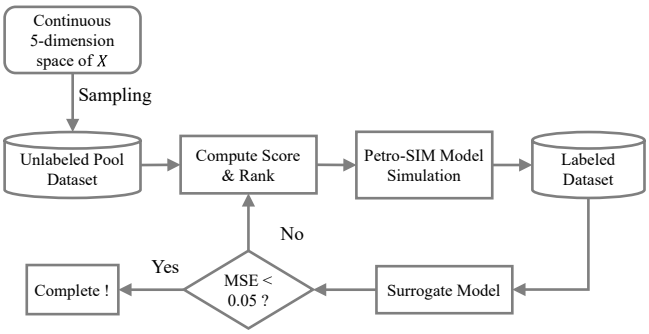

Figure 2: Framework of our active learning method.

Since the space of the operating variables is a continuous 5-dimensional space, we begin by performing sampling to obtain a discrete set of sample points. We randomly select 10,000 data points for simulation calculations, which serve as the validation set to verify the accuracy of the surrogate model. Initially, 3,000 samples

are randomly chosen to train the initial surrogate model. In each active learning iteration, we randomly select 10,000 samples from the unlabeled pool, evaluate them using our predefined active learning metric, and rank them according to their scores. The two active learning metrics are described below. The top 300 samples with the highest scores are then selected as inputs for the mechanistic model simulation. Once the outputs from the mechanistic model are obtained, we update the surrogate model with these new samples and decide whether to terminate the process based on the error on the validation set.

We design two active learning metrics, one based on the diversity of X and the other based on the prediction' s uncertainty. For the active learning method based on the diversity of X, we query the nearest k labeled samples to a selected unlabeled sample and take the average of their distances as the score for the unlabeled sample. This method identifies unexplored regions by evaluating the distance between unlabeled data points and labeled samples, thereby promoting more effective exploration of the variable space. It helps minimize redundant local exploration and ensures that the model samples more diverse areas of the operational variable space. The pseudocode for the process is as follows:

```
Strategy - X diversity:
Randomly select 10000 samples from the Unlabeled_Pool
for each sample in selected_samples do
   Compute distances between the sample and all samples in Labeled_Pool
   Sort the distances and select the smallest k distances
   Diversity_score ← average of the k smallest distances
End for
```

For the uncertainty metric, the variance of the model's predictions for an unlabeled point is used to quantify its uncertainty. We perform T inference runs on a selected unlabeled sample using the surrogate model, incorporating Monte Carlo (MC) dropout during the inference process [8]. While dropout is typically disabled during standard inference, the key idea behind MC dropout is to enable dropout during the inference phase as well. This approach allows for the estimation of prediction variance through multiple forward passes, facilitating the assessment of uncertainty. High uncertainty for an unlabeled point suggests that the model has not adequately explored the surrounding space. Such data points are precisely the ones that need to be labeled. The pseudocode for the calculation process is as follows:

```
Strategy - Uncertainty:
Randomly select 10000 samples from Unlabeled_Pool
for each sample in selected_samples do
   Perform T times inference with dropout enabled
   Uncertainty_score ← variance of the T predictions
End for
```

2.1.3 Surrogate Model

The surrogate model is based on a multi-layer perceptron (MLP) with two hidden layers. The model is trained using mean squared error (MSE) as the loss function, with a learning rate of 0.001. Each active learning iteration involves training the model for 100 epochs.

2.2 Hybrid Loss Function

For data-driven models used in FCC, the loss function typically includes mean squared error, mean absolute error, or other similar loss functions that quantify the deviation between predicted outputs and true values. However, for optimization tasks, solely considering the deviation between predicted values and true values is insufficient. Due to factors such as noise in factory data and overfitting, focusing only on errors in the yield variable during training can result in inaccurate gradients during the FCC optimization process. Therefore, it is crucial to integrate gradient information into the training process.

Consequently, we design a hybrid loss function to train the data-driven model, which not only considers the deviation between predicted values and true values of target variables but also accounts for the discrepancy in gradient directions between the data-driven model and the mechanistic model. The hybrid loss function is as follows:

$$\mathcal{L}_{Hybrid} = w_1\mathcal{L}_Y + w_2\mathcal{L}_\nabla \tag{1}$$

Where $\mathcal{L}_Y$ represents the deviation between the predicted values and the true values of the target variables, and $\mathcal{L}_\nabla$ is the loss function for the gradient component. w_1 and w_2 are the weights assigned to the two components.

2.2.1 Gradient Loss Function

We propose a gradient loss function to quantify the discrepancy between the data-driven model and the mechanistic model. In each epoch, for each training sample, we use the central difference method to compute the gradient of the yield with respect to the operational variables under the given conditions. The calculation formula is as follows:

$$\boldsymbol{\nabla} = \frac{\text{Model}(\boldsymbol{X}+\varepsilon\boldsymbol{I}) - \text{Model}(\boldsymbol{X}-\varepsilon\boldsymbol{I})}{2\varepsilon} \tag{2}$$

Where $\boldsymbol{I}$ is an $m \times m$ identity matrix, with m referring to the number of operational variables. The above formula is used to compute the gradient of a given training

point for both the data-driven model and the surrogate model of the mechanistic model, denoted as $\boldsymbol{\nabla_{pre}}$ and $\boldsymbol{\nabla_{mec}}$. The loss function for the gradient component is computed as follows:

$$\mathcal{L}_{\nabla} = softsign\left(\boldsymbol{\nabla_{pre}} \cdot \boldsymbol{\nabla_{mec}}\right) \cdot \boldsymbol{\nabla_{mec}} \tag{3}$$

We use a sign function as a criterion to determine the directionality of the two gradients. When the directions of the gradients from the data-driven model align with those of the mechanistic model, the sign function yields zero, resulting in a loss function value of zero. This indicates that no correction is needed for the data-driven model. Conversely, when the gradients of the data-driven model and the mechanistic model point in different directions, the sign function returns 1. The subsequent product term $\boldsymbol{\nabla_{mec}}$ serves as a weight term, which increases the importance of points with large gradients and incorrect gradient directions. We use the hyperbolic tangent function because its gradient during the backpropagation process is smoother. The scaling factor a affects the steepness of the hyperbolic tangent function. The $softsign$ formula is as follows:

$$softsign(x) = 0.5 \times \left(1 - \tanh(\alpha \cdot x)\right) \tag{4}$$

The accuracy of gradient direction determination for yield variables with respect to operational variables (denoted as Acc_{∇}) is also considered in the evaluation of our model, and is defined as follows:

$$Acc_{\nabla} = \frac{\sum_{i=1}^{m}\sum_{j=1}^{n} \mathbb{1}_{(\nabla_{pre}\cdot\nabla_{mec})_{ij}>0}}{m \times n} \times 100\% \tag{5}$$

Where m refers to the number of operational variables and n refers to the number of target variables. It calculates the proportion of gradient directions for all y with respect to all x that are consistent between the data-driven model and the mechanistic model, using the indicator function.

2.2.2 Data-Driven Model

Our hybrid loss function does not impose restrictions on the model type. In our experiments, we select an MLP model. The factory data is divided into training, validation, and test sets in a 0.6:0.2:0.2 ratio. The total size of the training set is 79,496 samples, while the validation and test sets each contain 26,499 samples. During the training process, we use our hybrid loss function.

3. RESULTS AND DISCUSSIONS

3.1 Active Learning Results

The MSE loss on the validation set for each iteration of the two active learning strategies is shown in Figure 3. A random labeling strategy is also included as a baseline. With the same number of labeled samples, the models trained using the active learning approaches show superior performance on the validation set compared to the random labeling strategy. Upon reaching the stopping criterion, the diversity-based and uncertainty-based strategies require only 5,400 and 5,700 labeled samples, respectively, while the random selection method demands 7,200 samples. These findings demonstrate that by leveraging our active learning strategies, it is possible to identify the most informative samples for simulation more efficiently, reducing labeling time by up to 25%.

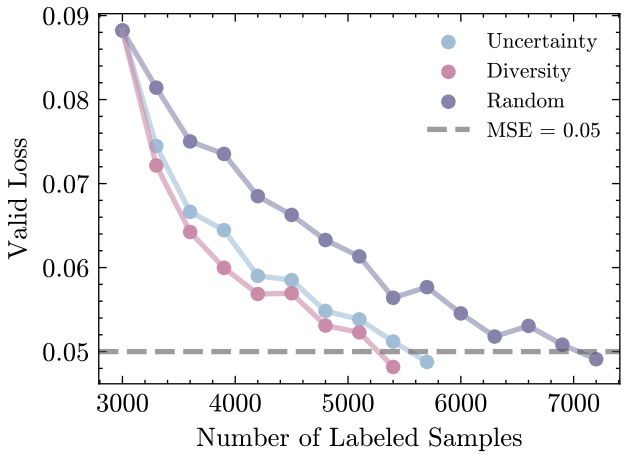

Figure 3: The MSE loss on the validation set for each iteration of active learning.

3.2 Surrogate Model Results

We select the model built using the diversity-based active learning strategy as our surrogate model. Its prediction performance on the validation set is shown in Figure 4. The accuracy of the surrogate model for the mechanistic model is high enough to ensure its reliability for subsequent precise gradient calculations.

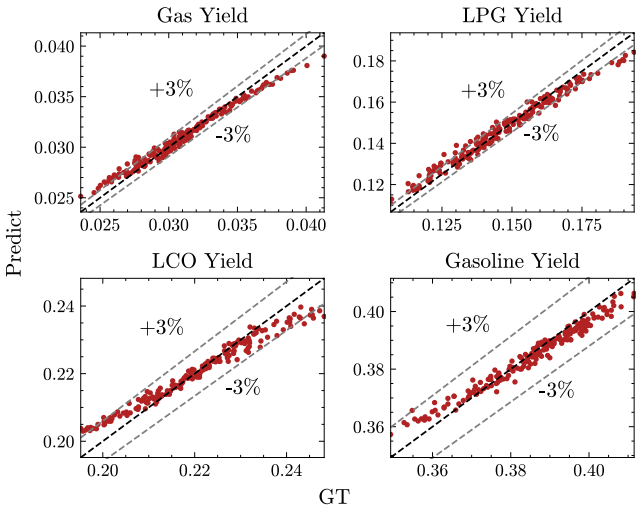

Figure 4: The prediction and Ground Truth (GT) on validation set of the surrogate model.

3.3 Data-Driven Model Results

The $\mathcal{L}_{MSE}$, $\mathcal{L}_{\nabla}$, and Acc_{∇} of the data-driven model trained with the hybrid loss function on the training and validation sets are shown in the Figure 5. During the training process, we first train the model using only the MSE loss. After 50 epochs, we gradually increased the weight of the gradient loss component to w_2.This training strategy ensures a more stable training process. It is observed that when the model is trained solely using the MSE loss function, the loss on both the training and validation sets consistently decreases as training progresses. However, the Acc_{∇} metric shows no improvement during the training processes. When the hybrid loss function is applied, $\mathcal{L}_{\nabla}$ decreases continuously as training progresses, and the gradient direction accuracy on both the training and validation sets shows consistent improvement. Furthermore, as the weight w_2 increases, the Acc_{∇} on the validation set improves correspondingly.

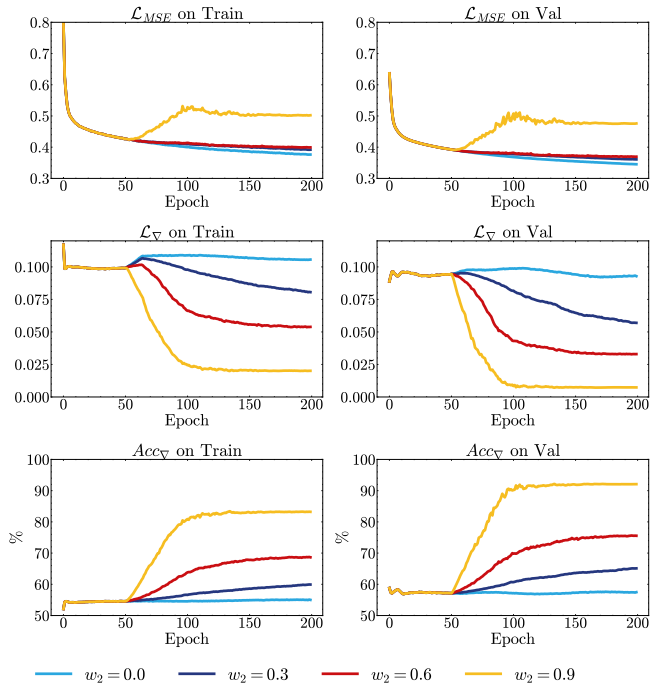

Figure 5: The $\mathcal{L}_{MSE}$, $\mathcal{L}_{\nabla}$, and Acc_{∇} of the training process.

Table 2: The $\mathcal{L}_{MSE}$, $\mathcal{L}_{\nabla}$, and ACC_{∇} on the test set.

w_2	$\mathcal{L}_{MSE}$	$\mathcal{L}_{\nabla}$	Acc_{∇} %
0.0	**0.332**	0.082	57.4
0.3	0.360	0.055	65.6
0.6	0.371	0.037	75.4
0.9	0.480	**0.008**	**92.0**

The $\mathcal{L}_{MSE}$, $\mathcal{L}_{\nabla}$, and ACC_{∇} on test set are shown in the Table 2. The use of the hybrid loss function results in a significant increase on the ACC_{∇} metric. When w_2 is set to 0.9, the gradient direction accuracy reached 92.0%, representing a 34.6% improvement compared to the model trained solely with the MSE loss function.

In addition, we randomly select 5 data points and visualize $\boldsymbol{\nabla_{pre}}$, as shown in Figure 6. Compared to the ∇_{pre} of the model trained using only MSE, the $\boldsymbol{\nabla_{pre}}$ of the model trained with the hybrid loss function aligns better to $\boldsymbol{\nabla_{mec}}$. The alignment with the mechanism model's gradient information enhances the reliability of the model, making it more suitable for optimization tasks.

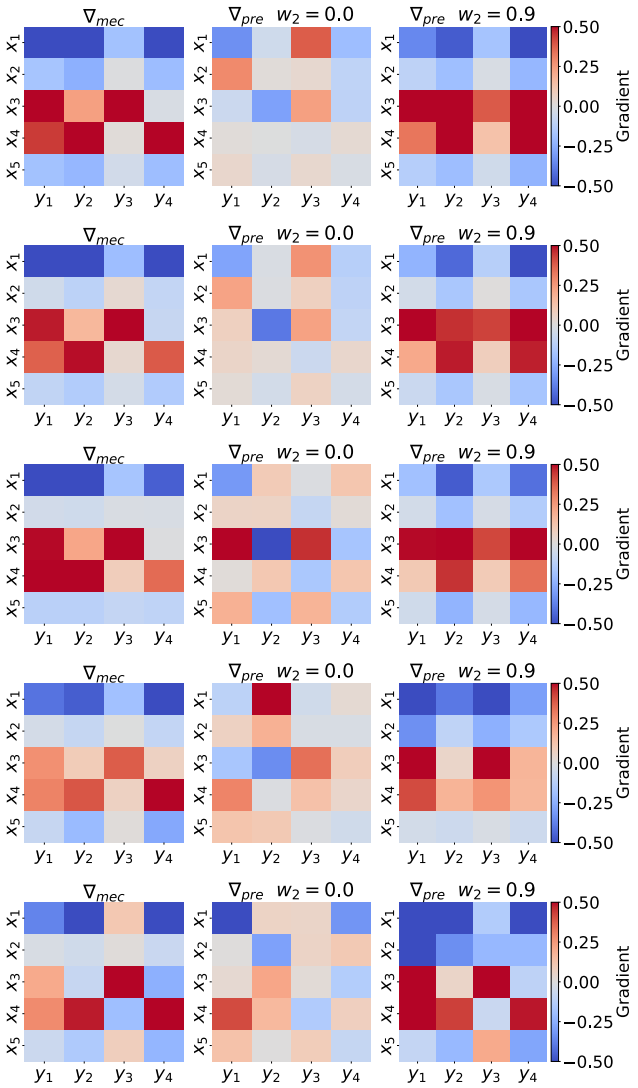

Figure 6: The $\boldsymbol{\nabla_{mec}}$ and $\boldsymbol{\nabla_{pre}}$ of the models with different loss function weights of different samples (each row).

4. Conclusions

We propose an active learning method based on diversity and uncertainty, significantly reducing the time and computational resources required for sampling from mechanistic models. By constructing a surrogate model from the sampled data, we can efficiently compute the gradient information for plant sample points. By incorporating the consistency of gradient information between

the data-driven model and the mechanistic model into the loss function, the accuracy of the gradient directions is improved by 34.6%. This loss function directly incorporates gradient information offers a novel approach to integrating mechanistic and data-driven models. It facilitates a more effective application of machine learning techniques to optimize operations in industrial plants.

REFERENCES

1. Wang, Z., Yang, B., Chen, C., Yuan, J., Wang, L. Modeling and optimization for the secondary reaction of FCC gasoline based on the fuzzy neural network and genetic algorithm. *Chemical Engineering and Processing: Process Intensification*. 46, 175–180 (2007). https://doi.org/10.1016/j.cep.2006.05.011
2. Weekman Jr., V.W., Nace, D.M. Kinetics of catalytic cracking selectivity in fixed, moving, and fluid bed reactors. *AIChE Journal*. 16, 397–404 (1970). https://doi.org/10.1002/aic.690160316
3. Hagelberg, P., Eilos, I., Hiltunen, J., Lipiäinen, K., Niemi, V.M., Aittamaa, J., Krause, A.O.I. Kinetics of catalytic cracking with short contact times. *Applied Catalysis A: General*. 223, 73–84 (2002). https://doi.org/10.1016/S0926-860X(01)00744-X
4. Dasila, P.K., Choudhury, I.R., Singh, S., Rajagopal, S., Chopra, S.J., Saraf, D.N. Simulation of an Industrial Fluid Catalytic Cracking Riser Reactor Using a Novel 10-Lump Kinetic Model and Some Parametric Sensitivity Studies. *Ind. Eng. Chem. Res*. 53, 19660–19670 (2014). https://doi.org/10.1021/ie5006433
5. Ghosh, P., Andrews, A.T., Quann, R.J., Halbert, T.R. Detailed Kinetic Model for the Hydro-desulfurization of FCC Naphtha. *Energy Fuels*. 23, 5743–5759 (2009). https://doi.org/10.1021/ef900632v
6. Long, J., Li, T., Yang, M., Hu, G., Zhong, W. Hybrid Strategy Integrating Variable Selection and a Neural Network for Fluid Catalytic Cracking Modeling. *Ind. Eng. Chem. Res.* 58, 247–258 (2019). https://doi.org/10.1021/acs.iecr.8b04821
7. Li, H., Zhao, Q., Wang, R., Xu, W., Qiu, T. Integrated Hybrid Modelling and Surrogate Model-Based Operation Optimization of Fluid Catalytic Cracking Process. *Processes*. 12, 2474 (2024). https://doi.org/10.3390/pr12112474
8. Gal, Y., Ghahramani, Z. Dropout as a Bayesian Approximation: Representing Model Uncertainty in Deep Learning. *Proceedings of The 33rd International Conference on Machine Learning*. pp. 1050–1059 (2016).