A hybrid deep learning and mechanistic kinetics model for the prediction of fluid catalytic cracking performance

Fan Yang [a], Chaonan Dai [a], Jianquan Tang [a], Jin Xuan [b,*], Jun Cao [c,*]

[a] Data Intelligence Application Lab, Lenovo Group, Chengdu, 610041, China
[b] Department of Chemical Engineering, Loughborough University, Loughborough, UK
[c] School of Mechanical and Power Engineering, East China University of Science and Technology, Shanghai, 200237, China

ARTICLE INFO

ABSTRACT

Fluid catalytic cracking (FCC) is one of the most important processes in the renewable energy as well as petrochemical industries. The prediction and understanding of the FCC performance in a real industrial environment is still challenging, as this is a highly complex process affected by many extremely non-linear and interrelated factors. In this paper, a novel hybrid predictive framework for FCC is developed by integrating a data-driven deep neural network with a physically meaningful lumped kinetic model, powered by orders of magnitude greater number of high-quality data from a modem automated FCC process. The results show that the novel hybrid model exhibits best predictions with regards to all the evaluation criteria such as Mean Absolute Percentage Error, Pearson coefficient, and standard deviation. It indicates that the hybrid data-driven deep learning with mechanistic kinetics model creates a better approach for fast prediction and optimization of complex reaction processes such as FCC.

© 2020 Institution of Chemical Engineers. Published by Elsevier B.V. All rights reserved.

1. Introduction

Fluidic catalytic cracking (FCC) is one of the most important chemicalprocess in the oil and gas industry (Souza et al., 2018; Salvado et al., 2017; Lu et al., 2018; Yang et al., 2017) as well as for renewable energy conversions (Santillan-Jimenez et al., 2019; Lovás et al., 2017). In FCC, heavy oil (hydrocarbon) undergoes a series of chemical reactions, mainly cracking reactions, to produce value-added light oil, gas, and coke in the presence of acidic catalysts, typically under 500 °C and 1–3 bar. In today's petrochemical industry, FCC produces the majority amount of the diesel and gasoline supplied to the market. FCC also plays a vital role in the renewable fuel sectors such as biofuel conversion and upgrading (Santillan-Jimenez et al., 2019) and waste plastics refining to produce fuels and chemicals (Lovás et al., 2017).

FCC is a highly complex process affected by many non-linear and underlying interactions. Numerous factors, such as the quality of feed oil, the catalyst degradation, and the operating conditions, will affect the reaction process and product yields (Shah et al., 2016; Naik et al., 2017). It is still challenging to predict and optimize the FCC process in a real industrial environment. There is a high desire to develop novel predictive framework which could shorten the R&D cycle of new process, optimize unit design & operation, and moreover, realize on-line optimization of the FCC process with the goal of improving raw materials utilization and enhance the yield and quality of the products.

The recent development of artificial intelligence (AI) has created huge opportunities to use machine learning as a powerful predictive tool to effectively solve the highly complex chemical processes, such as FCC, and achieve real-time optimization (Guan et al., 2019; Cai et al., 2018). The machine learning algorithms, typically Deep Neural Networks (DNN) or Artificial Neural Network (ANN) (Dasila et al., 2014; Jiang et al., 2015; Wang et al., 2014) can effectively resolve the relationship between the production process conditions and the products

---

yield and quality. It can also quickly capture the change of conditions through the data during the production process. With the recent progress of process automation and monitoring, the 'big data' needed to train the ANN become more and more available from the operating industrial units. However, the application of ML tool in the field of FCC is still at an early stage and very few literatures can be found. Michalopoulos et al. (2001) and Bollas et al. (2003) developed ANN models for determining the steady-state behavior of industrial FCC units. These excellent early works proved the concept of using ANN to predict the FCC products and optimize the operation conditions, where the hidden relationship between the input process condition and the output product yield were expounded through the ANN. However, the ANN is a data-driven "black box" method. Descriptions of the reaction process in the ANN model is not clear and with poor interpretability. Moreover, this method depends heavily on the quantity and quality of data samples, while these works in the early literature were severely limited by the data availability to effectively train the ANN. Thus, the model could be easily over or under fitted, resulting in poor prediction and generalization ability.

On the other hand, the FCC process can also be modelled by the classic mechanistic models such as correlation model (Xiong et al., 2015a), Lumped Kinetics Model (Chen et al., 2018), and Molecular Dynamic model (Cristina and Popa, 2015), which is a "white box" method based on a clear description of the process principles, and can effectively reflect the reaction phenomena with clear physical meaning and strong traceability (Chang et al., 2014; Ahsan, 2012). Lumped kinetics model (LKM) is the most commonly used method in FCC simulation. It classifies all the components in the reaction system into several virtual components with similar kinetic characteristics. The LKM effectively simplify the reaction system of a complex process, and, therefore, can analyze the product distribution and its governing mechanism, which is of great significance to optimize the operation process. However, it also has clear drawbacks. To describe the highly complex FCC process, mechanistic modeling such as the LKM needs simplified assumptions and neglects some factors selectively. These simplifications will lead to the loss of model accuracy. In addition, the modeling cycle for the physical system is long. Thus, the condition changes in the process can't be tracked in time, which results in the ineffectiveness of these mechanistic model.

To fill the research gaps and overcome shortcomings in both machine learning and classic mechanistic models for FCC process simulation, this study proposed a novel solution to hybridize the data-driven framework with the mechanism-driven model to create a new "gray box" tool that can largely improve the interpretability and traceability of the machine learning. The new hybrid model effectively embeds the physically-meaningful LKM into a data-driven framework, which enabled better data correlation guided by a clear reflection of the process principles. It therefore significantly improves the efficiency and accuracy. Moreover, benefiting from a modern FCC unit with full automation and monitoring, this paper addresses the data sample limitation issue in the previous work. For the first time, "big data" are employed in FCC modelling with two orders of magnitude higher number of high-quality data samples, which enables to develop, for the first time, a DNN (rather than few-layer ANN) to maximise the potential of deep learning as a powerful predictive tool for FCC. The study represents an advantage at the edge of machine learning applied to the prediction and optimization of chemical reaction engineering and processes.

2. Model development

In this research, three different models were derived and subsequently compared for the prediction of FCC performance, which are the purely data-driven machine-learning model based on a deep neural network, the purely physically-meaningful mechanistic model based on an 8-lumped kinetics scheme, and the hybrid data-driven and mechanistic model based on the integration of the above two models. The detailed
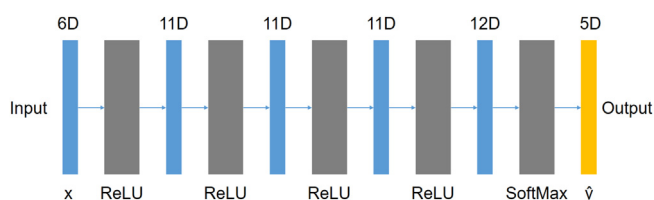

Fig. 1 – Structure of the DNN.

developments of the three models are described in the following sub-sections.

2.1. Development of the machine learning model

2.1.1. Deep neural network model
A standard neural network is developed as the representative of the purely data-driven models. The neural network has 4 layers, which is therefore classified as a deep neural network (DNN). Compared with simple neural network, the essence of DNN is the ability to learn more useful features by more hidden layers and massive training data, to ultimately improve the accuracy of classification or prediction. The structure of the DNN is shown in Fig.1.

The inputs of the DNN include 6 components, i.e., the mass weight percentage of components in the raw oil (saturates, aromatics, resins and asphaltenes), the weight of carbon residue, and the micro activity index. The four hidden layers contains 11, 11, 11, 12 nodes, respectively, with the ReLU (Rectified Linear Unit) (Nair and Hinton, 2010) as the activation function, which is defined as follows,

$$\text{ReLU}(x) = \begin{cases} x, & \text{if } x \geq 0, \\ 0, & \text{if } x < 0. \end{cases} \tag{1}$$

The output layer is a vector constructed by the weight percentages of 5 products: diesel, gasoline, liquified gas, dry gas, and coke. The activation function SoftMax (Goodfellow et al., 2016) maps each element of the vector $\boldsymbol{x}$ to a probability, defined as

$$\text{SoftMax}(\boldsymbol{x}) = \frac{1}{\sum\limits_{k} e^{x_k}} [e^{x_1}, e^{x_2}, \cdots]^{\text{T}}. \tag{2}$$

To summarize, the output of k th hidden layer is denoted as $\boldsymbol{z_k}$ and the corresponding weighted matrices and bias vectors as W_k and $\boldsymbol{b_k}$, respectively. Then, the forward propagation relationship is obtained as follows,

$$\begin{cases} \boldsymbol{z}_1 = ReLU(W_1\boldsymbol{x} + \boldsymbol{b}_1), \\ \boldsymbol{z}_2 = ReLU(W_2\boldsymbol{z}_1 + \boldsymbol{b}_2), \\ \boldsymbol{z}_3 = ReLU(W_3\boldsymbol{z}_2 + \boldsymbol{b}_3), \\ \boldsymbol{z}_4 = ReLU(W_4\boldsymbol{z}_3 + \boldsymbol{b}_4), \\ \hat{\boldsymbol{y}} = SoftMax(W_5\boldsymbol{z}_4 + \boldsymbol{b}_5) \end{cases} \tag{3}$$

The cost function MSE (Mean Square Error) (Goodfellow et al., 2016) is used to measure the difference between the real value and the predicted value, and then, the parameters

are updated in a back-propagation process (Goodfellow et al., 2016). The MSE is defined as follows,

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(x_i - \hat{x}_i)^2 \tag{4}$$

In the formula, the x_i and $\hat{x}_i$ represent real data and predictions respectively, and the n stands for the number of samples.

The DNN is then trained via the Adam (Kingma and Ba, 2015) optimization algorithm, in which the initial learning rate, β_1 and β_2 are set as 0.001, 0.9 and 0.999, respectively. After 1000 iterations, good prediction results of DNN can be obtained.

2.1.2.   Data source and preprocessing

In this research, the raw data is collected from an industrial FCC unit in a petrochemical factory in central China during 2017-2018. More specifically, the operation variables and material properties data analyzed in this study are from the DCS (Distributed Control System) and LIMS (Laboratory Information Management System) installed in the FCC plant. The collected data are arranged in the format of key-value pairs, in which the timestamp and indicator values are saved as the key and value, respectively. An initial database is formed via above arrangement, which is then pretreated based on the demand for modeling.

It is of great importance to obtain high quality data. Therefore, steps for data preprocessing have been developed. Firstly, the original data are made temporal aggregation to 1 sample per minute. Then, outlier detection methods, such as 3-sigma method (Bishop, 2006) and LOF algorithm (Breunig et al., 2000), are implemented on the data, and then the outlier data are set as missing values. Thirdly, B-spline representation (Bishop, 2006) of 1-D curves of degree 2 is used to fill the missing values of each index.

After the data pretreatment process, the whole dataset contains 4090 samples. Due to the unstable distribution of each index, the training and testing sets are randomly split to avoid unbalanced data. In the model, 1000 samples are picked out randomly with equal probabilities to construct the training set. The rest is used as the testing set.

2.2.   Development of 8-lumped kinetics model

The physically-meaningful mechanistic model is developed based on an 8-lumped kinetics scheme, as shown in Fig.2. In this model, the raw materials are lumped into saturated hydrocarbon (HS), aromatic hydrocarbon (HA), resins and asphaltenes (HR), and the products are lumped into gasoline, diesel, liquefied gas, dry gas, and coke, from which an 8-lump reaction network is formed. The following assumptions were made for the complex reaction system of the FCC process, which have been fully justified by previous studies (Xiong et al., 2015b):

(1) All reactions are first-order irreversible.
(2) Plug flow reactor condition is adopted.
(3) All reactions are carried out with the same catalyst activity.
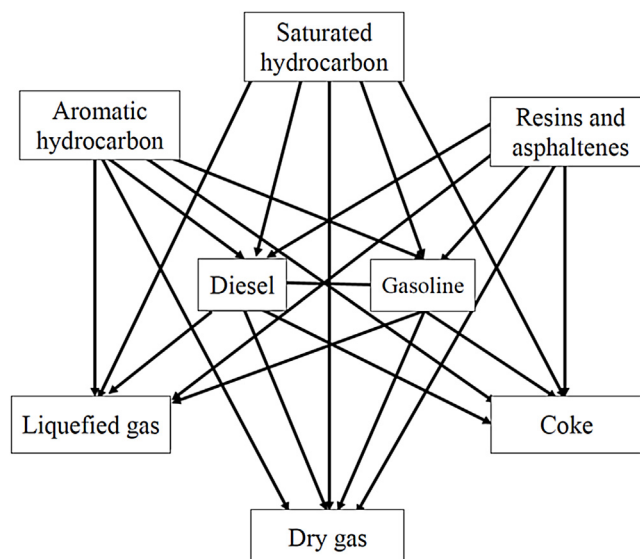(4) The deactivation of the catalyst during the reaction is only related to the residence time of the catalyst.

Fig. 2 – Schematic of the 8-lump kinetics model of FCC.

Considering the transformation relations between the 8 components of the LKM, the simplified state equation was obtained:

$$\frac{d\vec{a}}{d\overline{X}} = K\vec{a} \tag{5}$$

where $\overline{X}$ denotes the non-dimensional distance for the riser reactor, K denotes the reaction rate constant matrix, and the composition vectors of the 8-LKM are as follows,

$$\vec{a} \equiv (a_1, a_2, \ldots, a_8)^T \tag{6}$$

where a_i represents the mass percentage content of the i_{th} lump component. a_1 to a_8 denote for (1) saturates, (2) aromatics, (3) resins and asphaltenes, (4) diesel, (5) petrol, (6) liquefied gas, (7) dry gas and (8) coke, respectively.

It can be seen that the relative non-dimensional distance to the entrance of the reactor $\overline{X}$ is the only argument of the ODE. When $\overline{X} = 0$, it denotes the entrance of the reactor, where there only exists raw material and no products; when $\overline{X} = 1$, it denotes the outlet of the reactor, where the products reach the highest proportions.

For a sample $\vec{a}'$, the initial state can be denoted as $\vec{a}(\overline{X} = 0)$, and the final state as $\vec{a}(\overline{X} = 1)$. Then one sample can be showed as follows,

$$\vec{a}' = (a_1', a_2', \ldots, a_8')^T$$
$$\Rightarrow \begin{cases} \vec{a}\left(\overline{X} = 0\right) = & (a_1', a_2', a_3', 0, 0, 0, 0, 0)^T \\ \vec{a}\left(\overline{X} = 1\right) = & (0, 0, 0, a_4', a_5', a_6', a_7', a_8')^T \end{cases} \tag{7}$$

The final target of the LKM is to get the relationship between the initial state and the final state of all samples. To achieve that, the LKM searches for the optimal matrix K, which is used to represent the transformation rates, and is expressed as follows,

$$K = \begin{bmatrix} -\sum_{i=4}^{8}k_{1i} & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & -\sum_{i=4}^{8}k_{2i} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & -\sum_{i=4}^{8}k_{3i} & 0 & 0 & 0 & 0 & 0 \\ k_{14} & k_{24} & k_{34} & -\sum_{i=5}^{8}k_{4i} & 0 & 0 & 0 & 0 \\ k_{15} & k_{25} & k_{35} & k_{45} & -\sum_{i=6}^{8}k_{5i} & 0 & 0 & 0 \\ k_{16} & k_{26} & k_{36} & k_{46} & k_{56} & 0 & 0 & 0 \\ k_{17} & k_{27} & k_{37} & k_{47} & k_{57} & 0 & 0 & 0 \\ k_{18} & k_{28} & k_{38} & k_{48} & k_{58} & 0 & 0 & 0 \end{bmatrix} \tag{8}$$

For a given matrix K and initial input condition $\vec{a}(\overline{X}=0)$, the output solution $\hat{\vec{a}}(\overline{X}=1)$ of the ODE can be obtained. $\langle J(K) \rangle$ is defined in order to measure the effect of reaction rate constant matrix K as follows:

$$\langle J(K) \rangle \equiv \frac{1}{N}\sum_{k=1}^{N}J(\vec{a}_k', \hat{\vec{a}}_k') \tag{9}$$

where N is the number of samples in the training set, $J(\vec{a}_k', \hat{\vec{a}}_k')$ is a L2 norm, which is used to calculate the distance (difference) between the real value $\vec{a}_k'(\overline{X}=1)$ and model predicted value $\hat{\vec{a}}_k'(\overline{X}=1)$. It can be expressed as:

$$J(\vec{a}_k', \hat{\vec{a}}_k') = ||\vec{a}_k'(\overline{X}=1) - \hat{\vec{a}}_k'(\overline{X}=1)||_2 \tag{10}$$

The crucial step for the LKM is to search for the optimal K values. For this purpose, two Heuristic algorithms, i.e., Simulated Annealing (SA) and Particle Awarm Optimization (PSO), are used. The optimization target is to minimize the difference between the real and predicted values in the training set, i.e., to get $\min_{K} \langle J(K) \rangle$.

The SA algorithm (Kirkpatrick et al., 1983) searches for the optimal values by simulating the process of solid annealing. Firstly, the initial reaction rate, initial temperature, temperature change rate and minimum temperature are set as K_i, T, $\triangle T$ and T_{min} respectively. Then, iterations begin under the condition of $T > T_{min}$. In each iteration, temperature T is reduced and K_i is updated to its randomly-selected neighbor K_i' with a probability, which is denoted as p, and the expressions are

shown in Eqs. (11) and (12). Finally, a feasible reaction rate can be obtained.

$$T = T^* \Delta T \tag{11}$$

$$p = \begin{cases} 1, & \text{if } \langle J(K_i') \rangle < \langle J(K_i) \rangle \\ e^{-\dfrac{\langle J(K_i') \rangle - \langle J(K_i) \rangle}{T}}, & \text{if } \langle J(K_i') \rangle \geq \langle J(K_i) \rangle \end{cases} \tag{12}$$

On the other hand, the PSO algorithm (Kennedy and Eberhart, 1995) simulates the foraging behavior of birds. Firstly, M particles are generated as candidate solutions. Each particle has four attributes: reaction rate K_i, forward speed V_i (represents how much the change is to get the new reaction rate), historical optimal fitness F_{i_best} and historical optimal reaction rate K_{i_best}. Here, the fitness corresponding to the current reaction rate K_i is calculated from $\langle J(K_i) \rangle$, and the historical optimal fitness means the minimum value of the fitness in history. Moreover, there are two global attributes: the global historical optimal fitness of all particles F_{g_best} and the global historical optimal reaction rate K_{g_best}. In the iteration process, V_i and K_i for each particle are updated by Eqs. (13) and (14), in which the corresponding fitness, F_{i_best}, K_{i_best} are calculated, and then F_{g_best} and K_{g_best} are updated. After certain times of iterations, the global historical optimal reaction rate is returned as the result.

$$V_{i\_new} = w^*V_{i\_old} + C_1^*rand_1^*\left(K_{i\_best} - K_{i\_old}\right)$$
$$+ C_2^*rand_2^*\left(K_{g\_best} - K_{i\_old}\right) \tag{13}$$

$$K_{i\_new} = K_{i\_old} + V_{i\_new} \tag{14}$$

where $rand_1$ and $rand_2$ are random numbers between 0 and 1. w, C_1, C_2 are constants being set at the beginning of the algorithm.

With the same database described in Section 2.1.2, 100 out of the first 1000 samples are taken to construct the training set for the LKM. By performing SA and PSO algorithms iteratively, the matrix with the least value is chosen as the optimal one. The calculated transformation rates matrix is shown as follows:

$$\sim K = \begin{bmatrix} -31.8589 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & -20.1408 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & -34.0717 & 0 & 0 & 0 & 0 & 0 \\ 18.5006 & 4.6507 & 15.1141 & -0.7016 & 0 & 0 & 0 & 0 \\ 12.5756 & 10.771 & 8.4472 & 0.7016 & -0.4064 & 0 & 0 & 0 \\ 2.91 \times 10^{-10} & 2.5736 & 8.1144 & 2.20 \times 10^{-12} & 0.2498 & 0 & 0 & 0 \\ 0.0214 & 2.1322 & 1.6540 & 1.78 \times 10^{-12} & 3.81 \times 10^{-14} & 0 & 0 & 0 \\ 0.7613 & 0.0134 & 0.7420 & 1.27 \times 10^{-12} & 0.1566 & 0 & 0 & 0 \end{bmatrix} \tag{15}$$

2.3. Development of the hybrid model

A novel hybrid predictive framework for FCC is further developed by integrating the machine learning with the 8-lumped kinetics model. Here, the lump kinetics model predictions are embedded in the data-driven DNN to improve the performance of the machine learning algorithm. The new hybrid model has the same number of hidden layers and hyper
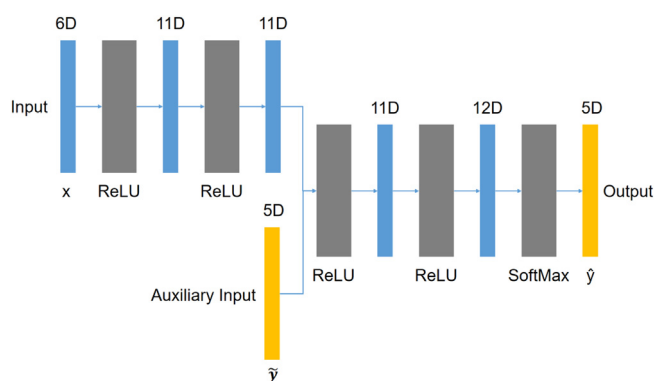
Fig. 3 – Structure of the hybrid model integrating DNN with the lump kinetics model as an auxiliary input.

parameters as the standard DNN developed above. An auxiliary input layer is concatenated with the second hidden layer, where the predictions obtained from the lump kinetics model is considered as the auxiliary input.

As shown in Fig.3, the inputs of the hybrid model are constructed with 2 parts: the main input x and the auxiliary input $\sim y$. The elements of x are the mass weight percentage of the components for the inlet raw oil (saturates, aromatics, resins and asphaltenes), the weight of carbon residue as well as the micro activity index, while the elements of $\sim y$ represents the mass weight percentage of the products (diesel oil, petrol, liquefied gas, dry gas, coke) predicted by the LKM. The output $\hat{y}$ is the final predicted result from the hybrid model, which is a 5-dimentional vector constructed by the weight percentage of 5 products.

To describe this process clearly, the output of the hidden layer is denoted as z_k. The corresponding weighted matrices and bias vectors are denoted as as W_k and b_k, respectively. The forward propagation relationship is obtained as follows,

$$
\begin{cases}
z_1 = & \text{ReLU}\left(W_1 x + b_1\right), \\
\tilde{z}_2 = & \text{ReLU}\left(W_2 z_1 + b_2\right), \\
z_2 = & \left[\tilde{z}_2^T, \tilde{y}^T\right]^T, \\
z_3 = & \text{ReLU}\left(W_3 z_2 + b_3\right), \\
z_4 = & \text{ReLU}\left(W_4 z_3 + b_4\right), \\
\hat{y} = & \text{SoftMax}\left(W_5 z_4 + b_5\right).
\end{cases}
\tag{16}
$$

The hybrid model uses the same cost function (MSE) and Adam optimization algorithm as the standard DNN introduced before. Again, the goal of the algorithm is to minimize the cost function, so as to maximize the accuracy of product yield prediction.

In next section, the calculation results for the LKM, DNN as well as hybrid models will be presented in detail to illustrate the advantage of the hybrid model.

3.   Results and discussion

3.1.   *Comparative analysis on model accuracy*

The three models developed in this study (DNN model, LKM model and hybrid model) are evaluated and compared with regards to four criteria, e.g., Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), Pearson Coefficient (Pear-

son R) and the Ratio of standard deviation (Rstd), they are defined briefly as following:

MAE and MAPE (Bishop, 2006) represent the absolute error between the prediction and the real samples. They reflect the accuracy of the model on a testing set. MAE and MAPE are defined as follows respectively.

$$
MAE = \frac{1}{n}\sum_{i=1}^{n}\left|x_i - \hat{x}_i\right|
\tag{17}
$$

$$
MAPE = \sum_{i=1}^{n}\frac{\left|x_i - \hat{x}_i\right|}{x_i} \times \frac{100}{n}
\tag{18}
$$

The x_i and $\hat{x}_i$ represent real data and predictions respectively, and n stands for the number of samples.

Bishop (2006) measures the linear relationship between variables and illustrates the correlation between predicted and real values. It is defined in Eq. (19), in which $\bar{x}$ stands for the mean value of the real data and $\bar{\hat{x}}$ denotes the average of the predictions. The value of Pearson R lies between -1 and 1. The closer its value is to 1, the more relevant the predicted value is to the real data.

$$
PearsonR = \frac{\displaystyle\sum_{i=1}^{N}(x_i - \bar{x})(\hat{x}_i - \bar{\hat{x}})}{\sqrt{\displaystyle\sum_{i=1}^{N}(x_i - \bar{x})^2}\sqrt{\displaystyle\sum_{i=1}^{N}(\hat{x}_i - \bar{\hat{x}})^2}}
\tag{19}
$$

The last criterion, Rstd (Bishop, 2006) reveals the tendency difference between $\hat{x}$ and x vectors, and is defined as:

$$
Rstd(x, \hat{x}) = \frac{std(x - \hat{x})}{std(x)}
\tag{20}
$$

In real industrial practice, the diesel and petrol are two most important products from FCC. Thus, their yields are selected for comparison among three models in this paper, and the results are shown in Table 1.

It can be seen that the MAPEs for the LKM predictions are much higher for both diesel and petrol yields than the DNN, and the Pearson Rs are much lower, indicating that the LKM exhibits much worse predictability than DNN. The performance of the LKM on the test set is very similar to that on the training set, showing that the LKM is under-fit, due to the assumptions induced in the LKM model.

It can also be found that the performance and prediction characterization of the DNN model and hybrid model are very similar, as they are both based on the machine learning algorithm. However, the hybrid model shows distinct improvements in terms of predictability. Comparatively, the hybrid model could reduce the prediction error by up to 9%, and improve the correlation of by up to 8% compared to the DNN model. It confirms that the auxiliary inputs from the LKM in the hybrid model can improve the performance of the DNN model, as the LKM provides additional physically-meaningful data to train the DNN to learn more useful features of the FCC process which are eventually governed by underlying physical laws.

Furthermore, the Pearson R of the hybrid model on the test set reaches 0.84 in the diesel yield and 0.73 in the petrol yield, and the MAE is less than 0.01 in both diesel and petrol yields.

Table 1 – Performance of LKM, DNN and the hybrid model on the training and testing set.

	Product rate	MAE		MAPE		Rstd		Pearson R	
		Train	Test	Train	Test	Train	Test	Train	Test
LKM	Diesel	0.0139	0.0136	0.0623	0.0615	0.9642	0.9669	0.3892	0.3923
	Petrol	0.0144	0.0146	0.0314	0.0321	1.3604	1.3948	0.1112	0.0902
DNN	Diesel	0.0071	0.0074	0.0313	0.0325	0.5312	0.5665	0.8513	0.8274
	Petrol	0.0068	0.0072	0.0148	0.0159	0.6854	0.7423	0.7672	0.7210
Hybrid model	Diesel	0.0069	0.0073	0.0305	0.0324	0.5086	0.5483	0.8638	0.8405
	Petrol	0.0062	0.0066	0.0135	0.0146	0.6291	0.6883	0.7821	0.7342

It indicates that the model is effective enough to be used in practical industry.

3.2. Comparative analysis on the product yield

To understand the predictability of the three models on the products yield, Fig. 4a (diesel yield) and Fig.4b (petrol yield) presented the diesel and petrol yields, respectively, with the predictions from the LKM, the DNN, the hybrid model, as well as the real industrial data. It is obvious that the results well support the analyses around Table 1 in Section 3.1, and both results from the DNN and hybrid models present better prediction compared with the LKM model, and they exhibit similar tendency with the real data. Moreover, the hybrid model shows better fitting trends than the DNN model. The reason is that the predictive results from LKM model bring more effective information into the hybrid model as auxiliary input, which allows the DNN to capture the variation of the FCC process behaviors governed by underlying physical laws. The conclusions are also verified from the higher Pearson R values for hybrid model both on training and testing sets, as shown in Table 1.
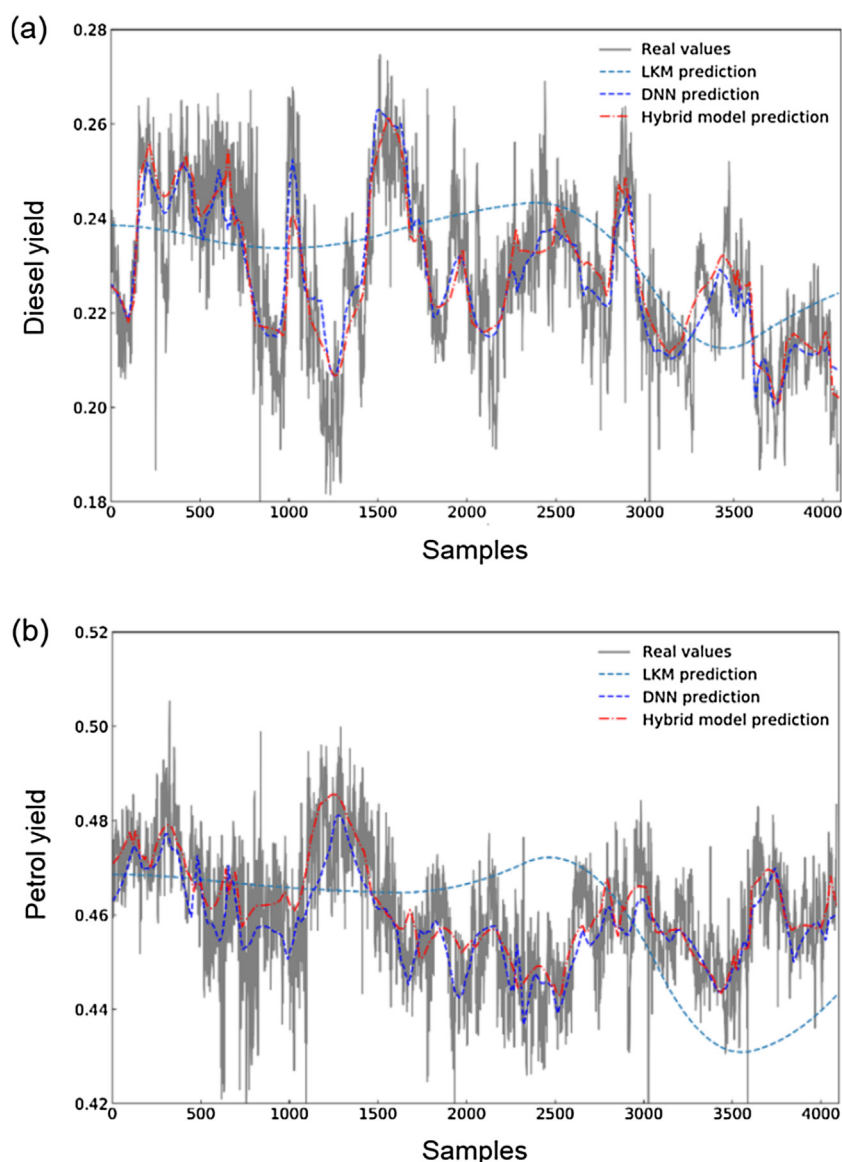

Fig. 4 – The predicted results of DNN, LKM, and the hybrid model on the whole dataset for (a) diesel and (b) petrol yield.
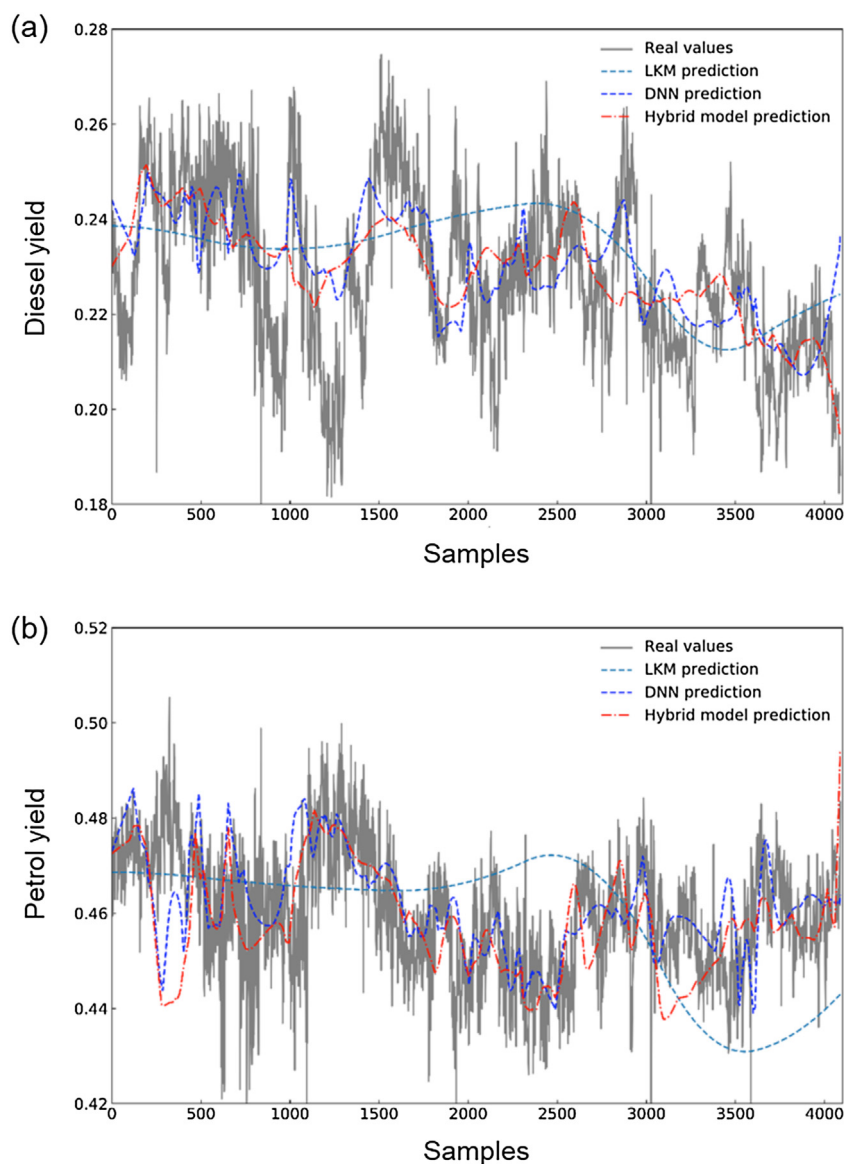
Fig. 5 – The predicted results of DNN, LKM, and the hybrid model with 100 samples' training set: (a) diesel yield; (b) petrol yield.

3.3. The effect of training set sample number on the predictability

The sample number of training set plays an important role in the accuracy of the data-driven model. The training set in previous data-driven models are 1000 samples, however, in common practical situation, the data for training are limited and not always sufficiently available. Therefore, the predictabilities of both DNN and hybrid models with different numbers of training data are studied and compared. Here, the size of the training data set is reduced from previously 1000 to 100 samples (with the sample processing method unchanged). The prediction results of the three models are shown in Fig. 5a (diesel yield) and Fig. 5b (petrol yield). It can be seen that the accuracy of all three models are getting worse for the 100-sample training set compared with the 1000 training samples. However, compared with the DNN model, the hybrid model still maintains a reasonable fitting trends, so the hybrid model can present better predictability even at small sample conditions.

In order to further investigate the effect of training sample number, the Pearson R of the diesel and gasoline yields of

the DNN and hybrid models with different training samples in the range from 100 to 1000 are also compared, which is shown in Fig.6. As expected, a larger training set could improve performances of both two models. More comprehensive features of the samples can be obtained in more training set samples, which leads to more accurate model. It can also be noted that at smaller training sample size, there exists a great difference between the Pearson R values of the petrol yield between two models. The hybrid model outperformed the DNN with doubled Pearson R, which can also attribute to the additional effective physically-meaningful information from LKM model as the in as auxiliary input to the hybrid model. The difference of pearson R between two models becomes smaller with the increase in the training data size, which means the pure DNN model requires more training data to reach comparable performance without the auxiliary inputs from LKM model. From above analysis, it is clear that the hybrid model can get better predictions in a wider range of training sample set numbers, which has a better generalization ability than the DNN model.

Last but not the least, it should be pointed out that the training and testing datasets are split randomly in this study considering that the distribution of each index is not bal-
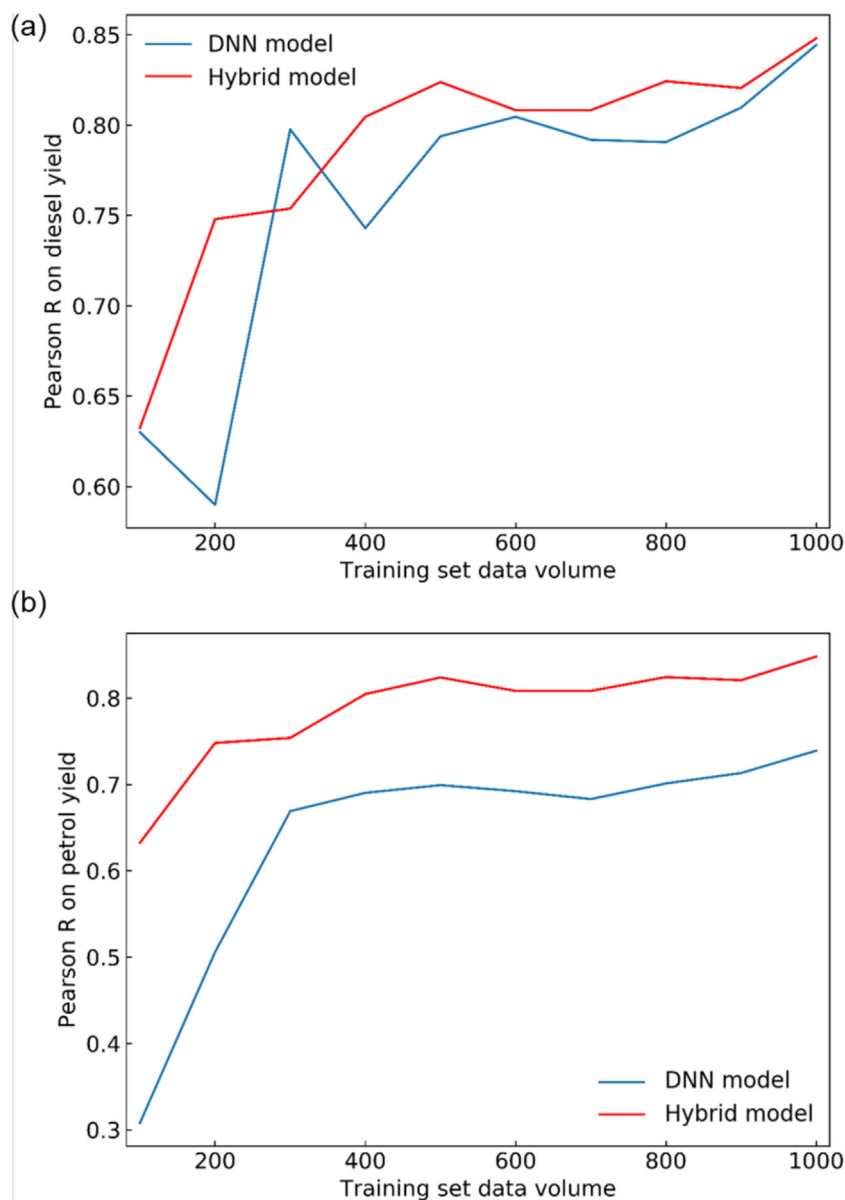
Fig. 6 – The Pearson R on DNN and the hybrid model vs. training set volumes for (a) diesel and (b) petrol yield.

anced. However, unbalanced data have to be dealt with in real scenario. In order to predict the behaviors of unbalanced systems, it is necessary to test the model on the time sequential splitting dataset, which will be the future research direction. To further improve the predictability, the Transfer Learning (Afridi et al., 2018) or Incremental Learning(Matsumori et al., 2018) methods could also be developed with slipping window in the future work.

4. Conclusions

In this research, a novel hybrid model integrating deep neural network with lumped kinetics model is developed to predict the product yields of a FCC unit. The prediction performance of the DNN, LKM as well as the hybrid model are compared. The results show that the hybrid model exhibits obvious improvements in all prediction indicators, e.g., MAE, MAPE, Pearson R and Rstd of products yields. Notably, the hybrid model reduces the prediction error by up to 9%, and improve the correlation of by up to 8% compared to the DNN model, which indicate that the hybrid model has the highest fitting ability and accu-

racy. Moreover, the results also prove that the newly developed hybrid model has advantaged predictability with limited training data availability. The hybrid model outperformed the DNN with doubled Pearson R value when the training size is reduced from 1000 to 100, showing a significantly greater generalization ability.

In summary, the new hybrid model effectively embeds the physically-meaningful LKM into the data-driven DNN, which enabled better data correlation guided by a clear reflection of the process physical principles. It provided a promising approach for process modeling and predictions not only for FCC, but can be widely employed by other complex chemical reaction engineering and processes.

Declaration of interests

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

References

Afridi, M.J., Ross, A., Shapiro, E.M., 2018. On automated source selection for transfer learning in convolutional neural networks. Pattern Recognit. 73, 65–75.

Ahsan, M., 2012. Computational fluid dynamics (CFD) prediction of mass fraction profiles of gas oil and gasoline in fluid catalytic cracking (FCC) riser. Ain Shams Eng. J. 3, 403–409.

Bishop, C.M., 2006. Pattern Recognition and Machine Learning. Springer.

Bollas, G.M., Papadokonstadakis, S., Michalopoulos, J., Arampatzis, G., Lygeros, A., 2003. Using hybrid neural networks in scaling up an FCC model from a pilot plant to an industrial unit. Chem. Eng. Process. 42, 697–713.

Breunig, M.M., Kriegel, H.P., Ng, R.T., 2000. LOF: identifying density-based local outliers. ACM Sigmod Rec. 29, 93–104.

Cai, S., Zhang, L., Hu, J., 2018. Scale-reasoning based risk propagation analysis: an application to fluid catalytic cracking unit. Process Saf. Environ. Prot. 120, 155–165.

Chang, J., Cai, W., Zhang, K., Meng, F., Wang, L., Yang, Y., 2014. Computational investigation of the hydrodynamics, heat transfer and kinetic reaction in an FCC gasoline riser. Chem. Eng. Sci. 111, 170–179.

Chen, Z., Feng, S., Zhang, L., 2018. Molecular-level kinetic modelling of fluid catalytic cracking slurry oil hydrotreating. Chem. Eng. Sci. 195, 619–630.

Cristina, Popa, 2015. Four – lump kinetic model vs. three – lump kinetic model for the fluid catalytic cracking riser reactor. Procedia Eng. 100, 602–608.

Dasila, P.K., Choudhury, I.R., Saraf, D.N., Kagdiyal, V., Rajagopal, S., Chopra, S.J., 2014. Estimation of FCC feed composition from routinely measured lab properties through ANN model. Fuel Process. Technol. 125, 155–162.

Goodfellow, I., Bengio, Y., Courville, A., 2016. Deep Learning. MIT Press.

Guan, H., Ye, L., Shen, F., Song, Z., 2019. Economic operation of a fluid catalytic cracking process using self-optimizing control and reconfiguration. J. Taiwan Inst. Chem. Eng. 96, 104–113.

Jiang, H., Tang, J., Ouyang, F., 2015. A new method for the prediction of the gasoline yield of the MIP process. Pet. Sci. Technol. 33, 1713–1720.

Kennedy, J., Eberhart, R., 1995. Particle swarm optimization (PSO). IEEE International Conference on Neural Networks.

Kingma, D., Ba, J., 2015. Adam: a method for stochastic optimization. In International Conference on Learning Representations (ICLR).

Kirkpatrick, S., Gelatt, C.D., Vecchi, M.P., 1983. Optimization by simulated annealing. Science 220, 606–615.

Lovás, P., Hudec, P., Jambor, B., Hájeková, E., Horňáček, M., 2017. Catalytic cracking of heavy fractions from the pyrolysis of waste HDPE and PP. Fuel 23, 244–252.

Lu, C., Fan, Y., Liu, M., 2018. Advances in key equipment technologies of reaction system in RFCC unit. Acta Petrolei Sinica (Pet. Process. Sect.) 34, 441–454.

Matsumori, S., Abe, Y., Osawa, M., Imai, M., 2018. Investigation of incremental learning as temporal feature extraction. Procedia Comput. Sci. 145, 342–347.

Michalopoulos, J., Papadokonstadakis, S., Arampatzis, G., Lygeros, A., 2001. Modelling of an industrial fluid catalytic cracking unit using neural networks. Chem. Eng. Res. Des. 79, 137–142.

Naik, D.V., Karthik, V., Kumar, V., Prasad, B., Garg, M.O., 2017. Kinetic modeling for catalytic cracking of pyrolysis oils with VGO in a FCC unit. Chem. Eng. Sci. 170, 790–798.

Nair, V., Hinton, G.E., 2010. Rectified linear units improve restricted boltzmann machines. Proceedings of the 27th International Conference on International Conference on Machine Learning, 807–814.

Salvado, F.C., Teixeira-Dias, F., Walley, S.M., Lea, L.J., Cardoso, J.B., 2017. A review on the strain rate dependency of the dynamic viscoplastic response of FCC metals. Prog. Mater. Sci. 88, 186–231.

Santillan-Jimenez, E., Pace, R., Morgan, T., Behnke, C., Crocker, M., 2019. Co-processing of hydrothermal liquefaction algal bio-oil and petroleum feedstock to fuel-like hydrocarbons via fluid catalytic cracking. Fuel Process. Technol. 188, 164–171.

Shah, M.T., Utikar, R.P., Pareek, V.K., Evans, G.M., Joshi, J.B., 2016. Computational fluid dynamic modelling of FCC riser: a review. Chem. Eng. Res. Des. 111, 403–448.

Souza, N.L.A., Tkach, I., Morgado, E., Krambrock, K., 2018. Vanadium poisoning of FCC catalysts: a quantitative analysis of impregnated and real equilibrium catalysts. Appl. Catal. A Gen. 560, 206–214.

Wang, Z., Yang, B., Chen, C., Yuan, J., Wang, L., 2014. Modeling and optimization for the secondary reaction of FCC gasoline based on the fuzzy neural network and genetic algorithm. Chem. Eng. Process. Process Intensif. 46, 175–180.

Xiong, K., Lu, C., Wang, Z., Gao, X., 2015a. Quantitative correlations of cracking performance with physiochemical properties of FCC catalysts by a novel lump kinetic modelling method. Fuel 161, 113–119.

Xiong, K., Lu, C., Wang, Z., Gao, X., 2015b. Quantitative correlations of cracking performance with physiochemical properties of FCC catalysts by a novel lump kinetic modelling method. Fuel 161, 113–119.

Yang, C., Chen, X., Li, C., 2017. Challenges and opportunities of fluid catalytic cracking technology. J. China Univ. Pet. (Edition of Natural Science) 41, 171–177.