

# Neural network model for atmospheric residue desulfurization conversion prediction using real plant data

Yungun Jung\* Hyungjun Kim\* Gyeongwan Jeon\*  
Yeonsoo Kim\*,<sup>1</sup>

\* Department of Chemical Engineering, Kwangwoon University, 20  
Kwangwoon-ro, Nowon-gu, Seoul, 01897, Republic of Korea

---

**Abstract:** Atmospheric residue desulfurization (RDS) is used to treat the invaluable atmospheric residue from crude distillation column before it is sent to the fluid catalytic cracking (FCC) process. RDS is operated by increasing the operation temperature of catalysts to keep the desirable conversion as the catalyst aging progresses. To use the catalyst as long as possible while satisfying the criterion on the impurities in the RDS product, we need to predict the conversion based on the operation histories. In this study, we propose neural network models for predicting removal amount of the impurities in RDS. As the temperature, hydrogen flow rate, and impurities feed flow rates are selected as the features for neural network models. In addition, an approximated aging factor for the catalyst is considered with the cumulative amount of treated impurities and trainable parameters. The neural networks are trained in a moving horizon manner and tested using the real plant data during 123 days. If we train the neural networks in a general way, the trained model shows unrealistic results that are physically and chemically unexpected. When we increase the temperature with the other factors fixed, the catalyst conversion predicted by the trained neural network decreases. This absurd prediction is because the model is trained using the data where the temperature increases when the conversion is low with aging catalyst. To address this issue, we use a clipping method that the weights of neural networks are restricted to positive values. This strategy always ensures the increasing (decreasing) tendencies of the removal amount with respect to the increase (decrease) of the temperature and hydrogen flow rate.

*Keywords:* Neural network, Desulfurization, Conversion Prediction, Atmospheric residue

---

## 1. INTRODUCTION

Atmospheric residue desulfurization (RDS) process is widely used to upgrade heavy petroleum oils and residues to valuable fuels (Hauser et al. (2005); Almutairi et al. (2007); Chehadeh et al. (2018)). RDS takes the high sulfur atmospheric residue (AR) produced from the crude distillation unit (CDU) as the feed, converts it into a more valuable substance, and sends it to the fluid catalytic cracking (FCC) process. In the RDS, substances harmful to the FCC process such as sulfur, nitrogen, and metals are removed. The addition of hydrogen at high temperature and pressure is the main method for this process (Marafi et al. (2006)). The hydrogenation method helps to produce a higher quality product than the heat treatment method, and the resulting oil has lower levels of aromatics, sulfur, nitrogen, and other contaminants (Chehadeh et al. (2020)). In the hydrogenation method, multiple reactors with graded catalyst bed are used for hydrodesulfurization (HDS), hydrodemetalization (HDM), hydrodeconradson-carbon-residue (HD-CCR), and hydrodenitrogenation (HDN). The catalysts substantially improve the removal performance of unde-

sirable impurities. However, the activity of the catalysts is degraded rapidly by coke and metal (Al-Dalama and Stanislaus (2006); Furimsky and Massoth (1999)). The catalysts need to be replaced about once every 6 months, and this consumes a lot of time and money.

To use the catalysts as long as possible while satisfying the requirements on the RDS products, it is necessary to predict the conversion of catalysts depending on the operation duration of the catalysts along with the operation conditions. If we can predict whether the catalysts provide a satisfactory conversion under some operation conditions during the future duration, we can use the catalysts longer with slightly reduced feed flow rate, increased hydrogen flow rate, or optimally designed operation temperature scenario. However, it is not easy to develop mathematical models that can describe the catalytic kinetics and to predict the extent of catalyst aging (Oh et al. (2021)). The experiments for determining the order of reactions and the inhibition terms are time consuming, and the properties should be re-examined each time the supplier changes. In addition, in the case of carbon compounds, the composition of saturated and unsaturated carbon compounds is not separately measured.

---

<sup>1</sup> Corresponding author. Tel.: +82-940-5182; Fax: ; E-mail address: kimy3@kw.ac.kr

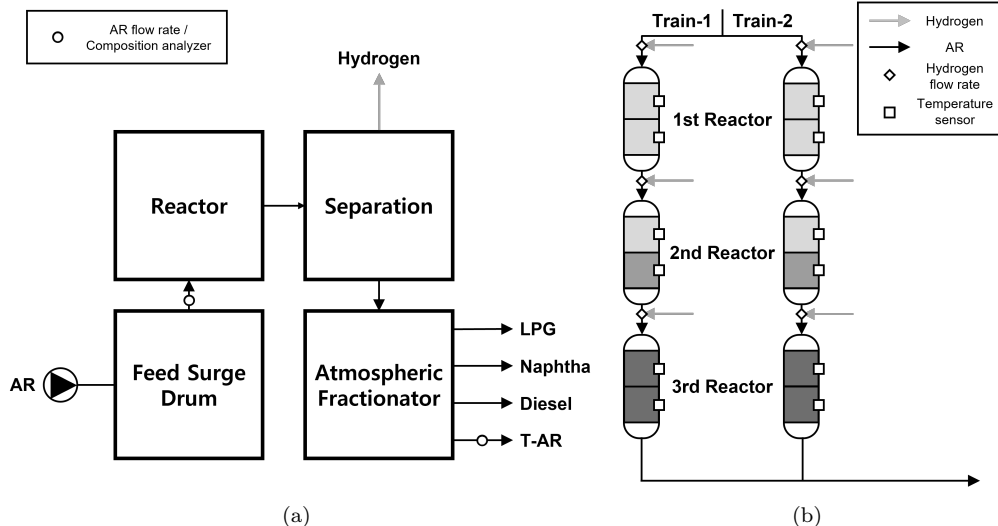


Fig. 1. (a) Simplified block diagram of RDS process. (b) Detailed flow of in the reactor part.

To address the issues, we develop a data-driven neural network model using the real RDS plant data. Several neural networks are trained to predict the removal amount of each impurities because the catalysts for each component are separately located in the reactors. The features for each neural network are selected accordingly with the position of catalysts and based on the key operating conditions.

To train the neural networks, we need the data for their input and output variables. Although temperature, pressure, and flow rate at so many points are measured in real-time and stored in the PI system, the composition is not analyzed in real-time and it is analyzed at limited points. Even with this practical limited situation, we can specify the composition of reactants and products in the reactor section by ignoring the section for just separating the hydrogen and by noticing that the metals are not evaporated much.

Based on the data, the neural networks are trained in receding horizon manner to consider the catalytic degradation effectively. Because the RDS is operated by increasing the temperature as the catalysts are degraded (Al Bazzaz et al. (2015); Marafi et al. (2009)), the typical training of neural network gives contradictory results physically and chemically. That is, the neural networks learn the trends where the conversion is reduced with the high temperature, which is the opposite of the real tendency. We resolve this issue by restricting the weights of the neural networks to be positive and introducing the aging model separately. By restricting the weights as the positive values, the neural networks show the increase of removal amount as the temperature and hydrogen flow rate increase.

## 2. RDS PROCESS

We use the real plant data to construct neural network models for the conversion prediction. The RDS is designed to treat 66,000 barrels per day of AR, and the simplified block diagram is shown in Fig. 1(a). The AR from the CDU enters the feed surge drum to prevent the surging. The outlet of the drum then flows into two parallel trains, each consisting of three reactors. The impurities

are removed through HDM, HDS, HDCCR, and HDN processes. Hydrogen gas is additionally supplied before each reactor and it is used to cool down the reactor properly. Separation part includes Hot High Pressure Separator (HHPS), Hot Low Pressure Separator (HLPS), Cold High Pressure Separator (CHPS), Cold Low Pressure Separator (CLPS). The separation process is to perform gas-liquid separation by the temperature. In detail, the impurities treated in the reactor part are absorbed and discharged, and the hydrogen gas is purified and recycled to the reaction part. Finally, the stream is distilled through atmospheric fractionator into treated atmospheric residue (T-AR), diesel, naphtha, and liquefied petroleum gas (LPG). The T-AR is fed to the FCC for further treatment. Note that the impurities are treated at the reactors; in the separation part,  $H_2S$  is mainly absorbed and  $H_2$  is separated for recycling.

In this study, the objective is to predict the removal amount of impurities from AR to T-AR. However, the composition is analyzed in detail at very few points as shown in Fig. 1(a); at the outlet from the feed surge drum and at the T-AR stream. Most of sulfur (S), nitrogen (N), metals (Ni and V), and Conradson carbon residue (CCR) are removed in the reactors; thus, we assume that the removal amount of each component is calculated using the composition data of the feed surge drum outlet and the T-AR with associated flow rates. The amount of impurities contained in the other product of atmospheric fractionator such as diesel, naphtha, and LPG is negligible. In particular, most of metals are not evaporated and remain in the bottom with CCR. Fortunately, temperature, pressure, and flow rate are measured at many points in real-time and stored. We use the temperature data of upper and lower parts in each reactor and the hydrogen flow rate data before the each reactor.

## 3. PROPOSED PREDICTION NEURAL NETWORK

### 3.1 Feature Selection

To select features related to the output to be predicted is significant especially for the real applications of neural

networks. In the process system engineering, we always consider the mass and heat balances based on the concentrations, temperature, pressure, and flow rate to describe each equipment. That is, we use the input variables directly affecting the equipment along with the state variables. Thus, to construct the neural network models, we need to only consider those variables as the features. If not, the feature that actually does not affect the removal amount can cause a bias or disrupt the training.

The mass flow rates of each component in the feed are selected as features, because the removal amounts increase to some extent as the amount of impurities in the inlet feed increase. The positions of each reaction catalysts are shown in Table 1, which are the same for the trains 1 and 2. The temperatures and the hydrogen flow rates associated for each catalyst are selected as the features. In other words, for each neural network predicting the removal amount of each component, the associated temperature and hydrogen flow rate are selected and summarized in Table 2. In addition, we used  $\exp(-\frac{1}{T})$  with each temperature by domain knowledge that the kinetic rates are usually proportional to the term.

To consider the aging effects of each catalyst, we used an exponential decaying function with trainable weights as Eq. (1). The maximum capacity for each catalyst is set by assuming that the catalysts can be used at the most six months with the initial feed composition and the typical flow rate. When training  $w_1$  and  $w_2$  in each neural network, we restrict them greater than and equal to 0.01. The positive values for them are required to describe the aging effects and the lower bounds for them are set to avoid the conversion to be zero.

$$\text{Aging factor} = w_1(1 - \exp(-(C_{max} - C_{treat}))) + w_2 \quad (1)$$

Here,  $C_{max}$  denotes the expected maximum amount of catalyst and  $C_{treat}$  represents the cumulative treated amount.

Table 1. Embedded position of catalysts for each reaction

		HDS	HDN	HDM	HDCCR
1st reactor	Upper			○	
	Lower			○	
2nd reactor	Upper			○	
	Lower	○		○	○
3rd reactor	Upper	○	○	○	○
	Lower	○	○	○	○

Table 2. Selected features for prediction of each component removal amount

		S	N	Metal	CCR	
Feed flow rate		○	○	○	○	
Hydrogen flow rate	1st reactor			○		
	2nd reactor	○		○	○	
	3rd reactor	○	○	○	○	
Temperature	1st reactor	Upper		○		
		Lower		○		
	2nd reactor	Upper			○	
		Lower	○		○	○
	3rd reactor	Upper	○	○	○	○
		Lower	○	○	○	○

### 3.2 Prediction model structure

The AR stream is divided to each train almost in the same amount, and each feed is treated by the three reactors having the same catalysts embedding. Thus, we introduce two neural networks for the prediction of each component removal amount. In other words, total eight neural networks are trained to predict the removal amount of S, N, metals, and CCR. The simplified structure of the neural network for a component is shown in Fig. 2. Two fully connected neural networks (FCNN) are used to calculate the removal amount in each train. The outputs from the FCNNs are multiplied with each aging factor, and then the summation of them is regarded as the total removal amount from the reaction part for the component. The final value is the model prediction value and this is compared with the real plant data to train the weights in neural networks and aging factors.

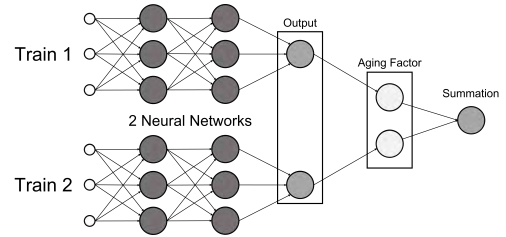


Fig. 2. Prediction model structure of removal amount for a component.

Table 3. Hyperparameters setting

The number of hidden layers	2	
The number of node for each hidden layer	17, 17	
Training epochs	3000	
Learning rate	1st	0.0001
	2nd	0.00005
	3rd	0.00003
	4th	0.00001
Batch size	33	
Optimizer	Adam	
Regularization	L2, 0.0005	

### 3.3 Training strategies

The objective function for training is the mean squared errors for the removal amount of each component. To address overfitting issue, we add L2 regularization term to the objective function. The hyperparameters are shown in Table 3.

To ensure physical and chemical compatibility and to train the neural networks with time series data, we use two strategies. First, we need to ensure that the removal amount increases as the temperature, hydrogen flow rate, and the impurities amount in the feed increase. In addition, the removal amount and the  $\exp(-\frac{1}{T})$  also have the positive correlation. To this end, the weights in the neural networks are restricted to the positive values. If some updated weights are negative after updated in a general way, the weights are clipped to zero. As will be shown in results section, if we do not use the clipping strategy, unreasonable cases happen with the trained neural network. For example, the removal amount is increased

as the temperature increases. This is because the plant is operated by increasing the temperature when the conversion is not satisfied with the lower temperature due to the catalyst aging (Seki and Yoshimoto (2001); Forzatti and Lietti (1999); Chandrasekaran and Sharma (2019)). If we train the neural network without any guidance, the neural network learn the wrong correlation that the removal amount decreases as the temperature increases. To address this issue, it is not sufficient to provide and learn the aging factor separately from the neural network. Similarly, the increase of the hydrogen flow rate and the feed amount affects the removal amount in the direction of increasing. The positive correlations from the features to the removal amount are enforced by the clipping strategy.

Second, we train the FCNNs in a moving horizon manner instead of using recurrent neural network (RNN) or long short-term memory (LSTM). To apply the clipping method, the correlations from the inputs to the outputs can be interpretable. RNN or LSTM includes hidden states storing the past data implicitly, which is difficult to extract any consistent tendency of hidden states to the removal amount. Thus, we use FCNNs in the moving horizon manner to describe the time series effects with keeping the interpretability. We have the operation data of 123 days (Dec. 17, 2020 – Apr. 18, 2021). The training is conducted based on the data of the first 67 days. Then, the trained FCNNs are used to test the data of 14 days after the first 67 days. Because the catalysts keep aged, we need to update the FCNNs periodically using the latest data. After the 14 days, we update the FCNNs using the data of the latest 67 days. Then, the removal amount of another 14 days are tested with this updated FCNNs. The train data and test data are clearly shown in Fig. 3. We have 123 days so that we train FCNNs four times. The learning rate for the first training is set as 0.0001 and the learning rate is reduced for the next pieces of training as shown in Table 3. After the first training, we only need to update the FCNNs slightly to describe the aging effects and time-varying effects and to reflect the past data to some extent.

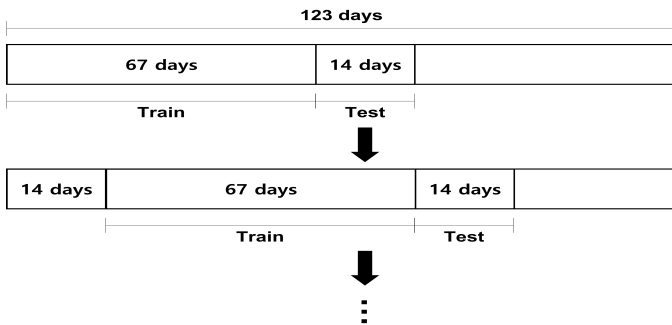


Fig. 3. Train and test data in a moving horizon manner. For the first training, the data of the first 67 days is used. The first trained FCNNs are used to test the 14 days after the duration. For the second training, the data from 15 day to 81 day is used to update the FCNNs slightly. Then, the data from 82 day to 96 day is tested by these second trained set of FCNNs. The rest of data is tested with the FCNNs trained in the moving horizon manner.

## 4. RESULTS AND DISCUSSION

### 4.1 Prediction accuracy for test data

Predicted removal amount and the real plant data are compared. In Fig. 4.1, the percent errors between them are shown for the test data. From 68 day to 81 day, 14 days, the first trained FCNNs are used to calculate the predicted values. For each next 14 days, the second, third, and fourth trained FCNNs are used, respectively. The results without weight clipping strategy and with weight clipping are shown in Figs. 4.1(a) and (b), respectively. In addition, after taking the absolutes on the error percents, the average, maximum value, and minimum values are shown in Table 4.

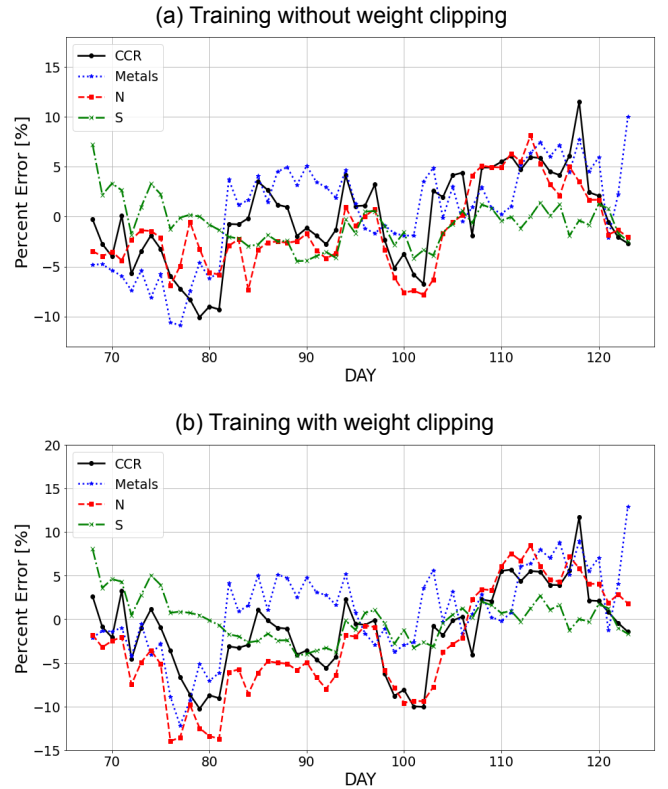


Fig. 4. Percent error for test data set (a) without weight clipping and (b) with weight clipping

The prediction accuracy is similar for both without the weight clipping and with the weight clipping. Even without the clipping method, the maximum percent error is 11.19%, and the average values are 1.63% – 3.78% for the components. With weight clipping, the maximum percent error is 13.98%, and the average values are 1.87% – 5.84%. However, as discussed in the next subsection, the consistency problem occurs without the clipping.

### 4.2 Validation of consistency to physical and chemical tendencies

To validate the physical and chemical compatibility, we increase or decrease each one of the features by 2% and then calculate the percent change in the removal amount. In other words, at each time, only one feature is changed by 2% and the other features are kept the same. Here,

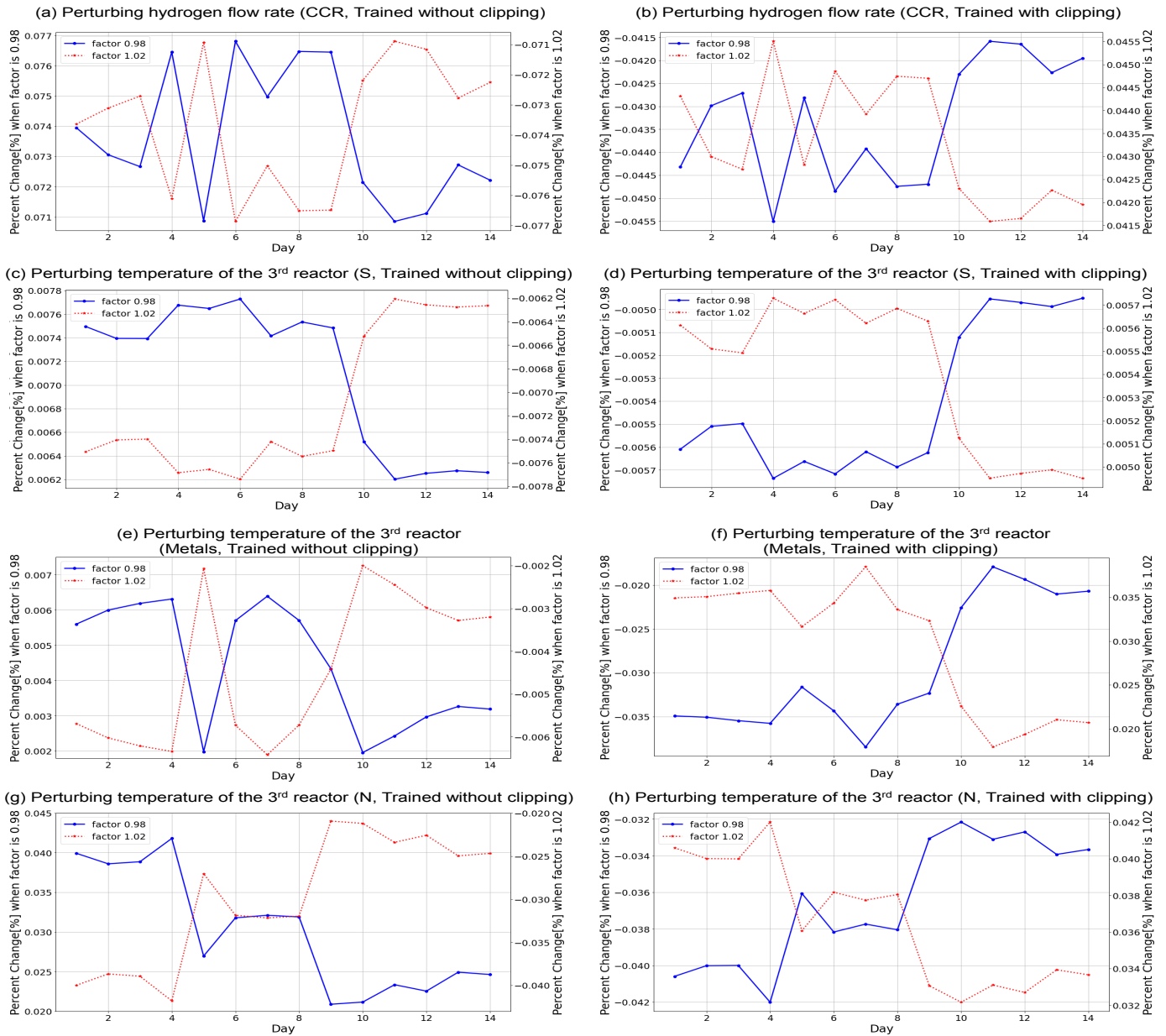


Fig. 5. Percent change in removal amount of (a) CCR, (c) S, (e) Metals, (g) N predicted by FCNNs trained without clipping method. Percent change in removal amount of (b) CCR, (d) S, (f) Metals, (h) N predicted by FCNNs trained with the proposed clipping strategy.

Table 4. Maximum, minimum, and average values of the absolute values of percent errors [%] when testing the FCNNs trained without clipping method and with clipping method.

	Training without weight clipping			
	S	N	Metals	CCR
Max	7.23	8.71	10.88	11.19
Min	0.001	0.16	0.04	0.04
Mean	1.63	3.50	3.33	3.78
	Training with weight clipping			
	S	N	Metals	CCR
Max	8.06	13.98	12.56	11.36
Min	0.012	0.58	0.24	0.067
Mean	1.87	5.84	3.90	3.91

we conduct the calculations for the first 14 days test data using the first trained FCNNs.

As shown with blue lines in Figs. 5(b), (d), (f), and (h), the trained FCNNs with the clipping method always show the decrease in the removal amount when decreasing the temperature or hydrogen flow rate. In addition, as shown with red lines in the same figures, the removal amount increases with the increase of temperature or hydrogen flow rate. However, as shown in Figs. 5(a), (c), (e), and (g), the incompatible results are obtained by the FCNNs trained without clipping method. In addition, we test the consistency validation for all the temperature and hydrogen flow rate features, respectively. As shown in Table 4.1, the removal amount predicted by the FCNNs trained without clipping method changes in contrast to the physical and chemical fact when perturbing several

Table 5. Unrealistic tendencies occur in removal amount prediction when the checked feature is perturbed and the removal amount is calculated by FCNNs trained without clipping strategy.

		Train 1				Train 2			
		S	N	Metals	CCR	S	N	Metals	CCR
1st reactor	Upper temperature			X				X	
	Lower temperature			X				X	
	Hydrogen flow rate			X					
2nd reactor	Upper temperature			X					
	Lower temperature				X	X		X	
	Hydrogen flow rate								
3rd reactor	Upper temperature	X	X		X	X			X
	Lower temperature		X			X		X	X
	Hydrogen flow rate			X					

features respectively. However, the FCNNs obtained with the clipping method always ensure the compatibility to the physical and chemical fact. Because all the weights are positive, the outputs of the FCNNs are always increasing with the input feature increase.

## 5. CONCLUSIONS

In this study, neural network models was constructed based on the real plant data of the RDS process. Based on the catalysts position, the features for the removal reactions of each impurity are selected. The associated temperature, hydrogen flow rate, and each component feed flow rate are used as the inputs of the neural networks, in addition to the general term in the kinetic rate constant. We trained the neural networks multiplied with the aging factors in the moving horizon manner, which can capture the time series effects and has the interpretability. In addition, the weights of the neural networks are clipped to the positive values to ensure the compatibility to the physical and chemical facts; the removal amounts increase (decrease) with the operating temperature or hydrogen flow rate increases (decreases). The percent errors in the predicted removal amount for the test data are smaller than 6% on average. Furthermore, we validated the agreement with physical and chemical facts by checking the percent change in the removal amount of each component with perturbing the features, temperature and hydrogen flow rate. In future works, we will integrate the neural networks with a process simulator to describe the whole RDS process and the hybrid models will be used to design the optimal operation.

## ACKNOWLEDGEMENTS

This work was supported by the National Research Foundation of Korea(NRF) grant funded by the Korea government(MSIT) (No. 2021R1C1C1004217) and by the Korea Agency for Infrastructure Technology Advancement(KAIA) grant funded by the Ministry of Land, Infrastructure and Transport (No. 21ATOG-C162087-01).

## REFERENCES

Al Bazzaz, H., Kang, J.L., Chehadeh, D., Bahzad, D., Wong, D.S.H., and Jang, S.S. (2015). Robust predictions of catalyst deactivation of atmospheric residual desulfurization. *Energy & Fuels*, 29(11), 7089–7100.

Al-Dalama, K. and Stanislaus, A. (2006). Comparison between deactivation pattern of catalysts in fixed-bed and

ebullating-bed residue hydroprocessing units. *Chemical Engineering Journal*, 120, 33–42.

Almutairi, A., Bahzad, D., and Halabi, M.A. (2007). A comparative study on the performance of a catalyst system for the desulfurization of two kinds of atmospheric residues, kuwait export and eocene residual oils. *Catalysis Today*, 125(3), 203–210. Catalysts and Processes for Heavy Oil Upgrading.

Chandrasekaran, S.R. and Sharma, B.K. (2019). 7 - fuel properties associated with catalytic conversion of plastics. In *Plastics to Energy*, 173–220. William Andrew Publishing.

Chehadeh, D., Albazzaz, H., and Bahzad, D. (2018). Predicted effects of heavy feeds on the deactivation of a commercial atmospheric residue desulfurization catalyst system. *Energy & Fuels*, 32(4), 5480–5488.

Chehadeh, D., Khajah, M., Albazzaz, H., and Bahzad, D. (2020). Effect of catalyst and feed properties on the deactivation parameters of an ards process model for hds and hdm reactions. *Chemical Engineering Communications*, 1–12.

Forzatti, P. and Lietti, L. (1999). Catalyst deactivation. *Catalysis Today*, 52(2), 165–181.

Furimsky, E. and Massoth, F.E. (1999). Deactivation of hydroprocessing catalysts. *Catalysis Today*, 52(4), 381–495.

Hauser, A., Marafi, A., Stanislaus, A., and Al-Adwani, A. (2005). Relation between feed quality and coke formation in a three-stage atmospheric residue desulfurization process. *Energy & Fuels*, 19(2), 544–553.

Marafi, A., Hauser, A., and Stanislaus, A. (2006). Atmospheric residue desulfurization process for residual oil upgrading: an investigation of the effect of catalyst type and operating severity on product oil quality. *Energy & Fuels*, 20(3), 1145–1149.

Marafi, M., Al-Sheeha, H., Al-Omani, S., and Al-Barood, A. (2009). Activity of hydroprocessing catalysts prepared by reprocessing spent catalysts. *Fuel Processing Technology*, 90(2), 264–269.

Oh, D., Adams, D., Vo, N., Gbadago, D., Lee, C., and Oh, M. (2021). Actor-critic reinforcement learning to estimate the optimal operating conditions of the hydrocracking process. *Computers & Chemical Engineering*, 149.

Seki, H. and Yoshimoto, M. (2001). Deactivation of hds catalyst in two-stage rds process: II. effect of crude oil and deactivation mechanism. *Fuel Processing Technology*, 69(3), 229–238.