nature portfolio

Peer Review File

Scale-up of Complex Molecular Reaction System by Hybrid Mechanistic Modeling and Deep Transfer Learning

Corresponding Author: Professor Linzhou Zhang

This file contains all reviewer reports in order by version, followed by all author rebuttals in order by version.

Version 0:

Reviewer comments:

Reviewer #1

(Remarks to the Author)
This work presents a new methodology for scale up process design coupling mechanistic, hybrid modelling, and machine/deep learning with transfer learning approaches. As a scale up is persistent challenge in many industries and the use of digital tools seeing greater adoption, this article sets a significant development as how combination of these approaches could be used to benefit process design.

Overall, the article is well written and structured. The approaches are well documented and easy to follow. Furthermore, the article does well to set itself against previous work highlight where it improved. One time that could be improved for clarity is focusing/highlighting more on the general approach benefits, rather than the FCC example. Changes in process mechanisms due to scale-up is a challenge in many industries. This is recognised in the abstract and introduction when introducing the proposed modelling approach. This is then stated to be demonstrated with the FCC example. In the discussion, there is a mix of points relating to the FCC example and the wider approach. I think the order/structure of this section needs to be changed to better highlight the wider approach, as that is what is more impactful in this reviewer's opinion. For example, start with any points specific to the FCC example, then talk about the general applicability – what would be adjusted for another application? What are the minimum data requirements?

Data availability: 'The paper contains information on how to obtain code and data after publication' is checked in the supplementary forms, but this statement doesn't appear to be in the manuscript.

Minor typos:
Page 18, line 321 – polit should be pilot.
Page 20, line 343 – Figure is labelled as Figure 10, but from text and order should be Figure 7.

I'm unable to comment if the article has major significance in the field of FCC. However, I do think the overall approach does have significance in many process industries that suffer from scale up challenges, and this should be highlighted in the article as such.

(Remarks on code availability)
Repository is well structured with a clear and detailed README, along with datasets, trained models, and example usage.

Reviewer #2

(Remarks to the Author)
The attached manuscript addresses a very important problem in chemical engineering: the need to construct pilot plants to scale up laboratory processes to industrially relevant outputs. If successful, such a method would be both impactful for the industry and could help other researchers identify if their work is worthy of considering commercialization. The paper's main claim is that using the proposed algorithm, one can limit to performing lab-scale experiments and generate reliable pilot plat data entirely in-silico. Unfortunately, this claim is not supported by data in the current manuscript form, and the manuscript needs to address all points below to meet the basic criteria for publication in Nature Communications.
1. Scope of reported data: there is no data on the experimental set-up used, which is unacceptable for any scientific

publication. Include details of the reactors (both on lab and pilot scales), used catalysts, flow of products, analytical methods, scheme, pictures of the reactors, etc.
Even if the reactor discusses an AI algorithm, this data must be reported, and it is impossible to finalize the review without access to it.
2. The authors postulate that the proposed method allows them to predict pilot-scale data and compare their outcomes to "optimal" experimental observations. However, there is no information on the optimization of the pilot scale reactor. Therefore, the only thing that the current comparison tells us is that the reaction and the AI model yield similar results, but not that these results are actually meaningful for the scale-up. Given that the proposed case-study process is well-studied, the authors should also compare their results to the available pilot/industrial scale data.
3. Pilot plant data is used in industry not only to predict the molecular composition on a larger scale but, very importantly, to assess the effect of the impurities in the feedstock material, study catalyst deactivation and selectivity over time. How could the proposed method be used to answer these critical questions?
4. An important feature of the proposed algorithm is that it uses an AI-based description of mass transfer, making the entire modeling task less resource-intensive. However, the author has chosen a very simple problem from a mass-transfer perspective (high-temperature gas phase processes are typically not mass-transfer limited). How can the authors guarantee that the same approach would work for liquid-phase reactions or other systems that are indeed mass transfer limited?
5. The article is very repetitive in description, and thus, it isn't easy to access the most important information. E.g. this entire paragraph offers a very basic introduction to the topic and does not need to be repeated as late as line 104: "Scaling up a novel chemical process from the laboratory to industrial application typically involves three stages: laboratory experiments, pilot-scale testing, and industrial trials. Each stage serves distinct objectives, often requiring reactors with different configurations. " All writing should be reviewed and edited to be more concise.
6. The authors left a number of unchecked boxes on their checklist documents, and the review suggests double-checking if that's accurate.

(Remarks on code availability)
The code is accessible, yet it consists of many separate files and it is unclear which one to run to check the reported results.

Reviewer #3

(Remarks to the Author)
Review - Scale-up of Complex Molecular Reaction System by Hybrid
Mechanistic Modeling and Deep Transfer Learning

The paper proposes an approach for dealing with scale-up of chemical processes based on a hybrid (e.g., first-principles and machine learning) and multi-scale modeling framework, using deep transfer learning aided by residual neural networks (ResNets). The proposed approach was applied to a fluid catalytic cracking (FCC) system. The premise is that the hybrid model, which is generated using mechanistic-based (first-principles) and data-driven-based techniques using data from different scales (laboratory-scale, pilot-scale and industrial-scale) coupled with transfer learning, generates a model that can potentially perform well across scales with respect to product molecular properties and compositions, when compared against purely mechanistic and/or purely data-driven approaches.

Overall Evaluation

The objective of the work is explicit, with a clear picture of the scope and applications in which the proposed method could benefit the community. The case study (FCC) is representative in that the scale-up and the different intrinsic properties determination across scales is challenging and could benefit from a hybrid modeling approach. The discussion and methods section could provide more explanation and the contribution could benefit from more specificity in several sections. We recommend this article for publication after major revisions as outlined in the comments below:

1. Different communities use different definitions for "hybrid model". The paper should make the definition clear. In particular, is the hybrid model composed of both mechanistic and machine-learning-based components, or is it a purely machine learning model where the training step combines both experimental data and simulated/generated data (e.g., surrogate model with data augmentation) or both. The paper should make this distinction clear to the reader early in the article.

2. Related to the previous comment, Lines 97-98: "The mechanistic model served as a data generator, producing a large dataset to train the neural network. On this basis, network parameters were fine-tuned using the pilot plant data to calculate the FCC product distribution accurately with limited data samples." Did the authors consider embedding the mechanistic model in the final form instead of only using it for training purposes? Also, is the "bulk property calculator" used only during training, or is it embedded when the model is used for optimization. The paper would benefit from further clarification of the model structure and additional detail on the training procedure used.

3. From Figure 2 (a), we see that the proposed model structure separates components for process conditions and feedstock. However, it appears that the only outputs are after the "Integrated ResNet". How do we know that this structure is better than a single neural network? Have the authors compared against this simpler model structure? Please clarify the benefits of this structure and provide additional details if there are data corresponding to outputs from the "Process-based" and "Molecule-based" ResNets.

4. It seems that the "Property-informed Transfer Learning Strategy for Process Scale-up" has similarities with PINNs since it

embeds first-principles knowledge in the loss function. Since this is an important novelty of the paper, it would be helpful if the authors could clarify the differences or similarities?

5. It is not clear why the authors chose Residual Networks (ResNet) as the data-driven modeling paradigm instead of other approaches such as Neural ODEs and Universal Differential Equations, for example. The paper would benefit from a clearer description of this selection.

6. Lines 198-200: “To validate the proposed hybrid model, a molecular-level mechanistic model for the naphtha FCC process was developed, and cross-scale computation from the laboratory to the pilot plant was performed.” Please provide more detail here. It can be very challenging to carry a molecular-level model from laboratory to pilot scale, and it is not clear how “cross-scale computation” was performed.

7. Regarding the case study, lines 249-250: “it was found that the FCC process converts olefins into i-paraffins and aromatics.” If this is a new finding for these applications, then it needs further clarification. If this is being mentioned to indicate that the proposed hybrid model is matching expectations, then this should be clarified.

8. Lines 274-275: “The laboratory-scale hybrid model was trained, and its hyperparameters were optimized using a Bayesian optimization (BO) algorithm.” The paper would benefit from a more thorough discussion of how this hyperparameter tuning was performed. If an existing package was used, then please provide those details.

9. Regarding the two-layer ResNet architecture Lines 166-168 that “This architecture mirrors the logic of the mechanistic model, facilitating more precise identification of parameter fine-tuning locations during transfer learning.” It is not clear to this reviewer what is meant here or why this should facilitate “more precise identification”. The authors should provide clarification here.

10. In the optimization results, “The results indicate that increasing the feedstock temperature by 8°C and reducing the catalyst temperature by 8°C can lead to higher yields of gasoline, i-paraffins, and aromatics. Simultaneously, both olefin and coke contents are decreased”. This seems like an interesting finding for improving operation. Can the authors expand on the control structure that would be used to achieve this temperature control?

11. The optimization is performed using the hybrid model, and a single optimal operating point is found. As with any hybrid or surrogate-based approach, it is important to evaluate the error (and ideally the optimality) using a rigorous model or experimental data. The paper would also benefit from some discussion of how one might account for uncertainty in the hybrid model.

12. The paper would benefit from a more thorough introduction on transfer learning as it relates to the task at hand. Further discussion of ResNets would also be a welcome addition.

13. The paper lacks a formal mathematical description of the optimization being performed, including the constraints and objective function. This is very important for the reader to understand the specifics of problem being solved.

14. Lines 482-483: “In this work, R2 is used as the objective function of the BO algorithm. When R2 is maximal, the optimal hyperparameters can be obtained.” Is R2 the most appropriate metric for this, especially given that the underlying phenomena is nonlinear and involving an optimization/training step? Could other metrics such as MAE, RMSE, or likelihood also be used. Can the authors clarify their choice of objective?

15. The “Transfer Learning” section needs to be more specific, providing substantially more detail about how this approach is applied in this particular problem.
Additional minor comments to further improve the manuscript
16. Some sentences could be reworked for better readability. For example, lines such as 76-77, “Qiu and her colleagues introduced an AI-mechanism integrated model for the naphtha steam cracking process based on the convolutional neural network (CNN)33” could be “Qiu and collaborators33…” with the reference immediately after the researchers’ names. The same is true for line 79-80, 84. Please check these occurrences and modify accordingly.

17. Please consider adding additional clarification for jargon in the manuscript. Terms such as “data augmentation”, “parameter transfer strategy”, “NN-NSGA” and others are mentioned but would benefit from definition, clarification, or additional specifics.

18. The “software policy checklist” *.pdf does not seem to be properly compiled, with a placeholder text which says, “If this message is not eventually replaced by the proper contents of the document, your PDF viewer may not be able to display this type of document.”

(Remarks on code availability)

Reviewer #4

(Remarks to the Author)

I co-reviewed this manuscript with one of the reviewers who provided the listed reports. This is part of the Nature Communications initiative to facilitate training in peer review and to provide appropriate recognition for Early Career Researchers who co-review manuscripts.

(Remarks on code availability)

Version 1:

Reviewer comments:

Reviewer #1

(Remarks to the Author)
Many thanks to the authors for addressing my previous comments about the focus of the article between it's general appeal and FCC focus. The structure of the conclusion and incorporation of generalizability and how the network structure might be adapted for different cases is a great addition. I'm happy the author's have addressed my previous comments on the article.

I do not have a further question with regards to terminology for the network architecture. It's referred to throughout as ResNet which implies a specific deep learning CNN architecture and the authors cite (REF 36) a paper that compared different CNN architectures for naphtha cracking processing. However, the architecture (described in ESI and shared code) used in this work uses MLP layers not convolutional ones. Unless there is a clear connection to the original ResNet design, I do not think this terminology should be used.

Additional minor comment:
Figure 3 - The first column (commonality and difference) should use more descriptive labels. For example, mode of operation, reaction type, etc.

(Remarks on code availability)
README for code provided is quite extensive, but could be improved. For example:
Installation - Why not include a requirements.txt to setup the environment.
System requirements - 'Standard computer' could mean anything. This should be more specific, along with examples of the memory requirements.
Running the code - Details given are ok, but again could be clearer for a user. Include example commands for running the 'Bayes_for_HNN_gasoline_FFC.py' etc. I also think this might benefit from a juptyer notebook version with additional markdown and comments alongside the code.

Streamlit app is interesting, but not clear what it's purpose is? Is it for just displaying the training results of the models? i.e. it's not for actually training the lab scale or pilot scale models?

Reviewer #2

(Remarks to the Author)
The authors provided detailed answers to my questions. While there are some points where I don't 100% agree with the argumentation, I do believe that this paper, in the current form, is a valuable tool for the community, and I recommend its publication.

(Remarks on code availability)
I check the basic functionality of the code but didn't review line by line. the new README file is very helpful.

Reviewer #3

(Remarks to the Author)
Please note, "I co-reviewed this manuscript with one of the reviewers who provided the listed reports. This is part of the Nature Communications initiative to facilitate training in peer review and to provide appropriate recognition for Early Career Researchers who co-review manuscripts."

Overall comment: Thank you for this response. The authors have addressed all of our major comments, and we believe the paper can be published after fixing the minor comments shown below.

Comment 1:
In the literature we are familiar with, a “hybrid” model must contain both mechanistic and data-driven components. Given the additional explanation, we believe that the laboratory-scale process is not a hybrid model, but a data-driven surrogate of the original mechanistic model. On the other hand, the pilot-scale machine learning model is indeed a hybrid model as it includes bulk property calculations based on first-principles in addition to the data-driven components. The authors have defined their intention with hybrid and data-driven models in the paper, so we leave this change as a choice for the author and editor.

Comment 2: This reviewer agrees with the response and action provided by the authors.

Comment 3: Thank you for the response. No further changes requested.

Comment 4: Thank you for the response. No further changes requested.

Comment 5: Thank you for the response. Is it true that the the bulk property calculation is a post-processing calculation? This could be better clarified in the paper.

Comment 6: Thank you for the response. No further changes requested.

Comment 7: Thank you for the response. No further changes requested.

Comment 8: No further comment, except that the order of the new references added to support the findings are flipped when comparing the “response to reviewers” document and updated manuscript. You might want to double check these.

Comment 9: Thank you for the response. No further changes requested.

Comment 10: Thank you for the response. No further changes requested.

Comment 11: Thank you for the response. No further changes requested.

Comment 12: Interesting analysis. No further changes requested.

Comment 13: Thank you for the response. No further changes requested.

Comment 14: The first occurrence of the NSGA-II acronym should be next to its definition (non-dominated sorting genetic algorithm II). Otherwise, no further comment.

Comment 15: Table R13, it looks like there is an MAE that is negative? Can you clarify this?

Comment 16: The text added in “Model structure for transfer learning” is different between the updated manuscript and the “response to reviewers” document. You might want to double-check.

Comment 17: Thank you for the response. No further changes requested.

Comment 18: Thank you for the response. No further changes requested.

Comment 19: Thank you for the response. No further changes requested.

(Remarks on code availability)

Response to Reviewer 1:

Comment 1:

One time that could be improved for clarity is focusing/highlighting more on the general approach benefits, rather than the FCC example. Changes in process mechanisms due to scale-up are a challenge in many industries. This is recognised in the abstract and introduction when introducing the proposed modelling approach. This is then stated to be demonstrated with the FCC example. In the discussion, there is a mix of points relating to the FCC example and the wider approach. I think the order/structure of this section needs to be changed to better highlight the wider approach, as that is what is more impactful in this reviewer's opinion. For example, start with any points specific to the FCC example, then talk about the general applicability – what would be adjusted for another application? What are the minimum data requirements?

Response:

Thanks to the reviewer's comments in the discussion section. We strongly agree with the reviewer's comment. The change in mechanism between different process scales is indeed a critical challenge for model development. The machine learning based on the cross-scale transfer learning method will benefit both the acceleration of process scale-up and the rapid modeling of pilot/industrial plants. The FCC reactor is a very good research subject since it is a typical gas-solid fluidized bed reaction system involving complex flow regimes and mass transfer phenomena. The method developed for FCC scale-up is also applicable to the other process.

We regret that the previous manuscript mainly focused on the FCC process. As the reviewer pointed out, the structure of the manuscript and the discussion section are not well organized for the readers to understand the general applicability of the proposed method. According to the reviewer's suggestion, we have revised the discussion section. In the revised version, we first summarized the application of the developed methodology to the FCC process. On this basis, we propose a generalizable modeling paradigm that can be applied to the scale-up of other complex chemical processes. Finally, we outline potential directions for further development of the methodology.

Modification:

The discussion section was rewritten in the revised manuscript, and the revised discussion section is as follows:

To address the cross-scale computational challenges in the process scale-up of complex molecular reaction systems, this study proposed a hybrid model integrating molecular

conversion mechanisms with deep transfer learning. The method was applied to the naphtha FCC process from the laboratory to the pilot plant. A molecular-level kinetic model was initially developed based on laboratory-scale data, serving as a data generator to produce a molecular conversion dataset with a size of ~10^4. A neural network architecture suitable for transfer learning was then designed. It incorporated two ResNet modules to extract features from process conditions and molecular composition. This architecture achieved highly accurate predictions of molecular conversion behavior within a laboratory-scale reactor, with an MAE for all molecular contents of less than 10^{-4}. To address data discrepancies at different scales, this work proposed a property-informed transfer learning strategy. The strategy can simultaneously predict product molecular composition and bulk properties using a neural network architecture, and it is beneficial to reduce time for experimental evaluation and model development/parameter tuning during the FCC process scale-up. Compared to traditional molecular-level mechanistic models, the proposed method avoids modeling for each scale separately. It also improved computational accuracy over the parameter transfer strategy proposed by Chen et al.[1]. Compared to conventional single neural network surrogate models, the proposed architecture not only improves prediction accuracy but also allows for flexible parameter freezing and fine-tuning during transfer learning.

Building on the FCC case study, we propose a generalizable modeling paradigm for process scale-up. Initially, high-throughput experiments are performed to generate data for developing the mechanistic model. The mechanistic model will be used to generate numerous data for training the hybrid model. Following this, a limited number of pilot/industrial-scale data points (as few as 4 ~ 5 points per process condition) is utilized to fine-tune the hybrid model. After fine-tuning, the model can be integrated with advanced process control (APC) systems to optimize process conditions in real-time. Notably, applying this model to other processes requires developing corresponding mechanistic models based on process characteristics. For processes dominated by transport limitations, replacing the ResNet with PINN or Neural ODE may further enhance model accuracy, since the key transport equations can be incorporated into the neural network. Furthermore, incorporating catalyst descriptors into the hybrid model will be crucial when scale-up involves catalyst variations. Overall, this work presents a novel modeling framework for the scale-up of complex reaction systems. It offers a promising pathway for complex process industries that suffer from scale-up challenges.

Reference:

1. Zhengyu Chen, et al. Molecular-Level modeling for naphtha olefin reduction in FCC subsidiary Riser: From laboratory reactor to pilot plant. *Chemical Engineering Journal* 437, 135429 (2022).

Comment 2:

Data availability: 'The paper contains information on how to obtain code and data after publication' is checked in the supplementary forms, but this statement doesn't appear to be in the manuscript.

Response:

The information on how to obtain the code and data has been added to the revised manuscript.

Data availability:

All data used for model training and prediction in this study are publicly available at: https://github.com/chenzhengyuAI/complex-reaction-system-scale-up.

Code availability:

The code of the hybrid model and all scripts used to train and predict product distribution are publicly available at: https://github.com/chenzhengyuAI/complex-reaction-system-scale-up.

Modification:

The statement for data availability and code availability was added to the revised manuscript.

Comment 3:

Page 18, line 321 – polit should be pilot.

Response:

We made a mistake in the previous manuscript. In the revised manuscript, 'polit' on page 18, line 32 has been revised into 'pilot'.

Modification:

'polit' has been revised into 'pilot' in the revised manuscript.

Comment 4:

Page 20, line 343 – Figure is labelled as Figure 10, but from text and order should be Figure 7.

Response:

We have corrected the label of Figure 10 to Figure 7.

Modification:

The label of Figure 10 has been changed to Figure 7 in the revised manuscript.

Comment 5:

I'm unable to comment if the article has major significance in the field of FCC. However, I do think the overall approach does have significance in many process industries that suffer from scale-up challenges, and this should be highlighted in the article as such.

Response:

We appreciate the reviewer's comment on the significance of our work. We hope our model will help researchers to rapidly build precise models for the optimization of operation conditions during process scaling-up. According to the reviewer's suggestion, we have revised the manuscript, and the importance of the method at the end of the discussion section was highlighted as follows:

"Overall, this work presents a novel modeling framework for the scale-up of complex reaction systems. It offers a promising pathway for complex process industries that suffer from scale-up challenges."

The reviewer also mentioned the significance of the FCC field. This method is also important for the FCC processes, since the apparent reaction rate of molecules was changed significantly for different scales. Different models were required to be developed for different reactor scales/structures. The proposed transfer learning method can automatically capture the influence of process scale-up, enabling accurate prediction of the key properties while reducing time for experimental evaluations and model development/parameter tuning.

Historically, the gas-solid fluidization theory, which is extremely essential in the chemical industry, was developed along with the invention of the FCC process. Thus, we believe that using the FCC process as a case study is both appropriate and meaningful. The successful application of this method also has important guiding significance for the accelerated process scale-up of other new processes. In the further, we aim to extend this modeling framework to other complex process industries facing similar scale-up challenges.

Modification:

The importance of the method for FCC and other processes facing scale-up challenges was added to the Discussion section of the revised manuscript.

Response to Reviewer 2:

Comment 1:

Scope of reported data: there is no data on the experimental set-up used, which is unacceptable for any scientific publication. Include details of the reactors (both on lab and pilot scales), used catalysts, flow of products, analytical methods, scheme, pictures of the reactors, etc.

Even if the reactor discusses an AI algorithm, this data must be reported, and it is impossible to finalize the review without access to it.

Response:

We understand the reviewer's concern about the experimental data accessibility. We have added all the information that was used in the manuscript to guarantee that all the readers are able to check experimental data points or build a similar machine-learning model based on the data in the submitted manuscript.

- Laboratory-scale and pilot-scale reactor

In this work, the laboratory-scale experiments for naphtha FCC were carried out in a fixed fluidized bed (FFB) reactor, which is a batch reaction device. The lower part of the reactor is conical, while the upper part is cylindrical. The main parameters of FFB are listed in Table R1.

The naphtha FCC pilot-scale experiments were performed in a riser, and its main process parameters are also presented in Table R1.

Table R1. Main process parameters of laboratory-scale and pilot-scale reactors

Parameter	Data
Fixed fluidized bed reactor	
Reactor volume, cm^3	600
Reactor diameter at top, cm	5.5
Reactor height, cm	41
Catalyst loading, g	150
Riser	
Riser length, cm	665
Riser diameter, cm	1.6
Mass flow rate for feedstock, kg/h	2
Mass flow rate of atomizing steam, kg/h	0.3
Pressure, kPa	135
Pressure drop, kPa	2.64

- Catalyst

The catalyst used in this work is an industrial FCC zeolite catalyst, and the catalyst properties are listed in Table R2.

Table R2. Catalyst properties for laboratory-scale and pilot-scale FCC experiments

Catalyst property		Data
Micro-activity		66
Specific surface area, cm^2/g		95.3
Pore volume, cm^3/g		0.028
Bulk density, g/cm^3		0.9
Particle size distribution, wt%	0 ~ 40 μm	5
	40 ~ 60 μm	23
	60 ~ 80 μm	28
	80 ~ 105 μm	23
	> 105 μm	21

- Experimental scheme

1. Laboratory-scale experiment

Laboratory-scale experiments of naphtha FCC were performed in an FFB reactor. 150 g of catalyst was initially loaded in the reactor and pre-fluidized with steam. Then, the reactor was preheated to the reaction temperature, while the naphtha feedstock was pumped into the reactor bottom. The mass of naphtha feedstock was calculated by the CTO. At the reactor bottom, the feedstock contacted the catalysts and underwent the catalytic cracking reaction to generate the gaseous product. When the reaction was completed, 20 minutes steam stripping process with a constant rate was implemented to separate the products. Liquid fractions were collected at the condenser bottom, while gaseous products were obtained at the top. The liquid products were fractionated via micro-distillation into gasoline and diesel fractions, and the gaseous products were divided into dry gas ($C_0 \sim C_2$) and LPG ($C_3 \sim C_4$) according to the carbon number. After cooling the reactor to ambient temperature, the coked catalyst was unloaded to determine coke yield. The FFB reactor and experimental scheme are illustrated in Figure R1.

2. Pilot-scale experiment

Pilot-scale experiments of naphtha FCC were performed in a riser reactor. Initially, the feedstock and regenerator were preheated to set the temperature according to process conditions. 2 kg/h naphtha was pumped into the bottom of the riser, where it contacted high temperature catalyst from the regenerator. The catalyst circulation rate was calculated based

on the CTO. The FCC reaction occurred in the riser. The generated gaseous substances and deactivated catalyst were delivered into the disengager for gas-solid separation. The gaseous phase was collected at the disengager top, while the deactivated catalyst was recycled to the regenerator for coke combustion and regeneration, enabling continuous operation of the unit. The products from the disengager were condensed into gas-liquid phases. The gaseous products were divided into dry gas (C_0 ~ C_2) and LPG (C_3 ~ C_4) according to the carbon number, while the liquid phase was fractionated via micro-distillation into gasoline and diesel fractions. The riser reactor and pilot-scale experimental scheme are illustrated in Figure R2.

- Product flow

The product flow of laboratory-scale and pilot-scale contains dry gas, LPG, gasoline, diesel, and coke. The specific descriptions are as follows:

When FCC products were cooled in a condenser, the gas and liquid products were obtained. The gaseous products were collected via the water displacement method, and the mass can be calculated by volume. Liquid phase products were directly weighed to determine their mass. The coke content was converted according to the generated CO_2.

The gaseous products were divided into dry gas (C_0 ~ C_2) and LPG (C_3 ~ C_4) according to the carbon number. The liquid products were fractionated via micro-distillation into gasoline fraction (Initial boiling point ~ 200 ^oC) and diesel fraction (200 ^oC ~ 350 ^oC).

- Pictures of Reactor

Laboratory-scale reactor, pilot-scale reactor, and detailed experimental scheme are shown in Figures R1 and R2.

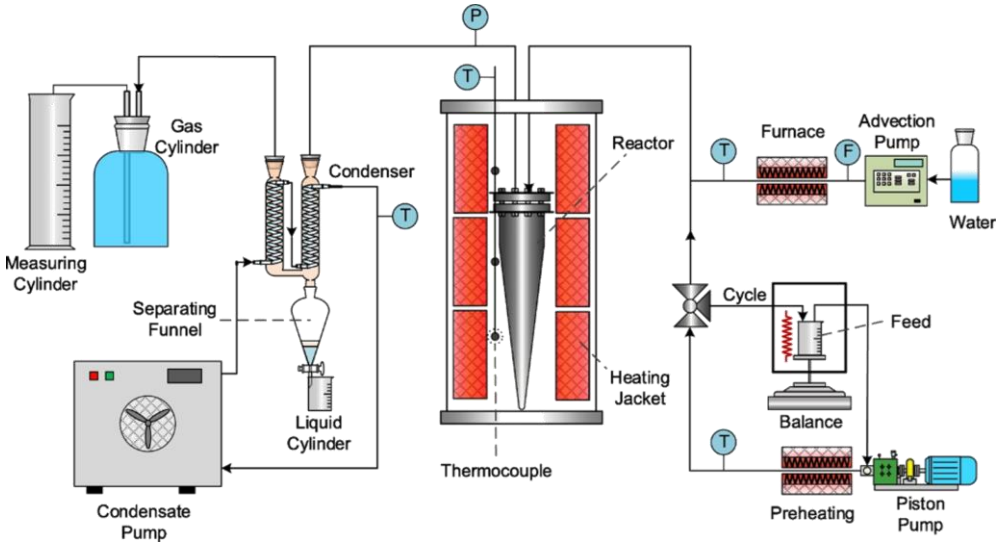

Figure R1. Laboratory-scale reactor and experiment scheme for naphtha FCC.

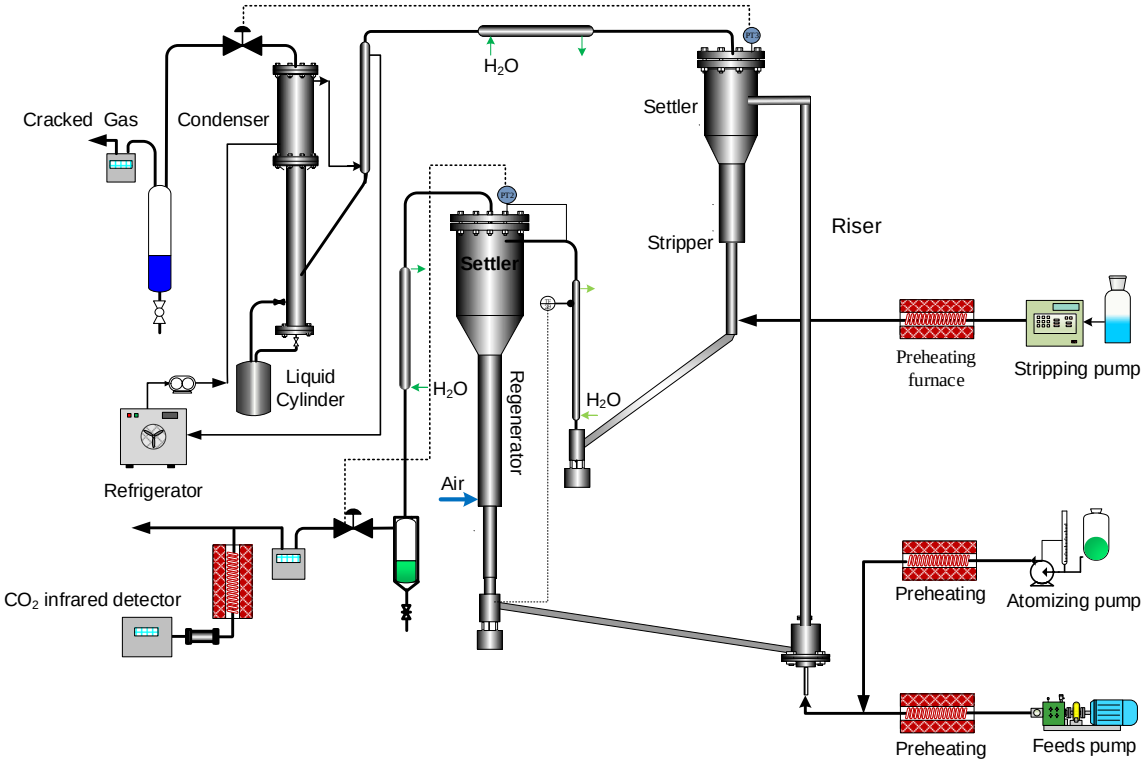

Figure R2. Pilot-scale reactor and experiment scheme for naphtha FCC.

- Analytical methods

1. Analytical method for gaseous product

The gaseous products from the FCC were quantitatively analyzed using an Agilent 6890 gas chromatograph (GC). The GC with flame ionization detector (FID) was employed for analyzing hydrocarbons, while the GC with thermal conductivity detector (TCD) was used for analyzing hydrogen. The analytical method used the UOP 539-2012 standard method.

2. Analytical method for liquid product

The molecular composition of the naphtha and gasoline fractions was analyzed using an Agilent 6890 GC equipped with an FID, following the ASTM D5134-21 standard method.

3. Analytical method for the solid product

The solid product mainly consists of coke deposited on the catalyst, which was analyzed using an HIR-944B high-frequency infrared carbon-sulfur analyzer. The catalyst was subjected to high-temperature combustion, and the CO_2 content in the mixed gas was measured via infrared analysis to determine the coke content.

Modification:

The details of the experiment scheme, product flow, and analytic method were added to the Method section.

Parameters for laboratory-scale and pilot-scale reactors are presented in Supplementary Table 9.

Catalyst information is listed in Supplementary Table 10.

Laboratory-scale and pilot-scale experimental schemes were added to Supplementary Figs. 7 and 8.

Comment 2:

The authors postulate that the proposed method allows them to predict pilot-scale data and compare their outcomes to "optimal" experimental observations. However, there is no information on the optimization of the pilot scale reactor. Therefore, the only thing that the current comparison tells us is that the reaction and the AI model yield similar results, but not that these results are actually meaningful for the scale-up. Given that the proposed case-study process is well-studied, the authors should also compare their results to the available pilot/industrial scale data.

Response:

Thank you for the reviewer's comment. We acknowledge that the original manuscript lacked information on the optimization of the pilot-scale reactor, which may raise concerns about the significance of our model. To address this issue, we have added the information on the pilot-scale reactor, as shown in Comment 1. Simultaneously, the significance of achieving consistency between AI model calculation results and reaction data was also discussed as follows:

The core of process scale-up is to address the inconsistencies in apparent reaction behavior between laboratory-scale reactors and pilot-scale plants. These changes are mainly caused by changes in fluid flow regime and mass transfer. To investigate the effects of flow regime and mass transfer on the reaction during scale-up, it is usually necessary to extrapolate pilot-scale reaction behavior based on laboratory-scale experimental results. In this research process, model guidance plays a crucial role. Conventional approaches involve reconstructing mechanistic models (e.g., CFD) for pilot-scale plants and then optimizing the model parameters to match experimental observations. However, this approach is overly complex, as mechanistic models may struggle to fully capture the actual flow regime and transport behavior. Moreover, their computational cost is relatively high.

To explore a more efficient method, we used transfer learning to automatically extract features related to process scale-up. By integrating mechanistic models with transfer learning, the developed model can accurately predict the pilot-scale product distribution using only a

limited number of samples. It is beneficial to accurately predict product distribution while reducing time for experimental evaluations and model development/parameter tuning.

The proposed method followed the same research strategy (extrapolating the behavior of the pilot plant from the laboratory reaction behavior), but it adopted a novel modeling framework to achieve this goal. The effects of fluid flow and mass transfer on the reaction process were captured implicitly as changes in the neural network weights. In this work, the agreement between the AI model and the experimental data was achieved by fine-tuning the parameters in the Process-based ResNet and Integrated ResNet. These parameter adjustments reflected the changes associated with process scale-up.

Moreover, we also strongly agree with the reviewer's comments that the calculated results of the AI model should be validated. Therefore, we performed pilot-scale experiments and developed a pilot-scale mechanistic model for validation.

- Pilot-scale experiment validation：

Due to confidentiality agreements, product data from industrial units cannot be publicly disclosed in academic publications. However, in our previous study on vacuum gas oil (VGO) FCC, the pilot-scale product distribution was compared with that of an industrial plant, as shown in Table R3. The results demonstrated that the pilot-scale product data agreed well with the industrial-scale data.

Table R3. The product distribution of the industrial FCC unit and the pilot FCC unit

	Industrial FCC unit	Pilot-scale FCC unit
Reaction temperature, ^{o}C	500	500
CTO	4.2	5
Reaction time, s	2.95	2.99
Product yield		
Dry gas, wt%	3.1	1.2
LPG, wt%	8.9	9.9
Gasoline, wt%	53.1	51.5
Diesel, wt%	27.7	26.4
Heavy oil, wt%	1.5	4.6
Coke, wt%	4.7	5.1
Loss, wt%	1.0	1.0

Building on this established consistency, this study utilized pilot-scale data to validate the AI model results. In addition to the naphtha composition used in the original manuscript

(naphtha 2), we selected another naphtha (naphtha 3) to further validate the model. The composition for naphtha 2 and 3 is presented in Table R4.

Table R4. Detailed PIONA composition of naphtha 2 and naphtha 3

Carbon number	3	4	5	6	7	8	9	10	11	12	13	Total
Naphtha 2												
n-Paraffins		0.4	0.8	1	1	1.2	0.4	0.4	0.2	0.1	0.1	5.6
i-Paraffins		0.7	3.9	6.3	5	3.9	1.9	1.9	1.1	0.3	0.3	25.3
Olefins	0.2	2.4	6.9	9.3	9	5.9	2.8	2.2	0.2			38.9
Naphthenes			0.1	0.2	3.5	1.4	1.3	0.3	0.1	0.1		7
Aromatics					3.1	7.6	6.1	4.9	1.5			23.2
Naphtha 3												
n-Paraffins		0.5	0.7	0.8	0.9	1.1	0.3	0.2	0.1			4.6
i-Paraffins		1	4	5.9	3.3	2.9	1.5	2.1	0.8	0.1	0.1	21.7
Olefins	0.1	4.3	10.7	12.5	10.6	6.4	3.6	1.7	0.1			50
Naphthenes			0.1	0.2	3.2	1.4	2.9	0.5				8.3
Aromatics					2.2	4.8	5.3	2.7	0.4			15.4

1. Taking naphtha 2 as feedstock, we conducted 15 sets of pilot-scale FCC experiments. In the original manuscript, we used these 15 sets of data to generate a series of low-fidelity data to fine-tune the neural network, while the 15 sets of experimental data were not utilized to train the AI model. After completing the model training, the AI model was directly applied to predict the product distributions under the 15 experimental conditions. The predicted results were then compared with the experimental data, as shown in Figure R3. The results demonstrate that the trained AI model can accurately predict the product distributions of the pilot-scale unit. The mean absolute error (MAE) is 0.3 wt%. The robustness of the model is validated.

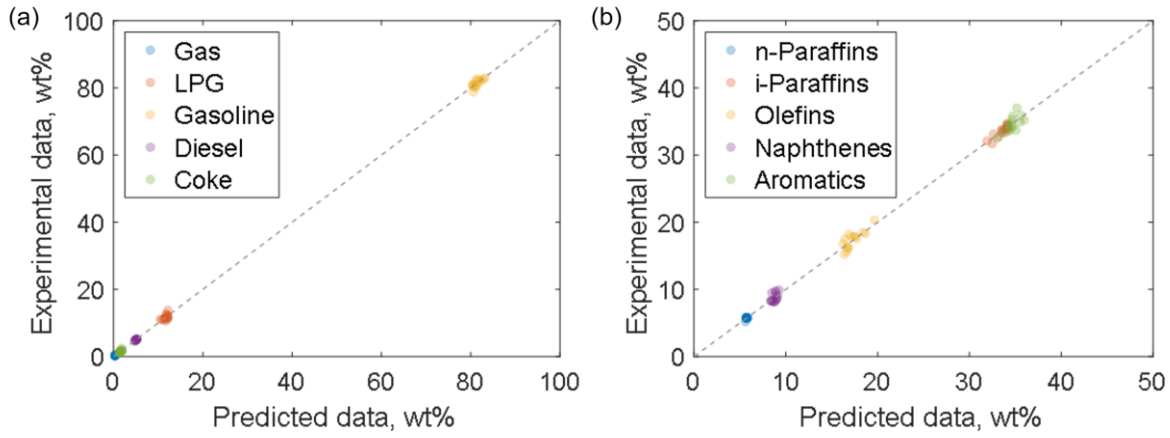

Figure R3. Comparison between predicted data and experimental data of naphtha 2 FCC. (a) Fraction yield; (b) Gasoline composition.

2. To further evaluate the generalization capability of the AI model, we conducted an additional pilot-scale FCC experiment using a different feedstock, referred to as naphtha 3. The process conditions were as follows: feedstock temperature of 200 °C, regenerated catalyst temperature of 623 °C, reaction time of 3.6 s, and a catalyst-to-oil ratio (CTO) of 5.5. Notably, the molecular composition of naphtha 3 differs significantly from that of naphtha 2, with a considerably higher olefin content, as detailed in Table R4. The trained AI model was directly applied (without any retraining) to predict the FCC product distributions of naphtha 3. The predicted results are shown in Figure R4. Despite significant changes in the feedstock molecular composition, the developed AI model can still accurately predict product distribution, demonstrating its strong generalization ability.

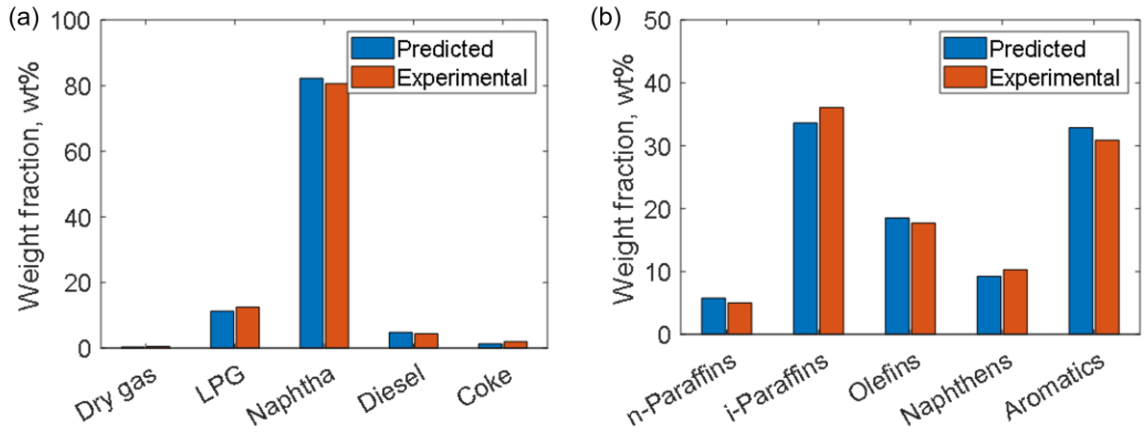

Figure R4. Comparison between predicted data and experimental data for naphtha 3 FCC. (a) Fraction yield; (b) Gasoline composition.

- Pilot-scale mechanistic model validation：

As rightly pointed out by the reviewer, validation of the optimized process conditions is meaningful for the scale-up. Unfortunately, the FCC experiments for naphtha 2 have already been completed, and the oil samples have been exhausted, making further experimental validation challenging. However, we were also trying our best to validate the optimized process conditions and developed a molecular-level mechanistic model for the pilot plant. Compared to the laboratory-scale mechanistic model, the pilot-scale mechanistic model requires integrating the molecular-level kinetic model with the riser reactor model. The mass and heat balance equations in the riser model are presented in Eqs. (R1) ~ (R3).

$$\frac{dC_i}{dz} = \frac{\rho_c \varepsilon_c}{v_g} R_i \varphi_{ck} \tag{R1}$$

$$\frac{dT_c}{dz} = \frac{A h_p A_p}{F_c C_{p\_c}} \left(T_g - T_c\right) \tag{R2}$$

$$\frac{dT_g}{dz} = \frac{A}{F_g C_{p\_g}} \left[h_p A_p \left(T_c - T_g\right) + \rho_c \varepsilon_c \sum\nolimits_j \left(-\Delta H_j r_j\right) \right] \tag{R3}$$

The mass and heat balance equations in the riser reactor model involve numerous bulk properties, and their correlation formulas are shown in Table R5.

After developing the molecular-level kinetic model for the naphtha FCC model, the kinetic parameters were regressed using pilot-scale data. The developed model accurately calculated and predicted the product distribution for the pilot-scale unit, and partial results are shown in Figure R5.[1]

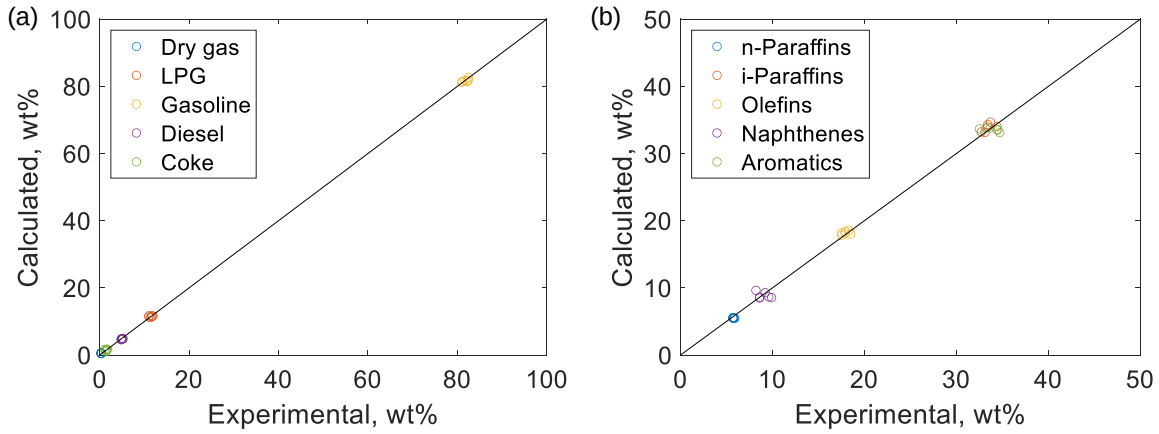

Figure R5. Calculated results for Pilot-scale mechanistic model (a) Product yield; (b) Gasoline composition

Table R5. Equations for properties in the riser model

Property	Equations	
Catalyst deactivation	$\varphi_{ck} = \frac{1}{1 + a_{ck}(C_{ck})^{b_{ck}}}$	(R4)
Volume fraction for catalyst phase	$\varepsilon_c = \frac{F_c}{v_c \rho_c A}$	(R5)
Volume fraction for oil phase	$\varepsilon_g = 1 - \varepsilon_c$	(R6)
Heat transfer coefficient	$h_p = 0.03 \frac{K_g}{d_c^{2/3}} \left[\frac{\lvert (v_g - v_c) \rvert \rho_g \varepsilon_g}{\mu_g} \right]^{1/3}$	(R7)
Heat transfer area	$A_p = (1 - \varepsilon_g) \frac{6}{0.72 d_c}$	(R8)
Density for oil stream	$\rho_g = \frac{F_g}{\varepsilon_g v_g A}$	(R9)
Heat conductivity	$K_g = 1 \times 10^{-6}(1.9469 - 0.374 M_w + 1.4815 \times 10^{-3} M_w^2 + 0.1028 T_g)$	(R10)
Oil viscosity	$\mu_g = 3.515 \times 10^{-8} \mu_{pr} \frac{\sqrt{M_w P_{pc}^{2/3}}}{T_{pc}^{1/6}}$	(R11)
Pseudo-reduced viscosity	$\mu_{pr} = 0.435 exp[(1.3316 - T_{pr}^{0.6921}) P_{pr}] T_{pr} + 0.0155$	(R12)
Pseudo-reduced pressure	$P_{pr} = \frac{P}{P_{pc}}$	(R13)
Pseudo-reduced temperature	$T_{pr} = \frac{T_g}{T_{pc}}$	(R14)

To validate the optimization results, the optimized process conditions obtained from the hybrid model were input into the rigorous mechanistic model to calculate the product distribution, as shown in Table R6. There is a good agreement between the fraction yield and bulk properties predicted by the hybrid model and the rigorous mechanistic model, demonstrating the accuracy of the hybrid model.

Table R6. Tuned process conditions and product distribution

	Experiment data	Tuned data	Model validation
	Process conditions		
Feedstock temperature, °C	70	78	78
Catalyst temperature, °C	678	670	670
CTO	5.6	5.7	5.7
Reaction time, s	3.6	3.6	3.6
	Bulk properties		
LPG yield, wt%	11.8	12.1	12.2
Gasoline yield, wt%	80.3	80.6	80.5
Coke yield, wt%	2.5	1.8	1.8
i-Paraffin yield, wt%	33.9	34.2	34.2
Olefin yield, wt%	18.2	17.2	17.2
Aromatics yield, wt%	33.9	33.9	33.9
Density at 293K, °C		0.7428	0.7432
ASTM D86 IBP, °C		26.2	25.8
ASTM D86 50%, °C		84.0	84.8
ASTM D86 FBP, °C		171.7	170.3
RON		89.8	91.2
Pc, kPa		3,260.2	3,287.3
Tc, °C		271.35	269.0

Reference:

1. Zhengyu Chen, et al. Molecular-Level modeling for naphtha olefin reduction in FCC subsidiary Riser: From laboratory reactor to pilot plant. *Chemical Engineering Journal* 437, 135429 (2022).

Modification:

We noted that 15 sets of pilot-scale experimental data were not used for model training in the revised manuscript, and the revised text can be found in the second paragraph of the “Deep Transfer Learning from Laboratory-Scale to Pilot-Scale” section.

The predicted product distribution of naphtha 3 was added to Fig. 8.

The details of the pilot-scale molecular-level mechanistic model were added to Supplementary Note 13 in the revised Supplementary Information.

The predicted product distribution by the mechanistic model was added to Table 1 of the revised manuscript.

Comment 3:

Pilot plant data is used in industry not only to predict the molecular composition on a larger scale but, very importantly, to assess the effect of the impurities in the feedstock material, study catalyst deactivation and selectivity over time. How could the proposed method be used to answer these critical questions?

Response:

We agree with the reviewer's comment. In addition to predicting the molecular composition of the product, other information, such as impurities, catalyst deactivation, and selectivity over time, is also very important. To address this, we discussed the above issues one by one.

- Impurity effect：

In the naphtha FCC process, coke is a critical impurity. It is both a by-product and a contaminant. It is mainly composed of polycyclic aromatic hydrocarbons (PAHs) and forms through the dehydrogenation and condensation of aromatic compounds present in the feedstock. Once formed, coke readily deposits on the active sites of the catalyst, leading to catalyst deactivation.

In the mechanistic model, the formation pathway of coke and its impact on catalytic activity were described. The reaction pathways of coke formation were generated by reaction rules for ring condensation, as listed in Supplementary Table 2, and detailed reaction pathways are provided in Supplementary Data 1 and 2.

Moreover, the effect of coke on the catalyst is further discussed in the following section ("Catalyst deactivation"). Because the mechanistic model can accurately describe the formation of coke impurities, it ensures that the pilot-scale hybrid model trained on this mechanistic data can also accurately calculate the distribution of coke. Figure R6 presents the predicted coke content under various process conditions and feedstock compositions, demonstrating that the model effectively captured the relationship between molecular feed characteristics and coke formation.

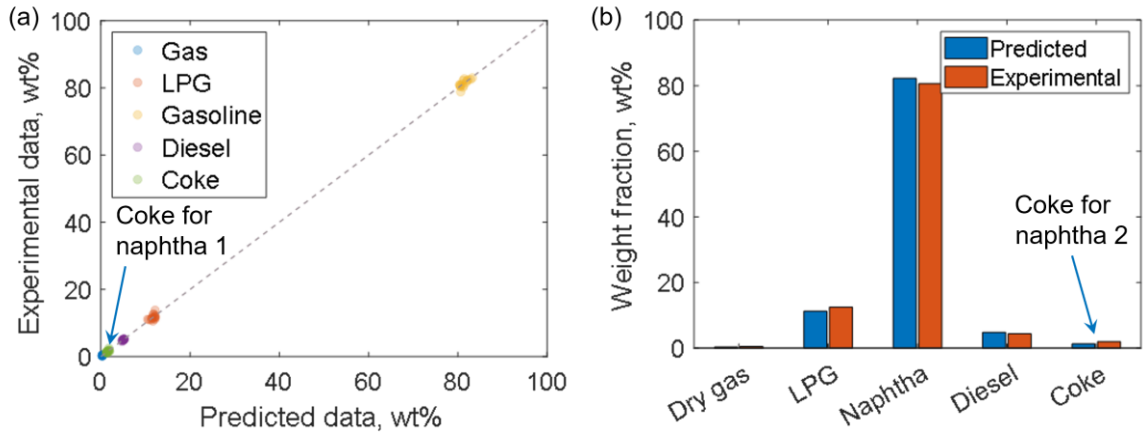

Figure R6. Predicted results for impurities (coke) distribution under various reaction conditions and feedstock. (a) Coke predicted results for naphtha 2 FCC under various reaction conditions; (b) Coke predicted results for naphtha 3 FCC.

- Catalyst deactivation:

Catalyst deactivation was explicitly considered in the development of the laboratory-scale mechanistic model and incorporated into the mass balance equation, as shown in Eqs. (R15) and (R16). Eq. (R15) represents the mass balance, and Eq. (R16) describes catalyst deactivation caused by coke deposition on the catalyst surface.

$$\frac{dC_i}{dX} = \frac{PM_w}{S_w RT} r_i \varphi_c \tag{R15}$$

$$\varphi_c = (1 + \alpha C_C)^{-\beta} \tag{R16}$$

In Eq. (R16), φ_c is the deactivation factor, and C_C represents the coke content per unit mass of catalyst and can be calculated using Eq. (R17). Model parameters (α and β) were previously determined through regression of experimental data and have values of 0.48 and 1.4, respectively (Mei Yang, et al. Fuel 2021: 121226). These parameters are intrinsic to the catalyst and remain constant across reactor scales.

$$C_C = \frac{w_{coke} \cdot m_{oil}}{m_{cat}} = \frac{w_{coke}}{CTO} \tag{R17}$$

In Eq. (R17), w_{coke} represents the coke yield, m_{oil} is the mass of oil, and m_{cat} is the catalyst mass. Substituting Eq. (R17) into Eq. (R16) enables a quantitative evaluation of the effect of coke on catalytic activity. This relationship is visualized in Figure R7.

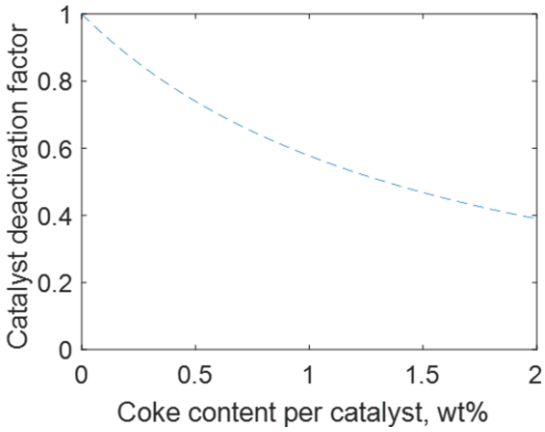

Figure R7. The effect of coke content on reaction rate

Overall, the mechanistic model developed in this work incorporated catalyst deactivation, ensuring that the pilot-scale hybrid model trained on this mechanistic data also accounted for this effect. If one intends to investigate the effect of coke on the catalyst separately, it can be directly evaluated using Eq. (R16). Additionally, it should also be noted that if catalyst properties change, the parameters (α and β) must be re-evaluated through new experimental data.

- Product yield over time:

We are also interested in the variation of product yield (selectivity) over time. However, the pilot-scale experimental data available lacked temporal distribution data, making it challenging to generate product distribution data at different time points for neural network fine-tuning. Despite this, we believe the proposed hybrid model can calculate product distributions at various reaction times. To address this, we developed a pilot-scale molecular-level mechanistic model for naphtha FCC, which allowed us to generate product data at different time intervals. The transfer learning strategy presented in this study was then applied to fine-tune the neural network using the generated data. The tuned results are shown in Table R7. Additionally, the product distribution over time was predicted by the hybrid model, as shown in Figure R8. In this figure, lines represent predictions from the hybrid model, while points denote mechanistic model calculations. The predicted results from the hybrid model show excellent agreement with the mechanistic model, demonstrating the ability of the proposed architecture to capture product variation over time. In future work, if the model can be trained by data under various reaction time, the model will enhance its capability to predict product distribution over time.

Table R7. Fine-tuned results based on temporal distribution data

Metrics	Value	Hyperparameter	Value
Training set MAE	0.03	Block/Layer number	2/2/3
Testing set MAE	0.08	Neuron number	128/512/256
Training set MSE	0.01	Learning rate	1.93E-05
Testing set MSE	0.01	Dropout parameter	0.25
		Batch size	8
		Epoch number	300
		Weight decay	4.03E-06

Note: x/x/x corresponds to process-based ResNet, molecule-based ResNet, and integrated ResNet, respectively.

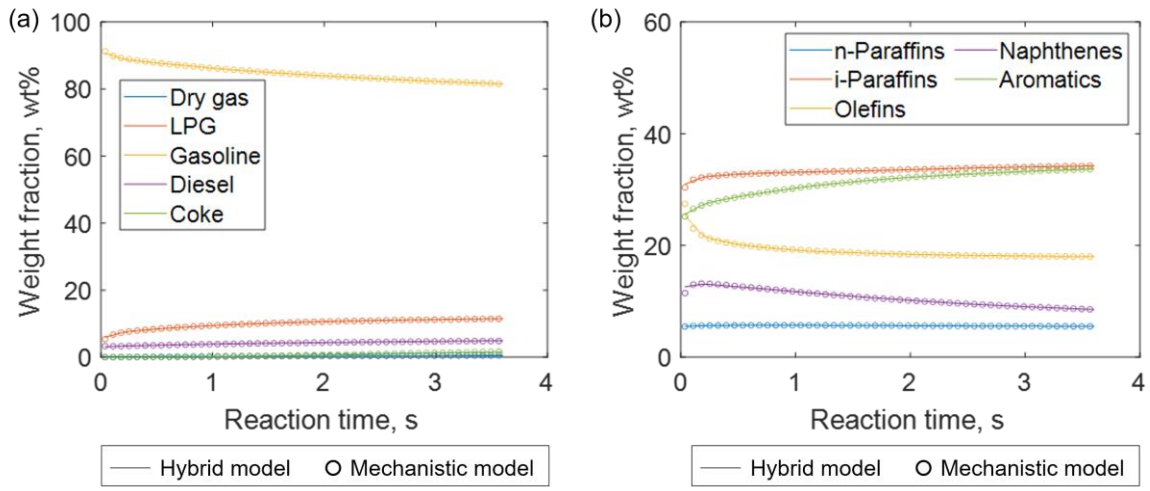

Figure R8. Predicted pilot-scale product distribution over time. (a) Fraction yield; (b) Gasoline composition.

Modification:

The discussion about impurities and catalyst deactivation is presented in Supplementary Note 4.

Product yield over time was added to Supplementary Note 11.

Comment 4:

An important feature of the proposed algorithm is that it uses an AI-based description of mass transfer, making the entire modeling task less resource-intensive. However, the author has chosen a very simple problem from a mass-transfer perspective (high-temperature gas phase processes are typically not mass-transfer limited). How can

the authors guarantee that the same approach would work for liquid-phase reactions or other systems that are indeed mass transfer limited?

Response:

We agree with the reviewer's comment that the impact of mass transfer on the FCC process is not very significant. However, the FCC reactor configuration undergoes substantial changes from the laboratory to the pilot scale. Although the variation in mass transfer is relatively minor, the changes in fluid flow behavior are considerable, which also significantly affect the reaction process. Our proposed approach can accurately predict pilot-scale product distribution for various process conditions and feedstock composition. The results further demonstrate that the approach can simultaneously capture the characteristics of both mass transfer and fluid flow behavior.

For chemical processes, the fundamental theories of gas-solid and liquid-solid mass transfer are the same. The primary difference lies only in calculation methods and the calculated value of the mass transfer coefficient. Neural networks exhibit strong fitting capabilities for such nonlinear data. Thus, the gas-solid reaction system can be calculated, the method can theoretically also be applied to liquid-phase reactions.

In our previous work,[1-4] we have developed numerous molecular-level mechanistic models for liquid-solid or gas-liquid-solid catalytic reaction systems, such as hydrofining and hydrotreating. We found that during scale-up, the modification strategies for liquid-solid or gas-liquid-solid mechanistic models were the same as those for gas-solid mechanistic models. In the above scenarios, the cross-scale computation from laboratory to pilot/industry plant can be achieved by fine-tuning the kinetic parameters. The difference between gas-solid and liquid-solid reactions lies mainly in the different coefficients or weights used to fine-tune the kinetic parameters. In the future, we will further apply this method to liquid-phase and other mass transfer-limited reaction systems to further validate the model's generalizability.

Reference:

1. Zhengyu Chen, et al. Molecular-level kinetic modelling of fluid catalytic cracking slurry oil hydrotreating. *Chemical Engineering Science* 195, 619-630 (2019).
2. Zhengyu Chen, et al. A molecular-level kinetic model for high-efficient aromatic-enriched oil conversion via the catalytic cracking and hydrofning processes. *Chemical Engineering Journal* 500, 157127 (2024).
3. Zhengyu Chen, et al. A mass-temperature decoupled discretization strategy for large-scale molecular-level kinetic model. *Chemical Engineering Science* 249, 117348 (2022).
4. Dong guan, et al. Molecular-level heavy petroleum hydrotreating modeling and comparison with high-resolution mass spectrometry. Fuel 297, 120792 (2021).

Comment 5:

The article is very repetitive in description, and thus, it isn't easy to access the most important information. E.g., this entire paragraph offers a very basic introduction to the topic and does not need to be repeated as late as line 104: "Scaling up a novel chemical process from the laboratory to industrial application typically involves three stages: laboratory experiments, pilot-scale testing, and industrial trials. Each stage serves distinct objectives, often requiring reactors with different configurations. "All writing should be reviewed and edited to be more concise.

Response:

Thank you for pointing out this issue. The basic introduction mentioned by the reviewer has been deleted in the revised manuscript. Moreover, to make the manuscript more concise, the entire manuscript was reviewed and edited.

Modification:

The sentence mentioned by the reviewer was deleted in the revised manuscript, and the entire manuscript was reviewed and edited for brevity.

Comment 6:

The authors left a number of unchecked boxes on their checklist documents, and the review suggests double-checking if that's accurate.

Response:

We've double-checked and modified the checklist, and the remaining unchecked boxes are not relevant to this work.

Modification:

Checklist documents were checked and modified in the revised version.

Comment 7:

The code is accessible, yet it consists of many separate files and it is unclear which one to run to check the reported results.

Response:

To make the reader better check our results, we have added detailed instructions for running the code in the "README" file. In addition, we have also programmed a web version of the interface to check the results. The two methods are as follows:

- Check the results through the code

All codes and datasets used in this study have been uploaded to a GitHub repository and organized into three folders, corresponding to the laboratory-scale model, pilot-scale model,

and process optimization, as illustrated in Figure R9.

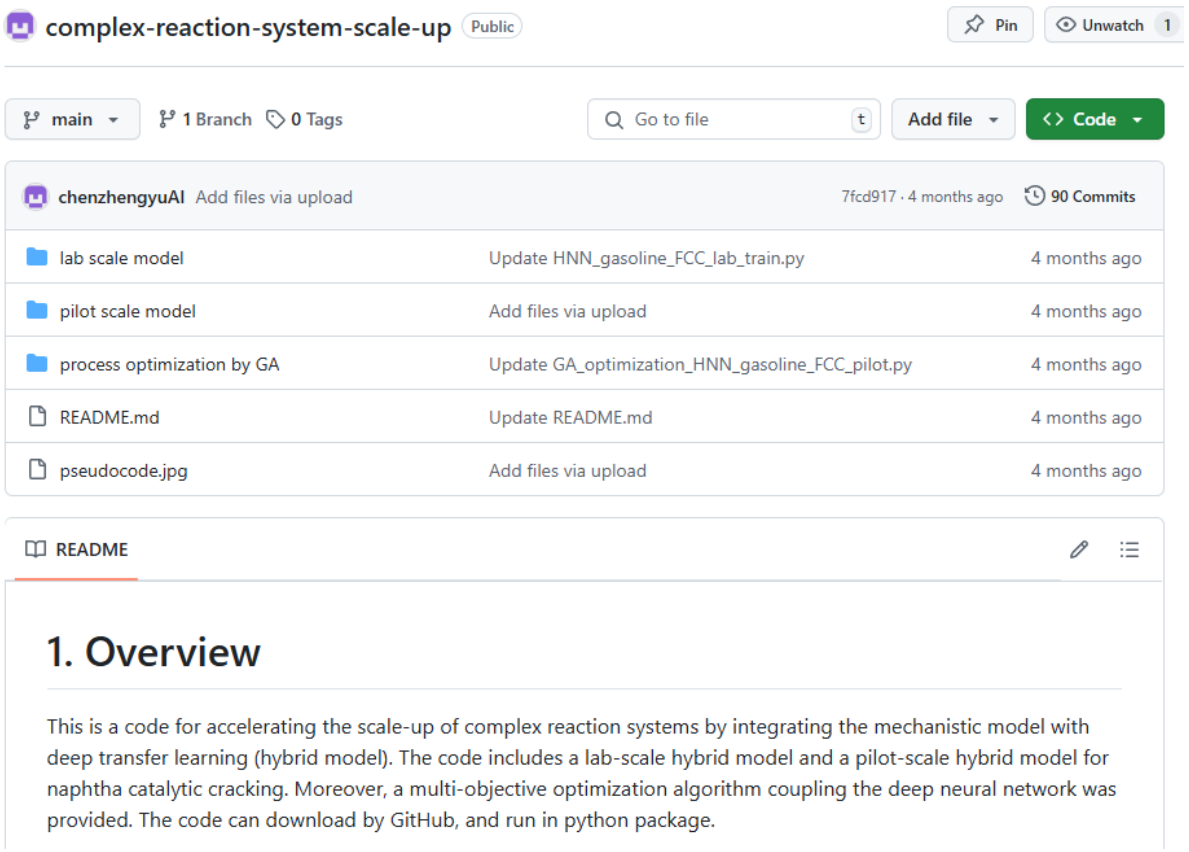

Figure R9. Screenshot of the GitHub page for this working

1. Laboratory-scale model validation

To validate the lab-scale model, please navigate to the "lab scale model" folder in Figure R9 and run the script "HNN_gasoline_FCC_lab_pred.py". The datasets for training and validation are provided in .csv files.

Note:

- Lines 93–94: Inputting "input_train_lab.csv" and "targets_train_lab.csv" will generate results for the laboratory-scale model training set.
- Lines 93–94: Inputting "input_test_lab.csv" and "targets_test_lab.csv" will generate results for the laboratory-scale model test set.

2. Pilot-scale model validation

To validate the pilot-scale model, open the "pilot scale model" folder in Figure R9 and execute the script "HNN_gasoline_FCC_pilot_pred_prop.py". The training, validation, and testing datasets are also provided in .csv format.

Note:

- Lines 104–105: Inputting "input_train_pilot.csv" and "targets_train_pilot.csv" will

generate results for the pilot-scale model training set.

- Lines 104–105: Inputting “input_validation_pilot.csv” and “targets_validation_pilot.csv” will generate results for the pilot-scale validation set.
- Lines 104–105: Inputting “input_test_pilot.csv” and “targets_test_pilot.csv” will generate results for the pilot-scale testing set.

3. Process optimization validation

To validate the process optimization results, go to the “process optimization by GA” folder shown in Figure R9 and run the script “GA_optimization_HNN_gasoline_FCC_pilot.py”. The relevant datasets are included in the folder as .csv files.

- Check the results through the interface

We programmed a web version of the interface based on Streamlit. The code to run the interface is uploaded to GitHub. “APP_lab_model.py” in the “lab scale model” folder is an interface for the laboratory-scale model, while “APP_pilot_model.py” in the “pilot scale model” folder is an interface for the pilot-scale model.

For example, if one wants to run the “APP_lab_model.py”, one can execute the command “streamlit run APP_lab_model.py” in the Python environment to open the corresponding web page, as shown in Figure R10. The left side in Figure R10 is used for data input, such as “ input_train_lab.csv” and targets_train_lab.csv”. The right side of Figure R10 shows the model training results.

Similarly, if one wants to run the “APP_pilot_model.py”, one can execute the command “streamlit run APP_pilot_model.py” in the Python environment to open the corresponding web page. “input_train_lab.csv” and “targets_train_lab.csv” should be input on the left side of the interface, and the calculated results will be presented on the right side of the interface.

Overall, readers can check the results of the model by using the two methods described above, and both methods of running the program have been added to the "README.md" file in Figure R9.

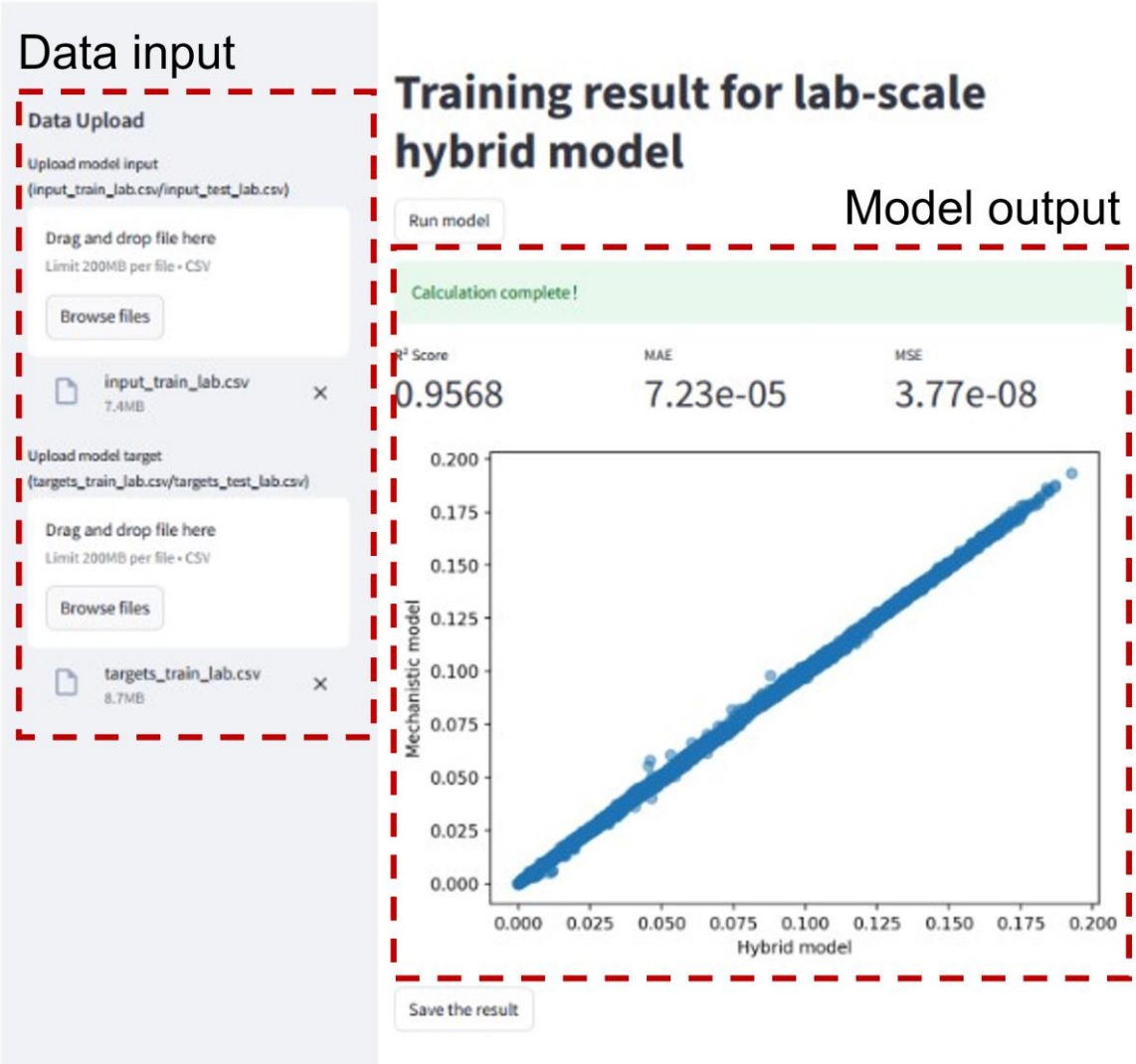

Figure R10. Screenshot of the interface for the calculated results of the laboratory-scale hybrid model

Modification:

The above content was added to sections 6 and 7 of the "README.md" file in the GitHub repository.

Response to Reviewer 3:

Comment 1:

Different communities use different definitions for "hybrid model". The paper should make the definition clear. In particular, is the hybrid model composed of both mechanistic and machine-learning-based components, or is it a pure machine learning model where the training step combines both experimental data and simulated/generated data (e.g., surrogate model with data augmentation) or both. The paper should make this distinction clear to the reader early in the article.

Response:

Thank you for pointing out this issue. We strongly agree that the "hybrid model" should be clearly defined at the beginning of the article. As the reviewer points out, there are two primary approaches for developing hybrid models: surrogate/data-driven models and hybrid models integrating mechanistic components and machine-learning-based components. In fact, both approaches were implemented in this work:

1. For the laboratory-scale process, a data-driven model was employed to accurately calculate product distributions. The approach is to generate molecular conversion data by a molecular-level mechanistic model and train the neural network with the generated data. In the revised manuscript, we will point out that the laboratory-scale hybrid model is a data-driven model.

2. For the pilot-scale process, a hybrid model combining mechanistic equations and machine learning was developed. In this case, we incorporated the mechanistic equation of bulk property directly into the neural network. On this basis, partial neural network parameters were fine-tuned using augmented pilot-scale data, enabling the model to accurately calculate the product bulk properties.

Moreover, we also introduced two methods in the introduction section:

"AI-mechanism hybrid models have emerged as a promising approach to address these challenges. Integrating mechanistic models directly into neural networks to build hybrid models has attracted broad attention, such as physics-informed neural network (PINN) and neural ordinary differential equation (Neural ODE). These methods have also been applied in some simple reaction systems. However, extending them to more complex systems remains difficult. Incorporating a large number of ODEs into neural networks still poses significant challenges for model training. An alternative hybrid modeling strategy is data-driven models. This model is trained using data generated from mechanistic models. Hough et al. utilized the artificial neural network (ANN) to build a data-driven model for biomass pyrolysis, facilitating

a rapid solution for large-scale reaction networks. Qiu and collaborators introduced a data-driven model for the naphtha steam cracking process based on the convolutional neural network (CNN). Compared to ANN, CNN can recognize features in the reaction network and accelerate computations by approximately 300 times."

Modification:

An introduction of the two hybrid models was added to the fourth paragraph of the Introduction section in the revised manuscript.

Moreover, we also explained that the laboratory-scale hybrid model is a data-driven model, while the pilot-scale hybrid model incorporates a mechanistic model in the last paragraph of the Introduction section.

Comment 2:

Related to the previous comment, Lines 97-98: "The mechanistic model served as a data generator, producing a large dataset to train the neural network. On this basis, network parameters were fine-tuned using the pilot plant data to calculate the FCC product distribution accurately with limited data samples." Did the authors consider embedding the mechanistic model in the final form instead of only using it for training purposes?

Response:

Yes, we did consider the possibility of directly embedding mechanistic models into the final model. However, due to model complexity, this approach has not been implemented in this work for the following reasons:

Embedding ordinary differential equations (ODEs) into neural network loss functions to develop hybrid models, such as Physics-Informed Neural Networks (PINN), is an innovative approach. However, naphtha FCC is an extremely complex reaction system, and the reaction network comprises 129 species and 757 reactions (Supplementary Data 1 and 2). The molecular-level kinetic model derived from this network not only contains 129 ODEs but also involves numerous equations for calculating product bulk properties. This significantly increases computationally intensive, and addressing such computational bottlenecks through surrogate modeling is also one of the primary objectives of this work. If one intends to embed these ODEs and numerous equations for calculating bulk properties into neural networks, it will make training the model highly challenging and computationally expensive. Therefore, this study does not integrate the full mechanistic model into the final model.

However, we also desire to build a hybrid model that contains some mechanistic information. To achieve this, we embedded equations for bulk property calculations into the

loss function. It allows for directly computing product bulk properties without overcomplicating the training process. In future work, we also plan to develop a new hybrid model that integrates the molecular-level kinetic model into the neural network framework.

Modification:

We explain why the work did not integrate all the mechanistic models into the neural network in the fourth paragraph of the introduction section.

Moreover, in the last paragraph of the Discussion section, we point out that hybrid models embedding mechanistic models into neural networks will be constructed in the future.

Comment 3:

Also, is the "bulk property calculator" used only during training, or is it embedded when the model is used for optimization. The paper would benefit from further clarification of the model structure and additional detail on the training procedure used.

Response:

During model training and process optimization of the pilot-scale unit, the bulk property calculator was embedded within the neural network. However, for the laboratory-scale hybrid model, the bulk property calculator was not integrated. The primary reason is that laboratory-scale experimental data provide detailed product molecular composition, and we can use the molecular composition data to train the model. For the pilot/industrial plant, experimental data are often limited by analytical instrumentation, and only partial bulk properties can be provided. Therefore, to calculate and optimize the product composition of the pilot plant, we embedded the bulk property calculator into the neural network during the transfer learning.

The model structure and training details were provided as follows:

- Model structure

The proposed hybrid model consists of three multilayer fully connected neural networks. The first network takes process conditions as inputs, including feedstock temperature, reaction temperature, reaction time (WHSV), and CTO. The second neural network takes molecular composition as input, and there are 129 molecules in the naphtha. The outputs of these two networks are then concatenated and delivered into the third network, which predicts the distribution of 129 product molecules. Each hidden layer uses the ReLU activation function, followed by a dropout layer to prevent overfitting. Naphtha FCC involves an extremely complex molecular reaction network, requiring a deep neural network to capture the nonlinear relationships between feedstock and product molecules. However, increasing network depth often leads to gradient vanishing or explosion. To overcome this limitation, residual connections were introduced between the two fully connected layers to form a ResNet.

Compared to other neural networks incorporating mechanistic models, the ResNets allow for efficient forward and backward propagation, and the training time can be significantly reduced. The ResNet was built using the PyTorch framework in Python. The detailed model structure optimized by the BO algorithm is shown in Figure R11.

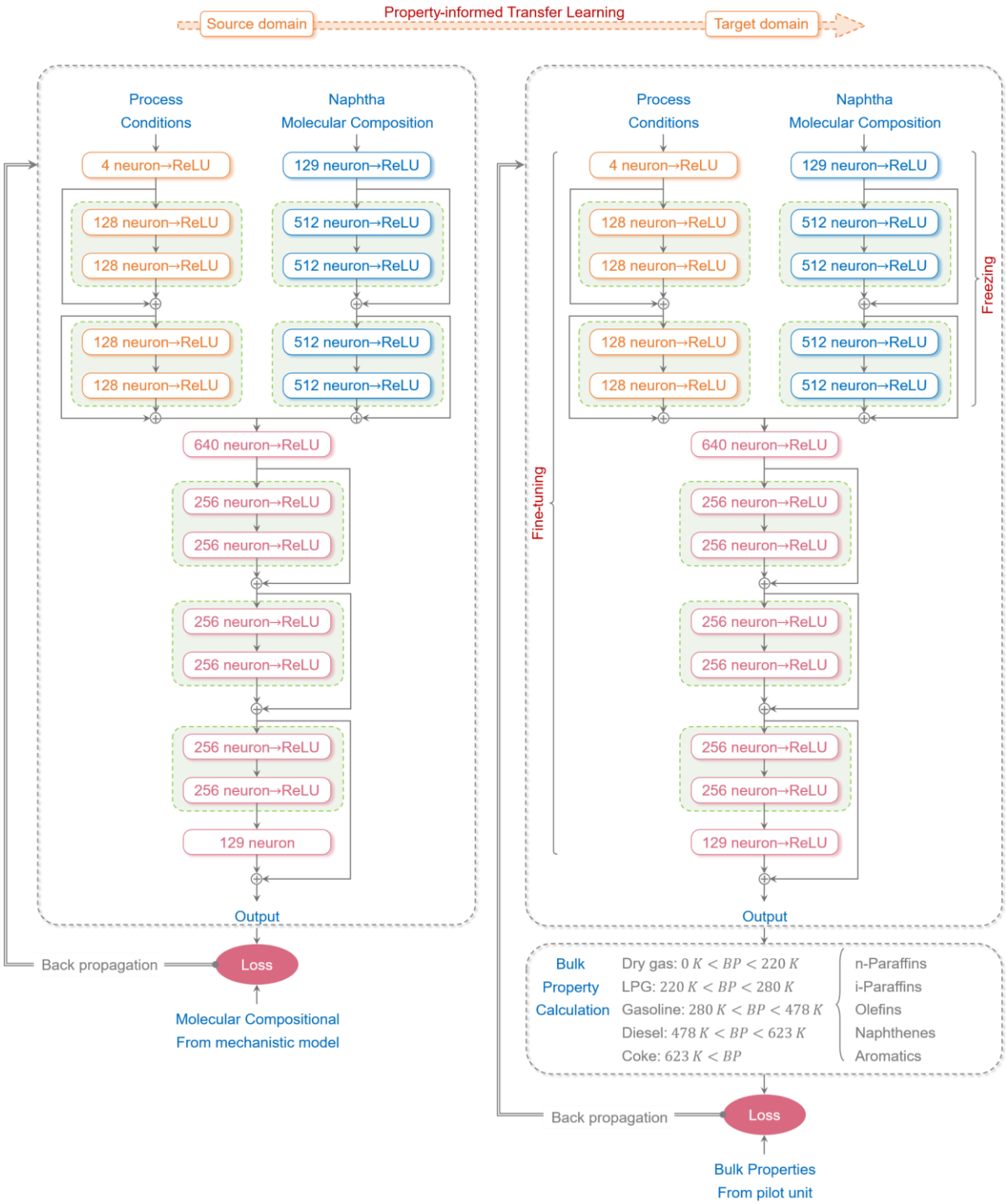

Figure R11. Optimal ResNet architecture of the hybrid model by the BO algorithm. On the left side of the figure is the laboratory-scale hybrid model, while on the right side of the figure is the pilot-scale hybrid model

This architecture enabled the calculation of product molecular composition by inputting feedstock molecular composition and process conditions. However, during chemical process scale-up, experimental data from pilot or industrial plants are often constrained by analytical instrumentation, providing only partial key bulk properties instead of detailed product molecular composition. To address this limitation, we incorporated bulk property equations (such as fraction yield and gasoline composition) into the loss function during transfer learning. This model structure can both calculate bulk properties and predict detailed product molecular composition.

The code has been uploaded to https://github.com/chenzhengyuAI/complex-reaction-system-scale-up.

Laboratory-scale model structure code: HNN_gasoline_FCC_lab_pred.py

Pilot-scale model structure code: HNN_gasoline_FCC_pilot_pred.py

- Model Training

1. Training for laboratory-scale hybrid model: The laboratory-scale hybrid model was trained on 10,000 sets of data generated by the mechanistic model. The data were randomly divided into training and testing sets in the ratio of 8:2. All input data (feedstock molecular composition and process condition) were normalized to the range [-1, 1]. The normalized feedstock molecular composition and laboratory-scale process conditions were fed into Molecule-based ResNet and Process-based ResNet, respectively. The outputs of two ResNets were concatenated and delivered into the Integrated ResNet to calculate the product molecular composition. The molecular composition calculated by ResNet and the mechanistic model were compared in the MSE loss function to compute the error. Then, backpropagation was performed, and the Adam optimizer with weight decay was used to adjust the ResNet parameters. A learning rate decay strategy was used, with a decay rate of 0.9 and a decay step of 10 epochs. This work applied three error metrics to evaluate the model performance, including R^2 score, MAE, and MSE, as shown in Eqs. (R18) ~ (R20). The training was stopped after 300 epochs, and the laboratory-scale hybrid model was obtained.

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_{cal} - y_{true})^2}{\sum_{i=1}^{n}(\overline{y_{cal}} - y_{true})^2} \tag{R18}$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_{cal} - y_{true}| \tag{R19}$$

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_{cal} - y_{true})^2 \tag{R20}$$

The code has been uploaded to https://github.com/chenzhengyuAI/complex-reaction-system-scale-up.

Training code: HNN_gasoline_FCC_lab_train.py".

2. Training for pilot-scale hybrid model: Although 15 sets of pilot-scale experimental data were obtained, training a robust model directly with such a limited dataset was a challenging task. To overcome this limitation, data augmentation was performed, and pilot-scale product data were generated through interpolation within the range of available pilot data. The augmented data were randomly split into training and validation sets with an 8:2 ratio, while the original 15 sets of experimental data were reserved as the testing set and excluded from training. All input data (feedstock molecular composition and process condition) were normalized to the range [-1, 1]. The normalized feedstock molecular composition and pilot-scale process conditions were input into Molecule-based ResNet and Process-based ResNet, respectively. Since the feedstock molecular composition and catalyst remained almost unchanged during the scale-up from laboratory to pilot plant, the network parameter of Molecule-based ResNet was frozen, while the Process-based and Integrated ResNets were fine-tuned. Then, the calculated molecular composition was input into the bulk property equations to estimate the product yield and gasoline composition. The calculated and measured bulk properties were compared in the MSE loss function to compute the error. Similar to training the laboratory-scale hybrid model, backpropagation was performed, and the Adam optimizer with weight decay was used to fine-tune the ResNet parameters. The performance of the transfer learning was evaluated using the R^2 score, MAE, and MSE. The fine-tuning was stopped after 300 epochs, and the pilot-scale hybrid model was obtained. Moreover, the bulk property calculator remained embedded in the final model during subsequent process optimization for the pilot plant.

The code has been uploaded to https://github.com/chenzhengyuAI/complex-reaction-system-scale-up.

Training code: HNN_gasoline_FCC_pilot_train.py

Optimization code: GA_optimization_HNN_gasoline_FCC_pilot.py

Modification:

Model structure and details for model training were added to the Method section of the revised manuscript.

Optimal ResNet architecture using the BO algorithm was added to the revised Supplementary Information, as shown in Supplementary Fig. 2.

Comment 4:

From Figure 2 (a), we see that the proposed model structure separates components for process conditions and feedstock. However, it appears that the only outputs are after

the "Integrated ResNet". How do we know that this structure is better than a single neural network? Have the authors compared against this simpler model structure? Please clarify the benefits of this structure and provide additional details if there are data corresponding to outputs from the "Process-based" and "Molecule-based" ResNets.

Response:

We understand the reviewer's concern about the proposed model structure. To validate the model structure, we compared the training results of single neural networks and highlighted the advantages of our hybrid structure. Additionally, we have provided the outputs from the "Process-based" and "Molecule-based" ResNets. We hope these discussions will be helpful to address your concerns.

- Compared the training results with single neural networks

We compared the training results of single neural networks, as shown in Table R8. All hyperparameters in the neural network were optimized using the Bayesian optimization algorithm, with the number of optimization iterations set to 100. Compared to the training results of the single neural network, the network structures proposed by this work can obtain better training performance.

Table R8. Training results for various neural network structures

	This work	Single ResNet	Single neural network
Training set R^2	0.956	0.912	0.737
Testing set R^2	0.953	0.910	0.736
Training set MAE	7.23E-05	1.13E-04	2.80E-04
Testing set MAE	7.75E-05	1.15E-04	2.78E-04
Training set MSE	3.85E-08	9.44E-08	7.43E-07
Testing set MSE	4.33E-08	9.50E-08	7.27E-07
	Hyperparameter		
Block/Layer number	2/2/3	4	2
Neuron number	128/512/256	256	1st layer: 128 2nd layer:256
Learning rate	3.47E-05	2.04E-05	1.22E-05
Dropout parameter	0.14	0.27	0.15
Weight decay	1.61E-06	7.11E-06	4.64E-06
Batch size	16	16	16
Epoch number	300	300	300

Note: x/x/x corresponds to process-based ResNet, molecule-based ResNet, and integrated ResNet, respectively.

- Advantages of our hybrid structure

In addition to having higher computational accuracy than simple neural networks, the proposed method can also flexibly fine-tune and freeze network parameters during the transfer learning. The specific transfer learning strategy is as follows:

1. When scaling up a chemical process, researchers typically maintain the same feedstock and catalyst while adjusting reactor size, type, or process conditions. In this case, since the feedstock remains unchanged, the molecule-based ResNet can be fixed, and only the process-based and integrated ResNets require fine-tuning.

2. When reactor and process conditions were determined, researchers often investigated the product distribution under different feedstocks. In this case, the process-based ResNet should remain unchanged, and only the molecule-based and integrated ResNets need fine-tuning.

Traditional methods cannot determine the location of fine-tuning parameters during process scale-up, and the transfer learning strategy proposed by this work is presented in Figure R12.

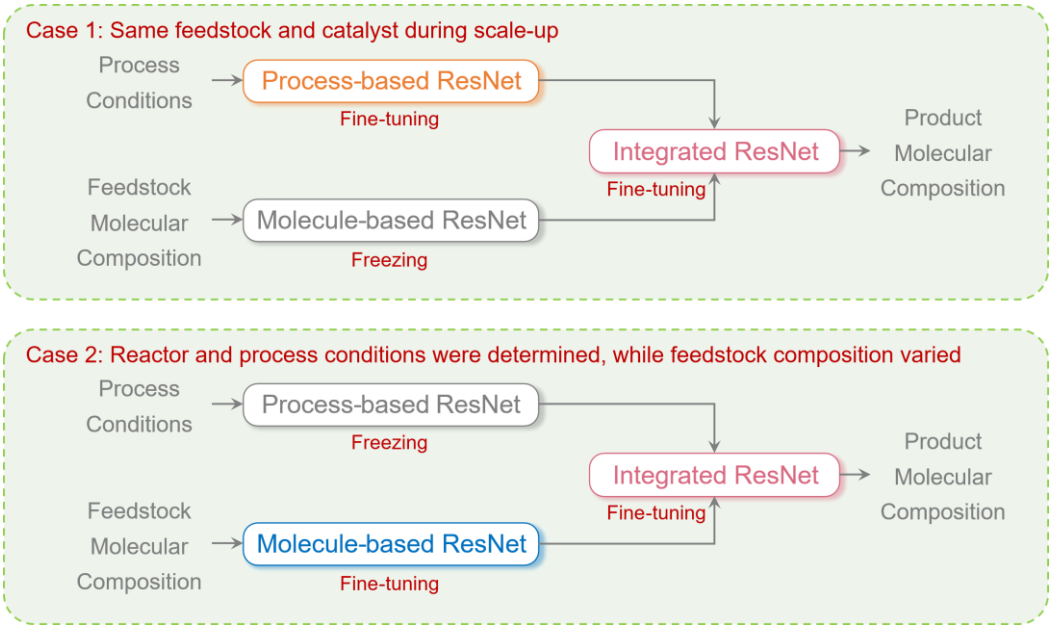

Figure R12. The transfer learning strategy proposed by this work

- Outputs from the “Process-based” and “Molecule-based” ResNets

Since the outfile is too large, we uploaded outputs of the test set from the “Process-based” and “Molecule-based” ResNets to https://github.com/chenzhengyuAI/complex-reaction-system-scale-up , Please see the “Supplementary Data” folder.

Modification:

A comparison of training results between this work and a single neural network was added in Supplementary Table 5 of the revised Supplementary Information.

Outputs from the "Process-based" and "Molecule-based" ResNets were uploaded to https://github.com/chenzhengyuAI/complex-reaction-system-scale-up.

Comment 5:

It seems that the "Property-informed Transfer Learning Strategy for Process Scale-up" has similarities with PINNs since it embeds first-principles knowledge in the loss function. Since this is an important novelty of the paper, it would be helpful if the authors could clarify the differences or similarities?

Response:

Thanks for your constructive suggestion, and the Property-informed Transfer Learning Strategy proposed in this work was inspired by Physics-Informed Neural Networks (PINNs). Similar to PINN, this method also integrated some mechanistic equations (fraction yield and gasoline composition) into the neural network. The developed pilot-scale model can directly compute product bulk properties while also predicting detailed molecular composition.

Although this method was inspired by PINN, there are still some differences between the two approaches. In PINNs, physical equations are incorporated as regularization terms in the loss function, as shown in Eq. (R21). Our method directly calculates bulk properties based on the molecular composition derived from the integrated ResNet. The calculated and measured bulk properties are then compared using a mean squared error (MSE) loss function, as shown in Eq. (R22). Compared to the PINN, our approach does not introduce additional regularization terms into the loss function.

$$\mathcal{L} = \mathcal{L}_{data} + \mathcal{L}_{physics} \tag{R21}$$

$$\mathcal{L} = \mathcal{L}_{bulk\ property} = \frac{1}{n}\sum_{i=1}^{n}(y_{cal} - y_{true})^2 \tag{R22}$$

Modification:

The similarities and differences between the two models were added to the revised manuscript, and the revised text can be found in the "Property-informed Transfer Learning Strategy for Process Scale-up" section, as follows:

"This work proposed a novel property-informed transfer learning strategy, as shown in Fig. 2(b). Inspired by the PINN, we also integrated the mechanistic model (bulk property calculation equations) into the neural network. However, compared to PINN, which adds an

additional regularization term to the loss function, this method directly modified the loss function to compute bulk properties."

Comment 6:

It is not clear why the authors chose Residual Networks (ResNet) as the data-driven modeling paradigm instead of other approaches such as Neural ODEs and Universal Differential Equations, for example. The paper would benefit from a clearer description of this selection.

Response:

We agree with the reviewer's comment that clarifying the reasons for choosing ResNet will benefit the paper. We selected ResNet to build the hybrid model for the following reasons:

1. Accurate calculation of product molecular composition

Accurately computing laboratory-scale product molecular composition is essential for predicting and optimizing pilot-scale product distribution by transfer learning. Naphtha FCC involves a highly complex molecular reaction network, requiring a deep neural network to capture the nonlinear relationships between feedstock and product molecules. However, increasing network depth often leads to gradient vanishing or explosion. ResNet addresses these challenges through residual connections, ensuring numerical stability and robustness. Previous studies (Xin Zhou, et al. AIChE Journal, 2023; 69(7): e18083.) have demonstrated that ResNet can accurately predict product distributions in complex chemical processes, making it a suitable choice for this study.

2. Computational efficiency for complex reaction systems

The molecular-level mechanistic model contains 129 ODEs, and the model solution is time-consuming. Searching for a neural network architecture that can quickly compute the product distribution is important for process optimization. However, Neural ODEs and Universal Differential Equations require embedding ODEs into the neural network and solving them by numerical methods, such as Runge-Kutta. It is a challenging task to embed 129 ODEs into a neural network and train them in a short time. In contrast, the discrete layer structure of ResNet allows for efficient forward and backward propagation, and the training time can be significantly reduced compared to the ODE-based model.

3. Code Reproducibility and Accessibility:

A key objective of this study was to develop a model that is easily reproducible and accessible to researchers in chemistry and chemical engineering. These researchers may have limited expertise in machine learning. Implementing Neural ODEs or Universal Differential Equations requires specialized knowledge, making reproducibility challenging. In contrast,

mainstream deep learning frameworks, such as PyTorch and TensorFlow, provide user-friendly tools for building ResNet architectures. Researchers with basic AI knowledge can also reproduce and extend this work.

Overall, considering computational accuracy, computational efficiency, and code reproducibility, we selected ResNet to build the hybrid model. In future work, we will also develop a hybrid model based on Neural ODEs and Universal Differential Equations.

Modification:

Reasons for choosing ResNet were added to the Method section of the revised manuscript, and this text can be found in the "Laboratory-scale Hybrid Model Structure" section.

We explain why the work did not choose other approaches, such as Neural ODE, in the fourth paragraph of the Introduction section.

Comment 7:

Lines 198-200: "To validate the proposed hybrid model, a molecular-level mechanistic model for the naphtha FCC process was developed, and cross-scale computation from the laboratory to the pilot plant was performed." Please provide more details here. It can be very challenging to carry a molecular-level model from laboratory to pilot scale, and it is not clear how "cross-scale computation" was performed.

Response:

We appreciate the reviewer's comments, this sentence in the manuscript was not sufficiently clear and could potentially lead to misunderstanding. What we are trying to say is as follows:

To validate the proposed hybrid model, we selected the naphtha FCC as a case study. A molecular-level mechanistic model was developed as a data generator to produce large-scale laboratory data for training the deep neural network. On this basis, a property-informed transfer learning strategy was employed to fine-tune partial network parameters, and cross-scale computation from the laboratory to the pilot plant was performed.

Notably, in the revised manuscript, we deleted this sentence to make the paper more concise.

Modification:

This sentence has been deleted in the revised manuscript.

Comment 8:

Regarding the case study, lines 249-250: "It was found that the FCC process converts olefins into i-paraffins and aromatics." If this is a new finding for these applications, then it needs further clarification. If this is being mentioned to indicate that the proposed

hybrid model is matching expectations, then this should be clarified.

Response:

Thanks to the reviewers for pointing this out. This is not a new finding, but simply demonstrates that the proposed hybrid model is matching expectations.

According to the reviewer's suggestion, this sentence was changed to "It was found that the FCC process converts olefins into i-paraffins and aromatics, and the calculated product results were in agreement with the published work. [1, 2]"

Reference:

1. Xiaolin Zhu, et al. Conceptual Fluid Catalytic Cracking Process with the Additional Regenerated Catalyst Circulation Path for Gasoline Reprocessing and Upgrading with Minimum Loss. *Energy & Fuels* 34(1), 235-244 (2020).
2. Youhao Xu, et al. A novel fluid catalytic cracking process for maximizing iso-paraffins: from fundamentals to commercialization. *Frontiers of Chemical Science and Engineering* 12, 9-23 (2018).

Modification:

The clarification of this sentence was added to the revised manuscript, and the revised text can be found in the fourth paragraph of the "Molecular-level kinetic model for Naphtha FCC" section.

Comment 9:

Lines 274-275: "The laboratory-scale hybrid model was trained, and its hyperparameters were optimized using a Bayesian optimization (BO) algorithm." The paper would benefit from a more thorough discussion of how this hyperparameter tuning was performed. If an existing package was used, then please provide those details.

Response:

We agree with the reviewer's comment that a deeper discussion of the hyperparameter optimization process will benefit the paper. This work used the Optuna package for hyperparameter tuning, and the details of hyperparameter optimization are as follows:

The proposed network architecture for the hybrid model contains a large number of hyperparameters such as residual block number, neuron number of each residual block, dropout rate, learning rate, weight decay, and batch size. To efficiently determine optimal hyperparameters, the Bayesian optimization (BO) algorithm was employed based on the Optuna package. The model training and hyperparameter optimization were performed based on NVIDIA's 4090 GPU. The hyperparameter optimization process for the laboratory-scale hybrid model is as follows:

Before hyperparameter optimization, the full dataset derived from the mechanistic model

was randomly split into training (80%) and testing (20%) sets. All input data was normalized to the range [-1, 1], and the search space for each hyperparameter was predefined, as shown in Table R9. After that, the BO algorithm was called to tune the hyperparameter. Initially, a series of candidate hyperparameters were generated based BO algorithm. The training set, testing set, and hyperparameters were input to the ResNet model for training. When the model training was completed, the R^2 score of the testing set was calculated and selected as the objective function for the BO algorithm. According to the R^2 score, the BO algorithm was repeatedly called to tune the hyperparameters. The optimization process was stopped after 100 iterations. The optimal hyperparameters were obtained from the above 100 iterations when the R^2 score was maximized. In addition to R^2, other evaluation metrics, such as MAE and root mean square error (RMSE), can also be used as the objective function in the BO algorithm.

The hyperparameters in transfer learning were also tuned by the BO algorithm. Since the neural network architecture has been determined, only four hyperparameters need to be tuned, including dropout rate, learning rate, weight decay, and batch size. Their ranges are listed in Table R10. The hyperparameter optimization process followed the same procedure as that of the laboratory-scale hybrid model.

Table R9. The hyperparameter range setting in the BO algorithm

Hyperparameter	Method	Range
Residual block number for process-based ResNet	Integer-type hyperparameters	1 ~ 2
Residual block number for molecule-based ResNet	Integer-type hyperparameters	1 ~ 2
Residual block number for integrated ResNet	Integer-type hyperparameters	1 ~ 5
Neuron number for process-based ResNet	Discrete categorical hyperparameters	128, 256, 512
Neuron number for molecule-based ResNet	Discrete categorical hyperparameters	128, 256, 512
Neuron number for integrated ResNet	Discrete categorical hyperparameters	128, 256, 512, 1024
Learning rate	Floating-point hyperparameters	0 ~ 0.5
Dropout parameter	Floating-point hyperparameters	1.0E-06 ~ 1.0E-03
Weight decay	Floating-point hyperparameters	1.0E-06 ~ 1.0E-02
Batch size	Discrete categorical hyperparameters	16, 32, 64, 128, 256

Table R10. The hyperparameter range setting in the BO algorithm

Hyperparameter	Method	Range
Learning rate	Floating-point hyperparameters	0 ~ 0.5
Dropout parameter	Floating-point hyperparameters	1.0E-06 ~ 1.0E-03
Weight decay	Floating-point hyperparameters	1.0E-06 ~ 1.0E-02
Batch size	Discrete categorical hyperparameters	4, 8, 16, 32, 64, 128, 256

Modification:

The hyperparameter optimization method for the laboratory-scale hybrid model was added to the "Hyperparameter optimization" section of the Method section in the revised manuscript.

The hyperparameter optimization method for the pilot-scale hybrid model was added to the "Parameter Fine-tuning in Transfer Learning" section of the Method section.

The hyperparameter range for the laboratory-scale model and pilot-scale model (Tables R9 and R10) was added to Supplementary Tables 13 and 14, respectively.

Comment 10:

Regarding the two-layer ResNet architecture Lines 166-168 that "This architecture mirrors the logic of the mechanistic model, facilitating more precise identification of parameter fine-tuning locations during transfer learning." It is not clear to this reviewer what is meant here or why this should facilitate "more precise identification". The authors should provide clarification here.

Response:

Our proposed network architecture employed two ResNets to separately capture features of feedstock molecular composition and process conditions, facilitating more precise identification of parameter fine-tuning locations during transfer learning. The specific transfer learning strategy is as follows:

1. When scaling up a chemical process, researchers typically maintain the same feedstock and catalyst while adjusting reactor size, type, or process conditions. In this case, since the feedstock remains unchanged, the Molecule-based ResNet can be fixed, and only the Process-based and Integrated ResNets require fine-tuning.

2. When reactor and process conditions were determined, researchers often investigated the product distribution under different feedstocks. In this case, the Process-based ResNet should remain unchanged, and only the Molecule-based and Integrated ResNets need fine-tuning.

Thus, this architecture allows precise identification of parameter fine-tuning locations based on specific scale-up requirements, as shown in Figure R12. In contrast, conventional

architecture integrates feedstock molecular composition and process conditions into a single neural network, making it difficult to determine fine-tuning locations using prior knowledge.

Modification:

The clarification of "more precise identification" was added to the revised manuscript, and the revised text can be found in the "Network Architecture of Transfer Learning for Complex Molecular Reaction Systems" section.

Comment 11:

In the optimization results, "The results indicate that increasing the feedstock temperature by 8°C and reducing the catalyst temperature by 8°C can lead to higher yields of gasoline, i-paraffins, and aromatics. Simultaneously, both olefin and coke contents are decreased". This seems like an interesting finding for improving operation. Can the authors expand on the control structure that would be used to achieve this temperature control?

Response:

We appreciate the reviewer's positive comment on our findings, and we explored the control structure to achieve temperature control according to the reviewer's suggestion. In this finding, we need to control the feedstock temperature and the regenerated catalyst temperature. The specific control strategies are as follows:

1. Feed Temperature Control

The feedstock temperature is regulated by adjusting the flow rate of the heat carrier in the preheater. A temperature indicator (TI) is installed at the preheated naphtha pipe to monitor the feed temperature. A control valve is installed at the inlet of the heat carrier. On this basis, a temperature indicating controller (TIC) was used to adjust the valve opening, increasing the feed temperature by 8°C, as illustrated in Figure R13.

2. Regenerated Catalyst Temperature Control

The regenerated catalyst temperature is related to the reaction temperature in the regenerator. To regulate the regenerated catalyst temperature, we intend to control the reaction temperature in the regenerator by changing the air flow at the regenerator inlet. A temperature indicator (TI) is installed on the sloped pipe of the regenerated catalyst to monitor the catalyst temperature, while a control valve is placed at the air pipe. Then, a temperature indicating controller (TIC) was used to adjust the valve opening to reduce the air flow rate, reducing the regenerated catalyst temperature by 8°C, as shown in Figure R13.

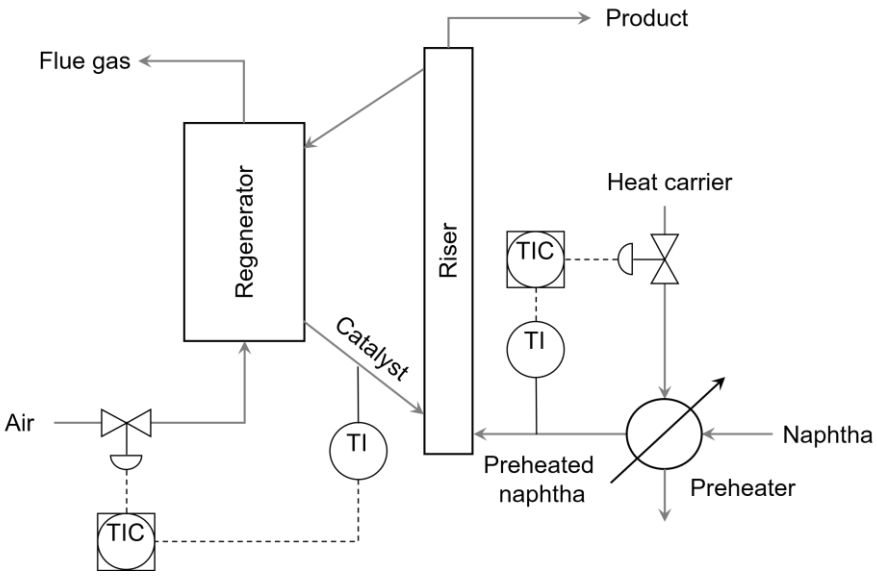

Figure R13. Control structure for feedstock temperature and regenerated catalyst temperature

Modification:

Control strategies for feedstock temperature and regenerated catalyst temperature were added to Supplementary Note 12 in the revised Supplementary Information.

Comment 12:

The optimization is performed using the hybrid model, and a single optimal operating point is found. As with any hybrid or surrogate-based approach, it is important to evaluate the error (and ideally the optimality) using a rigorous model or experimental data. The paper would also benefit from some discussion of how one might account for uncertainty in the hybrid model.

Response:

We agree with the reviewer's comments that the optimization results of the hybrid model should be validated. Therefore, we employed a rigorous mechanistic model for evaluation. In our previous work, we developed a molecular-level kinetic model for a pilot-scale naphtha FCC process. Compared to the laboratory-scale FCC model, the pilot-scale FCC model requires integrating the molecular-level kinetic model with the riser reactor model. The mass and heat balance equations in the riser model are presented in Eqs. (R23) ~ (R25).

$$\frac{dC_i}{dz} = \frac{\rho_c \varepsilon_c}{v_g} R_i \varphi_{ck} \tag{R23}$$

$$\frac{dT_c}{dz} = \frac{A h_p A_p}{F_c C_{p\_c}} \left(T_g - T_c\right) \tag{R24}$$

$$\frac{dT_g}{dz} = \frac{A}{F_g C_{p\_g}} \left[h_p A_p \left(T_c - T_g\right) + \rho_c \varepsilon_c \sum\nolimits_j \left(-\Delta H_j r_j\right) \right] \tag{R25}$$

The mass and heat balance equations in the riser reactor model involve numerous bulk properties, and their correlation formulas are shown in Table R11.

Table R11. Equations for properties in the riser model

Property	Equations	
Catalyst deactivation	$\varphi_{ck} = \frac{1}{1 + a_{ck}(C_{ck})^{b_{ck}}}$	(R26)
Volume fraction for catalyst phase	$\varepsilon_c = \frac{F_c}{v_c \rho_c A}$	(R27)
Volume fraction for oil phase	$\varepsilon_g = 1 - \varepsilon_c$	(R28)
Heat transfer coefficient	$h_p = 0.03 \frac{K_g}{d_c^{2/3}} \left[\frac{\lvert(v_g - v_c)\rvert \rho_g \varepsilon_g}{\mu_g}\right]^{1/3}$	(R29)
Heat transfer area	$A_p = (1 - \varepsilon_g)\frac{6}{0.72 d_c}$	(R30)
Density for oil stream	$\rho_g = \frac{F_g}{\varepsilon_g v_g A}$	(R31)
Heat conductivity	$K_g = 1 \times 10^{-6}(1.9469 - 0.374 M_w + 1.4815 \times 10^{-3} M_w^2 + 0.1028 T_g)$	(R32)
Oil viscosity	$\mu_g = 3.515 \times 10^{-8} \mu_{pr} \frac{\sqrt{M_w P_{pc}^{2/3}}}{T_{pc}^{1/6}}$	(R33)
Pseudo-reduced viscosity	$\mu_{pr} = 0.435 exp[(1.3316 - T_{pr}^{0.6921}) P_{pr}] T_{pr} + 0.0155$	(R34)
Pseudo-reduced pressure	$P_{pr} = \frac{P}{P_{pc}}$	(R35)
Pseudo-reduced temperature	$T_{pr} = \frac{T_g}{T_{pc}}$	(R36)

After developing the molecular-level kinetic model for the naphtha FCC model, the kinetic parameters were regressed using pilot-scale data. The developed model accurately calculated and predicted the product distribution for the pilot-scale unit, and partial results were shown in Figure R14.

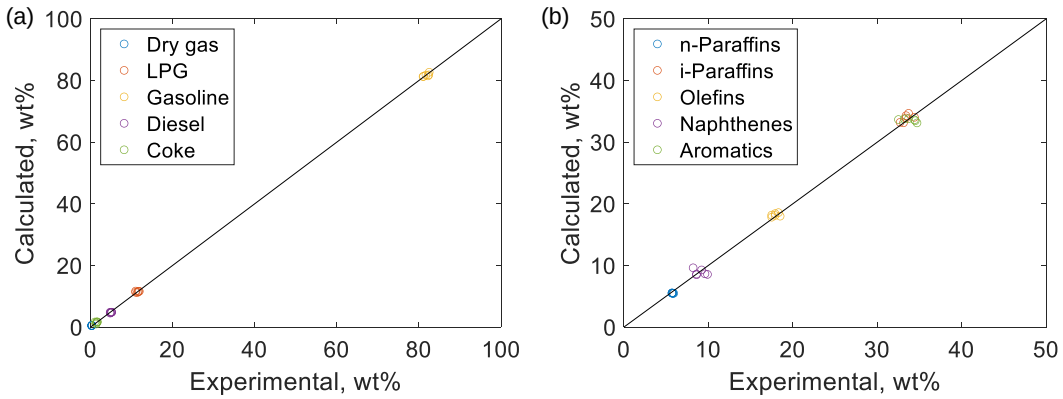

Figure R14. Calculated results for Pilot-scale mechanistic model (a) Product yield; (b) Gasoline composition

To validate the optimization results, the optimized process conditions obtained from the hybrid model were input into the rigorous mechanistic model to calculate the product distribution, as shown in Table R12. There is a good agreement between the fraction yield and bulk properties predicted by the hybrid model and the rigorous mechanistic model, demonstrating the accuracy of the hybrid model.

Table R12. Tuned process conditions and product distribution

	Experiment data	Tuned data	Model validation
	Process conditions		
Feedstock temperature, °C	70	78	78
Catalyst temperature, °C	678	670	670
CTO	5.6	5.7	5.7
Reaction time, s	3.6	3.6	3.6
	Bulk properties		
LPG yield, wt%	11.8	12.1	12.2
Gasoline yield, wt%	80.3	80.6	80.5
Coke yield, wt%	2.5	1.8	1.8
i-Paraffin yield, wt%	33.9	34.2	34.2
Olefin yield, wt%	18.2	17.2	17.2
Aromatics yield, wt%	33.9	33.9	33.9
Density at 293K, °C		0.7428	0.7432

ASTM D86 IBP, ºC	26.2	25.8
ASTM D86 50%, ºC	84.0	84.8
ASTM D86 FBP, ºC	171.7	170.3
RON	89.8	91.2
Pc, kPa	3,260.2	3,287.3
Tc, ºC	271.35	269.0

Moreover, according to the reviewer's suggestion, we performed an uncertainty analysis of the hybrid model:

1. Experimental data: The product composition analysis for both laboratory-scale and pilot-scale units relied on offline analytical methods, introducing uncertainties in experimental data acquisition. Additionally, the composition of feedstock and product fraction was analyzed using a GC-FID, following the ASTM D5134-21 standard method. The uncertainty of this analytical method is less than ±0.01 (mass fraction). The FCC products are separated using a micro distillation device, and the uncertainty is less than 1 mL.

2. Mechanistic model: The molecular-level kinetic model developed in this work decomposed reversible reactions into two irreversible reactions, which may introduce deviations in the generated molecular conversion data.

3. Hybrid model: The developed neural network may introduce uncertainties when calculating product distributions outside the boundary process conditions.

Modification:

The details on the pilot-scale mechanistic model were added to Supplementary Note 13 of the revised Supplementary Information.

The evaluation result of process optimization was added to Table 1 of the revised manuscript.

The uncertainty in the hybrid model was also added to Supplementary Note 14 of the revised Supplementary Information.

Comment 13:

The paper would benefit from a more thorough introduction on transfer learning as it relates to the task at hand. Further discussion of ResNets would also be a welcome addition.

Response:

We strongly agree reviewer's suggestion. We performed a thorough introduction on

transfer learning and further discussion of ResNets. Firstly, a thorough introduction to transfer learning in chemical engineering was summarized as follows:

The core concept of transfer learning is to transfer knowledge from a well-learned task (source domain) to a related but distinct task (target domain), enhancing model performance or reducing data requirements. It has been widely applied in chemical engineering for property prediction[1], fault diagnosis[2], and process simulation[3]. For example, Buterez et al.[1] developed a graph neural network (GNN)-based transfer learning method that leveraged readily available, inexpensive molecular property data to predict complex and expensive molecular properties. Xiao et al.[3] proposed a recurrent neural network (RNN)-based transfer learning framework for continuous stirred-tank reactors (CSTRs). They used a simplified mechanistic model to generate training data for the RNN, followed by fine-tuning to simulate industrial-scale CSTR operations. In chemical process scale-up, although apparent reaction rates vary due to changes in transport phenomena, intrinsic reaction mechanisms remain unchanged across scales. This characteristic agrees well with the principles of transfer learning. If transfer learning can effectively capture flow regime variations across scales, it is expected to solve the cross-scale computational challenges in reaction processes.

Reference:

1. David Buterez. et al. Transfer learning with graph neural networks for improved molecular property prediction in the multi-fidelity setting. *Nature Communications* 15, 1517 (2024).
2. Hao Wu, et al. Fault detection and diagnosis based on transfer learning for multimode chemical processes. *Computers & Chemical Engineering* 135, 106731 (2020).
3. Ming Xiao, et al. Modeling and predictive control of nonlinear processes using transfer learning method. *AIChE Journal* 69(7), e18076 (2023).

The advantages and limitations of ResNets were further discussed as follows:

Advantage: Compared to conventional single neural network surrogate models, the proposed ResNet-based architecture not only improves prediction accuracy but also enables efficient parameter freezing and fine-tuning during transfer learning, as shown in Table R8.

Limitation: For chemical processes dominated by transport limitations, replacing the ResNet with Physics-Informed Neural Network (PINN) or Neural ODE may enhance model accuracy, as the key transport equations can be incorporated into the neural network.

Modification:

The introduction of transfer learning was added to the fifth paragraph of the Introduction section in the revised manuscript.

The discussion of ResNets was added to the Discussion section in the revised manuscript.

Comment 14:

The paper lacks a formal mathematical description of the optimization being performed, including the constraints and objective function. This is very important for the reader to understand the specifics of problem being solved.

Response:

Thank you for pointing out this issue, and we added the details and mathematical description for process condition optimization as follows:

This study utilized the Non-dominated Sorting Genetic Algorithm II (NSGA-II) to optimize the process conditions of the pilot-scale unit, including feedstock temperature, regenerated catalyst temperature, and CTO. The maximum and minimum values of these conditions, derived from pilot experiments, were defined as constraints, as shown in Eqs. (R37) ~ (R39).

$$T_{feed}^{min} \leq T_{feed} \leq T_{feed}^{max} \quad \text{(R37)}$$

$$T_{cat}^{min} \leq T_{cat} \leq T_{cat}^{max} \quad \text{(R38)}$$

$$CTO^{min} \leq CTO \leq CTO^{max} \quad \text{(R39)}$$

After that, the NSGA-II algorithm was coupled with the hybrid model to optimize the process condition. NSGA-II initially generated a range of process conditions within the defined upper and lower bounds. These conditions were input into the trained pilot-scale hybrid model to calculate the product yield and gasoline composition. The objective functions were to maximize gasoline and i-paraffin yields, as defined in Eqs. (R40) and (R41). Then, calculated yields of gasoline and i-paraffins were fed back into the NSGA-II, and the algorithm iteratively searched for tuned process conditions. After 50 iterations, the Pareto-optimal solution was obtained, and the optimal process conditions were identified from the Pareto front. The calculation process is shown in Figure R15.

$$\min y_{gasoline} = -\sum_{i=1}^{n} x_i \quad \text{(280 K < Molecular boiling point < 473 K)} \quad \text{(R40)}$$

$$\min y_{i-paraffin} = -\sum_{i=1}^{n} x_i \quad \text{(i-paraffin molecules)} \quad \text{(R41)}$$

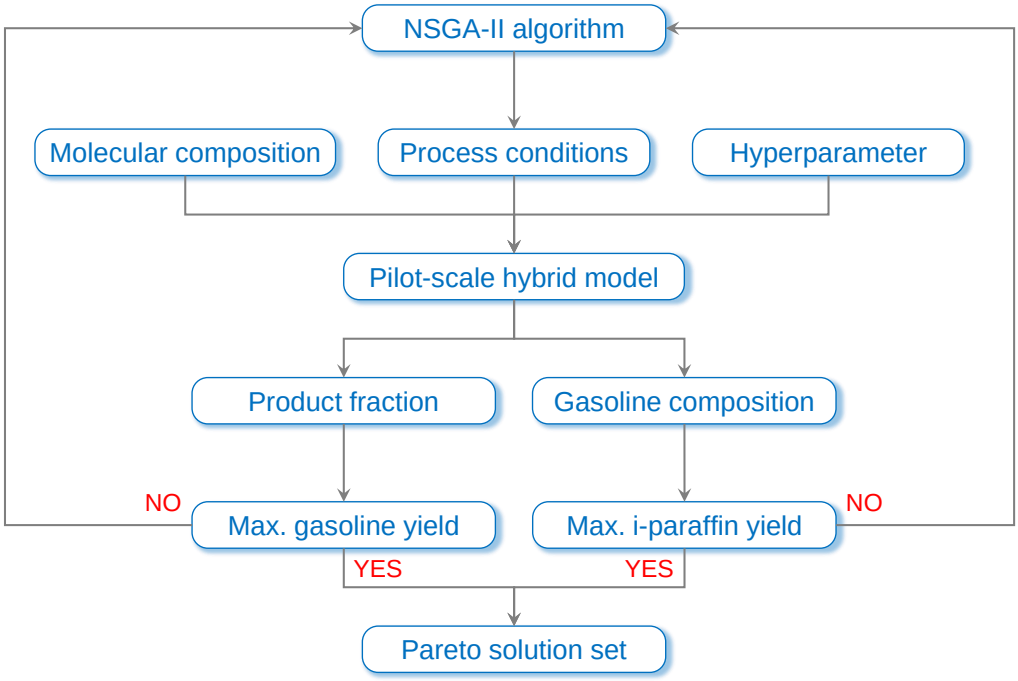

Figure R15. Calculation process for NN-NSGA algorithm

Modification:

The details about process optimization were added to the "Process Condition Optimization by NSGA-II" section in the Method section.

Figure R15 was added to the revised Supplementary Information, as shown in Supplementary Fig. 9.

Comment 15:

Lines 482-483: "In this work, R2 is used as the objective function of the BO algorithm. When R2 is maximal, the optimal hyperparameters can be obtained." Is R2 the most appropriate metric for this, especially given that the underlying phenomena is nonlinear and involving an optimization/training step? Could other metrics such as MAE, RMSE, or likelihood also be used. Can the authors clarify their choice of objective?

Response:

We understand the reviewer's concern about the choice of the objective function. According to the reviewer's suggestions, we evaluated different metrics as the objective function for the Bayesian Optimization (BO) algorithm. During the optimization process, the same upper and lower bounds for hyperparameters were applied, with the number of optimization trials set to 100. The optimal results are presented in Table R13. Although using MAE as the objective function resulted in the lowest error, its corresponding R^2 score is relatively poor. This is mainly due to the poor calculated results for some outliers when the

MAE was used as an optimization objective. According to optimization results, R^2 was selected as the objective function of the BO algorithm to ensure both accuracy and overall model reliability.

Table R13. Hyperparameter optimization results based on various metrics

	R^2	MAE	RMSE	likelihood
Training set R^2	0.956	-2.32	0.955	0.55
Testing set R^2	0.953	-2.99	0.952	0.56
Training set MAE	7.23E-05	7.08E-05	8.21E-05	3.92E-04
Testing set MAE	7.75E-05	7.61E-05	8.69E-05	3.88E-04
Training set MSE	3.85E-08	3.13E-08	5.61E-08	1.45E-06
Testing set MSE	4.33E-08	3.58E-08	6.24E-08	1.41E-06
		Hyperparameter		
Block/Layer number	2/2/3	2/2/2	2/1/4	1/1/1
Neuron number	128/512/256	512/512/1024	128/256/1024	256/512/512
Learning rate	3.47E-05	1.63E-05	9.12E-05	1.08E-05
Dropout parameter	0.14	0.01	0.01	0.03
Weight decay	1.61E-06	1.20E-06	1.02E-06	2.49E-04
Batch size	16	32	16	16
Epoch number	300	300	300	300

Note: x/x/x corresponds to process-based ResNet, molecule-based ResNet, and integrated ResNet, respectively.

For the reason of choosing R^2 as the objective function, it is mainly because R^2 is often used as an evaluation metric in some similar studies[1-3]. We agree with their work, and R^2 was also used in this work.

Reference:

1. Xin Zhou, et al. A hybrid deep learning framework driven by data and reaction mechanism for predicting sustainable glycolic acid production performance. *AIChE Journal* 69(7), e18083 (2023).
2. Jiong Du, et al. Development of hybrid surrogate model structures for design and optimization of CO2 capture processes: Part I. Vacuum pressure swing adsorption in a confined space. *Chemical Engineering Science* 283, 119379 (2024).

3. Qingchun Yang, et al. An auto-configurable machine learning framework to optimize and predict catalysts for CO_2 to light olefins process. *AIChE Journal* 70(8), e18437 (2024)

Modification:

Hyperparameter optimization results based on various metrics were listed in Supplementary Table 4 in the revised Supplementary Information.

Comment 16:

The "Transfer Learning" section needs to be more specific, providing substantially more detail about how this approach is applied in this particular problem.

Response:

We agreed with the reviewer's comment that the "Transfer Learning" section needs to provide more detail. Thus, the "Transfer Learning" section was divided into two sections, namely, "Model Structure for Transfer Learning" and "Parameter Fine-tuning in Transfer Learning".

- Model Structure for Transfer Learning

In this work, transfer learning was applied to calculate the product distribution under the pilot plant. Due to the similarity of the reaction mechanisms between the laboratory and pilot scales, transfer learning was primarily employed to capture the differences in fluid flow regimes within the reactor and calculate the pilot-scale product distribution. The source domain was laboratory-scale molecular composition data, containing 10,000 samples with 129 features. The target domain was 15 sets of pilot-scale bulk property data for naphtha 2 FCC. Due to data discrepancies between the source domain and target domain, directly calculating the pilot-scale product distribution using conventional transfer learning posed a challenge. To address this, this work incorporated some bulk property equations into the loss functions, including fraction yield and gasoline composition. The fraction yields were divided primarily based on the molecular boiling points as follows:

- Dry gas：0 K < Molecular boiling point < 220 K
- LPG：220 K < Molecular boiling point < 280 K
- Gasoline：280 K < Molecular boiling point < 473 K
- Diesel：473 K < Molecular boiling point < 623 K
- Coke：623 K < Molecular boiling point, and molecules with more than 5 aromatic rings and less than 2 carbons in the side chain.

The molecular boiling points were calculated by the group contribution method, followed by dividing them according to the scheme described above. Once the fraction yields were calculated, gasoline was further classified into n-paraffins, i-paraffins, olefins, naphthenes, and aromatics based on molecular structure. The specific classification scheme is as follows:

- n-Paraffins: Molecules without branched chains, double bonds, and rings
- i-Paraffins: Molecules with branched chains and without double bonds and rings
- Olefins: Molecules with double bonds and without rings
- Naphthenes: Molecules with naphthenic rings and without aromatic rings
- Aromatics: Molecules with aromatic rings.

According to the above method, the gasoline composition can be calculated. When the property information is incorporated into the hybrid model, the calculated property data can be directly compared to experimental data during model training.

- Parameter Fine-tuning in Transfer Learning

Although 15 sets of pilot-scale experimental data were obtained, training a robust model directly with such a limited dataset was a challenging task. To overcome this limitation, data augmentation was performed, and pilot-scale product data were generated through interpolation within the range of available pilot data. The augmented data were randomly split into training and validation sets with an 8:2 ratio, while the original 15 sets of experimental data were reserved as the testing set and excluded from training. All input data (feedstock molecular composition and process condition) were normalized to the range [-1, 1]. The normalized feedstock molecular composition and pilot-scale process conditions were input into Molecule-based ResNet and Process-based ResNet, respectively. Since the feedstock molecular composition and catalyst remained almost unchanged during the scale-up from laboratory to pilot plant, the network parameter of Molecule-based ResNet was frozen, while the Process-based and Integrated ResNets were fine-tuned. Then, the calculated molecular composition was input into the bulk property equations to estimate the product yield and gasoline composition. The calculated and measured bulk properties were compared in the MSE loss function to compute the error. Similar to training the laboratory-scale hybrid model, backpropagation was performed, and the Adam optimizer with weight decay was used to fine-tune the ResNet parameters. The performance of the transfer learning was evaluated using the R2 score, MAE, and MSE. The fine-tuning was stopped after 300 epochs, and the pilot-scale hybrid model was obtained.

Moreover, the hyperparameters for transfer learning were also tuned by the BO algorithm. Since the neural network architecture has been determined, only four hyperparameters need to be tuned, including dropout rate, learning rate, weight decay, and batch size. Their ranges are listed in Table R14. The hyperparameter optimization process was the same as the laboratory-scale hybrid model.

Table R14. The hyperparameter range setting in the BO algorithm

Hyperparameter	Method	Range
Learning rate	Floating-point hyperparameters	0 ~ 0.5
Dropout parameter	Floating-point hyperparameters	1.0E-06 ~ 1.0E-03
Weight decay	Floating-point hyperparameters	1.0E-06 ~ 1.0E-02
Batch size	Discrete categorical hyperparameters	4, 8, 16, 32, 64, 128, 256

Modification:

We modified the "Transfer Learning" section, and the "Transfer Learning" section was divided into the "Model Structure for Transfer Learning" section and the "Parameter Fine-tuning in Transfer Learning" section in the Method section of the revised manuscript.

Comment 17:

Some sentences could be reworked for better readability. For example, lines such as 76-77, "Qiu and her colleagues introduced an AI-mechanism integrated model for the naphtha steam cracking process based on the convolutional neural network (CNN)33" could be "Qiu and collaborators33…" with the reference immediately after the researchers' names. The same is true for line 79-80, 84. Please check these occurrences and modify accordingly.

Response:

Thank you for making us notice this, and we have modified "Qiu and her colleagues…" to "Qiu and collaborators[33]…".

Moreover, when a researcher's name is present, the reference is repositioned after the researcher's name.

Modification:

"Qiu and her colleagues…" was changed to "Qiu and collaborators[33]…".

Moreover, when a researcher's name is present, the reference is repositioned after the researcher's name in the revised manuscript.

We have also revised the entire draft to remove duplicates and increase readability.

Comment 18:

Please consider adding additional clarification for jargon in the manuscript. Terms such as "data augmentation", "parameter transfer strategy", "NN-NSGA" and others are mentioned but would benefit from definition, clarification, or additional specifics.

Response:

We agree with the reviewer's comment and explained the jargon in the manuscript one by

one.

Data augmentation: Pilot-scale product data were generated through interpolation within the range of available pilot data.

Parameter transfer strategy: The kinetic parameters regressed from laboratory-scale data were fine-tuned using pilot-scale experimental data, ensuring that the calculated product distribution closely matched the pilot-scale experimental results. This strategy was proposed by Chen et al. (Zhengyu Chen, et al. Chemical Engineering Journal, 2022, 437: 135429).

NN-NSGA: An optimization algorithm that couples neural networks with the NSGA-II to optimize process conditions.

Modification:

Clarification for the parameter transfer strategy was added to the fifth paragraph of the introduction section in the revised manuscript.

Clarification for data augmentation was added to the method section in the revised manuscript and can be found in the "Parameter Fine-tuning in Transfer Learning" section.

Clarification for NN-NSGA was added to the revised manuscript and can be found in the second paragraph of the "Multi-objective Optimization via NN-NSGA" section.

Comment 19:

The "software policy checklist" *.pdf does not seem to be properly compiled, with a placeholder text which says, "If this message is not eventually replaced by the proper contents of the document, your PDF viewer may not be able to display this type of document."

Response:

The "software policy checklist" *.pdf was double-checked, and all information was filled out.

Modification:

The "software policy checklist" *.pdf was rechecked and completed in the revised version.

Response to Reviewer 1:

Comment 1:

I do not have a further question with regards to terminology for the network architecture. It's referred to throughout as ResNet which implies a specific deep learning CNN architecture and the authors cite (REF 36) a paper that compared different CNN architectures for naphtha cracking processing. However, the architecture (described in ESI and shared code) used in this work uses MLP layers not convolutional ones. Unless there is a clear connection to the original ResNet design, I do not think this terminology should be used.

Response:

We thank the reviewer for pointing out this issue. Indeed, ResNet is fundamentally based on CNNs, whereas our model employs exclusively MLP layers. Accordingly, we have revised the term "ResNet" to "Residual MLPs (ResMLPs)" in the revised manuscript.

Modification:

"ResNet" was modified to "Residual MLPs (ResMLPs)" in the revised manuscript.

Comment 2:

Figure 3 - The first column (commonality and difference) should use more descriptive labels. For example, mode of operation, reaction type, etc.

Response:

We agreed with the reviewer's comment that the first column in Figure 3 should use more descriptive labels. The revised Figure 3 was shown in Figure R1.

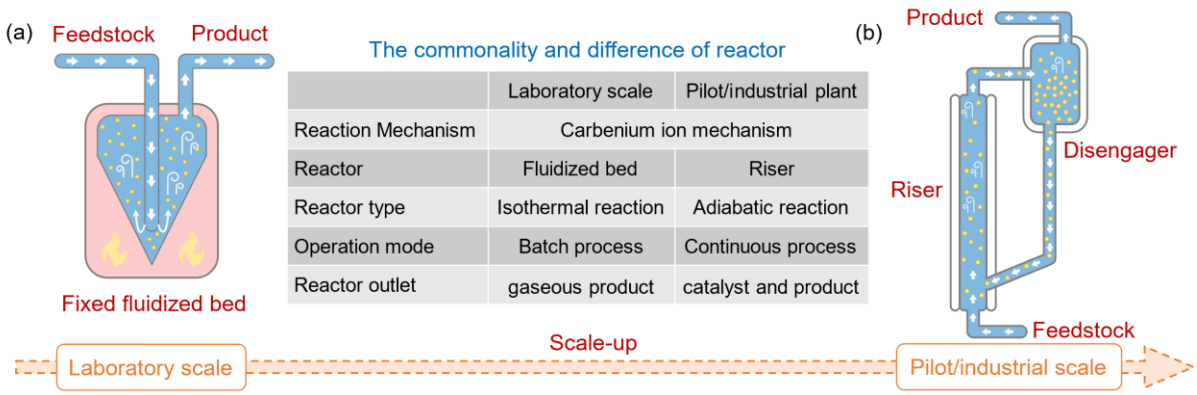

Figure R1. Fig. 3. Characterization of FCC reactors at different scales. (a) Laboratory scale; (b) Pilot/industrial scale.

Modification:

We modified Figure 3 using Figure R1 in the revised manuscript.

Comment 3:

README for code provided is quite extensive, but could be improved. For example:

Installation - Why not include a requirements.txt to setup the environment.

System requirements - 'Standard computer' could mean anything. This should be more specific, along with examples of the memory requirements.

Running the code - Details given are ok, but again could be clearer for a user. Include example commands for running the 'Bayes_for_HNN_gasoline_FFC.py' etc. I also think this might benefit from a juptyer notebook version with additional markdown and comments alongside the code.

Response:

We agreed with the reviewer's comment and improved the README.

- Requirements.txt

torch==2.0.0+cu118

numpy==1.26.4

optuna==3.6.1

pymoo==0.6.1.3

pandas==2.2.1

scikit-learn==1.6.1

matplotlib==3.8.4

joblib==1.4.0

- System requirements

OS: Windows 10/11

CPU: Intel Core i7-13700H

RAM: 32GB

GPU: NIVIDA RTX 3070

Storage: 1TB SSD

- Running the code

The provided code consists of runnable scripts. Taking script "Bayes_for_HNN_gasoline_FCC.py" as an example:

Opening it in PyCharm, the user can run the code by simply clicking the run button.

Moreover, converting all the code to Jupyter format requires significant effort, but there are now many tools available for .py-to-Jupyter conversion, such as PyCharm (Professional

Edition) and ChatGPT. ChatGPT can also add markdown and comments alongside the code. Thus, users can transform the code according to their needs.

Modification:

The details about Requirements.txt, System requirements, and Running the code were added to the revised README.

Comment 4:

Streamlit app is interesting, but not clear what it's purpose is? Is it for just displaying the training results of the models? i.e. it's not for actually training the lab scale or pilot scale models?

Response:

Yes, Streamlit is only for the audience to validate and display our current training results. If one intends to retrain the model, please refer to the corresponding source code ("HNN_gasoline_FCC_lab_train.py").

Response to Reviewer 3:

Comment 1:

In the literature we are familiar with, a "hybrid" model must contain both mechanistic and data-driven components. Given the additional explanation, we believe that the laboratory-scale process is not a hybrid model, but a data-driven surrogate of the original mechanistic model. On the other hand, the pilot-scale machine learning model is indeed a hybrid model as it includes bulk property calculations based on first-principles in addition to the data-driven components. The authors have defined their intention with hybrid and data-driven models in the paper, so we leave this change as a choice for the author and editor.

Response:

Currently, there is indeed controversy surrounding the definition of hybrid models, and we agree with the reviewer's comments. Therefore, we clearly defined the laboratory-scale process as a data-driven model, while the pilot-scale model is a hybrid model in the manuscript. We believe that as research progresses, the definition of hybrid models will become clearer.

Moreover, the definition for the laboratory-scale model can be found in the last paragraph of the Introduction section and the second paragraph of the "Hybrid Mechanistic Modeling and Deep Transfer Learning for Process Scale-up" section.

Comment 2:

Thank you for the response. Is it true that the bulk property calculation is a post-processing calculation? This could be better clarified in the paper.

Response:

Thanks for the reviewer's comment. This is a post-processing step for the output layer results in the neural network training.

We incorporated the bulk property calculation into the fine-tuning process during the transfer learning. When the molecular composition is output from the last layer of the neural network, the calculated molecular composition will be delivered into the bulk property calculator to estimate the bulk properties.

Modification:

We have clarified the calculation process in the "Property-informed Transfer Learning Strategy for Process Scale-up" section of the revised manuscript.

Comment 3:

No further comment, except that the order of the new references added to support the findings are flipped when comparing the "response to reviewers" document and updated manuscript. You might want to double check these.

Response:

We sincerely thank the reviewers for their careful review of the manuscript. Both of these references were cited in the same location. In the revised manuscript, we cited them in order of publication date. Please refer to the current version of the manuscript.

Comment 4:

The first occurrence of the NSGA-II acronym should be next to its definition (non-dominated sorting genetic algorithm II). Otherwise, no further comment.

Response:

We agreed with the reviewer's comment. When the NSGA-II acronym first appeared, we added its definition (non-dominated sorting genetic algorithm II).

Modification:

We have made modifications to the content in the revised manuscript as follows:

"the non-dominated sorting genetic algorithm II (NSGA-II)" was changed into "the NSGA-II (non-dominated sorting genetic algorithm II)" in "Multi-objective Optimization via NN-NSGA" section of the revised manuscript.

Comment 5:

Table R13, it looks like there is an MAE that is negative? Can you clarify this?

Response:

We regret that Table R3 caused a misunderstanding for reviewers. In Table R13, the table header "MAE" indicates that MAE is used as the optimization objective in the hyperparameter optimization process. The negative numbers in the table do not mean that MAE is negative, but rather that R^2 is negative.

This phenomenon indicates that when MAE is used as the optimization objective, although the overall MAE can be minimized, the fitting result for some extreme values is poor, resulting in a discrepancy between the predicted trend and the actual trend. To prevent this phenomenon, this paper chooses R^2 as the optimization objective for the Bayesian hyperparameter instead of MAE.

Comment 6:

The text added in "Model structure for transfer learning" is different between the

updated manuscript and the "response to reviewers" document. You might want to double-check.

Response:

We sincerely appreciate the reviewer's careful review. We have also conducted a double-check for this section. When we put the content in the "Model structure for transfer learning" section into the manuscript, we removed the first sentence in the "response to reviewers" file (In this work, transfer learning was applied to calculate the product distribution under the pilot plant). The rest of the content is consistent. The reason for deleting this sentence is that it has already been mentioned earlier and will not be repeated in the "Model Structure for Transfer Learning" section.