

# Data-driven Studies on Automation and Operational Monitoring of Fine Chemical Processes

Botao Yu<sup>1,✉</sup>

<sup>1</sup> Heze Emergency Management Support and Technical Service Center, Heze 274000, Shandong, China

## ABSTRACT

Fine chemical processes are integral to modern industries such as automotive, environmental protection, aviation, and new energy. However, these processes involve highly toxic substances and complex chemical interactions, making them vulnerable to uncontrollable circumstances and posing significant risks to human safety and the environment. This work proposes an enhanced GA-LVW algorithm for reliability assessment of fine chemical processes, focusing on essential operating units. The method utilizes global-local structure analysis to extract features from operating unit variables, reducing data noise, simplifying the construction of fuzzy rules, and improving model resilience. The extracted features are integrated into a fuzzy inference system. The proposed approach is validated using the Tennessee Eastman (TE) process model and the R-22 production process in a fluoride facility. Results demonstrate that the enhanced GA-LVW algorithm significantly improves the system's efficiency and maintainability compared to conventional fuzzy inference systems.

*Keywords:* Fine chemical process, PHM, FIS, GA-LVW algorithm, Chemical process monitoring

## 1. Introduction

Modern industry places significant importance on fine chemical processes due to their extensive applications across various sectors, including automotive, environmental protection, aviation, new energy, and others [15]. The increasing demand for product variety and the expansion of application domains have heightened the significance of the fine chemical industry. However, the synthesis of fine chemicals often involves highly toxic substances and complex chemical interactions, making the processes highly susceptible to uncontrollable factors, which pose significant risks to human safety and the environment [10, 14]. For example, even small leaks of highly toxic compounds used in

✉ Corresponding author.

*E-mail addresses:* [hzybt666@163.com](mailto:hzybt666@163.com) (Y. Botao).

Received 31 March 2024; accepted 01 September 2024; published 31 December 2024.

DOI: [10.61091/jcmcc123-22](https://doi.org/10.61091/jcmcc123-22)

© 2024 The Author(s). Published by Combinatorial Press. This is an open access article under the CC BY license (<https://creativecommons.org/licenses/by/4.0/>).

fluorochemical processes can severely threaten worker safety, equipment functionality, and public property [5]. Traditional monitoring and maintenance methods are struggling to meet the safety and reliability requirements of modern industry, necessitating new strategies for system monitoring and automation [2].

Conventional fine chemical process monitoring systems primarily focus on fault diagnosis and detection. While these systems can minimize losses and promptly repair faults, they are often effective only after the fault has caused irreversible financial losses and safety risks [12]. Recent research highlights that preventing breakdowns alone cannot guarantee system safety and reliability, emphasizing the need for predictive health management (PHM) techniques in the fine chemical sector [8]. These techniques aim to evaluate reliability and implement corrective actions before failures occur.

Reliability evaluation techniques in PHM can generally be categorized into knowledge-based, analysis-based, and data-driven approaches [9]. Knowledge-based approaches rely on expert experience and understanding, but their applicability is limited by production constraints and the unpredictability of expert knowledge. Analysis-based methods leverage physical and chemical principles but struggle to handle the complexity of fine chemical processes, such as time-variation and multivariate coupling [1, 16]. On the other hand, data-driven methods overcome these limitations by using extensive process data to build black-box models, offering broader applicability [4].

Although mainstream reliability assessment methods such as reliability block diagrams (RBD), fault tree analysis, Monte Carlo simulations, and Bayesian networks have been widely used in fields like electrical equipment, electronics, aviation, and energy, their application in the chemical industry remains limited [13]. This limitation arises from the multidimensional, strongly coupled, time-lagged, and nonlinear nature of chemical processes, which makes direct application of conventional techniques challenging. Moreover, the reliance on expert knowledge in most existing studies introduces judgmental biases, leading to uncertain reliability evaluation results [11].

To address these challenges, this study introduces a fuzzy inference system (FIS) for the reliability assessment of essential operational units in fine chemical processes, based on global-local structural analysis using an enhanced GA-LVW algorithm. FIS can handle uncertainty and logical reasoning, reducing the dependency on precise expert knowledge. However, its practical application in large-scale chemical processes is often hindered by the excessive number of variables and the complexity of fuzzy rule formulation [3]. By integrating global-local structural analysis, the proposed GA-LVW algorithm extracts features from operating unit variables, replacing the original variables with global-local features as inputs to the FIS. This approach reduces data noise, enhances model resilience, and significantly simplifies the fuzzy rule design, thereby reducing reliance on expert knowledge and expediting the system's logical design [7].

The methodology is validated using the Tennessee Eastman (TE) process model and the R-22 production process from a domestic fluorination plant. Real-time assessments of the reactor operating unit's reliability under various operating conditions demonstrate the efficacy of the enhanced GA-LVW algorithm in precise chemical process monitoring. Compared to the conventional FIS approach, the enhanced GA-LVW algorithm not only simplifies fuzzy rule design but also improves system efficiency and maintainability, making it highly applicable in real-world scenarios [6].

The enhanced GA-LVW algorithm proposed in this paper offers a novel approach to predictive health management and reliability evaluation in fine chemical processes. By integrating global and local structural elements, the approach effectively addresses the limitations of existing methods and demonstrates significant potential for broader applications in chemical process monitoring and au-

tomation. With advancements in industrial big data, this approach is expected to be further expanded, providing robust support for ensuring the safety and reliability of chemical production.

## 2. Improved GA-LVW Algorithm

### 2.1. GA Algorithm

The Genetic Algorithm (GA) is a heuristic search technique used to find optimal solutions by mimicking natural evolutionary processes. The steps for using GA in feature selection are as follows [11]:

1. **Feature Representation**: Each feature is represented by a binary value, where 1 indicates that the feature is selected, and 0 indicates it is not. A random function generates the initial population. The random function  $f_r$  in Python's random module is used in this paper, and it selects elements randomly from the sequence:

$$f_r = \text{random choice}(0, 1). \quad (1)$$

2. **Fitness Computation**: Compute the fitness value ( $f_i$ ) for each individual  $i$  in the population  $pop(t)$ . The fitness value is determined by the fitness function, which evaluates the superiority or inferiority of an individual:

$$f_i = \text{fitness}(pop_i(t)). \quad (2)$$

3. **Selection**: The selection process involves retaining the fittest individuals from the population to generate the next generation. Using the random competition method, two individuals  $random(pop_i)$  and  $random(pop_j)$  are selected, and the one with the higher fitness value is chosen for crossover:

$$\text{winner} = \max\{\text{random}(pop_i), \text{random}(pop_j)\}. \quad (3)$$

4. **Crossover**: Two individuals are selected at random, and their genetic material is exchanged to create offspring with better traits. The parent chromosomes before crossover are represented by:

$$\begin{cases} b_1 b_2 b_3 \cdots b_i \mid b_{i+1} b_{i+2} \cdots b_n \\ c_1 c_2 c_3 \cdots c_i \mid c_{i+1} c_{i+2} \cdots c_n \end{cases} . \quad (4)$$

5. **Mutation**: During the search process, mutation alters individual genes to escape local minima and accelerate convergence. A single-point mutation, or bit mutation, changes one bit of the DNA sequence. For instance, 0 is converted to 1, and 1 to 0. The mutation of the fifth bit is shown below:

$$0111110110. \quad (5)$$

### 2.2. LVW Algorithm

The LVW (Las Vegas Wrapper) algorithm is a feature selection method that uses a stochastic search strategy within the Las Vegas framework and employs the classifier's error as the evaluation criterion. The following steps outline the LVW algorithm:

1. **Sub-dataset Creation**: Given a dataset  $D$ , LVW randomly selects a feature subset  $A'$  to create a sub-dataset  $D^{A'}$ . The dataset is then divided into training and validation sets. Using a designated learning machine  $\xi$ ,  $k$ -fold cross-validation computes the cross-validation error:

$$E_i = \text{CrossValidation} \left( \zeta \left( D^{A'} \right) \right), \quad (6)$$

and the average error is calculated as:

$$E'_i = \frac{1}{k} \sum_{i=1}^k E_i. \quad (7)$$

Alternatively, the generalization error  $E'$  on the test dataset can be used as the evaluation criterion:

$$E' = \text{Generalization} \left( \zeta_{\text{train}} \left( D_{\text{test}}^{A'} \right) \right), \quad (8)$$

where the feature subset is part of the test sub-dataset.

2. **Updating Global Values**: If the conditions of Eq. (9) are met, update the global minimum error  $E$ , the optimal feature subset  $A^*$ , the number of feature subsets  $d'$ , and the global optimal features  $d$ :

$$(E' < E), (E' = E), (d < d'). \quad (9)$$

3. **Termination**: Repeat the above steps until the termination criterion is satisfied.

### 2.3. The GA-LVW Algorithm

The GA-LVW algorithm combines the global search capability of GA with the feature optimization efficiency of LVW. The algorithm operates as follows:

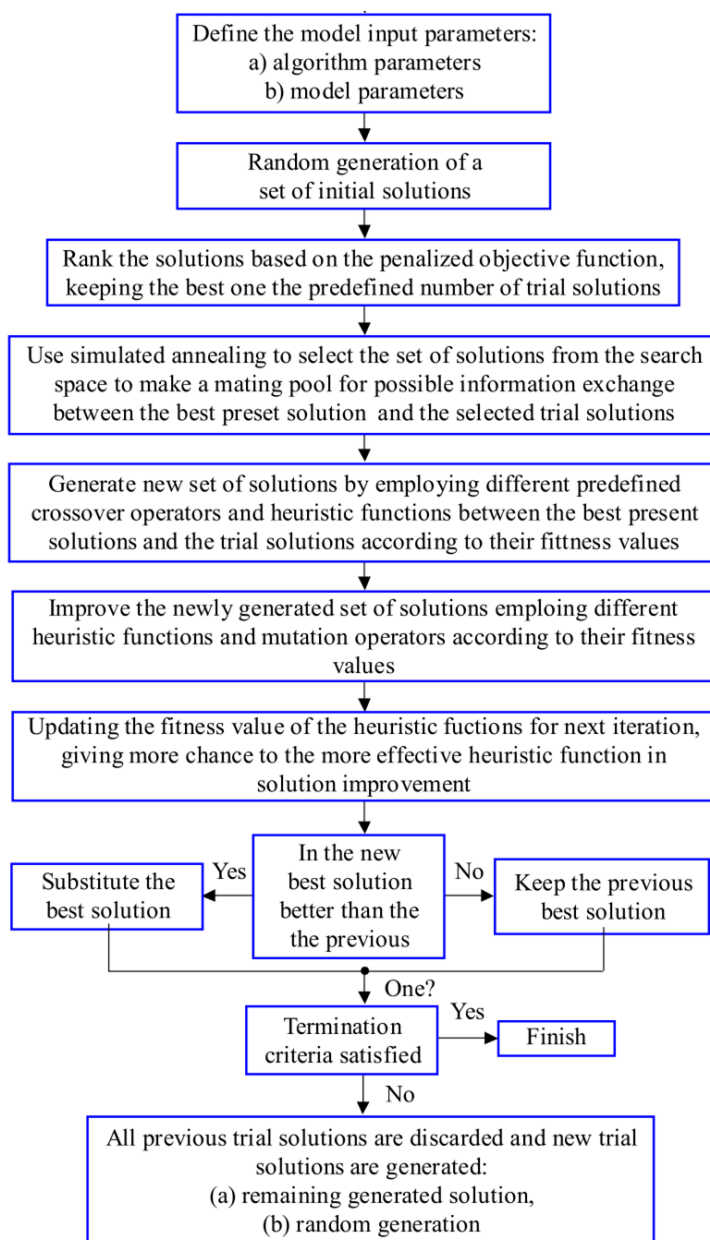
1. The features of the population are encoded and initialized using GA.
2. LVW randomly selects a feature subset to create a sub-dataset, which is divided into training and validation sets.
3. The fitness of each individual in the population is calculated after selection, crossover, and mutation. Individuals with higher fitness are retained as the current optimal population, and the optimal subset is updated.
4. The algorithm evaluates whether crossover or mutation is needed. If not, the target feature set is output. Otherwise, the process continues until the termination condition of the GA algorithm is satisfied.

Figure 1 illustrates the flow of the GA-LVW algorithm.

## 3. Performance Assessment of the Enhanced GA-LVW Algorithm Method

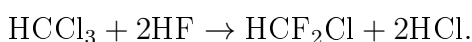
### 3.1. Fluorochemical R-22 Process Reactor Operating Unit Reliability Estimates

The proposed enhanced GA-LVW algorithm was applied to the R-22 manufacturing process dataset of a fluoride plant in China to evaluate its effectiveness. Due to confidentiality agreements, this study introduces the application procedure and results using the reactor operation unit of the R-22 production process as an example.



**Fig. 1.** Flowchart of the GA-LVW algorithm

Dichlorodifluoromethane (R-22) is a commonly used intermediate in fluorochemicals and is widely used as a raw material for propellants and refrigerants. The primary raw materials for R-22 production are hydrofluoric acid and chloroform. The chemical reaction equation is:



The R-22 production unit includes the reactor, separator, water washing tower, distillation towers, and other equipment. Hydrofluoric acid and chloroform are pumped into the reactor, where they react to produce R-22, R-21, R-23, and HCl. The subsequent process involves fractional distillation, washing, and further distillation to obtain the main product, R-22.

This research utilizes data from January 2023 to October 2023, sampled every minute. The dataset is representative, covering nearly a year and accounting for seasonal changes and feedstock variations. Rigorous safety controls were implemented, and data from shutdowns and maintenance periods were

included to simulate abnormal production conditions and validate the method.

To train the GLSA model, 1,000 samples from regular reactor operations were collected. The top three eigenvalues were selected from the 10 reactor operational unit variables, and the remaining samples were projected into three dimensions. These reduced variables served as inputs for the fuzzy inference system. Figure 2 illustrates the fuzzy inference system design.

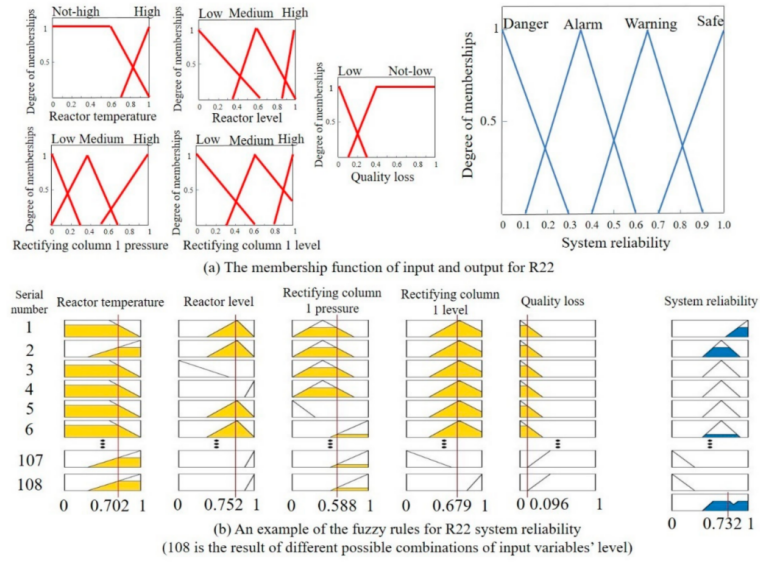


Fig. 2. An illustration of the enhanced GA-LVW algorithm’s use in the R-22 reactor’s operating unit.

Figure 3 shows the reliability of the reactor operating unit and the critical feedstock input under three abnormal conditions. In Case 1, the critical feedstock initially remains near the target value, and the reactor’s reliability exceeds 90%. However, during a routine maintenance halt, the feedstock drops to zero, and the reliability falls to around 20%. Once the feed is restored, reliability gradually recovers above 90%.

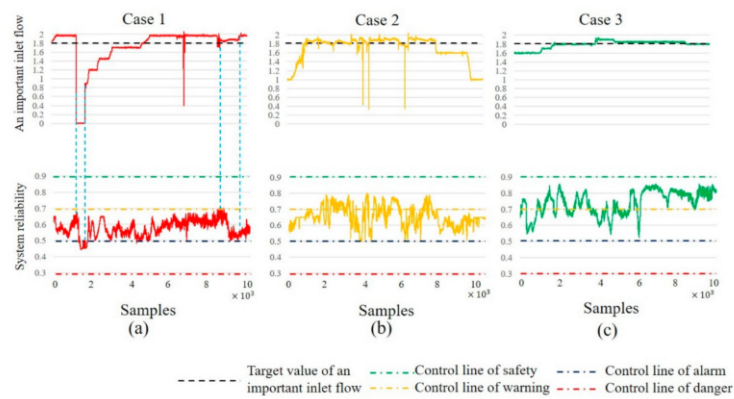


Fig. 3. Reliability of reactor operational unit and critical feedstock input under three scenarios of R-22.

In Case 2, a brief decrease in critical feedstock results in a decline in reactor reliability, approaching the warning limit of 60%. As the feedstock stabilizes, reliability recovers. In Case 3, the feedstock remains stable, and reliability fluctuates between the safety and warning control limits.

A comparison of the enhanced GA-LVW algorithm with expert-guided fuzzy inference systems is shown in Figure 4. While the trends in reliability estimates align, the expert-guided results exhibit

larger deviations from the actual system behavior, demonstrating the advantages of the enhanced GA-LVW method.

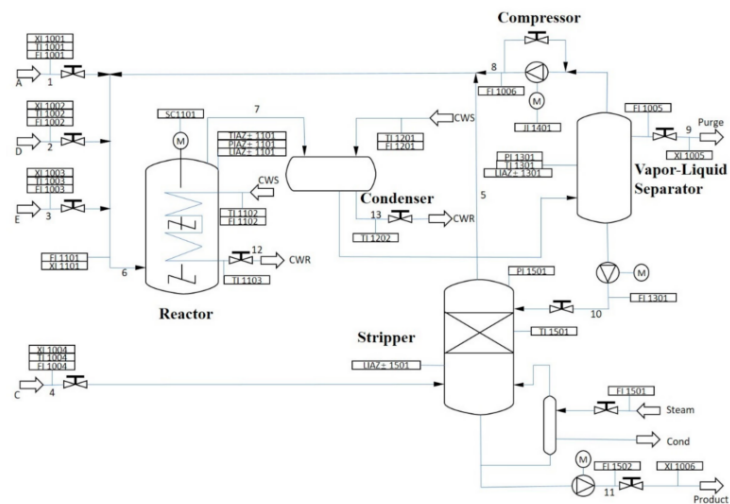


**Fig. 4.** Comparison of reliability estimation results for the R-22 reactor operating unit.

The enhanced GA-LVW algorithm requires only 27 fuzzy rules compared to the 59,045 rules needed for standard FIS, significantly reducing system complexity and enhancing maintainability.

**3.2. Evaluation of TE Model Reactor Operating Units for Reliability**

The proposed method was further validated using the Tennessee Eastman (TE) process model. The TE process comprises five operational units: reactor, condenser, separator, stripper, and compressor. The process flow diagram is shown in Figure 5.



**Fig. 5.** Flowchart of the Tennessee Eastman (TE) process.

The TE model includes 12 operational variables and 22 process variables. The enhanced GA-LVW algorithm was used to evaluate reactor and separator reliability. The fuzzy inference system required only 27 rules for reactor reliability assessment, compared to  $4.30 \times 10^7$  rules in the standard FIS approach, highlighting the method’s practicality.

Figure 6 illustrates the affiliation functions of the input and output variables for the TE reactor and separator operational units.

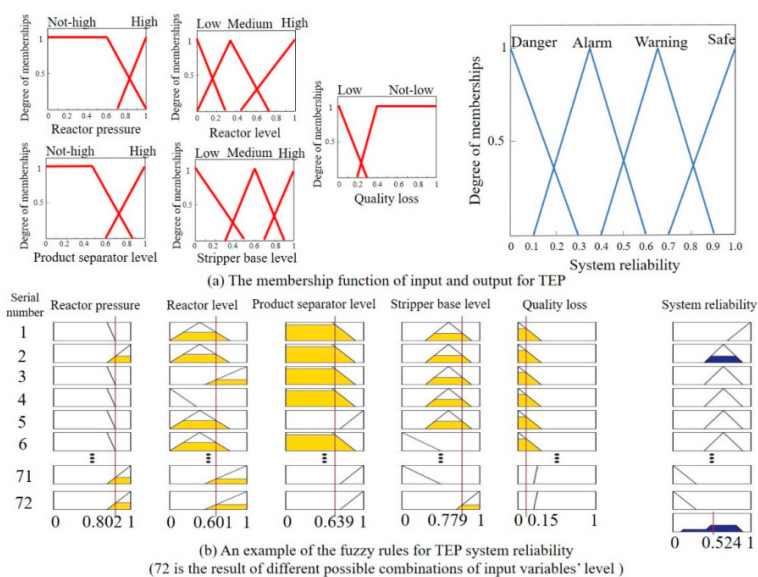


Fig. 6. Affiliation functions for the TE reactor and separator operational units.

Figure 7 depicts the reliability variation of the TE reactor and separator under three fault scenarios. The enhanced GA-LVW algorithm accurately reflects unit reliability changes, demonstrating its effectiveness in fault detection and system monitoring.

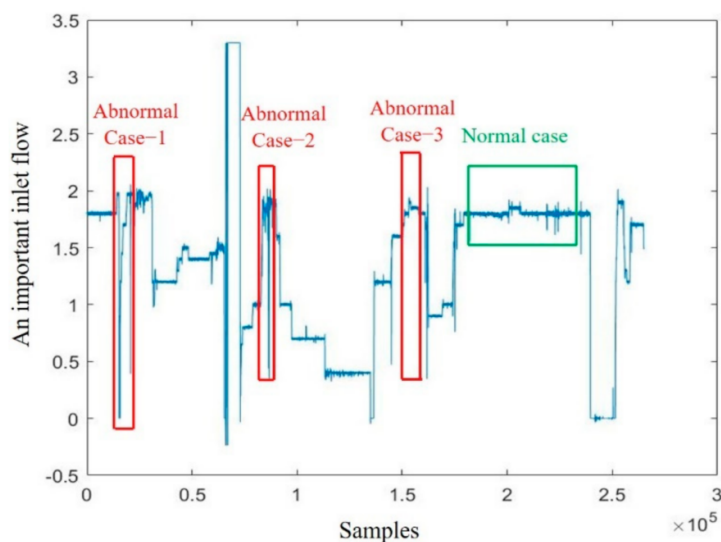


Fig. 7. Reliability of TE reactor and separator operational units under three fault scenarios.

## 4. Conclusion

This study proposes an enhanced GA-LVW algorithm for reliability assessment of fine chemical process operating units, combining global-local structure analysis and fuzzy inference systems. The method reduces reliance on expert knowledge, lowers design complexity, and enhances system robustness. Validation using the R-22 process and TE model demonstrates the algorithm's superiority in



improving reliability assessment and maintainability, making it a valuable tool for predictive health management in complex chemical processes.

## References

- [1] S. Agrawal, S. H. Rawal, V. R. Reddy, and J. M. Merritt. Use of automation, dynamic image analysis, and process analytical technologies to enable data rich particle engineering efforts at the drug substance/drug product interface: a case study using lovastatin. *Chemical Engineering Research and Design*, 205:578–590, 2024. <https://doi.org/10.1016/j.cherd.2024.04.032>.
- [2] M. Bortz, K. Dadhe, S. Engell, V. Gepert, N. Kockmann, R. Müller-Pfefferkorn, T. Schindler, and L. Urbas. Ai in process industries—current status and future prospects. *Chemie Ingenieur Technik*, 95(7):975–988, 2023. <https://doi.org/10.1002/cite.202200247>.
- [3] L. Guo, Y. Sun, et al. Economic forecasting analysis of high-dimensional multifractal action based on financial time series. *International Journal for Housing Science and Its Applications*, 45(2):11–19.
- [4] D. Liu, Y. Wang, C. Liu, K. Wang, and X. Yuan. Intelligent operational optimization method for salt lake chemical industrial processes based on data-driven modeling. In *2023 IEEE 18th Conference on Industrial Electronics and Applications (ICIEA)*, pages 1833–1838. IEEE, 2023. <https://doi.org/10.1109/ICIEA58696.2023.10241917>.
- [5] S. Meier, S. Klarmann, N. Thielen, C. Pfefferer, M. Kuhn, and J. Franke. A process model for systematically setting up the data basis for data-driven projects in manufacturing. *Journal of Manufacturing Systems*, 71:1–19, 2023. <https://doi.org/10.1016/j.jmsy.2023.08.024>.
- [6] K. Nag and N. R. Pal. A multiobjective genetic programming-based ensemble for simultaneous feature selection and classification. *IEEE Transactions on Cybernetics*, 46(2):499–510, 2015. <https://doi.org/10.1109/TCYB.2015.2404806>.
- [7] T. X. Nguyen, Z.-Y. Wei, T.-M. Zheng, Y.-H. Su, K.-S. Chuang, and J.-M. Ting. Enhancement of urea oxidation reaction in alkaline condition via heterointerface engineering. *Chemical Engineering Journal*, 496:153841, 2024. <https://doi.org/10.1016/j.cej.2024.153841>.
- [8] T. H. Oh. Quantitative comparison of reinforcement learning and data-driven model predictive control for chemical and biological processes. *Computers & Chemical Engineering*, 181:108558, 2024. <https://doi.org/10.1016/j.compchemeng.2023.108558>.
- [9] L. Peterson, J. Bremer, and K. Sundmacher. Challenges in data-based reactor modeling: a critical analysis of purely data-driven and hybrid models for a cstr case study. *Computers & Chemical Engineering*, 184:108643, 2024. <https://doi.org/10.1016/j.compchemeng.2024.108643>.
- [10] J. Sansana, R. Rendall, M. N. Joswiak, I. Castillo, G. Miller, L. H. Chiang, and M. S. Reis. A functional data-driven approach to monitor and analyze equipment degradation in multiproduct batch processes. *Process Safety and Environmental Protection*, 180:868–882, 2023. <https://doi.org/10.1016/j.psep.2023.10.041>.
- [11] N. Sifakis, N. Sarantinoudis, G. Tsinarakis, C. Politis, and G. Arampatzis. Soft sensing of lpg processes using deep learning. *Sensors*, 23(18):7858, 2023. <https://doi.org/10.3390/s23187858>.
- [12] Y.-N. Sun, W. Qin, J.-H. Hu, H.-W. Xu, and P. Z. Sun. A causal model-inspired automatic feature-selection method for developing data-driven soft sensors in complex industrial processes. *Engineering*, 22:82–93, 2023. <https://doi.org/10.1016/j.eng.2022.06.019>.

- 
- [13] S. Wang, N. Schultheiss, and S. Kim. Dddas within the oil and gas industry. *Handbook of Dynamic Data Driven Applications Systems: Volume 2*:331–351, 2023. [https://doi.org/10.1007/978-3-031-27986-7\\_12](https://doi.org/10.1007/978-3-031-27986-7_12).
- [14] F. Yan, L. Kong, Y. Li, H. Zhang, C. Yang, and L. Chai. A survey of data-driven soft sensing in ironmaking system: research status and opportunities. *ACS Omega*, 9(24):25539–25554, 2024. <https://pubs.acs.org/doi/10.1021/acsomega.4c01254>.
- [15] F. Yan, X. Zhang, C. Yang, B. Hu, W. Qian, and Z. Song. Data-driven modelling methods in sintering process: current research status and perspectives. *The Canadian Journal of Chemical Engineering*, 101(8):4506–4522, 2023. <https://doi.org/10.1002/cjce.24790>.
- [16] H. Yu, M. Gao, H. Zhang, G. Yue, and Z. Zhang. Data-driven optimization of pollutant emission and operational efficiency for circulating fluidized bed unit. *Energy*, 281:128338, 2023. <https://doi.org/10.1016/j.energy.2023.128338>.