Data Article

Data on propylene/propane separation by the externally heat-integrated distillation column (EHIDiC) using data-driven analysis

Peng Qiu\textsuperscript{a,b}, Bo Huang\textsuperscript{a,b}, Zhenghua Dai\textsuperscript{a,b,*}, Fuchen Wang\textsuperscript{a,b}

\textsuperscript{a} Key Laboratory of Coal Gasification of Ministry of Education, East China University of Science and Technology, Shanghai 200237, PR China
\textsuperscript{b} Shanghai Engineering Research Center of Coal Gasification, East China University of Science and Technology, P.O. Box 272, Shanghai 200237, PR China

\textbf{Article history:}
Received 19 October 2019
Revised 29 February 2020
Accepted 6 April 2020
Available online 20 April 2020

\textbf{Keywords:}
Data-driven
Heat-integrated distillation column
Propylene-propane separation

\textbf{A B S T R A C T}

Nine hundred data were collected from rigorously simulated EHIDiC for propylene/propane separation using the Aspen Plus-Matlab communication platform. Statistical analysis was performed to give a further understanding of EHIDiC, a highly coupled system. The input variables, integrated stages, and the corresponding heat of two couples of external exchangers were generated stochastically in a wide range to cover the completely reasonable operation window. The results of blocks and streams in the flowsheet were calculated as the output data, including the flow rate, the temperature, and the pressure, as well as Total Annualized Cost (TAC). The data can be reused to design and optimize both the steady and dynamic schemes of EHIDiC, especially with machine learning methods. This manuscript gave a full description of collecting and analyzing of the data used in the research article “Data-driven analysis and optimization of externally heat-integrated distillation columns (EHIDiC)” [1].

© 2020 Published by Elsevier Inc.
This is an open access article under the CC BY-NC-ND license. (http://creativecommons.org/licenses/by-nc-nd/4.0/)

\* Corresponding author at: Key Laboratory of Coal Gasification of Ministry of Education, East China University of Science and Technology, Shanghai 200237, PR China.
E-mail address: chinadai@ecust.edu.cn (Z. Dai).

https://doi.org/10.1016/j.dib.2020.105559
2352-3409/© 2020 Published by Elsevier Inc. This is an open access article under the CC BY-NC-ND license. (http://creativecommons.org/licenses/by-nc-nd/4.0/)
Specifications table

| Subject                        | Chemical Engineering          |
|-------------------------------|-------------------------------|
| Specific subject area         | Energy, Separation technology, Process engineering, Modelling and simulation |
| Type of data                  | Table                         |
|                               | Figure                        |
|                               | Code                          |
| How data were acquired        | Data were acquired from rigorously simulated models by Aspen Plus (V 7.3)-Matlab (R2018a) communication platform. |
| Data format                   | Raw                           |
|                               | Analyzed                      |
|                               | Filtered                      |
| Parameters for data collection| RadFrac blocks at 80% flooding using Fair correlation with the Peng-Robinson method were simulated and verified with the literature results. The feed and operating conditions were the same as employed by Olujic [2]. Distributed computing with 50 clients that each has an Intel i5 CPU of 2.4 GHz was conducted by Matlab to generate a large dataset (5000 data) in a short time. Each converged sample takes about 1 minute. |
| Description of data collection| Firstly, six input variables, i.e., the integrated heat and stages of two external exchangers, were generated stochastically in Matlab and passed to Aspen Plus to run a simulation. Then, the calculated results were read by Matlab only when the simulation was converged. Finally, the collected data were cleaned to eliminate boundary results (confined by design-specifications) and the data that not make sense. |
| Data source location          | Shanghai, PR China            |
| Data accessibility            | With the article              |
| Related research article      | Peng Qiu, Bo Huang, Zhenghua Dai, Fuchen Wang, Data-driven analysis and optimization of externally heat-integrated distillation columns (EHIDIC), Energy. https://doi.org/10.1016/j.energy.2019.116177 (In Press) |

Value of the data

- The present data is useful to analyze EHIDiC for splitting propylene/propane.
- Researchers in the communities of EHIDiC, propylene/propane separation, and other highly coupled systems can benefit from these data.
- These data can be used for further understanding of both the steady and dynamic performances of EHIDiC.
- These data can be trained using machine learning to construct reliable reduced-order models.

1. Data description

The EHIDiC scheme to be optimized is illustrated in Fig. 2. Heat exchanger HX0 is arranged between the top of the rectifier and the bottom of the stripper, whose heat load is Q0. E0 is used to reduce the vapor fraction of the stream that flows into the top of the stripper through the valve [3]. HX1 and HX2 are two external heat exchangers, which randomly connect paired stages along the column length of the rectifier (NR1 and NR2) and the stripper (NS1 and NS2), respectively. The amount of integrated heat via HX1 and HX2 are named as Q1 and Q2. The raw data provided in this article are in the Supplementary Material, which includes 900 samples that each with six input variables, 23 output variables (see in Table 1), and the TACs, as well as their compositions (i.e., the capital costs of the rectifier, stripper, compressor, E0, HX0, HX1, HX2, and the operation cost of the compressor, E0). In addition, the Matlab codes used in this work are provided in the Supplementary Material.

To gain insights into TAC optimization and heat-integration mechanism, we divide the two heat-integrated pairs, NS1-Q1-NR1 and NS2-Q2-NR2, into four sets (i.e., NS1-Q1, NR1-Q1, NS2-Q2, and NR2-Q2) (see in Fig. 1). Although Q0 is calculated given a set of input variables and
the design-specification of products, it is also considered here because Q0 is a part of heat-integration.

Color bubbles in the top-left quarter stand for the effects of NS1-Q1 on TAC (see in Fig. 1a); its counterpart, effects of NR1-Q1 on TAC, are in the top-right quarter (see in Fig. 1b). Similarly, the impacts of NS2-Q2 and NR2-Q2 on TAC are presented in the bottom left quarter (see in Fig. 1c) and the bottom-right quarter (see in Fig. 1d), respectively. In other words, the parameters of the first heat-integrated pair, NS1-Q1-NR1, can be found in Fig. 1a and b; the parameters of the second heat-integrated pair, NS2-Q2-NR2, can be found in Fig. 1c and d. By organizing input variables this way, it will be more convenient to target a specific sample. For example, given a sample in Table 2, its variables of NS1-Q1, NR1-Q1, NS2-Q2, and NR2-Q2 can be found at p1 (see in Fig. 1a), p2 (see in Fig. 1b), p3 (see in Fig. 1c), and p4 (see in Fig. 1d), respectively. These four bubbles, which have the same TAC and Q0, and symmetrically locate on dashed circle r3, make up one sample. As a result, all samples can be projected into Fig. 1 by organizing the variables into four sets of parameters.

For simplicity, TAC is divided into three regions (i.e., low TAC region in the middle shadow, high TAC region in the left and the right shadow, and middle TAC region for the other samples). In the low TAC region, only green bubbles (see in Fig. 1b and d) can be discovered, suggesting that heat rejected from the top of the rectifier is essential for reducing TAC. Meanwhile, only red bubbles (see in Fig. 1a and c) in the low TAC region are shown, indicating that heat must be injected into the bottom of the stripper for reducing TAC. Although little bubbles in other colors in the low TAC region can also be found, the integrated heat in these samples is too weak to affect TAC. However, these small bubbles have a common characteristic of high Q0, suggesting that HX0 has evolved to be the dominant source of heat-integration of EHIDIc. We thus conclude that decreasing TAC of EHIDIc prefers to integrate heat between the bottom of the stripper and the top of the rectifier, which is consistent with the work by Chen et al. [4] and Shahandeh et al. [5]. Besides, the diversity of color bubbles in the low TAC region suggests that there are different scenarios for an optimal configuration.

In the high TAC region in the right shadow (see in Fig. 1b and d), only big red bubbles can be found, while multi-colored bubbles are shown in the left shadow (see in Fig. 1a and c). It suggests that if the heat is rejected from the bottom of the rectifier, TAC will be brought to a high level, no matter which stage is integrated on the stripper side. In this situation, not enough vapor will be generated to satisfy the driving force. Thus, more heat must be added to the bottom of the rectifier by the vapor through the compressor. However, based on the operation conditions of propylene/propane splitter, the rectifier bears more separation duty than the stripper, which makes rectifier more sensitive to the patterns of heat-integration. As a result, heat rejection from the bottom of the rectifier must be avoided in the design process.

---

### Table 1

The 23 output variables simulated from EHIDIc schemes.

| Rectifier | Qc | xD | RR | T1 | T8 | V1 | L1 | V8 | L8 |
|-----------|----|----|----|----|----|----|----|----|----|

| Stripper  | Q8 | xW | RR | T1 | T8 | V1 | L1 | V8 | L8 |
|-----------|----|----|----|----|----|----|----|----|----|

| Compressor | BP | T0 | E0 | Q0 | T0 | Valve | HX0 |
|------------|----|----|----|----|----|--------|-----|

* The condenser duty of the rectifier is equivalent with the rebolier duty of the stripper.

### Table 2

The input variables and TAC of a given sample.

| Variables | NS1 | Q1 (Gcal h⁻¹) | NR1 | NS2 | Q2 (Gcal h⁻¹) | NR2 | TAC (× 10⁻⁷) |
|-----------|-----|--------------|-----|-----|--------------|-----|---------------|
| Sample    | 47  | 22.67218     | 4   | 4   | 31.64612     | 37  | 1.26453       |
Heat-integration between the bottom of the stripper and the top of the rectifier is preferred to minimize TAC. This pattern gains the most significant driven force between the operation line and the equilibrium line, which require the least heat duty for separation. Besides, rejecting heat from the bottom of the rectifier may increase TAC dramatically. Moreover, the searching space is non-convex that more than one set of input variables can get the minimum TAC, which has been discussed in the research paper [1].

2. Experimental design, materials, and methods

The optimization of EHIDiC was a Non-convex Mixed Integer Non-Liner Programming problem with multiple variables in a huge searching space. Thus, the data-driven analysis framework was proposed to deal with a large amount of data. As shown in Fig. 2, to collect data within a short time, distributed computing was used by Matlab. Given an amount of data, i.e., 5000
Fig. 2. The framework of data-driven analysis based on simulation for propylene/propane separation using EHIDiC with three external exchangers.

Table 3
Principal component analysis of the effects of input variables on TAC.

| Variables | Eigenvalue | Percentage of Variance (%) |
|-----------|------------|----------------------------|
| NR1       | 1.6496     | 27.49338                   |
| NR2       | 1.12661    | 18.77688                   |
| NS1       | 1.03939    | 17.32323                   |
| NS2       | 0.93196    | 15.53273                   |
| Q1        | 0.90081    | 15.01344                   |
| Q2        | 0.35162    | 5.8034                     |

in this experiment, the input variables were generated stochastically in the host machine and were assigned to 50 clients as jobs. Then, in each client machine, a set of 6 input variables was passed to Aspen Plus via COM to run a simulation. Only when the calculation was converged, the output variables can be returned to Matlab in a client which finally gathered by the host.

The collected data must be pre-processed before further analysis. Firstly, the data were cleaned carefully to eliminate abnormal samples, e.g., duplicate data. Besides, during the simulation, the purities of the products were confined by specific-designs in Aspen Plus, which made some results reach their limits. Although all the simulations were converged, these boundary data were still abandoned to make data reliable and robust. Secondly, data distribution must be checked. The EHIDiC was a highly coupled system, which may appreciate specific input variables that made the simulations converged, no matter how uniformly the input variables were generated. Finally, the dimensionality reduction was conducted to make the following analysis more efficient. Six independent variables affect TAC. It will be easier for designing the HIDiC if we know the main factors. Therefore, we want to find out whether the dimension could be reduced by PCA. The covariance matrix is calculated by Eq. (1). Then, the eigenvalue decomposition is conducted by Eq. (2). After that, the percentage of variance is obtained by Eq. (3). According to Table 3, although the dimension could not be reduced, we know that the NR1 takes nearly one-third of the variance of TAC, suggesting that TAC is more sensitive to the paired stages on the rectifier. Thus, heat-integrated stages on rectifier should be carefully chosen.

\[ C = \frac{X^T X}{N-1} \]
Table 4  
Formulas and corresponding parameters of capital cost calculation (M&S=1463.2).

| Cost components | Formulas | Parameter values |
|------------------|----------|------------------|
| column shell     | $CSC_{int} = \left( \frac{MKS}{280} \right)\Delta d_{col}^{1.066}h_{col}^{0.802} \times 1.5$ | rectifier: $C = 4106.85$ |
| condenser/reboiler/heat exchanger | $HEX_{int} = \left( \frac{MKS}{280} \right)CA^{0.65}$ | condenser: $C = 1609.13$, $U = 800$ W/m² K |
| compressor       | $CC_{int} = \left( \frac{MKS}{280} \right) \times 2047.24bp^{0.82}$ | reboiler: $C = 1775.26$, $U = 1000$ W/m² K |

$$C = P_1A P_T^T$$  \hspace{1cm} (2)  
$$V_i = \frac{P_i}{\sum P_i}$$  \hspace{1cm} (3)  

The TAC is considered as the target for optimization, and is calculated according to Table 4. Consequently, input variables were investigated according to their univariate and multivariate effects on TAC. Besides, a large amount of data can be used to construct a highly reliable data-driven model using machine learning. Moreover, the learned reduced-order models can be used for both steady state and dynamic simulation in high computational efficiency. In addition, the data-driven models can be used for global optimization due to their large searching space.

Acknowledgments  
The authors are grateful for the financial support from the National Key R&D Plan (Grant no. 2018YFC0808500; 2017YFB0602600) and the National Natural Science Foundation of China (Grant no. 21776086).

Conflict of Interest  
The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Supplementary materials  
Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.dib.2020.105559.

References  
[1] P. Qiu, B. Huang and Z. Dai et al., Data-driven analysis and optimization of externally heat-integrated distillation columns (EHIDiCs), Energy (2019) 116177, doi:10.1016/j.energy.2019.116177.  
[2] Ž. Olujić, L. Sun, A. de Rijke, P.J. Jansens, Conceptual design of an internally heat integrated propylene-propane splitter, Energy 31 (15) (2006) 3083–3096, doi:10.1016/j.energy.2006.03.030.  
[3] L. Zhang, J. Liu, X. Li, B. Jiang, H. Li, An improved pressure swing thermally coupled distillation, Acta Petrol. Sin. Petrol. Process. Sect. 29 (2) (2013) 312–317, doi:10.16085/j.issn.1000-6613.2016.01.007.  
[4] D. Chen, X. Yuan, L. Xu, K.T. Yu, Comparison between different configurations of internally and externally heat-integrated distillation by numerical simulation, Ind. Eng. Chem. Res. 52 (16) (2013) 5781–5790, doi:10.1021/ie400112k.  
[5] H. Shahandeh, J. Ivakpour, N. Kasiri, Internal and external HI DiCs (heat-integrated distillation columns) optimization by genetic algorithm, Energy 64 (2013) 875–886, doi:10.1016/j.energy.2013.10.042.