



Article

Multi-Objective Collaborative Optimization of Distillation Column Group Based on System Identification

Renchu He ^{1,2} , Keshuai Ju ¹, Linlin Li ³ and Jian Long ^{1,*} ¹ Key Laboratory of Smart Manufacturing in Energy Chemical Process, Ministry of Education, East China University of Science and Technology, Shanghai 200237, China² Engineering Research Center of Process System Engineering, Ministry of Education, East China University of Science and Technology, Shanghai 200237, China³ School of Automation and Electrical Engineering, University of Science and Technology Beijing, Beijing 100083, China

* Correspondence: longjian@ecust.edu.cn

Abstract: In this paper, a multi-objective collaborative optimization (MOCO) strategy is proposed for making decisions on a distillation column group. Firstly, based on data preprocessing, the operating modes of the tower group are determined by use of the fuzzy C-means clustering method. Secondly, based on the proposed concept of a collaborative variable, the discrete state-space model of the main towers are constructed by the subspace identification method. Then, a MOCO optimization model is designed for the ethylene plant. Finally, NSGA-III is used to solve the optimization problem. An analysis of a Pareto-optimal frontier and population is carried out. To illustrate the superiority of the proposed strategy, the results are compared with historical data and the appealing operation area is finally obtained.

Keywords: multi-objective optimization; ethylene; propylene; clustering; distillation column group



Citation: He, R.; Ju, K.; Li, L.; Long, J. Multi-Objective Collaborative Optimization of Distillation Column Group Based on System Identification. *Processes* **2023**, *11*, 436. <https://doi.org/10.3390/pr11020436>

Academic Editors: Blaž Likozar and Vincenzo Russo

Received: 10 January 2023

Revised: 28 January 2023

Accepted: 30 January 2023

Published: 1 February 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

In the real world, most optimization problems have multiple conflicting objectives. Single objective optimization (SOO) which gives a single solution has difficulty weighing multiple objectives. Unlike SOO, multi-objective optimization (MOO) gives a set of optimal solutions, each of which corresponds to different values of the objective functions, which, in turn, can form a Pareto frontier to provide rational guidance to decision-makers. Therefore, MOO has many applications in renewable energy resources production optimization [1,2], energy saving optimization [3,4], and emission reduction optimization [5–7]. In trade-off MOO methods, priori methods, interactive methods, Pareto-dominated methods and new dominance methods are available [8].

In recent years, there has been a wealth of MOO research on single towers and tower groups. Zhou et al. [9] optimized the decision variables of a methanol-to-propylene plant under different operating conditions based on a MOO framework and achieved a balance between total heating, cooling costs and product purity. For Fischer–Tropsch synthesis, Zhang et al. [10] transformed the design of a reactive distillation column into a MOO problem and determined the column structure based on the Pareto-optimal frontier. In [11], MOO was used to weigh the total annual cost and energy of a cyclohexane–isopropanol–water extractive distillation process. In [12], by considering three conflicting objectives, the total annual cost, the CO₂ emissions and the molar purity of a biodiesel product, an algal biodiesel reactive distillation unit was optimized. Gu et al. [13] optimized heat-integrated pressure-swing distillation based on MOO to reduce the total annual cost and CO₂ emissions, while saving energy. For similar purposes, MOO was also used for heteroazeotropic distillation processes [14], dividing wall columns [15,16]. It is worth noting that ethylene is one of the most produced chemical products in the world—the ethylene industry is also

the core of the petrochemical industry, but little research has been undertaken in this area. Shen et al. [17] targeted the chilling train and demethanization system in ethylene manufacturing according to the Pareto-optimal frontier, and, by maximizing the exergy efficiency and minimizing the operational cost, the optimal operation was selected. Multi-objective adaptive surrogate model-assisted optimization was proposed by [18] et al. with energetic, economic and environmental MOO performed for a practical ethylene separation process. More research should be devoted to the ethylene industry.

MOO studies of single towers and tower groups are essentially based on Aspen [9,11–18]. To build an Aspen model of a tower or tower group, it is necessary to have in-depth knowledge of the physical parameters and it is difficult to obtain satisfactory Aspen models in the absence of sufficient parameters. The question arises of how to conduct a MOO study of a tower group in the absence of the necessary physical parameters or without access to Aspen software. This is a problem. In addition, the optimization of Aspen-based MOO studies is separated from the simulation of the process; interaction between software is required. Before population iteration, the data in Aspen needs to be updated to the optimization platform, either manually or with the help of plug-ins, which is inconvenient and inefficient.

Under the above circumstances, it is important to study the MOO problem based on historical data for the manipulated variables and control variables, but this also raises the following questions: (1) How to handle contaminated data for modeling; (2) What kind of model to build in order to optimize; (3) How to design a MOO problem for a distillation column group.

In order to address these problems, a multi-objective collaborative optimization (MOCO) strategy based on system identification is proposed for an ethylene tower group. For the preprocessing of historical data, the Hampel method [19,20] is used to eliminate missing values and outliers. Gaussian filtering [21] is used to smooth the noise present in the data. Considering the direct or indirect connection between towers, a collaborative variable is proposed and applied. Fuzzy C-means clustering [22–24], a soft clustering method, is used to distinguish the operating modes of the ethylene plant. Selecting the data near the main operating mode, collaborative-variable-based linear discrete state-space models of the target towers are constructed using a subspace identification method [25]. A MOCO problem with the objectives of minimizing the impurity content in the ethylene product, the impurity content in the propylene product, and the operating cost is designed to optimize the top reflux flow rates and steam flow rates of bottom reboilers. NSGA-III [26] is used to solve the optimization problem and the effectiveness of MOCO is analyzed.

The remainder of this paper is organized as follows: Section 2 describes the ethylene plant under study. Section 3 describes a data-preprocessing scheme, determines the operating modes, and constructs state-space models of the target towers. The MOCO problem is described in Section 4. Section 5 provides a MOCO case study and discussion based on historical data. The conclusions are presented in Section 6.

2. Description of Ethylene Plant

The ethylene plant studied in this paper has a target annual production of 1 million tons of ethylene and 500,000 tons of propylene. Without considering the product systems of ethylene and propylene, the tower group for the distillation process consists of a de-ethanizer (T_1), a C2 hydrogenation reactor (R_1), a green oil tower (T_2), an ethylene distillation column (T_3), a low pressure depropanizer (T_4), a high pressure depropanizer (T_5), a C3 hydrogenation reactor (R_2), a propylene distillation column A (T_6), a propylene distillation column B (T_7), a debutanizer (T_8) and several dryers (DR) and heat exchangers (EX).

The tower group is shown in Figure 1. The main control variables of T_1 are the bottom C2 concentration (ethylene and ethane) and the top C3 concentration, which are measured by online analyzers on the bottom and top lines, respectively. The feed to T_1 is the demethanizer bottom product in the upstream tower group. Its top product goes to R_1 , where acetylene is selectively hydrogenated to produce ethylene and ethane. The bottom product of R_1 enters T_3 after passing through T_2 and a dryer. The main controlled variable

of T_3 is the impurity content in the top ethylene product; the ethylene extracted from the tower side line is pumped out partially to the cryogenic tank area. The T_4 bottom product is fed to T_8 and the top product is used as another feed to T_5 . The T_5 top product is fed to R_2 after a dryer and an arsenic-protected bed, and is eventually provided as a feed to T_7 . R_2 selectively hydrogenates the methylacetylene (MA) and propadiene (PD) from the T_5 top product to produce propylene and propane. The propylene distillation column is a two-tower system, with the top product of T_6 fed to T_7 and the bottom product used for recirculation. The main controlled variable of T_7 is the content of impurities in the top product, the bottom product is fed to T_6 and the product extracted from the tower side line is used for polymerization.

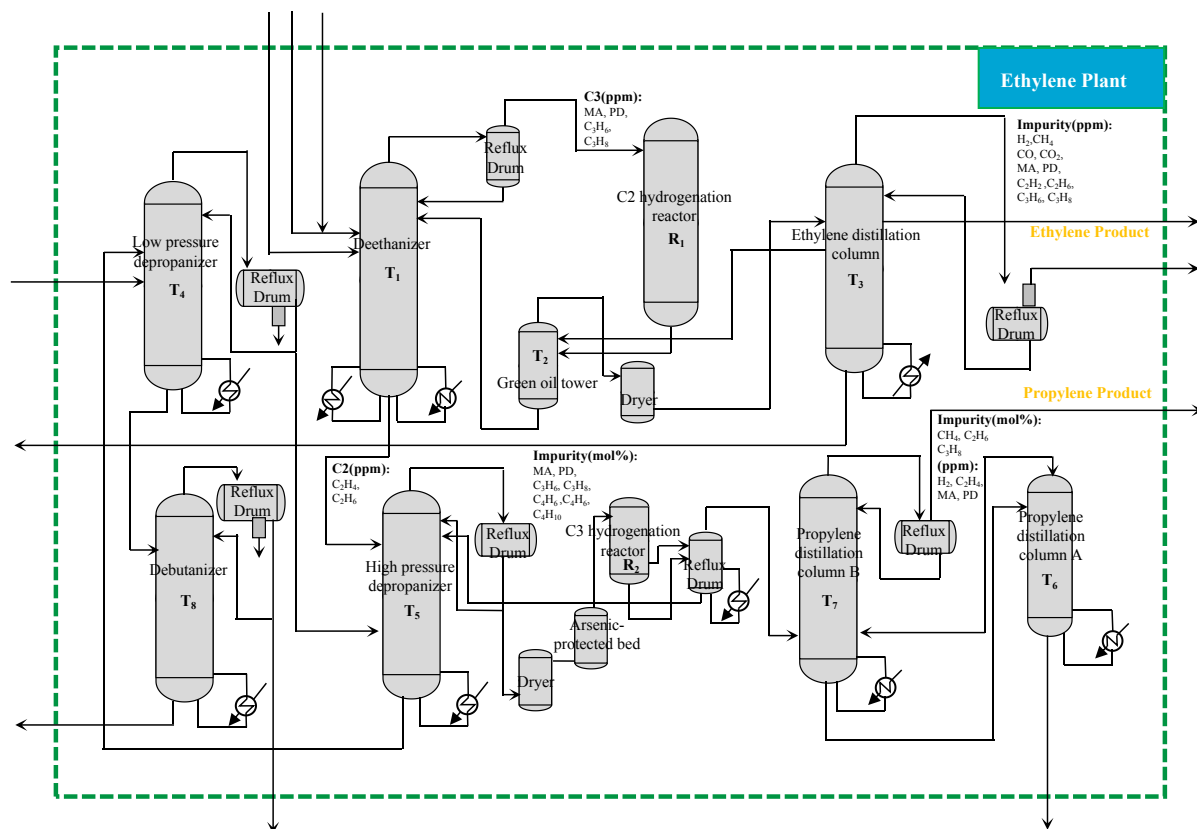


Figure 1. Diagrammatic sketch of an ethylene plant.

3. Model Construction

Data preprocessing is required before modeling with historical data, otherwise, it is difficult to obtain accurate models. On the premise of ensuring the stable operation of the tower group, the operators need to complete the target according to the plan. Since the actual operators change shifts by turns, the operation of the tower group will fluctuate, so it is necessary to identify the operating modes of the tower group. In addition, there is coupling between towers, which makes it difficult to build a complex model of a tower group. Establishing a linear model of each tower near a certain operating point is a good choice.

3.1. Data Preprocessing

The Hampel method [19,20] was used to process the missing values and outliers in one year's data for the manipulated and controlled variables for T_1 , T_3 , T_5 , T_6 , T_7 . Given the data

$\{x_1, x_2, \dots, x_n\}$, and a sliding window of length k , the i -th median and standard-deviation estimate are shown in (1), (2), respectively.

$$m_i = \text{median}(x_{i-k}, x_{i-k+1}, x_{i-k+2}, \dots, x_i, \dots, x_{i+k-2}, x_{i+k-1}, x_{i+k}) \quad (1)$$

$$\sigma_i = \kappa \text{ median}(|x_{i-k} - m_i|, \dots, |x_{i+k} - m_i|) \quad (2)$$

where $\kappa = \frac{1}{\sqrt{2}\text{erfc}^{-1}(1/2)} \approx 1.4826$. Near the sequence endpoints, the function truncating the window is used to compute (1), (2).

If a sample x_i satisfies (3) at a given threshold n_σ , then the sample is considered as an outlier and m_i is used to replace it.

$$|x_i - m_i| > n_\sigma \sigma_i \quad (3)$$

Since the controlled variables are all impurity contents measured by online analyzers, there is a certain amount of noise in the data. Therefore, the data of the controlled variables are smoothed using the Gaussian-weighted moving average [21] shown in (4).

$$\hat{x}_i = \sum_{j=\frac{p-1}{2}}^{i+\frac{p-1}{2}} w_j x_j \quad (4)$$

where \hat{x}_i is the filtered value, p is the window size, and w_j is the weighting factor of the Gaussian distribution.

3.2. Determination of Operating Modes

Since the towers show certain linear characteristics near the stable operating point, it is necessary to determine the operating modes of the tower group in order to facilitate the construction of the linear model of each tower. Considering that the target products of the ethylene plant are ethylene and propylene, the preprocessed data for the impurity content of the T_3 top product (D1), the impurity content on the 73rd tray of T_6 (D2), and the impurity content of the T_7 top product (D3) are selected and analyzed by a fuzzy C-means clustering method [22–24]. Using the membership function (MF), the degree of belongingness of a sample to a cluster can be quantified.

Figure 2 shows the performance of clustering when the number of clusters is set to two. The reason for setting the number of clusters to two is that, during the one-year production process, the operation of a tower group may be affected by people or equipment and deviate from the main operating mode, that is, the ideal operating conditions of the tower group.

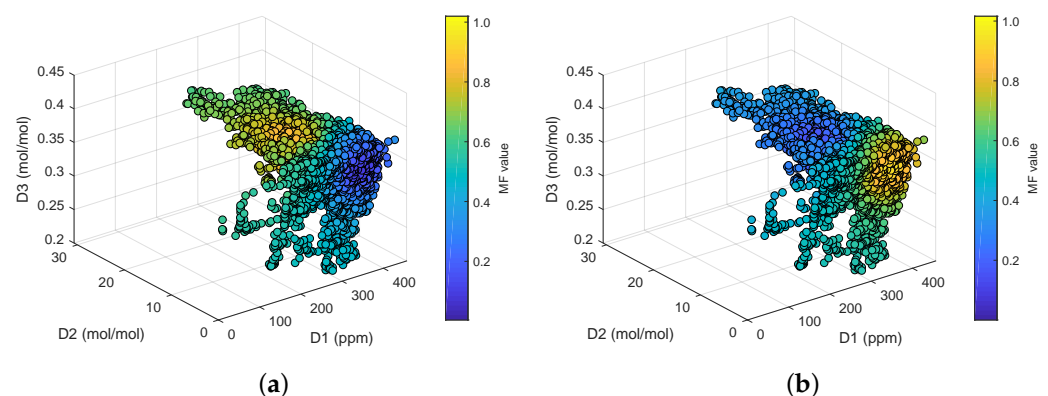


Figure 2. Performance of fuzzy C-means clustering: (a) Operating mode 1; (b) Operating mode 2 (main operating mode).

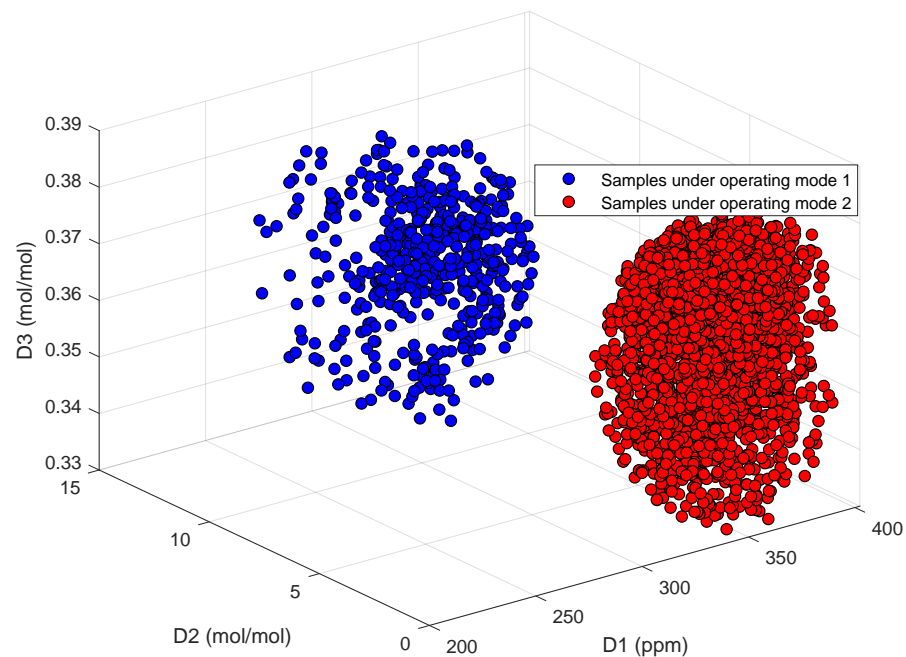


Figure 3. Samples with MF value above 0.8 under two operating modes.

Two clusters correspond to two operating modes; the data with an MF value greater than 0.8 are presented in Figure 3. The number of samples with an MF value above 0.8 under operating mode 1 is 490. In comparison, the number of data under operating mode 2 is larger, which is 2634. Therefore, we regard operating mode 2 as the main operating mode; the 2634 data are used to construct the subsequent tower models.

3.3. Subspace Identification

The linear discrete state-space model of each tower is constructed using the method of subspace identification [25]. Consider the following state-space model

$$\begin{bmatrix} \bar{x}(t+1) \\ y(t) \end{bmatrix} = \begin{bmatrix} A & B \\ C & D \end{bmatrix} \begin{bmatrix} \bar{x}(t) \\ u(t) \end{bmatrix} + \begin{bmatrix} \eta(t) \\ \nu(t) \end{bmatrix} \quad (5)$$

where $\bar{x} \in \mathbb{R}^n$ is the estimated state, $u \in \mathbb{R}^m$ is the input, $y \in \mathbb{R}^p$ is the output. η and ν are residuals. The system parameter of the regression model is $\Theta := \begin{bmatrix} A & B \\ C & D \end{bmatrix} \in \mathbb{R}^{(n+p) \times (n+m)}$, and its least squares estimation form is shown in (6).

$$\Theta = \left(\sum_{t=0}^{N-1} \begin{bmatrix} \bar{x}(t+1) \\ y(t) \end{bmatrix} \begin{bmatrix} \bar{x}^T(t) & u^T(t) \end{bmatrix} \right) \left(\sum_{t=0}^{N-1} \begin{bmatrix} \bar{x}(t) \\ u(t) \end{bmatrix} \begin{bmatrix} \bar{x}^T(t) & u^T(t) \end{bmatrix} \right)^{-1} \quad (6)$$

Equation (7) presents the covariance matrices of the residuals.

$$\begin{bmatrix} Q & S \\ S^T & R \end{bmatrix} = \frac{1}{N} \sum_{t=0}^{N-1} \begin{bmatrix} \eta(t) \\ \nu(t) \end{bmatrix} \begin{bmatrix} \eta^T(t) & \nu^T(t) \end{bmatrix} \quad (7)$$

The stable Kalman gain K can be obtained by solving an algebraic Riccati equation.

$$A^T X A - E^T X E - (A^T X B + S) (B^T X B + R)^{-1} (B^T X A + S^T) + Q = 0 \quad (8)$$

Finally, an innovation model (9) is obtained.

$$\begin{bmatrix} \hat{x}(t+1) \\ y(t) \end{bmatrix} = \begin{bmatrix} A & B \\ C & D \end{bmatrix} \begin{bmatrix} \hat{x}(t) \\ u(t) \end{bmatrix} + \begin{bmatrix} K \\ I_p \end{bmatrix} \hat{e}(t) \quad (9)$$

where \hat{x} is the estimate of the state vector, and \hat{e} is the estimate of the output prediction error.

3.4. Collaborative-Variable-Based State-Space Model

The towers considered for optimization are T_1 , T_3 , T_5 , T_7 . T_6 is the upper tower of the propylene final product. The content of its propylene product is associated with T_7 , so only T_7 is considered in the two-tower system propylene distillation column.

There often exists a direct (connection through pipelines) or indirect (connection without pipelines) connection between towers and this situation needs to be taken into account when building a single tower model. Therefore, we propose the concept of a collaborative variable—a variable that reflects the direct or indirect connection between two towers, which is a controlled variable of one tower and a “manipulated variable” of another tower. For example, the bottom product of T_1 is directly used as a feed to T_5 , the bottom C2 content of T_1 can be considered as a direct collaborative variable to T_5 , the top product of T_1 is hydrogenated and dried as a feed to T_3 , and the top C3 content of T_1 can be considered as an indirect collaborative variable to T_3 .

Variables in the state-space models are given in Table 1. For T_1 , the manipulated variables are the top reflux flow rate $u_{T_1}^1$ and the bottom reboiler steam flow rate $u_{T_1}^2$. The controlled variables are the bottom C2 content $y_{T_1}^1$ and the top C3 content $y_{T_1}^2$. For T_3 , the manipulated variables are the top reflux flow rate $u_{T_3}^1$ and the bottom reboiler steam flow rate $u_{T_3}^2$. The controlled variable is the impurity content y_{T_3} of the lateral line extraction product. For T_5 , the manipulated variables are the top reflux flow rate $u_{T_5}^1$ and the bottom reboiler steam flow rate $u_{T_5}^2$. The controlled variable is the impurity content y_{T_5} of the top product. For T_7 , the manipulated variables are the top reflux flow rate $u_{T_7}^1$ and the bottom reboiler steam flow rate $u_{T_7}^2$. The controlled variable is the impurity content y_{T_7} of the top product. Considering the connections between towers, the bottom C2 content of T_1 can be considered as a direct collaborative variable $c_{T_1}^1$ to T_5 , the top C3 content of T_1 can be considered as an indirect collaborative variable $c_{T_1}^2$ to T_3 , and the impurity content y_{T_5} can be considered as an indirect collaborative variable c_{T_5} to T_7 .

Table 1. Variables in models.

Tower	Manipulated Variables	Collaborative Variable(s)	Controlled Variable(s)
T_1	$[u_{T_1}^1, u_{T_1}^2]^T$	$c_{T_1}^1, c_{T_1}^2$	$y_{T_1}^1, y_{T_1}^2$
T_3	$[u_{T_3}^1, u_{T_3}^2]^T$		y_{T_3}
T_5	$[u_{T_5}^1, u_{T_5}^2]^T$	c_{T_5}	y_{T_5}
T_7	$[u_{T_7}^1, u_{T_7}^2]^T$		y_{T_7}

u : t/h; $c, y_{T_1}, y_{T_3}, y_{T_5}$: ppm; y_{T_7} : %/%

The *n4sid* function in MATLAB was used to construct the multi-input and single-output (MISO) model of the above towers, and the model with optimal normalized root mean squared error (NRMSE) was selected from the 2nd–10th-order model, respectively. The formulation of NRMSE is shown in (10).

$$NRMSE = \left(1 - \frac{\|y_{\text{measured}} - y_{\text{model}}\|}{\|y_{\text{measured}} - y_{\text{measured}}\|} \right) \times 100\% \quad (10)$$

The details of each MISO state-space model identified are shown in Table 2. It can be seen that the identified model has high accuracy and the adopted fuzzy C-means clustering method distinguishes the operating modes well.

Table 2. Parameters of models.

Output	Input	Initial State	e	NRMSE
$y_{T_1}^1$	$[u_{T_1}^1, u_{T_1}^2]^T$	$[-0.1287; 0.0088; -0.0987; 0.1041]$	4.57×10^{-4}	99.6%
$y_{T_1}^2$	$[u_{T_1}^1, u_{T_1}^2]^T$	$[-0.1749; 0.0151; -0.3372]$	3.36×10^{-2}	97.8%
y_{T_3}	$[u_{T_3}^1, u_{T_3}^2, c_{T_1}^1]^T$	$[1.4254; -0.0611; 0.7460; -12.1675]$	3.24×10^{-2}	97.6%
y_{T_5}	$[u_{T_5}^1, u_{T_5}^2, c_{T_1}^2]^T$	$[2.7991; 0.1510; -0.4182]$	5.24×10^{-5}	98.4%
y_{T_7}	$[u_{T_7}^1, u_{T_7}^2, c_{T_5}^1]^T$	$[-2.7744; 4.1782; 56.4836; 835.3279]$	6.70×10^{-9}	98.2%

e: Variance of white noise.

4. Construction of MOCO Problem

In order to facilitate giving the form of the optimization problem, considering one-step prediction, the expression of (9) is adjusted to (11).

$$\begin{aligned} x(T_s) &= Ax(0) + Bu(0) + Ke(0) \\ y(T_s) &= Cx(T_s) + Du(T_s) + e(T_s) \end{aligned} \quad (11)$$

where T_s denotes the discretization time, which in our case is 0.5 h. $x(0)$ denotes the initial state, $x(T_s)$ denotes the next state, and $y(T_s)$ denotes the predicted output of the next step.

The operating cost of the tower group considered is obtained by weighting the top reflux flow rate as well as the bottom reboiler steam flow rate of each tower. Since the data obtained is the content of impurities in the product, minimizing the top impurity content of T3 and T7 is equivalent to maximizing the ethylene and propylene content. The form of the optimization objective is shown in (12), which is a three objective optimization problem, i.e., minimizing the content of impurities in the propylene product f_1 , the content of impurities in the ethylene product f_2 and the total operating cost f_3 .

$$\min_{u_{T_i}^j, i \in I, j \in J} (f_1, f_2, f_3) \quad (12)$$

where $u_{T_i}^j$ denotes the j th manipulated variable of the i th tower, and I, J are the corresponding index sets in the form shown in (13). The form of each sub-objective in (12) is shown in (14).

$$i \in I = \{1, 3, 5, 7\}, \quad j \in J = \{1, 2\} \quad (13)$$

$$\begin{aligned} f_1 &= y_{T_7}(T_s) = C_{T_7}x_{T_7}(T_s) + D_{T_7}[u_{T_7}^1, u_{T_7}^2, c_{T_5}^1]^T + e_{T_7}(T_s) \\ f_2 &= y_{T_3}(T_s) = C_{T_3}x_{T_3}(T_s) + D_{T_3}[u_{T_3}^1, u_{T_3}^2, c_{T_1}^1]^T + e_{T_3}(T_s) \\ f_3 &= \sum_{i,j} R_{T_i}^j u_{T_i}^j \end{aligned} \quad (14)$$

R in (14) is the weight, c is the collaborative variable, and all collaborative variables are shown as

$$\begin{aligned} c_{T_5} &= y_{T_5}(T_s) = C_{T_5}x_{T_5}(T_s) + D_{T_5}[u_{T_5}^1, u_{T_5}^2, c_{T_1}^1]^T + e_{T_5}(T_s) \\ c_{T_1}^1 &= y_{T_1}^1(T_s) = C_{T_1}^1x_{T_1}^1(T_s) + D_{T_1}^1[u_{T_1}^1, u_{T_1}^2]^T + e_{T_1}^1(T_s) \\ c_{T_1}^2 &= y_{T_1}^2(T_s) = C_{T_1}^2x_{T_1}^2(T_s) + D_{T_1}^2[u_{T_1}^1, u_{T_1}^2]^T + e_{T_1}^2(T_s) \end{aligned} \quad (15)$$

The constraints are divided into manipulated variable constraints (16) and content constraints (17).

$$u_{T_i}^j \otimes [-1 \quad 1]^T \leq F_{T_i}^j \quad (16)$$

$$\begin{aligned} y_{T_1}^1 \otimes [-1 \quad 1]^T &\leq G_{T_1}^1 \\ y_{T_1}^2 \otimes [-1 \quad 1]^T &\leq G_{T_1}^2 \end{aligned} \quad (17)$$

where \otimes is the Kronecker product, and F and G are parameter vectors. The constraints in (17) are the bottom C2 and top C3 content constraint of T_1 .

Combining the above equations, the MOCO problem takes the form shown in (18).

$$\begin{aligned}
 & \min_{u_{T_i}^j, i \in I, j \in J} (f_1, f_2, f_3) \\
 & \text{s.t.} \quad \begin{aligned}
 & \text{Index sets defined by (13)} \\
 & \text{Sub-objectives defined by (14)} \\
 & \text{Collaborative variables defined by (15)} \\
 & \text{Manipulated variable constraints defined by (16)} \\
 & \text{Content constraints defined by (17)}
 \end{aligned}
 \end{aligned} \tag{18}$$

5. Case Study and Discussions

This section describes a MOCO case study for an ethylene plant based on historical data of a chemical plant. The data-sampling interval is 0.5 h, and the data for one year are collected, with 17,523 samples in total. The *Hampel* and *smoothdata* function in MATLAB are used for data preprocessing. k in *Hampel* is set to 20 and the standard deviations are set to 2. A *Gaussian* method is applied in *smoothdata* and the window size p is set to 20. A total of 7420 data points are left after data preprocessing, and 2634 data points are used for system identification. The parameters and parameter vectors in the constraints are shown in Table 3 and the coefficient matrices are given in Appendix A.

Table 3. Parameters in constraints.

Parameter		Parameter	
$R_{T_1}^1$	971.0	$F_{T_1}^2$	$[-775, 1703]^T$
$R_{T_1}^2$	27.2	$F_{T_3}^1$	$[-519, 591]^T$
$R_{T_3}^1$	1087.0	$F_{T_3}^2$	$[-135, 151]^T$
$R_{T_3}^2$	27.2	$F_{T_5}^1$	$[-7.9, 11.2]^T$
$R_{T_5}^1$	870.0	$F_{T_5}^2$	$[-15.7, 19.1]^T$
$R_{T_5}^2$	27.2	$F_{T_7}^1$	$[-738, 910]^T$
$R_{T_7}^1$	1004.3	$F_{T_7}^2$	$[-1963, 2372]^T$
$R_{T_7}^2$	27.2	$G_{T_1}^1$	$[-6.5, 34.2]^T$
$F_{T_1}^1$	$[-93, 116]^T$	$G_{T_1}^2$	$[-30.8, 116.5]^T$

R: \$/h; F: t/h; G: ppm.

Platemo, an evolutionary MOO platform developed by Tian et al. [27], is used to design and solve the MOCO problem. The algorithm chosen is NSGA-III [26], which uses adaptive updating of multiple reference points to maintain the population diversity. Compared to MOEA/D [28] and NSGA-II [29], it can give satisfactory results on more 2–15 objective problems.

The number of populations is 105 and 100 iterations are performed using NSGA-III. The final populations all meet the constraint requirements; the obtained Pareto frontiers are shown in Figure 4.

After 100 iterations, the final Pareto frontier approximation is presented as a plane which is shown in Figure 4. In order to increase the purity of the propylene product, it is necessary to increase the reflux flow rates and the reboiler steam flow rates of towers, which leads to an increase in cost. Therefore, the impurity content in the propylene product is negatively related to the total operating cost. However, the impurity content in the ethylene product does not show the above phenomenon, which is unreasonable. A MOO of the T_1 and T_3 is carried out to minimize the operating cost of both towers and the impurity content in the ethylene product; the Pareto frontier after 100 iterations is shown in Figure 5. Analysis of this figure shows that the impurity content in ethylene is negatively related

to the cost. The possible reasons for the above anomaly are that the cost of T_5 and T_7 represents a greater proportion of the total operating cost and the purity of propylene product has more room for improvement. Therefore, in our case, the impurity content in the ethylene product has no significant relationship with the total operating cost of the tower group. As for the relationship between propylene and ethylene content, since there is no connection between T_3 and the two propylene distillation columns, these two objectives appear to be relatively independent.

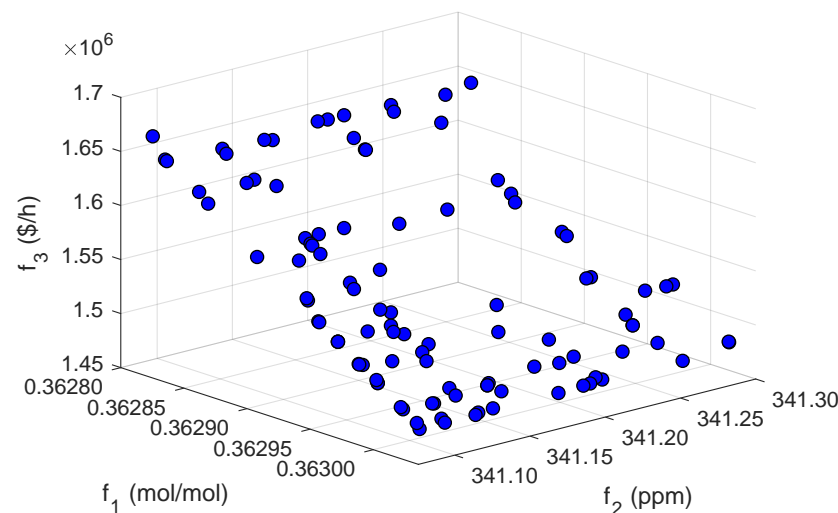


Figure 4. Pareto-optimal frontier for the three-objective optimization problem.

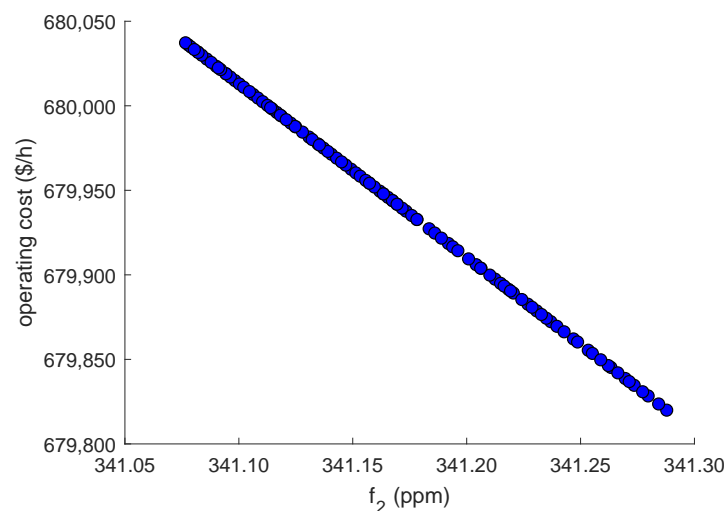


Figure 5. Pareto-optimal frontier for the two-objective optimization problem.

Corresponding to Figure 4, the optimal values of the decision variables after 100 iterations are shown in Figure 6. The subplots in each row of the figure correspond to the relationship between the same decision variable and different objectives, and the subplots in each column represent the relationships between the same objective and different decision variables. Analyzing the above three figures from the perspective of decision variables, it can be seen that the top reflux flow rate of T_3 and the bottom reboiler steam flow rate of T_7 are close to the lower boundary. For the objective of total operating cost f_3 , the weight of the T_3 and T_7 reflux flow rates are the largest and the magnitude of these two variables is also large. Thus, the smaller reflux flow rates of T_3 and T_7 imply lower cost. In our case, for the purpose of weighing the objectives and meeting the constraints of T_1 , NSGA-III tends to form a small T_3 reflux flow rate and to reduce the T_7 reflux flow rate.

Since the bottom reboiler steam flow rate magnitude of T_7 is the largest of all the decision variables in the solution set, this variable is also near the lower boundary. In the case of the three objective trade-offs, the solutions for the reflux flow rate of T_1 and T_5 , the reboiler steam flow rates of T_1 and T_5 are mostly close to the lower boundary, with a small number of solutions distributed in other regions. The solutions for the T_3 bottom reboiler steam flow rate, as well as the T_7 reflux flow rate, are the most uniformly distributed of all the manipulated variables.

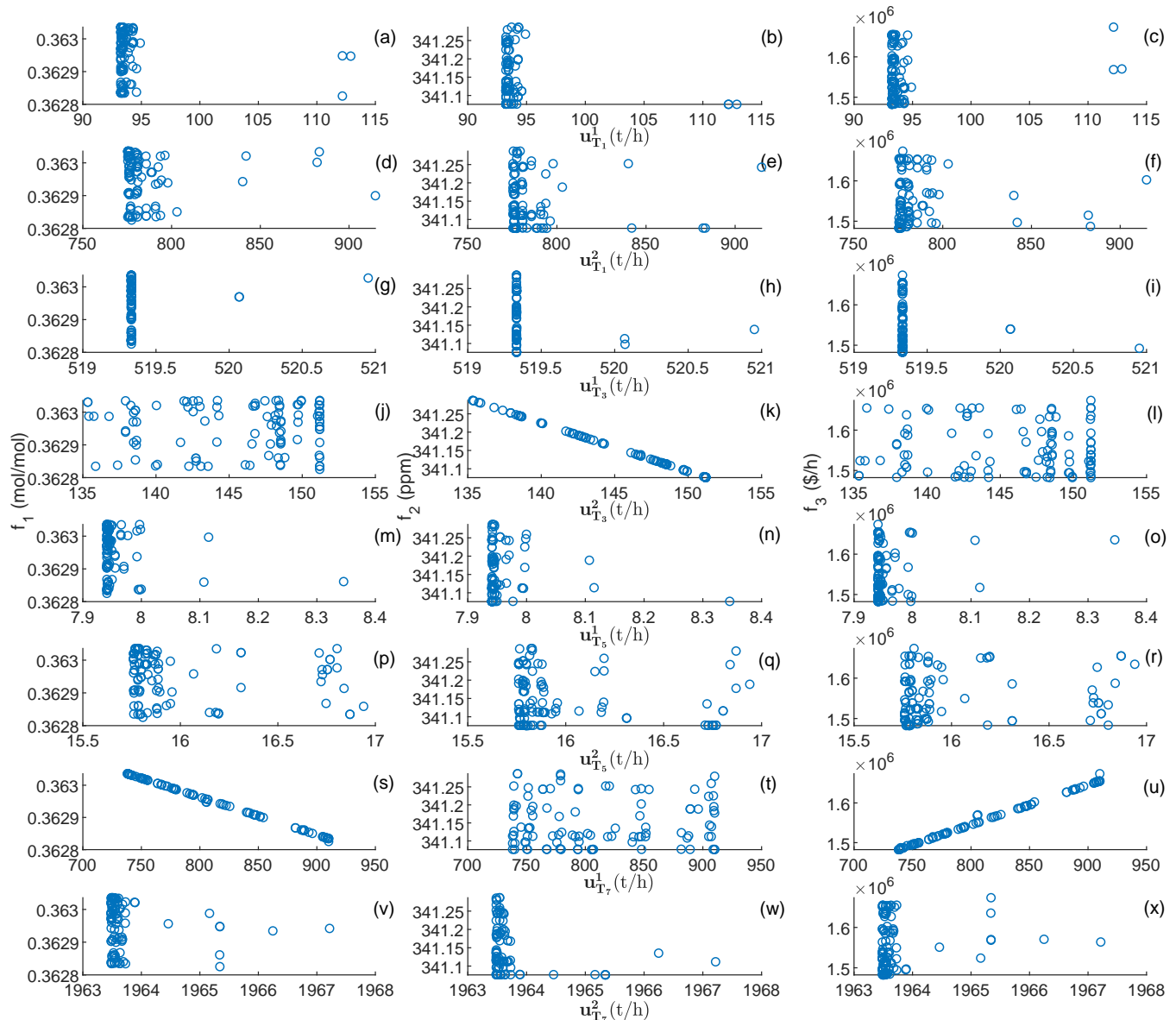


Figure 6. The solution set of the considered decision variables. (a–x) Optimal values of the considered decision variables corresponding to the Pareto-optimal front.

Analyzing Figure 6 from the perspective of the objective, an interesting phenomenon is that, although no connection exists between T_3 and T_7 , there is a complementary relationship between the T_3 bottom reboiler steam flow rate and the T_7 reflux flow rate, i.e., under the same objective, the solutions for one variable are irregularly related to the objective, and the other shows a linear relationship with the objective. The possible reason for the above phenomenon is that the NSGA-III algorithm selects the above two manipulated variables directly related to the objective to maintain the diversity of the population. Based on the small variation in the other decision variables, the algorithm selects the variable

with the larger influence on the target of the two variables to increase the range of target variation. Since both the constraints and the model are linear, the remaining one variable shows a certain linear relationship with the objective.

To verify the effectiveness of the proposed method, we further explored the sample data. Figure 7 shows the scatterplot corresponding to the sample data. It can be clearly seen that the total operating costs corresponding to the sample data are all above 1.5×10^6 \$/h. A sample data point with low cost and moderate impurity content is selected as the most preferred point. Figure 8 gives a comparison of this preference point with the Pareto frontier of the MOCO. The part in the dashed box is the appealing operation area because the total operating cost under this area is below 1.5×10^6 \$/h and the content of impurities in both ethylene and propylene products is lower than the preference point.

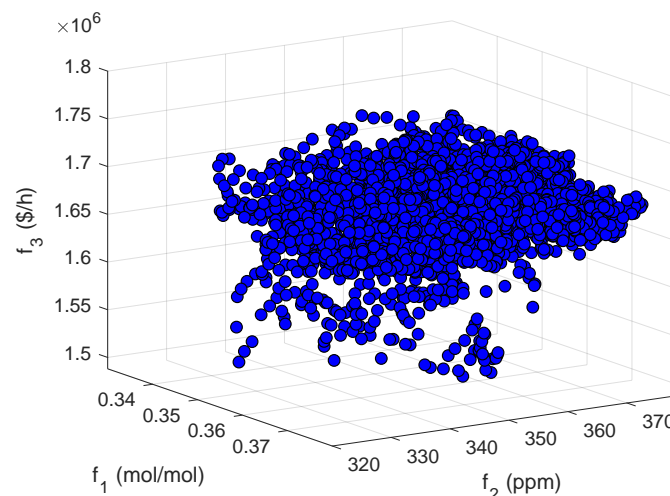


Figure 7. Distribution of objective values corresponding to data for modeling.

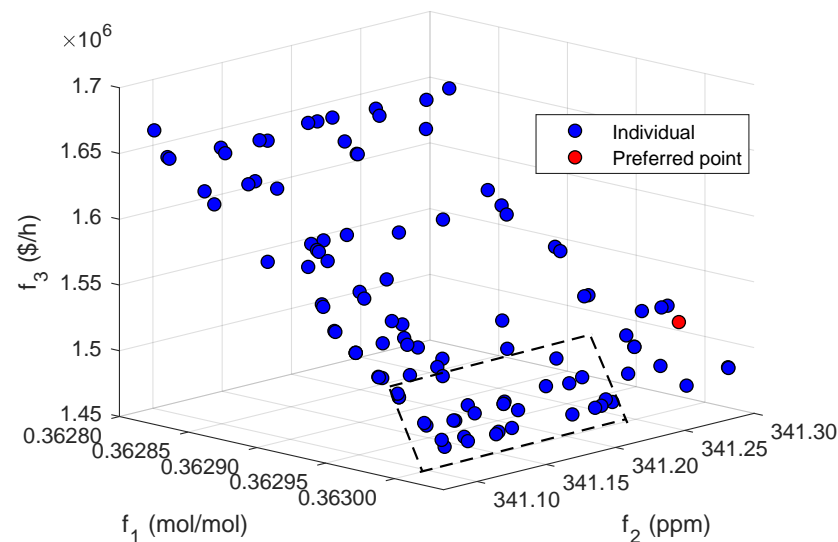


Figure 8. The appealing operation area in the optimal frontier under the guidance of a preference point.

The above analysis shows that the proposed MOCO strategy for the ethylene plant can provide rational support for decision-makers to obtain higher product purity at lower cost.

6. Conclusions

A multi-objective collaborative optimization (MOCO) strategy is proposed for a distillation column group. Based on the preprocessing of historical data, the data for the impurity

content of the ethylene and propylene products and the propylene content on the 73rd tray, are selected to determine the operating modes of the tower group by the fuzzy C-means clustering method. Data with a membership function value greater than 0.8 under the main operating mode are selected for subsequent modeling. The state-space models of towers are constructed based on a subspace identification method and collaborative variables. A MOCO case study is performed using MATLAB and Platemo,. The superiority of the proposed method in decision-making is verified by the Pareto-optimal frontier, population analysis and comparison with sample data.

Author Contributions: Formal analysis, K.J. and R.H.; funding acquisition, R.H. and J.L.; investigation, L.L.; methodology, R.H. and K.J.; project administration, R.H.; resources, R.H.; software, K.J.; supervision, R.H., J.L. and L.L.; validation, R.H.; writing—original draft, K.J.; writing—review and editing, L.L. and J.L. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by National Key Research and Development Program of China (2022YFB3305901), National Natural Science Foundation of China (62073142, 62136003, 61973124), Fundamental Research Funds for the Central Universities and the Programme of Introducing Talents of Discipline to Universities (the 111 Project) under Grant B17017.

Data Availability Statement: The data presented in this study are available in the manuscript.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A

Coefficient matrices of models.

$$A_{T_1}^1 = \begin{bmatrix} 0.9977 & -0.042 & 4.2 \times 10^{-3} & -1.7 \times 10^{-3} \\ 0.0381 & 0.9720 & -0.217 & -0.023 \\ 3.4 \times 10^{-4} & 0.1954 & 0.8943 & -0.383 \\ -0.012 & 4.5 \times 10^{-3} & 0.3584 & 0.2068 \end{bmatrix} \quad (A1)$$

$$A_{T_1}^2 = \begin{bmatrix} 0.9982 & -0.069 & 7.9 \times 10^{-3} \\ 0.0678 & 0.9676 & -0.272 \\ -7.7 \times 10^{-4} & 0.2259 & 0.7152 \end{bmatrix} \quad (A2)$$

$$A_{T_3} = \begin{bmatrix} 0.9923 & -0.150 & 9.4 \times 10^{-3} & -2.2 \times 10^{-3} \\ 0.1480 & 0.9510 & -0.334 & -9.6 \times 10^{-3} \\ 0.0204 & 0.2218 & 0.6909 & -0.026 \\ -0.247 & 0.2824 & 1.9029 & -0.033 \end{bmatrix} \quad (A3)$$

$$A_{T_5} = \begin{bmatrix} 1.0011 & -0.020 & -1.8 \times 10^{-3} \\ 0.0340 & 0.9868 & 0.2708 \\ -0.018 & -0.222 & 0.7400 \end{bmatrix} \quad (A4)$$

$$A_{T_7} = \begin{bmatrix} 0.9879 & -0.129 & 0.0131 & -2.6 \times 10^{-3} \\ 0.1312 & 0.9710 & -0.241 & -0.011 \\ 7.9 \times 10^{-3} & 0.1992 & 0.6855 & -0.1001 \\ -0.178 & -0.165 & -1.173 & -0.4213 \end{bmatrix} \quad (A5)$$

$$B_{T_1}^1 = \begin{bmatrix} 3.5 \times 10^{-6} & 1.1 \times 10^{-8} \\ -1.5 \times 10^{-4} & 1.3 \times 10^{-7} \\ 1.4 \times 10^{-4} & -5.6 \times 10^{-6} \\ 8.0 \times 10^{-4} & -1.4 \times 10^{-5} \end{bmatrix} \quad (A6)$$

$$B_{T_1}^2 = \begin{bmatrix} 3.1 \times 10^{-5} & 1.4 \times 10^{-8} \\ -6.7 \times 10^{-4} & -5.1 \times 10^{-7} \\ -9.0 \times 10^{-4} & 4.1 \times 10^{-6} \end{bmatrix} \quad (A7)$$

$$B_{T_3} = \begin{bmatrix} 2.0 \times 10^{-6} & -1.9 \times 10^{-5} & -4.3 \times 10^{-4} \\ 1.3 \times 10^{-5} & 3.3 \times 10^{-4} & -2.1 \times 10^{-3} \\ 2.7 \times 10^{-4} & 1.1 \times 10^{-4} & -4.0 \times 10^{-3} \\ -2.6 \times 10^{-3} & -0.023 & -0.126 \end{bmatrix} \quad (A8)$$

$$B_{T_5} = \begin{bmatrix} -6.0 \times 10^{-5} & -2.4 \times 10^{-5} & 1.1 \times 10^{-5} \\ 9.3 \times 10^{-4} & -1.1 \times 10^{-3} & 5.5 \times 10^{-4} \\ -1.4 \times 10^{-3} & -8.7 \times 10^{-5} & -4.3 \times 10^{-4} \end{bmatrix} \quad (A9)$$

$$B_{T_7} = \begin{bmatrix} 3.7 \times 10^{-6} & -6.5 \times 10^{-7} & 0.0194 \\ -2.7 \times 10^{-5} & 6.3 \times 10^{-6} & 0.2329 \\ 5.4 \times 10^{-6} & -2.3 \times 10^{-6} & 1.0189 \\ -3.5 \times 10^{-4} & 5.8 \times 10^{-5} & 12.696 \end{bmatrix} \quad (A10)$$

$$\begin{aligned} C_{T_1}^1 &= [-239.9 \quad 5.0507 \quad -0.902 \quad 0.1595] \\ C_{T_1}^2 &= [-396.2 \quad 13.521 \quad -2.921] \\ C_{T_3} &= [232.48 \quad -17.16 \quad 3.7044 \quad 0.1595] \\ C_{T_5} &= [35.386 \quad -0.357 \quad -0.071] \\ C_{T_7} &= [-0.146 \quad 9.5 \times 10^{-3} \quad -1.7 \times 10^{-3} \quad 9.5 \times 10^{-5}] \end{aligned} \quad (A11)$$

$$\begin{aligned} D_{T_1}^1 &= [-5.8 \times 10^{-3} \quad -5.5 \times 10^{-6}] \\ D_{T_1}^2 &= [-0.012 \quad -2.5 \times 10^{-5}] \\ D_{T_3} &= [5.0 \times 10^{-4} \quad -6.3 \times 10^{-3} \quad 0.2502] \\ D_{T_5} &= [-1.1 \times 10^{-3} \quad 1.5 \times 10^{-4} \quad -8.9 \times 10^{-3}] \\ D_{T_7} &= [-3.4 \times 10^{-7} \quad 1.1 \times 10^{-7} \quad -6.7 \times 10^{-4}] \end{aligned} \quad (A12)$$

$$\begin{aligned} E_{T_1}^1 &= [-5.6 \times 10^{-3} \quad 0.0937 \quad -0.186 \quad 0.0519]^T \\ E_{T_1}^2 &= [-2.2 \times 10^{-3} \quad 0.0166 \quad -0.010]^T \\ E_{T_3} &= [4.8 \times 10^{-3} \quad -0.025 \quad 0.0305 \quad 0.0135]^T \\ E_{T_5} &= [0.0245 \quad -0.596 \quad -0.407]^T \\ E_{T_7} &= [-6.958 \quad 53.443 \quad -60.08 \quad 206.98]^T \end{aligned} \quad (A13)$$

References

1. Mayer, M.; Szilágyi, A.; Gróf, G. Environmental and economic multi-objective optimization of a household level hybrid renewable energy system by genetic algorithm. *Appl. Energy* **2020**, *269*, 115058.
2. Nuvvula, R.; Devaraj, E.; Madurai Elavarasan, R.; Iman Taheri, S.; Irfan, M.; Teegala, K. Multi-objective mutation-enabled adaptive local attractor quantum behaved particle swarm optimisation based optimal sizing of hybrid renewable energy system for smart cities in India. *Sustain. Energy Technol. Assess.* **2022**, *49*, 101689. [\[CrossRef\]](#)
3. Chen, T.; Cheng, C.; Chou, Y. Multi-objective genetic algorithm for energy-efficient hybrid flow shop scheduling with lot streaming. *Ann. Oper. Res.* **2020**, *290*, 813–836. [\[CrossRef\]](#)
4. Wang, X.; Mao, X.; Khodaei, H. A multi-objective home energy management system based on internet of things and optimization algorithms. *J. Build. Eng.* **2021**, *33*, 101603. [\[CrossRef\]](#)
5. Karimi, H.; Jadid, S. Optimal energy management for multi-microgrid considering demand response programs: A stochastic multi-objective framework. *Energy* **2020**, *195*, 116992. [\[CrossRef\]](#)
6. Khanali, M.; Akram, A.; Behzadi, J.; Mostashari-Rad, F.; Saber, Z.; Chau, K.; Nabavi-Pelesaraei, A. Multi-objective optimization of energy use and environmental emissions for walnut production using imperialist competitive algorithm. *Appl. Energy* **2020**, *284*, 116342. [\[CrossRef\]](#)
7. Sohani, A.; Dehnavi, A.; Sayyaadi, H.; Hoseinzadeh, S.; Goodarzi, E.; Garcia, D.; Groppi, D. The real-time dynamic multi-objective optimization of a building integrated photovoltaic thermal (BIPV/T) system enhanced by phase change materials. *J. Energy Storage* **2022**, *46*, 103777. [\[CrossRef\]](#)

8. Cui, Y.; Geng, Z.; Zhu, Q.; Han, Y. Review: Multi-objective optimization methods and application in energy saving. *Energy* **2017**, *125*, 681–704. [\[CrossRef\]](#)
9. Zhou, L.; Liao, Z.; Wang, L.; Zhang, L.; Ji, X.; Jiao, H.; Wang, J.; Yang, Y.; Dang, Y. Simulation-Based Multiobjective Optimization of the Product Separation Process within an MTP Plant. *Ind. Eng. Chem. Res.* **2019**, *58*, 12166–12178. [\[CrossRef\]](#)
10. Zhang, Y.; He, N.; Masuku, C.; Biegler, L. A multi-objective reactive distillation optimization model for Fischer–Tropsch synthesis. *Comput. Chem. Eng.* **2020**, *135*, 106754. [\[CrossRef\]](#)
11. Zhang, H.; Wang, S.; Tang, J.; Li, N.; Li, Y.; Cui, P.; Wang, Y.; Zheng, S.; Zhu, Z.; Ma, Y. Multi-objective optimization and control strategy for extractive distillation with dividing-wall column/pervaporation for separation of ternary azeotropes based on mechanism analysis. *Energy* **2021**, *229*, 120774.
12. Mondal, B.; Rangaiah, G.; Jana, A. Optimizing algal biodiesel production from a novel reactive distillation based unit: Reducing CO₂ emission and cost. *Chem. Eng. Process. Process. Intensif.* **2022**, *176*, 108948. [\[CrossRef\]](#)
13. Gu, J.; Lu, S.; Shi, F.; Wang, X.; You, X. Economic and Environmental Evaluation of Heat-Integrated Pressure-Swing Distillation by Multiobjective Optimization. *Ind. Eng. Chem. Res.* **2022**, *61*, 9004–9014.
14. Kruber, K.; Skiborowski, M. Topology-Based Initialization for the Optimization-Based Design of Heteroazeotropic Distillation Processes. *Processes* **2022**, *10*, 1482.
15. Deshpande, G.; Shrikhande, S.; Patle, D.; Sawarkar, A. Simultaneous optimization of economic, environmental and safety criteria for algal biodiesel process retrofitted using dividing wall column and multistage vapor recompression. *Process Saf. Environ. Prot.* **2022**, *164*, 1–14. [\[CrossRef\]](#)
16. Pandit, S.; Jana, A. Transforming conventional distillation sequence to dividing wall column: Minimizing cost, energy usage and environmental impact through genetic algorithm. *Sep. Purif. Technol.* **2022**, *297*, 121437. [\[CrossRef\]](#)
17. Shen, F.; Wang, M.; Huang, L.; Qian, F. Exergy analysis and multi-objective optimisation for energy system: A case study of a separation process in ethylene manufacturing. *J. Ind. Eng. Chem.* **2021**, *93*, 394–406.
18. Dai, M.; Yang, F.; Zhang, Z.; Liu, G.; Feng, X. Energetic, economic and environmental (3E) multi-objective optimization of the back-end separation of ethylene plant based on adaptive surrogate model. *J. Clean. Prod.* **2021**, *310*, 127426. [\[CrossRef\]](#)
19. Liu, H.; Shah, S.; Jiang, W. On-line outlier detection and data cleaning. *Comput. Chem. Eng.* **2004**, *28*, 1635–1647. [\[CrossRef\]](#)
20. Suomela, J. Median Filtering is Equivalent to Sorting. *arXiv* **2014**, arXiv:1406.1717.
21. Sharifi, S.; Hendry, M.; Macciotta, R.; Evans, T. Evaluation of filtering methods for use on high-frequency measurements of landslide displacements. *Nat. Hazards Earth Syst. Sci.* **2022**, *21*, 411–430. [\[CrossRef\]](#)
22. Bezdek, J. *Pattern Recognition with Fuzzy Objective Function Algorithms*, 1st ed.; Springer: Boston, MA, USA, 1981; pp. 65–79.
23. Sharifi, S.; Mazumdar, A.; Pal, S. Fuzzy Clustering with Similarity Queries. *Adv. N.A. Inf. Process. Syst.* **2021**, *34*, 789–801.
24. Huleihel, W.; Miriyala, S.; Mitra, K. An Evolutionary Neuro-Fuzzy C-means Clustering Technique. *Eng. Appl. Artif. Intell.* **2020**, *89*, 103435.
25. Katayama, T. *Subspace Methods for System Identification*, 1st ed.; Springer: Berlin, Germany; London, UK, 2005; pp. 6–9.
26. Deb, K.; Jain, H. An Evolutionary Many-Objective Optimization Algorithm Using Reference-Point-Based Nondominated Sorting Approach, Part I: Solving Problems With Box Constraints. *IEEE Trans. Evol. Computat.* **2014**, *18*, 577–601. [\[CrossRef\]](#)
27. Tian, Y.; Cheng, R.; Zhang, X.; Jin, Y. PlatEMO: A MATLAB Platform for Evolutionary Multi-Objective Optimization [Educational Forum]. *IEEE Comput. Intell. Mag.* **2017**, *12*, 73–87. [\[CrossRef\]](#)
28. Zhang, Q.; Li, H. MOEA/D: A Multiobjective Evolutionary Algorithm Based on Decomposition. *IEEE Trans. Evol. Computat.* **2007**, *11*, 712–731. [\[CrossRef\]](#)
29. Deb, K.; Agrawal, S.; Pratap, A.; Meyarivan, T. A Fast Elitist Non-dominated Sorting Genetic Algorithm for Multi-objective Optimization: NSGA-II. *IEEE Trans. Evol. Computat.* **2002**, *6*, 182–197. [\[CrossRef\]](#)

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.