Chinese Journal of Chemical Engineering 85 (2025) 140–157

Contents lists available at ScienceDirect

Chinese Journal of Chemical Engineering

journal homepage: www.elsevier.com/locate/CJChE

Full Length Article

Establishment of normal operating zone models by boundary points for CSTR-DC-recycle chemical processes

Poku Gyasi , Jiandong Wang *, Mengyao Wei , Hao Jing

Department of Electrical Engineering and Automation, Shandong University of Science and Technology, Qingdao 266590, China

ARTICLE INFO

Article history:
Received 5 December 2024
Received in revised form
16 March 2025
Accepted 21 March 2025
Available online 24 April 2025

Keywords:
Chemical processes
Grey-box model
Normal operating zone
Bayesian estimation
Model uncertainty measurement
Boundary points

ABSTRACT

Integrated continuous stirred-tank reactors and distillation columns with recycle (CSTR-DC-recycle) are essential components in chemical processes. This paper proposes a method to establish a normal operating zone (NOZ) model to represent allowable variations of the CSTR-DC-recycle chemical processes. The NOZ is a geometric space containing all safe operating points of the CSTR-DC-recycle chemical processes, so that it is an effective model for process monitoring. The novelty of the proposed method is to establish the NOZ model based on boundary points. The boundary points make it possible to capture the actual geometric space irrespective of the space shape. In contrast, existing methods represent the NOZ of processes by fixed mathematical models such as ellipsoidal and convex-hull models; they are not suitable for the CSTR-DC-recycle chemical processes whose NOZs cannot be exactly defined by fixed mathematical structures. Simulated case studies based on Aspen Hysys software are given to illustrate the proposed method.

1. Introduction

The integrated continuous stirred-tank reactors (CSTR) and distillation column (DC) with recycle (CSTR-DC-recycle) chemical processes are central to many industrial chemical plants [1]. Such processes are economical since they allow for simultaneous conversion of reactants, separation of desired products, and recycling of unreacted reactants [2–4]. Monitoring the CSTR-DC-recycle chemical processes is essential to ensure safety process and product quality [5–8]. Hence, efficient monitoring of the CSTR-DC-recycle chemical processes has attracted the of academic and industrial communities [9–11].

A common way to monitor chemical processes is to separately monitor each process variable by comparing their measured values against predefined allowable limits, referred to as univariate monitoring [9,12–15]. The allowable limits of process variables formulate a multidimensional rectangular box or hypercube without an emphasis on the inter-relationship among the process variables [9]. For illustration, Fig. 1 gives two related process variables x_1 and x_2 that are monitored based on allowable limits $[x_{1,\min}, x_{1,\max}]$ and $[x_{2,\min}, x_{2,\max}]$, respectively. The allowable limits formulate a rectangular box given by the red dash lines. This rectangular space does not consider the relationship between x_1 and x_2. However, chemical process variables are usually inter-related *via* physical models, *e.g.*, mass and energy balances, such that the interactions between the process variables give a non-rectangular normal operating zone (NOZ) [9,16–18] shown by the blue lines in Fig. 1.

The NOZ is defined as a geometric space consisting of all safe operating points. Consequently, the operating points inside and outside the NOZ are respectively in the normal and abnormal conditions. This means that the NOZ is an effective model for monitoring the process operating conditions. In Fig. 1, the NOZ correctly detects an abnormality represented by the red dot at *A* since it considers the interaction between x_1 and x_2. In contrast, the univariate rectangular box wrongly detects *A* being in the normal condition. Thus, the univariate rectangular box is not appropriate for monitoring multivariate chemical processes whose process variables are related like the CSTR-DC-recycle processes.

Hence, it is reasonable to consider the interactions between the process variables in defining related limits to monitor chemical processes such as the CSTR-DC-recycle processes.

* Corresponding author.
E-mail address: jiandong@sdust.edu.cn (J. Wang).

https://doi.org/10.1016/j.cjche.2025.03.012

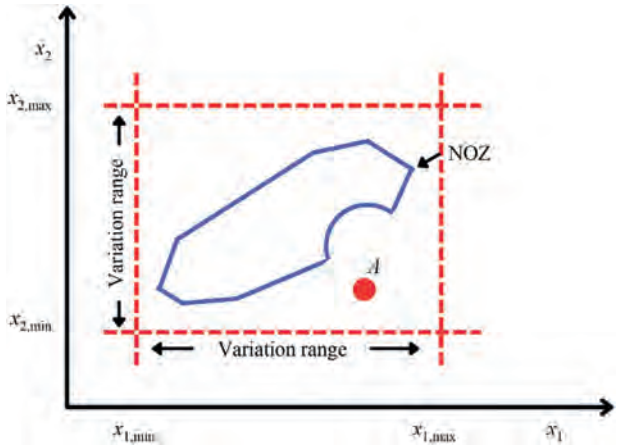

Fig. 1. Illustrating the NOZ (given by blue lines) and the rectangular box formed by the individual allowable limits of the process variables x_1 and x_2.

To address the above limitations of univariate monitoring methods, Brooks *et al.* [9] monitored multivariate processes using best operating zone formulated from historic data by considering the process operational requirements. They identified the manipulated variables to be adjusted in the parallel coordinate. However, they did not provide specific technical details. Inspired by the works of [9], Yu *et al.* [17] and Yu and Wang [18] respectively monitored multivariate processes based on NOZ represented by fixed ellipsoidal and convex-hull models established from historical data without paying attention to the process knowledge. However, by considering the process knowledge, the NOZ model of the CSTR-DC-recycle chemical processes may be irregular in shape that cannot be exactly defined by the fixed ellipsoidal and convex-hull models. Hence, using fixed ellipsoidal and convex-hull models may reduce the monitoring accuracy for the CSTR-DC-recycle chemical processes.

The main contribution of this paper is to establish the NOZ models of the CSTR-DC-recycle chemical processes based on boundary points by considering the process knowledge. This is achieved in four major steps. First, the grey-box model of the CSTR-DC-recycle chemical process required to establish the NOZ model is built from material balances and historic data samples of process variables. Second, the uncertainty of the grey-box model is measured based on Bayesian estimation rule to ensure that the grey-box model accurately represents the CSTR-DC-recycle chemical process. Third, the NOZ is formulated by simulating operating points using the grey-box model based on historic data samples of input process variables and selecting the operating points that satisfy the process operational requirement. The boundary points of the formulated NOZ are then extracted to establish the NOZ model of the CSTR-DC-recycle chemical process. Fourth, the accuracy of the established NOZ model is measured *via* Bayesian estimation rule and Euclidean distances.

The key feature that distinguishes the proposed method from existing methods is the exploitation of boundary points based on process knowledge to represent the NOZ model of the CSTR-DC-recycle chemical processes. In contrast, existing methods represent the NOZ models of processes by fixed mathematical models such as the ellipsoidal and convex-hull models. The fixed mathematical models are unable to represent the true NOZ model of the CSTR-DC-recycle chemical processes whose NOZ models cannot be exactly defined by fixed mathematical models. By exploiting the boundary points, the proposed method is able to capture the true NOZ model of the CSTR-DC-recycle chemical processes irrespective of its shape. The proposed method accurately detects abnormalities, whereas the fixed mathematical models lead to misclassification of abnormalities for the CSTR-DC-recycle chemical process. As shown later in Fig. 17, the normal operating point H is misclassified as abnormal by the fixed ellipsoidal model, while the abnormal operating point G is incorrectly identified as normal by the fixed convex-hull model.

The paper is arranged as follows. The problem to be addressed is described in Section 2. The details of the proposed method are given in Section 3. Section 4 presents simulated case studies, while Section 5 makes concluding remarks.

2. Problem Description

Fig. 2 gives the schematic diagram of CSTR-DC-recycle chemical process plant for propylene glycol production, having 29 process variables given in Table 1. In the CSTR reactor, propylene oxide (C_3H_6O) undergoes hydrolysis in excess water (H_2O) to produce propylene glycol [19–21]. Eq. (1a) gives the balanced chemical reaction for producing propylene glycol with two unwanted side reactions in Eq. (1b) and Eq. (1c) [20,22].

$$\underset{(A)}{C_3H_6O} + \underset{(B)}{H_2O} \xrightarrow{H_2SO_4} \underset{(C)}{C_3H_8O_2} \tag{1a}$$

$$C_3H_6O + C_3H_8O_2 \xrightarrow{H_2SO_4} C_6H_{14}O_3 \tag{1b}$$

$$C_3H_6O + C_6H_{14}O_3 \xrightarrow{H_2SO_4} C_9H_{20}O_4 \tag{1c}$$

The reaction is irreversible, exothermic, and first-order with respect to the limiting reactant C_3H_6O [3]. Since the reaction is exothermic, the reactor is covered with a cooling water jacket to keep the temperature within the desired temperature range [19]. The reacting mixture in the reactor is uniformly mixed to ensure no gradients in density, concentration, or temperature. The reactor holdup is held constant so that the inlet flow rate is equal to the reactor output flow rate.

The distillation column receives inputs from the reactor and separates the propylene glycol ($C_3H_8O_2$) from the unreacted reactants C_3H_6O and H_2O [19,23,24]. It is composed of a reboiler that provides energy for vaporizing the low boiling point component C_3H_6O and a condenser that condenses the distillate into liquid. The distillate is received downstream of the condenser, while the bottom product is received at the bottom of the column [19]. A portion of the distillate is refluxed back to the distillation column while the remaining fraction is recycled back to the reactor.

The critical process variables that influence the entire CSTR-DC-recycle chemical process are F_i, C_{A_i}, T_o and C_{C_L}. The operating points of F_i, C_{A_i}, T_o and C_{C_L} that conform to the process operational requirement $C_{C_L,min} \leq C_{C_L} \leq C_{C_L,max}$ formulate the NOZ which is a 4-dimensional geometric space encompassing all the normal operating points of the process [9,25,26]. Thus, the NOZ model can be used for process monitoring [9,18]. The NOZ models are usually represented by fixed mathematical models such as ellipsoid and convex-hull models. However, these fixed mathematical models do not consider the process mechanism and operational requirements; hence, they are unable to represent the true NOZ model of the CSTR-DC-recycle chemical process whose NOZ is not in the shape of ellipsoid and convex-hull models.

The main objective of this paper is to establish a NOZ modeled by boundary points for the CSTR-DC-recycle chemical process by considering the process knowledge so that the operational requirement $C_{C_L,min} \leq C_{C_L} \leq C_{C_L,max}$ can be monitored.

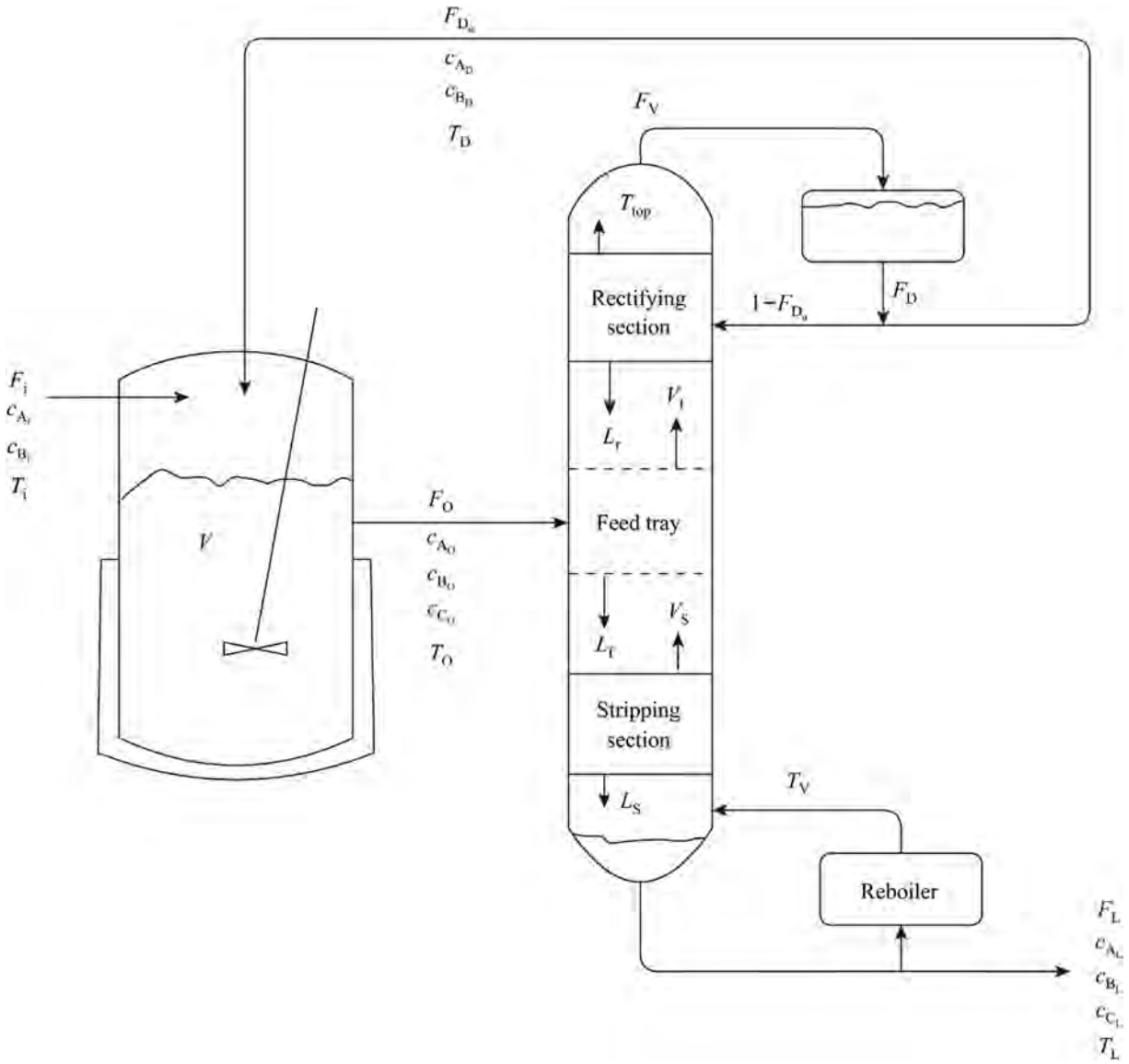

Fig. 2. Schematic diagram of propylene glycol production plant with recycle.

3. Proposed Method

The specific steps of the proposed method are presented in this section. First, the grey-box model is built for the CSTR-DC-recycle chemical process for propylene glycol production. Second, the uncertainty of the grey-box model is measured. Third, the NOZ model is established. Fourth, the accuracy of established NOZ model is determined.

3.1. Grey-box modeling for the CSTR-DC-recycle chemical process

The CSTR-DC-recycle chemical process model is required for establishing the NOZ model and is therefore built in this subsection.

The CSTR-DC-recycle chemical process model is obtained based on decomposition, whereby each unit operation is modeled separately, and the input of the proceeding unit operation is the output of the preceding unit operation. Since the reactor holdup (V) is constant, the output flow rate (F_o) is equal to the sum of input and recycle flow rates. Thus, the overall material balance around the CSTR is given by

$$F_o = F_i + F_{D_\alpha} \tag{2}$$

Here $F_{D_\alpha} = F_D \times \alpha$ is recycle flow rate where F_D and α are the distillate flow rate and recycle ratio, respectively. For the CSTR with C_3H_6O, H_2O and $C_3H_8O_2$ respectively represented by the symbols A, B and C in Eq. (1a), the dynamic component balances including the recycle flow are

$$V\frac{dC_{A_o}}{dt} = F_i C_{A_i} + F_{D_\alpha} C_{A_D} - F_o C_{A_o} + V\left(-A_o e^{-\frac{E}{RT_o}} C_{A_o}\right) \tag{3}$$

$$V\frac{dC_{B_o}}{dt} = F_i C_{B_i} + F_{D_\alpha} C_{B_D} - F_o C_{B_o} + V\left(-A_o e^{-\frac{E}{RT_o}} C_{A_o}\right) \tag{4}$$

$$V\frac{dC_{C_o}}{dt} = -F_o C_{C_o} + V\left(-A_o e^{-\frac{E}{RT_o}} C_{A_o}\right) \tag{5}$$

Table 1
Process variables.

Variable	Description	Type	Value
F_i	Input flow rate/$m^3 \cdot h^{-1}$	Input	Manipulated
C_{A_i}	Input concentration of propylene oxide/$kmol \cdot m^{-3}$	Input	Manipulated
C_{B_i}	Input concentration of water/$kmol \cdot m^{-3}$	Input	Fixed
T_i	Input temperature/K	Input	Fixed
F_m	Mixer output flowrate/$m^3 \cdot h^{-1}$	Output	Depends on input
C_{A_m}	Mixer output concentration of propylene oxide/$kmol \cdot m^{-3}$	Output	Depends on input
C_{B_m}	Mixer output concentration of water/$kmol \cdot m^{-3}$	Output	Depends on input
C_{C_m}	Mixer output concentration of propylene glycol/$kmol \cdot m^{-3}$	Output	Depends on input
T_m	Mixer output temperature/K	Output	Depends on input
F_o	Reactor output flowrate/$m^3 \cdot h^{-1}$	Output	Depends on input
C_{A_o}	Reactor output concentration of propylene oxide/$kmol \cdot m^{-3}$	Output	Depends on input
C_{B_o}	Reactor output concentration of water/$kmol \cdot m^{-3}$	Output	Depends on input
C_{C_o}	Reactor output concentration of propylene glycol/$kmol \cdot m^{-3}$	Output	Depends on input
T_o	Reactor output temperature/K	Input	Manipulated
F_L	Bottom flow rate/$m^3 \cdot h^{-1}$	Output	Depend on input
C_{A_L}	Bottom concentration of propylene oxide/$kmol \cdot m^{-3}$	Output	Depends on input
C_{B_L}	Bottom concentration of water/$kmol \cdot m^{-3}$	Output	Depends on input
C_{C_L}	Bottom concentration of propylene glycol/$kmol \cdot m^{-3}$	Output	Controlled
T_L	Bottom output temperature/K	Input	Fixed
F_{D_α}	Recycle flow rate/$m^3 \cdot h^{-1}$	Output	Depends on input
C_{A_D}	Recycle concentration of propylene oxide/$kmol \cdot m^{-3}$	Output	Depends on input
C_{B_D}	Recycle concentration of water/$kmol \cdot m^{-3}$	Output	Depends on input
T_D	Recycle temperature/K	Output	Depends on input
P_{CSTR}	CSTR pressure/kPa	Input	Fixed
F_d	Condenser pressure/kPa	Input	Fixed
P_R	Reboiler pressure/kPa	Input	Fixed
T_{top}	Column top temperature/K	Input	Fixed
T_d	Condenser temperature/K	Input	Fixed
T_R	Reboiler temperature/K	Input	Fixed

Here A_o is the preexponential constant, E is the activation constant, R is the ideal gas constant, T_o is the reactor temperature, and V is the reactor hold-up.

At steady-states, *i.e.*, $dC_{A_o}/dt = 0$, Eqs. (3), (4) and (5) become

$$F_i C_{A_i} + F_{D_\alpha} C_{A_D} - A_o e^{-\frac{E}{RT_o}} C_{A_o} V - F_o C_{A_o} = 0 \tag{6}$$

$$F_i C_{B_i} + F_{D_\alpha} C_{B_D} - A_o e^{-\frac{E}{RT_o}} C_{A_o} V - F_o C_{B_o} = 0 \tag{7}$$

$$F_{D_\alpha} C_{C_D} + A_o e^{-\frac{E}{RT_o}} C_{A_o} V - F_o C_{C_o} = 0 \tag{8}$$

The steady-state component balances of A, B, and C yield

$$C_{A_o} = \frac{F_i C_{A_i} + F_{D_\alpha} C_{A_D}}{A_o e^{-\frac{E}{RT_o}} V + F_i + F_{D_\alpha}} \tag{9}$$

$$C_{B_o} = \frac{F_i C_{B_i} + F_{D_\alpha} C_{B_D} - A_o e^{-\frac{E}{RT_o}} C_{A_o} V}{F_i + F_{D_\alpha}} \tag{10}$$

$$C_{C_o} = \frac{F_{D_\alpha} + A_o e^{-\frac{E}{RT_o}} C_{A_o} V}{F_i + F_{D_\alpha}} \tag{11}$$

Given the outputs of the CSTR, it is ready to obtain the material balances around the distillation column. The total steady-state material balance around the distillation column is given by

$$F_o - F_{D_\alpha} - F_L = 0 \tag{12}$$

Substituting Eq. (2) into Eq. (12) yields $F_i = F_L$ so that component balances of A, B and C around the distillation column are

$$F_i C_{A_i} + F_{D_\alpha} C_{A_D} - A_o e^{-\frac{E}{RT_o}} C_{A_o} V - F_{D_\alpha} C_{A_D} - F_L C_{A_L} = 0 \tag{13}$$

$$F_i C_{B_i} + F_{D_\alpha} C_{B_D} - A_o e^{-\frac{E}{RT_o}} C_{A_o} V - F_{D_\alpha} C_{A_D} - F_L C_{B_L} = 0 \tag{14}$$

$$F_i C_{C_i} + F_{D_\alpha} C_{C_D} + A_o e^{-\frac{E}{RT_o}} C_{A_o} V - F_{D_\alpha} C_{C_D} - F_L C_{C_L} = 0 \tag{15}$$

By putting $F_i = F_L$ in Eq. (13) and rearranging, the bottom concentrations of propylene oxide C_{A_L}, water C_{B_L} and propylene glycol C_{C_L} are

$$C_{A_L} = C_{A_i} - \frac{A_o e^{-\frac{E}{RT_o}} C_{A_o} V}{F_i} \tag{16}$$

$$C_{B_L} = C_{B_i} - \frac{A_o e^{-\frac{E}{RT_o}} C_{A_o} V}{F_i} \tag{17}$$

$$C_{C_L} = \frac{A_o e^{-\frac{E}{RT_o}} C_{A_o} V}{F_i} \tag{18}$$

The recycle flow rate and component balances are obtained as follows. From Eq. (12), $F_o = F_{D_\alpha} + F_L$ so that the component balances for A and B are

$$F_o C_{A_o} = F_L C_{A_L} + F_{D_\alpha} C_{A_D} \tag{19}$$

$$F_o C_{B_o} = F_L C_{B_L} + F_{D_\alpha} C_{B_D} \tag{20}$$

The component balance for C is not considered here because the distillate does not contain C (*i.e.*, propylene glycol). Since $F_o = F_{D_\alpha} + F_L$ in Eq. (12) and $F_i = F_L$.

$$(F_{\mathrm{i}}+F_{\mathrm{D}_\alpha})C_{\mathrm{A_o}}=F_{\mathrm{i}}C_{\mathrm{A_L}}+F_{\mathrm{D}_\alpha}C_{\mathrm{A_D}} \tag{21}$$

Some simplification gives

$$F_{\mathrm{i}}(C_{\mathrm{A_o}}-C_{\mathrm{A_L}})=F_{\mathrm{D}_\alpha}(C_{\mathrm{A_D}}-C_{\mathrm{A_o}}) \tag{22}$$

Thus, the recycle flow rate is given by

$$F_{\mathrm{D}_\alpha}=\frac{F_{\mathrm{i}}(C_{\mathrm{A_o}}-C_{\mathrm{A_L}})}{(C_{\mathrm{A_D}}-C_{\mathrm{A_o}})} \tag{23}$$

In Eq. (2) $F_{\mathrm{o}}=F_{\mathrm{i}}+F_{\mathrm{D}_\alpha}$. Hence, Eqs. (19) and (20) become

$$C_{\mathrm{A_o}}\left(F_{\mathrm{i}}-F_{\mathrm{D}_\alpha}C_{\mathrm{A_D}}\right)=F_{\mathrm{L}}C_{\mathrm{A_L}}-F_{\mathrm{D}_\alpha}C_{\mathrm{A_D}} \tag{24}$$

$$C_{\mathrm{B_o}}\left(F_{\mathrm{i}}-F_{\mathrm{D}_\alpha}C_{\mathrm{B_D}}\right)=F_{\mathrm{L}}C_{\mathrm{B_L}}-F_{\mathrm{D}_\alpha}C_{\mathrm{B_D}} \tag{25}$$

$$F_{\mathrm{i}}C_{\mathrm{A_o}}-F_{\mathrm{L}}C_{\mathrm{A_L}}+F_{\mathrm{D}_\alpha}C_{\mathrm{A_o}}=F_{\mathrm{D}_\alpha}C_{\mathrm{A_D}} \tag{26}$$

$$F_{\mathrm{i}}C_{\mathrm{B_o}}-F_{\mathrm{L}}C_{\mathrm{B_L}}+F_{\mathrm{D}_\alpha}C_{\mathrm{B_o}}=F_{\mathrm{D}_\alpha}C_{\mathrm{B_D}} \tag{27}$$

Since $F_{\mathrm{i}}=F_{\mathrm{L}}$, it is ready to obtain

$$C_{\mathrm{A_D}}=\frac{F_{\mathrm{i}}(C_{\mathrm{A_o}}-C_{\mathrm{A_L}})}{F_{\mathrm{D}_\alpha}}+C_{\mathrm{A_o}} \tag{28}$$

$$C_{\mathrm{B_D}}=\frac{F_{\mathrm{i}}(C_{\mathrm{B_o}}-C_{\mathrm{B_L}})}{F_{\mathrm{D}_\alpha}}+C_{\mathrm{B_o}} \tag{29}$$

The bottom concentration of propylene glycol $C_{\mathrm{C_L}}$ is the main critical value that measures the status of the process which is affected by the critical process variable $C_{\mathrm{A_i}}$, F_{D_α}, $C_{\mathrm{A_D}}$, F_{i} and T_{o}. Thus, the model of the CSTR-DC-recycle chemical process is given by Eqs. (9), (16), (18), (23) and (28). Eqs. (9)–(28) involve the unknown parameters E and A_{o} to be estimated since they may be influenced by different process conditions. Since Eqs. (9), (16) and (18) share the same functional dependency on E and A_{o}, (18) is selected to estimate the parameters. At an instant t the bottom product $C_{\mathrm{C_L}}$ is given by the function

$$\widehat{C}_{\mathrm{C_L}}(t;\beta_{\mathrm{p}})=f(C_{\mathrm{A_i}},F_{\mathrm{i}},\beta_{\mathrm{p}}) \tag{30}$$

where $\beta_{\mathrm{p}}=[E,A_{\mathrm{o}}]$ are the unknown parameters to be estimated and the function f is defined by Eq. (18). β_{p} is estimated by the Genetic algorithm [27] by minimizing the difference between the historical data $C_{\mathrm{C_{L(}}t\mathrm{)}}$ and the predicted data $\widehat{C}_{\mathrm{C_L}}(t;\beta_{\mathrm{p}})$

$$\widehat{\beta}_{\mathrm{p}}=\underset{\beta_{\mathrm{p}}}{\mathrm{argmax}}(F(\beta_{\mathrm{p}})) \tag{31}$$

where

$$F(\beta_{\mathrm{p}})=1-\frac{\sum_{t=1}^{T}\left(C_{\mathrm{C_L}}(t)-\widehat{C}_{\mathrm{C_L}}(t;\beta_{\mathrm{p}})\right)^2}{\sum_{t=1}^{T}\left(C_{\mathrm{C_L}}(t)-\overline{C}_{\mathrm{C_L}}(t;\beta_{\mathrm{p}})\right)^2} \tag{32}$$

Here $\widehat{C}_{\mathrm{C_L}}(t;\beta_{\mathrm{p}})$ is the predicted value of $C_{\mathrm{C_L}}$ at an instant t by Eq. (30) using β_{p} and $\overline{C}_{\mathrm{C_L}}\left(t;\beta_p\right)=\sum_{t=1}^{T}(C_{\mathrm{C_L}}(t))/T$ is the mean of $\{C_{\mathrm{C_L}}(t)\}_{t=1}^{T}$. The values of E and A_{o} that give the maximum value of $F(\beta_{\mathrm{p}})$ in Eq. (32) are selected as the estimated parameter values. The historical data samples $C_{\mathrm{C_L}}$ used in Eq. (30) and Eq. (31) represent steady-state values so that the estimates $\widehat{C}_{\mathrm{C_L}}$ also represent steady-state values.

3.2. Measuring the uncertainty of the grey-box model

The parameters $\beta_{\mathrm{p}}=[E,A_{\mathrm{o}}]$ of the grey-box model are estimated from measured data samples of the process variables. These measured values are usually affected by noise and disturbances which increase the uncertainty level of the grey-box model and consequently the NOZ to be established. Hence, the uncertainty of the grey-box model is hereby assessed based on the uncertainty of the estimated model parameters to ensure its reliability in providing accurate NOZ model of the CSTR-DC-recycle process.

Let $\{C_{\mathrm{C_L}}(t)\}_{t=1}^{T}$ be the measured data samples of $C_{\mathrm{C_L}}$ where t is the sampling index. According to Bayesian estimation, β_{p} can be treated as a random variable $\Theta_{\beta_{\mathrm{p}}}$ with $\{\beta_{p,i}\}_{\mathrm{i}}^{\mathrm{I}}=\{E_{\mathrm{i}},A_{\mathrm{o_i}}\}_{\mathrm{i}}^{\mathrm{I}}$ as the data samples of the possible values of E and A_o [27,28]. Here, i is the index for the i-th value of E and A_{o}. The data samples in $\{\beta_{\mathrm{p,i}}\}_{\mathrm{i}}^{\mathrm{I}}$ are generated by estimating β_{p} for different data sets of $C_{\mathrm{C_L}}$ using Eq. (31). Since the predicted value $\widehat{C}_{\mathrm{C_L}}(t)$ at an instant t by the model with $\{\beta_{\mathrm{p,i}}\}_{\mathrm{i}}^{\mathrm{I}}=\{E_{\mathrm{i}},A_{\mathrm{o_i}}\}_{\mathrm{i}}^{\mathrm{I}}$ can either be in the measured data samples $\{C_{C_L}(\mathrm{T})\}_{t=1}^{T}$ or not, the process can be treated as the Bernoulli process.

Let $z_{\beta_{\mathrm{p,i}}}$ be the least number of data samples in $\{C_{\mathrm{C_L}}(t)\}_{t=1}^{T}$ that should be explained by $\beta_{\mathrm{p,i}}=[E_{\mathrm{i}},A_{\mathrm{o_i}}]$. The conditional probability of $z_{\beta_{\mathrm{p,i}}}$ given $\Theta_{\beta_{\mathrm{p,i}}}$ for the respective realizations $z_{\beta_{\mathrm{p,i}}}$ and $\theta_{\beta_{\mathrm{p,i}}}$ is computed by

$$P_{Z_{\beta_{\mathrm{p,i}}}\left|\Theta_{\beta_{\mathrm{p,i}}}\right.}\left(z_{\beta_{\mathrm{p,i}}}\middle|\theta_{\beta_{\mathrm{p,i}}}\right)=\frac{z_{\beta_{\mathrm{p,i}}}!}{z_{\beta_{\mathrm{p,i}}}!\left(T-Z_{\beta_{\mathrm{p,i}}}\right)!}\theta_{Z_{\beta_{\mathrm{p,i}}}}^{Z_{\beta_{\mathrm{p,i}}}}\times\left(1-\theta_{\beta_{\mathrm{p,i}}}\right)^{T-Z_{\beta_{\mathrm{p,i}}}} \tag{33}$$

Hence, for $z_{\beta_{\mathrm{p,i}}}$ the posterior distribution of $\Theta_{\beta_{\mathrm{p,i}}}$ is

$$P_{\Theta_{\beta_{\mathrm{p,i}}}\left|Z_{\beta_{\mathrm{p,i}}}\right.}\left(\theta_{\beta_{\mathrm{p,i}}}\middle|z_{\beta_{\mathrm{p,i}}}\right)=\frac{\frac{z_{\beta_{\mathrm{p,i}}}!}{z_{\beta_{\mathrm{p,i}}}!\left(T-Z_{\beta_{\mathrm{p,i}}}\right)!}\theta_{Z_{\beta_{\mathrm{p,i}}}}^{Z_{\beta_{\mathrm{p,i}}}}\left(1-\theta_{\beta_{\mathrm{p,i}}}\right)^{T-Z_{\beta_{\mathrm{p,i}}}}p_{\Theta_{\beta_{\mathrm{p,i}}}}\left(\theta_{\beta_{\mathrm{p,i}}}\right)}{\int_{\beta_{\mathrm{p}}}^{0}\frac{z_{\beta_{\mathrm{p,i}}}!}{z_{\beta_{\mathrm{p,i}}}!\left(T-Z_{\beta_{\mathrm{p,i}}}\right)!}\theta_{Z_{\beta_{\mathrm{p,i}}}}^{Z_{\beta_{\mathrm{p,i}}}}\left(1-\theta_{\beta_{\mathrm{p,i}}}\right)^{T-Z_{\beta_{\mathrm{p,i}}}}p_{\Theta_{\beta_{\mathrm{p,i}}}}\left(\theta_{\beta_{\mathrm{p,i}}}\right)\mathrm{d}\theta_{\beta_{\mathrm{p,i}}}} \tag{34}$$

where $p_{Z_{\beta_{\mathrm{p,i}}},\Theta_{\beta_{\mathrm{p,i}}}}(z_{\beta_{\mathrm{p,i}}},\theta_{\beta_{\mathrm{p,i}}})=p_{Z_{\beta_{\mathrm{p,i}}}\left|\Theta_{\beta_{\mathrm{p,i}}}\right.}\left(z_{\beta_{\mathrm{p,i}}}\middle|\theta_{\beta_{\mathrm{p,i}}}\right)\cdot p_{\Theta_{\beta_{\mathrm{p,i}}}}(\theta_{\beta_{\mathrm{p,i}}})$ is the joint distribution of $z_{\beta_{\mathrm{p,i}}}$ and $\Theta_{\beta_{\mathrm{p,i}}}$. Here $p_{\Theta_{\beta_{\mathrm{p,i}}}}(\theta_{\beta_{\mathrm{p,i}}})$ is the prior distribution assumed to be uniform and continuous as

$$p_{\Theta_{\beta_{\mathrm{p,i}}}}\left(\theta_{\beta_{\mathrm{p,i}}}\right)=\begin{cases}1, & 0<\theta_{\beta_{\mathrm{p,i}}}<1\\ 0, & \text{otherwise}\end{cases} \tag{35}$$

Thus, the Bayesian estimate of $\theta_{\beta_{\mathrm{p,i}}}$ is

$$\widehat{\theta}_{\beta_{\mathrm{p}}}=\int_{\beta_{\mathrm{p}}}\theta_{\beta_{\mathrm{p,i}}}\frac{\frac{z_{\beta_{\mathrm{p,i}}}!}{z_{\beta_{\mathrm{p,i}}}!\left(T-Z_{\beta_{\mathrm{p,i}}}\right)!}\theta_{Z_{\beta_{\mathrm{p,i}}}}^{Z_{\beta_{\mathrm{p,i}}}}\left(1-\theta_{\beta_{\mathrm{p,i}}}\right)^{T-Z_{\beta_{\mathrm{p,i}}}}p_{\Theta_{\beta_{\mathrm{p,i}}}}\left(\theta_{\beta_{\mathrm{p,i}}}\right)}{\int_{\beta_{\mathrm{p}}}\frac{Z_{\beta_{\mathrm{p,i}}}!}{z_{\beta_{\mathrm{p,i}}}!\left(T-Z_{\beta_{\mathrm{p,i}}}\right)!}\theta_{Z_{\beta_{\mathrm{p,i}}}}^{Z_{\beta_{\mathrm{p,i}}}}\left(1-\theta_{\beta_{\mathrm{p,i}}}\right)^{T-Z_{\beta_{\mathrm{p,i}}}}p_{\Theta_{\beta_{\mathrm{p,i}}}}\left(\theta_{\beta_{\mathrm{p,i}}}\right)\mathrm{d}\theta_{\beta_{\mathrm{p,i}}}}\mathrm{d}\theta_{\beta_{\mathrm{p,i}}} \tag{36}$$

The confidence interval $[\widehat{\theta}_{\beta_{\mathrm{p}-}}, \widehat{\theta}_{\beta_{\mathrm{p}+}}]$ bounding $\widehat{\theta}_{\beta_{\mathrm{p}}}$ for a specific significance level α is the region enclosed by $[\widehat{E}_-, \widehat{E}_+]$ [29] and $[\widehat{A}_{\mathrm{o}-}, \widehat{A}_{\mathrm{o}+}]$ computed as the region that satisfies

$$\int_{\widehat{E}_-}^{\widehat{E}_+}\int_{\widehat{A}_{o-}}^{\widehat{A}_{o+}} \frac{\dfrac{z_{\beta_{\mathrm{p,i}}}!}{z_{\beta_{\mathrm{p,i}}}!\left(T-Z_{\beta_{\mathrm{p,i}}}\right)!}\theta_{z_{\beta_{\mathrm{p,i}}}}^{z_{\beta_{\mathrm{p,i}}}}\left(1-\theta_{\beta_{\mathrm{p,i}}}\right)^{T-z_{\beta_{\mathrm{p,i}}}} p_{\Theta_{\beta_{\mathrm{p,i}}}}\left(\theta_{\beta_{\mathrm{p,i}}}\right)}{\displaystyle\int_{\beta_{\mathrm{p}}}\frac{z_{\beta_{\mathrm{p,i}}}!}{z_{\beta_{\mathrm{p,i}}}!\left(T-Z_{\beta_{\mathrm{p,i}}}\right)!}\theta_{z_{\beta_{\mathrm{p,i}}}}^{z_{\beta_{\mathrm{p,i}}}}\left(1-\theta_{\beta_{\mathrm{p,i}}}\right)^{T-z_{\beta_{\mathrm{p,i}}}} p_{\Theta_{\beta_{\mathrm{p,i}}}}\left(\theta_{\beta_{\mathrm{p,i}}}\right)\mathrm{d}\theta_{\beta_{\mathrm{p,i}}}} \times \mathrm{d}\theta_{\beta_{\mathrm{p,i}}} = 1-\alpha \tag{37}$$

Let $[\widehat{E}_-, \widehat{E}_+]$ and $[\widehat{A}_{o-}, \widehat{A}_{o+}]$ be the $(100-\alpha)\%$ confidence intervals of $\widehat{E}$ and $\widehat{A}_{\mathrm{o}}$, respectively. The parameters and are defined as the maximum confidence interval widths, *i.e.*,

$$\varepsilon_E = \max(|\widehat{E}-\widehat{E}_-, \widehat{E}_+-\widehat{E}|) \tag{38}$$

$$\varepsilon_{A_o} = \max(|\widehat{A}_{\mathrm{o}}-\widehat{A}_{\mathrm{o}-}, \widehat{A}_{\mathrm{o}+}-\widehat{A}_{\mathrm{o}}|) \tag{39}$$

Given the estimates $\widehat{E}_{\mathrm{i}}$, $\widehat{A}_{\mathrm{o}}$, $\varepsilon_{\hat{E}}$ and $\varepsilon_{\hat{A}_o}$, an index $\eta_{\hat{\theta}_{\beta_{\mathrm{p}}}}$ is defined as

$$\eta_{\hat{\theta}_{\beta_{\mathrm{p}}}} = \frac{\widehat{E}}{\varepsilon_{\hat{E}}} + \frac{\widehat{A}_o}{\varepsilon_{\hat{A}_o}} \tag{40}$$

The index ratios $\widehat{E}/\varepsilon_{\hat{E}}$ and $\widehat{A}_o/\varepsilon_{\hat{A}_o}$ are equivalent to ICV which is the ratio of the estimated mean of a data set to its standard deviation, *i.e.*,

$$\mathrm{ICV} = \frac{\overline{m}}{s} \tag{41}$$

The ICV calculates the sample mean's deviation from the population mean. If the standard deviation is larger, the ICV becomes smaller, indicating that the sample mean largely deviates from the population mean [29]. Here the standard deviations are equivalent to $\varepsilon_{\hat{E}}$ and $\varepsilon_{\hat{A}_o}$. Hence, the accuracy of the estimates $\widehat{E}$ and $\widehat{A}_{\mathrm{o}}$ increases with a decrease in $\varepsilon_{\hat{E}}$ and $\varepsilon_{\hat{A}_o}$ [29]. To determine the overall model accuracy, the combined accuracies $\eta_{\hat{\theta}_{\beta_{\mathrm{p}}}}$ in Eq. (40) and the

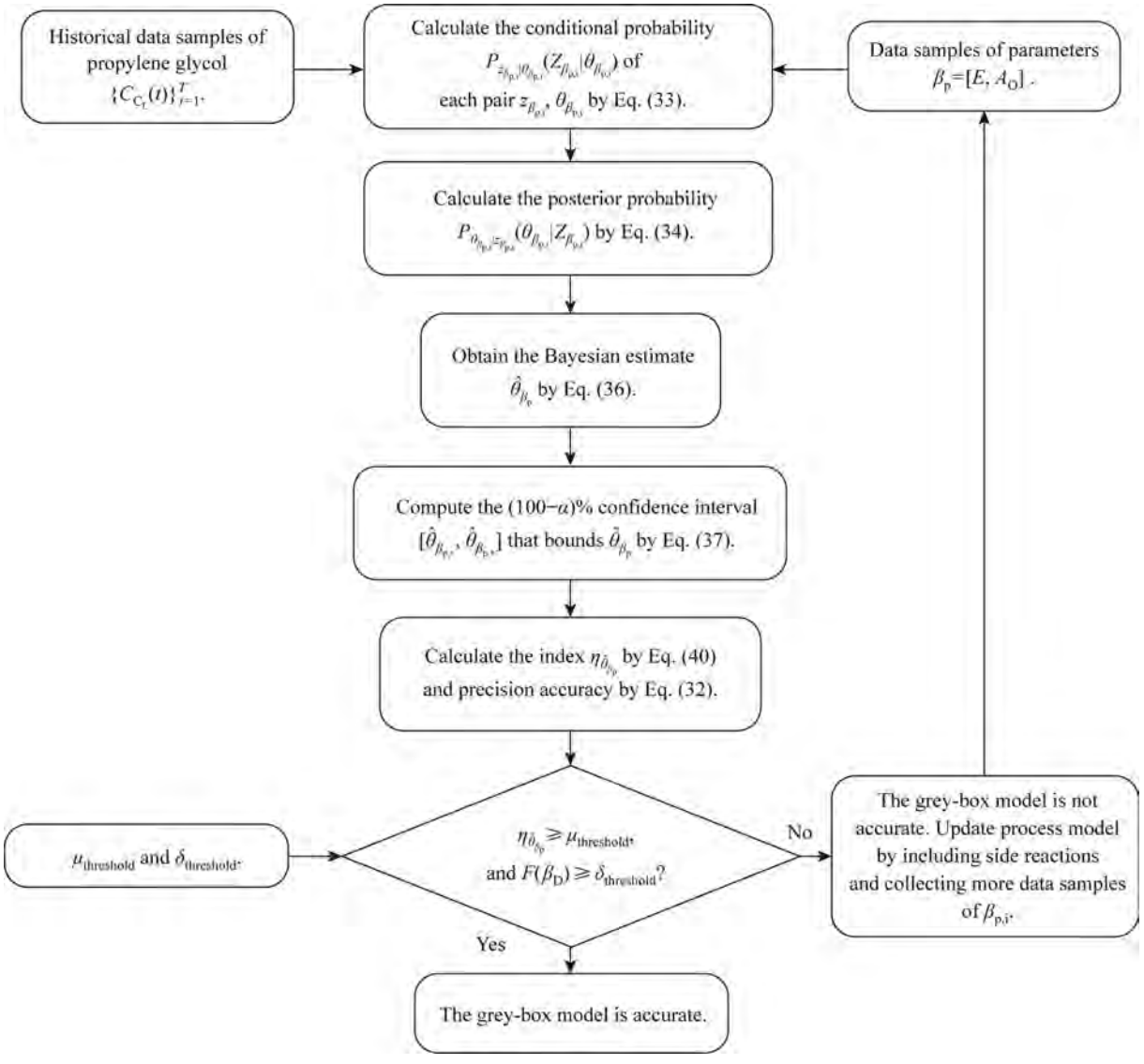

Fig. 3. Flowchart of the uncertainty measurement of the grey-box model.

prediction accuracy in Eq. (32) are used. The model is accurate if the following conditions are satisfied

$$\eta_{\hat{\theta}_{\beta_p}} \geq \mu_{\text{threshold}} \tag{42}$$

$$F(\beta_p) \geq \delta_{\text{threshold}} \tag{43}$$

Here $\mu_{\text{threshold}}$ and $\delta_{\text{threshold}}$ are user defined thresholds. $\mu_{\text{threshold}}$ can be decided based on the significance level α used in determining the confidence interval of $\widehat{E}$ and $\widehat{A}_o$. Given α, the z-score $z_{1-\alpha}$ at $1-\alpha$ is calculated. Each of the estimates $\widehat{E}$ and $\widehat{A}_o$ should be $z_{1-\alpha}$ larger than their uncertainties at α significance level. By combining the respective z-scores for $\widehat{E}$ and $\widehat{A}_o$, $\mu_{\text{threshold}}$ is given by

$$\mu_{\text{threshold}} = 2 \times z_{1-\alpha} \tag{44}$$

For an example, if $\alpha = 0.05$, the z-score at 0.95 is calculated as 1.96. Thus, $\mu_{\text{threshold}} = 3.92$. Similarly, $\delta_{\text{threshold}}$ can be decided by selecting a preferred limit value such as 0.90 on $F(\beta_p)$ in Eq. (32). The method is summarized in Algorithm 1 and Fig. 3.

Algorithm 1 Measuring the uncertainty of the grey-box model.

procedure
INPUT : $z_{\beta_{p,i}}, \{C_{C_L}(t)\}_{t=1}^{T}, \beta_{p,i}, \Theta_{\beta_{p,i}}, z_{\beta_{p,i}}, \theta_{\beta_{p,i}}, \mu_{\text{threshold}}, \delta_{\text{threshold}}$
procedure OUTPUT : $\eta_{\hat{\theta}_{\beta_p}}, F(\beta_p)$
for $\{C_{C_L}(t)\}_{t=1}^{T}, \beta_{p,i}, z_{\beta_{p,i}}, \theta_{\beta_{p,i}}$ **do**
Calculate the conditional probability $P_{Z_{\beta_{p,i}}|\Theta_{\beta_{p,i}}}(z_{\beta_{p,i}}|\theta_{\beta_{p,i}})$ of each pair $z_{\beta_{p,i}}, \theta_{\beta_{p,i}}$ by Eq. (33).
Calculate the posterior probability $P_{\Theta_{\beta_{p,i}}|Z_{\beta_{p,i}}}(\theta_{\beta_{p,i}}|z_{\beta_{p,i}})$ by Eq. (34).
Obtain the Bayesian estimate $\hat{\theta}_{\beta_p}$ by Eq. (36).
Compute the $(100-\alpha)\%$ confidence interval $[\hat{\theta}_{\beta_{p-}}, \hat{\theta}_{\beta_{p+}}]$ that bounds $\hat{\theta}_{\beta_p}$ by Eq. (37).
Calculate the maximum confidence intervals $\varepsilon_{\hat{E}}$ and $\varepsilon_{\hat{A}_o}$ by Eq. (38) and Eq. (39), respectively.
Calculate the index $\eta_{\hat{\theta}_{\beta_p}}$ by Eq. (40) and precision accuracy by Eq. (32).
if $\eta_{\hat{\theta}_{\beta_p}} \geq \mu_{\text{threshold}}$ and $F(\beta_p) \geq \delta_{\text{threshold}}$ **then**
The grey-box model is accurate.
end if
end for
end procedure
end procedure

It is very necessary to validate the identified grey-box model to assess how it fits well with observed data. A method is adapted from Ref. [30] to validate the grey-box model based on surrogate estimates of C_{C_L} Under weak assumptions, $\hat{\beta}_p$ converges to the Gaussian distribution,

$$\hat{\beta}_p \sim \mathcal{N}(\beta_p, \mathbf{Cov}(\hat{\beta}_p)) \tag{45}$$

Here $\mathbf{Cov}(\hat{\beta}_p)$ is the covariance matrix of $\hat{\beta}_p$. Data samples are obtained for surrogate estimates $\{C_{C_L}(n)\}_{n=1}^{N}$ by simulating the grey-box model by sampling $\{\beta_p(n)\}_{n=1}^{N}$ data samples from the distribution in Eq. (45) based on the input data of F_i, C_{A_i} and T_o. The $(100-\alpha)\%$ (α takes small values such as 5%) confidence interval $[\widehat{C}_{C_L,-}, \widehat{C}_{C_L,+}]$ enclosing the mean $\widehat{C}_{C_L}$ of the surrogate estimates $\{C_{C_L}(n)\}_{n=1}^{N}$ is computed by

$$\left[\widehat{C}_{C_L} - \frac{t_{\alpha/2,N-1}s}{\sqrt{N}}, \widehat{C}_{C_L} + \frac{t_{\alpha/2,N-1}s}{\sqrt{N}}\right] \tag{46}$$

where s is the standard deviation of $\{C_{C_L}(n)\}_{n=1}^{N}$ and $t_{\alpha/2,N-1}$ is the critical value from the t-distribution. Let $\widehat{C}_{C_L,o}$ be the observed value based on the input data of F_i, C_{A_i} and T_o. If $\widehat{C}_{C_L,o}$ is within the interval $[\widehat{C}_{C_L,-}, \widehat{C}_{C_L,+}]$, then the grey-box model is valid and fits well with the observed data with $(100-\alpha)\%$ confidence. If the grey-box model does not fit well with observed data, the process model is no longer valid, possibly due to the ignorance of the side reactions in Eq. (1b) and Eq. (1c).

The proposed method for the grey-box model is a multi-parameter estimation based on Bayesian estimation method and inverse coefficient of variation which are well established methods for parameter estimation and uncertainty measurements [29]. Thus, given other chemical processes, the proposed method can be used to develop the grey-box model. For instance, a grey-box model was built for condensers in Ref. [31], based on the energy balance between hot steam and cooling water; a NOZ model was developed from the grey-box model by considering allowable operating ranges of the condenser pressure and temperatures.

3.3. Establishing the normal operating zone model

After obtaining the grey-box model of the CSTR-DC-recycle process, it is ready to establish the NOZ model which is presented in this section. The grey-box model in Eqs. (9), (16), (18), (23) and (28) is used to simulate the values of C_{C_L} based on historic data samples of the input variables F_i, C_{A_i} and T_o. Given the process operational requirement $C_{C_L,\min} \leq C_{C_L} \leq C_{C_L,\max}$, the simulated data samples of F_i, C_{A_i}, T_o and C_{C_L} that agree to the operational requirement are selected to form the NOZ.

To establish the NOZ model, the boundary points of the NOZ are extracted by adopting a boundary point factor (BPF) method from Refs. [32,33]. Let $\{X_i(t)\}_{t=1}^{T} = \{C_{A_o}(t), C_{A_D}(t), C_{A_L}(t), C_{C_L}(t)\}_{t=1}^{T}$ be the data samples in the NOZ. The k-nearest neighborhood of all the data samples in $\{X_i(t)\}_{t=1}^{T}$ is calculated. By assuming that there are no outliers in $\{X_i(t)\}_{t=1}^{T}$, a boundary index I_p is formulated to determine the boundary points in $\{X_i(t)\}_{t=1}^{T}$ as follows. Let X_p be a data sample in $\{X_i(t)\}_{t=1}^{T}$ having a neighbor $X_n \in Z_k(X_p)$ where $Z_k(X_p)$ is the set of k-nearest neighbors of X_p and the norm $\left\|\overrightarrow{X_pX_n}\right\|$. I_p is calculated by

$$I_p = \frac{1}{|Z_k(X_p)|}\left\|\sum_{X_n \in Z_k(X_p)} \frac{\overrightarrow{X_pX_n}}{\left\|\overrightarrow{X_pX_n}\right\|}\right\| \tag{47}$$

Here $|Z_k(X_p)|$ is the number of k-nearest neighbors of X_p in $Z_k(X_p)$, $\overrightarrow{X_pX_n}$ is the vector from X_p to X_n and $\left\|\overrightarrow{X_pX_n}\right\|$, is the Euclidean distance between X_p and X_n. In general sense, I_p indicates how the k-neighborhood sample set $Z_k(X_p)$ is distributed around X_p. To explain how the index I_p is used to determine the boundary points in $\{X_i(t)\}_{t=1}^{T}$, Fig. 4 shows cluster of samples with the sample X_p and another point X_q. Since number of k-nearest neighbors of X_p in $|Z_k(X_p)| = 3$ is smaller than that of X_p in $|Z_k(X_p)| = 8$, the index I_p will be larger than I_q, *i.e.*,

$$I_p > I_q \tag{48}$$

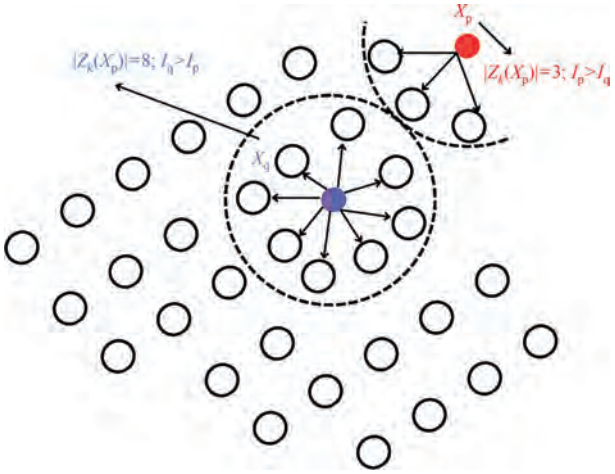

Fig. 4. Illustrating data distribution around a boundary point X_p and core point X_q.

The inequality in Eq. (48) suggests that core data samples in $\{X_{\mathrm{i}}(t)\}_{t=1}^{T}$ (such as X_{q}) are surrounded by more data samples than those on the boundary. Therefore, the boundary indices calculated by Eq. (47) for those on the boundary are greater than the index for the core samples since the index is computed by having the number of k-nearest neighbors as the denominator. Hence, the boundary points in $\{X_{\mathrm{i}}(t)\}_{t=1}^{T}$ are determined by calculating the index for each data sample in $\{X_{\mathrm{i}}(t)\}_{t=1}^{T}$ by Eq. (47). Given the data samples of the indices $\{I_{\mathrm{p}}(t)\}_{t=1}^{T}$, the boundary points are selected by arranging $\{I_{\mathrm{p}}(t)\}_{t=1}^{T}$ in ascending order and selecting those operating points whose indices are larger than a user selected threshold $I_{\mathrm{p}} = \mathrm{m}$ to represent the NOZ model. Let $\{B_{\mathrm{p}}^{*}(k)\}_{k=1}^{K} = [b_{1,\mathrm{p}}^{*}(k), \cdots, b_{n,\mathrm{p}}^{*}(k), \cdots, b_{N,\mathrm{p}}^{*}(\mathrm{k})]_{k=1}^{K}$ be the selected operating points whose I_{p}'s are larger than where N is the number of process variables and K is the sample size. The NOZ model is given by

$$\mathrm{NOZ}_{\mathrm{model}} := \left\{B_{\mathrm{p}}^{*}(k)\right\}_{k=1}^{K} \tag{49}$$

The boundary points of the NOZ represent the worse safety operating points of the process such that the operating points inside are in the normal condition while those outsides are in the abnormal condition. Hence, the NOZ can be used for monitoring the CSTR-DC-recycle process. Since the NOZ model is obtained from uncertain measured process data, it is very important to determine the accuracy of the NOZ model to ensure effective process monitoring. The following section determines the accuracy of the NOZ model.

3.4. Determining the accuracy of the NOZ model

The NOZ model is established based on uncertain process data samples. The uncertainty of the NOZ model will be large if the size of data samples is small thus reducing its accuracy for monitoring. Hence, the accuracy of the NOZ model should be determined to ensure its effectiveness for monitoring the process. Let $\{\{B_{\mathrm{p},y}^{*}(k)\}_{k=1}^{K}\}_{y=1}^{Y}$ be the data set of established NOZ models where Y is the total number of established NOZ models. To ensure uniformity in the different measurement units of the process variables, $B_{\mathrm{p},y}^{*}$ is normalized into $\left\{B_{\mathrm{p},y}^{*}(k)\right\}_{k=1}^{K} = \left[b_{1,\mathrm{p},y}^{*}(k), \cdots, b_{n,\mathrm{p},y}^{*}(k), \cdots, b_{N,\mathrm{p},y}^{*}(k)\right]_{k=1}^{K}$ by

$$b_{n,\mathrm{p},y}(k) = \frac{b_{n,\mathrm{p},y}^{*}(k) - \min\left(b_{n,\mathrm{p},y}^{*}(k)\right)}{\max\left(b_{n,\mathrm{p},y}^{*}(k)\right) - \min\left(b_{n,\mathrm{p},y}^{*}(k)\right)} \tag{50}$$

Let $B_{\mathrm{p},y,\mathrm{i}} = [b_{1,\mathrm{p},y,\mathrm{i}}, \cdots, b_{n,\mathrm{p},y,\mathrm{i}}, \cdots, b_{N,\mathrm{p},y,\mathrm{i}}]$ and $B_{\mathrm{p},y,\mathrm{j}} = [b_{1,\mathrm{p},y,\mathrm{j}}, \cdots, b_{n,\mathrm{p},y,\mathrm{j}}, \cdots, b_{N,\mathrm{p},y,\mathrm{j}}]$ be two randomly selected points in $\{B_{\mathrm{p},y}(k)\}_{k=1}^{K}$. The maximum distance $d_{\mathrm{m},y}$ between the boundary points $B_{\mathrm{p},y}$ of the y-th NOZ is given by

$$d_{\mathrm{m},y} = \max_{1\le i<j\le K} \sqrt{\sum_{n=1}^{4} \left(B_{\mathrm{p},n,y}(i) - B_{\mathrm{p},n,y}(j)\right)^{2}} \tag{51}$$

Let $\{D_{\mathrm{m},y}\}_{y=1}^{Y}$ be the data set of the maximum distances for NOZs in $\{\{B_{\mathrm{p},y}^{*}(k)\}_{k=1}^{K}\}_{y=1}^{Y}$. Each y-th NOZ has a unique $d_{\mathrm{m},y}$ so that the distribution of $\{D_{\mathrm{m},y}\}_{y=1}^{Y}$ can be used to measure the uncertainty of the established NOZ model based on Bayesian estimation.

Let $P_{d_{\mathrm{m},y}}$ be the probability of the most occurred outcome $d_{\mathrm{m},y}$ in $\{D_{\mathrm{m},y}\}_{y=1}^{Y}$ given by

$$P_{d_{\mathrm{m},y}} = \frac{Q}{Y} \tag{52}$$

Here $Q = |D_{\mathrm{m},y} = d_{\mathrm{m},y}|$ is the number of outcomes in $D_{\mathrm{m},y}$ that are equal to $d_{\mathrm{m},y}$. The $P_{d_{\mathrm{m},y}}$ can be treated by the Bayesian estimation rule as a random variable $\Theta_{d_{\mathrm{m},y}}$ taking values in the range [0,1]. Hence, the Bayesian estimate of $P_{d_{\mathrm{m},y}}$ is given by

$$\widehat{\theta}_{\mathrm{d_{m,y}}} = \int_{0}^{1} \theta_{\mathrm{d_{m,y}}} \frac{\frac{q!}{q_{d_{\mathrm{m},y}}!(Y-q)!}\theta_{d_{\mathrm{m},y}}^{q}\left(1-\theta_{d_{\mathrm{m},y}}\right)^{Y-q} p_{\Theta_{\mathrm{d_{m,y}}}}\left(\theta_{d_{\mathrm{m},y}}\right)}{\int_{0}^{1} \frac{q!}{q_{d_{\mathrm{m},y}}!(Y-q)!}\theta_{d_{\mathrm{m},y}}^{q}\left(1-\theta_{d_{\mathrm{m},y}}\right)^{Y-q} p_{\Theta_{\mathrm{d_{m,y}}}}\left(\theta_{d_{\mathrm{m},y}}\right)\mathrm{d}\theta_{d_{\mathrm{m},y}}} \mathrm{d}\theta_{d_{\mathrm{m},y}} \tag{53}$$

Where q and $\Theta_{\mathrm{d_{m,y}}}$ are outcomes in Q and $D_{\mathrm{m},y}$ respectively and $p_{\Theta_{d_{\mathrm{m},y}}}(\Theta_{d_{\mathrm{m},y}})$ is a uniform prior probability of $\Theta_{d_{\mathrm{m},y}}$. The $100-\alpha$ confidence interval of $\widehat{\theta}_{d_{\mathrm{m},y}}$ for a particular significance level α is given by

$$\int_{\hat{\theta}_{d_{\mathrm{m},y}-}}^{\hat{\theta}_{d_{\mathrm{m},y}+}} \frac{\frac{q!}{q_{d_{\mathrm{m},y}}!(Y-q)!}\theta_{d_{\mathrm{m},y}}^{q}\left(1-\theta_{d_{\mathrm{m},y}}\right)^{Y-q} p_{\Theta_{\mathrm{d_{m,y}}}}\left(\Theta_{d_{\mathrm{m},y}}\right)}{\int_{0}^{1} \frac{q!}{q_{d_{\mathrm{m},y}}!(Y-q)!}\theta_{d_{\mathrm{m},y}}^{q}\left(1-\theta_{d_{\mathrm{m},y}}\right)^{Y-q} p_{\Theta_{\mathrm{d_{m,y}}}}\left(\Theta_{d_{\mathrm{m},y}}\right)d\theta_{d_{\mathrm{m},y}}} \times \mathrm{d}\theta_{d_{\mathrm{m},y}} = 1-\alpha \tag{54}$$

Let $[\widehat{\theta}_{d_{\mathrm{m},y}-}, \widehat{\theta}_{d_{\mathrm{m},y}+}]$ be the confidence interval bounding $\widehat{\theta}_{d_{\mathrm{m},y}}$. NOZ_y is closest to the true NOZ if the following condition is satisfied

$$\frac{\widehat{\theta}_{d_{\mathrm{m},y}}}{\max\left(\left[\widehat{\theta}_{d_{\mathrm{m},y}} - \widehat{\theta}_{d_{\mathrm{m},y}-}, \widehat{\theta}_{d_{\mathrm{m},y}+} - \widehat{\theta}_{d_{\mathrm{m},y}}\right]\right)} \ge \beta_{d_{\mathrm{m},y}} \tag{55}$$

Here the ratio $\widehat{\theta}_{d_{\mathrm{m},y}}/\max\left([\widehat{\theta}_{d_{\mathrm{m},y}} - \widehat{\theta}_{d_{\mathrm{m},y}-}, \widehat{\theta}_{d_{\mathrm{m},y}+} - \widehat{\theta}_{d_{\mathrm{m},y}}]\right)$ is synonymous to the definition of ICV in Eq. (41) and $\beta_{d_{\mathrm{m},y}}$ is a user defined value. The value of $\beta_{d_{\mathrm{m},y}}$ is taken as 1 to indicate that the deviation $\max\left([\widehat{\theta}_{d_{\mathrm{m},y}} - \widehat{\theta}_{d_{\mathrm{m},y}-}, \widehat{\theta}_{d_{\mathrm{m},y}+} - \widehat{\theta}_{d_{\mathrm{m},y}}]\right)$ is not larger than the estimate

$\widehat{\theta}_{d_{m,y}}$. The Algorithm 2 and Fig. 5 summarizes the method for determining the accuracy of the NOZ model.

Algorithm 2 Determining the accuracy of the NOZ model.

procedure INPUT : $\{D_m(y)\}_{y=1}^{Y}, Q, q, D_{m,y}, d_{m,y}, \Theta_{d_{m,y}}, \theta_{d_{m,y}}, \beta_{d_{m,y}}$
  procedure OUTPUT : $\widehat{\theta}_{d_{m,y}}, \widehat{\theta}_{d_{m,y}-}, \widehat{\theta}_{d_{m,y}+}$
    for $q, D_{m,y}$ **do**
      Compute Bayesian estimate $\widehat{\theta}_{d_{m,y}}$ by Eq. (53)
      Calculate Bayesian confidence interval $\left[\widehat{\theta}_{d_{m,y}-}, \widehat{\theta}_{d_{m,y}+}\right]$ by Eq. (54)
      Select a user defined value $\beta_{d_{m,y}}$
    if $\widehat{\theta}_{d_{m,y}} / \max\left(\left[\widehat{\theta}_{d_{m,y}} - \widehat{\theta}_{d_{m,y}-}, \widehat{\theta}_{d_{m,y}+} - \widehat{\theta}_{d_{m,y}}\right]\right) \geq \beta_{d_{m,y}}$ **then** NOZ_{model} in Eq. (49) is accurate.
    end if
  end for
  end procedure
end procedure

The main steps of the proposed method are summarized as follows.

(1) Build the grey-box model of the CSTR-DC-recycle chemical process given by Eqs. (9), (16), (18), (23) and (28) based on material balances and historic data samples of process variables.
(2) Measure the uncertainty of the grey-box model by Eq. (42) and Eq. (43). If Eq. (42) and Eq. (43) are not satisfied, repeat step 1 by updating the process model to include the side reactions and collecting more historic data samples of the process variables to estimate the grey-box model that satisfies Eq. (42) and Eq. (43).
(3) Validate the quality of the grey-box model by Eq. (46). If the grey-box model is valid proceed to step 4 else repeat steps 1–2 until a valid grey-box model is established.
(4) Formulate the NOZ by simulating operating points using the grey-box model in Eqs. (9), (16), (18), (23) and based on historic data samples of input process variables; and selecting the operating points that satisfy the process operational requirement. Extract the boundary points based on Eq. (47) to establish the NOZ model in Eq. (49).
(5) Determine the accuracy of the established NOZ model by Eq. (55). According to the law of large number [29], Bayesian estimates will converge to true values as the number of data samples used in estimation goes to infinity. Hence, when more data samples are used to establish the NOZ model, the uncertainty of the NOZ model will decrease in general. Thus, if Eq. (55) is not satisfied, repeat step 3 by collecting more historic data samples.

Though the proposed NOZ model is based on steady-states and the CSTR-DC-recycle chemical process may exhibit dynamic characteristics, it has a good potential to be used for monitoring the dynamic process. The operating zone model for process monitoring is similar to a map used for navigating through mountains. While the map is a static tool and mountain travel is a dynamic activity, the map offers essential information that ensures safe and efficient navigation.

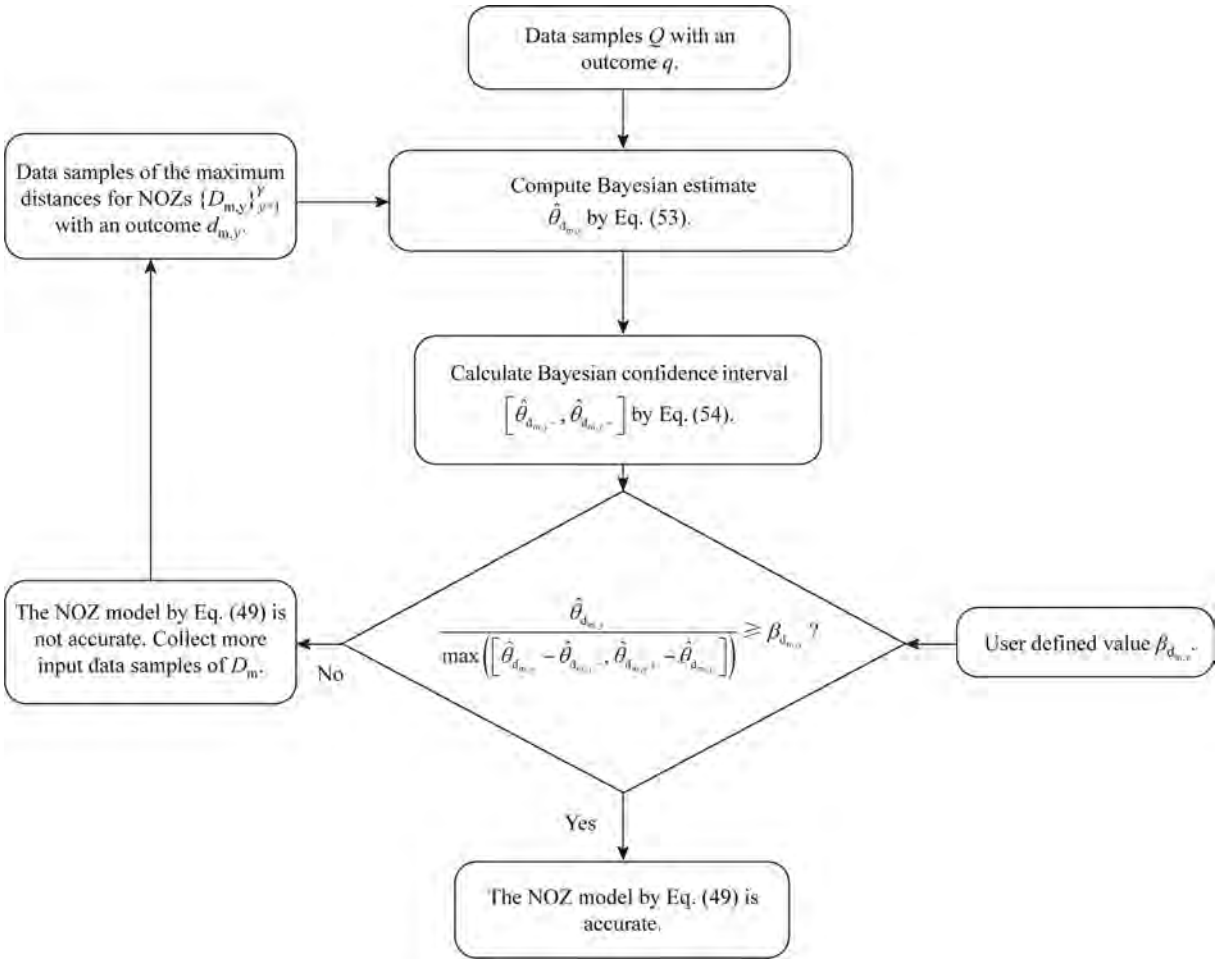

Fig. 5. Flowchart for determining the accuracy measurement of the NOZ model.

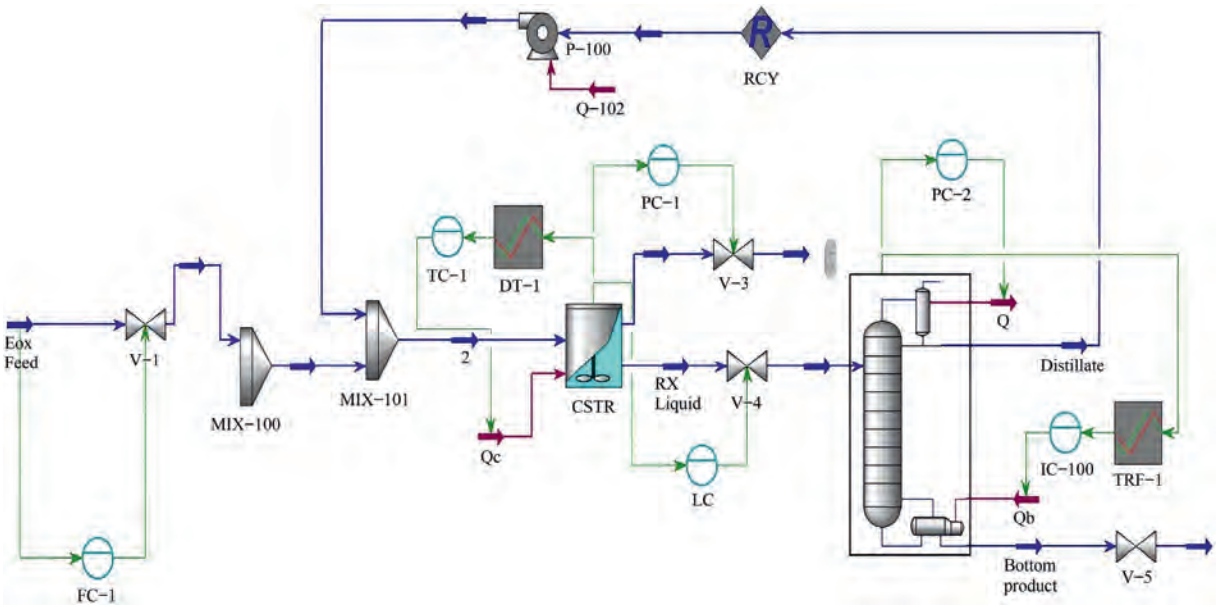

Fig. 6. Schematic diagram of ASPEN Hysys model for propylene glycol production.

4. Case Studies

This section presents the proposed method for establishing the NOZ model of the CSTR-DC-recycle chemical processes through simulated case studies. First, the grey-box model parameter estimation is illustrated. Second, the NOZ modeling and its accuracy are demonstrated.

4.1. Grey-box model parameter estimation

In this case study, the parameter estimation of the grey-box model is illustrated. A steady-state Aspen Hysys model is built for the CSTR-DC-recycle chemical process for producing propylene glycol from hydrolysis of propylene oxide, as shown in Fig. 6. Aspen Hysys is a widely recognized software for modeling chemical processes [34]. The Aspen Hysys model consists of a uniformly mixed jacketed CSTR reactor and a horizontal distillation column with a recycle loop. The output from the reactor is the input to the distillation column. The phase behavior is modeled using the UNIQUAC thermodynamic model in Aspen Hysys. The reactor holdup V is set at 5.5 m^3 and a recycle ratio is given as 0.1. Table 2 gives the sizing of the reactor and distillation column values while the rest process variables and parameters are provided in Table 3 and Table 4. The Aspen Hysys model is simulated at different input values of F_i, C_{A_i} and T_o by taking step sizes $\Delta F_i = 0.3\ m^3 \cdot h^{-1}$ and to obtain $\Delta C_{A_i} = 0.1\ kmol \cdot m^{-3}$ and $\Delta T_o = 1$ K steady-state output values of C_{A_o} and C_{C_L} based on other input variables and parameters in Tables 2–4. The data samples of F_i, C_{A_i} and T_o are provided in Fig. 7 (a), (b), and (c), whereas the collected steady-state data samples of C_{A_o} and C_{C_L} are presented in blue line by Fig. 7(d) and (e). Since steady-state values are collected, the in Fig. 7 is used as an index for different steady-state operating conditions rather than as a time variable. Hence, variations of data samples in Fig. 7 do not represent frequent fluctuations of process variables in consecutive time instants.

A set of different pairs of parameters $\beta_{p,i} = [E, A_o]_{i=1}^{500}$ are estimated based on Eq. (31) by randomly selecting 1000 data samples iteratively from the data samples provided in Fig. 7. The optimal values of E and A_o with accuracy measurements are estimated using Algorithm 1. Fig. 8 presents the posterior probabilities of $\beta_{p,i}$. The estimated optimal pair are $\widehat{E} = 72335.86$ and $\widehat{A}_o = 1.71 \times 10^{13}$ bounded by 95% the confidence intervals $[72305, 72365]$ and $[1.69 \times 10^{13}, 1.72 \times 10^{13}]$, respectively, with the highest posterior probability of 0.89 as shown by Fig. 8. At a significance level $\alpha = 0.05$, the estimated optimal parameters $\widehat{E} = 72335.86$ and $\widehat{A}_o = 1.71 \times 10^{13}$ give the index $\eta_{\hat{\theta}_{\beta_p}}$ from Eq. (40) to be 24.9, which is greater than the threshold 3.96 in Eq. (42). Based on $\widehat{E} = 72335.86$ and $\widehat{A}_o = 1.71 \times 10^{13}$, the grey-box model is simulated using the same input data samples of F_i, C_{A_i}, C_{A_o} and T_o rom Fig. 7, producing the output C_{C_L} represented by red dashed lines in Fig. 7(e). The model achieves an accuracy of $F(\beta_p) = 0.937$ by Eq. (32) which is greater than the selected threshold $\delta_{threshold} = 0.9$ indicating that the grey-box model is accurate. Another independent data samples of $F_i, C_{A_i}, T_o, C_{A_o}$ and C_{C_L} are generated by the Aspen Hysys model, as given in Fig. 9. Based on these data samples,

Table 2
Sizing of the CSTR reactor and distillation column.

	Reactor	Distillation column separator	Distillation column bottom
Diameter/m	4.39	1.67	2.00
Height/m	6.59	3.34	2.27
Volume/m^3	100.00	7.34	8.56

Table 3
Fixed input variables.

Variable	Description	Value
T_i	Input temperature/K	300
T_d	Condenser temperature/K	306.18
T_L	Bottom temperature/K	490
T_D	Recycle temperature/K	306.21
T_R	Reboiler temperature/K	486.8
P_c	Distillation column pressure/kPa	213
P_{CSTR}	CSTR pressure/kPa	221.3

Table 4
Process parameters.

Parameter	Definition	Value
ρ	reactor input density/kg·m^{-3}	952
H_Δ	enthalpy of reaction/kJ·kmol^{-1}	−90000
C_p	reactor input specific heat capacity/kJ·kg^{-1}·K^{-1}	125
R	ideal gas constant/J·K^{-1}	8.314
E	activation energy/J·mol^{-1}	72362
U_A	heat transfer coefficient of cooling jacket multiplied by surface area/kJ·K^{-1}	360000
A_o	preexponential constant	1.7×10^{13}
α	recycle ratio	0.1

the grey-box model provides the estimated values of C_{C_L} with an accuracy of 0.972 (given by red dash lines in Fig. 9(e) which is also greater than $\delta_{\text{threshold}} = 0.9$. This further reinforces that the grey-box model can reliably represent the CSTR-DC-recycle process.

To validate whether the grey-box-model fits well with the current observed data, the surrogate data samples $\{C_{C_L}(t)\}_{t=1}^{100}$ shown in Fig. 10(a) are iteratively estimated based on the input values $F_i = 15.1, C_{A_i} = 38.15$, and $T_o = 301.65$ by sampling 100 model parameters from the Gaussian distribution in Eq. (45) with the mean and covariance matrix as,

$$\widehat{\beta}_p = \left[\widehat{E} = 72335.86, \widehat{A}_o = 1.71 \times 10^{13}\right]$$

$$\mathbf{Cov}_{\widehat{\beta}_p} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

The mean of $\{C_{C_L}(t)\}_{t=1}^{100}$ is 32.2278 (given by red vertical line) enclosed by 95% confidence interval $\left[22.3796, 38.3165\ \text{kmol}\cdot\text{m}^{-3}\right]$ (given by black dash lines) in Fig. 10-(b). The observed value for C_{C_L} based on the same inputes $F_i = 15.1, C_{A_i} = 38.15$, and $T_o = 301.65$ in Aspen Hysys is 31.9981 kmol·m^{-3} which is within the confidence interval $\left[22.3796, 38.3165\ \text{kmol}\cdot\text{m}^{-3}\right]$ indicating that the grey-box model fits well with the current observed data with 95% confidence. In other words, the grey-box model represents the process in a good manner.

To evaluate the grey-box model's interpolation and extrapolation capabilities, data samples inside and outside the NOZ model (presented later in Subsection 4.2 by Fig. 15) are generated by the Aspen Hysys model and shown in Figs. 11 and 12. In Fig. 15, the data samples for estimating the grey-box model presented in Fig. 7 are given in blue stars. First, 7 operating points selected in Table 5(a) are inside the NOZ, and 7 operating points in Table 5(b) are outside the NOZ, selected from data samples presented in Figs. 11 and 12. The outputs C_{C_L} of the fourteen operating points are simulated by the grey-box model. As shown in Fig. 15, among the fourteen operating points, only one operating point in the pink circle overlaps with the counterparts for

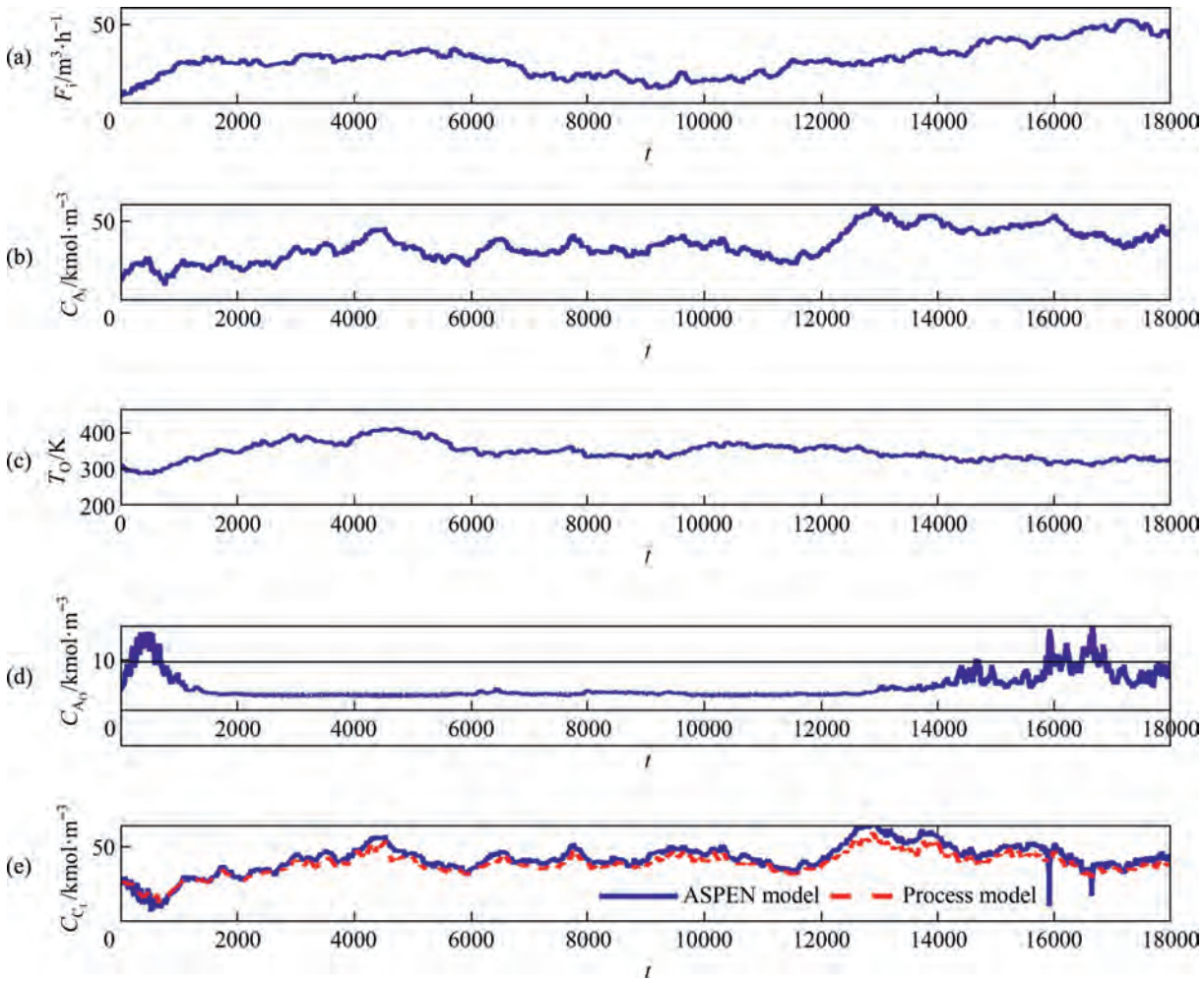

Fig. 7. Process data samples in the normal condition for estimating the grey-box model: (a) input flow rate F_i, (b) input concentration of propylene oxide C_{A_i} (c), reactor temperature T_o, (d) reactor output concentration of propylene oxide C_{A_o} and (e) bottom concentration of propylene glycol C_{C_L} (blue solid) and its estimate (red dash).

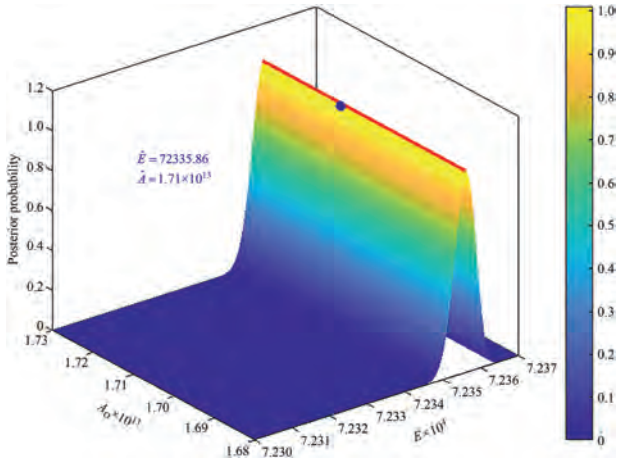

Fig. 8. Plot of posterior probabilities of the estimates E and A_o.

estimating the grey-box model, while the rest thirteen operating points in the red stars and green pluses do not overlap with the counterparts for estimating the grey-box model. Second, for the 14 operating points in Table 5, the simulated values of C_{C_L} from Aspen Hysys model and the grey-box model are very close, saying that the grey-box model is accurate and can be used to generated data to represent the process for modeling the NOZ.

4.2. Modeling the NOZ and determining its accuracy

In this case study, modeling of the NOZ and its accuracy are demonstrated. The NOZ is generated by the 4-dimensional geometric space formed by the input variables F_i, C_{A_i}, T_o and the simulated output C_{C_L} by the grey-box model provided in Fig. 9. Since direct visualization of this 4-dimensional space is not feasible, projections between C_{C_L} and F_i, C_{A_i} and T_o are given in blue dots in Fig. 13. According to process requirements, the

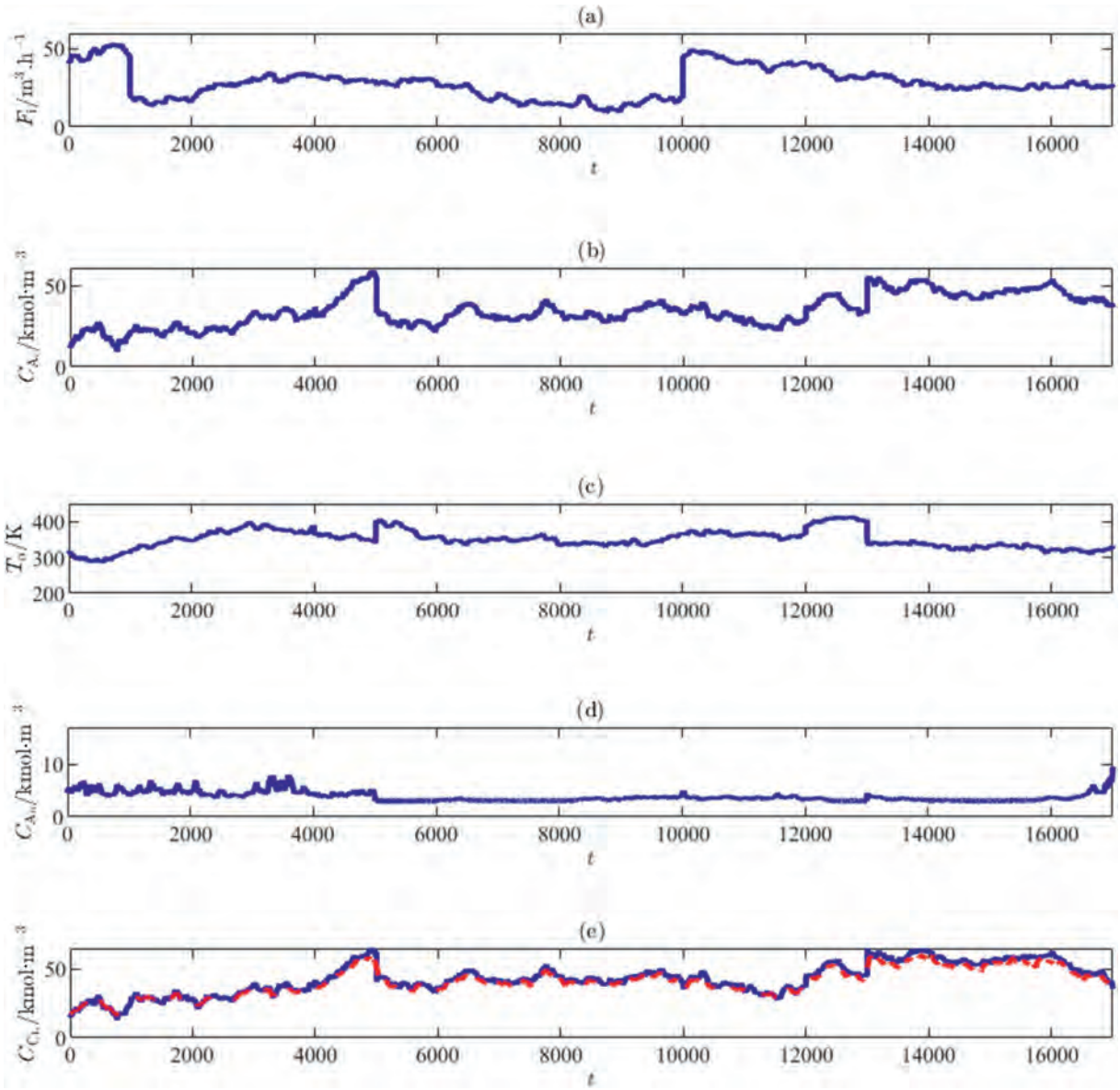

Fig. 9. Process data samples in the normal condition for estimating the grey-box model: (a) input flow rate F_i, (b) input concentration of propylene oxide C_{A_i}, (c) reactor temperature T_o, (d) reactor output concentration of propylene oxide C_{A_o} and (e) bottom concentration of propylene glycol C_{C_L} (blue solid) and its estimate (red dash).

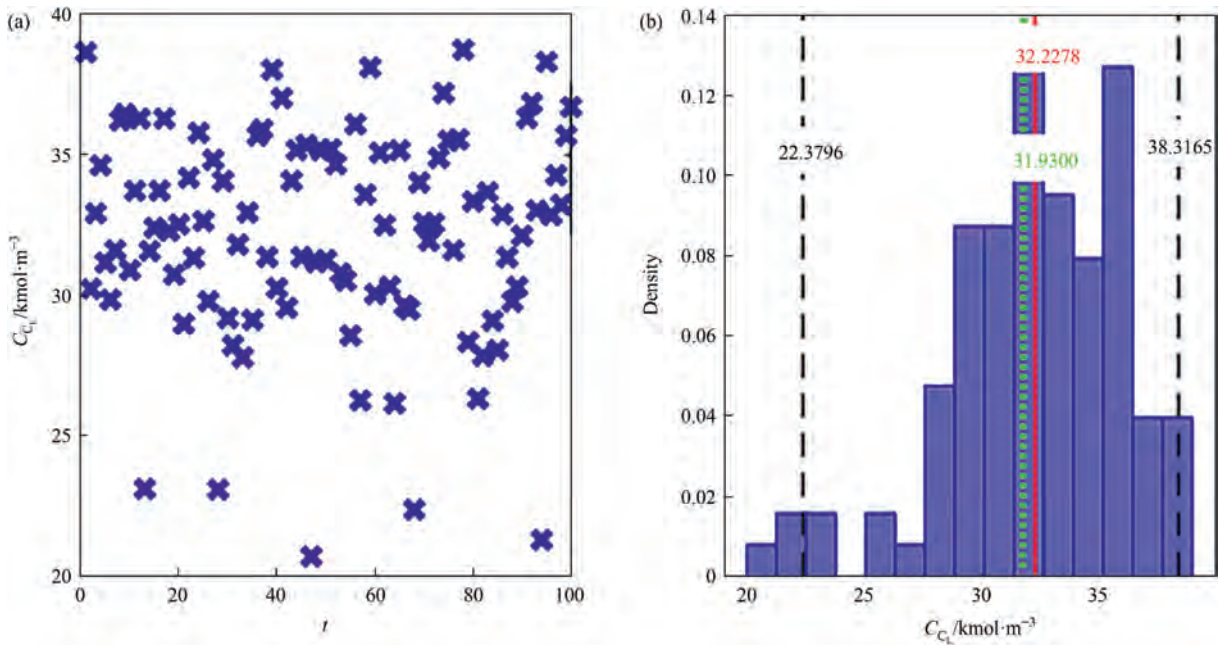

Fig. 10. (a) Scatter plot of surrogate estimates of C_{C_L} and (b) histogram plot of C_{C_L} with estimated mean, confidence interval and observed data given by red line, black dash lines and green dot line, respectively.

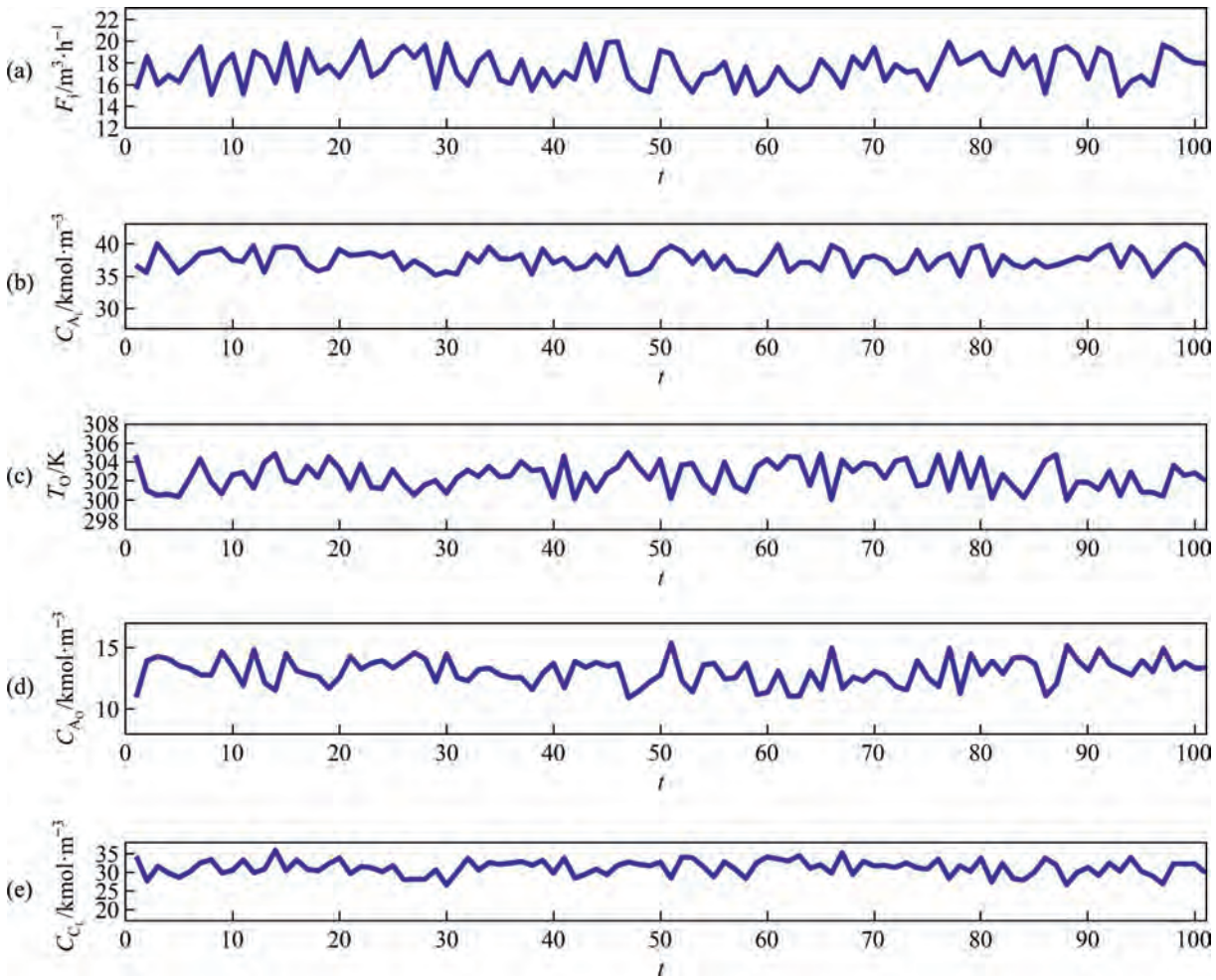

Fig. 11. Process data samples inside the NOZ for assessing the interpolation ability of the grey-box model: (a) input flow rate F_i, (b) input concentration of propylene oxide C_{A_i}, (c) reactor temperature T_o and (d) reactor output concentration of propylene oxide C_{A_o} and (e) bottom concentration of propylene glycol C_{C_L}.

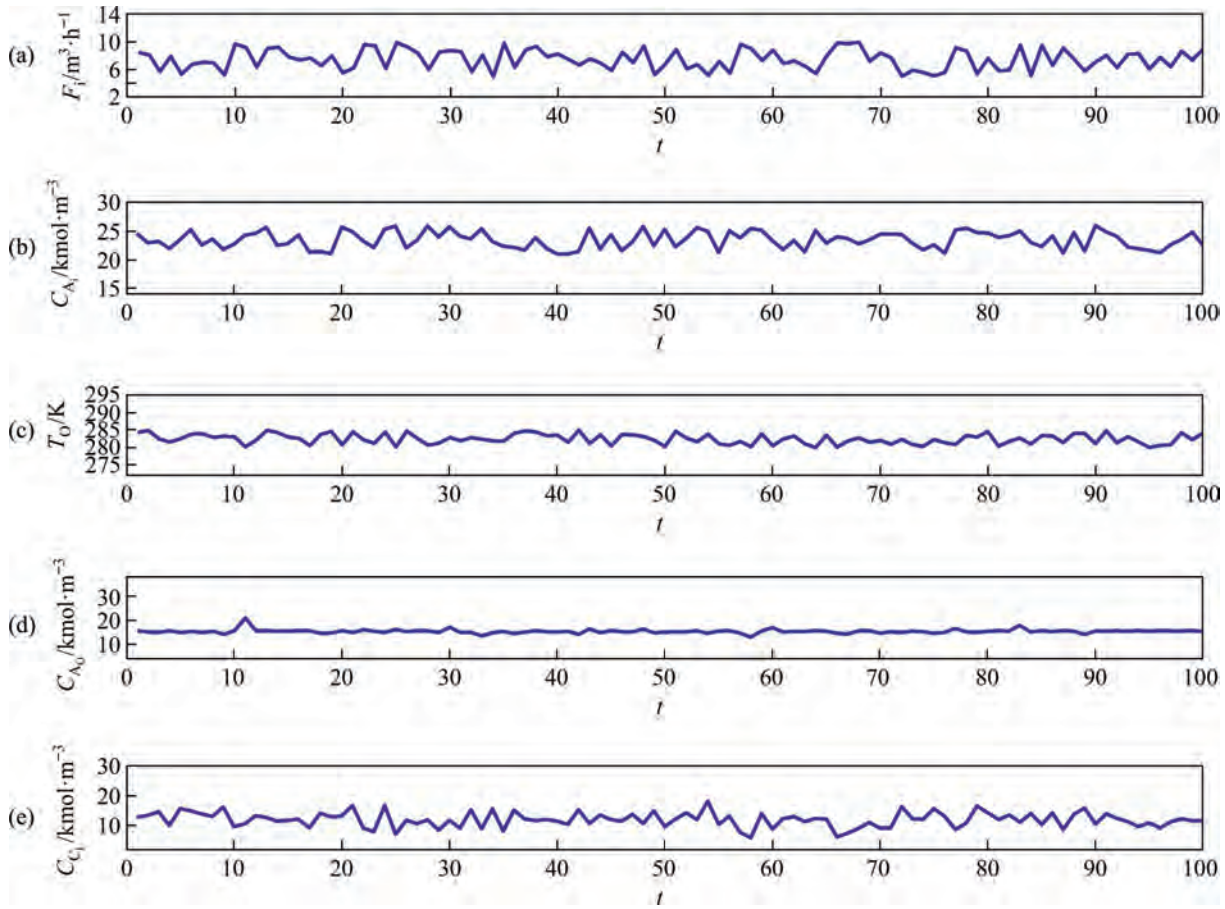

Fig. 12. Process data samples outside NOZ for assessing the extrapolation ability of the grey-box model: (a) input flow rate F_i, (b) input concentration of propylene oxide C_{A_i}, (c) reactor temperature T_o and (d) reactor output concentration of propylene oxide C_{A_o} and (e) bottom concentration of propylene glycol C_{C_L}.

Table 5
Operating points simulated by the grey-box model to: (a) illustrate interpolation ability of the grey-box model shown by the red and pink stars in Fig. 15 and (b) illustrate extrapolation ability of the grey-box model given in green pluses.

(a)							
Variable	Value						
F_i / m^3· h^{-1}	15.10	19.20	20.00	17.15	16.00	15.00	16.95
T_o/ K	301.65	303.50	301.50	304.75	301.60	301.85	303.00
C_{A_i}/ kmol·m^{-3}	38.15	37.20	38.5	36.75	36.95	36.60	37.30
C_{A_o}/ kmol·m^{-3}	9.6882	9.5025	11.9318	11.0494	10.9360	9.1165	9.6713
$C_{A_{L\text{-}ASPEN}}$/kmol·m^{-3}	31.9300	31.1010	28.9730	32.6500	31.2000	31.9900	32.1280
$C_{C_{L\text{-}model}}$/kmol·m^{-3}	32.7984	31.0492	28.9636	32.6539	31.1417	31.9893	32.1287
(b)							
Variable	Value						
F_i / m^3· h^{-1}	7.05	9.20	5.90	7.35	8.35	6.40	5.10
T_o/ K	284.00	280.20	284.70	281.40	282.8.0	284.80	280.65
C_{A_i}/ kmol·m^{-3}	23.85	24.80	25.00	21.60	22.00	23.80	21.80
C_{A_o}/ kmol·m^{-3}	12.1664	13.0500	11.6170	12.4840	11.9723	11.5794	11.9748
$C_{A_{L\text{-}ASPEN}}$/kmol·m^{-3}	14.3805	7.8105	17.6916	10.6634	14.7181	16.4321	13.5721
$C_{C_{L\text{-}model}}$/kmol·m^{-3}	14.3801	7.8100	17.6920	10.6632	14.7183	16.4400	13.5716

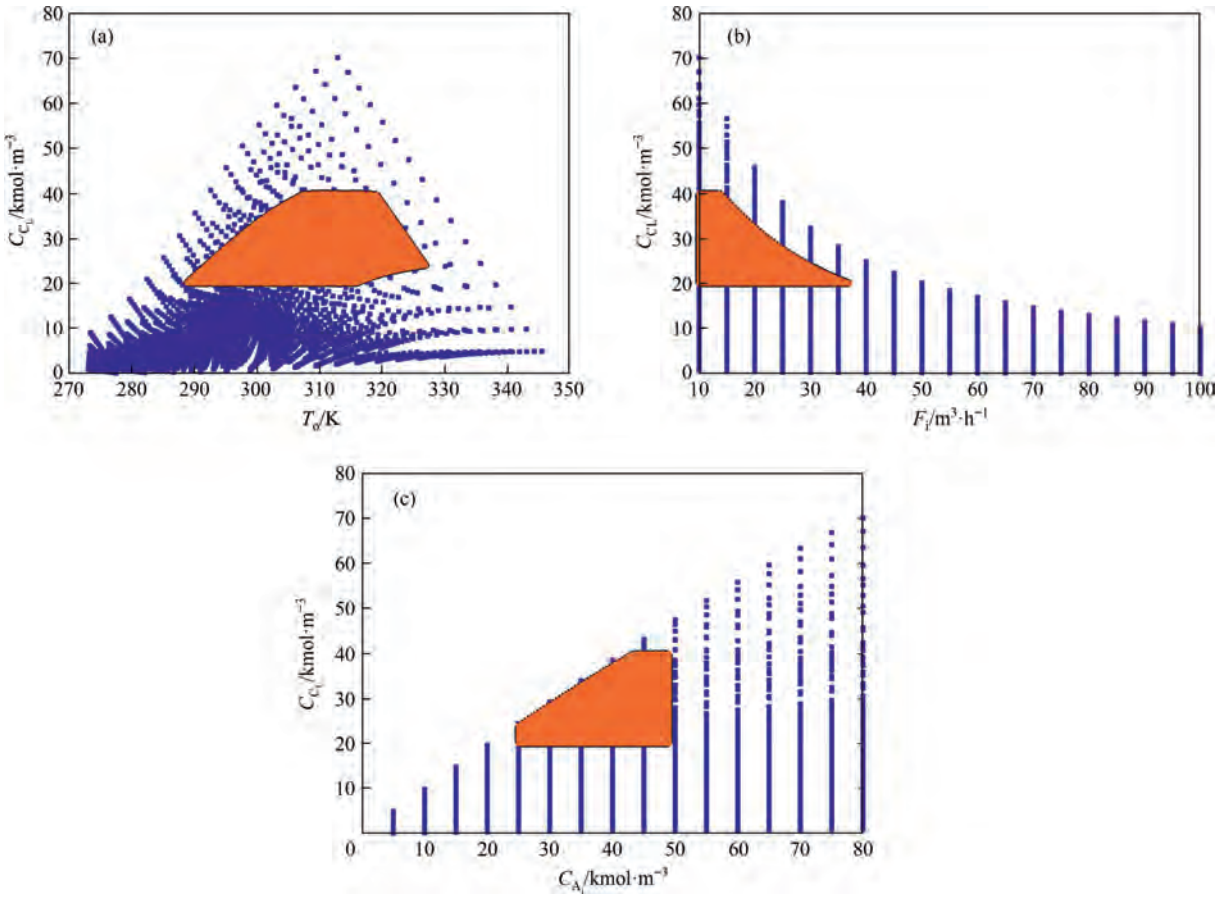

Fig. 13. Projected NOZ between: (a) T_o and C_{C_L}, (b) F_i and C_{C_L}, and (c) C_{A_i} and C_{C_L}.

range 20 kmol·m^{-3} ≤ C_{C_L} ≤ 40 kmol·m^{-3} is specified for normal operation in the CSTR-DC-recycle process. The simulated data samples that conform to this operational requirement are selected to formulate the NOZ, demarcated by the orange region in Fig. 13.

To model the NOZ, boundary operating points are extracted using Eq. (47). A plot of I_p for the NOZ data samples (orange region in Fig. 13) is presented in Fig. 14. A distinct inflection at $I_p = 0.67$ leads to the selection of this value for. Operating points with values above 0.67 are then used to represent the NOZ model. The extracted boundary points for the CSTR-DC-recycle NOZ model are shown as black lines in Fig. 15. To analyze model uncertainty, a data set of 100 different NOZs is generated from randomly selected input values. Table 6 lists the maximum distances $d_{m,y}$ of the NOZs with their occurrence numbers. The

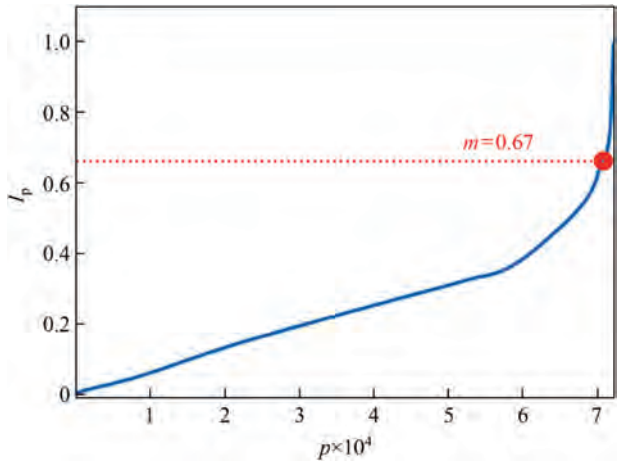

Fig. 14. Plot of boundary index I_p and p.

Table 6
Maximum distances $d_{m,y}$ in Eq. (51) and their occurrence numbers.

$d_{m,y}$	0.51	0.67	0.22	0.3	0.11
$q_{m,y}$	15	45	10	5	25

Table 7
Calculated values of process variables at the operating points G and H in Fig. 17 using Eqs. (9), (16), (18), (23) and (28).

Variable	Symbols	Value	
		G	H
Input flow rate/m^3·h	F_i	25	20
Input propylene oxide concentration/kmol·m^{-3}	C_{A_i}	32	30
Reactor temperature/K	T_o	329	297
Bottom propylene glycol concentration/kmol·m^{-3}	C_{C_L}	29	22

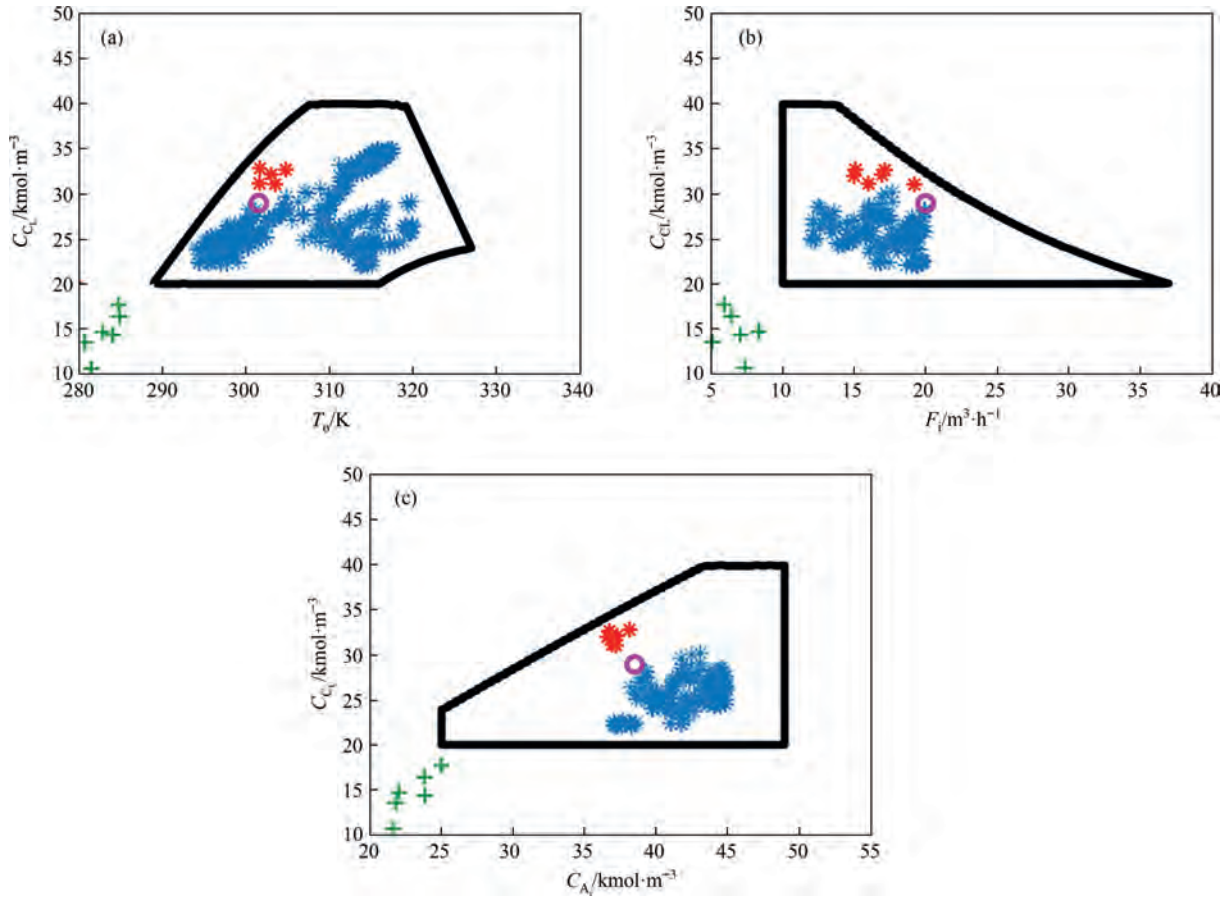

Fig. 15. Operating points used for modeling the NOZ (blue stars), illustrating the interpolation (red and pink stars) and extrapolation (green pluses) abilities of the grey-box model with projected NOZ between: (a) C_{C_L} and T_o, (b) C_{C_L} and F_i and (c) C_{C_L} and C_{A_i}.

probability estimates $\theta_{0.67}$ of the most frequent maximum distance $d_{m,y} = 0.67$ is 0.4509 bounded by $[0.35, 0.54]$ as shown in Fig. 16, with a ratio of 1.4 in Eq. (55) greater than 1 to confirm the accuracy of the NOZ model represented by the black lines in Fig. 15 .

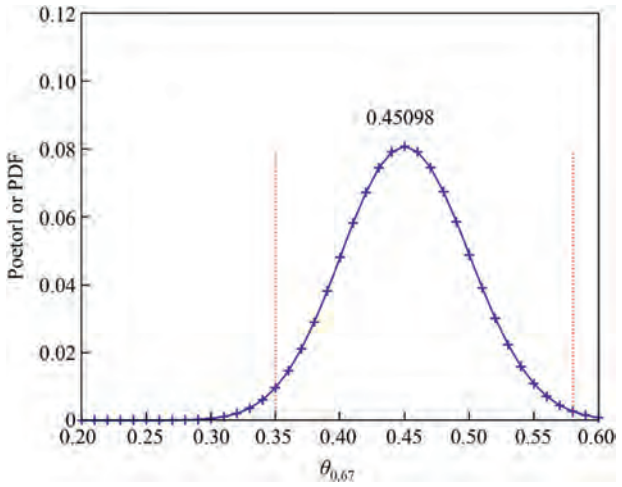

Fig. 16. Estimated probability of the most occurred maximum distance $d_{m,y} = 0.67$.

In comparison, Fig. 17 shows the proposed NOZ model (black line) against the convex-hull model (green dashed line) and the 95% ellipsoidal model (blue line). The fixed convex-hull and ellipsoidal models deviate from the true NOZ of the CSTR-DC-recycle process, which can lead to incorrect abnormality detection if they are used in place of the true NOZ. For an example, Table 7 gives values of abnormal operating points G and H. In Fig. 17, G (blue dot) falls outside the proposed NOZ and ellipsoidal models, indicating abnormality, but is detected as normal by the convex-hull model. Similarly, a normal operating point H (red star) is detected as normal by the proposed NOZ and convex-hull models but as abnormal by the ellipsoidal model.

5. Conclusions

The paper proposed a method to establish the NOZ model for CSTR-DC-recycle chemical processes based on boundary points by considering the process knowledge. In contrast, existing methods represented the NOZ model of chemical processes by fixed ellipsoidal and convex-hull models which did not consider the process knowledge. These fixed mathematical models were not suitable for the CSTR-DC-recycle chemical processes whose NOZs could not be represented by fixed mathematical models. By exploiting the boundary points to represent the NOZ model, the proposed method

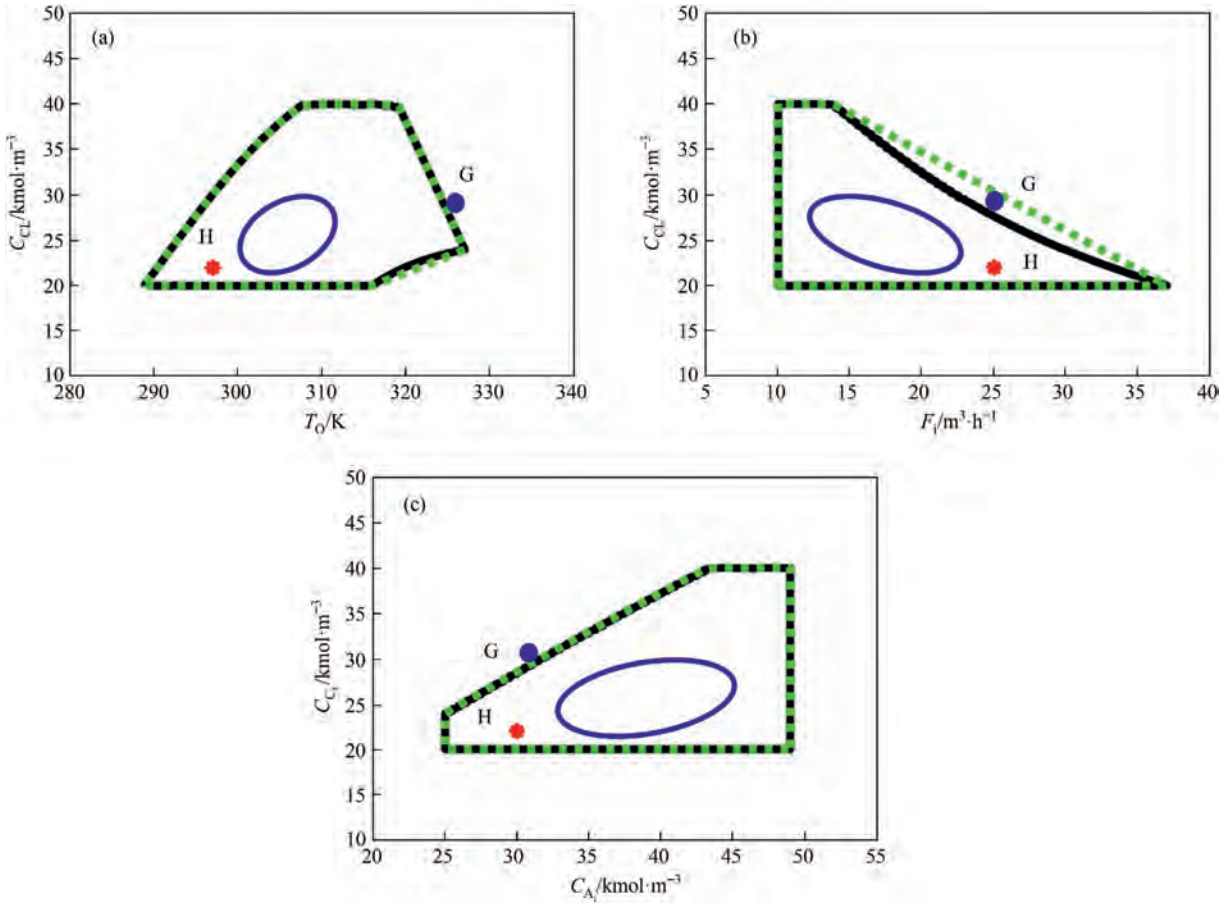

Fig. 17. Projected NOZ (given in black line) between: (a) T_o and C_{C_L}, (b) F_i and C_{C_L}, and (c) C_{A_i} and C_{C_L}. The purple line and green dot lines respectively give the 95% ellipsoid and convex-hull models.

gave a flexible method to model the NOZ of the CSTR-DC-recycle chemical processes whose NOZs could not be represented by fixed mathematical models. Simulated case studies based on Aspen Hysys showed that the proposed method was able to capture the true NOZ model of the CSTR-DC-recycle chemical process compared to fixed ellipsoidal and convex-hull models.

CRediT Authorship Contribution Statement

Poku Gyasi: Writing – original draft, Validation, Software, Methodology, Investigation, Formal analysis, Data curation. Jiandong Wang: Writing – review & editing, Visualization, Validation, Supervision, Resources, Project administration, Methodology, Funding acquisition, Formal analysis, Conceptualization. Mengyao Wei: Writing – review & editing, Visualization, Validation, Software, Methodology, Investigation, Formal analysis, Data curation. Hao Jing: Writing – review & editing, Validation, Software, Methodology, Investigation, Formal analysis, Data curation.

Declaration of Competing Interest

The authors whose names are listed above certify that they have NO affiliations with or involvement in any organization or entity with any financial interest (such as honoraria; educational grants; participation in speakers' bureaus; membership, employment, consul-tancies, stock ownership, or other equity interest; and expert testimony or patent-licensing arrangements), or non-financial interest (such as personal or professional relationships, affiliations, knowledge or beliefs) in the subject matter or materials discussed in this manuscript.

Acknowledgements

This paper was partially funded by the National Natural Science Foundation of China (62273215).

References

[1] V. Kumar, N. Kaistha, Invariants for optimal operation of a reactor-separator-recycle process, *J. Process. Contr.* 82 (2019) 1–12.
[2] M. Misra, H.H. Yue, S.J. Qin, C. Ling, Multivariate process monitoring and fault diagnosis by multi-scale PCA, *Comput. Chem. Eng.* 26 (9) (2002) 1281–1293.
[3] H.S. Fogler, Elementos de Ingenier'la de Las Reacciones Qu' Imicas, fourth ed., USA, Pearson Education, 2008.
[4] B. Betlem, B. Roffel, Process Dynamics and Control: Modeling for Control and Prediction, John Wiley and Sons, England, UK, 2009.
[5] J. Chen, C.M. Liao, Dynamic process fault monitoring based on neural network and PCA, *J. Process. Contr.* 12 (2) (2002) 277–289.
[6] H. Seki, Y.J. Naka, Optimizing control of CSTR/distillation column processes with one material recycle, *Ind. Eng. Chem. Res.* 47 (22) (2008) 8741–8753.
[7] D.E. Seborg, T.F. Edger, D.A. Mellichamp, F.J. Doyle III, Process Dyn. Control, John Wiley and Sons, England, UK, 2016.
[8] W.D. Tian, G.X. Zhang, X. Zhang, Y.X. Dong, PCA weight and Johnson transformation based alarm threshold optimization in chemical processes, *Chin. J. Chem. Eng.* 26 (8) (2018) 1653–1661.
[9] R. Brooks, R. Thorpe, J. Wilson, A new method for defining and managing process alarms and for correcting process operation when an alarm occurs, *J. Hazard Mater.* 115 (1–3) (2004) 169–174.

[10] D.H. Rothenberg, Alarm Management for Process Control: a best-practice Guide for Design, Implementation and Use of Industrial Alarm Systems, Momentum Press, New Jersey, USA, 2009.
[11] EEMUA (Engineering Equipment and Materials Users' Association), Alarm Systems: a Guide to Design, Management and Procurement vol. 191, EEMUA Publication, London, 2013.
[12] M. Kano, S. Hasebe, I. Hashimoto, H. Ohno, A new multivariate statistical process monitoring method using principal component analysis, *Comput. Chem. Eng.* 25 (7–8) (2001) 1103–1113.
[13] ISA, ISA-18.2, Management of Alarm Systems for the Process Industries, International Society of Automation, Durham, NC, USA, 2009.
[14] W.Y. Du, Y.W. Zhang, W. Zhou, Modified Non-Gaussian multivariate statistical process monitoring based on the Gaussian distribution transformation, *J. Process. Contr.* 85 (2020) 1–14.
[15] X.Z. Zhu, D.L. Gao, C. Yang, C.J. Yang, A blast furnace fault monitoring algorithm with low false alarm rate: ensemble of greedy dynamic principal component analysis-gaussian mixture model, *Chin. J. Chem. Eng.* 57 (2023) 151–161.
[16] L. Xie, J.M. Zhang, S.Q. Wang, Investigation of dynamic multivariate chemical process monitoring, *Chin. J. Chem. Eng.* 14 (5) (2006) 559–568.
[17] Y. Yu, J.D. Wang, Z.J. Ouyang, Designing dynamic alarm limits and adjusting manipulated variables for multivariate systems, *IEEE Trans. Ind. Electron.* 67 (3) (2020) 2314–2325.
[18] Y. Yu, J.D. Wang, Alarm monitoring for multivariate processes based on a convex-hull normal operating zone, *IEEE Trans. Contr. Syst. Technol.* 28 (6) (2020) 2649–2656.
[19] R. Smith, Chemical Process: Design and Integration, John Wiley and Sons, England, UK, 2005.
[20] M.Y. Wei, J.D. Wang, S. Gao, J. Li, X.K. Pang, An operating zone model for safety and efficiency monitoring of power generation units in thermal power plants, *Contr. Eng. Pract.* 153 (2024) 106101.
[21] E. Alshbuki, Simulation of the optimum operating conditions for A propylene glycol production unit using aspen hysys software, *J. Pure Appl. Sci.* 22 (3) (2023) 120–123.
[22] N. Olivier-Maget, F. Berdouzi, C. Murillo, N. Gabas, Deviation propagation along a propylene glycol process using dynamic simulation: an innovative contribution to the risk evaluation, *J. Loss Prev. Process. Ind.* 70 (2021) 104435.
[23] N.H. El-Farra, A. Gani, P.D. Christofides, Fault-tolerant control of process systems using communication networks, *AIChE J.* 51 (6) (2005) 1665–1682.
[24] G. Towler, R. Sinnott, Chemical Engineering Design: Principles, Practice and Eco Nomics of Plant and Process Design, Butterworth-Heinemann, Oxford, UK, 2021.
[25] Z. Wang, J.D. Wang, J.L. Hou, Multivariate alarm monitoring for non-convex normal operating zones based on search cones, *IEEE Trans. Autom. Sci. Eng.* 21 (1) (2024) 452–462.
[26] A. Chan, W.D. Seider, Batch Manufacturing of Propylene Glycol, Department of Chemical and Biomedical Engineering University of Pennsylvania, Pennsylvania, USA, 2004.
[27] K.K. Sharma, H. Krishna, Asymptotic sampling distribution of inverse coefficient-of-variation and its applications, *IEEE Trans. Reliab.* 43 (4) (1994) 630–633.
[28] H. Pishro-Nik, Introduction to Probability, Statistics and Random Processes, Kappa Research LLC, USA, 2016.
[29] D.C. Montgomery, G.C. Runger, Applied Statistics and Probability for Engineers, John Wiley and Sons, New Jersey, 2010.
[30] J.D. Wang, J.J. Su, Y. Zhao, X.K. Pang, J. Li, Z.F. Bi, Performance assessment of primary frequency control responses for thermal power generation units using system identification techniques, *Int. J. Electr. Power Energy Syst.* 100 (2018) 81–90.
[31] Z. Wang, J.D. Wang, J.D. Liu, Multivariate process monitoring for safe operation of condensers in thermal power plants based on normal operating zones, *IEEE Trans. Contr. Syst. Technol.* 32 (4) (2024) 1399–1409.
[32] R.A. Bates, H.P. Wynn, E.S. Fraga, Feasible region approximation: a comparison of search cone and convex hull methods, *Eng. Optim.* 39 (5) (2007) 513–527.
[33] V. Khalique, H. Kitagawa, T. Amagasa, BPF: a novel cluster boundary points detection method for static and streaming data, *Knowl. Inf. Syst.* 65 (7) (2023) 2991–3022.
[34] P.S. Roy, M.R. Amin, Aspen-HYSYS simulation of natural gas processing plant, *J. Chem. Eng.* 26 (2012) 62–65.