ACS OMEGA

http://pubs.acs.org/journal/acsodf

Article

Data-Knowledge-Driven Modeling and Operational Adjustment for the Pharmaceutical Tablet Manufacturing Process via Wet Granulation

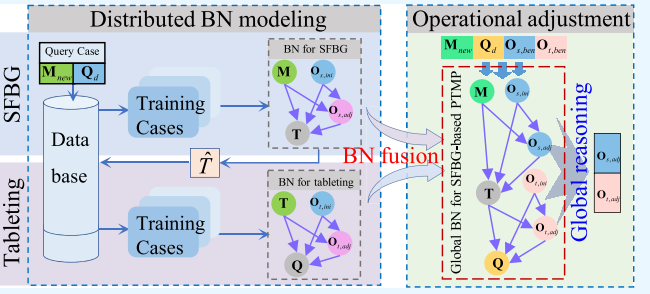
Zhengsong Wang, Shengnan Tang, Yanqiu Yang,* Yeqiu Chen, and Le Yang

Cite This: *ACS Omega* 2023, 8, 24441−24453

Read Online

ACCESS | Metrics & More | Article Recommendations

ABSTRACT: In the context of Pharma 4.0, pharmaceutical quality control (PQC) is beset by issues such as uncertainties from ever-changing critical material attributes and strong coupling between variables in the multi-unit pharmaceutical tablet manufacturing process (PTMP), and how to timely adjust the operational variables to deal with such challenges has become a key problem in PQC. In this study, we propose a novel data-knowledge-driven modeling and operational adjustment framework for PTMP by integrating Bayesian network (BN) and case-based reasoning (CBR). At the modeling level, first, a distributed concept is introduced, i.e., the BN model for each subunit of PTMP is established in accordance with the operation process sequence, and the transition variables are given by the BN model established first and retrieved as the new query for the next unit. Once the BN models of all subunits are built, they are integrated into a global BN model. At the operational adjustment level, by taking the expected critical quality attributes (CQAs) and related prior information as evidence, the operational adjustment is achieved through global BN reasoning. Finally, the case study in a sprayed fluidized-bed granulation-based PTMP demonstrates the feasibility and effectiveness in improving the terminal CQAs of the proposed method, which is also compared with other methods to showcase its efficacy and merits.

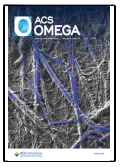


1. INTRODUCTION

Drugs are special commodities used to cure diseases and save lives, and their quality is of vital importance to people's well-being. Pharmaceutical manufacturing is therefore an area consisting of strict regulation and precise quality control of drug products.[1] However, owing to the potential variations or disturbance in the incoming raw materials, equipment conditions, or environmental factors, the departures of critical quality attributes (CQAs) of drugs from their acceptable scopes may be generated. In the context of Pharma 4.0,[2] in particular, the increasingly strong trend of and demand for drug customization lead to complex uncertainties manifesting as frequently changing critical material attributes (CMAs), which make it extremely challenging for traditional "experience + experiment"- or model-based pharmaceutical quality control (PQC) methods to successfully reduce costs and improve efficiency during pharmaceutical development.[1,3] Therefore, in response to the possible variations or disturbances, the automatic and optimal adjustment for operational variables would be urgently necessary to bring the CQA variables back to the established acceptance criteria while improving resource allocation efficiency.

Researchers have directed many efforts toward proposing precise and efficient PQC approaches for operational adjustment in order to ensure the drug quality. The related methodologies can roughly be divided into two categories of model-based control (MBC) and model-free control (MFC) depending on whether predictive models are used. As a typical representative of MBC, model predictive control (MPC) has found broad application in PQC studies for various pharmaceutical process units,[4] such as crystallization,[5] blending,[6,7] hot-melt extrusion,[8,9] continuous tableting,[10−12] etc. However, such studies generally rely on process mechanism models, and the real-time decision making are seriously hampered by whose modeling complexities and analytical solutions. Therefore, researchers tend to use data-driven models instead of mechanism models for MPC-based PQC, with related works including application studies in crystallization,[13] and tableting processes.[14,15] In contrast, MFC is a kind of heuristic and model-free control methodology based on process analysis and measurement technologies, and

Received: April 2, 2023
Accepted: June 20, 2023
Published: June 29, 2023



the related studies in PQC mainly focus on the applications of data-driven feedback control in granulation and tableting processes, such as the proportional−integral−derivative (PID) control[16] and data-driven iterative learning control (ILC).[1]

Nevertheless, due to the rising costs and falling efficiency resulting from the large number of fragmented, inefficient, and repetitive experiments, the infeasibility of MBC has been asserted in the PQC tasks facing the challenge of ever-changing CMAs.[1] Although MFC methods, in which the controller is designed directly using the measured input−output (I/O) data instead of explicit information from models, can alleviate this problem to a certain extent, it is a pity that such approaches simultaneously ignore some visible or latent knowledge captured in material streams, process mechanisms, and topological structures. This knowledge contains much additional information that precedes and is independent of data and meanwhile objectively reflects some of the laws and trends governing pharmaceutical process operations, which can greatly contribute to making up for any insufficiencies in the information provided by process data. Therefore, it is critically necessary to break through the traditional PQC modes and design more powerful PQC strategies by integrating data and knowledge from diverse sources, thus effectively solving the above PQC problems.

In recent years, data-knowledge-driven approaches have attracted increasing attention and interest in academia and have achieved the broad applications in solving various engineering problems such as system modeling and control, process monitoring, fault diagnosis, and so on.[17−23] Specifically, as an important artificial intelligence (AI) technique, the Bayesian network (BN) is a type of probabilistic graphical model that is capable of effectively integrating data and knowledge to simulate human reasoning. It represents the causal relations of variables by a directed acyclic graph, and so it has better interpretability than other AI methods.[24,25] Owing to the advantages in interpretability, probabilistic modeling, and dealing with data uncertainties, BNs have been widely applied to a variety of industrial systems and processes in different areas to successfully solve problems such as process monitoring, fault diagnosis, prognosis, risk assessment, decision making, etc.[26,27,30] However, the BN-based PQC studies in the pharmaceutical field are rarely reported. Because of its significant advantages and application potential, we attempt to utilize BN for PQC (or operational adjustment) tasks for the first time.

As far as our research background is concerned, although BN is showing great industrial application values, it still faces great challenges if ever-changing CMAs, strong coupling between variables and high nonlinearities exist. Additionally, the traditional discrete BN modeling has low adaptability to the frequently changing working conditions, and its network performance is also greatly affected by discretization.[30] Fortunately, as another important AI technique representing the knowledge in form of a case, case-based reasoning (CBR) is an empirical and knowledge-based reasoning method that draws on human thinking to deal with uncertain problems and is becoming widely popular for implementing the intelligence in various engineering areas.[30,31] CBR can make use of its idea of selecting similar (even the same) cases to cope with the frequently changing working conditions and provide targeted solutions. In view of the excellent combined performance of CBR and BN, as well as its sound application effects, the integration of BN and CBR has attracted increasing attentions and interests from lots of experts and scholars in various areas.[30−35] Nevertheless, none of the existing literature considers the integration of BN and CBR to study the operational adjustment for the PQC problem of pharmaceutical manufacturing processes.

Inspired by the advantages of integrated strategies, the motivation for this article is to introduce the idea of data-knowledge-driven approaches into PQC and present a modeling and operational adjustment framework by integrating BN and CBR based on two phases: distributed BN modeling and global operational adjustment. To demonstrate the preliminary implementation of above framework, a case study in a sprayed fluidized-bed granulation (SFBG)-based pharmaceutical tablet manufacturing process (PTMP) is presented, and the feasibility and effectiveness of the proposed framework are verified through a data experimental study. The main contributions of this case study are three-fold:

(1) A CBR-based data selection approach is proposed to pick similar cases out from database to construct the local incremental datasets used for dynamically training local BNs for different CMAs. In this way, the features of ever-changing CMAs can be substantially reflected by their similar cases and will further be learned by local BNs, thus effectively adapting to the variations from CMAs and offsetting their influences.

(2) To overcome the strong coupling between variables that seriously affects the modeling and control effects, SFBG and tableting processes are separated for distributed BN modeling, in which the process knowledge is used to preconfigure BN structures and the constructed local incremental datasets are responsible for the identification of BN parameters.

(3) According to the inheritance relationship between SFBG and tableting, the developed BN models for two units are integrated into a global BN, based on which the adjustment suggestions for each subunit's operational variables can be deduced by feeding the evidences to global BN, such as new CMAs, desired CQAs, and initial operational variables.

The rest of this article is organized as follows. Section 2 introduces the problem description and the preliminaries. Section 3 proposes a data-knowledge-driven modeling and operational adjustment framework for SFBG-based PTMP. In Section 4, a data experimental study is explored to demonstrate the feasibility and effectiveness of proposed framework. Finally, Section 5 concludes this article.

2. PROBLEM DESCRIPTION AND PRELIMINARIES

2.1. Problem Description. The most typical and common PTMP via wet granulation involves several successive units,[37] as is shown in Figure 1. Among the various types of wet granulation, SFBG is superior to others due to its inherent advantages of integrating blending, granulation and drying for granules in a single unit, which helps to improve material flows and densification, strengthen dust containment, and facilitate to achieve homogenous mixtures,[36] and thus SFBG is widely used in the pharmaceutical industry. For more details about the process description please refer to ref 37.

The CQAs are the target variables to represent drug quality and ensure the safety of medication for patients. The key factors affecting CQAs are CMAs and operating variables. In an SFBG-based PTMP, CMAs, operational variables, CQAs of

https://doi.org/10.1021/acsomega.3c02199
ACS Omega 2023, 8, 24441−24453

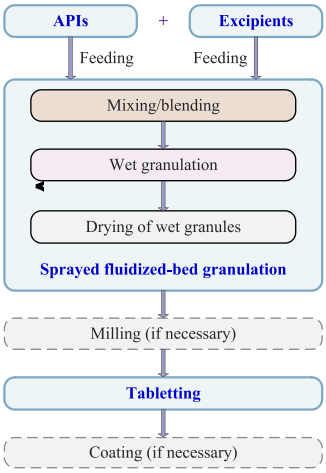


Figure 1. Schematic diagram of typical PTMP via wet granulation.

tablets and their relationships are shown in Figure 2, in which their symbolic definitions are also given for ease of expression.

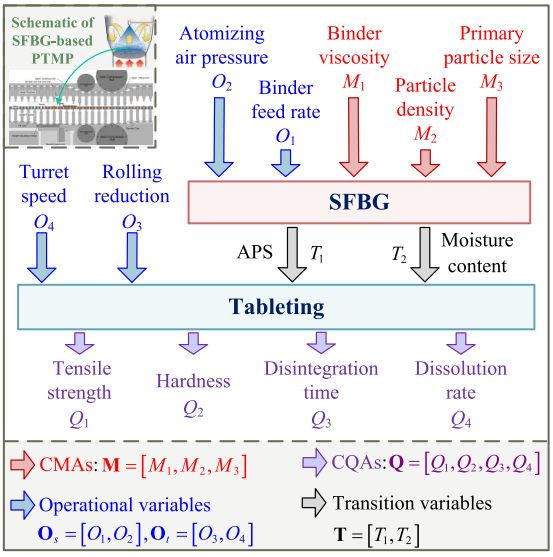


Figure 2. Diagram of variable relationships in the SFBG-based PTMP.

Note that as the CQAs of SFBG, average particle size (APS) and moisture content are also the CMAs of tableting, which relate SFBG and tableting by acting as the transition variables. The SFBG-based PTMP is briefly described as a general repeatable nonlinear system as follows:

$$\mathbf{T}_k = f_s(\mathbf{M}_k, \mathbf{O}_{s,k}) \tag{1}$$

$$\mathbf{Q}_k = f_t(\mathbf{T}_k, \mathbf{O}_{t,k}) \tag{2}$$

where $k \in Z_+$ is the iteration number, $f_s(\cdot)$ and $f_t(\cdot)$ are the unknown nonlinear functions.

The operational adjustment aims to adjust $\mathbf{O}_{s,k}$ and $\mathbf{O}_{t,k}$ so that CQAs follow a desired range of $[\mathbf{Q}_d - \boldsymbol{\delta}, \mathbf{Q}_d + \boldsymbol{\delta}]$, even when internal or external variations (or disturbances) occur, such as the changing CMAs. Here, $\mathbf{Q}_d$ is the median vector of desired range and is also the CQAs most expected to be achieved. The process is assumed as currently operating under an initial working condition of CMAs, i.e., $\mathbf{M}_{ini}$, with its operational variables defined as $\mathbf{O}_{s,ini}$ and $\mathbf{O}_{t,ini}$. When a new working condition ($\mathbf{M}_{new}$) comes, the operational adjustment should be activated to respond to such variations and give the operational adjustment values of $\mathbf{O}_{s,adj}$ and $\mathbf{O}_{t,adj}$, thus ensuring the conformity of drug quality. After compensation, the desired operational variables giving qualified CQAs under $\mathbf{M}_{new}$ can be obtained as follows:

$$\mathbf{O}_{s,d} = \mathbf{O}_{s,ini} + \mathbf{O}_{s,adj} \tag{3}$$

$$\mathbf{O}_{t,d} = \mathbf{O}_{t,ini} + \mathbf{O}_{t,adj} \tag{4}$$

2.2. Preliminaries. A BN is a type of acyclic graphical model that is widely utilized in the field of uncertain knowledge expression and reasoning. It is composed of variables, causal relationships between these variables, and the associated transfer probabilities. For a BN comprising n variables or nodes, its universal mathematical representation can be expressed as[38]

$$P(X_1, X_2, \ldots, X_n) = \prod_{i=1}^{n} P(X_i|\mathbf{Pa}_i) \tag{5}$$

where $P(X_1, X_2, \ldots, X_n)$ is the joint probability distribution of a BN and$P(X_i \mid \mathbf{Pa}_i)$ is the conditional probability distribution of X_i conditioned on its parent nodes denoted as $\mathbf{Pa}_i$.

Based on the data types of node variables, BN can be classified into discrete, continuous and hybrid BNs. In pursuit of more precise drug quality indexes, all the node variables are continuous types, which allows the BN model to be regarded as a continuous BN, or, namely, a Gaussian BN.[30] In this case, the conditional probability distribution of each continuous node with respect to its continuous parent nodes can be defined as follows:[39]

$$P(X_i|Pa_i) = N(\varepsilon_0 + \varepsilon_1 x_{1(i)} + \cdots + \varepsilon_n x_{n(i)}, \sigma^2) \tag{6}$$

where $\mathbf{Pa}_i = \{X_{1(i)}, X_{2(i)}, \ldots, X_{n(i)}\}$ is the set of parents of node X_i, $x_{1(i)}, x_{2(i)}, \ldots, x_{n(i)}$ are the values of $X_{1(i)}, X_{2(i)}, \ldots, X_{n(i)}$, ε_0 is the independent coefficient, $\varepsilon_1, \ldots, \varepsilon_n$ are the coefficients assigned to each parent, and σ^2 is the variance of X_i.

In BN modeling, there exist two fundamental components: the structure learning and parameter learning. There are three primary methods for constructing a BN model: utilizing expert knowledge, utilizing data, or a combination of both. Since a large amount of knowledge, such as process mechanism, potential relationship between subunits, and coupling relationship between variables, is contained in PTMP, this paper uses process knowledge to acquire BN structure. After the structure is established, parameter learning for each BN node should be considered immediately in the next step, which is carried out in a data-driven way. Given the sufficient and complete historical data available in this paper, we choose to employ the maximum likelihood estimation (MLE) algorithm to estimate BN parameters, and the objective function can be defined as follows:[28]

$$J(\Theta) = \sum_{i=1}^{n} LL_i(\theta_{X_i}|Pa_i; D) \tag{7}$$

where Θ represents the parameter of BN, $\theta_{X_i} \in \Theta$ is the conditional probability distribution associated with X_i, $LL_i(\theta_{X_i} \mid Pa_i; D)$ is the logarithm of the likelihood function, $L_i(\theta_{X_i} \mid Pa_i; D)$ is the likelihood function with respect to the parent node

https://doi.org/10.1021/acsomega.3c02199
ACS Omega 2023, 8, 24441−24453

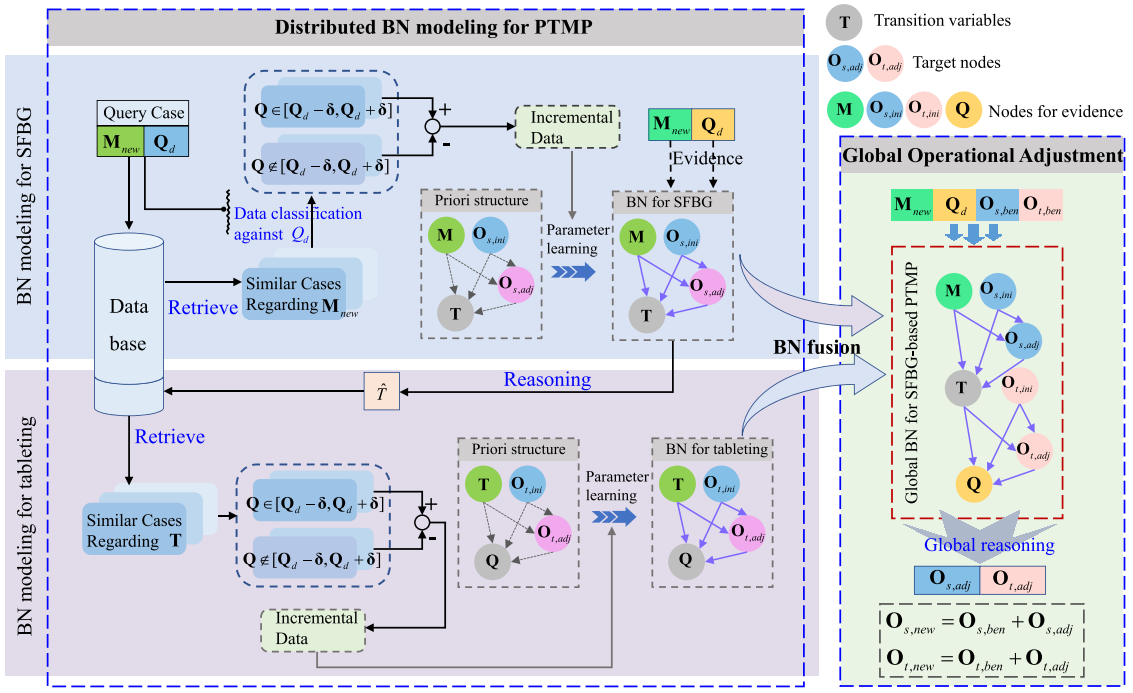


Figure 3. Schematic diagram of data-knowledge-driven modeling and operational adjustment framework.

set $\mathbf{Pa}_i$ and parameters θ_{X_i}, D is the observed data points $\{x_{i,k}\}$, where k denotes the node index, namely, training data.

CBR is an analogical reasoning method that extracts the solutions of previous cases to cope with the problems in new cases. For multi-unit processes such as SFBG-based PTMP, their historical operation information can be directly called by CBR, making the solution implementation process relatively simple and easy to implement in the real industrial scenarios. On this basis, this paper further introduces CBR as a supplement to solve the problem of inaccurate BN modeling in SFBG-based PTMP facing ever-changing CMAs. Please refer to refs 30 and 32 for more details about solving a new problem by CBR.

3. DATA-KNOWLEDGE-DRIVEN MODELING AND OPERATIONAL ADJUSTMENT FRAMEWORK FOR SFBG-BASED PTMP

Here, a data-knowledge-driven modeling and operational adjustment framework for dealing with SFBG-based PTMP is presented by integrating BN and CBR, as illustrated in Figure 3, which broadly divided into two key components: (1) distributed BN modeling and (2) global operational adjustment. As previously mentioned, SFBG and tableting processes are strongly coupled and highly nonlinear after integration. So, to reduce the caused modeling complexity, a distributed BN modeling idea is adopted. Concretely, SFBG and tableting processes are modeled separately, and the models are finally fused together for global operational adjustment.

3.1. Database Formulation. As the pharmaceutical manufacturing process proceeds and is accompanied by PQC, historical data under different CMAs ($\mathbf{M}_{\#1}$, $\mathbf{M}_{\#2}$, ⋯, $\mathbf{M}_{\#n}$) will be accumulated. A database is assumed to have been established and to include a certain amount of historical data, whose schematic structure is illustrated in Figure 4. Based on whether CQAs fall within the desired range of $[\mathbf{Q}_d - \boldsymbol{\delta}, \mathbf{Q}_d + \boldsymbol{\delta}]$, there are two types of data in the database: one is called

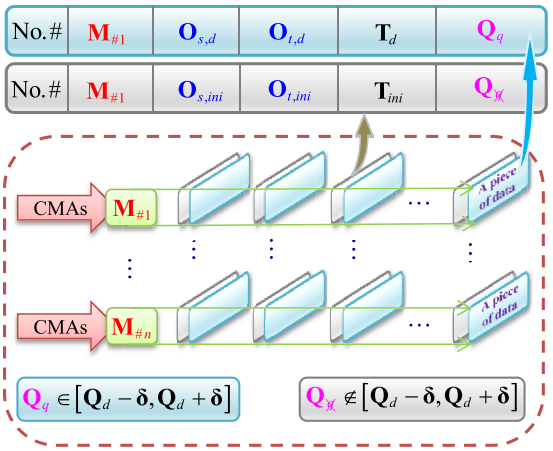


Figure 4. Schematic structure of the database.

successful data (see the blue block data in Figure 4), that is, the operational variables after adjustment result in well qualified CQAs ($\mathbf{Q} \in [\mathbf{Q}_d - \boldsymbol{\delta}, \mathbf{Q}_d + \boldsymbol{\delta}]$) under specific CMAs $\mathbf{M}_{\#1}$, ⋯, $\mathbf{M}_{\#n}$; the other is unsuccessful, that is, the CQAs are still unqualified after operational adjustment, as is represented by the gray block in Figure 4. These two types of data are chosen to be half of each for balance. That is, the data needed to train BN comes in pairs, and they impart implicit experience from operational failure to success to the BN. Therefore, these two types of data form the basis for proposing effective operational adjustment schemes. Finally, the database will also be dynamically updated when a complete piece of data under $\mathbf{M}_{new}$ is acquired.

3.2. BN Modeling for SFBG-Based PTMP. Here, the BN modeling for SFBG-based PTMP is given based on the following several steps.

3.2.1. Step 1: Data Standardization. When $\mathbf{M}_{new}$ comes, standardize $\mathbf{M}_{new}$ and $\mathbf{M}_{\#i}$ to $\mathbf{V}_{new}$ and $\mathbf{V}_i$, respectively,

https://doi.org/10.1021/acsomega.3c02199
ACS Omega 2023, 8, 24441−24453

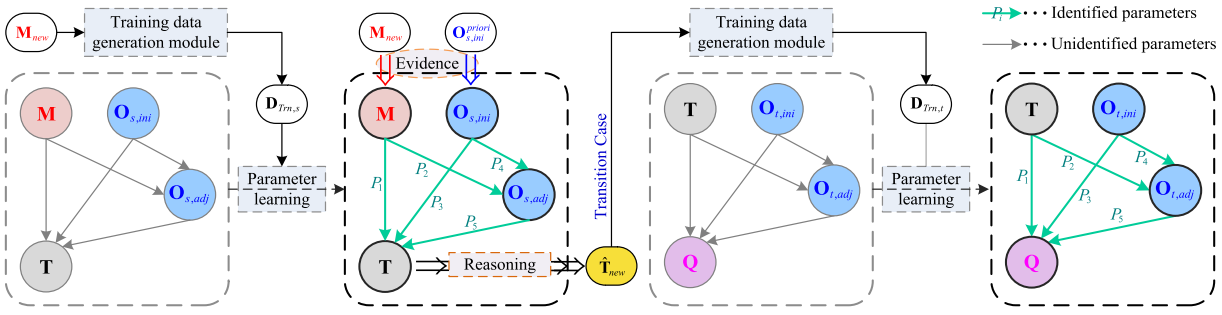


Figure 5. Schematic diagram of BN modeling for SFBG-based PTMP.

according to the mean vector $\boldsymbol{\mu}_M$ and standard deviation vector $\boldsymbol{\sigma}_M$ of $\mathbf{M}_{\#i}$ in the database:

$$\mathbf{V}_{new} = \frac{\mathbf{M}_{new} - \boldsymbol{\mu}_M}{\boldsymbol{\sigma}_M} \tag{8}$$

$$\mathbf{V}_i = \frac{\mathbf{M}_{\#i} - \boldsymbol{\mu}_M}{\boldsymbol{\sigma}_M} \tag{9}$$

where $i = 1,2, \cdots, n$.

3.2.2. Step 2: Training Data Generation for BN of SFBG. Access the database and calculate the similarity $J_{s,i}$ between $\mathbf{V}_{new}$ and $\mathbf{V}_i$ using the following similarity index to select the similar cases regarding $\mathbf{M}_{new}$:

$$d(\mathbf{V}_{new}, \mathbf{V}_i) = \|\mathbf{V}_{new} - \mathbf{V}_i\| \tag{10}$$

$$\cos\langle \mathbf{V}_{new}, \mathbf{V}_i \rangle = \frac{\mathbf{V}_{new}^T \mathbf{V}_i}{\|\mathbf{V}_{new}\|_2 \|\mathbf{V}_i\|_2} \tag{11}$$

$$J_{s,i} = \zeta \sqrt{e^{-d^2(\mathbf{V}_{new}, \mathbf{V}_i)}} + (1 - \zeta)\max(\cos\langle \mathbf{V}_{new}, \mathbf{V}_i \rangle, 0) \tag{12}$$

where $\zeta > 0$ a is a weight factor.

All the selected case data pairs are stored in a temporary local dataset $\mathbf{D}_1^s$. Now, the desired CQAs $\mathbf{Q}_d \in [\mathbf{Q}_d - \boldsymbol{\delta}, \mathbf{Q}_d + \boldsymbol{\delta}]$ are given, according to which each pair of case data $\mathbf{C}_1^l$ in $\mathbf{D}_1^s$contains two aspects of operation: one is the initial operation $\mathbf{O}_{s,ini}^l$ giving unqualified CQAs and another is the satisfied operation $\mathbf{O}_{s,d}^l$ generating qualified CQAs after adjustment, that is,

$$\mathbf{C}_1^l = \{\mathbf{M}^l, \mathbf{O}_{s,ini}^l, \mathbf{O}_{s,d}^l, \mathbf{T}_q^l\} \tag{13}$$

where $l = 1,2, \cdots, N$, N is the number of case data pairs in $\mathbf{D}_1^s$and $\mathbf{T}_q^l$is the CQAs of SFBG.

The operational adjustment values $\mathbf{O}_{s,adj}^l$ can be obtained as follows:

$$\mathbf{O}_{s,adj} = \mathbf{O}_{s,d}^l - \mathbf{O}_{s,ini}^l \tag{14}$$

and then a training dataset comprising incremental data is formulated as:

$$\mathbf{D}_{Trn,s} = \{\mathbf{M}^l, \mathbf{O}_{s,ini}^l, \mathbf{O}_{s,adj}^l, \mathbf{T}_q^l\} \tag{15}$$

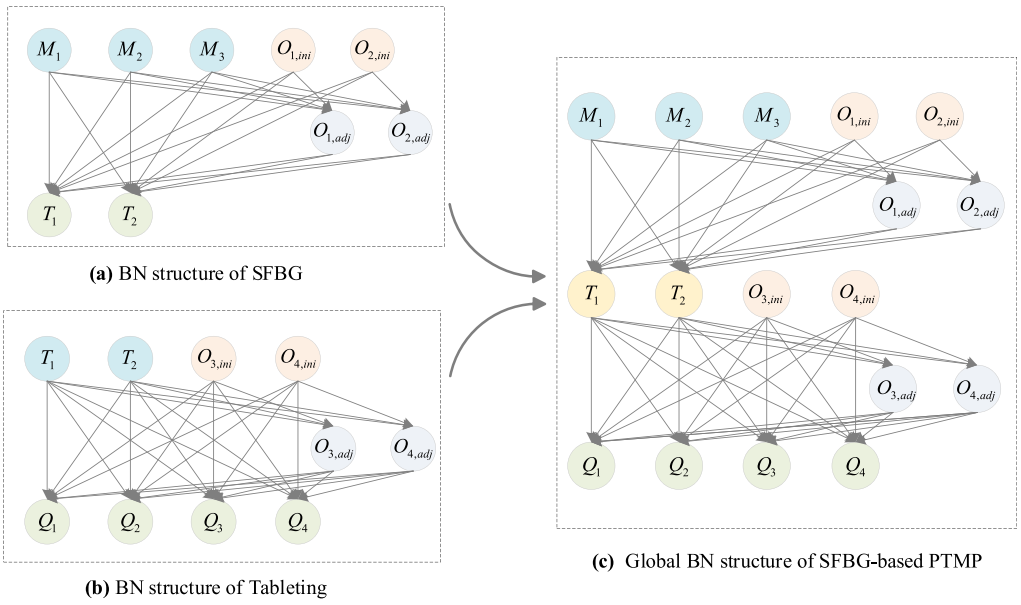
where $\mathbf{M}^l$ and $\mathbf{T}_q^l$are the original data from $\mathbf{D}_1^s$.

Remark 1: $J_{s,i}$ is the similarity between two cases, which is used to screen out a batch of data similar to the new case for modeling. In general, the more variables involved in a BN, the greater the number of similar samples required for accurate model establishment. The larger the value of $J_{s,i}$, the higher the similarity of the selected cases, but the number of similar cases will decrease, which is not enough to support BN modeling; in the contrary, if $J_{s,i}$ is too small, although more similar data can be retrieved, the similarity of the cases will decrease, resulting in an inaccurate BN model. The two similarity judgment methods based on angle and Euclidean distance are linked by the weight factor $\zeta \in [0,1]$. When $\zeta = 1$, (12) is completely transformed into a similarity calculation method based on Euclidean distance; when $\zeta = 0$, (12) becomes a similarity calculation method based on angle. Generally, when $\zeta = 0.5$, (12) is suitable for most similarity calculation scenarios, providing the advantages of both Euclidean distance-based and angle-based methods.

3.2.3. Step 3: BN Modeling for SFBG. According to the definition of BN, its modeling usually encompasses two learning elements: structure learning and parameter learning. In the field of industrial processes, the structure of BN can be presupposed as a priori by process knowledge. In the case of CMAs or process parameters changing, the operational variables must be promptly adjusted to direct unqualified CQAs caused by changing working conditions toward a qualified level (for SFBG, the CQAs here are what we call the transition variables). To this end, an essential BN structure that incorporates these key variables is critical. This paper presents a three-layer BN structure, as illustrated in Figure 5a. The top layer represents CMAs and initial states of operational variables, the middle layer represents adjustment values of operational variables, and the bottom layer represents adjusted CQAs of SFBG. The nodes in such BN structure correspond to four variables in the training dataset in (15). To achieve more precise control over the terminal output of entire process, this work adopts a BN model where all nodes are continuous variables. With the BN structure predetermined, the BN parameters of SFBG are identified based on the training dataset $\mathbf{D}_{Trn,s}$ generated in **Step 2** through MLE algorithm.

3.2.4. Step 4: Construction of Transition Case. The SFBG-based PTMP is naturally composed of two units, SFBG and tableting, which are modeled separately. To connect their BN models, it is crucial to pay attention to the coupling variables between subunits and use them to construct a transition case to link two units with an inheritance relationship. Here, we have learned that the CQAs of SFBG are the transition variables, which are transferred to the tableting process as the CMAs. We define $\mathbf{T}_{new}$ as the transition case when new CMAs of SFBG is given by $\mathbf{M}_{new}$, and its estimated value $\hat{\mathbf{T}}_{new}$ is obtained through BN reasoning, as illustrated in Figure 5b. To be more specific, the given $\mathbf{M}_{new}$ and a priori operation $\mathbf{O}_{s,ini}^{priori}$

https://doi.org/10.1021/acsomega.3c02199
ACS Omega 2023, 8, 24441−24453



(a) BN structure of SFBG

(b) BN structure of Tableting

(c) Global BN structure of SFBG-based PTMP

Figure 6. BN structures of SFBG-based PTMP.

are first fed into BN of SFBG as evidences for reasoning to obtain $\hat{\mathbf{T}}_{new}$, where $\mathbf{O}_{s,\,ini}^{priori}$ is the priori initial operations. Then, $\hat{\mathbf{T}}_{new}$ is regarded as a new query case for BN modeling of tableting, and the role of $\hat{\mathbf{T}}_{new}$ in BN modeling for tableting is the same as that of $\mathbf{M}_{new}$ in the BN modeling for SFBG.

3.2.5. Step 5: BN Modeling for Tableting. After estimating, $\hat{\mathbf{T}}_{new}$ is used as a query case for BN modeling of tableting. By emulating steps 1 and 2, the CBR-based methodology is employed to selectively extract a specified quantity of similar cases from the database to construct another temporary local dataset $\mathbf{D}_2^s$containing the following cases:

$$\mathbf{C}_2^l = \{\mathbf{T}^l, \mathbf{O}_{t,ini}^l, \mathbf{O}_{t,d}^l, \mathbf{Q}_q^{\,l}\} \tag{16}$$

based on which a BN training dataset is built as below:

$$\mathbf{D}_{Trn,t} = \{\mathbf{T}^l, \mathbf{O}_{t,ini}^l, \mathbf{O}_{t,adj}^l, \mathbf{Q}_q^{\,l}\} \tag{17}$$

where $\mathbf{O}_{t,\,adj}^l = \mathbf{O}_{t,\,d}^l - \mathbf{O}_{t,\,ini}^l$. Finally, the BN model of tableting is also developed by repeating step 3 with $\mathbf{D}_{Trn,\,t}$, as is shown in Figure 5c,d.

Remark 2: During the distributed BN modeling process, SFBG is subjected to the same modeling approach as that of tableting. However, there exists a precedence constraint in the modeling order, which is implied in Figure 2. In the modeling for SFBG, $\mathbf{T}^l$ is the CQAs of SFBG and also the CMAs needed to be retrieved in tableting, so $\mathbf{T}^l$ is named as the transition variable. If there are remaining units after the tableting, the CQAs of the tableting, namely, $\mathbf{Q}_q^l$, are still taken as the CMAs to be queried in the next unit. Figure 5 shows the cohesion of BN with two units as an example, and if there are more units, they can be connected in an orderly manner in that way.

3.3. Global Operational Adjustment Based on Distributed Model Fusion. After developing the distributed BN models for two units, they are further fused together for global operational adjustment, as is shown in Figure 3. The global BN incorporates two distributed models into its framework through the transition variables. Following the establishment of global BN, it is imperative to augment the pertinent evidence of each subunit in order to culminate in a comprehensive global reasoning of BN. In the SFBG layer, $\mathbf{M}_{new}$ and $\mathbf{O}_{s,\,ini}^{priori}$ are fed into the corresponding nodes as evidences, and $\mathbf{T}_{new}$ is estimated as a data bridge between two units. For tableting, the evidence nodes are $\mathbf{O}_{t,\,ini}$ and $\mathbf{Q}$, which are fed into $\mathbf{O}_{t,\,ini}^{priori}$ and $\mathbf{Q}_d$, respectively. After the reasoning with a junction tree algorithm,[29,40] the operational adjustment values are given as $\mathbf{O}_{s,\,adj}$ and $\mathbf{O}_{t,\,adj}$, and the final desired operational variables can be obtained as follows:

$$\mathbf{O}_{s,d} = \mathbf{O}_{s,ini}^{priori} + \mathbf{O}_{s,adj} \tag{18}$$

$$\mathbf{O}_{t,d} = \mathbf{O}_{t,ini}^{priori} + \mathbf{O}_{t,adj} \tag{19}$$

Remark 3: The method of selecting similar case data based on similarity criterions for local BN modeling can be referred to as just in time (JIT) learning, also known as lazy learning, employing global modeling techniques for localized domains. The prediction process of JIT can be executed in three stages: (1) selection of similar cases from a historical database to construct a dataset of comparable instances, based on a similarity criterion; (2) utilization of the constructed dataset to develop a localized model for the process; and (3) application of the localized model for prediction under new conditions, subsequently discarding the current localized model.

Remark 4: The steps for distributed BN modeling and global operational adjustment in Section 3.2 and Section 3.3 only target the adjustment suggestions for a single $\mathbf{M}_{new}$. For SFBG-based PTMP with ever-changing CMAs, whenever another new CMA, i.e., $\mathbf{M}'_{new}$, is encountered, the proposed framework initiates a fresh round of distributed BN modeling and global operational adjustment, dynamically generating the distinct local (or global) BN models completely independent of $\mathbf{M}_{new}$. Therefore, in addition to the distributed characteristics, another of BN modeling is the dynamic characteristics with the change of CMAs.

4. DATA EXPERIMENTAL STUDY: RESULTS AND DISCUSSIONS

In this section, a data experimental study is performed to verify the effectiveness of the proposed framework in the decision-making of operational adjustment for SFBG-based PTMP. The

https://doi.org/10.1021/acsomega.3c02199
ACS Omega 2023, 8, 24441−24453

Table 1. The Evidence Nodes Used for Reasoning and the Nodes Being Reasoned About

evidence nodes	reasoning nodes
M_1, M_2, M_3, $O_{1,ini}$, $O_{2,ini}$, $O_{3,ini}$, $O_{4,ini}$, Q_1, Q_2, Q_3, Q_4	$O_{1,adj}$, $O_{2,adj}$, $O_{3,adj}$, $O_{4,adj}$

historical data used for verification are generated from previous research on PQC problems through mechanistic models that have been shown to effectively simulate real processes. These models have already been summarized in our previous work in ref 37, and they will not be covered here. The BN modeling process is coded using the BN toolbox within Matlab 2016b, which can be downloaded from https://github.com/bayesnet/bnt.

The data experimental study is carried out from three levels: (1) historical data are used to build a case base. Part of the data is used as training samples to train BN model, while the other part is used as test samples to verify the effectiveness of BN. (2) The proposed method is compared with other methods to showcase the superiority. (3) A PQC running test is implemented to further illustrate the feasibility of BN modeling and global operational adjustment.

4.1. Performance Evaluation of Proposed Method on Test Samples. First, a total of 1200 adjusted operating samples, namely, historical data, are collected from a previous work as mentioned before, which contains the historical cases where the terminal CQAs conformed to the requirements after operational adjustments. Then, 1200 samples are divided into six partitions randomly, with one partition serving as the test set (200 samples) and the remaining five partitions serving as the case base (1000 samples). It should be additionally noted that the training set of the BN model is a set of cases that meet a certain degree of similarity filtered online by CBR from the aforementioned case base. To avoid the established BN model performing exceptionally well on a particular test set when comparing with other methods and to further verify the proposed method is statistically superior to other methods and not by chance, we conducted 10 repeated experiments where each experiment involved randomly re-dividing 1200 samples into the test set and the case base based on a certain proportion. Due to the page limitation, only 2 of the 10 repeated experiments are selected for presentation. The test samples in the two experiments are named test samples A and test samples B, respectively. In this section, simulation explanations will be made based on these two test samples.

Based on the process knowledge, the PTMP can naturally be segregated into two main subunits of SFBG and tableting. According to the BN structure in Section 3.2, 17 process variables can be finally involved for modeling, where four variables about the compensation value of operational variables are included, and the other variables are shown in Figure 2 in detail. Once subunits and variables are defined, the BN model structure of each subunit can be derived in advance using causality between variables. The BN model structures of SFBG and tableting are shown in Figure 6a,b. Next, the samples from test samples A or B are taken as query cases to test the performance of the proposed method. By applying the modeling approach in Section 3, the global BN model can be obtained, and the parameters and structures of each subunit are integrated into it through the transition variables. The structure of the global BN model is displayed in Figure 6c. It is worth noting that for each test sample, a global BN model with the same structure and different parameters will be built, as the CMAs are different for each test sample, that is, to simulate the ever-changing CMAs.

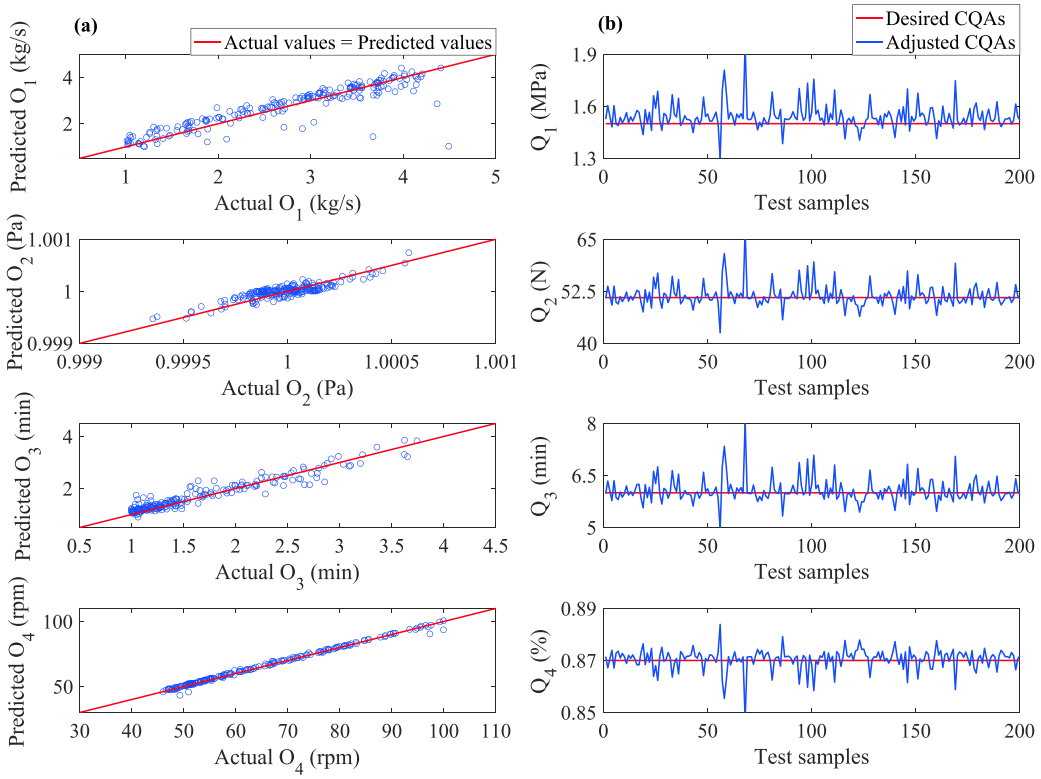


Figure 7. Adjustment results of test samples A; (a) operational variables; (b) terminal CQAs.

https://doi.org/10.1021/acsomega.3c02199
ACS Omega 2023, 8, 24441−24453

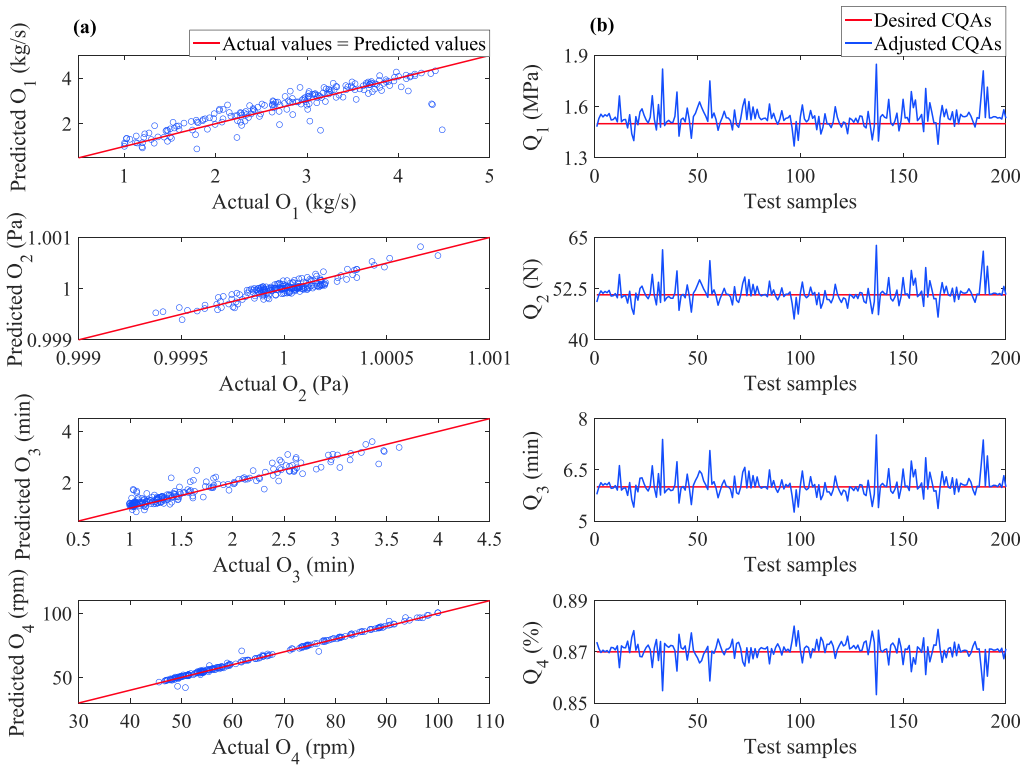


Figure 8. Adjustment results of test samples B; (a) operational variables; (b) terminal CQAs.

Table 2. Desired CQAs

$Q_{1,d}$ (MPa)	$Q_{2,d}$(N)	$Q_{3,d}$(min)	$Q_{4,d}$(%)
1.5	51	6.0	0.87

Table 3. Evaluation Results of Operational Variables and Terminal CQAs for Test Samples A

O_i/Q_i	MAPE (%)	RMSE	MAE
O_1/Q_1	9.8962/3.6288	0.4146/0.0820	0.2428/0.0544
O_2/Q_2	0.0107/3.2425	0.0002/2.6936	0.0001/1.6537
O_3/Q_3	8.5834/3.5639	0.1905/0.3446	0.1343/0.2138
O_4/Q_4	0.8134/0.3275	1.0757/0.0040	0.5180/0.0028

Table 4. Evaluation Results of Operational Variables and Terminal CQAs for Test Samples B

O_i/Q_i	MAPE(%)	RMSE	MAE
O_1/Q_1	9.8061/3.5952	0.3791/0.0773	0.2405/0.0539
O_2/Q_2	0.0126/3.1216	0.0002/2.5292	0.0001/1.5920
O_3/Q_3	9.7714/3.4327	0.2105/0.3241	0.1493/0.2060
O_4/Q_4	0.9145/0.3160	1.2693/0.0039	0.5524/0.0027

Table 5. Definition of each Method in Different Scenarios

scenarios	SFBG	tableting	method name
1	CBR is used	CBR is used	the proposed method
2	CBR is used		SMF-method I
3		CBR is used	SMF-method II
4			SMF-method III

Once distributed BN modeling is completed, then the relevant prior information is fed into the evidence nodes of the global BN network (as described in Section 3.3), operational adjustment results can be obtained through global BN reasoning. The evidence nodes used for reasoning and the variables being reasoned about are given in Table 1. The reasoning results $\mathbf{O}_{adj} = [\mathbf{O}_{s,adj}, \mathbf{O}_{t,adj}]$ are added to the initial operational variables to acquire the adjusted operational variables (called the predicted values) based on eqs 18 and 19 and compared with the actual adjusted operational variables (called the actual values). Note that the actual adjusted operational variables in test samples A or B give rise to the qualified CQAs.

The comparison results are shown in the parity plots for Figures 7a and 8a, where the *X*-axis represents the actual values and the *Y*-axis represents the predicted values. It can be seen that they are very close to each other. Figures 7b and 8b show the terminal CQAs after operational adjustment, with the desired CQAs are shown in Table 2. To more clearly demonstrate the feasibility of the adjusted results in Figures 7 and 8, three evaluation indexes denoted as MAPE, RMSE, and MAE are introduced:

$$\mathrm{MAPE} = \sum_{i=1}^{n} \left| \frac{s^i - \hat{s}^i}{s^i} \right| \times \frac{100}{n} \tag{20}$$

$$\mathrm{RMSE} = \sqrt{\frac{\sum_{i=1}^{n} (s^i - \hat{s}^i)^2}{n}} \tag{21}$$

$$\mathrm{MAE} = \frac{1}{n} \sum_{i=1}^{n} |s^i - \hat{s}^i| \tag{22}$$

where $\hat{s}^i$ are the actual or desired values, $\hat{s}^i$ is the predicted values after adjustment, and n represents the number of test samples. The evaluation results of operational variables and terminal CQAs for test samples A and B are given in Tables 3 and 4, respectively. From a practical engineering perspective, the precisions of the adjusted results meet the production

https://doi.org/10.1021/acsomega.3c02199
ACS Omega 2023, 8, 24441−24453

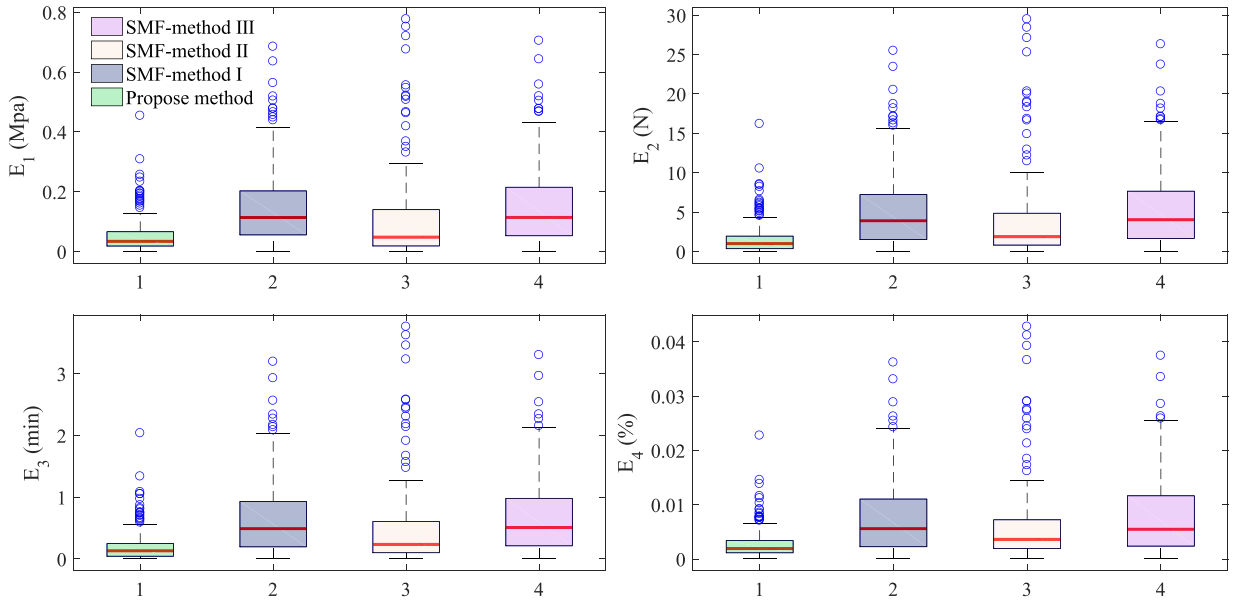


Figure 9. Absolute values of the errors of the CQAs for test samples A.

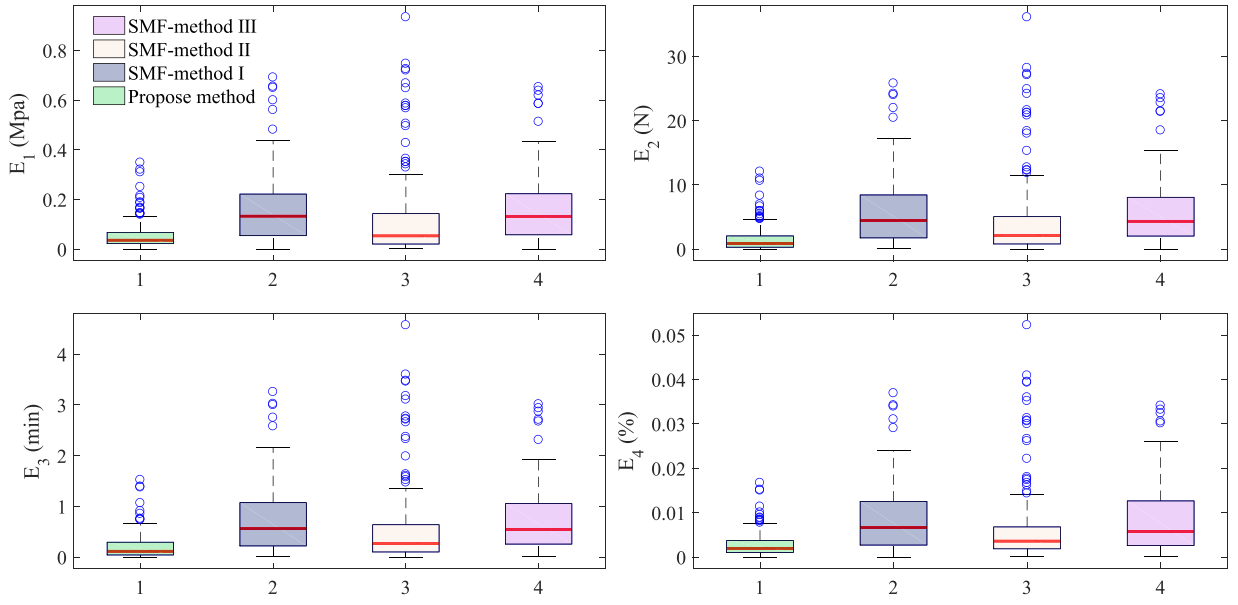


Figure 10. Absolute values of the errors of the CQAs for test samples B.

requirements, which proves the feasibility and effectiveness of the proposed method.

4.2. Comparative Study with Other Methods. The compensation of operating variables aims to optimize terminal CQAs, and the adjusted CQAs evidently have greater research interest. Therefore, the forthcoming analysis will emphasize on the adjusted CQAs.

To showcase the superior performance of the proposed method, a comparative study with several semi-finished methods is performed. In the modeling process of two subunits, the term "semi-finished methods" refers to whether CBR is used in constructing the BN model of each subunit. The detailed definitions of each method are exhibited in Table 5. Consistency in BN modeling, including network structure and parameter learning, is maintained across all methods, regardless of the use of CBR. Next, the definition of the absolute values of the errors between adjusted CQAs and desired CQAs are given as:

$$E_i = |Q_{i,d} - Q_{i,adj}|, i = 1, 2, 3, 4 \tag{23}$$

where $Q_{i,adj}$ is the ith adjusted CQA. Figures 9 and 10 show the box plots of the E_i of different methods for test samples A and B, and one can observe from plots that the errors obtained by the proposed method are generally smaller than those of other methods.

To transform each index under CQAs into the same order of magnitude for presentation, the following formula is given for exponential transformation,

$$L_{\mathrm{MAPE}} = \log_{1.001}(2 + \mathrm{MAPE}) \tag{24}$$

$$L_{\mathrm{RMSE}} = \log_{1.001}(2 + \mathrm{RMSE}) \tag{25}$$

https://doi.org/10.1021/acsomega.3c02199
ACS Omega 2023, 8, 24441−24453

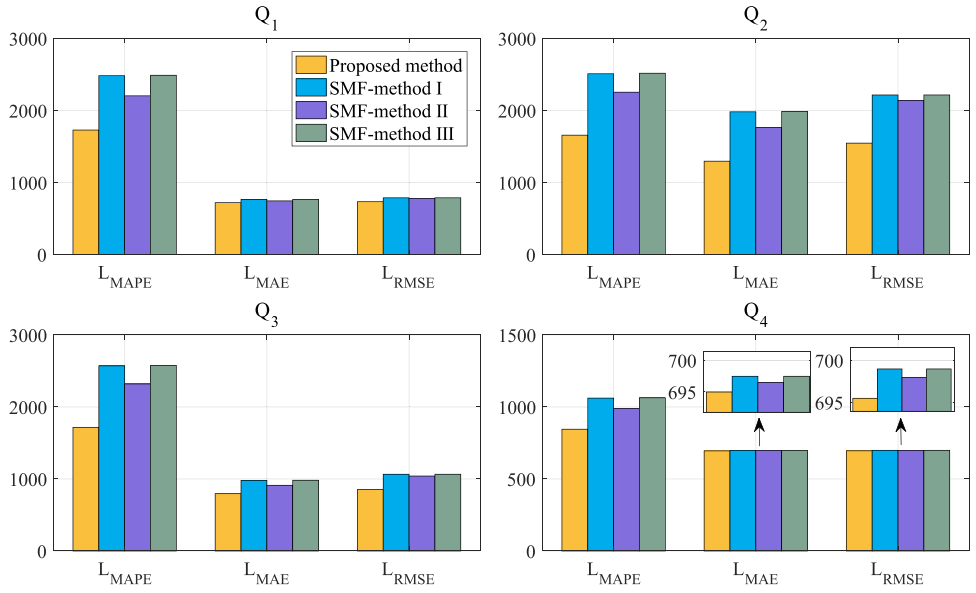


Figure 11. Comparison results of different indexes by the four methods for test samples A.

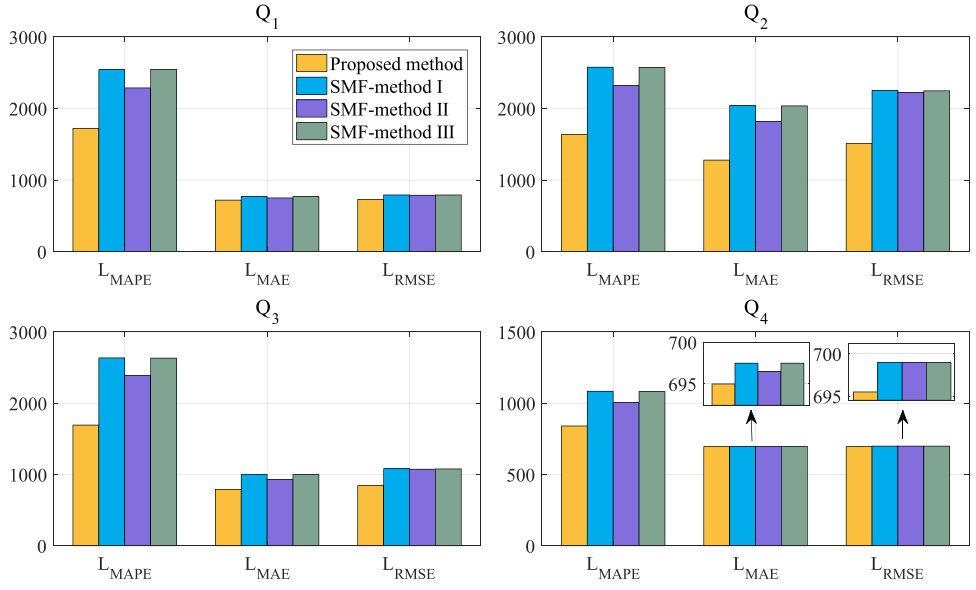


Figure 12. Comparison results of different indexes by the four methods for test samples B.

$$L_{\mathrm{MAE}} = \log_{1.001}(2 + \mathrm{MAE}) \tag{26}$$

When $x > 0$, the function $f(x) = \log_{1.001}(2 + x)$ shows explosive monotonically increasing, which can effectively increase the order of magnitude of smaller indexes, thus showing more clearly the difference between different indexes under different methods. In addition, the number 2 in $\log_{1.001}(2 + x)$ is to make sure that all the transformed indexes are greater than 0. Eventually, one can see that the proposed method yields smaller indexes, as is shown in Figures 11 and 12, whose results prove the better performance of proposed method.

In addition, it can be derived from Figures 9−12 that SMF-method III and SMF-method I have similar performance, and SMF-method II performs better than SMF-method III. These results are due to the fact that the tableting is the last unit of PTMP, which directly affects the terminal CQAs. If CBR is applied in tableting, a local BN model that is more consistent with the characteristics of current transition variables can be built, and a more accurate mapping relationship between the PTMP and global BN model can be established as long as CBR is applied in SFBG; otherwise, the BN model of tableting will immediately become inaccurate, thus affecting the accuracy of the global BN model and whether CBR is adopted in the SFBG has minute effect on the adjustment results.

Moreover, a statistical hypothesis test is also performed to allay the potential concern that the proposed method might perform well on a particular dataset but poorly on other unseen datasets. As described at the beginning of Section 4.1 for the 10-replicate experimental method, we compared the different evaluation indexes of each CQA under each replicate experiment for each method. Taking MAPE of tensile strength of the proposed method and SMF-method III as an example to show the hypothesis testing process. One can respectively obtain the MAPE of the proposed method and the SMF-method III after each experiment, namely, $\mathrm{MAPE}_i^{\mathrm{pro}}$ and

https://doi.org/10.1021/acsomega.3c02199
ACS Omega 2023, 8, 24441−24453

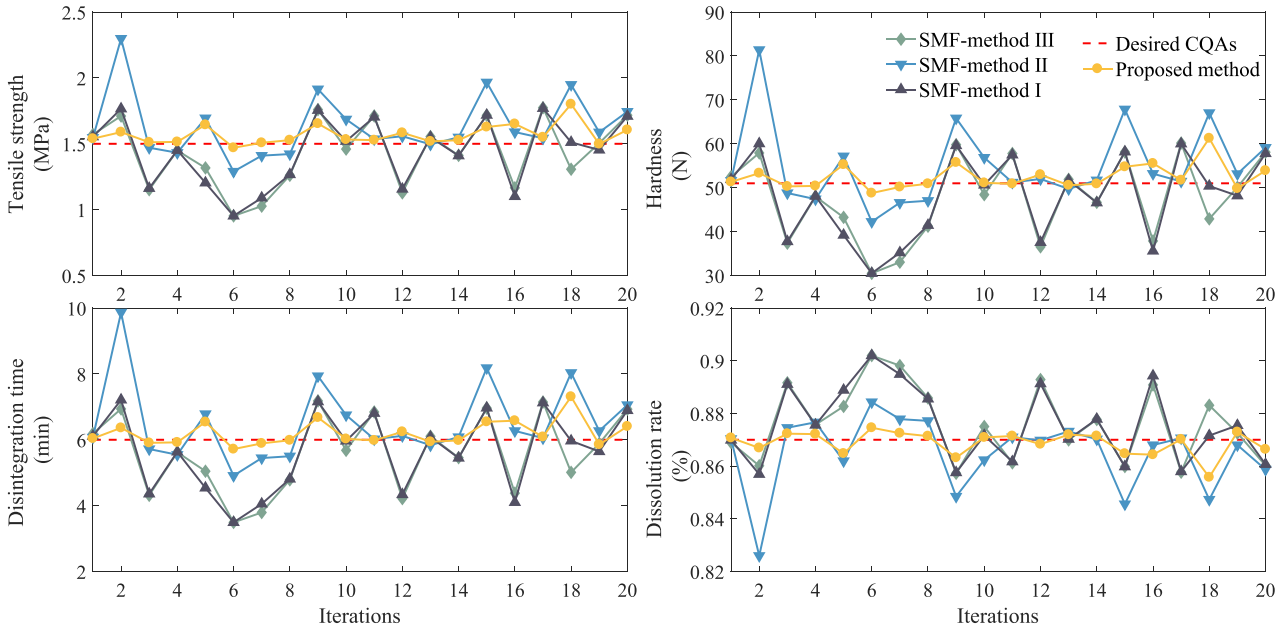


Figure 13. CQAs adjusted by different methods.

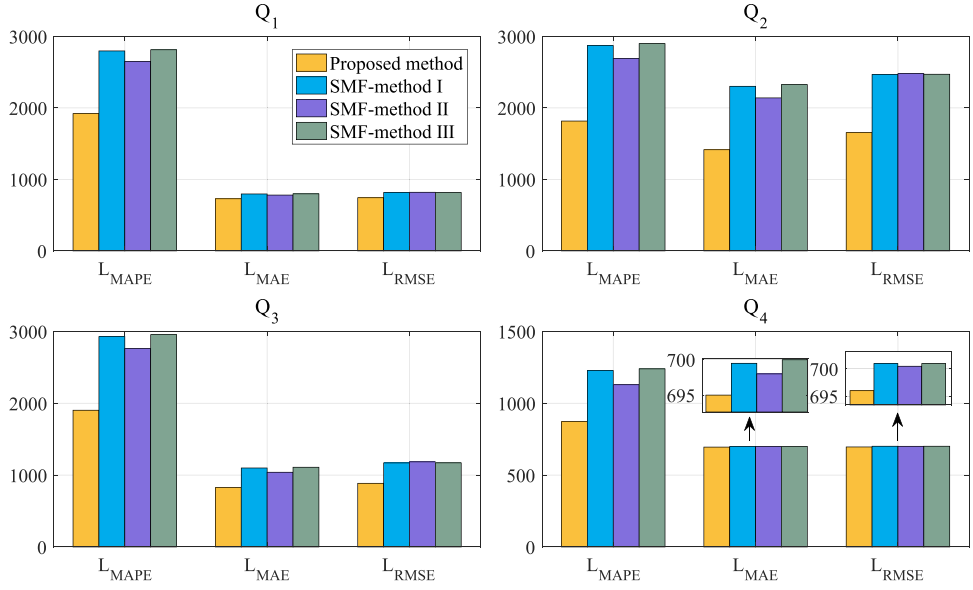


Figure 14. Comparison results of different indexes by the four methods.

$MAPE_i^{III}$, where i = 1, ...,10. Here, Student's test is used to check whether there are significant differences between the two methods. The difference in MAPE between the two methods is first calculated as follows:

$$\Lambda_i = MAPE_i^{pro} - MAPE_i^{III} \tag{27}$$

and then the mean and standard deviation of Λ_i is obtained respectively according to the following formula,

$$\mu_\Lambda = \sum_{i=1}^{10} \Lambda_i/10 \tag{28}$$

$$\sigma_\Lambda = \sqrt{\sum_{i=1}^{10} (\Lambda_i - \mu_\Lambda)^2/9} \tag{29}$$

Next, the null hypothesis $H_0 : \mu_\Lambda = 0$ is given, which assume no significant difference in performance between the two methods. The alternative hypothesis is$H_1 : \mu_\Lambda \neq 0$, representing a significant performance difference between the two methods. When the significance level is selected, if the t-statistic t_Λ is less than the critical value $t_{\alpha/2,\, n-1}$, the null hypothesis H_0 cannot be rejected; conversely, if the t-statistic t_Λ is greater than the critical value $t_{\alpha/2,\, n-1}$, it is reasonable to believe the alternative hypothesis H_1. Give the following formula to calculate the t-statistic,

$$t_\Lambda = \sqrt{10}\,|\mu_\Lambda/\sigma_\Lambda| \tag{30}$$

By calculating, the t-statistic t_Λ = 32.205. The critical value $t_{\alpha/2,\, n-1}$ for the two-sides test with n = 10 and α = 0.05 is $t_{0.025,9}$ = 2.262. Therefore, the null hypothesis H_0 is rejected and the alternative hypothesis H_1 is accepted because t_Λ =

https://doi.org/10.1021/acsomega.3c02199
ACS Omega 2023, 8, 24441−24453

32.205 > $t_{0.025,9}$ = 2.262. We have statistical evidence to indicate that there is a significant difference between the two methods. Based on the above results, one can conclude that the proposed method outperforms the SMF-method III. The comparisons with remaining methods can be evaluated using the same approach.

4.3. Further Validations by the Running Experimental Test. In this section, a running experimental test is performed to further verify the feasibility of the proposed method. In this experimental setup, the CMAs continuously change, and these new cases are completely unfamiliar to the PQC process. The real-time CQAs adjustment results by different operational adjustment methods are shown in Figure 13, and Figure 14 gives the comparative results of different indexes. As evidenced by the figures, the proposed method has a better performance than other methods.

In the actual pharmaceutical scenario, PQC is a challenging task for researchers. Typically, trial-and-error experiments are conducted to grope for operational conditions, and rough adjustments are made to ensure that the terminal CQAs meet requirements. Obviously, this approach is inefficient and imprecise. The proposed method develops a data and knowledge driven model for PTMP with the help of historical data and has a better effect.

5. CONCLUSIONS

In this work, we have used a combination of BN and CBR for operational adjustment to improve terminal CQAs in a SFBG-based PTMP. A distributed BN modeling method to alleviate the complexity of variable coupling in SFBG-based PTMP is first explored, in which the process knowledge is utilized to pre-establish the BN structures and the local similar data screened out by CBR is used for BN parameter learning. The developed BN models for all subunits are then integrated into a global BN model for global operational adjustments, which effectively responds to dynamically changing CMAs and further assures the qualified CQAs. Finally, extensive data experimental results in a SFBG-based PTMP demonstrate the feasibility and effectiveness of the proposed method, highlighting its potential to deliver practical benefits in real-world applications. The proposed method results are also compared to the other methods and achieve much superior performances than other methods.

AUTHOR INFORMATION

Corresponding Author

Yanqiu Yang − *College of Life and Health Sciences, Northeastern University, Shenyang 110169, China*; Email: 524448897@qq.com

Authors

Zhengsong Wang − *School of Control Engineering, Northeastern University at Qinhuangdao, Qinhuangdao 066004, China; College of Information Science and Engineering, Northeastern University, Shenyang 110819, China*; orcid.org/0000-0002-7614-7987

Shengnan Tang − *School of Control Engineering, Northeastern University at Qinhuangdao, Qinhuangdao 066004, China; College of Information Science and Engineering, Northeastern University, Shenyang 110819, China*

Yeqiu Chen − *School of Control Engineering, Northeastern University at Qinhuangdao, Qinhuangdao 066004, China*

Le Yang − *School of Control Engineering, Northeastern University at Qinhuangdao, Qinhuangdao 066004, China; College of Information Science and Engineering, Northeastern University, Shenyang 110819, China*

Complete contact information is available at: https://pubs.acs.org/10.1021/acsomega.3c02199

Funding

Funding was received from National Natural Science Foundation of China (grant # 62203095), Natural Science Foundation of Hebei Province (grant # F2021501009), and Fundamental Research Funds for the Central Universities of China (grant # N2223024).

Notes

The authors declare no competing financial interest.

REFERENCES

(1) Wang, Z.; He, D.; Hou, Y. Data-Driven Adaptive Quality Control Under Uncertain Conditions for a Cyber-Pharmaceutical-Development System. *IEEE Trans. Ind. Inf.* **2021**, *17*, 3165−3175.

(2) Arden, N. S.; Fisher, A. C.; Tyner, K.; Yu, L. X.; Lee, S. L.; Kopcha, M. Industry 4.0 for pharmaceutical manufacturing: preparing for the smart factories of the future. *Int. J. Pharm.* **2021**, *602*, No. 120554.

(3) Kapoor, Y.; Meyer, R. F.; Meyer, B. K.; DiNunzio, J. C.; Bhambhani, A.; Stanbro, J.; Ploeger, K. J. M.; Guidry, E. N.; Troup, G. M.; Procopio, A. T.; Templeton, A. C. Flexible manufacturing: the future state of drug product development and commercialization in the pharmaceutical industry. *J. Pharm. Innovation* **2021**, *16*, 2−10.

(4) Jelsch, M.; Roggo, Y.; Kleinebudde, P.; Krumme, M. Model predictive control in pharmaceutical continuous manufacturing: A review from a user's perspective. *Eur. J. Pharm. Biopharm.* **2021**, *159*, 137−142.

(5) Szilágyi, B.; Borsos, Á.; Pal, K.; Nagy, Z. K. Experimental implementation of a Quality-by-Control (QbC) framework using a mechanistic PBM-based nonlinear model predictive control involving chord length distribution measurement for the batch cooling crystallization of l-ascorbic acid. *Chem. Eng. Sci.* **2019**, *195*, 335−346.

(6) Celikovic, S.; Kirchengast, M.; Rehrl, J.; Kruisz, J.; Sacher, S.; Khinast, J.; Horn, M. Model predictive control for continuous pharmaceutical feeding blending units. *Chem. Eng. Res. Des.* **2020**, *154*, 101−114.

(7) Rehrl, J.; Kruisz, J.; Sacher, S.; Khinast, J.; Horn, M. Optimized continuous pharmaceutical manufacturing via model-predictive control. *Int. J. Pharm.* **2016**, *510*, 100−115.

(8) Sacher, S.; Celikovic, S.; Rehrl, J.; Poms, J.; Kirchengast, M.; Kruisz, J.; Sipek, M.; Salar-Behzadi, S.; Berger, H.; Stark, G.; Horn, M.; Khinast, J. G. Towards a novel continuous HME-Tableting line: Process development and control concept. *Eur. J. Pharm. Sci.* **2020**, *142*, No. 105097.

(9) Grimard, J.; Dewasme, L.; Wouwer, A. V. Dynamic model reduction and predictive control of hot-melt extrusion applied to drug manufacturing. *IEEE Trans. Control Syst. Technol.* **2021**, *29*, 2366−2378.

(10) Bhaskar, A.; Barros, F. N.; Singh, R. Development and implementation of an advanced model predictive control system into continuous pharmaceutical tablet compaction process. *Int. J. Pharm.* **2017**, *534*, 159−178.

(11) Mesbah, A.; Paulson, J. A.; Lakerveld, R.; Braatz, R. D. Model predictive control of an integrated continuous pharmaceutical manufacturing pilot plant. *Org. Process Res. Dev.* **2017**, *21*, 844−854.

(12) Huang, Y.-S.; Sheriff, M. Z.; Bachawala, S.; Gonzalez, M.; Nagy, Z. K.; Reklaitis, G. V. Evaluation of a combined MHE-NMPC approach to handle plant-model mismatch in a rotary tablet press. *Processes* **2021**, *9*, 1612.

(13) Öner, M.; Montes, F. C. C.; Ståhlberg, T.; Stocks, S. M.; Bajtner, J. E.; Sin, G. Comprehensive evaluation of a data driven

https://doi.org/10.1021/acsomega.3c02199
ACS Omega 2023, 8, 24441−24453

control strategy: Experimental application to a pharmaceutical crystallization process. *Chem. Eng. Res. Des.* **2020**, *163*, 248−261.
(14) Kirchengast, M.; Celikovic, S.; Rehrl, J.; Sacher, S.; Kruisz, J.; Khinast, J.; Horn, M. Ensuring tablet quality via model-based control of a continuous direct compaction process. *Int. J. Pharm.* **2019**, *567*, No. 118457.
(15) Mäki-Lohiluoma, E.; Säkkinen, N.; Palomäki, M.; Winberg, O.; Ta, H. X.; Heikkinen, T.; Kiljunen, E.; Kauppinen, A. Use of machine learning in prediction of granule particle size distribution and tablet tensile strength in commercial pharmaceutical manufacturing. *Int. J. Pharm.* **2021**, *609*, No. 121146.
(16) Tian, G.; Wei, Y.; Zhao, J.; Li, W.; Qu, H. Application of pulsed spray and moisture content control strategies on quality consistency control in fluidized bed granulation: a comparative study. *Powder Technol.* **2020**, *363*, 232−244.
(17) Deng, Z.; Jiang, Y.; Choi, K.-S.; Chung, F.-L.; Wang, S. Knowledge-leverage-based TSK fuzzy system modeling. *IEEE Trans. Neural Networks Learn. Syst.* **2013**, *24*, 1200−1212.
(18) Bahari, M.; Nejjar, I.; Alahi, A. Injecting knowledge in data-driven vehicle trajectory predictors. *Transp. Res. Part C: Emerging Technol.* **2021**, *128*, No. 103010.
(19) Han, H.; Liu, Z.; Liu, H.; Qiao, J. Knowledge-data-driven model predictive control for a class of nonlinear systems. *IEEE Trans. Syst., Man, Cybern.: Syst.* **2021**, *51*, 4492−4504.
(20) Han, H.-G.; Dong, L.-X.; Qiao, J.-F. Data-knowledge-driven diagnosis method for sludge bulking of wastewater treatment process. *J. Process Control* **2021**, *98*, 106−115.
(21) Zhao, H.; Zhao, J.; Zheng, Y.; Qiu, J.; Wen, F. A hybrid method for electric spring control based on data and knowledge integration. *IEEE Trans. Smart Grid* **2020**, *11*, 2303−2312.
(22) Luo, L.; Bao, S. Knowledge-data-integrated sparse modeling for batch process monitoring. *Chem. Eng. Sci.* **2018**, *189*, 221−232.
(23) Li, D.; Zhou, Y.; Hu, G.; Spanos, C. J. Identifying unseen faults for smart buildings by incorporating expert knowledge with data. *IEEE Trans. Autom. Sci. Eng.* **2019**, *16*, 1412−1425.
(24) Zeng, L.; Ge, Z. Pyramid Dynamic Bayesian Networks for Key Performance Indicator Prediction in Long Time-Delay Industrial Processes. *IEEE Trans. Artif. Intell.* **2023**, 1.
(25) Zeng, L.; Ge, Z. Bayesian network for dynamic variable structure learning and transfer modeling of probabilistic soft sensor. *J. Process Control* **2021**, *100*, 20−29.
(26) Amin, M. T.; Khan, F. Dynamic Process Safety Assessment Using Adaptive Bayesian Network with Loss Function. *Ind. Eng. Chem. Res.* **2022**, *61*, 16799−16814.
(27) Kumari, P.; Wang, Q.; Khan, F.; Kwon, J. S.-I. I. A Direct Transfer Entropy-Based Multiblock Bayesian Network for Root Cause Diagnosis of Process Faults. *Ind. Eng. Chem. Res.* **2022**, *61*, 16166−16178.
(28) Yan, H.; Wang, F.; He, D.; Zhao, L.; Wang, Q. Bayesian network-based modeling and operational adjustment of plantwide flotation industrial process. *Ind. Eng. Chem. Res.* **2020**, *59*, 2025−2035.
(29) Yan, H.; Wang, F.; He, D.; Wang, Q. An operational adjustment framework for a complex industrial process based on hybrid Bayesian network. *IEEE Trans. Autom. Sci. Eng.* **2020**, *17*, 1699−1710.
(30) Yan, H.; Wang, F.; Yan, G.; He, D. Hybrid approach integrating case-based reasoning and Bayesian network for operational adjustment in industrial flotation process. *J. Process Control* **2021**, *103*, 34−47.
(31) Guo, Y.; Chen, W.; Zhu, Y.-X.; Guo, Y.-Q. Research on the integrated system of case-based reasoning and Bayesian network. *ISA Trans.* **2019**, *90*, 213−225.
(32) Chergui, O.; Begdouri, A.; Groux-Leclet, D. Integrating a Bayesian semantic similarity approach into CBR for knowledge reuse in Community Question Answering. *Knowl.-Based Syst.* **2019**, *185*, No. 104919.
(33) Nikpour, H.; Aamodt, A. Fault diagnosis under uncertain situations within a bayesian knowledge-intensive cbr system. *Prog. Artif. Intell.* **2021**, *10*, 245−258.
(34) Yeguas, E.; Luzón, M. V.; Pavón, R.; Laza, R.; Arroyo, G.; Díaz, F. Automatic parameter tuning for evolutionary algorithms using a bayesian case-based reasoning system. *Appl. Soft Comput.* **2014**, *18*, 185−195.
(35) Kaedi, M.; Ghasem-Aghaee, N. Biasing Bayesian optimization algorithm using case based reasoning. *Knowl.-Based Syst.* **2011**, *24*, 1245−1253.
(36) Zhalehrajabi, E.; Lau, K. K.; KuShaari, K. Z.; Tay, W. H.; Hagemeier, T.; Idris, A. Modelling of urea aggregation efficiency via particle tracking velocimetry in fluidized bed granulation. *Chem. Eng. Sci.* **2020**, *223*, No. 115737.
(37) Wang, Z.; Pan, Z.; He, D.; Shi, J.; Sun, S.; Hou, Y. Simulation modeling of a pharmaceutical tablet manufacturing process via wet granulation. *Complexity* **2019**, *2019*, 1−16.
(38) Chen, G.; Ge, Z. Robust Bayesian networks for low-quality data modeling and process monitoring applications. *Control Eng. Practice* **2020**, *97*, No. 104344.
(39) Quesada, D.; Valverde, G.; Larrañaga, P.; Bielza, C. Long-term forecasting of multivariate time series in industrial furnaces with dynamic Gaussian Bayesian networks. *Eng. Appl. Artif. Intell.* **2021**, *103*, No. 104301.
(40) Chin, George, Jr.; Sutanay, Choudhury; Lars, Kangas; Sally, McFarlane; Andres, Marquez Evaluating in-clique and topological parallelism strategies for junction tree-based Bayesian network inference algorithm on the Cray XMT. In *IEEE International Symposium on Parallel and Distributed Processing Workshops and Phd Forum*; IEEE 2011: 1710−1719.

https://doi.org/10.1021/acsomega.3c02199
ACS Omega 2023, 8, 24441−24453