Biotechnology and Bioengineering

Online prediction for contamination of chlortetracycline fermentation based on Dezert–Smarandache theory

Jianwen Yang, Xiangguang Chen *, Huaiping Jin

School of Chemical Engineering and Environment, Beijing Institute of Technology, Beijing 100081, China

ABSTRACT

Fermentative production of chlortetracycline is a complex fed-batch bioprocess. It generally takes over 90 h for cultivation and is often contaminated by undesired microorganisms. Once the fermentation system is contaminated to certain extent, the product quality and yield will be seriously affected, leading to a substantial economic loss. Using information fusion based on the Dezer–Smarandache theory, self-recursive wavelet neural network and unscented kalman filter, a novel method for online prediction of contamination is developed. All state variables of culture process involving easy-to-measure and difficult-to-measure variables commonly obtained with soft-sensors present their contamination symptoms. By extracting and fusing latent information from the changing trend of each variable, integral and accurate prediction results for contamination can be achieved. This makes preventive and corrective measures be taken promptly. The field experimental results show that the method can be used to detect the contamination in time, reducing production loss and enhancing economic efficiency.

© 2015 The Chemical Industry and Engineering Society of China, and Chemical Industry Press. All rights reserved.

1. Introduction

Chlortetracycline (CTC) is an important antibiotic and a secondary metabolite of *Streptomyces aureofaciens*. It is characterized by bacterial inhibition, promotion of animal growth, high availability in animal feed, minimal residue in animal tissues, and low production cost. In recent decades, CTC has been the most consumed antibiotic in animal feed industry [1]. The most important considerations for biochemical industry are high yield, quality, and profit. Therefore, research efforts have often focused on in-depth physiological characteristics of cells to optimize industrial production. However, the risk of bacterial contamination is inevitable despite the maintenance of strict aseptic conditions during the production process. The biochemical plant we surveyed loses nearly 20 million CNY annually due to contamination-related issues. Therefore, the detection and prevention of bacterial contamination have been an active research area for the last two decades.

Contamination is defined by the migration of an undesired microorganism along with the desired microorganism, which affects the normal growth of the latter. These fast-growing, bacterial contaminants and phages soon outnumber culture strains, producing large amounts of byproducts, severely inhibiting the growth and metabolism of the culture strain of interest. Furthermore, a large proportion of nutrients, especially glucose, for supporting the growth and CTC production by

the culture strain, are diverted to the contaminants. Additionally, destruction of the culture strain will lead to disastrous consequences for the CTC plant.

Thus, early and accurate prediction of contamination of culture broth is of vital importance to biological fermentation and several methods are available for detecting or evaluating the contamination in laboratory or large-scale fermentation plant. These are broadly classified into physical and biological methods. The use of physical methods such as light, radiometry, and chromatography facilitates the rapid, precise, and non-invasive evaluation of broth [2–4]. However, their disadvantages lie in the cost of equipment, which are very expensive and require a high level of maintenance, and the tests are time-consuming and do not allow online application. On the other hand, biological methods, which exploit the genetic, immunological, and morphological characteristics of microorganisms [5–7], afford high accuracy, but the requirement of operator expertise and time-consuming procedures for prediction present significant drawbacks. In contrast, the soft-sensor prediction method that works on the principle of cause and effect reveals the intrinsic biological relation between measured and unmeasured states, and has been employed by several investigators [8,9]. This system generates data-driven black-box models on the basis of data from history of fermentation batches and captures underlying changes in process state, judging whether the broth is contaminated [10,11]. Several popular approaches such as principle component analysis, partial least squares, and clustering have been used as references in recent literature on process monitoring for detecting and diagnosing errors in the culture process [12–15]. Multisensor data fusion is widely applied in sensor networks,

robotics, video and image processing, and intelligent system design, combining information from several sources to form a unified picture. It introduces a novel approach to information processing and our work is an offshoot of this idea. Each variable state obtained from the CTC fermentation process is considered as a frame of discernment, namely source evidence, in which three elementary propositions are defined, which constitute seven focal elements with union operators [16–19]. Primitive contamination information is represented as a mass function of seven focal elements for each source, explored from the latent information, using the self-recursive wavelet neural network (SRWNN) calculation [20]. Next, by virtue of the Dezer–Smarandache theory (DSmT) methodology, all contamination information can be integrated into a comprehensive decision that facilitates necessary preventive and corrective measures.

In this work, we introduce the concept of DSmT, the principle of SRWNN and kalman filter (UKF filter) algorithms [21]. Specific procedure is presented to achieve online prediction of contamination in the culture process. The proposed method can be employed in industrial scale CTC plants.

2. Preliminaries

2.1. Dezert–Smarandache theory

The DSmT of plausible and paradoxical reasoning overcomes inherent limitations of the classical Dempster–Shafer theory (DST), which is based on the refutation of the principle of the third excluded middle, and is the generalization that DST can formally combine different information sources (rational, uncertain, or paradoxical). Owing to the vague, relative, and imprecise nature of the hyper-powerset D^Θ of the general frame of discernment Θ, the DSmT can solve these complex fusion problems where the DST or other methods often fail, especially when conflicts between sources become strong and the refinement of the frame of discernment Θ is inaccessible.

To understand the algorithm-based DSmT, three important concepts are introduced briefly, namely, hyper-powerset D^Θ, generalized basic belief mass, and proportional conflict redistribution rule.

The cornerstone of the DSmT is the concept of hyper-powerset D^Θ. In order to fuse information, one defines the frame of discernment Θ, representing a source set of n finite elements, $\Theta = \{\theta_1, \theta_2, ..., \theta_n\}$, where θ_i represents a concrete hypothesis, which is impossible to be defined and separated precisely. Here, D^Θ is considered as the set of all propositions from Θ with $\cap$ and $\cup$ operators, and these propositions must satisfy the following three conditions: (i) $\varnothing, \theta_1, \theta_2, ..., \theta_n \in D^\Theta$, (ii) if $A, B \in D^\Theta$, then $A \cap B \in D^\Theta$ and $A \cup B \in D^\Theta$, and (iii) no other elements belong to D^Θ, except those obtained by using rule 1 or 2.

The second concept is of generalized basic belief mass. For every evidential source Θ of the frame of discernment, mapping $m(\cdot)$: $D^\Theta \to [0, 1]$ associated to it is defined, which satisfies the following condition

$$m(\varnothing) = 0 \quad \text{and} \quad \sum_{A \in D^\Theta} m(A) = 1. \tag{1}$$

Considering the inherent nature of element θ_i, it is possible that the non-exclusive and non-refinement elements of Θ turn into a new, finer, exclusive frame of discernment. Quantity $m(A)$ represents the level of trust for proposition 'A' and the support to 'A' directly. Mapping $m(\cdot)$ is referred to as a generalized basic belief mass (gbbm).

The crux of the proposed method is the proportional conflict redistribution (PCR) rule. PCR can be applied to DST and DSmT framework dealing with the combination of belief functions and working for any degree of conflict under static or dynamical fusion situations. PCR rule redistributes the partial conflicting mass to the elements involved in the partial conflict, considering the conjunctive normal form of the partial conflict. PCR is considered as the most mathematically exact

redistribution of conflicting mass to non-empty sets following the logic of the conjunctive rule. PCR redistributes the conflicting mass only to the sets involved in the conflict and proportionally to their masses placed in the conflict. The general PCR formula for $s \geq 2$ sources is given by [16]. For $m_{PCR}(\varnothing) = 0$ and $\forall X \in G/\{\varnothing\}$

$$m_{PCR}(X) = m_{12...s} + \sum_{\substack{2 < t \leq s \\ 1 \leq r_1, ..., r_t \leq s \\ 1 \leq r_1 < r_2 < ... < r_{t-1} < (r_t = s)}} \sum_{\substack{X_{j_2}, ..., X_{j_t} \in G \; \{X\} \\ \{j_2, ..., j_t\} \in P^{t-1}(\{1, ..., n\}) \\ c(X \cap X_{j_2} \cap ... \cap X_{j_t}) = \varnothing \\ \{i_1, ..., i_s\} \in P^s(\{1, ..., s\})}} \tag{2}$$

$$\frac{\left[\prod_{k_1=1}^{r_1} m_{i_{k_1}}(X)^2 \right] \cdot \left[\prod_{k_i=r_{i-1}+1}^{t} m_{i_{k_i}}\left(X_{j_i}\right) \right]}{\left[\prod_{k_1=1}^{r_1} m_{i_{k_1}}(X) \right] + \left[\sum_{l=2}^{t} \prod_{k_i=r_{i-1}+1}^{t} m_{i_{k_i}}\left(X_{j_i}\right) \right]}$$

where G corresponds to a constrained hyper-power set D^Θ; i, j, k, r, s and t are all integers; $m_{12...s} \equiv m_\cap(X)$ corresponds to the conjunctive consensus on X between s sources, where all denominators are not equal to zero; the set of all subsets of k elements from $\{1, 2, ..., n\}$ (permutations of n elements taken by k) is denoted as $P^k(\{1, ..., n\})$ and the order of elements does not count; $c(X)$ is the canonical form (conjunctive normal form) of X.

2.2. SRWNN model for gbbm

The SRWNN model, which combines the attractor dynamic property of recurrent neural networks and good convergence performance of wavelet neural networks, can deal with time-varying input or output and shows good identification performance. In this section, taking the fermentation process into account, we briefly describe the application of SRWNN to achieve gbbm value for input variables of CTC fermentation process. Firstly, we assume that there are only three hypotheses in terms of each state variable available from the culture process, that is, $\Theta = \{\theta_1, \theta_2, \theta_3\}$, referred to as a frame of discernment. Next, we formulate hyper-powerset D^Θ by building it from the elements of Θ with operators $\cup$ and $\cap$. In order to decrease the complexity of calculation, we assume that D^Θ contains the following composite propositions: $X_1 = \theta_1, X_2 = \theta_2, X_3 = \theta_3, X_4 = \theta_1 \cup \theta_2, X_5 = \theta_1 \cup \theta_3, X_6 = \theta_2 \cup \theta_3$, and $X_7 = \theta_1 \cup \theta_2 \cup \theta_3$. Meanwhile, the focal elements from D^Θ satisfy the following constraint condition: $\sum_{X_i \in D^\Theta} m(X_i) = 1$, where the quantity $m(X_i)$ is a gbbm of X_i. Thus, the seven gbbm values, reduced for each source of evidence, can be computed based on SRWNN structure and least squared error-based learning algorithm. A schematic diagram of the SRWNN structure is shown in Fig. 1, where $N_2 =$ wavelets. The SRWNN structure consists of four layers: input layer, mother wavelet layer with a self-feedback loop, wavelet layer, and output layer. The details for formulation and calculation of SRWNN have been described previously [20].

2.3. UKF algorithm for smoothing

The collected signals generated from CTC fermentation process are susceptible to the environment for various reasons. Accurate and reliable results of desired contamination information rely on removal of noise from the sampled primitive signals.

Normally, the nonlinear discrete-time system considered is of the form

$$\begin{cases} \boldsymbol{x}(k+1) = f[\boldsymbol{x}(k)] + \boldsymbol{w}(k) \\ \boldsymbol{y}(k) = H[\boldsymbol{x}(k)] + \boldsymbol{v}(k) \end{cases} \tag{3}$$

where k denotes discrete time, $k \in N_0$ (N_0 denotes the set of natural numbers including zero), $\boldsymbol{x}(k) \in R^n$ is the state vector, and $\boldsymbol{y}(k) \in R^m$ is the measurement vector; the nonlinear mapping $f(\cdot)$ and $H(\cdot)$ are assumed to be continuously differentiable with respect to $\boldsymbol{x}(k)$; $\boldsymbol{v}(k)$
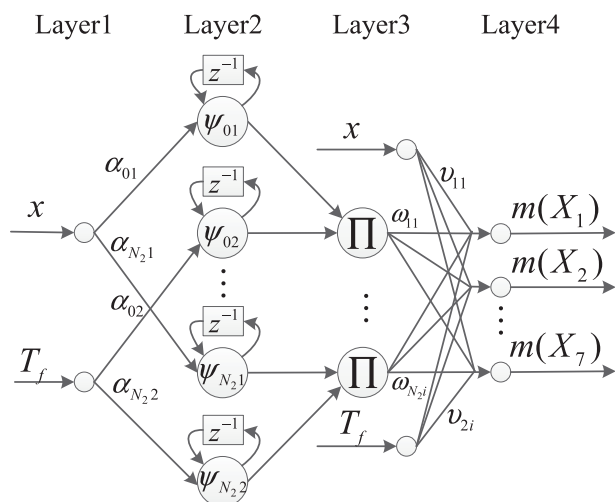
Layer1   Layer2   Layer3   Layer4

Fig. 1. Structure of SRWNN.

and $\boldsymbol{w}(k)$ are system state noise and output noise, respectively. Similar to Eq. (3), the UKF algorithm is considered as the most suitable filter algorithm for CTC fermentation process due to its ability to approximate nonlinear process and measurement model. The rationale behind the UKF algorithm, based on unscented transformation, is to use a minimal set of sample points to capture the true mean and covariance of nonlinear process, and then estimate the posterior mean and covariance with errors introduced in the second or higher orders when the set of sample points is run through the nonlinear system. The specific use of UKF algorithm as a nonlinear filter has been reported previously [21].

3. Procedures

Development of an estimator system to detect contamination during CTC fermentation is explained in this section. The available variables, including online and offline, of the CTC fermentation process are as follows: temperature (TM), dissolved oxygen (DO), agitator current intensity (CI), ammonia accumulation (AA), glucose accumulation (GA), liquid volume (LV), air flow accumulation (AF), carbon dioxide concentration in exhaust (CO), fermentation time (T_f), amino nitrogen concentration (AC), viscosity of culture broth (VS), titer of CTC (TI), and glucose concentration (GC). The essence of the proposed method is to combine all contamination information obtained from online sensors and soft-sensors into an accurate decision. A schematic diagram for realizing this method is illustrated in Fig. 2. The following sections describe in detail the procedure to preprocess historical data, establish the SRWNN model, detect unmeasured but important variables with a soft-sensor, fuse all contamination information with DSmT, and take the required decision.

3.1. Data preprocessing

The intensive, data-driven nature of the proposed method requires a sufficient amount of data. The preliminary work, which includes gathering, arrangement, and normalization of data, is crucial for building a robust and accurate model. Firstly, the selected batch data must be complete, without missing any key state variables, and the duration range should cover the entire fermentation process, especially for the normal process data set. To facilitate subsequent application, all batch data sets are normalized to the 0–1 range and classified into three groups, *i.e.*, normal process data set, *Bacillus* infection data set, and phage infection data set (the first part of Fig. 2). Meanwhile, all data in the three groups are filtered using the UKF algorithm. Finally, 150 batches of typical data set, which covered the whole year, are deliberately selected

from the preprocessed data set to constitute the training set of 120 batches and the test set of 30 batches.

3.2. Measuring difficult-to-measure variables with soft-sensor

Invariably, there are some key measurements and quantities reflecting cellular metabolism, energy transaction, microbial growth, product yield, and so on that cannot be simply measured online by an instrument, due to the unavailability of the instrument, high cost of hardware sensors and their maintenance, or the reliability of the sensors. An alternative is to measure them by laboratory analysis, by sampling the culture broth during the fermentation process. However, this procedure is time-consuming and arduous, which in turn increases the cost of production, high off-spec products, and risk to environment safety. To solve the problem, the alternative is to use a soft-sensor that can measure and predict important variables difficult to measure physically. Using these variables measured with soft-sensor, four evidence sources, indirectly reflecting the contamination of CTC fermentation, can be obtained. Based on the proposed method, four soft-sensor units have been built to measure the viscosity of culture broth, titer of CTC, amino nitrogen concentration, and glucose concentration. To enhance the reliability of the soft-sensor, correction units have been developed for the four variables, so that the model parameters showing a difference between the value from laboratory analysis and that from the soft-sensor can be corrected [22]. Fig. 3 depicts the working principle of the four variables with soft-sensor, in which inputs of the model include known, online, continuous variables. The results generated from soft-sensors are real-time and change during the CTC fermentation process.

3.3. Fusing information based on DSmT

Whether the CTC fermentation process is contaminated cannot be read directly by measuring the culture broth online. However, some information indicating that broth state may contain *Bacillus* or phages can be obtained by comparing the current trend of process variables with the normal control at the same time point. During the culture process, different state variables have varying susceptibility or response time for the same source of contamination and this is an inherent feature of CTC fermentation characteristics. It is recommended to combine all available information from state variables to capture accurate and comprehensive information. In terms of fusing uncertain, imprecise, and conflicting information, the DSmT methodology is more advantageous than DST. For the convenience of combining contamination information, each state variable is considered to have the same frame of discernment, where three hypothesis elements exist within a frame. We then define a mapping set associated with each source of evidence and construct the gbbm as follows: $m(\theta_1)$ is defined as the gbbm for non-contamination, denoted by $m(X_1)$; $m(\theta_2)$ is defined as the gbbm for definite *Bacillus* contamination, denoted by $m(X_2)$; $m(\theta_3)$ is defined as the gbbm for definite phage infection, denoted by $m(X_3)$; $m(\theta_1 \cup \theta_2)$ is defined as the gbbm for probable *Bacillus* contamination, denoted by $m(X_4)$; $m(\theta_1 \cup \theta_3)$ is defined as the gbbm for probable phages infection, denoted by $m(X_5)$; $m(\theta_2 \cup \theta_3)$ is defined as the gbbm for probable *Bacillus* contamination or phages infection, denoted by $m(X_6)$; and $m(\theta_1 \cup \theta_2 \cup \theta_3)$ is defined as the gbbm for probable *Bacillus* contamination and phage infection, denoted by $m(X_7)$. Based on the SRWNN method, the model structure relating each input process variable to the seven gbbms is formulated. The overall procedure adapts a methodology of trial and error, not stopping the tests and continually revising until the performance requirement of the application is met (see the second part of Fig. 2).

In terms of the CTC fermentation process, 12 process state variables exist in all, which consists of the online and offline variables, each of which is considered as a body of evidence for contamination. To find some symptom of contamination from those process state quantities,
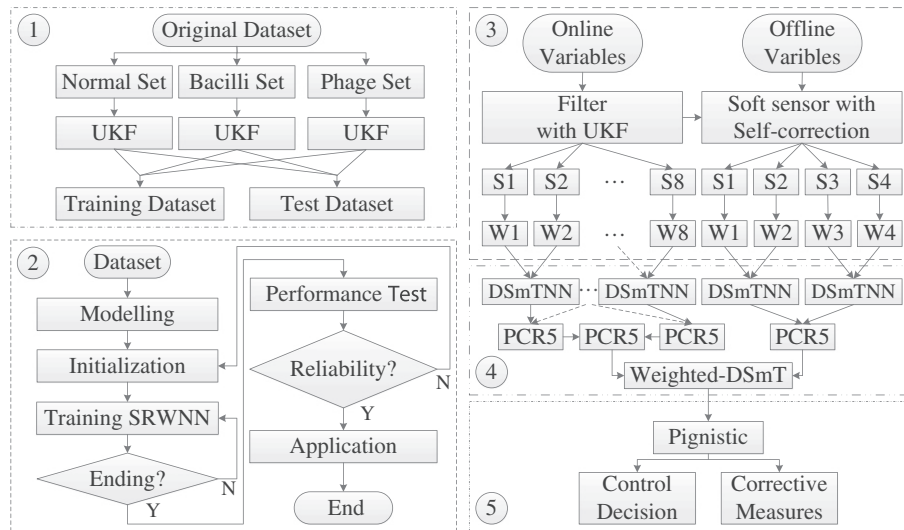
Fig. 2. Implement flowchart of the proposed method.

it is necessary to combine information from these 12 diverse variables, making a final judgment or decision for the next specific steps. The illustration of information fusion architecture is shown in the third part of Fig. 2.

Finally, the information for general basic belief mass of every process variable generated from SRWNN is combined and normalized. The input value of PCR begins to be interpreted. A schematic diagram of the DSmT neural network structure (DSmTNN), given in Fig. 4, has two evidence sources, 14 inputs, and seven normalized outputs, and is composed of four layers.

Layer 1 is the input layer: each input node corresponds to the gbbm of a focal element of a single source. The input layer accepts the gbbm values and transmits them to layer 2.

Layer 2 is the multiplication layer: each node performs multiplication for the two incoming masses from layer 1. For instance, node k has two input masses, $m_1^0(X_i)$ and $m_2^0(X_j)$, and it produces its output, $M_k = m_1^0(X_i) \times m_2^0(X_j)$, so the number of multiplications in layer 3 is $i \times j$.

Layer 3 is the summation layer: it consists of seven sum nodes, each adding its respective incoming mass from layer 2. Each node output corresponds to the non-normalized combined mass as

$$m_{12}(X) = \sum_{X_i \cap X_j = X} \alpha M_k \tag{4}$$

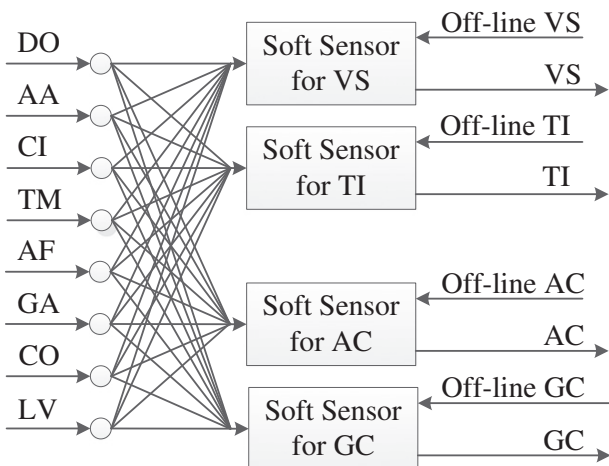where $\alpha = 1$ if $X_i \cap X_j = X$, else $\alpha = 0$.

Layer 4 is the normalization layer: in DSmT, if $\mathrm{Card}(\Theta) = |\Theta| = 3$, then $|D^\Theta| = 19$. Obviously, this size is too big to establish a moderate model structure and compute the fusion information accurately. To simplify the model structure and decrease the operational burden, we assume that all focal elements in this study are void except the seven combination relationships between the three propositions in the frame of discernment. Consequently, in the fourth layer, a normalization method is used to bring the values that come from the summation layer, lying outside the boundary, within the range of 0 to 1. The output of the mth node is the following ratio, which denotes the normalized combined mass

$$m_{12}^1(X_i) = \frac{m_{12}(X_i)}{\sum\limits_{i=1}^{7} m_{12}(X_i)}. \tag{5}$$

Per part 4 of Fig. 2, performing the initial fusion of output information from part 3 of Fig. 2 by means of DSmTNN results in a six-group

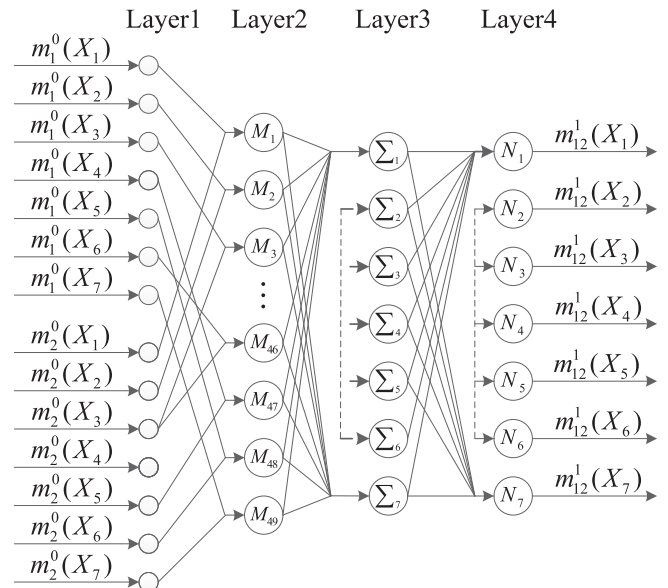**Fig. 3.** Topological structure of soft-sensor.

Fig. 4. Cascade DSmTNN model for information fusion.

primary result: {VS, TI}, {AC, GC}, {DO, AA}, {CI, TM}, {GA, AF}, and {CO, LV}. Among these six data pairs, the latter two are obtained from soft-sensor data, while the remaining four are real-time data pairs. Each pair result would require further information fusion, in the form of two new evidences. Thus the result is more accurate and reliable than the initial outcome.

Sometimes, the primary result may present some conflict if it is determined solely from the gbbm values. In view of this challenge, PCR is considered the best combination rule to synthesize relative information, which then transfers (total and partial) conflicting mass to non-empty sets involved in the conflicts proportional to the mass assigned to them by the source. PCR includes five versions, PCR1 through PCR5, with increasing complexity of rules and precision of redistributing conflicting masses.

The method to combine the primary fusion information using PCR5 is described in brief. First, the conjunctive rule as follows is applied: $\{m_{12}(X_1), ..., m_{12}(X_7)\}$.

$$m_{\cap}(X) = \sum_{\substack{X_i, X_j \in D^{\theta} \\ X_i \cap X_j = X}} m_1(X_i) m_2\left(X_j\right) \tag{6}$$

Next, with the following PCR5 formula, the set of $\{m_{PCR5}(X_1), ..., m_{PCR5}(X_7)\}$ may be obtained.

$$m_{PCR5}(X_i) = m_{12}(X_i) + \sum_{X_j \in G \setminus \{X_i\}} \left[\frac{m_1(X_i)^2 m_2\left(X_j\right)}{m_1(X_i) + m_2\left(X_j\right)} + \frac{m_2(X_i)^2 m_1\left(X_j\right)}{m_2(X_i) + m_1\left(X_j\right)} \right] \tag{7}$$

In the proposed method, applying the PCR5 rule repeatedly may yield the combination information of both online and offline data. Owing to the difference between the reliability of the two data types by virtue of the nature of the CTC fermentation plant, the reliability weightage should be added to the fusion information before making the final decision. Therefore, the finial fusion result between the online and offline data can be obtained as follows.

$$m(X) = \sum_{\substack{X_i, X_j \in D^{\theta} \\ X_i \cap X_j = X}} \left[m^{on}(X_i)\right]^{\alpha} \left[m^{off}(X_i)\right]^{\beta} \tag{8}$$

where α and β are all statistical determinants, satisfying $\alpha + \beta = 1$, while in general $\alpha < \beta$.

3.4. Making the final decision

Once the previous steps in the information combination procedure are complete, obtaining six pairs of initial combination based on the DSmTNN, four groups of secondary combination base using PCR5, and a final weighted combination, an algorithm are chosen to make the final decision in the form of probability. The Pignistic algorithm is selected for this purpose, representing intuitive and operable properties of probability as follows:

$$\forall A \in D^{\theta}, P\{A\} = \sum_{X \in D^{\theta}} \frac{|A \cap B|}{|X|} m(X) \tag{9}$$

where $|X|$ is the cardinal of proposition X in the DSm model. With the improvement via Eq. (10), each gbbm is calculated according to the sequence $\{\theta_1, \theta_2, \theta_3\}$ to obtain $\{\dot{P}\{\theta_1\}, \dot{P}\{\theta_2\}, \dot{P}\{\theta_3\}\}$.

$$\begin{cases} \dot{P}\{\theta_1\} = m(\theta_1) + 1/2m(\theta_1 \cup \theta_2) + 1/2m(\theta_1 \cup \theta_3) + 1/3m(\theta_1 \cup \theta_2 \cup \theta_3) \\ \dot{P}\{\theta_2\} = m(\theta_2) + 1/2m(\theta_1 \cup \theta_2) + 1/2m(\theta_2 \cup \theta_3) + 1/3m(\theta_1 \cup \theta_2 \cup \theta_3) \\ \dot{P}\{\theta_3\} = m(\theta_3) + 1/2m(\theta_2 \cup \theta_3) + 1/2m(\theta_1 \cup \theta_3) + 1/3m(\theta_1 \cup \theta_2 \cup \theta_3) \end{cases} \tag{10}$$

Then, the normalized set $\{P\{\theta_1\}, P\{\theta_2\}, P\{\theta_3\}\}$ is computed as follows.

$$P\{\theta_i\} = \frac{\dot{P}\{\theta_i\}}{\sum_{i=1}^{3} \dot{P}\{\theta_i\}} \tag{11}$$

From an application point of view, the set of $\{P\{\theta_1\}, P\{\theta_2\}, P\{\theta_3\}\}$ provides the contamination information of the CTC fermentation process expressed as a probability, as described in part 5, Fig. 2. Thus, if a key constraint or condition goes beyond the predefined limitation of contamination, the system will warn the operator. Based on the experienced worker's judgment using the information, preventive and corrective measures can be taken in time.

3.5. Results and discussion

In the CTC fermentation process, several challenges, such as mechanical failures, process disruption, operational or instrument errors, contribute to data records of contamination. A breach of aseptic conditions in any part of the operation would expose the fermentation to a high risk of infection by undesired microorganism. Before discussing the eventuality of contamination, we first identify the issues that can result in infection using a flowchart of CTC culture procedure, illustrated in Fig. 5. The flowchart contains four phases including strain preparation, primary seed amplification, secondary seed amplification, and final fermentation process. During these steps, the likelihood of introducing harmful microorganisms into production process involves both situations and equipment, namely, leakage in the pipes carrying sterile air, agitator malfunction, failure of gasket or o-ring valves, deviation from vessel and media sterilization procedures, contamination during initial or mid-cycle inoculations, contamination during tank-to-tank transfers, contamination during offline collection of broth at the sampling port, failure of exhaust outlet fan, and contamination of water, air, defoamer, and so on. These causes of contamination in the CTC fermentation plant have been determined based on the data obtained from actual plant operations, but several other undetectable phenomena may occur as well.

The proposed scheme based on DSmT was carried out experimentally to predict contamination occurrence online and make corrective measures in a 130 m^3 fermenter of the CTC plant. We divide the experimental results into three types of situations as those of microbial infection. Figs. 6–8 present three curves for the normal state, denoted as 'Normal', *Bacillus* infection, denoted as 'Bacilli', and phage infection, denoted as 'Phage', where the horizontal axis denotes the duration of culture process while the vertical axis denotes the probability percentage. In the following section, we will elaborate on the typical distribution for the three trajectories.

The normal fed-batch process plot shows that the culture process is either contaminated or the extent of contamination within the predefined, lowest threshold of detection. In Fig. 6, 'Normal' trajectory is over 60% though it gets close to 50% at approximately 37 h, but lasts for less than 3 h for a single batch. The other two representative trajectories fluctuate independently by approximately 20% from the beginning to end. Thus the culture process can be considered as a normal batch only if the percentage of the 'Normal' curve is larger than 50%.

Fig. 7 illustrates the probability variation of the three curves once the fermentation process is contaminated, i.e., the broth is infected by a *Bacillus* species, which is the most common and infectious contaminant of the CTC fermentation. Therefore, 'Bacilli' is considered as a representative contaminant, outside of the phages.

In Fig. 7, the average probability of 'Bacilli' trajectory exceeds the lowest threshold by approximately 40% at about 40 h, which is predefined as the determination condition of contamination, while the trajectory of 'Normal' is less than 50% and that of 'Phage' is at 15%. This result indicates that the overgrowth of *Bacillus* would threaten the fermentation strain so that necessary measures should be taken to inhibit the growth of contaminant. If this process is unchecked, large amounts of glucose and other
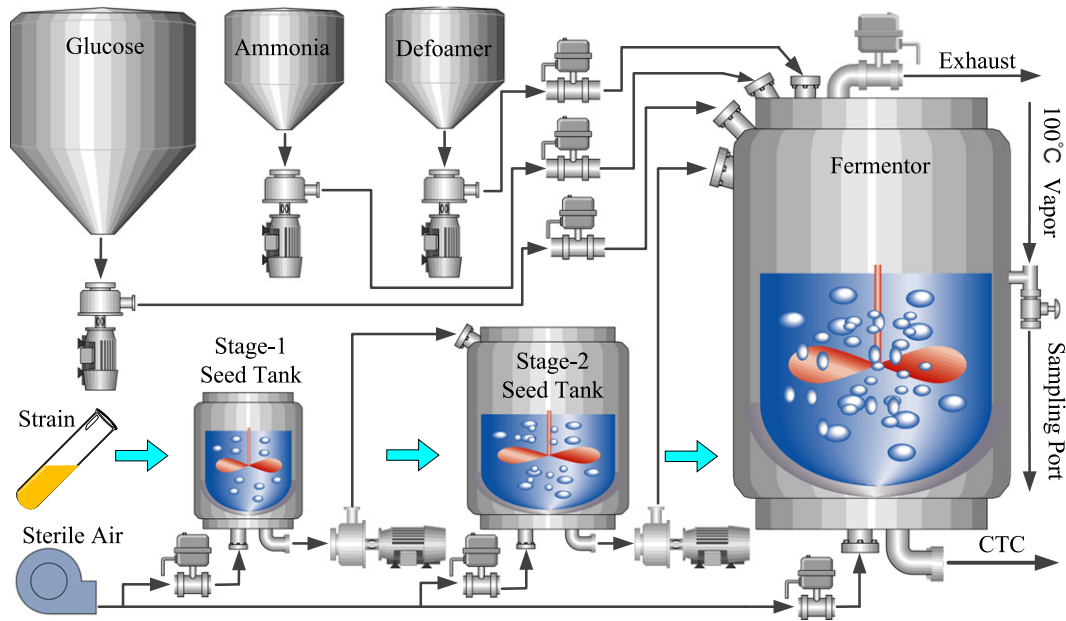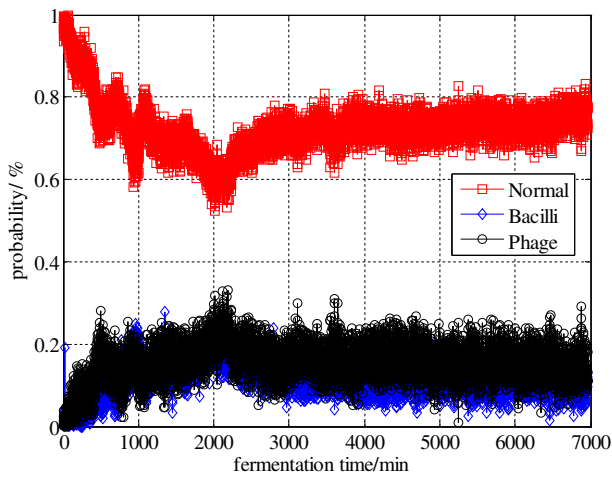
Fig. 5. CTC fermentation process phages.

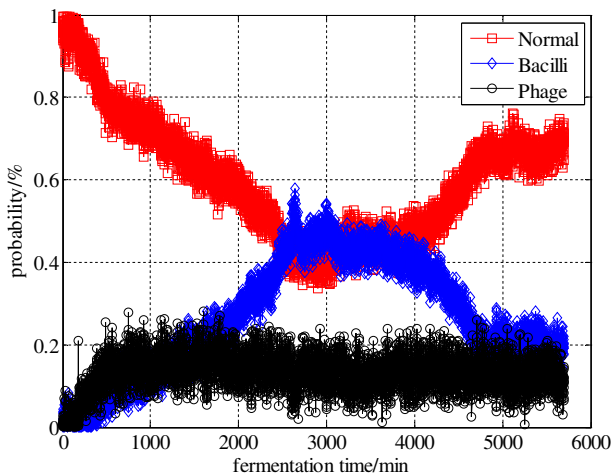**Fig. 6.** Probability variation of a normal batch.

nutrients in the fermenter will be converted into lactic acid, acetic acid, or other undesired products, resulting in the death of fermentation strain due to unfavorable conditions and the CTC fermentation product could be degraded by contaminant-derived enzymes. Discharging the broth early is the only viable option to protect the system, which results in a large amount of wastage and increases production cost. If such a situation can be predicted by the proposed method, the following measures can control the damage to some extent: (1) lower or raise the temperature outside the range of 20 °C to 40 °C, (2) lower the pH outside the range of 5 to 8, (3) decrease the feed rate, (4) reduce airflow to lower the DO level, and (5) add specific antibiotics into fermenter. Advanced and accurate prediction will make corrective action be taken in time and better control of *Bacillus* contamination will be achieved. From previous statistics of recorded *Bacillus* infection, it is found that the beginning of CTC fermentation is the most vulnerable stage for contamination. Since CTC itself is an antibiotic, which can inhibit contaminant growth, adding a moderate amount of CTC into the substrate is a preventive measure against *Bacillus* contamination.
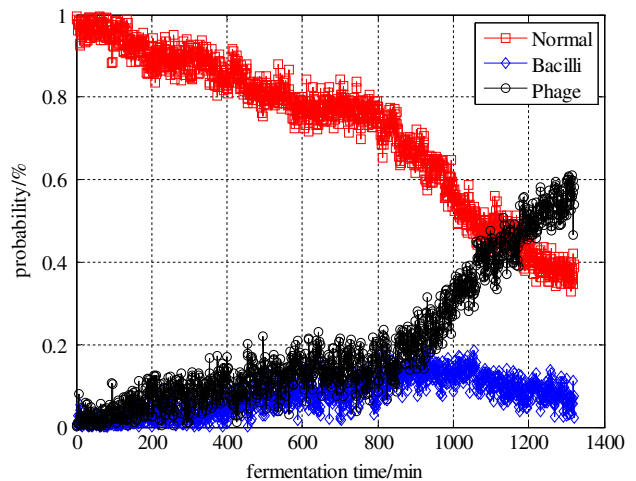

Fig. 7. Probability variation of a Bacillus-infected batch.

Fig. 8. Probability variation of a phages-infected batch.

Finally, infection of the CTC fermentation process by phages (bacteriophages) is illustrated in Fig. 8. A phage is a kind of virus that can infect and replicate within bacteria and is widely distributed in areas populated by bacterial hosts. 'Phage' is also a general name for microbes and viruses causing infection of bacteria, fungi, actinomycetes, spirochetes, etc. If the culture process is infected by phages, common symptoms include a slow or sluggish pace of growth with the fermentation time and the final product is reduced to a great extent. Other indicators include reduced carbon dioxide in the exhaust, gradually lighter substrate, progressively increasing pH, dark color of broth instead of normal yellowish-brown, and large amounts of residual glucose. Fig. 8 describes a phage-infected batch, where probability of 'Phage' rapidly increases to 60% from approximately 15–26 h of the culture process. Similar to identifying a *Bacillus* infection, the average probability of 'Phage' trajectory exceeding by 30% is considered as infection by phages, where a 30% probability is predefined as the lowest threshold and the trajectory of 'Normal' is also less than 60%.

Compared with a *Bacillus* infection, the control and preventive measures for phage infection are much more challenging, once the culture is infected by a virulent phage. Symptoms of a low-grade phage infection are often invisible, but the concentration of glucose will begin to rise with time keeping the feed rate constant, until a majority of cells are lysed within hours of infection resulting in the fermentation stopping. This phenomenon relies on several factors including the type of phages, the stage of fermentation occurring infection, the quantum of phages in contrast with that of its host, the substrate components, as well as the physical and chemical environment in the fermenter. Thus infections with the same phage may show diverse symptoms.

In actual operation, the tendency of phage infection can be estimated through experience or with the aid of software, such as the proposed method in this study. By analyzing previous records of phage infections, we can draw the following conclusions: infections occurring in the seed culture phase may spread to all production fermenters, infections in the early phase of culture can make industry-scale fermentation process difficult, but if the infection occurs towards the end of the culture process, it generally does not exhibit any obvious symptoms. Methods for preventing and attenuating the harm caused by phage infection are still a subject for intensive research. Routine methods such as addition of chelating agents, non-ionic detergents, and antibiotics may abate phage propagation, but when an infection occurs in a fed-batch due to poor equipment, the best choice is to stop the infected batch, discard all contaminated material, and conduct a thorough cleaning and sterilization of equipment. Such actions must be undertaken even at the expense of closure of the entire plant.

In some CTC fermentation plants, whether the culture process is contaminated previously depends on the judgment, with the help from two methods: microscopy and bacterial culture, both carried out offline, which is not only time-consuming and laborious, but also increases the risk of contamination accidents. The proposed method as an alternative strategy can predict the real-time state of contamination based on process data, thus overcoming the challenges of manual, offline tests. The prediction performance for *Bacillus* and phage contaminants is listed in Table 1. We can see that the prediction performance for phage infection is clearly superior to that for *Bacillus* infection. This can be attributed to that phage infection itself displays characteristics of destructive force: short latent period and large burst scale, so the proposed method can easily capture the marked changes, while *Bacillus* infection is characterized by mild reaction and a gradual process.

Table 1
Prediction performance for contamination

Term	Accuracy rate/%	False alarms/%	Missing alarms/%
Bacillus	61	11	28
Phages	84	9	17

Note: all data are collected from two plants for five month in 2013.

Since the occurrence of contamination is impossible to eradicate completely, one hopes to minimize the impact of contamination on the yield and quality of products as well as the cost of production in the CTC fermentation process as a more viable solution. The controlled culture can be modified adequately based on the proposed method. Using the results of prediction, several corrective measures are possible depending on the time of infection. If the infection occurs at the beginning, corrective measures may be taken to curb the extent of contamination, such as decreasing temperature, pH, feed rate, airflow rate or agitator rate, adding a moderate amount of antibiotic or similar sterilizing agent to the culture broth, or sterilizing the medium and re-inoculating the current batch. If the infection occurs in the steady stage, the above-mentioned measures as well as altering the culture environment and adding an antibacterial agent can be attempted to keep the culture running, but the broth should be discharged earlier, once the above measures become invalid. If the infection occurs towards the end of the culture, nothing can be done, but observe the trend of infection.

Compared to the performance of control, based on the conventional justification for contamination, the proposed method shows an improved prediction capability with enhanced economic benefit, by virtue of online predictions. Table 2 shows the results of comparison between the proposed and conventional methods, which are of average value.

Table 2
Comparison between new and conventional method

Term	Prediction time for *Bacillus*/h	Prediction time for phages/h	CTC yield/μg·ml^{-1}	Discarded batch/%
New method[1]	31.5	23.1	22,673	4.74
Con. method[2]	41.3	29.7	21,354	6.13

Note: all data are collected from two plants for five months in 2013.
[1] New method is the proposed method in the context.
[2] Con. method is the conventional method in the context.

Several aspects of the proposed method using information fusion based on DSmT knowledge still have scope for improvement, for example, additional case records would increase the accuracy and stability of the system.

4. Conclusions

Formulation of a mechanistic model for CTC fermentation is challenging, due to its intrinsic, nonlinear nature and time-dependent variability. At the same time, limited online methods are available to monitor contamination of culture process. Subsequently, infection control in CTC fermentation is equally challenging as that in large-scale plants. With the help of information fusion, based on DSmT, the tendency of contamination during CTC fermentation process can be predicted accurately in time, indirectly utilizing both measured and unmeasurable variables. Applied to the actual plant, results show that the proposed method can reduce the risk of infection, maintenance costs, as well as labor required, while improving the yield, quality, and economic efficiency.

References

[1] Z. Hošťálek, Relationship between the carbohydrate metabolism of *Streptomyces aureofaciens* and the biosynthesis of chlortetracycline, *Folia Microbiol.* 9 (2) (1964) 78–88.
[2] A.O. Onadipe, Prediction of microbial contamination in cell culture using a laser scanning system, *BioPharm* 14 (2001) 38–40.
[3] L. Jimenez, Light up the life in microbial contamination, *Pharm. Formul. Qual.* (August/September 2001) 54–55.
[4] I. Elmroth, A. Fox, O. Holst, L. Larsson, Prediction of bacterial contamination in cultures of eukaryotic cell by gas chromatography–mass spectrometry, *Biotechnol. Bioeng.* 42 (1993) 421–429.
[5] B.A. Plantz, J. Anderen, L.A. Smith, M.M. Meagher, V.L. Schlegel, Prediction of non-host viable contamination in *Pichia pastoris* culture and fermentation broth, *J. Ind. Microbiol. Biotechnol.* 30 (2003) 643–650.

[6] J.R. Swartz, N. McFarland, Genetic approaches to the prediction of contaminants in *Escherichia coli* fermentations, *Biotechnol. Prog.* 14 (1998) 88–91.

[7] R. Van Beurden, R. Mackintosh, New developments in rapid microbiology using immunoassays, *Food Agric. Immunol.* 6 (1994) 209–214.

[8] O. Xu, Y. Fu, H. Su, *et al.*, A selective moving window partial least squares method and its application in process modeling, *Chin. J. Chem. Eng.* 22 (7) (2014) 799–804.

[9] C.Q. Cheryl, H. Juergen, Process monitoring and parameter estimation via unscented kalman filtering, *J. Loss Prev. Process Ind.* 22 (2009) 703–709.

[10] G. Yang, X. Li, Y. Qian, A real-time updated model predictive control strategy for batch processes based on state estimation, *Chin. J. Chem. Eng.* 22 (3) (2014) 318–329.

[11] J. Wang, L. Zhao, T. YU, On-line estimation in fed-batch fermentation process using state space model and unscented kalman filter, *Chin. J. Chem. Eng.* 18 (2) (2010) 258–264.

[12] Y.W. Zhang, Z.Y. Hu, On-line batch process monitoring using hierarchical kernel partial least squares, *Chem. Eng. Res. Des.* 89 (2011) 2078–2084.

[13] W.W. Dong, Y. Yao, F.R. Gao, Phase analysis and identification method for multiple batch process with partitioning multi-way principal component, *Chin. J. Chem. Eng.* 20 (6) (2012) 1121–1127.

[14] J. Mori, J. Yu, Quality relevant nonlinear batch process performance monitoring using a kernel based multiway non-gaussian latent subspace, *J. Process Control* 24 (2014) 57–71.

[15] X. Zhao, W.W. Yan, H.H. Shao, Nonlinear statistical process monitoring and fault diagnosis based on kernel fisher discriminant analysis, *J. Chem. Ind. Eng.* 58 (4) (2007) 951–956 (in Chinese).

[16] X.D. Li, X.H. Huang, J. Dezert, L. Duan, M. Wang, A successful application of DSmT in sonar grid map building and comparison with DST-based approach, *Int. J. Innov. Comput. Inf. Control* 3 (3) (2007) 539–549.

[17] A. Tchamova, T. Semerdjive, J. Dezert, Estimation of target behavior tendencies using DSmT, Prco. Of Fusion 2003, Cairns, Australia, July 8–11, 2003.

[18] S. Corgne, L. Hubert-Moy, J. Dezert, J. Mercier, Land cover change prediction with a new theory of plausible and paradoxical reasoning, Proc. Of Fusion 2003 Conf., Cairns, Australiz, July 8-11, 2003.

[19] F. Smarandache, J. Dezert, Information fusion based on new proportional conflict redistribution rules, Proceedings of Fusion 2005 Conf., Philadelphia, July 26–29, 2005.

[20] J.Y. Sung, H.C. Yoon, B.P. Jin, Generalized Predictive Control Based on Self-recurrent Wavelet Neural Network for Stable Path Tracking of Mobile Robots: Adaptive Learning Rates Approach, IEEE Transactions on Circuits and System—I: Regular Papers, 53(6), 2006, pp. 1381–1394.

[21] K. Xiong, H.Y. Zhang, C.W. Chan, Performance evaluation of UKF-based nonlinear filtering, *Automatica* 42 (2006) 261–270.

[22] J.W. Yang, X.G. Chen, H.P. Jin, L. WU, Study on the industrial chlortetracycline fermentation feed rate adjustment method based on soft sensor, *Chin. J. Sci. Instrum.* 35 (2) (2014) 468–474.