Journal of Hazardous Materials 185 (2011) 262–268

Contents lists available at ScienceDirect

Journal of Hazardous Materials

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

Fault detection using dynamic time warping (DTW) algorithm and discriminant analysis for swine wastewater treatment

B.H. Jun*

School of Fire and Disaster Prevention, Kangwon National University, 1 Joongang-ro, Samcheok 245-711, South Korea

A R T I C L E   I N F O

Article history:
Received 26 May 2010
Received in revised form 6 September 2010
Accepted 7 September 2010
Available online 17 September 2010

Key words:
Sequencing batch reactor
Fault detection
Swine wastewater treatment
Dynamic time warping
Discriminant analysis

A B S T R A C T

This paper proposes a diagnosis system using dynamic time warping (DTW) and discriminant analysis with oxidation–reduction potential (ORP) and dissolved oxygen (DO) values for swine wastewater treatment. A full-scale sequencing batch reactor (SBR), which has an effective volume of 20 m^3, was auto-controlled, and the reaction phase was performed by a sub-cycle operation consisting of a repeated short cycle of the anoxic–aerobic step. Using ORP and DO profiles, SBR status was divided into four categories of normal and abnormal cases; these were influent disturbance, aeration controller fault, instrument trouble and inadequate raw wastewater feeding. Through the DTW process, difference values (D) were determined and classified into seven cases. In spite of the misclassification of high loading rates, the ORP profile provided good diagnosis results. However, the DO profiles detected five misclassifications that indicated different statuses. After the DTW process, several statistical values, including maximum value, minimum value, average value, standard deviation value and three quartile values, were extracted and applied to establish the discriminant function. The discriminant analysis allows one to classify seven cases with a percentage of 100% and 92.7% for ORP and DO profiles, respectively. Consequently, the study showed that ORP profiles are more efficient than DO profiles as diagnosis parameters and DTW diagnosis algorithms and discriminants.

1. Introduction

Sequencing batch reactors (SBRs) are simple to operate, have low space requirements, and are appropriate for dynamic loading rates and biological nutrient removal [1]. They have become very popular for treating swine wastewater, which has high amounts of organics, nutrients, fibre and minerals when compared to domestic sewage. Control strategies for sequential on–off aeration are very important to SBR operation, and many control applications can be used to detect the endpoint of the nitrification and denitrification reactions as well as the shortening of the reaction stages based on oxidation–reduction potential (ORP), pH and dissolved oxygen (DO) measurements [2–4].

Although control techniques have resolved many problems, when an uncommon fault occurs in the process, an SBR cannot operate under control. Therefore, the development of an automated fault detection method is of significant practical value to the effectiveness and robustness of wastewater treatment processes.

Due to the complexity and the time-varying behaviour of biological reactions, any process management must be insensitive to time-varying fluctuations and unexpected parameters. If an accurate process model is not available, then fault detection and diagnosis can be applied using a pattern recognition approach [5]. According to this approach, historical data from normal operation and past faults must be collected. One can extract the fault cases from normal operation as separate classes to which the patterns belong. These training data must then be processed so that important information is derived; these procedures are known as feature extraction.

In this paper, we develop a diagnosis method of deterministic faults in the SBR process using dynamic time warping (DTW) and discriminant analysis (DA). DTW is a general time alignment and similarity measurement for two temporal sequences that was introduced by Bellman [6]. DTW is a flexible pattern recognition method that can appropriately translate, compress, and expand patterns so that when the magnitude is invariant, similar features are matched. The DTW consists of two steps. In the first step, a set of dynamic patterns of known past faults is collected. The patterns are then scaled to remove the magnitude information, which enables the diagnosis scheme to operate independently of the magnitude of the fault. When the pattern of an unknown fault is obtained, the same scaling procedure is applied and the scaled pattern is compared with all reference patterns. DTW has been extended to handle speech recognition [7,8] and connected word recognition [9,10]. More recent research on DTW has focused on applying it to mining patterns from batch reactors in the industrial field [11] and

* Tel.: +82 33 570 6432; fax: +82 33 570 6819.
E-mail address: bhjun@kangwon.ac.kr.

0304-3894/$ – see front matter © 2010 Elsevier B.V. All rights reserved.
doi:10.1016/j.jhazmat.2010.09.027

Table 1
Characteristics of raw swine wastewater collected from scraper-type barns (unit: mg/l).

SCOD	BOD_5	TSS	NH_4^+-N	T-P	Alkalinity	pH
11,000 ± 2000	8000 ± 1500	1500 ± 100	3800 ± 400	20 ± 5	13,000 ± 2000	8.7 ± 0.2

Table 2
Fault cases in SBR operation.

Fault cases		Fault number
Influent loading rate	Normal	F1
	High	F2
	Extremely high	F3
Aeration control		F4
Instrument	Raw water feeding	F5
	Carbon feeding	F6
Raw water (slurry)		F7

in bioprocess [12] and other spectroscopic fields [13]. Multivariate statistical data analysis methods, such as discriminant analysis and principal component analysis (PCA), were used to classify the data. These techniques have proven to be useful for environmental, chemical and biological case studies [14,15].

This study develops a knowledge-based diagnosis system using DTW and discriminant analysis with ORP and DO values as variables for swine wastewater treatment. ORP and DO values are the most widely used parameters for SBR control, and most wastewater treatment plants have sensors that can detect these values. Therefore, our proposed diagnosis system based on ORP and DO can improve the stability of process management without additional costs. This study can be applied in the field for rapid fault detection using only a sub-cycle signal. This study focuses on developing a DTW-based discriminant analysis to detect faults using ORP and DO profiles obtained from SBRs.

2. Materials and methods

2.1. Operation of full-scale SBR

A full-scale swine wastewater treatment SBR (effective volume of 20 m^3) of rectangular shape with W 3 m × L 5.5 m × H 2.5 m was installed in Kimhae City. A ring blower supplied 3.64 m^3 air/min through 30 units of disk-type diffusers at the bottom of reactor. An impeller-type mixer was installed and operated during the anoxic stage for proper mixing. Raw swine wastewater was collected from a storage tank of scrapper-type barns. The influent characteristics are shown in Table 1.

The purpose of an SBR is mainly nitrogen removal in a two-step process of nitrification and denitrification. During nitrification, ammonia is converted aerobically to nitrate by aerobic microorganisms. In denitrification, nitrate is converted to nitrite or nitrogen gas under anaerobic conditions by anoxic microorganisms. Because the C/N ratio of the raw wastewater was less than three, an external carbon source was required to complete denitrification. The operating schedule and typical ORP profile for the SBR are shown in Fig. 1. One main cycle consisted of four sub-cycles, settling, decant phase and idle phase within 24 h. A classical SBR operation is performed with a fixed time cycle. However, this is a disadvantage because the cycle length cannot be adapted to process deviations and influent loading rate changes. Each sub-cycle was fixed at 1 h of anoxic phase and 3 h of aerobic phase. In automation-mode, the operation time of the aerobic phase was changed for denitrification and nitrification. This sub-cycle step performs most of the biological reactions and shows typical profiles with important points that provide information about the beginning and ending of the biological reactions. The DO and ORP profiles of the sub-cycle were collected and used for dynamic time warping and discriminant analysis.

The influent swine wastewater was fed at the beginning of each anoxic period, except during the final anoxic period, at 0.2–0.4 m^3 per batch with a feeding rate of 20 ml/min for 10–20 min. During the final anoxic period, methanol was fed to enhance denitrification. The hydraulic retention time was maintained from 10 to 25 days depending on the strength of the wastewater. Although the concentration of NH_4^+–N was high enough to cause substrate inhibition against nitrification, inhibition could be avoided by increasing dilution rate using the intermittent feeding of wastewater during the sub-cycle operation. The nominal operating loading rate was 0.22 kg NH_4^+–N/m^3/day as suggested by Kim et al. [16], and the volume of the feed was gradually increased to evaluate maximum removal capacity while maintaining an effluent quality of 60 TN mg/l.

Because most of the operation time is used for aeration, the optimisation and control of aeration are important. In a previous study, the process control using ORP or DO was performed with a threshold method including a set point of dORP/dt or dDO/dt. However, because these set points are affected by reactor and influent conditions, periodically fine tuning the set point was required for stable control [17]. DO and ORP profiles were obtained from a full-scale SBR. ORP (U.S. filter, Strantrol 880, USA), and DO (Knick,

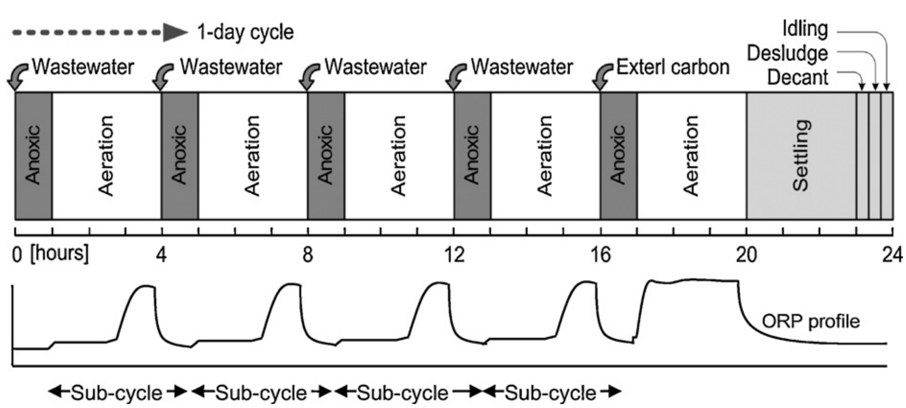

Fig. 1. Full cycle operation and typical ORP profile of SBR with fixed time operation; in automation-mode the aeration time is variable with the influent loading rate.

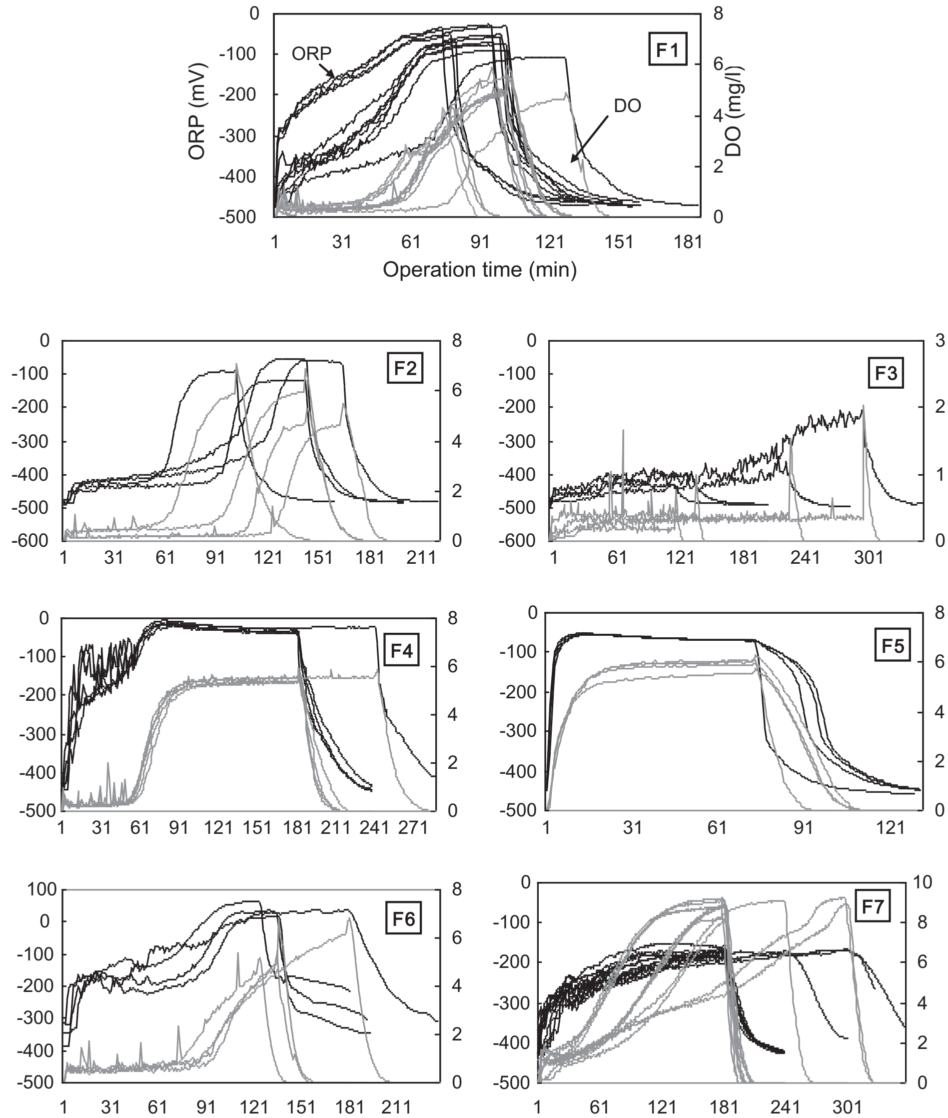

Fig. 2. Typical DO and ORP profiles corresponding to each fault case from F1 to F7.

Stratos 2401OXY, Germany) sensors were installed in the reactor for on-line monitoring and control.

2.2. Target faults

The on-line data patterns of an SBR are extremely dynamic depending on the operational mode and influent characteristics, such as influent loading rate and C/N ratio. A simple statistical model cannot easily detect abnormal behaviour in an SBR. However, a pattern-comparing method could be able to detect a point where an SBR deviates from a template profile that was generated from normal operation. The detection and diagnosis algorithm were established on the principle that the ORP and DO profile patterns during normal operation were different than those of abnormal operation caused by each fault. For the purpose of this study, we selected fault types in four categories: influent disturbance, aeration controller faults, instrument trouble and inadequate raw wastewater feeding. Table 2 lists the detailed definitions of the selected faults.

The most important disturbance in an SBR comes from the variations of influence loading rate that affect the biological reactions and the efficiency of the nutrient removal. Influent faults are divided into loading rate (F2, F3) and quality disturbance (F7). Although high loading rates within an adequate range do not disturb reactor operation, for stable operation, they should be monitored to avoid overloading (F2). An influent loading rate that is too high can seriously disturb biological reactions due to the toxic effect of ammonia in influent (F3). If influent is not fed, the operator must receive this information as soon as possible. Swine wastewater can be classified as scraper or slurry type. Scraper type wastewater is produced as faeces are separated from urine during collection by a mechanical device and has relatively low organic concentration. Slurry type wastewater is produced as faeces are mixed with urine and cleaning water, and its organic carbon concentration is about two times higher than the scraper type. A public treatment plant is subjected to only scraper type wastewater, and if slurry type wastewater is fed, the operator should be informed.

Aeration time in the sub-cycle was controlled with the threshold method using dORP/d*t* or dDO/d*t* values as control parameters. If the set point of a control parameter is too high, the mode change from the aerobic to anoxic phase will occur more rapidly and cause imperfect ammonium oxidation. Conversely, a low set value results in over-aeration during the aerobic phase (F4). These problems are

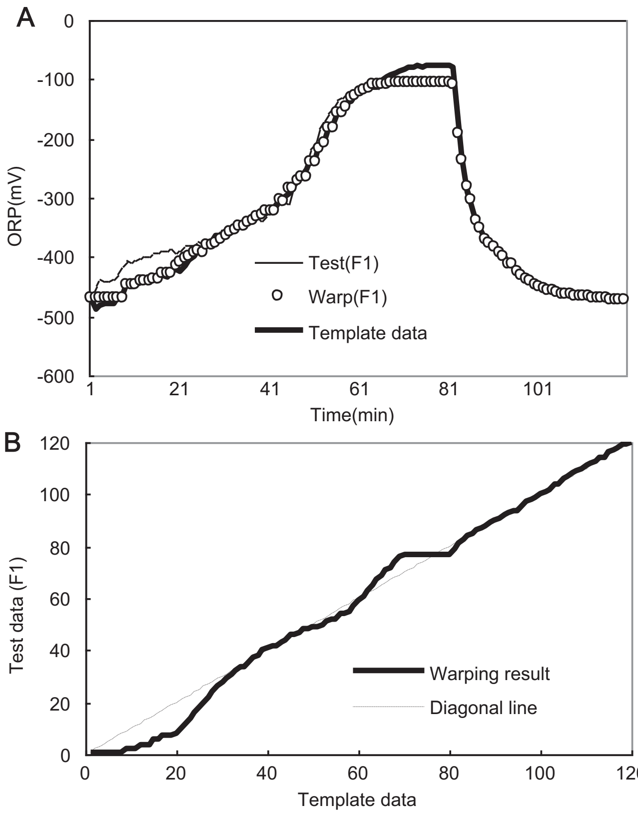

Fig. 3. Example of DTW alignment path for the F1.

included in the controller fault, and we recommend a feed-back response for fault diagnosis.

Instrument trouble (F5, F6) means malfunctioning equipment such as the influent feeding pump, blower, mixer and chemical (methanol) feeding pump. Influent feeding faults occur in plant quite frequently by clogging due to a high concentration of solids and debris in wastewater. However, chemical feeding pump faults do not occur often during the operation period.

2.3. Dynamic time warping (DTW)

DTW is used to find an optimal alignment between two given time-dependent sequences. DTW uses the principle of dynamic programming to nonlinearly warp two sequences. Consider a test sequence $T(i \times 1)$ of length I and a reference sequence $R(j \times 1)$ of length J. To measure the similarity between these two sequences, an $I \times J$ distance table D is constructed, where $d(i, j)$ is the local distance between $T(i)$ and $R(j)$. Typically, the Euclidean distance is used to measure the local distances, thus $d(i, j) = (T(i) - R(j))^2$. A warping path W is then calculated from a distance table that consists of a set of table elements that defines the mapping and alignment between $T(i)$ and $R(j)$. The overall distance $D(T, R)$ between the test sequence and the reference sequence is then calculated by summing the local distances over the warping path W. One popular choice for finding the best alignment between the test sequence and the reference sequence is to search for the path with the smallest distance of all possible warping paths. A detailed description of the DTW algorithm can be found in [18]. All programs used for calculation in the study were written in the Matlab 7.1 (MathWorks) computing environment.

2.4. Discriminant analysis

Discriminant analysis is a technique used for classifying a set of observations into predefined classes. Its purpose is to determine the class of an observation based on a set of variables known as predictors or input variables. The model is built based on a set of observations for which the classes are known. This set of observations is sometimes referred to as the training set. Based on the training set, the technique constructs a set of linear functions of the predictors, known as discriminant functions, such that

$$L = b_1x_1 + b_2x_2 + \cdots + b_nx_n + c \tag{1}$$

where b is the discriminant coefficient, x is the input variable or predictor and c is a constant. These discriminant functions are used to predict the class of a new observation when its class is unknown. For a k class problem, k discriminant functions are constructed. Given a new observation, all of the k discriminant functions are evaluated, and the observation is assigned to class i if the i^{th} discriminant function has the highest value. Discriminant analysis constructs a discriminant function for each group as follows:

$$f(G_i) = k_i + \sum_{j=1}^{n} w_{ij}p_{ij} \tag{2}$$

where i is a number of group (G), k_i is the constant inherent to each group, n is the number of parameters used to classify a set of data into a given group, w_i is the weight coefficient and p_{ij} is a selected parameter. In this study, we selected seven groups for evaluation, and a number of analytical parameters were used to measure the SBR. Data analysis was carried out using the SPSS v. 12.0 statistical software packages.

3. Results and discussion

3.1. DTW-based classification

The DO values respond to microbial oxidation reactions, so the DO profile provides a good indication of the ongoing biological reactions. ORP has a direct correlation with nitrification rates and other biological reactions in anoxic conditions. ORP can immediately show the changes of state in situ, but its range of values changes depending on environmental conditions. Under normal conditions, ORP is positive in aerobic stages and negative in anoxic stages. ORP profiles do not always display normal behaviour during the cycles. When a problem occurs, the most significant difference between the normal profiles is the ORP value range and slopes during aerobic or anoxic stages. While the ORP and DO profiles in full cycle operation were obtained, the profiles in sub-cycles only were used to DTW and discriminant analysis as shown in Fig. 1.

Some typical ORP and DO profiles corresponding to each fault case named F1–F7 are shown in Fig. 2. Because the operation time of the sub-cycle in automatic mode varies with the circumstances, some profiles were extended to over 180 min. The normal DO profile can be divided into three different phases including the DO lag phase where DO is under 2 mg/l, the rapidly increasing phase and the DO saturating phase. The normal operation of F1 shows typical DO profiles, which consist of three different phases; these are the lag phase, the rapidly increasing phase and the saturating phase. When the influent loading rate is increased, the lag phase of DO becomes longer than that of F1. If the influent loading rate is above an acceptable range, the lag phase exceeded the permitted aeration time, and the rapidly increasing phase did not occur. The aeration control fault of F4 shows a very long DO saturation phase. If influent was not fed due to influent feeding pump trouble, the DO lag phase became much shorter. Both F6 and F7 show a unique profile pattern, which may be caused by restricted biological reactions.

Table 3
D values between reference profiles and test profiles.

ORP	RE_1	RE_2	RE_3	RE_4	RE_5	RE_6	RE_7	DO	RE_1	RE_2	RE_3	RE_4	RE_5	RE_6	RE_7
F1	514	1108	5048	7686	6890	9363	5669	F1	4.6	4.3	58.3	47.4	183.1	13.7	211.4
	490	1080	6311	6123	5585	8378	5663		6.3	8.1	73.4	9.9	138.3	11.3	167.6
	416	1113	5452	6292	5776	8759	5912		4.4	8.3	57.5	25.9	155.2	9.9	197.2
	719	1835	5866	5906	5323	8574	5815		8.3	12.9	34.3	81.5	200.0	11.7	262.8
	548	660	5537	5781	5950	8668	6435		5.1	7.3	61.6	27.7	157.0	10.0	193.7
	1879	2717	9437	5804	8142	6396	4190		5.8	13.3	81.8	37.8	189.4	11.0	199.8
F2	1015	859	3367	1022	10,892	11,100	7207	F2	5.9	4.1	48.1	29.3	192.9	19.1	200.2
	1594	1859	3026	13,979	12,742	14,078	7811		19.3	24.1	82.6	10.0	139.7	20.7	123.9
	863	996	3375	10,899	9956	12,272	8092		19.5	26.6	88.7	9.6	117.1	20.6	104.8
F3	11,675	8837	2423	31,963	23,554	30,607	20,958	F3	73.2	74.5	4.1	252.3	303.1	74.7	430.6
	14,239	11,541	4388	35,323	26,002	33,966	24,318		99.8	80.4	6.9	321.0	372.9	111.4	515.9
F4	5740	7947	14,558	942	2086	3915	6898	F4	50.8	66.5	146.3	6.8	91.0	60.1	143.8
	5525	8361	14,524	562	3436	2918	5940		53.5	69.1	149.2	7.9	92.8	63.3	151.0
	7368	10,299	16,210	807	3774	3030	6902		73.3	93.1	172.1	5.8	91.8	81.8	150.1
F5	5765	8596	13,402	3540	731	6450	7448	F5	123.8	144.3	219.2	22.5	5.4	119.7	124.2
	3270	8451	11,157	4231	921	7674	10,481		93.2	108.5	189.0	20.5	0.7	113.9	153.8
	5371	7263	12,610	3438	559	7622	8339		101.9	121.1	195.1	10.6	21.0	97.0	148.7
	5718	8484	13,336	3515	697	6989	7707		113.7	134.5	209.0	18.2	8.2	113.3	136.7
	5765	8596	13,402	3540	731	6450	7448		123.7	144.3	219.2	22.5	5.4	119.7	124.3
	3287	5421	11,132	4292	36	7731	10,534		91.7	107.1	187.7	21.4	1.5	112.1	155.3
F6	14,230	16,501	23,940	3936	9670	949	7936	F6	28.6	40.7	116.6	16.7	122.0	25.0	101.1
	15,598	17,254	23,611	6585	14,752	2987	7694		11.1	22.1	64.5	25.7	154.6	6.0	174.8
F7	8367	10,436	13,403	12,094	12,938	7274	491	F7	136.4	156.6	231.2	72.8	93.3	138.1	6.6
	12,582	12,848	13,826	14,445	18,088	7702	1514		140.7	161.7	237.2	72.5	103.2	150.0	12.0
	4996	6696	7745	11,887	10,447	9185	1021		88.6	109.1	187.5	52.9	97.3	92.0	34.3
	4065	5225	7716	11,509	10,248	8822	1204		166.9	185.5	263.5	84.8	90.6	171.1	16.5
	3957	5364	7855	10,425	9511	8356	1579		178.8	197.9	274.0	92.2	89.8	194.1	9.5
	6389	8561	11,342	10,575	11,652	6699	1564		235.9	256.1	332.5	119.5	127.7	243.6	11.3

The normal profiles of each sub-cycle were used to generate template profiles, which were the mean value of the normal profiles at time *t*. The reference set consisted of seven patterns that correspond to seven major deterministic upsets. The method simply uses the DTW technique to align several training samples and then averages them to create the reference template. Only two or three training samples are usually necessary. The DTW-based classifier uses three training samples. The DO and ORP profiles are normalised so that they have a length of 120 points. The length feature is used to pre-select seven reference templates.

The performances of independent new sub-cycles are then compared with the reference set. In the DTW model, the reference and test data must have equal durations; however, achieving this is almost impossible in practice. The operation time of a sub-cycle in automation-mode may be different. In such cases, sub-cycle synchronisation is required. For this, the reference and independent test data are synchronised to a length of 120 points.

Fig. 3 shows an example of the DTW alignment path (A) and warping result in F1. The cumulative distance (*D*) from the beginning to the end points (120, 120) is calculated between two points of the test and reference. These calculated *D* values are summarised in Table 3. The gray shaded portions indicate the minimal *D* values among the results of *D* value calculation. The objective of DTW is to find the optimal path that connects the beginning and ending points of the grid with a minimal *D* value as shown in Fig. 3B. The high similarity between the reference template and the test profile results in a linear alignment path and a consequently low value of dissimilarity (*D*). As shown in Table 3, with the exception of pattern F2, all other ORP-based diagnoses are correct. Discrimination between various fault works better when the test pattern is identical to that of the reference patterns and becomes worsen when the test pattern is similar to another such as F1 and F2.

On the other hand, the DO-based diagnosis also revealed five misdiagnosis results. The DO results show that the process variable is the most important factor in the diagnosis. It was thought that using a discriminatory power as the classifier with the DTW method is more sensitive to profile shape patterns than phase length, such as the lag phase of the DO profile. As result, it was shown that ORP profiles can be more efficiently applied than DO as a diagnosis parameter.

3.2. Discriminant analysis

The time-shift problem of each sub-cycle process was solved by preprocessing the ORP and DO profiles using the DTW algorithm. After the DTW process, seven statistical values, including maximum value (Max), minimum value (Min), average value (Ave), standard deviation value (Stdev) and three quartile values (Q1, Q2, Q3) in sub-cycle operation were extracted and applied to establish the discriminant function.

In the DTW model, the DO and ORP profiles were normalised to a length of 120 points. The statistic variables that were used in the discriminant analysis were calculated with the same range of 120 points. The discriminant function is defined as the discriminant standard for the classification of the SBR operating state. The main purpose of a discriminant function analysis is to predict group membership based on a linear combination of the interval variables, which are Max, Min, Ave, Stdev, Q1, Q2 and Q3 in this study. The procedure begins with a set of observations where both group membership and the values of the interval variables are known. The end result of the procedure is a model that allows one to predict group membership when only the interval variables are known.

In stepwise discriminant function analysis, a discrimination model is built step-by-step. Specifically, at each step, all variables are reviewed and evaluated to determine which one contributes most to the discrimination between groups. That variable will then be included in the model, and the process begins again.

In this study, all of the variables were included in stepwise variable selection, and then some were eliminated in a stepwise manner. An iteration of the procedure consists of the evaluation of

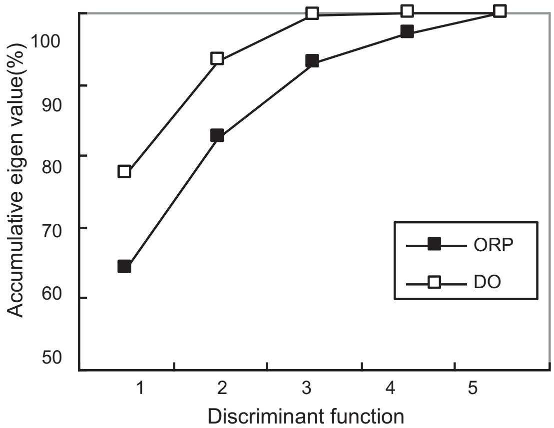

Fig. 4. Accumulative Eigen values of discriminant function.

the Wilks' Lambda (λ) value of each variable at every step, and the elimination of the variable which has the highest λ value. As the result of evaluation, five variables, Min, Stdev, Q1, Q2 and Q3, were selected for discriminant function construction using ORP. For DO, five variables, Ave, Max, Q1, Q2 and Q3 were selected. The discriminant coefficients obtained from the canonical discriminant analysis were used to compute the discriminant scores as follows:

$$\text{Factor\_ORP}\ (1) = 14.369 + 0.024 * \text{Min} + 0.034 * \text{Stdev} - 0.004 * \text{Q1} + 0.066 * \text{Q2} - 0.037 * \text{Q3}$$
$$\text{Factor\_ORP}\ (2) = 19.656 + 0.073 * \text{Min} + 0.133 * \text{Stdev} + 0.033 * \text{Q1} - 0.029 * \text{Q2} - 0.014 * \text{Q3}$$
$$\text{Factor\_ORP}\ (3) = 21.563 + 0.051 * \text{Min} + 0.011 * \text{Stdev} + 0.011 * \text{Q1} + 0.008 * \text{Q2} - 0.034 * \text{Q3}$$
$$\text{Factor\_ORP}\ (4) = -4.642 + 0.004 * \text{Min} + 0.089 * \text{Stdev} + 0.026 * \text{Q1} + 0.014 * \text{Q2} - 0.055 * \text{Q3}$$
$$\text{Factor\_ORP}\ (5) = -9.301 - 0.044 * \text{Min} - 0.019 * \text{Stdev} + 0.024 * \text{Q1} - 0.010 * \text{Q2} + 0.013 * \text{Q3}$$

$$\text{Factor\_DO}\ (1) = 19.773 - 16.161 * \text{Ave} - 9.436 * \text{Stdev} + 3.422 * \text{Q1} + 3.713 * \text{Q2} + 6.865 * \text{Q3}$$
$$\text{Factor\_DO}\ (2) = 0.468 - 2.014 * \text{Ave} - 3.148 * \text{Stdev} - 1.988 * \text{Q1} + 2.046 * \text{Q2} + 1.679 * \text{Q3}$$
$$\text{Factor\_DO}\ (3) = 2.272 - 8.336 * \text{Ave} - 1.179 * \text{Stdev} + 5.937 * \text{Q1} + 0.957 * \text{Q2} + 3.670 * \text{Q3}$$
$$\text{Factor\_DO}\ (4) = 0.172 - 11.580 * \text{Ave} + 6.588 * \text{Stdev} + 7.357 * \text{Q1} + 3.847 * \text{Q2} + 0.400 * \text{Q3}$$
$$\text{Factor\_DO}\ (5) = -0.060 - 13.745 * \text{Ave} + 6.271 * \text{Stdev} + 5.945 * \text{Q1} + 3.308 * \text{Q2} + 2.395 * \text{Q3}$$

Factor_ORP (1) was dependent largely on Q2 and Stdev. Factor_ORP (2) was mainly dependent on Stdev and Min. The canonical correlation coefficient for Factor_ORP (1), 0.99, was larger than the corresponding canonical correlation coefficient for Factor_ORP (2), 0.97. This result shows that Factor_ORP (1) had the most discriminant functions.

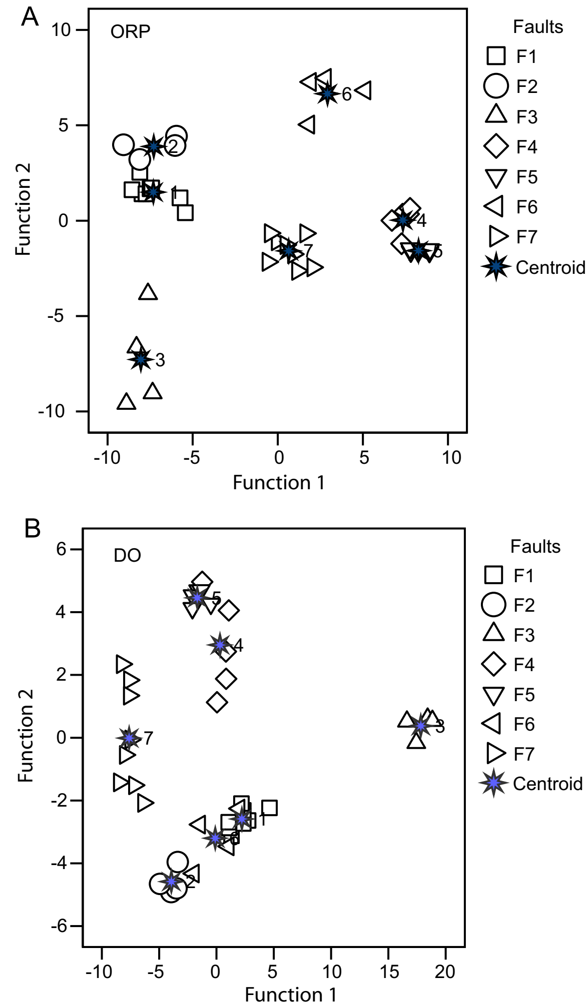

Fig. 5. Scatter plot of discriminant score.

As shown in Fig. 4, the accumulative percentages of the third discriminant function Eigen value of ORP and DO were 92.9% and 99.6%, respectively, and showed that dimension reduction using principal component analysis (PCA) could be applied effectively.

The number of misclassified sites was determined, and the misclassification rate is given in Table 4. The shaded cells indicate an incorrect diagnosis. None of the ORP group memberships was misclassified. One out of F4 and two out of F6 were misclassified. The discriminant analysis of DO data was able to classify correctly 92.7% of the original group cases. The discriminant score of F4 had a wide spread range, which resulted in one overlapping point in the F5 group. The centroid of F6 was located between F1 and F2, which caused a high possibility of misclassification. Abnormal anoxic conditions caused by chemical feeding pump faults are reflected in

Table 4
Classification results in each fault cases.

Variety	Predicted group membership (ORP)							Predicted group membership (DO)						
	1	2	3	4	5	6	7	1	2	3	4	5	6	7
1	8	0	0	0	0	0	0	8	0	0	0	0	0	0
2	0	4	0	0	0	0	0	0	4	0	0	0	0	0
3	0	0	4	0	0	0	0	0	0	4	0	0	0	0
4	0	0	0	5	0	0	0	0	0	0	4	1	0	0
5	0	0	0	0	8	0	0	0	0	0	0	8	0	0
6	0	0	0	0	0	4	0	1	1	0	0	0	2	0
7	0	0	0	0	0	0	8	0	0	0	0	0	0	8

the ORP pattern, whereas the DO profile is less sensitive to anoxic biological reactions.

The values of the two canonical discriminant functions of Function (1) and Function (2) can be seen in Fig. 5. In Fig. 5, the fault case of ORP data can be divided into seven different groups. There was no misclassification of ORP data in spite of the close proximity between F1 and F2. Because the operating status of F2 required the operator's increased attention but was not itself a serious fault, the proximate discriminant group was assumed to be acceptable.

Consequently, we concluded that using ORP profiles is more efficient than DO and DTW algorithms, and discriminant analysis can be applied to stable SBR operation for swine wastewater treatment. In fact, effective diagnosis was possible using a discriminant analysis with the simple statistical values of maximum, minimum, average, standard deviation and quartile values.

4. Conclusion

The diagnosis method of deterministic faults in a full-scale SBR for swine wastewater treatment was studied using dynamic time warping and discriminant analysis. Based on this study, we were able to draw the following conclusions:

1. Using ORP and DO profiles that were extracted from sub-cycle, full-scale SBR status can be divided into a normal case and four abnormal categories—influent disturbance, aeration controller fault, instrument trouble and inadequate raw wastewater feeding.
2. Using the DTW process for seven reference and operation profiles, difference values (*D*) were determined and classified into seven cases. Although the DTW process misclassified high loading rate (F2), the ORP profile provided good diagnosis results. However, the DO profiles detected five misclassifications.
3. After the DTW process, several statistical values, including maximum value (Max), minimum value (Min), average value (Ave), standard deviation value (Stdev) and three quartile values (Q1, Q2, Q3), were extracted and applied to establish the discriminant function. The accumulative percentages of the third discriminant function Eigen value of ORP and DO were 92.9% and 99.6%, respectively; this finding proved that dimension reduction using principal component analysis (PCA) could be applied effectively.
4. Discriminant analysis allows us to classify seven cases with a percentage of 100% and 92.7% for ORP and DO profiles, respectively. This study showed the ability of the method to classify process faults independently of their magnitude or the time origin of the fault. The use of PCA as an additional extraction step reduced the dimensions of each profile and resulted in an improvement of the discriminatory power of the classifier. Consequently, we concluded that the DTW algorithm and discriminant analysis can be applied to stable SBR operation for swine wastewater treatment.

Acknowledgement

This research was supported by a grant from Kangwon National University in 2009.

References

[1] N. Bernet, F. Beline, Challenges and innovations on biological treatment of livestock effluents, Bioresour. Technol. 100 (2009) 5431–5436.
[2] G. Andreottola, P. Foladori, M. Ragazzi, On-line control of a SBR system for nitrogen removal from industrial wastewater, Water Sci. Technol. 43 (2001) 93–100.
[3] J.H. Kim, M. Chen, N. Kishida, R. Sudo, Integrated real-time control strategy for nitrogen removal in swine wastewater treatment using sequencing batch reactors, Water Res. 38 (2004) 3340–3348.
[4] C.H. Ga, C.S. Ra, Real-time control of oxic phase using pH(mV)–time profile in swine wastewater treatment, J. Hazard. Mater. 172 (2009) 61–67.
[5] B. Lamrini, M.V. Lann, A. Benhammou, K. Lakhal, Detection of functional states by the 'LAMDA' classification technique: application to a coagulation process in drinking water process, C. R. Phys. 6 (2005) 1161–1168.
[6] R. Bellman, Dynamic Programming, Princeton University Press, Princeton, NJ, 1957.
[7] L. Rabiner, B.H. Juang, Fundamentals of Speech Recognition, Prentice Hall, Englewood Cliffs, NJ, 1993.
[8] K. Chen, T.Y. Wu, H.J. Zhang, On the use of nearest feature line for speaker identification, Pattern Recognit. Lett. 23 (2002) 1735–1746.
[9] I. Guler, M. Meghdadi, A different approach to off-line handwritten signature verification using the optimal dynamic time warping algorithm, Digit. Signal Process. 18 (2008) 940–950.
[10] V.P. Carlos, F.Z. Marcos, M.P. Juan, An efficient low cost approach for on-line signature recognition based on length normalization and fractional distances, Pattern Recognit. 42 (2009) 183–193.
[11] H.J. Ramaker, E.N.M. van Sprang, J.A. Westerhuis, A.K. Smilde, Dynamic time warping of spectroscopic BATCH data, Anal. Chim. Acta 498 (2003) 133–153.
[12] K. Gollmer, C. Posten, Supervision of bioprocesses using a dynamic time warping algorithm, Control Eng. Pract. 4 (1996) 1287–1295.
[13] V. Pravdova, B. Walczak, D.L. Massart, A comparison of two algorithms for warping of analytical signals, Anal. Chim. Acta 456 (2002) 77–92.
[14] K.P. Singh, A. Malik, D. Mohan, S. Sinha, Multivariate statistical techniques for the evaluation of spatial and temporal variations in water quality of Gomti River (India)—a case study, Water Res. 38 (2004) 3980–3992.
[15] X.T. Doan, R. Srinivasan, Online monitoring of multi-phase batch processes using phase-based multivariate statistical process control, Comput. Chem. Eng. 32 (2008) 230–243.
[16] B.C. Kim, S.P. Cheon, S.S. Kim, H.J. Woo, C.W. Kim, Fuzzy control optimization based on operator's action for the activated sludge process, in: The 1st IWA Conf. on Instrum. Control and Automation, June 3–7, Malmo, Sweden, 2001.
[17] B.H. Jun, D.H. Kim, E.H. Choi, H. Bae, S.S. Kim, C.W. Kim, Control of SBR operation for Piggery Wastewater Treatment with DO and ORP, J. Korean Soc. Water Qual. 18 (2002) 545–551.
[18] H. Ramaker, E.N. van Sprang, J.A. Westerhuis, A.K. Smilde, Dynamic time warping of spectroscopic BATCH data, Anal. Chim. Acta 498 (2003) 133–153.