15th Triennial World Congress, Barcelona, Spain

Online Fault Detection in Virginiamycin Production

S. Shioya*, J.H. Huang and H. Shimizu

Dept. of Biotechnology, Graduate School of Engineering,Osaka University, Suita, Osaka 565-0871, Japan
*: Corresponding author

Abstract: It is difficult to measure online substrate, biomass, and product concentrations, due to the lack of reliable sensors in the fermentation. In view of this, the pH, dissolved oxygen (DO) concentration, and CO_2 production, among others, are usually utilized in bioprocess analysis. With these easily obtained online measurements, it is possible to reconstruct the evolution of the state variables or estimate the bioprocess parameters. Neural networks, which rely on the efficacious nonlinear multivariate analysis capacity and its favorable black-box feature, are most widely applied to bioprocess analysis and fault detection. In this study, an artificial autoassociative neural network (AANN) has been used online to detect deviations from normal antibiotic production fermentation with ordinary state variables. To improve the efficiency of extracting hidden information contained in multidimensional state variables, and finally to render the AANN adequate for fault detection, we have explored the following methods: preprocessing of the data that involved normalizing the training data of the AANN; evaluation of the data that involved assessing the output of the AANN; and selection of state variables. A method for fault detection for virginiamycin production by *Streptomyces virginiae* was developed.

Keywords: neural network, data preprocessing, data evaluation, antibiotics production

1. INTRODUCTION

It is very important to detect deviations or faults in a bioprocess in real time. An optimization strategy is usually applied to industrial operations. However, monitoring the bioprocess (Bastin and Dochain, 1990, Thibault *et al*., 1990), for which an optimization strategy is applied, from the initial to the final stage, and predicting the process under normal or abnormal conditions, are major challenges. Many approaches have been used to detect various deviations from or faults in normal fermentations by extracting the relevant information from multidimensional data (Stephanopoulos and Han, 1996). Principal component analysis (PCA), one of the multivariate statistical projection methods, has been applied to fault detection (Saner and Stephanopoulos, 1992). Online fault detection for primary metabolite production using artificial auto-associative neural network (AANN) has been tested by Shimizu *et al*. (1998). Ignova *et al*. (1999) used a self-organizing map (SOM), which is an unsupervised artificial neural network (ANN) and facilitates visualization of complex data, to analyze fermentation seed quality with routinely measured plant data.

When bioprocess analysis or fault detection via a neural network is considered, three steps are usually necessary, they are: (1) selecting the structure of the neural network, (2) designing a method for preprocessing of the data before the neural network is applied and (3) applying the neural network with preprocessed data and evaluating the result. Once the structure of the neural network is fixed, the characteristics of the preprocessed data, such as the information density and validity, directly affect the analysis of the neural network; in other words, designing a method to preprocess the data relates to the efficiency of extracting hidden information. Moreover, since the results derived directly from the output of the neural network are difficult to understand, it is necessary to assess the output data. Consequently, when using a neural network, bioprocess analysis efficiency depends on evaluation of the output of the neural network. In this study, an AANN was adopted to detect faults

in secondary metabolite production with ordinary online state variables. We have developed a method for preprocessing of the data, evaluation of the output of the AANN, and selection of state variables. As a practical example, virginiamycin fermentation of *Streptomyces virginiae* was performed for this research.

2. MATERIALS AND METHODS

2.1 Culture Conditions

Streptomyces virginiae MAFF10-06014 (National Food Research Institute, Ministry of Agriculture, Forestry and Fisheries, Tsukuba, Japan) was used as a virginiamycin M- and S-producing microorganism. Optimal production strategies for virginiamycins have previously been studied (Yang *et al*., 1996, Shioya *et al*., 1999). The main fermentations were conducted under batch mode, in a 5-L jar fermentor (KMJ-5A, Mitsuwa Co., Japan) fitted with temperature (T), dissolved oxygen (DO) concentration and pH sensors and a laser turbid meter (LA-300LT, Automatic System Research Ltd., Japan) for measuring the optical density (OD). The DO concentration was maintained above 50% of saturation, and DO control was achieved by adjusting the agitation speed (RPM) when the DO concentration fell below the desired value. The feed weights of hydrochloric acid (HCl) and sodium hydroxide (NaOH) for pH control were measured, using the electrical balance (FX3000, A&D Ltd., Japan). The concentrations of CO_2 (CO_2%) and O_2 (O_2%) in the exhaust gas were measured online with a CO_2 analyzer (ZFD, Fuji Electric Ltd., Japan) and an O_2 analyzer (PMA200, Horiba Ltd., Japan). The sampling interval was 1 min.

2.2 Selecting State Variables

In the fermentation used in this study, the ordinary state variables were detected online, including T, DO concentration, pH, OD, the CO_2% and O_2%, the HCl and NaOH addition, and the RPM. The virginiamycin production, dry cell mass, and glycerol concentration were measured offline. Based on virginiamycin production, fermentations can be divided into two groups of high and low production (shown in Figure 1). Here, the virginiamycin quantum was represented by virginiamycin M because the amount of virginiamycin S produced was very low compared with virginiamycin M.

Among the nine online state variables mentioned above, T, pH and DO concentration were used as process control variables. For utilizing limited online state variables as fully as possible, the six remaining variables were all subjected to multivariate analysis at first. The problem caused by selecting the NaOH addition as an analyzing variable is discussed in the text.

2.3 Wavelet Filter Bank (WFB)

In general, online raw fermentation data that is obtained directly includes a lot of noise. Noise results in low efficiency and a longer time to train the AANN. In this study, online raw data were smoothed via a wavelet filter bank (WFB) before the AANN was implemented. The filter bank was composed of a set of high-pass filters, low-pass filters, and sampling operators (Strang and Nguyen, 1996). *Daubechies* second-order wavelet function was combined with the WFB that was utilized in this study, because it is considered the simplest and most useful function for multiresolution analysis and noise filtering (Daubechies, 1988).

2.4 Artificial Autoassociative Neural Network

Kramer (1991) developed an AANN using a nonlinear statistical approach for multivariate analysis of chemical processes. The AANN operates by training a feed-forward neural network (FFNN) to identify the mapping from where the network inputs are reproduced at the output layer. The AANN consists of five layers of input, mapping, bottleneck, demapping and output. The number of nodes in the output layer is the same as that in the input layer. The number of nodes in the bottleneck layer should be lower than that in the input and output layers. This means that the mapping layer compresses the system data while the demapping layer works to restore the data. The AANN was termed nonlinear principal component analysis by Kramer (1991), because it is similar in concept to linear principal component analysis (PCA) but it uses nonlinear functions as the base functions. In this study, two nonlinear sigmoid functions were used (equations 1 and 2).

$$f(u) = \frac{1 - e^{-u}}{1 + e^{-u}} \qquad (1)$$

$$g(u) = \frac{1}{1 + e^{-u}} \qquad (2)$$

Equation 1 with a wider output interval of (-1, 1) was employed as a transfer function of the mapping, bottleneck, demapping layers since it supplies a wider space for mapping and demapping. Equation 2 with an output interval of (0, 1) was

only used in the output layer for transforming.

Firstly, the six variables of OD, the CO_2% and O_2% in the exhaust gas, the weight of HCl and NaOH fed into the fermentor for pH control, and the RPM for DO concentration control were selected as the input variables of the AANN. As a concise and efficient structure is desired in the construction of an AANN, here, the mapping and demapping layers contained seven nodes each, and the bottleneck layer contained two nodes, with the AANN being termed AANN-67276. AANN-56265 was also constructed (exclude NaOH addition rate), which is explained mainly in this report. The AANN was trained only using data from a normal fermentation process. The training calculation for the AANN was performed by a commercial software package, Skill Tran (Chiyoda Kako Co. Ltd., Tokyo), on a Sun Classic workstation (Sun Microsystems, U.S.)

3. RESULTS AND DISCUSSION

3.1 Virginiamycin Fermentation Process

Batch cultures of antibiotic-producing filamentous microorganisms (actinomyces and fungi) usually undergo two phases of development, the growth phase and the production phase (Yang *et al*., 1996). The growth phase lasted for about 800 min from the start of culture. Figure 1 shows the starting time and the amounts of virginiamycin produced for all of the twelve runs. The energy was supplied by yeast extract and Bacto-casitione in the growth phase, whereas glycerol was used predominantly in the production phase (Yang *et al*., 1996). The arrow shown in Figure 2A indicates that the CO_2 concentration in the exhaust gas decreased suddenly and formed a characteristic peak at about 600 min from the start of culture. Correspondingly, the other state variables (O_2 concentration, DO concentration and so forth) also gave rise to a similar peak at the same time. This characteristic feature might be related not only to the change in the energy source, but also to the initiation of the biosynthesis of autoregulators of *S. virginiae*, virginiae butanolides (VBs) (Yang *et al*., 1996). Generally, autoregulators are synthesized by the microorganism itself during culture and subsequently induce antibiotic biosynthesis. Thus, this characteristic peak related to VB biosynthesis and virginiamycin production level is very important and it was associated with a novel preprocessing of data to normalize the training data of the AANN in this study. In addition, another characteristic which should be noted is that NaOH which was added for pH control at the late stage of the production phase in all of the abnormal fermentations (Figure 2D), did not need to be added in any of the normal fermentations (data not shown). The addition of NaOH at the late stage of the production phase could possibly be used as an

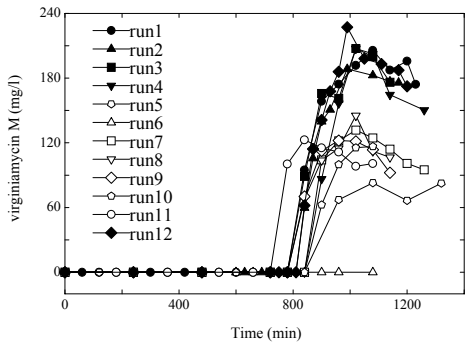

Fig. 1 Time courses of virginiamycin M production in normal and abnormal fermentations.

Solid symbols show optimal VM production, and blank symbols show that maximum production was not attained. Runs 1, 2, 3, 4 and 12 are normal fermentations. In runs 5 to 11 the DO concentration and pH were not controlled, the fermentation was contaminated, the culture temperature diverged from the optimal of 28 ºC to 29.5 ºC, the concentration of the autoregulators of *S. virginiae*, virginiae butanolides (VBs), was abnormal from the start of culture, the DO concentration was controlled at a low level of 25% of saturation, the low quality yeast extract was used and the inoculum was of low cell activity with serially transferred culture, respectively.

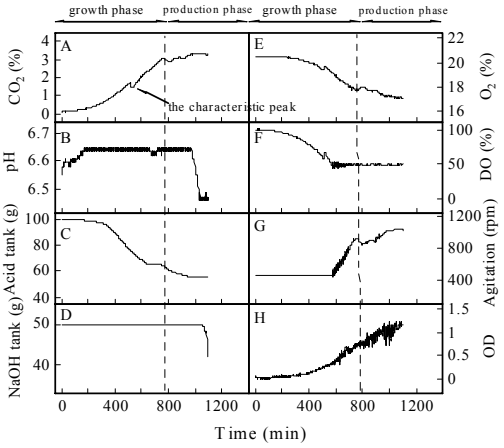

Fig. 2 Time courses of online state variables in an abnormal fermentation.

The DO concentration, pH, OD, CO_2% and O_2% in the exhaust gas, the feed weight of HCl and NaOH for pH control, and the RPM for DO control are shown. The arrow indicates the characteristic peak of CO_2 concentration in the fermentation process of *Streptomyces virginiae*.

indicator for distinguishing normal or abnormal fermentations, though this signal came too late for detecting faults for an industrial plant.

3.2 Standard Data-Preprocessing

In this study, online raw data were filtered and compressed by the WFB at a resolution scale of three before the AANN was implemented. Furthermore, it was necessary to scale state variable data into an appropriate unified interval before the AANN was implemented since state variable units differ. An ordinary normalization is adopted by dividing each state variable datum by its maximum value simply to scale data into the range of (0, 1). Figure 3a shows the normalized data of CO_2(%) for a normal process and run 11, which was an example of an abnormal fermentation process. In addition to this ordinary normalization, a novel standardizing method, related to the characteristic peak of the virginiamycin fermentation process, will be described later in this paper.

3.3 Fault Detection by the AANN

The normal fermentation data of runs 1, 2, 3, which had been filtered by the WFB and standardized by ordinary normalization, were used to train the AANN-67276. It is also expected that the error between the input and output values will be very small because the AANN was trained only using the data pattern from the normal fermentation state variables. Based on this concept, a statistical index, J, was introduced to detect deviations from the normal process operation. J is defined in Equation 3.

$$J = \sum \left\{ (input - output)^2 \right\}, \quad (3)$$

where J is the summation of the squared error between input and output for the state variables of OD, CO_2%, O_2% , HCl, NaOH and RPM, and it was evaluated at each sampling time. As mentioned above, the J of a normal fermentation should be low during fermentation, and will rise in an abnormal fermentation. The deviation from normality can be detected in real time by setting an appropriate threshold value, ε, to evaluate J. When $J \leq \varepsilon$, the fermentation process is considered normal, and when $J > \varepsilon$, the corresponding process is judged abnormal. Using AANN-67276, fault operation or low product operation was detected but the time to detect was lather late, which resulted in that the virginiamycin production had already started (data not shown). So, the improvement of AANN was tried to by changing the data preprocessing.

3.4 A Novel Preprocessing Method

As previously mentioned, this characteristic peak should be related not only to the change in the energy source utilized, but also to the initiation of VB biosynthesis, and is associated with the start of virginiamycin production (Yang *et al*.,1995, 1996a). In the proposed normalization, the characteristic peak was regarded as a unified criterion, and various minimum values at the time of the sudden decrease in the CO_2 concentration were set at the unified level by multiplying their respective factors. Then, these factors were multiplied as an influence element while normalizing the other state variables, such as OD and the amount of HCl added. To differentiate an ordinary normalization, we arbitrarily named the proposed normalization a concrete preprocess (CPP). Figure 3b shows the CO_2(%) data in normal fermentation processes and in run 11 normalized

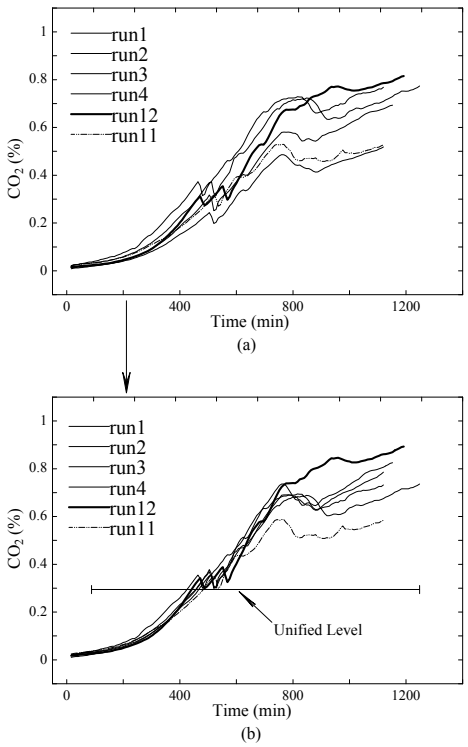

Fig. 3 Training data of the AANN for CO_2 (%)
Normalized by ordinary preprocessing, namely, generating an interval of (0, 1) for original data by dividing by the maximum value of this state variable (a); normalized by the CPP, in which various minimum value of a sudden decrease in the CO_2 concentration were set at the unified level by multiplying their respective factors (b).

by the CPP. Compared with Figs. 3a and 3b, it is apparent that the curves for the normal fermentation process clustered while the curve for run 11 separated from those for the normal runs (Fig. 3b). This indicated that the characteristics of the normal fermentation process had been pre-extracted.

3.5 *AANN-56265 and Evaluation of Output*

In this study, the five variables of OD, CO_2(%), O_2(%), HCl addition and RPM were selected as the input variables of the AANN. Correspondingly, the most concise AANN, AANN-56265 was constructed, in which the mapping and demapping layers contained six nodes each, and the bottleneck layer contained two nodes. This AANN-56265 was trained using training data obtained via the CPP. Figure 4A shows two-dimensional plots of normal and abnormal data with respect to the Z1 and Z2 at this AANN-56265, where the variables, Z1 and Z2, are ones at the bottle neck layer. It is obvious that the data for the normal and abnormal fermentation process plots were clustered in different groups, and the dashed line in Fig.4A indicates the boundary between the two groups. Similarly, the Z1 data of normal runs were not only clustered but also clearly separated from the data for run 11 (Fig. 4B). This indicated that different information hidden in the state variables of normal and abnormal fermentation processes had been extracted and classified efficiently.

The outputs at the bottleneck layer of the AANN are usually considered as the principal components in correspondence with nonlinear PCA. Here, to set a reasonable value of ε, the h in Fig. 4B, which represented the difference between normal and abnormal runs, was investigated in real time. As a new evaluation index, H was defined in,

$$H_i = (Z1_i - \tilde{Z1}_{\text{run 1,2,3}})^2, \qquad (4)$$

where $\tilde{Z1}_{\text{run 1,2,3}}$ is the average of the training data set of the AANN, and was used as the standard of Z1 for normal fermentation processes. $Z1_i$ is Z1 data for every fermentation process. H_i represents the difference between the standard for a normal operation and an actual running process, and it was evaluated at each sampling time. To assess the deviations from the standard for normal operation, at first the maximum distance between actual normal runs and the standard for a normal fermentation process was calculated, and termed $h_{\text{max (normal)}}$. The $h_{\text{max (normal)}}$ was selected at the maximum value among the training data (runs 1, 2 and 3) by Equation 4. Then, a warning line

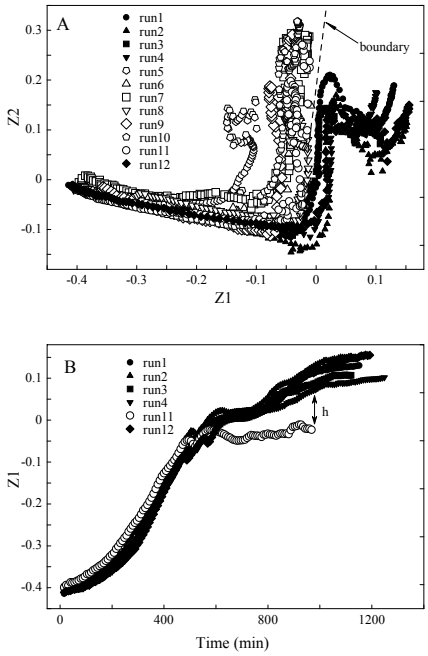

Fig. 4 Two-dimensional plots of Z1 and Z2 data using the trained AANN-56265.

The normal and abnormal fermentation process plots were clustered into different groups, and the dashed line indicates the boundary between the two groups (A). Time course of Z1 for normal runs and run 11 (B). Solid and blank symbols indicate normal and abnormal data, respectively. The AANN-56265 was trained by three normal data sets of runs 1, 2 and 3, and the training data were obtained by the CPP.

referring to the $h_{\text{max (normal)}}$ was set at 6-fold $h_{\text{max (normal)}}$ and smoothed by polynomial curve termed θ_{warn}. Similarly, 8-fold $h_{\text{max (normal)}}$ was adopted as an emergency line and its polynomial curve was termed θ_{emerg}. The emergency line indicates that the running process is going into an abnormal state when its H value goes beyond this line. As observed in Figs. 5A and 5B, the H values of the normal fermentation processes of run 4 and 12 were under the θ_{warn} for almost the entire fermentation; the H value of the abnormal fermentation of run 11 overstepped the warning line and finally crossed through the gray region to go into the area representing an abnormal fermentation state. Illustrations such as in Figs. 5B nd 5C, referring to the intersection of the H value of abnormal runs with θ_{warn} and θ_{emerg}, can be used to access the $\varepsilon_{\text{warn}}$ and $\varepsilon_{\text{emerg}}$ at the J coordinate axis. The average of $\varepsilon_{\text{warn}}$ and $\varepsilon_{\text{emerg}}$ was adopted as the threshold value of J, namely ε. In this study,

the ε was rounded at 0.04.

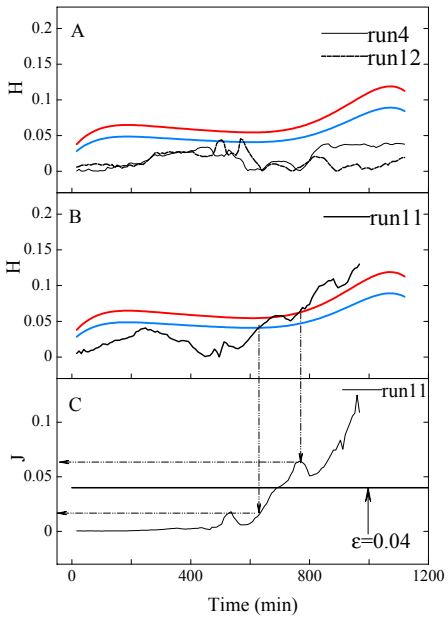

Fig. 5 Illustrations of fault detection via H evaluation.

The H values of the normal fermentation processes of run 4 and 12 were under θ_{warn} for almost the entire cultivation (A); and the H of the abnormal fermentation of run 11 overstepped the warning line and crossed through the gray region into the area representing an abnormal fermentation state (B). Referring to the intersection of H of the abnormal fermentation process with θ_{warn} and θ_{emerg}, the ε_{warn} and ε_{emerg} can be accessed at the J coordinate axis (C). The average of ε_{warn} and ε_{emerg} was adopted as the threshold value of J, namely, ε. In this study, the ε was rounded at 0.04.

4. CONCLUSION

Utilizing ordinary online measurable state variables, such CO_2 concentration in the exhaust gas, deviations from the normality and faults have been detected in an antibiotic fermentation process before the starting time of virginiamycin production using a trained AANN-56265. The strategy developed here was proved to be used well for online fault detection of the primary metabolite production process by our previous study. Now the method is applicable to the online fault detection of other complex secondary metabolite productions, particularly to antibiotic fermentations that usually undergo two phases of development, the growth phase and the production phase.

REFERENCES

,.

Bastin, G., Dochain, D. 1990. On line estimation and adaptive control of bioreactors. Elsevier, Amsterdam.

Daubechies, I. 1988. Orthonormal bases of compactly supported wavelets. Communication on Pure and Applied Mathematics. **61**: 909-996.

Ignova, M., Montague, G. A., Ward, A. C., Glassey, J. 1999. Fermentation seed quality analysis with self-organising neural networks. Biotechnol. Bioeng. **64**: 82-91.

Kramer, N. A. 1991. Nonlinear principal component analysis using auto-associative neural networks. AIChE J. **37**: 233-243.

Saner, U., Stephanopoulos, G. Application of pattern recognition techniques to fermentation data analysis. In *Proceedings of the 5th International Conference on Computer Application in Biotechnology*. Karim, N. M., Stephanopoulos, G., Eds.; Pergamon: New York, 1992. 123-128.

Shimizu, H., Yasuoka, K., Uchiyama, K., Shioya, S. 1998. Bioprocess fault detection by nonlinear multivariate analysis: Application of an Artificial Autoassociative Neural Network and Wavelet Filter Bank. Biotechnol. Prog. **14**: 79-87.

Shioya, S., Morikawa, M., Kajihara Y., Shimizu, H. 1999. Optimization of agitation and aeration conditions for maximum virginiamycin production. Appl. Microbiol. Biotechnol. **51**: 164-169.

Stephanopoulos, G., Han, C. 1996. Intelligent systems in process engineering. A review. Comput Chem Eng. **20**: 743-791.

Strang, G., Nguyen, T. Wavelets and filter banks. Wellseley Cambridge Press. Wellseley, MA. 1996.

Thibault, J., Breusegem, V.V., Cheruy, A. 1990. On-line prediction of fermentation variables using Neural Networks. Biotechnol. Bioeng. 36:1041-1048.

Yang, Y. K., Morikawa, M., Shimizu, H., Shioya, S., Suga, K., Nihira, T., Yamada, Y. 1996. Maximum virginiamycin production by optimization of cultivation conditions in batch culture with autoregulator addition. Biotechnol. Bioeng. **49**: 437-444.