BIOENGINEERING ANALYSIS OF INCOMPLETE DATA FROM LARGE-SCALE ANAEROBIC BIOREACTOR FOR WASTE WATER TREATMENT BY FUZZY EXPERT SYSTEM

Vu Lan HUONG[a], Jaroslav VOTRUBA[a,*] and Ivan STUCHL[b]
[a] *Institute of Microbiology,*
Academy of Sciences of the Czech Republic, 142 20 Prague 4, The Czech Republic
[b] *Sugar Research Institute,*
Cukrspol, a.s., 140 00 Prague 4, The Czech Republic

Received June 1, 1993
Accepted August 13, 1993

The knowledge-based model of an anaerobic waste water treatment plant was designed when the incompleteness of industrial records excluded the application of mass balances. New two step information technology that permits to excerpt the knowledge from industrial data as a set of heuristic rules is described. In the first step, the real numbers are converted to the set of fuzzy ones and arranged as a knowledge data base of an expert system. Secondly, the fuzzy expert system SENECA was used for fuzzy modelling of the process. Three examples of bioengineering analysis are presented. First, the average efficiency of biological degradation was determined. Second, the effect of pH in the input stream was evaluated. Third, the process dynamics and time required to reach the maximum activity of microbial population were estimated. It was possible to detect and evaluate the decrease of the biodegradation efficiency caused by unexpected leakage of xenobiotics.

There is no doubt that the mathematical models either deterministic or stochastic are basic tools used by process engineers for process control and optimization in chemical and related technologies. For most biological unit operations, this approach fails owing to the complexity of the process and uncertainty or incompleteness of measured data. Usually, it deserves ill either because of the biological nature of a process or because of the accuracy of analytical method used. To face these difficulties caused by inherent uncertainty of data, which excludes the application of deterministic models[1 – 3], we used fuzzy modelling, introduced by Dohnal[4,5], to analyze the behavior of real large-scale anaerobic biological waste water treatment plant in sugar factory at Brodek near Prerov.

---

[*] The author to whom correspondence should be addressed.

THEORETICAL

The details of plant construction including the limited amount of experimental data were published in a series of papers[6 – 12]. The flowsheet in Fig. 1 illustrates the linking of anaerobic reactor with the aerobic part of the waste water treatment plant.

The waste water from the sugar factory contains organic pollutants whose total amount may be represented as chemical oxygen demand (COD)[13]. The anaerobic microorganisms living in the waste water degrade the sugar-like compounds by the reaction:

$$C_6H_{12}O_6 \longrightarrow 3\,CH_4 + 3\,CO_2 + \text{biosludge} + \text{heat}\,. \qquad (A)$$

We expressed the efficiency of biological degradation as the difference of COD of input and output stream. The available data, COD and pH of output and input streams, are summarized in Figs 2 and 3. Evidently, such a data set is incomplete and it excludes to carry out balance calculations[1]. First, we wished to estimate, how long it takes the bioreactor to reach the maximum of biological purification efficiency. Second, we wanted to analyze the effects of the input COD and pH on the bioreactor behavior. To excerpt the desired information from industrial records, we developed an information technology based on the fuzzy modelling.

Fuzzy Model of the Process and Expert System SENECA

Earlier, Dohnal[4,5] described a fuzzy expert system shell that was implemented as computer program SENECA. The application of the program supposes that the state variables are represented by fuzzy numbers in special trapezoidal form[14]. Recently, Sterbacek and Votruba[15] have described in detail a simple method that permits to convert the process records from real numbers to fuzzy ones. Similar approach was used

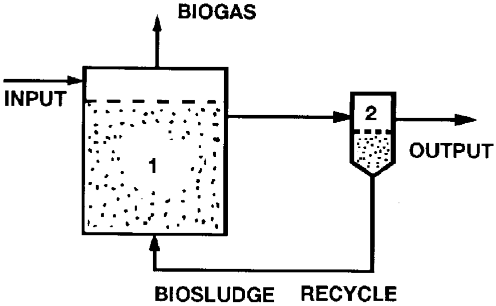

FIG. 1
Technological scheme of anaerobic waste water treatment plant: 1 anaerobic, 2 aerobic bioreactor

here to prepare the fuzzy knowledge data base that fitted to the input data structure of the expert system SENECA.

Generally, it assumes that the knowledge is represented by a vector of n independent state variables $x_1, \ldots x_n$ and one dependent variable Y. For the set of state variables $\boldsymbol{X}$, the fuzzy model of the process allocates the proper value of Y and its corresponding measure of uncertainty, a so-called grade membership.

In fact, the knowledge data base of the program SENECA is splitted into two data files[5]. The first one contains the description of all $\boldsymbol{X}$ and Y fuzzy set including an appropriate name and range of values. For example, when a name of fuzzy value of pH is PH5 then it means that the measured value of pH changes its value from 4.8 to 5.2. The second file contains the multivariable set of fuzzy numbers, coding the logical structure of following production rules:

$$\begin{array}{l}
\text{IF} \quad X_1 = A_{11} \quad \text{and} \quad \ldots \quad \text{and} \quad X_\text{n} = A_{1n} \quad \text{THEN} \quad Y = B_1 \quad \text{with} \quad W_1 \\
\quad \text{ELSE} \\
\text{IF} \quad X_1 = A_{21} \quad \text{and} \quad \ldots \quad \text{and} \quad X_n = A_{2n} \quad \text{THEN} \quad Y = B_2 \quad \text{with} \quad W_2 \\
\quad \text{ELSE} \\
\quad -------------- \\
\text{IF} \quad X_1 = A_{m1} \quad \text{and} \quad \ldots \quad \text{and} \quad X_n = A_{mn} \quad \text{THEN} \quad Y = B_m \quad \text{with} \quad W_m \, ,
\end{array} \tag{1}$$

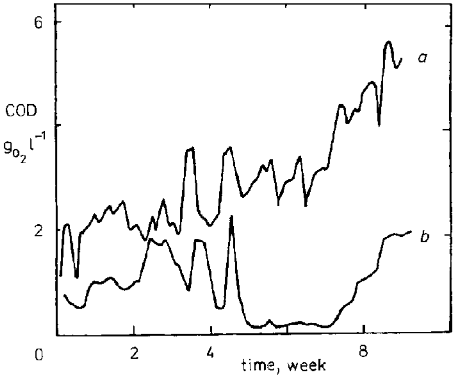

FIG. 2
Time course of chemical oxygen demand (COD) in input (*a*) and output (*b*) streams

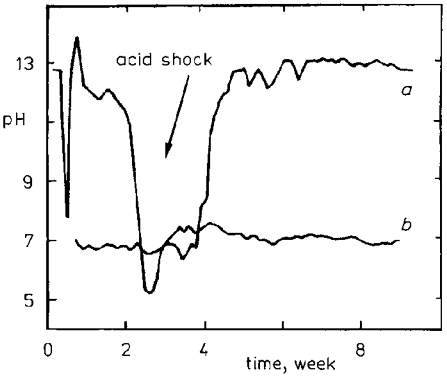

FIG. 3
Time course of pH in input (*a*) and output (*b*) streams. The unexpected leakage of xenobiotics is denoted as “acid shock”

where A_{ij}, B_i are the fuzzy values of j-th independent and dependent state variable in i-th statement, respectively. The symbol W_i is the weight of i-th statement (grade of membership) defined by user as a real number within 0 and 1. A lower value W_i means a higher uncertainty in information. When a value in the vectors of production rules A or B is missing, than it is denoted as "–" (unknown) in the date file.

The knowledge data base should be regularly updated when new data (production rules) are assembled. In our example, the knowledge base was a somewhat simple form of a question. For a chosen process time, pH in the bioreactor and for COD and pH on its input, the expert system returns the value of COD on the output from bioreactor. The formulation of proper production rule (question in antecedent and answer in consequent) follows from

$$\text{IF} \quad (\text{time} \ \& \ \text{pH}_{\text{input}} \ \& \text{pH}_{\text{output}} \ \& \ \text{COD}_{\text{input}}) \quad \text{THEN COD}_{\text{output}} \ . \tag{2}$$

Desired value of $\text{COD}_{\text{output}}$ is received by a calculation of Cartesian product of the expressions behind the key word "IF" and "THEN" with fuzzy sets defined in the first file of the knowledge data base. Dohnal[16] described an effective algorithm for the calculation of Cartesian product that was also used in our study.

It was useful to rewrite the original SENECA code from BASIC to Turbo Pascal programming language. It allows us to apply the sophisticated tools of object programming such as the Turbo Vision to make the expert system shell SENECA user friendly.

At first, according to Sterbacek and Votruba[15], we modified the pre-processing part of original computer code when we introduced the subroutine BASE which converts the input data from real numbers into fuzzy ones.

Secondly, we used the Turbo Pascal programming tools to create the brand-new system of flexible menu which permits:

a) To READ the vector of questions as a set of independent and dependent variables either from keyboard or from existing data files.
b) To ADD new production rules and extend the knowledge data base.
c) To BROWSE and EDIT input data.
d) To DRAW and DISPLAY the answers of expert system as a graph.

These new modifications essentially improved the performance of the original version of the fuzzy expert system shell SENECA and permitted its implementation on all types of IBM PC.

RESULTS AND DISCUSSION

The original industrial data, which were converted into fuzzy numbers and production rules by subroutine BASE, are depicted in Figs 2 and 3. The pH and COD values were sorted into 15 regular intervals that covered the range of experimental data. The effect of process time was analyzed for nine (one week long) intervals. In such a way, both knowledge base data files, which are the information input of expert system SENECA, were created. As we have explained above, the fuzzy modelling of the process is realized by repeated search of answers for a certain group of questions. In our case, the questions posed to SENECA have three different purposes.

Estimation of the Effect of pH_{input} on COD_{output}

The value of pH in input stream is the only one control variable which can be influenced by process operator. Therefore it is useful to evaluate its effect on COD_{output}. In this case the value of COD_{output} represents the efficiency of biological degradation of organic pollutants.

For the whole period of 9 weeks, prescribed pH_{input} in the range from 5 to 14, arbitrary COD_{input} and pH_{output}, we evaluated COD_{output} [g of oxygen/l]. Figure 4 shows that for pH_{input} in the range from 8 to 10 there exists a minimum in COD_{output}.

According to Dohnal[4,5] there are few ways how to represent the answer of expert system. We used the numerical representation based on the estimation of the center of gravity for the calculated Cartesian product. When the center of gravity is evaluated for the level of grade of membership 0.9, then we received an “optimistic” answer (dotted line). This approach simulates the behavior of a person which accentuated the “positive features” of the fuzzy answer. On the contrary, when evaluating the center of gravity for the level of grade of membership 0.1 (full line), then we obtain a “pragmatic” representation of the fuzzy answer because it simulates the behavior of a person that bears in mind most of existing knowledge.

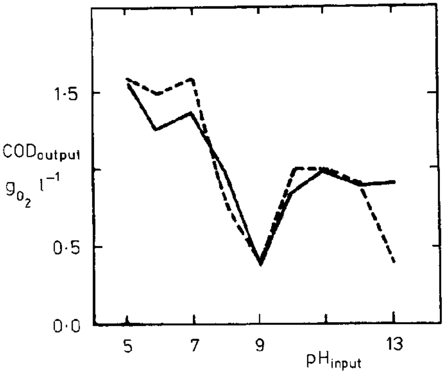

FIG. 4
Effect of pH_{input} on COD_{output}: (– – – –) “optimistic”, (——) “pragmatic” prediction

As it can be concluded from the results summarized in Fig. 4, the value of pH_{input} has to be kept close to 9 to achieve the maximum of biodegradability potential of microbial consortium responsible for biogas production.

Evaluation of Average COD_{output}

The second goal of our analysis was to estimate the average efficiency of anaerobic bioreactor. As shown in Fig. 1, the output from the anaerobic bioreactor is an input to the aerobic one. Therefore, it is useful to know how much of organic pollutants must be removed in the aerobic stage of waste water treatment process. The electrical power required for mixing and aeration in aerobic bioreactors is most frequently proportional to the COD value in the input.

For the whole process time and arbitrary pH in input and output streams, we evaluated COD_{output} as a function of COD_{input}. The results for the "pragmatic" and "optimistic" representation are summarized in Fig. 5. The statistical mean for the original set of data was 0.86 g of oxygen/l with confidence interval from 0.20 to 1.5. The "pragmatic" estimation (full line) was different and displayed much narrow range of COD_{output} from 1.2 to 1.3 g of oxygen/l. For the level of grade of membership up to 0.5, we received similar results. The "optimistic" prediction for the center of gravity at the level of grade of membership higher then 0.7 (dotted line) predicted a worse efficiency of biodegradation when compared with the "pragmatic" one. The threshold level of COD_{output} varies between 1.2 – 1.8 g of oxygen/l. Seemingly, the "optimistic" prediction of COD_{output} overestimates the risk of "acid shock" (see Fig. 3) that caused the decrease of biodegradation activity of anaerobic stage. Figure 5 shows that for the "pragmatic" prediction, which takes into account most of information, the correlation between COD_{output} and

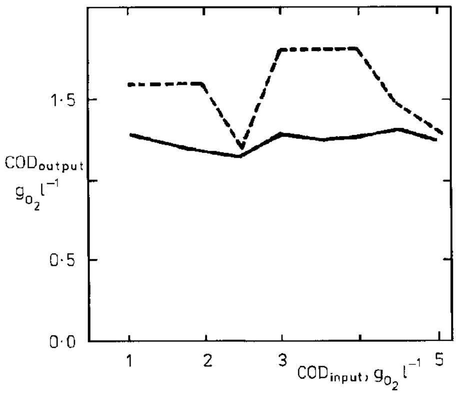

FIG. 5
Relation between average COD in input and output streams (symbols as in Fig. 4)

COD_{input} is insignificant. This can be explained by huge biodegradation potential of microbial consortium, which was limited by concentration of carbon source, only.

Estimation of the Process Dynamics

The dynamics of anaerobic digestor is a combination of very fast and slow processes[2,3]. The time constant of physico-chemical processes such as acid-base reactions or liquid-gas mass transfer may be estimated as a few seconds. On the contrary, the time constant of slow processes such as the growth of at least three physiologically different groups of microorganisms has to be considered as a few days. That dynamics of microbial consortium determine the observed rate of biodegradation of organic pollutants. When the dynamics are not taken into account, the anaerobic bioreactor may be overloaded and the microbial activity depressed or eliminated. Therefore, we used the expert system SENECA to estimate the magnitude of that "biological bottlenecks" of process dynamics as time dependence of COD_{output}.

For the chosen process time, arbitrary COD_{input}, pH_{input} and PH_{output}, we evaluated the COD_{output}. The results summarized in Fig. 6 show that the worst efficiency of cleaning process was between 3rd and the 4th week of plant operation. Probably, this loss of biodegradation activity of microbial consortium was caused by an unexpected massive leakage of xenobiotics. The xenobiotics were used during the start-up period of the sugar factory for cleaning and induced the "acid shock" in pH_{input} (see Fig. 3).

Figure 6 shows that the value of COD_{output} achieved its minimum between the 6th and 7th week. The apparent time constant describing the recovery of "acid shocked" microbial consortium may be estimated as one and a half week.

The increase of COD_{output} in the 8th and 9th week may be adjudged to the overloading of the anaerobic bioreactor. On the other hand, the total bioconversion of organic pollutant increased in the same time.

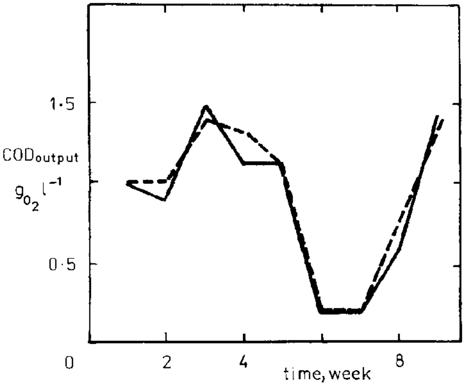

Fig. 6
Process dynamics excerpted from industrial records (symbols as in Fig. 4)

When analyzing the effect of COD_{input} on the process dynamics, we found that the fluctuations in COD_{input} were unimportant. Figure 6 shows that the differences between the "optimistic" (dashed line) and "pragmatic" (full line) strategy of the process dynamics evaluation are unimportant.

CONCLUSION

When we began to analyze the anaerobic waste water treatment process using the fuzzy modelling, we did not assume that the information technology was so powerful. Owing to the incompleteness of data, we have to ignore the deterministic models based on the mass and energy balances and to anticipate a new vision of information that was diluted in process records. We tried to "dig out" the information as heuristic rules from industrial records. The expert systems help us to excerpt the rules and they represent the arranged information – knowledge. The models of process analysis used here may be convenient for the self-learning process control and optimization when the knowledge data base is updated. The "aging of information" may easily performed by step-wise decrease of grade of membership when the knowledge data base is updated.

REFERENCES

1. Sobotka M., Votruba J., Havlik I., Minkevich I. G.: Folia Microbiol. (Prague) *28*, 195 (1983).
2. Havlik I., Votruba J., Sobotka M., Volesky B.: Biotechnol. Lett. *6*, 607 (1984).
3. Havlik I., Votruba J., Sobotka M.: Folia Microbiol. (Prague) *31*, 56 (1986).
4. Dohnal M.: Biotechnol. Bioeng. *27*, 1146 (1985).
5. Dohnal M.: Collect. Czech. Chem. Commun. *51*, 1027 (1986).
6. Budicek L., Rejsek J., Kubin M., Spoustova I.: Listy Cukrovarnicke *105*, 58 (1989).
7. Rejsek J., Budicek L., Spoustova I.: Listy Cukrovarnicke *105*, 151 (1989).
8. Budicek L., Rejsek J., Tinl J., Kubin M., Navratilova I.: Listy Cukrovarnicke *106*, 27 (1990).
9. Budicek L., Dufek K., Rejsek J., Kubin M., Navratilova I.: Listy Cukrovarnicke *106*, 135 (1990).
10. Rejsek J., Budicek L., Kubin M., Navratilova I.: Listy Cukrovarnicke *106*, 193 (1990).
11. Stuchl I., Budicek L., Kubin M., Navratilova I.: Listy Cukrovarnicke *107*, 34 (1991).
12. Stuchl I., Rejsek J., Budicek L., Kubin M., Navratilova I.: Listy Cukrovarnicke *107*, 201 (1991).
13. Hoffmann P., Havranek M., Cuta J., Chalupa J., Madera V., Hamackova J., Kohout M.: *Jednotne metody chemickeho rozboru vod*, p. 111. SNTL, Praha, 1965.
14. Dubois D., Prade H.: *Theorie des Possibilites*, p. 102. Masson, Paris, 1988.
15. Sterbacek Z., Votruba J.: Chem. Eng. J. *51*, B35 (1993).
16. Dohnal M.: Strojnicky Casopis *37*, 333 (1986).

Translation revised by J. Linek.