Elektrotehniški vestnik XX(Y): 1–5, YEAR
Electrotechnical Review, Ljubljana, Slovenija

Monitoring and sensor fault detection in a waste-water treatment process based on a fuzzy model

Dejan Dovžan[1], Vito Logar[2], Nadja Hvala[3], Igor Škrjanc[4]

[1,2,4] *Faculty of Electrical Engineering, Tržaška 25, 1000 Ljubljana*
[2] *Inštitut Jožef Štefan, Jamova cesta 39, 1000 Ljubljana*
E-mail: [1] *dejan.dovzan@fe.uni-lj.si*

Abstract. In this paper, monitoring and sensor fault detection in a waste-water treatment process are discussed. Monitoring is based on the Takagi-Sugeno fuzzy model of a plant process obtained by using Gustafson-Kessel fuzzy clustering algorithm.The paper also explains the principle of the Takagi-Sugeno fuzzy model. The main idea is to cope with the non-linearity of a monitored process. The output of the fuzzy-model in normal operation regime is compared with the current behavior. If the fault-detection index exceeds a certain predefined value (the fault-tolerant index), an alarm is triggered. The data treated in this paper are obtained with a simulation model of a waste-water treatment plant and by simulating sensor faults. The signals to be measured in the process monitoring are the following: influent ammonia concentration, dissolved-oxygen concentration in the first aerobic reactor tank, temperature, dissolved-oxygen concentration and ammonia concentration in the second aerobic reactor.

Key words: fuzzy clustering, fuzzy modelling, waste-water treatment plant, process monitoring, fault detection

1 Introduction

Recently, fault detection and management became a very popular area in the process industry. Fault-detection methods are mainly based either on a process model, expert system, statistical signal processing [3, 4] or on pattern recognition techniques [1, 2]. The today's fast and cheep sensors allow us to measure a vast amount of process variables on-line. By processing these signals, process monitoring can be used to evaluate the current performance of the process and to detect its early faults.

In this paper, sensor fault detection in a waste water treatment process (WWT) [5, 6, 7] is discussed. WWT processes are of a nonlinear nature with time-varying dynamics and relations changing on a daily, monthly and seasonal basis. They are affected by the outside air temperature, amount of rain and varying loads. Therefore, theoretical modelling of a process is a complex and difficult task potentially leading to questionable results. In this paper we propose fault detection method based on a fuzzy model. The fuzzy model is able to approximate a nonlinear system more accurately than a linear model, thus reducing the number of false alarms. To identify the number of clusters the, Gustafson-Kessel clustering algorithm was used and to identify the model parameters the least-square algorithm was used. After being identified on a set of training data, the model was used to monitor a WWT and to detect a simulated fault on a sensor.

Received d
Accepted d

1.1 Fuzzy model and Gustafson-Kessel clustering

In this section the methods used in data analyzing are explained. First the Gustafson-Kessel (GK) clustering algorithm is explained then the Takagi-Sugeno (TS) fuzzy model is derived and identified.

1.1.1 Gustafson-Kessel clustering algorithm

The GK clustering algorithm is used to identify clusters of different shapes. This is convenient for a WWT process with this kind of shapes present. The input data matrix is given as:

$$X \in \mathbb{R}^{n\times p}. \qquad (1)$$

The input vector is defined as:

$$x_k = [x_{k1}, \ldots, x_{kp}]\,,\; x_k \in \mathbb{R}^p. \qquad (2)$$

The set of n measurements is denoted as:

$$X = \{x_k \mid k = 1, 2, \ldots, n\} \tag{3}$$

and can be presented as $n \times p$ matrix:

$$X = \begin{bmatrix} x_{11} & x_{12} & \ldots & x_{1p} \\ x_{21} & x_{22} & \ldots & x_{2p} \\ \vdots & \vdots & \vdots & \vdots \\ x_{n1} & x_{n2} & \ldots & x_{np} \end{bmatrix}. \tag{4}$$

The main objective of clustering is to partition a set of data X into c partitions called clusters. The fuzzily partitioned set of data X is combined of fuzzy subsets (clusters) $\{A_i \mid 1 \le i \le c\}$. They are defined with the membership functions implicitly defined in the fuzzy partition matrix $U = [\mu_{ik}] \in \mathbb{R}^{c\times n}$. The i-th row of the matrix contains the membership degree of the i-th cluster A_i of data set X. The partition matrix satisfies the following conditions: the membership degrees are real numbers from the interval $\mu_{ik} \in [0,1]$, $1 \le i \le c$, $1 \le k \le n$, the total membership degree of sample x_k to all clusters is one ($\sum_{i=1}^{c} \mu_{ik} = 1$, $1 \le k \le n$), none of the clusters is neither empty nor containing all data ($0 < \sum_{k=1}^{n} \mu_{ik} < n$, $1 \le i \le c$). This means that the fuzzy partition matrix belongs to a fuzzy partition set defined as:

$$\begin{gathered} M = \{U \in \mathbb{R}^{c\times n} \mid \mu_{ik} \in [0,1], \forall i, k; \\ \sum_{i=1}^{c} \mu_{ik} = 1, \forall k;\ 0 < \sum_{k=1}^{n} \mu_{ik} < n,\ \forall i\}. \end{gathered} \tag{5}$$

The fuzzy partition matrix is obtained by applying the clustering method on a data-set matrix. The clustering algorithm is obtained by minimizing the fuzzy c-means criterion function using the constraint from Eq. 5:

$$J(X, U, V, \lambda) = \sum_{i=1}^{c}\sum_{k=1}^{n} \mu_{ik}^m d_{ik}^2 + \lambda \sum_{i=1}^{c}\sum_{k=1}^{n} (\mu_{ik} - 1), \tag{6}$$

where U is the fuzzy partition matrix of data matrix X, V is the vector of cluster centres

$$V = [v_1, v_2, \ldots, v_c],\ v_i \in \mathbb{R}^p, \tag{7}$$

d_{ik}^2 is the distance norm

$$d_{ik}^2 = (x_k - v_i)^T A_i (x_k - v_i).$$

Matrix A_i is defined as:

$$A_i = (\rho_i det(C_i))^{1/p} C_i^{-1},$$

where $\rho_i = 1, i = 1, ..., c$ and p is equal to the number of the measured variables and C_i is the fuzzy covariance matrix of the i-th cluster:

$$C_i = \frac{\sum_{k=1}^{n} \mu_{ik}^m (x_k - v_i)(x_k - v_i)^T}{\sum_{k=1}^{n} \mu_{ik}^m}.$$

This allows us to detect hiper-ellipsoidal clusters in the data distribution. If the data are distributed along nonlinear hyper-surface, the algorithm will find the clusters that are local linear approximations of this hyper-space. Overlapping of clusters is defined with fuzziness factor $m \in [1, \infty)$.

The number of clusters is defined by using cluster validity functions or iterative insertion and merging clusters depending on the model error. Factor m effects fuzziness of the cluster: from crisp $m = 1$ to completely fuzzy $m \to \infty$. In our example, the standard value $m = 2$ was used.

1.1.2 Steps of Gustafson-Kessel clustering algorithm

The GK clustering algorithm can be described with the following steps:

- **Initialization** Set the number of clusters c, define the overlapping/fuzziness factor m (usually $m = 2$) and stopping error $\epsilon_{end} > 0$ (in our case $\epsilon_{end} = 0.001$). Random initialization of fuzzy partition matrix $U \in M$. Epoch $r = 0$.
- **Loop**

$$r = r + 1$$

computation of the cluster center positions:

$$v_i^{(r)} = \frac{\sum_{k=1}^{n} \left(\mu_{ik}^{(r)}\right)^m x_k}{\sum_{k=1}^{n} \left(\mu_{ik}^{(r)}\right)^m},\ 1 \le i \le c. \tag{8}$$

computation of fuzzy covariance matrices and inner-product distance norm A_i:

$$C_i = \frac{\sum_{k=1}^{n} \mu_{ik}^m (x_k - v_i)(x_k - v_i)^T}{\sum_{k=1}^{n} \mu_{ik}^m}, \tag{9}$$

$$A_i = (\rho_i det(C_i))^{1/p} C_i^{-1},\ 1 \le i \le c \tag{10}$$

computation of the distance from the cluster centers

$$\begin{gathered} d_{ik}^2 = \left(x_k - v_i^{(r)}\right)^T A_i \left(x_k - v_i^{(r)}\right), \\ 1 \le i \le c,\ 1 \le k \le n \end{gathered} \tag{11}$$

updating of the partition matrix:

$$\text{if } d_{ik} > 0, \quad \mu_{ik}^{(r)} = \frac{1}{\sum_{j=1}^{c} \left(\frac{d_{ik}}{d_{jk}}\right)^{\frac{2}{m-1}}} \tag{12}$$

- **until** $||U^{(r)} - U^{(r-1)}|| < \epsilon_{end}$

1.1.3 The Takagi-Sugeno fuzzy model

The TS fuzzy model approximates the nonlinear system by using smoothly interpolating local linear models. Each local model contributes to the global model output depending on the membership degrees of the current input vector. We assume a set of input vectors:

$$X = [x_1, x_2, \ldots, x_n]^T \tag{13}$$

and a set of outputs

$$Y = [y_1, y_2, \ldots, y_n]^T. \tag{14}$$

A typical fuzzy model is given in the form of rules R_i:

$$\begin{aligned} &R_i: \\ &\text{If } x_k \text{ is } A_i \text{ then } \hat{y}_k = \phi_i(x_k),\ i = 1, \ldots, c \end{aligned} \tag{15}$$

x_k denotes the input vector (variables of premise), $\hat{y}_k$ is the output of a local linear model at time instant k. Input vector x_k belongs to each fuzzy subset $(A_1, \ldots, A_c)$ with a current membership degree $\mu_{A_i}(x_k)$ or $\mu_{ik} : \mathbb{R} \to [0, 1]$. Functions $\phi_i(\cdot)$ are arbitrary smooth functions, although linear or affine functions are normally used. The global model output is calculated as:

$$\hat{y}_k = \frac{\sum_{i=1}^{c} \mu_{ik} \phi_i(x_k)}{\sum_{i=1}^{c} \mu_{ik}}. \tag{16}$$

To simplify Eq. (16), partitioning the unity is considered, where the function of $\beta_i(x_k)$ (Eq. (17)) gives information of fulfilment of the respective fuzzy rule in a normalized form.

$$\beta_i(x_k) = \frac{\mu_{ik}}{\sum_{i=1}^{c} \mu_{ik}},\ i = 1, \ldots, c \tag{17}$$

The sum of this fulfilment over the clusters is one ($\sum_{i=1}^{c} \beta_i(x_k) = 1$) irrespective of x_k, as long as the denominator of $\beta_i(x_k)$ is not zero. This can be easily achieved by properly defining membership functions. Combining Eqs. (16) and (17), the following equation can be derived:

$$\hat{y}_k = \sum_{i=1}^{c} \beta_i(x_k) \phi_i(x_k),\ k = 1, \ldots, n \tag{18}$$

The local-model output is usually defined as a linear combination of the input vector:

$$\phi_i(x_k) = x_k \theta_i, \quad i = 1, \ldots, c,$$

$$\theta_i^T = \left[\theta_{i1}, \ldots, \theta_{i(p+q)}\right]. \tag{19}$$

The vector of fuzzified input variables at time instant k is defined as:

$$\psi_k = [\beta_1(x_k) x_k, \ldots, \beta_c(x_k) x_k],\ k = 1, \ldots, n, \tag{20}$$

the fuzzified data matrix is then written as:

$$\Psi^T = \left[\psi_1^T, \psi_2^T, \ldots, \psi_n^T\right]. \tag{21}$$

The matrix of the whole set of rules can be written as:

$$\Theta^T = \left[\theta_1^T, ..., \theta_c^T\right], \tag{22}$$

The global-model output (Eq. (18)) can then be written in a matrix form:

$$\hat{y}_k = \psi_k \Theta. \tag{23}$$

The relation of the input vectors and outputs can be written in a compact form:

$$\hat{Y} = \Psi \Theta, \tag{24}$$

where $\hat{Y}$ stands for the vector of model outputs $\hat{y}_k$ $(k = 1, \ldots, n)$

$$\hat{Y} = [\hat{y}_1, \hat{y}_2, \ldots, \hat{y}_n]^T. \tag{25}$$

The TS fuzzy model given with equation (23) is also called the affine TS model. The model is able to approximate an arbitrary nonlinear function with the desired degree of accuracy [8, 9, 10]. The generality can be proven with the Stone-Weierstrass theorem [11], suggesting that any continuous function can be approximated by a fuzzy basis function expansion [12].

1.1.4 Estimation of local linear parameters

To estimate local linear model parameters, the least-square method is used. Measurements satisfy the nonlinear equation of the system:

$$y_i = g(x_i), \quad i = 1, \ldots, n \tag{26}$$

According to the Stone-Weierstrassevm theorem, for any given function g on a compact set $U^c \subset \mathbb{R}^p$, there exists a fuzzy system f such that:

$$\max_{x_i \in X} |f(x_i) - g(x_i)| < \delta, \quad \forall i, \tag{27}$$

where $\delta > 0$ is an arbitrary small constant. When a continuous function is approximated with a fuzzy function from class $\mathcal{F}^p$, defined in Eq. (23), it should be noted that lower vales of δ imply higher values of clusters c to satisfy Eq. (27).

The error between the function (process outputs) and fuzzy approximation (fuzzy-model outputs) can be defined as:

$$e_i = y_i - f(x_i) = y_i - \hat{y}_i, \quad i = 1, ..., n, \tag{28}$$

where y_i stands for the measured output and $\hat{y}_k$ for the fuzzy model output at time instant k. The parameters of the proposed fuzzy function (Θ) are estimated by minimizing the sum of the squared errors over the whole input set of data:

$$\begin{aligned} E = \sum_{i=1}^{n} e_i^2 = \\ = (Y - \hat{Y})^T (Y - \hat{Y}) == (Y - \Psi\Theta)^T (Y - \Psi\Theta). \end{aligned} \tag{29}$$

Parameter Θ is obtained $\frac{\partial E}{\partial \Theta} = 0$:

$$\Theta = \left(\Psi^T \Psi\right)^{-1} \Psi^T Y.$$

2 Biological waste-water treatment process

The WWT plants are large nonlinear systems subject to large perturbations. Their dynamics and nonlinearity depend on outside air temperature, waste-water inflow and composition and other factors. A simulation benchmark has been developed for an unbiased evaluation of different control schemes. It consists of five sequentially connected reactors along with a 10-layer secondary settling tank. The plant layout, model equations and control strategy are described in detail on the web page (http://www.ensic.unancy. fr/costwwtp). In our approach waste-water is purified in a mechanical phase and a moving bed bio-film reactor is used. A schematic presentation of the used simulation benchmark is shown in Fig. 1. Signals used to build our

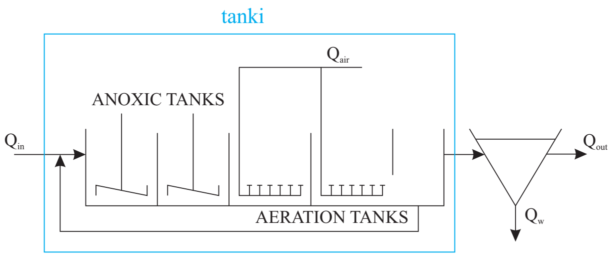

Figure 1: Schematic presentation of the simulation benchmark.

fuzzy model were: influent ammonia concentration in inflow Q_{in} defined as $C_{NH4N_{in}}$, dissolved-oxygen concentration in the first aerobic reactor tank $C^1_{O_2}$, dissolved-oxygen concentration in the second aerobic reactor tank $C^2_{O_2}$ and ammonia concentration in the second aerobic reactor tank $C_{NH4N_{out}}$. The model was implemented to approximate the relation between the ammonia concentration in the second aerobic reactor tank and the other measured variables:

$$C_{NH4N_{out}}(k) = \mathcal{G}\left(C_{NH4N_{in}}(k), C^1_{O_2}(k), C^2_{O_2}(k)\right), \tag{30}$$

where $\mathcal{G}$ stands for nonlinear relation between the measured variables. The whole set of measurements is shown in Fig. 2. The sampling time of the process was 120 s. The first 15000 samples were used

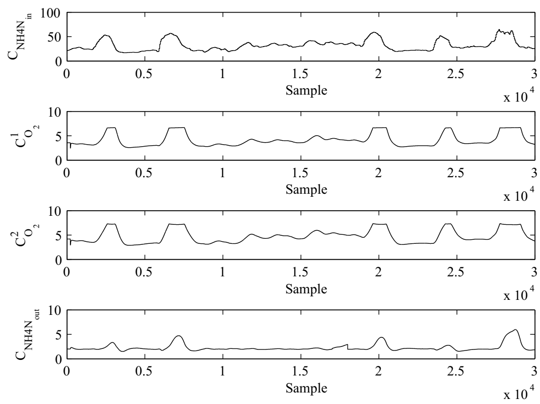

Figure 2: Whole set of measurements. The influent ammonia concentration $C_{NH4N_{in}}$, dissolved-oxygen concentration in the first aerobic reactor tank $C^1_{O_2}$, dissolved-oxygen concentration in the second aerobic reactor tank $C^2_{O_2}$ and ammonia concentration in the second aerobic reactor tank $C_{NH4N_{out}}$.

to identify the model. The output of the identified fuzzy model ($\hat{C}_{NH4N_{out}}$) and the process output ($C_{NH4N_{out}}$) are shown in Fig. 3. The identification stage was also used to calculate the threshold for alarm activation. To detect faults, the fault-detection index is defined as:

$$f = \left(\frac{C_{NH4N_{out}} - \hat{C}_{NH4N_{out}}}{\hat{C}_{NH4N_{out}}}\right)^2. \tag{31}$$

The fault-tolerance index is defined as a relative degree of the maximal value of the fault detection index in the identification phase $f_{tol} = \gamma \max f$. In our case, we chose $\gamma = 1.5$. Fault tolerance-index was $f_{tol} = 0.15$. The fault-detection index is shown in Fig. 4. The alarm is triggered when the fault-detection index is higher than the fault-tolerance index.

The fault was simulated at time sample 17000 on the $C_{NH4N_{out}}$ signal. A signal with an exponentially

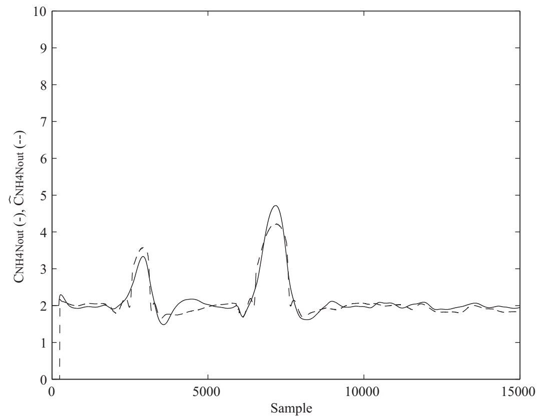

Figure 3: Verification of the developed fuzzy model with fuzzy-model output $\hat{C}_{NH4N_{out}}$ and process output $C_{NH4N_{out}}$ shown.

increasing value was added to the nominal signal in order to simulate the fault. The slowly increasing fault was eliminated at time sample 18000. The fault was detected at time sample 17556. Detection was a bit delayed, as usually experienced with slowly increasing faults.

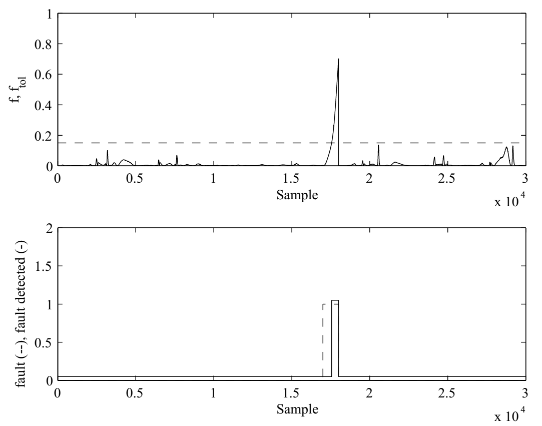

Figure 4: Fault-detection index, fault-tolerance index f_{tol} and actual and detected fault

3 Conclusion

In this paper, fault detection on a sensor in a WWT process is discussed. A fault-detection system was realized by using the Takagi-Sugeno fuzzy model. The model was identified on basis of the Gustafson-Kessel method for clustering and the least-square method for local-model parameter identification. The proposed concept was tested on a simulated model of a WWT process, with a fault simulated on one of the senors. The measurements used for building our fuzzy model were: influent ammonia concentration, dissolved-oxygen concentration in the first aerobic reactor tank, temperature, dissolved-oxygen concentration and ammonia concentration in the second aerobic reactor. Fault occurring on the ammonia concentration sensor in the second aerobic reactor was detected with no false alarms and with a small time-delay because of the nature of the fault.

Since the process dynamics and nonlinearity change depending on many factors (such as load, amount of rain, etc.), on-line fuzzy identification should be adopted. By employing the on-line identification method the fuzzy model should be able to adapt to new process dynamics thus increasing the accuracy of the model estimation.

4 References

[1] Chen, J., Liao, C. M., “Dynamic process fault monitoring based on neural network and PCA”. *Jour. of Process Control*, vol. 12, pp. 277–289, 2002.

[2] Klančar, G., Ð. Juričić, R. Karba., “Robust fault detection based on compensation of the modelling error”. *International journal of Systems Science*, vol. 33, no. 2, pp. 97–105, 2002.

[3] Johnson, R. A., Wichern, D. W., *Applied Multivariate Statistical Analysis*, Prentice-Hall, New Jersey, 1992.

[4] Daszykowski, M., Walczak, B., Massart, D. L., “Projection methods in chemistry”. *Chemometrics and Intelligent Laboratory Systems*, vol. 65, pp. 97–112, 2003.

[5] Vrečko, D., Hvala, N., Kocijan, J., Zec, M. “System analysis for optimal control of a wastewater treatment benchmark”. *Water sci. technol.*, vol. 43, pp. 199–206, 2001.

[6] Vrečko, D., Hvala, N., Kocijan, J., “Wastewater treatment benchmark : What can be achieved with simple control?”. *Water sci. technol.*, vol. 45, pp. 127–134, 2002.

[7] Hvala, N., Vrečko, D., Burica, O., Stražar, M., Levstek, M., “Simulation study supporting wastewater treatment plant upgrading”. *Water sci. technol.*, vol. 46, pp. 325–332, 2002.

[8] Kosko, B., “Fuzzy Systems as Universal Approximators”, *IEEE Transactions on Computers*, vol. 43, no. 11, pp. 1329–1333, 1994.

[9] Ying, H. GC., “Necessary conditions for some typical fuzzy systems as universal approximators”, *Automatica*, vol. 33, pp. 1333–1338, 1997.

[10] Wang, L.-X., Mendel, J. M., “Fuzzy basis functions, universal approximation, and orthogonal least-squares learning”, *IEEE Trans. Neural Networks*, vol. 3, no. 5, pp. 807–814, 1992.

[11] Goldberg, R. R., *Methods of Real Analysis*, John Wiley and Sons, 1976.

[12] Lin, C-H., “Siso nonlinear system identification using a fuzzy-neural hybrid system”, *Int. Jour. of Neural Systems*, vol. 8, no. 3, pp. 325–337, 1997.