

Zaznavanje napak in spremljanje čiščenja odpadnih voda na podlagi mehkega modela

Dejan Dovžan¹, Vito Logar², Nadja Hvala³, Igor Škrjanc⁴

^{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

Povzetek. V članku je predstavljen sistem za spremljanje in zaznavanje napak na procesu čištilne naprave. Zaradi nelinearnosti procesa smo kot osnovo za sistem spremljanja uporabili mehki model, s katerim lahko zelo dobro aproksimiramo nelinearnosti procesa in se s tem izognemo nepotrebnim lažnim alarmom. Model je dobljen s pomočjo Gustafson-Kesselovega algoritma za rojenje. Odziv mehkega modela v normalnem obratovanju je primerjan s trenutnim dogajanjem v procesu. Na podlagi odstopanja odzivov lahko zaznamo napako. Proces v našem primeru simuliramo. Simulirane pa so tudi napake na senzorjih. Merjeni signali so: vhodna koncentracija amonijaka in koncentracija kisika v prvem aerobnem tanku ter temperatura, koncentracija kisika in amonijaka v drugem tanku. Prikazani so rezultati spremljanja procesa in odkrivanja napak na podlagi mehkega modela.

Ključne besede: mehko rojenje, graditev mehkega modela, čiščenje odpadnih voda, spremljanje procesa, zaznavanje napak

1 Uvod

Spremljanje procesov skupaj z zaznavanjem napak in diagnozo je v zadnjem času postalo eno popularnejših področij procesne avtomatike. Za spremljanje procesa in odkrivanje napak se pogosto uporabljajo tehnike na podlagi modela procesa, ekspertni sistemi in sistemi za razpoznavanje vzorcev [1, 2]. Hitri senzorji in oprema za zajem podatkov nam omogočajo zajem velike količine podatkov, ki nam kažejo, kaj se dogaja s procesom. Za obdelavo in analizo zajetih podatkov je bilo razvitih veliko statističnih metod [3, 4]. S pomočjo meritev lahko ves čas spremljamo procese in zaznavamo napake ter temu primerno prožimo alarme. V našem primeru obravnavamo zaznavanje napake na senzorju simuliranega procesa čiščenja odpadne vode [5, 6, 7]. Ti procesi so zaradi dnevnih, mesečnih in sezonskih vplivov izpostavljeni nenehnemu spreminjanju dinamike. Nanje vplivajo sprememba temperature, količina padavin in različne obremenitve naprave. Teoretično modeliranje je pri čiščenju odpadnih voda zelo zapleten postopek, katerega rezultati so dvomljivi. Zato se za spremljanje takih procesov uporabljajo statistične metode, ki temeljijo na rudarjenju s podatki. Čiščenje odpadnih voda je zelo nelinearen proces, zato je treba njegovo nelinearnost upoštevati v modelu. Za predpro-

cesiranje podatkov smo v našem primeru uporabili algoritem za mehko rojenje podatkov. S tem smo se izognili napačnim alarmom, ki lahko nastanejo zaradi nelinearnosti procesa (če bi uporabili linearni model za zaznavanje), saj z mehkim modelom lahko poljubno natančno aproksimiramo nelinearnosti.

1.1 Mehki model z Gustafson-Kesselovim rojenjem

V tem poglavju bomo opisali algoritme in metode, ki so bile uporabljene za analizo podatkov. Razložen bo Gustafson-Kesselov algoritem rojenja in izpeljan ter identificiran Takagi-Sugeno mehki model procesa čiščenja odpadnih voda.

1.1.1 Gustafson-Kesselov algoritem mehkega rojenja

Gustafson-Kesselov algoritem rojenja nam omogoča, da definiramo roje različnih oblik, kar nam pri takih procesih, kot je obravnavani, pride prav.

Matrika vhodnih podatkov je podana kot

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

Vektor vhodnih podatkov v trenutku k je definiran kot

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

Nabor n meritev označimo kot

$$X = \{x_k \mid k = 1, 2, \dots, n\} \quad (3)$$

in je predstavljen kot matrika $n \times p$:

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

Cilj rojenja je razdeliti nabor podatkov X v c podmnožic, ki jih imenujemo roji. Mehko razdeljen nabor podatkov X je skupek mehkih podmnožic (rojov) $\{A_i \mid 1 \leq i \leq c\}$. Roji so definirani s pripadnostnimi funkcijami, ki so implicitno vsebovane v pripadnostni matriki $U = [\mu_{ik}] \in \mathbb{R}^{c \times n}$. i -ta vrsta matrike U vsebuje vrednosti pripadnostne funkcije i -tega roja A_i iz množice podatkov X . Matrika U izpolnjuje naslednje pogoje: pripadnostne stopnje so realna števila iz intervala od nič do ena ($\mu_{ik} \in [0, 1]$, $1 \leq i \leq c$, $1 \leq k \leq n$), pripadnost vzorca x_k k uniji rojev je ena ($\sum_{i=1}^c \mu_{ik} = 1$, $1 \leq k \leq n$), noben roj ni prazen oziroma ne vsebuje vseh podatkov ($0 < \sum_{k=1}^n \mu_{ik} < n$, $1 \leq i \leq c$). Zgoraj naštetu pomeni, da matrika U pripada množici, ki je definirana kot:

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

V našem primeru je matrika pripadnostni (U) dobljena z mehkim "c-means" algoritmom, ki temelji na Mahalanobisovi normi. Algoritem dobimo z minimizacijo kriterijske funkcije ob upoštevanju omejitev iz En. 5:

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

kjer je U matrika pripadnostni množice podatkov X , V je vektor centrov rojev

$$V = [v_1, v_2, \dots, v_c], \quad v_i \in \mathbb{R}^p, \quad (7)$$

d_{ik}^2 pa je kvadrat razdalje

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

Matrika A_i je definirana kot:

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

kjer je $\rho_i = 1$, $i = 1, \dots, c$ in p enak številu merjenih spremenljivk. C_i je mehka kovariančna matrika i -tega roja, definirana kot

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

Slednje nam omogoča detekcijo hiperelipsoidnega roja v porazdelitvi podatkov. Če so podatki porazdeljeni ob nelinearni hiperpovršini, algoritem najde roje, ki so lokalni linearni aproksimatorji hiperpovršine. Prekrivanje rojev je definirano s parametrom $m \in [1, \infty)$.

Za določanje števila rojev lahko uporabimo funkcije za merjenje veljavnosti rojev ali pa jih iterativno dodajamo oziroma odstranjujemo glede na odstopanje modela. Parameter prekrivanja m vpliva na mehkost roja; od trdega ($m = 1$) do čisto mehkega ($m \rightarrow \infty$). V našem primeru smo parameter nastavili $m = 2$, kar je tudi običajna vrednost tega parametra.

1.1.2 Koraki Gustafson-Kesselovega algoritma

Postopek Gustafson-Kesselovega algoritma za rojenje nabora podatkov X lahko opišemo z naslednjimi koraki:

- **Inicializacija** Izberemo število rojev c , faktor prekrivanja m (v našem primeru $m = 2$) in napako, pri kateri se algoritem konča $\epsilon_{end} > 0$ (v našem primeru $\epsilon_{end} = 0.001$). Inicializiramo matriko pripadnosti: $U \in M$ (naključno) in epoch $r = 0$.

- **Zanka**

$$r = r + 1$$

izračun centrov rojev:

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

izračun kovariančne matrike in matrike A_i :

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

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

izračun razdalje vzorcev do centrov rojev

$$d_{ik}^2 = \left(x_k - v_i^{(r)}\right)^T A_i \left(x_k - v_i^{(r)}\right), \quad (11)$$

$$1 \leq i \leq c, \quad 1 \leq k \leq n$$

osvežitev matrike pripadnosti:

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

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

1.1.3 Mehki model Takagi-Sugeno

Mehki TS-modeli aproksimirajo nelinearen sistem z interpolacijo lokalnih linearnih modelov. Vsak lokalni linearni model prispeva svoj delež k izhodu globalnega modela. Če imamo nabor vhodnih vektorjev

$$X = [x_1, x_2, \dots, x_n]^T \quad (13)$$

in nabor pripadajočih izhodov

$$Y = [y_1, y_2, \dots, y_n]^T, \quad (14)$$

potem je mehki model podan v obliki pravil R_i :

$$R_i : \quad (15)$$

če je x_k iz A_i potem $\hat{y}_k = \phi_i(x_k)$, $i = 1, \dots, c$

Vektor x_k je vektor vhodnih podatkov, \hat{y}_k je izhod lokalnega linearnega modela v trenutku k . Vektor x_k pripada mehkim množicam (A_1, \dots, A_c) z določeno pripadnostjo $\mu_{A_i}(x_k)$ oziroma $\mu_{ik} : \mathbb{R} \rightarrow [0, 1]$. Funkcije $\phi_i(\cdot)$ so lahko poljubne zvezne funkcije, čeprav so največkrat uporabljene linearne afine funkcije.

Izhod globalnega modela izračunamo po enačbi:

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

En. (16) poenostavimo tako, da definiramo funkcijo:

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

Funkcija $\beta_i(x_k)$ nam podaja normirano pripadnost vzorca k posameznemu roju. S tem nam pove tudi, kakšna je veljavnost pravila. Vsota funkcije po rojih je ena ($\sum_{i=1}^c \beta_i(x_k) = 1$) ne glede na x_k , če je imenovalec $\beta_i(x_k)$ različen od nič. To pa je lahko zagotoviti s pravilno definicijo pripadnostnih funkcij nad roji. Z združitvijo enačb (16) in (17) pridemo do slednje poenostavljene enačbe za izhod globalnega modela:

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

Izhod lokalnega modela je običajno definiran kot linearna kombinacija elementov podatkovnega vektorja:

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

$$\theta_i^T = [\theta_{i1}, \dots, \theta_{i(p+q)}]. \quad (19)$$

Vektor zmeščanih vhodov v trenutku k zapišemo kot:

$$\psi_k = [\beta_1(x_k)x_k, \dots, \beta_c(x_k)x_k], \quad k = 1, \dots, n, \quad (20)$$

matriko zmeščanih podatkov pa kot:

$$\Psi^T = [\psi_1^T, \psi_2^T, \dots, \psi_n^T]. \quad (21)$$

Če zapišemo matriko koeficientov celotnega nabora pravil kot

$$\Theta^T = [\theta_1^T, \dots, \theta_c^T], \quad (22)$$

potem lahko izhod globalnega modela (En. (18)) zapišemo v matrični obliki:

$$\hat{y}_k = \psi_k \Theta. \quad (23)$$

Relacija izhoda modela in vhodnih podatkov za celoten nabor zapišemo kot:

$$\hat{Y} = \Psi \Theta, \quad (24)$$

kjer je \hat{Y} vektor izhodov modela \hat{y}_k ($k = 1, \dots, n$)

$$\hat{Y} = [\hat{y}_1, \hat{y}_2, \dots, \hat{y}_n]^T. \quad (25)$$

Mehki model, podan z enačbo En. (23), imenujemo afini model Takagi-Sugeno. S tem modelom lahko s poljubno natančnostjo aproksimiramo poljubno funkcijo [8, 9, 10]. Splošnost modela lahko dokažemo s Stone-Weierstrassovim teoremom [11]. Ta pravi, da lahko vsako zvezno funkcijo aproksimiramo z razširitvijo osnovnih mehkih funkcij [12].

1.1.4 Ocena parametrov mehkega modela

Za oceno parametrov mehkega modela bo uporabljena metoda najmanjših kvadratov. Meritve zadoščajo nelinearni enačbi sistema:

$$y_i = g(x_i), \quad i = 1, \dots, n \quad (26)$$

Po Stone-Weierstrassevm teoremu obstaja za realno zvezno funkcijo g , definirano nad naborom $U^c \subset \mathbb{R}^p$, mehki sistem f tako, da je

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

kjer je $\delta > 0$ poljubno majhna konstanta. Pri aproksimaciji zveznih funkcij z mehkiimi funkcijami iz razreda \mathcal{F}^p , ki so definirane z En. (23), je treba opozoriti, da manjše vrednosti δ povečajo število rojev c , da bi zadovoljili En. (27). Pri aproksimaciji funkcije z mehkim modelom je napaka med meritvami in vrednostmi izhoda mehkega modela definirana kot:

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

kjer y_i pomeni merjene izhode sistema, \hat{y}_k pa izhode mehkega modela v trenutku k . Parametre mehke

funkcije (Θ) dobimo z minimizacijo vsote kvadratov napak:

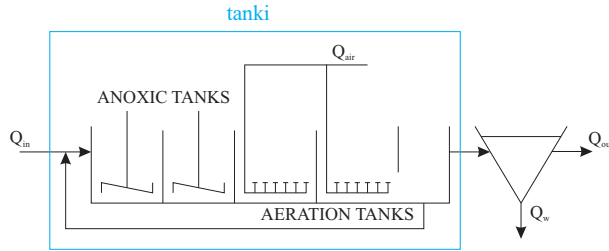
$$\begin{aligned} E &= \sum_{i=1}^n e_i^2 = \\ &= (Y - \hat{Y})^T (Y - \hat{Y}) = (Y - \Psi\Theta)^T (Y - \Psi\Theta). \end{aligned} \quad (29)$$

Parameter Θ dobimo kot parcialni odvod $\frac{\partial E}{\partial \Theta} = 0$:

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

2 Biološka čistilna naprava za čiščenje odpadnih voda

Procesi čiščenja odpadnih voda so veliki nelinearni sistemi, ki so izpostavljeni velikim motnjam predvsem v toku in obremenitvi. Hkrati pa se spreminja tudi sestava odpadne vode, ki pride v čistilno napravo. Za nepristransko primerjavo različnih sistemov oziroma strategij vodenja in sistemov spremljanja je bil razvit enoten simulacijski model. Sestoji iz petih zaporedno vezanih reaktorjev, ki se nato iztekaajo v desetplastni sekundarni tank, kjer se tvorijo usedline. Model procesa, pripadajoče enačbe in podroben opis najdemo na spletni strani <http://www.ensic.inpl-nancy.fr/COSTWWTP/>. V našem primeru smo spremljali del sistema, kjer je odpadna voda čiščena (*tanki*). Shema procesa je prikazana na sliki 1. Za izračun mehkega modela



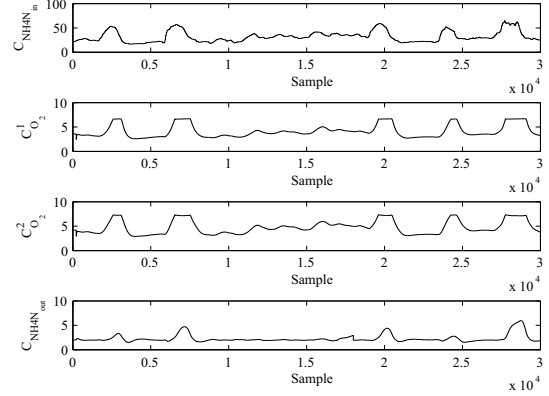
Slika 1: Shema simuliranega procesa.

obravnavanega dela procesa smo uporabili naslednje spremenljivke: koncentracija amonijaka v vhodnem toku v proces Q_{in} , ki je označena kot $C_{NH4N_{in}}$, koncentracija topljenega kisika v prvem tanku $C_{O_2}^1$, koncentracija topljenega kisika v drugem tanku $C_{O_2}^2$ in koncentracija amonijaka v drugem tanku $C_{NH4N_{out}}$. Mehki model je bil zgrajen tako, da aproksimira odvisnost koncentracije amonijaka v drugem tanku od preostalih merjenih spremenljivk:

$$C_{NH4N_{out}}(k) = \mathcal{G}(C_{NH4N_{in}}(k), C_{O_2}^1(k), C_{O_2}^2(k)), \quad (30)$$

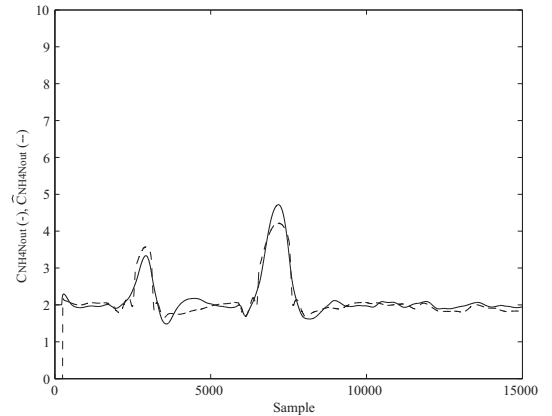
kjer \mathcal{G} označuje nelinearno povezavo med merjenimi spremenljivkami. Prvih 15000 vzorcev (s časom vzorčenja $T_s = 120s$) je bilo uporabljenih za gradnjo (identifikacijo) mehkega modela Takagi-Sugeno.

Ob 17000. vzorcu začne počasi naraščati napaka na sensorju, ki je odpravljena ob 18000. vzorcu. Sensor za merjenje koncentracije amonijaka v drugem tanku $C_{NH4N_{out}}$ se je torej pokvaril. K nominalnemu signalu sensorja smo dodali eksponentni signal za simulacijo napake. Celoten set meritev je prikazan na sliki 2.



Slika 2: Celoten nabor meritev. Koncentracija amonijaka v pritoku v sistem $C_{NH4N_{in}}$, koncentracija kisika v prvem tanku $C_{O_2}^1$ ter koncentracija kisika $C_{O_2}^2$ in amonijaka $C_{NH4N_{out}}$ v drugem tanku.

Mehki model smo zgradili na prvih 15000 vzorcih. Izhod mehkega modela $\hat{C}_{NH4N_{out}}$ in izhod procesa $C_{NH4N_{out}}$ sta prikazana na sliki 3. Indeks za detek-



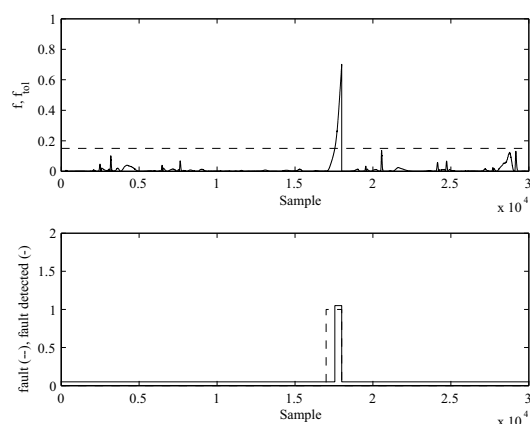
Slika 3: Verifikacija mehkega modela, kjer je $\hat{C}_{NH4N_{out}}$ izhod modela, $C_{NH4N_{out}}$ pa izhod procesa.

cijo napake je definiran kot:

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

Ko indeks detekcije preseže indeks tolerance napake, se sproži alarm. Indeks tolerance je definiran kot na-

jvečji indeks napake, pomnožen s konstanto: $f_{tol} = \gamma \max f$. Največji indeks napake smo zaznali pri učenju modela na podatkih, kjer ni bilo simulirane napake. V našem primeru smo pomnožili maksimalno vrednost indeksa napake s konstanto $\gamma = 1,5$. To pomeni, da smo za tolerančni indeks dobili vrednost $f_{tol} = 0,15$. Slika 4 prikazuje tolerančni indeks, indeks detekcije napake in signal za alarm. Napako, ki nastane pri vzorcu 17000, je predlagani sistem za spremljanje procesa detektiral pri 17556. vzorcu. Detekcija napake je zakasnjena, kar je pri napakah, ki naraščajo počasi običajno.



Slika 4: Indeks detekcije napake, tolerančni indeks f_{tol} dejanska in detektirana napaka.

3 Sklep

V članku smo obravnavali detekcijo napake na senzorju in spremljanje čiščenja odpadnih voda. Sistem za detekcijo je bil realiziran z mehkim modelom Takagi-Sugeno. Model smo identificirali s pomočjo Gustafson-Kesslovega algoritma rojenja, parametri pa so bili dobljeni s pomočjo najmanjših kvadratov. Razvit sistem je bil testiran na modelu procesa, kjer je bila simulirana napaka na enem od senzorjev. Za gradnjo mehkega modela smo uporabili naslednje meritve: vhodno koncentracijo amonijaka in koncentracijo kisika v prvem tanku ter temperaturo, koncentracijo kisika in koncentracijo amonijaka v drugem. Zaradi narave napake, simulirane na senzorju koncentracije amonijaka na drugem tanku, je bila le-ta detektirana brez lažnih alarmov in z majhno zakasnitvijo. Glede na to, da je dinamika sistema izpostavljena spreminjanju zaradi različnih obremenitev, padavin itd., bomo poskušali sistem za detekcijo napake implementirati z rekurzivno mehko identifikacijo, ki bo sposobna prilagajati dinamiko sistema tem spremembam, hkrati pa bo sistem sposoben detektirati tudi napake.

4 Literatura

- [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., Đ. 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., Włczak, 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. G.C., "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.