

Spremljanje delovanja senzorjev v čistilni napravi odpadnih voda z uporabo adaptivne metode glavnih komponent

Mina Žele, Darko Vrečko, Đani Juričić
Institut Jožef Stefan, Odsek za sisteme in vodenje
Jamova 39, 1000 Ljubljana, Slovenija
mina.zele@ijs.si

Monitoring of sensors performance in waste water treatment plant by adaptive principal component analysis

Abstract: The purpose of this work is to design and validate the procedure for sensor fault detection in pilot plant in Domžale-Kamnik Waste-Water Treatment Plant (WWTP). Waste-water treatment plants are typically subjected to daily, weekly and seasonal variations due to surrounding temperature, rain and varying process load. Therefore, fault detection procedures can result in false alarms. To avoid the problem of false alarms adaptive principal component analysis was applied.

Uvod

Pravilno delovanje senzorjev je ključnega pomena za optimalno obratovanje biološke čistilne naprave odpadnih voda. Pri obratovanju čistilnih naprav se pogosto uporabljajo senzorji kisika, v zadnjem času pa tudi amonijevega dušika. Senzorji so podvrženi različnim napakam, ki so posledica nabiranja usedlin, izpadov električnega toka in izklopov med rednim čiščenjem. Zadnji dve napaki je možno enostavno zaznati, saj se ponavadi kažeta kot padec izhoda senzorja na vrednost nič. Nepravilno delovanje senzorjev zaradi nabiranja usedlin, pa predstavlja kompleksnejši problem. V mnogih čistilnih napravah senzorje čistijo periodično (ponavadi enkrat tedensko) ne glede na dejansko stanje. Omenjeni problem je še posebej kritičen v primeru, ko izhod senzorja predstavlja regulirano veličino (senzor je uporabljen v povratni zanki), kar lahko pripelje do popolnoma neustreznega delovanja

regulacijskega algoritma. Cilj našega dela je načrtati in ovrednotiti sistem za nadzor in odkrivanje napak senzorjev v pilotni čistilni napravi, ki je postavljena v Čistilni napravi Domžale-Kamnik, saj bi s pravočasnim odkrivanjem napak lahko znatno izboljšali kvaliteto vodenja.

Statistične metode odkrivanja napak so dokaj enostavne za uporabo, zato smo v prispevku uporabili statistične modele, ki temeljijo na analizi kolinearnosti med signali čistilne naprave. V čistilni napravi je na na voljo veliko število merjenih signalov, ki so medsebojno korelirani. Z metodo glavnih komponent lahko na osnovi medsebojnih korelacij merjenih signalov sklepamo na pravilno delovanje senzorjev [3,5,6,8]. Spreminjanje signala ne opazujemo neodvisno, pač pa relativno glede na spreminjanje ostalih signalov, s katerimi je v korelaciji. Metoda analize glavnih komponent skrči število signalov na t.i. glavne komponente, ki so dejansko linearna kombinacija signalov in so medsebojno neodvisne. Glavne komponente zajamejo večino variance meritev, kar pomeni, da vsebujejo informacijo skrito v meritvah. Normalno delovanje procesa opisuje t.i. statistični model, ki ga tvorijo smeri glavnih komponent v prostoru meritev in varianca v smeri glavnih komponent. Motnje, napake na senzorjih in druge spremembe v procesu porušijo medsebojne korelacije med izmerjenimi veličinami (spremenijo se smeri glavnih komponent). Ko meritve rekonstruiramo iz glavnih komponent se bo napaka na senzorju odražala na residualih in jo bomo lahko detektirali s statističnimi testi [5]. Spremljanje delovanja procesa z analizo glavnih komponent temelji na predpostavki, da je proces

stacionaren (statistične lastnosti signalov se ne spreminjajo) [3]. Problem nestacionarnosti čistilne naprave, ki je posledica spreminjajočih vremenskih razmer, bomo rešili z uporabo t.i. adaptivne analize glavnih komponent [1, 7].

1 Analiza glavnih komponent

V procesni industriji pogosto merimo veliko število signalov, ki so korelirani. To pomeni, da imamo opravka z redundantno informacijo, ki otežuje spremljanje delovanja procesa. Z analizo glavnih komponent korelirane signale transformiramo v prostor nižje dimenzije, ki vsebuje večino variance meritev, nove koordinate pa niso več korelirane [3, 5, 8].

Vzemimo matriko meritev \mathbf{X} dimenzije $(n \times m)$, kjer m pomeni število merjenih signalov, n pa število vzorcev. Signali so normirani, kar pomeni, da imajo srednjo vrednost nič in varianco enako 1. Koordinate v novem prostoru, oziroma t.i. matriko zadetkov \mathbf{T} dobimo kot linearno kombinacijo signalov

$$\mathbf{T} = \mathbf{X}\mathbf{P} \quad (1)$$

kjer je $\mathbf{P} \in \mathfrak{R}^{m \times m}$ transformacijska matrika. Vektor zadetkov v trenutku k $\mathbf{t}(k)=[t_1(k), t_2(k), \dots, t_m(k)]^T$ dobimo s sledečo transformacijo

$$\mathbf{t}(k) = \mathbf{P}^T \mathbf{x}(k), \quad (2)$$

pri čemer vektor meritev v trenutku k $\mathbf{x}(k)=[x_1(k), x_2(k), \dots, x_m(k)]^T$ tvorijo elementi k vrstice matrike meritev \mathbf{X} .

Transformacijska matrika \mathbf{P} vsebuje lastne vektorje \mathbf{p}_i ($i=1,2,\dots,m$) kovariančne matrike meritev in jo dobimo z dekompozicijo

$$\mathbf{R} = \text{cov}(\mathbf{X}) = \frac{\mathbf{X}^T \mathbf{X}}{n-1} = \mathbf{P}\mathbf{\Sigma}\mathbf{P}^T \quad (3)$$

$\mathbf{P} \in \mathfrak{R}^{m \times m}$ je unitarna matrika $\mathbf{P}\mathbf{P}^T = \mathbf{I}$ in $\mathbf{\Sigma} \in \mathfrak{R}^{m \times m}$ je diagonalna matrika lastnih vrednosti kovariančne matrike v padajočem vrstnem redu ($\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_m$). Lastni vektorji so baza novega koordinatnega sistema in lastne vrednosti predstavljajo varianco meritev v teh smereh.

Prvi lastni vektor predstavlja smer, v katerih je varianca signalov največja, drugi lastni vektor smer, v kateri je največja varianca preostalega podprostora meritev i.t.d. Pri transformaciji v nov koordinatni sistem obdržimo samo prvih p lastnih vektorjev $\mathbf{P}_p \in \mathfrak{R}^{m \times p}$, ki ustrezajo največjim lastnim vrednostim

$$\mathbf{T}_p = \mathbf{X}\mathbf{P}_p \quad (4)$$

Optimalna izbira števila p je taka, da s p lastnimi vektorji zajamemo večino variance meritev, preostali $m-p$ lastni vektorji pa tvorijo podprostor, ki vsebuje šum meritev. V literaturi zasledimo več načinov določevanja števila lastnih vektorjev [7]. V praksi se pogosto uporablja mera CPV (cumulative percent variance), ki nam pove kolikšen procent celotne variance meritev vsebuje prvih a lastnih vektorjev

$$\text{CVP}(p) = \frac{\sum_{j=1}^p \lambda_j}{\sum_{j=1}^m \lambda_j} 100\% \quad (5)$$

Število glavnih komponent je izbrano tako, da mera CVP preseže neko vnaprej določeno vrednost, ki je ponavadi 95%.

S p glavnimi komponentami lahko meritve rekonstruiramo kot vsoto

$$\mathbf{X} = \tilde{\mathbf{X}} + \mathbf{E} = \mathbf{T}_p \mathbf{P}_p^T + \mathbf{E} = \sum_{i=1}^p \mathbf{t}_i \mathbf{p}_i^T + \mathbf{E}, \quad (6)$$

kjer je \mathbf{t}_i stolpec matrike \mathbf{T} . Prvi člen v vsoti predstavlja PCA model drugi del pa šum. Na ta način je prostor meritev razstavljen na dva ortogonalna prostora. Prvi ustreza sistematični varianci procesa, drugi pa šumu.

2 Statistično spremljanje delovanja procesa z analizo glavnih komponent

PCA model tvorijo smeri lastnih vektorjev in variance meritev v smereh lastnih vektorjev. Določimo ga v t.i. fazi učenja iz podatkov normalno delujočega sistema. Pri spremljanju delovanja procesa, oziroma v t.i. fazi odločanje pa se meritve v vsakem trenutku vzorčenja

projecirajo na lastne vektorje. Statistično spremljanje delovanje procesa temelji na predpostavki, da se med normalnim delovanjem procesa meritve preslikajo v določeno področje v prostoru zadetkov, ki ga določa PCA model. Ujemanje PCA modela in trenutne meritve testiramo z dvema statistikama:

Hotellingova T^2 statistika (ugotavljamo, ali je meritev znotraj predpisane envelope)

$$T^2(k) = \mathbf{t}(k)^T \boldsymbol{\Sigma}^{-1} \mathbf{t}(k) \quad (7)$$

Zgornja statistika ima F porazdelitev, ker je kovariančna matrika ocenjena iz končnega števila podatkov. Mejo zaupanja za α % verjetnost dobimo kot

$$T_{p,N,\alpha}^2 = \frac{p(N-1)}{N-p} F_{p,n-p,\alpha} \quad (8)$$

V zgornji enačbi je N število podatkov, ki smo jih uporabili za določitev PCA modela (kovariančne matrike).

SPE statistika (Q statistika) je vsota kvadratov elementov v posamezni vrstici matrike residualov. Testiramo, če varianca meritev, ki ni zajeta z glavnimi komponentami, presega predpisano vrednost

$$SPE(k) = \mathbf{e}(k)^T \mathbf{e}(k) = \mathbf{x}(k)^T (\mathbf{I} - \mathbf{P}_p \mathbf{P}_p^T) \mathbf{x}(k), \quad (9)$$

V enačbi (9) smo upoštevali, da je vektor residualov enak

$$\mathbf{e}(k) = \mathbf{x}(k) - \tilde{\mathbf{x}}(k) = \mathbf{x}(k) - \mathbf{P}_p \mathbf{t}(k).$$

Zgornja meja zaupanja SPE_α je za α % verjetnost podana v [2].

3 Adaptivna analiza glavnih komponent

Metodo analize glavnih komponent za spremljanje delovanja procesa lahko uporabljamo samo za stacionarne procese. V praksi pa se proces čiščenja odplak ponavadi počasi, a neprestano spreminjajo predvsem zaradi zunanjih vremenskih vplivov. Spreminjanje procesa se kaže kot sprememba srednjih vrednosti meritev, sprememba variance

meritev ter sprememba korelacij med spremenljivkami. Problem nestacionarnosti rešujemo z adaptivnim postopkom, ki omogoča prilagajanje srednjih vrednosti, varianc in kovariančne matrike meritev počasnim spremembam procesa. Tehnično to pomeni, da se pri odločanju omejimo na končni interval $[t-T, t]$, namesto da bi upoštevali vse pretekle podatke iz zgodovine. Rekurzivni postopek adaptivne analize glavnih komponent je opisan v [7].

S faktorjem pozabljanja a uravnavamo dinamiko prilagajanja parametrov (srednjih vrednosti in varianc meritev ter kovariančna matrike meritev). Izberemo ga tako, da se PCA model prilagaja počasnim spremembam procesa. V vsakem koraku izračunamo tudi število glavnih komponent in mejo zaupanja za T^2 in SPE statistiko. Prekoračitev meje zaupanja ene izmed statistik je indikator za spremembo opazovanega procesa, zato se sproži alarm. Ko pa se model sčasoma prilagodi spremenjeni situaciji, alarm izveni, čeprav se proces ni vrnil v normalno delujoče stanje. Opisanemu problemu se izognemo tako, da prilagajanje parametrov ustavimo (a postavimo na 1) takoj, ko SPE ali T^2 statistika prekoračita določeno mejo. S prilagajanjem PCA modela nadaljujemo, ko sta obe statistiki ponovno pod določeno mejo. Postopek je bolj natančno opisan s sledečim algoritmom:

1. korak : Izberemo ustrezen faktor pozabljanja $a=a_0$.

2. korak: V vsakem trenutku vzorčenja z adaptivnim postopkom izračunamo SPE in T^2 statistiko ter ustrezne meje.

3. korak: Če $SPE > SPE_\alpha$ ali $T^2 > T_{p,N,\alpha}^2$ potem $a=1$, sicer $a=a_0$.

4. korak: Nadaljujemo s korakom 2.

4 Odkrivanje napak senzorjev v čistilni napravi

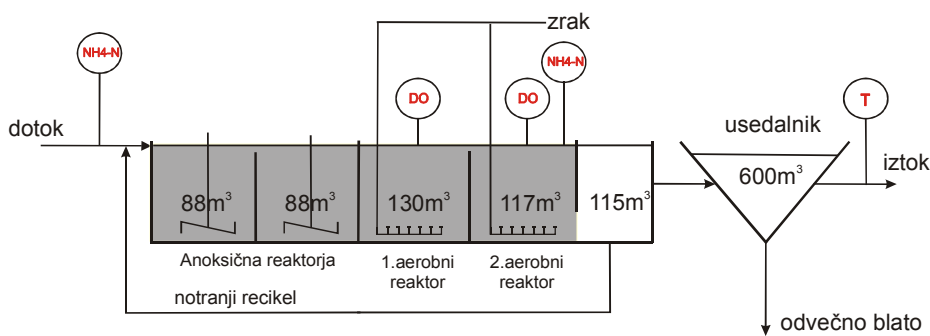
Rezultate odkrivanja napak z analizo glavnih komponent smo dobili s simulacijskimi poskusi z nelinearnim modelom pilotne čistilne naprave,

ki je realiziran v programskem paketu GPS-X [2]. Shema pilotne naprave, ki je postavljena v Čistilni napravi Domžale-Kamnik, je prikazana na sliki 1. Na čistilni napravi se odpadna voda po mehanski stopnji čisti z aktivnim blatom (mikroorganizmi) po MBBR (moving bed biofilm reactor) postopku. Pri analizi glavnih komponent smo uporabili sledeče signale: koncentracijo amonijevega dušika na dotoku, koncentracijo raztopljenega kisika v prvem aerobnem reaktorju, koncentracijo raztopljenega kisika v drugem aerobnem reaktorju, temperaturo in koncentracijo amonijevega dušika v drugem aerobnem reaktorju (simboli na sliki).

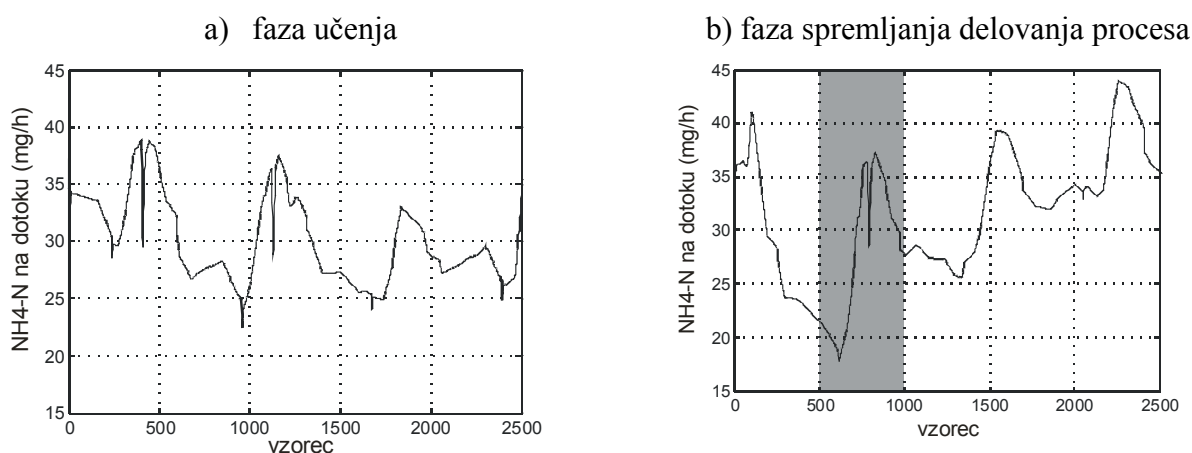
Amonijev dušik na dotoku je dejanski izmerjeni signal in ima značilen vzorec

periodičnih dnevnih sprememb (slika 2). Čas vzorčenja je 2 minuti. Število vseh vzorcev pa je 5040 (v enem tednu). Signal je nestacionaren, saj se njegove statistične lastnosti spreminjajo zaradi vpliva vremena. Napako na senzorju amonijevega dušika v drugem aerobnem reaktorju smo simulirali tako, da smo koncentraciji amonijevega dušika v reaktorju umetno dodali vrednosti prikazane na sliki 3.

Ker avtokorelacije signalov pri zakasnitvi T_s niso zanemarljive, moramo v matriko podatkov X dodati še zakasnjene signale z zakasnitvijo T_s [6] (matrika podatkov je vsebovala vsega skupaj 10 signalov).



Slika 1. Shema pilotne čistilne naprave odpadnih voda



Slika 2: Koncentracija amonijevega dušika na dotoku

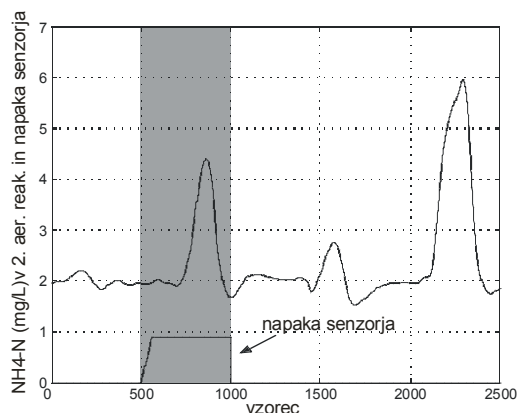
Približno polovico vzorcev (koncentracija amonijevega dušika na dotoku je na sliki 2 a)) smo uporabili za izračun začetnega PCA modela (enačbe (11), (12), (13)). Izkazalo se je, da s 3 glavnimi komponentami zajamemo več kot 95% variance meritev. Za preostale podatke (koncentracija amonijevega dušika na dotoku je prikazana na sliki 2 b)) smo sproti v vsakem trenutku vzorčenja izračunali T^2 in SPE statistiko.

Na sliki 4 sta prikazani obe statistiki skupaj z mejami in ustrezni alarmi v primeru, ko smo ves čas (naslednjih 2540 vzorcev) uporabljali začetni PCA model. Jasno je razvidno, da nastopijo napačni alarmi, ki so posledica spremembe delovne točke procesa. Zaradi nelinearnosti procesa sprememba delovne točke spremeni medsebojne korelacije signalov in začetni PCA model ni več verodostojna slika dejanskih razmer.

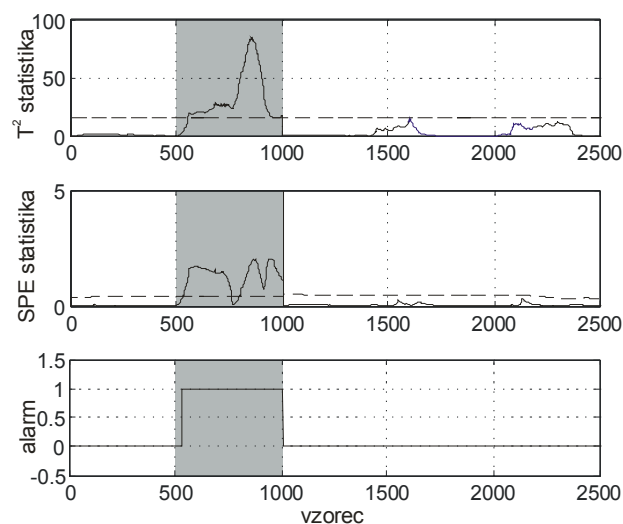
Slika 5 prikazuje rezultate pri uporabi adaptivnega PCA modela, ki smo ga rekurzivno računali s faktorjem pozabljanja 0.9995 (časovna konstanta prilagajanja je približno 3 dni). Z adaptivnim PCA modelom smo na osnovi obeh statistik detektirali napako senzorja. SPE statistika v 535. vzorcu prekorači mejno vrednost, kar sproži alarm. Kot smo omenili že zgoraj, se alarm sproži, ko ena izmed statistik prekorači mejno vrednost. Hkrati se faktor pozabljanja postavi na 1, dokler se proces ponovno ne vrne v normalno delujoče stanje.

Pri uporabi adaptacije nimamo napačnih alarmov, saj se PCA model prilagaja spreminjanju procesa. Vendar pa je z opisanim rekurzivnim postopkom možno odkriti samo napake, ki se pojavijo mnogo hitreje kot pa je dinamika prilagajanja PCA modela. Omenjeni

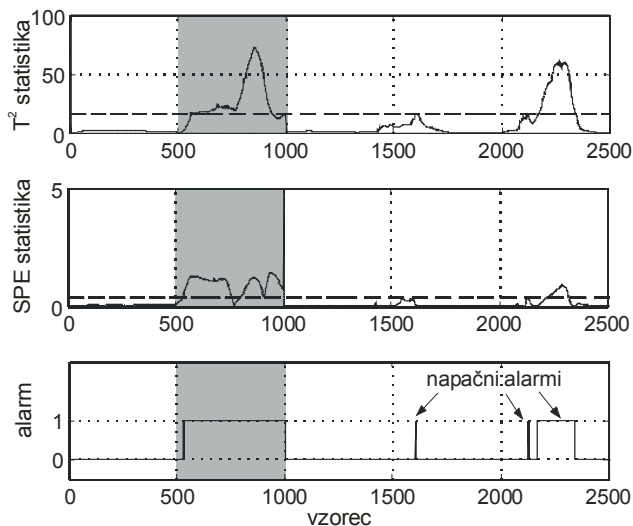
problem bomo poskušali rešiti tekom prihodnjih raziskav.



Slika 3: Koncentracija amonija v drugem aerobnem reaktorju



Slika 4: Rezultati statističnega spremljanja procesa z analizo glavnih komponent (T^2 statistika in ustrezna meja označena s črtkano črto; SPE statistika in ustrezna meja označena s črtkano črto; alarm)



Slika 5: Rezultati statističnega spremljanja procesa z adaptivno analizo glavnih komponent (T^2 statistika in ustrezna meja označena s črtkano črto; SPE statistika in ustrezna meja označena s črtkano črto; alarm)

5 Sklep

V prispevku smo obravnavali odkrivanje napak senzorjev v biološki čistilni napravi z metodo analize glavnih komponent. Pri tem smo se omejili zgolj na detekcijo napak, kar pomeni ugotavljanje prisotnosti napake ne pa tudi njeno lokalizacijo. Proces čiščenja odpadnih voda je zaradi vremenskih motenj časovno spremenljiv. Poleg tega se pri odkrivanju napak pojavi problem ločevanja vpliva motnje od vpliva napak senzorjev na SPE in T^2 statistiki. Omenjeni problem smo rešili z adaptivnim postopkom analize glavnih komponent. Le-ta omogoča sprotno prilagajanje statističnega modela počasnim spremembam procesa, s čimer se zmanjša število napačnih alarmov.

Pri simulacijski poskusih z nelinearnim modelom čistilne naprave smo senzorju koncentracije amonijevega dušika v drugem aerobnem reaktorju umetno dodali napako.

Izkazalo se je, da se v primeru izbrane napake s primerno izbiro faktorja pozabljanja pri rekurzivni analizi glavnih komponent lahko znebimo napačnih alarmov.

V prispevku je prikazan le prvi delni rezultat širše naloge, katere glavni cilj je sinteza in uporaba algoritma za nadzor delovanja senzorjev v Čistilni napravi Domžale- Kamnik.

6 Literatura

- [1] N.B. Gallagher, B.M. Wise, S.W. Butler, D. White, G.G. Barna, Development and benchmarking of multivariate statistical process control tools for a semiconductor etch process: Improving robustness through model updating. IFAC ADCHEM'97, str. 7883, Banff, Canada, June 1997.
- [2] Hydromantis (2001). GPS-X–Technical Reference, GPS-X Version 4. Ontario, Canada.
- [3] J. E. Jackson, G. S. Mudholkar, Control procedures for residuals associated with principal component analysis. *Technometrics*, Vol. 21, No. 3, str. 341-349, 1979.
- [4] D. R. Jensen, H. Solomon, A gaussian approximation to the distribution of a definite quadratic form, *Journal of the American Statistical Association*, Vol 67, No.340, str. 898-902, 1972.
- [5] J. V. Kresta, J. F. MacGregor, T. E. Marlin, Multivariate statistical monitoring of process operating performance. *The Canadian Journal of Chemical Engineering*, Vol. 69, str. 35-47, 1991.
- [6] W. Ku, R. H. Storer, C. Georgakis, Disturbance detection and isolation by dynamic principal component analysis, *Chemometrics and Intelligent Laboratory Systems*, Vol. 30, str. 179-196, 1995.
- [7] W. Li, H. H. Yue, S. Valle-Cervantes, S. J. Quin, Recursive PCA for adaptive process monitoring. *Journal of Process Control*, Vol. 10, str. 471-486, 2000.
- [8] A. Negiz and A. Cinar, Statistical monitoring of multivariable dynamic processes with state-space models. *AIChE Journal*, Vol. 43, No. 8, str. 2002-2020, 1997.