

Primjena kontinuum regresije za modeliranje procesa na temelju pogonskih podataka

UDK 658.5:519.246
665.63:658.5
IFAC 5.6.3; 2.8.1

Izvorni znanstveni članak

Važne procesne veličine koje daju informaciju o kakvoći izlaznog proizvoda često nije moguće mjeriti senzorom nego se njihov iznos utvrđuje laboratorijskom analizom. Kako bi se omogućilo kontinuirano praćenje tijekom procesa te efikasnije upravljanje proizvodnim procesom, ovu teško mjerljivu procesnu veličinu je potrebno estimirati, tj. odrediti na temelju matematičkog modela. Za izgradnju odgovarajućeg modela procesa vrlo često su na raspolaganju samo procesni mjerni podaci pohranjeni u procesnu bazu podataka.

U ovom se radu prikazuje prikladna metodologija za modeliranje procesa na temelju pogonskih podataka. Za izgradnju modela pri tome se predlažu regresijske metode zasnovane na preslikavanju ulaznog prostora u latentni potprostor. U radu se posebno istražuju svojstva kontinuum regresije (CR). Budući da neuronske mreže predstavljaju dobru osnovu za izgradnju modela na podacima, dopunski se istražuje mogućnost hibridizacije višeslojne perceptronske (MLP) neuronske mreže i CR metode, s ciljem iskorištavanja dobrih svojstava obju metoda te izbjegavanja njihovih nedostataka u izgradnji modela procesa na pogonskim podacima. Prednosti predloženih metoda izgradnje modela procesa nad uobičajeno korištenim regresijskim metodama prikazane su na primjeru modeliranja procesa destilacije nafte na raspoloživim mjernim podacima.

Ključne riječi: modeliranje procesa, pogonski podaci, estimacija teško mjerljive procesne veličine, preslikavanje u latentni prostor, kontinuum regresija, neuronske mreže

1. UVOD

Stalni rast zahtjeva u industrijskoj proizvodnji u pogledu kvalitete proizvoda i ekonomičnosti proizvodnje nameće visoke zahtjeve na mjerenje procesnih veličina. Nerijetko važne procesne veličine, koje daju informaciju o kvaliteti izlaznog proizvoda, senzorima nisu mjerljive ili su ovakva mjerenja suviše skupa i/ili nedovoljno pouzdana pa se ne koriste. Informacija o iznosu ovih teško mjerljivih veličina utvrđuje se laboratorijskom analizom uzoraka uzimanih iz procesa. Ovaj način mjerenja provodi se povremeno, uz veliko kašnjenje u dobivanju informacije, pa ne omogućava kontinuirani nadzor nad kvalitetom izlaznog proizvoda i primjenu automatskog upravljanja. Stoga je potrebno pronaći način kako i u ovakvim slučajevima pravovremeno raspolagati informacijom o iznosu teško mjerljive veličine.

Ako neku veličinu nije moguće mjeriti, alternativni postupak je njena estimacija. Kada se u postrojenju senzorima mjeri dovoljan broj procesnih veličina koje su u korelaciji s teško mjerljivom veličinom, estimaciju teško mjerljive procesne veličine je moguće provesti na temelju informacija o ovim lako mjerljivim procesnim veličinama [1, 2]. Pri tome je potrebno raspolagati matematičkim modelom

procesu koji povezuje ulazne veličine modela s izlaznom veličinom, na isti način kako su u procesu povezane odabrane lako mjerljive procesne veličine s teško mjerljivom veličinom koja se estimira. U praksi je uobičajena situacija da model nije na raspolaganju, niti je do modela procesa moguće doći kroz teorijsku analizu. Stoga se iznalaženje modela procesa temelji na raspoloživim mjernim podacima.

U suvremenim industrijskim postrojenjima mnoge procesne veličine se mjere kroz duže vremensko razdoblje, a podaci se pohranjuju u procesne baze podataka. Za pretpostaviti je da je u ovoj gomili podataka sadržano veliko znanje o procesu [3]. Iznalaženje metodologije koja bi omogućila da se na temelju podataka sadržanih u procesnim bazama dobije model procesa željenih svojstava je veliki izazov jer je ovakvo modeliranje u načelu ekonomski vrlo povoljno i primjenjivo na sve tipove procesa.

Mjerni podaci sadržani u procesnim bazama su pogonski podaci, koji u pravilu sadrže mnoštvo različitih smetnji i mjernih pogrešaka. Budući da kakvoća modela izgrađenog na podacima jako ovisi o informativnosti podataka za modeliranje, pri modeliranju procesa na pogonskim podacima od velike je važnosti pripremi dio modeliranja u kojem se provodi analiza, odabir i predobradba raspoloživih

mjernih podataka [4]. Osim ovog odabira reprezentativnih mjernih uzoraka, pri formiranju skupa podataka za izgradnju modela procesa potrebno je odabrati i odgovarajući skup lako mjerljivih procesnih veličina na temelju kojih će se estimirati teško mjerljiva procesna veličina. Pripremni dio modeliranja procesa često zahtijeva više vremena i napora nego neposredna izgradnja modela na odabranom skupu podataka [4], a uslijed njegove složenosti, često je u rad potrebno uključiti i eksperte s područja procesa koji se modelira.

S ciljem postizanja zadovoljavajuće točnosti i robusnosti estimatora, estimacija teško mjerljive veličine u pravilu se provodi na temelju većeg broja lako mjerljivih procesnih veličina. Stoga se u ovom slučaju izgradnja modela procesa provodi na temelju skupa visokodimenzionalnih, koreliranih i onečišćenih podataka. Uobičajeno korištene regresijske metode modeliranja na podacima, kao što su višestruka linearna regresija (MLR) [5] i njeno poopćenje na nelinearno modeliranje, višeslojna perceptronska (MLP) neuronska mreža [6], pod ovim uvjetima često zakazuju i rezultiraju modelima s lošim predikcijskim svojstvima.

Modeliranje procesa na temelju pogonskih podataka odavno je prisutno u kemometriji. Na ovom području su do kraja 80-ih razvijena različita poboljšanja MLR metode, uglavnom kroz dogradnju metodama statističke multivarijantne analize [7], kao što su PCA (*Principal Component Analysis*) i PLS (*Partial Least Squares*). Kao rezultat dobivene su PCR (*Principal Component Regression*) i PLSR (*Partial Least Squares Regression*) linearne metode izgradnje modela na podacima. Budući da su procesi u pravilu nelinearni, ove linearne metode se nastoji poopćiti na nelinearno modeliranje kroz primjenu neuronskih mreža [8], što rezultira NNPCA i NNPLS metodama [9]. Početkom 90-ih razvijena je CR (*Continuum Regression*) metoda [10, 11] koja predstavlja određeno poopćenje postojećih metoda multivarijantne analize, pri čemu PCR, PLSR i MLR predstavljaju samo specijalne slučajeve CR metode. S obzirom da u odnosu na ostale metode posjeduje dopunski mehanizam za utjecaj na preraspodjelu pomaka i varijance u predikciji, CR metoda je vrlo obećavajuća u pogledu izgradnje modela s dobrim predikcijskim svojstvima.

S obzirom da se u radu istražuju metode izgradnje modela procesa na temelju mjernih podataka preuzetih iz procesne baze, svi podaci za modeliranje su unaprijed dostupni, pa se primjenjuje off-line postupak izgradnje modela. Nadalje, budući da su podaci o teško mjerljivoj veličini dobiveni vrlo niskom i promjenjivom frekvencijom uzorkovanja, za izgradnju modela procesa u načelu nije moguće koristiti metode izgradnje dinamičkog modela.

Ovaj rad je organiziran na slijedeći način. U odjeljku 2 je naznačena struktura modela i naznačen smisao modeliranja zasnovanog na preslikavanju ulaza u latentno područje. Kriterij za procjenu parametara prema kontinuumu regresiji dan je u 3. odjeljku, dok se u 4. odjeljku analizira mogućnost udruživanja CR metode s MLP neuronskom mrežom. U 5. odjeljku je opisan proces koji se modelira, način formiranja skupova za izgradnju modela, te su prikazani rezultati ispitivanja izgrađenih modela. U posljednjem odjeljku je dana kratka rasprava te su rezimirani postignuti rezultati.

2. METODE ZASNOVANE NA PRESLIKAVANJU U LATENTNI PROSTOR

Problem izgradnje modela procesa na temelju podataka svodi se na pronalaženje aproksimacijske funkcije $f_m(\bullet)$ koja aproksimira nepoznatu prirodnu funkcionalnu ovisnost izlazne veličine procesa o odabranim ulaznim veličinama. Za aproksimacijsku funkciju obično se odabire neka funkcija parametrirana konačno-dimenzionalnim vektorom parametara, te se model može predstaviti sasvim općenito kao:

$$\hat{y} = f_m(\mathbf{x}, \Theta), \quad (1)$$

gdje je:

- \mathbf{x} – vektor ulaznih veličina (N odabranih lako mjerljivih procesnih veličina),
- Θ – vektor parametara aproksimacijske funkcije $f_m(\bullet)$, odnosno modela procesa,
- \hat{y} – izlaz modela (estimirana vrijednost teško mjerljive procesne veličine).

Budući da nema a priori znanja o procesu koje bi pomoglo u strukturiranju modela, u izgradnji modela procesa polazi se od neke opće strukture modela. Pri tome se za aproksimacijsku funkciju uobičajeno primjenjuje funkcija oblika zbroja više funkcija:

$$f_m(\mathbf{x}, \Theta) = \sum_{k=1}^K \nu_k(\mathbf{x}, \Theta_k), \quad (2)$$

gdje je:

- ν – temeljna funkcija aproksimacijske funkcije $f_m(\bullet)$,
- K – ukupni broj temeljnih funkcija $\nu(\bullet)$ u aproksimacijskoj funkciji $f_m(\bullet)$.

Ova opća struktura, predstavljena sustavom parametriranih funkcija $\nu(\bullet)$, zasnovana je na tzv. kompozitnoj funkciji koja prema teoremu Kolmogorova [12] može aproksimirati bilo koju kontinuiranu funkciju superpozicijom konačnog broja osnovnih nelinearnih kontinuiranih funkcija.

Za odabranu strukturu modela potrebno je odrediti njegove parametre Θ . Kada se projektira

estimator teško mjerljive veličine, model procesa treba imati što bolja predikcijska svojstva. Pri izgradnji predikcijskog modela uobičajeno je parametre određivati kroz regresiju, pri čemu se svi parametri modela procjenjuju po kriteriju minimizacije izlazne pogreške aproksimacije. Kako se kao mjera kakvoće aproksimacije najčešće koristi L_2 -norma, kriterij kakvoće regresijskog modela ima oblik:

$$\mathfrak{J}(\Theta) = \frac{1}{2} \sum_{i=1}^I [y_i - f_m(\mathbf{x}_i, \Theta)]^2 = \frac{1}{2} \sum_{i=1}^I e_i^2(\Theta), \quad (3)$$

gdje je:

- I – ukupni broj mjerenja u skupu podataka za modeliranje procesa,
- y – izlazna veličina procesa (teško mjerljiva procesna veličina),
- e – izlazna pogreška modela procesa (pogreška estimacije).

Kada se parametri modela određuju kroz regresiju na temelju slabo informativnih podataka, kao što su to pogonski podaci, pouzdanost procjene parametara u pravilu je mala. Ovo rezultira velikom varijancom predikcijske pogreške i time lošim svojstvima modela. Smanjivanjem broja parametara, kroz smanjivanje dimenzije modela (K), smanjuje se i varijanca. Međutim, pretjeranim smanjivanjem dimenzije, model postaje nedovoljno fleksibilan u aproksimaciji ulazno-izlaznih odnosa u procesu koji opisuje, pa raste pomak u predikciji. Stoga je za postizanje minimalne predikcijske pogreške potrebno postići optimalnu preraspodjelu pomaka i varijance u predikciji njegovog izlaza [6]. Pri izgradnji regresijskog modela zasnovanog na općoj strukturi (2) i temeljnim funkcijama nepromjenjivog oblika, optimalna preraspodjela pomaka i varijance traži se kroz optimalnu dimenziju modela.

Preporuka da se na ulaz estimatora dovedu sve lako mjerljive procesne veličine koje su u bilo kakvoj korelaciji s teško mjerljivom veličinom koja se estimira, rezultira izgradnjom modela procesa s velikim brojem ulaznih veličina, a time i velikog broja parametara. Usljed toga model (2) već i pri malom broju temeljnih funkcija može postići suviše veliku varijancu u predikciji. Ovaj problem je dopunski izražen kada se modelira izrazito nelinearan proces jer je za dobar opis procesa potrebno koristiti veći broj temeljnih funkcija. Stoga pri izgradnji modela procesa na pogonskim podacima i velikom broju međusobno koreliranih ulaznih veličina nije prihvatljivo parametre modela procjenjivati isključivo kroz minimizaciju izlazne pogreške.

Jedan od načina unaprjeđenja regresijskog modeliranja za ovu namjenu je primjena metoda zasnovanih na preslikavanju ulaznog prostora u latentni prostor. Na ovaj se način visokodimenzionalni, ko-

relirani ulazni prostor najprije preslikava u odgovarajući nižedimenzionalni potprostor, a regresija se provodi na novim (latentnim) varijablama dobivenim ovim preslikavanjem [7]. Pripadni model sada ima oblik kompozicije dviju funkcija i može se predstaviti kao:

$$\hat{y} = f_m(\mathbf{x}, \Theta) = f_r(\boldsymbol{\varphi}_j(\mathbf{x}; \boldsymbol{\alpha}_j); \boldsymbol{\beta}), \quad j = 1, 2, \dots, J, \quad (4)$$

gdje je:

- $\boldsymbol{\varphi}(\bullet)$ – funkcija preslikavanja ulaznog prostora u latentni prostor, $\boldsymbol{\varphi}(\bullet): \mathfrak{R}^{N+n(\alpha)} \rightarrow \mathfrak{R}^J$,
- $f_r(\bullet)$ – funkcija preslikavanja latentnog prostora na izlaz, $f_r(\bullet): \mathfrak{R}^{J+n(\beta)} \rightarrow \mathfrak{R}^J$,
- J – dimenzija preslikavanja ulaznog prostora (broj latentnih varijabli), $J < N$,
- $\boldsymbol{\alpha}_j$ – vektor parametara funkcije preslikavanja ulaza (u j -tom smjeru),
- $\boldsymbol{\beta}$ – vektor parametara funkcije preslikavanja latentnog prostora na izlaz.

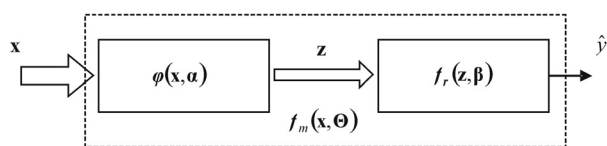
Pri odabiru tipa preslikavanja ulaza važno je odabrati preslikavanje koje rezultira dovoljno informativnim latentnim varijablama, na temelju kojih se izlazna veličina može estimirati uz zadovoljavajuću točnost. Latentne varijable predstavljaju transformirane ulazne veličine i izlazi su funkcije preslikavanja ulaza, te se mogu predstaviti kao:

$$z_j = \boldsymbol{\varphi}_j(\mathbf{x}; \boldsymbol{\alpha}_j). \quad (5)$$

Vektor parametara modela Θ u ovom slučaju je podijeljen na parametre ulazne transformacije α i regresijske parametre β , pa za j -ti smjer projekcije ulaza vrijedi:

$$\Theta_j = (\boldsymbol{\alpha}_j, \boldsymbol{\beta}_j). \quad (6)$$

Struktura modela zasnovanog na preslikavanju ulaznog prostora shematski je prikazana na slici 1.



Sl. 1. Načelna shema modela procesa podijeljenog na dvije razine preslikavanja

Za praktičnu primjenu model (4) se najčešće realizira u obliku težinskog zbroja funkcija:

$$\hat{y} = f_m(\mathbf{x}, \Theta) = \sum_{j=1}^J \beta_j \boldsymbol{\psi}_j \left[\boldsymbol{\varphi}_j(\mathbf{x}; \boldsymbol{\alpha}_j) \right], \quad (7)$$

i predstavlja specijalni slučaj modela (2), pri čemu je (uz $K=J$) temeljna funkcija $\boldsymbol{\nu}(\bullet)$ rastavljena na dvije funkcije: funkciju preslikavanja ulaznog prostora $\boldsymbol{\varphi}(\bullet)$ i aktivacijsku funkciju $\boldsymbol{\psi}(\bullet)$. Osim odabira

funkcija odgovarajućeg tipa, ovakva struktura modela zahtijeva i zasebnu definiciju kriterija za procjenu parametara, tj. funkcije $\mathfrak{Z}(\alpha)$ i $\mathfrak{Z}(\beta)$. Pri tome se kriterij $\mathfrak{Z}(\beta)$ fokusira isključivo na minimizaciju izlazne pogreške modela, tj. isključivo na izlazni prostor, dok se za $\mathfrak{Z}(\alpha)$ mogu koristiti različiti kriteriji. Sa stajališta izgradnje predikcijskog modela, pri definiciji kriterija $\mathfrak{Z}(\alpha)$ važan aspekt je kako eksplicitno definirati preraspodjelu naglaska na obuhvaćanje odnosa među ulaznim veličinama i minimizaciju izlazne pogreške modela. Budući da povećanje naglaska na ulazni prostor povećava pomak u predikciji, a povećanje naglaska na minimizaciju izlazne pogreške modela povećava varijancu u predikciji modela [13], ova preraspodjela izravno utječe na predikcijska svojstva modela, slično promjeni dimenzije modela.

Budući da kakvoća ulaznog preslikavanja u najvećoj mjeri određuje ukupna svojstva modela, u razvoju metoda modeliranja temeljenog na podacima težište istraživanja se stavlja na određivanje optimalnog tipa ulaznog preslikavanja, odnosno tipa funkcije $\varphi(\bullet)$, te načina procjene parametara α .

3. KONTINUUM REGRESIJA

Linearno preslikavanje ulaza u velikom broju slučajeva daje dobre rezultate [7]. Linearno preslikavanje ulaznog prostora su projekcije ulaza u smjerovima vektora α_j , pa je funkcija preslikavanja ulaza u j -tom smjeru oblika:

$$\varphi_j(\mathbf{x}; \alpha_j) = \sum_{n=1}^N \alpha_{nj} x_n, \quad (8)$$

gdje je α_{nj} ulazni parametar modela, s obzirom na n -ti ulaz i j -ti izlaz iz funkcije preslikavanja ulaza. Latentne varijable (z) su prema tome dobivene kao linearni težinski zbroj ulaznih veličina. Ukupno J vektora α formiraju matricu \mathbf{A} (dim. $N \times J$) koja čini vektorsku bazu za preslikavanje ulaznog prostora u latentni potprostor. Smjerovi projekcije α_j mogu se odrediti na mnoštvo različitih načina. Ako se odabere ortonormalna vektorska baza, kao u slučaju PCA i PLS metoda, gdje su bazni vektori međusobno ortogonalni i dopunski normirani, latentni prostor formira J također međusobno ortogonalnih vektora $\mathbf{z}_j = \mathbf{X} \alpha_j$, a postupak procjene parametara α bitno se pojednostavljuje. Ovo je ujedno i razlog zašto su ove metode preslikavanja ulaznog prostora općenito najšire korištene tehnike u modeliranju temeljenom na višedimenzionalnim podacima [7]. Uzimanjem u obzir svih J ortogonalnih smjerova projekcije α_j , mjerni vektor \mathbf{x} projicira se u koordinatu \mathbf{z} u latentnom prostoru kao:

$$\mathbf{z} = \mathbf{A}^T \mathbf{x}, \quad (9)$$

iz čega slijedi matrica latentnih varijabli $\mathbf{Z} = \mathbf{X} \cdot \mathbf{A}$ (dim. $I \times J$). Nakon odabira funkcije preslikavanja (8) potrebno je definirati kriterij za određivanje smjerova projekcije ulaza, tj. parametara α .

Prema PCA kriteriju u latentnom prostoru se nastoji očuvati što više varijacija ulaznog prostora. Stoga se PCA kriterij može definirati kao traženje smjerova projekcije α_j koji maksimiziraju varijacije u latentnom prostoru:

$$\max_{\alpha_j} [\text{var}(\mathbf{X} \alpha_j)]. \quad (10)$$

S obzirom na uvjet ortonormalnosti, može se pokazati da su ovako optimirani smjerovi projekcije u stvari svojstveni vektori ulazne kovarijancne matrice $\mathbf{X}^T \mathbf{X}$ [7], koji se još nazivaju glavnim komponentama. Pri tome su glavne komponente poredane po veličini, tj. na način da prvu formira svojstveni vektor što pripada najvećoj svojstvenoj vrijednosti, a posljednju svojstveni vektor što pripada najmanjoj svojstvenoj vrijednosti kovarijancne matrice. Sažimanje ulaznog prostora postiže se odabirom podskupa nekoliko prvih glavnih komponenata koje obuhvaćaju većinu varijacija u ulaznom prostoru, jer se pretpostavlja da nedominantne glavne komponente u latentni prostor većinom preslikavaju informaciju o šumu u ulaznim podacima te nekim manje važnim fenomenima u procesu.

Primjena PCA kriterija (10) omogućava izgradnju predikcijskog modela i na jako koreliranim visokodimenzionalnim podacima, a može dati dobre rezultate i kada je za izgradnju modela raspoloživ mali broj mjernih uzoraka. Međutim, uslijed neuzimanja u obzir ulazno-izlaznih odnosa pri procjeni parametara α , ovakvo preslikavanje ulaza može rezultirati modelom s lošim predikcijskim svojstvima.

PLS predstavlja važno proširenje PCA tehnike sa stajališta primjene u izgradnji predikcijskog modela na visokodimenzionalnim koreliranim podacima. Razlog tome je što PLS kriterij za procjenu parametara α , osim varijacija u ulaznom prostoru, obuhvaća i korelaciju između ulaznih veličina i izlaza. PLS kriterij se može definirati kao proširenje kriterija (10), na način:

$$\max_{\alpha_j} [\text{corr}^2(\mathbf{y}, \mathbf{X} \alpha_j) \text{var}(\mathbf{X} \alpha_j)]. \quad (11)$$

I u PLS slučaju rješenje problema optimizacije smjerova projekcije može se dati kroz rješenje svojstvenog problema. Za razliku od PCA, u PLS-u je središnji objekt analize međukorelacijska matrica ($\mathbf{X}^T \mathbf{Y}$), pa su optimalni smjerovi projekcije svojstveni vektori matrice $(\mathbf{X}^T \mathbf{Y})^T (\mathbf{X}^T \mathbf{Y})$. Komponenta u optimizacijskom kriteriju (11), koja se odnosi na maksimizaciju kvadrata korelacije između izlazne veličine i projiciranih ulaznih veličina, ekvivalentna je mini-

mizaciji srednje kvadratne pogreške modela (3), jer vrijedi [13]:

$$\max_{\alpha} [\text{corr}^2(\mathbf{y}, \mathbf{X}\alpha)] = \min_{\alpha} (\|\mathbf{y} - \hat{\mathbf{y}}\|^2). \quad (12)$$

Ovo znači da se PLS pri traženju optimalnih smjerova projekcije ulaznog prostora orijentira i na svojstva izlaznog prostora, odnosno da su PLS smjerovi projekcije za određeni iznos zakrenuti od onih za PCA.

Budući da PLS formira latentne varijable orijentirajući se i prema izlaznom prostoru, poželjno je provesti redukciju dimenzije preslikavanja jer regresija na latentnim varijablama dobivenim uz sve PLS smjerove projekcije u načelu rezultira prevelikom varijancom u predikciji. Pri tome se odabire nekoliko prvih smjerova projekcije koji formiraju latentne varijable koje će dati najbolju predikciju modela.

Kontinuum regresija, CR (*Continuum Regression*), je metoda nastala objedinjavanjem metoda temeljenih na linearnoj projekciji u jednu zajedničku metodu [10]. CR kriterij za procjenu parametara α obuhvaća čitav kontinuum od kriterija (10), preko (11), sve do kriterija (12) koji se u potpunosti orijentira na izlazni prostor. CR stoga omogućava optimizaciju smjera projekcije ulaznog prostora u latentni kroz proizvoljni odabir preraspodjele naglaska na obuhvaćanje odnosa među ulaznim veličinama i minimizaciju izlazne pogreške modela. CR kriterij se može definirati kao proširenje kriterija (11), na način:

$$\max_{\alpha_j} \left\{ [\text{corr}^2(\mathbf{y}, \mathbf{X}\alpha_j)]^{1+a-2a^2} [\text{var}(\mathbf{X}\alpha_j)]^{3a-2a^2} \right\}, \quad \text{za } a = 0 \dots 1. \quad (13)$$

Preraspodjela naglaska pri tome je određena kontinuum parametrom a . I u ovom slučaju rješenje problema optimizacije smjerova projekcije može se tražiti kroz rješenje svojstvenog problema. Pri tome je središnji objekt analize međukorelacijska matrica definirana kao $[(\mathbf{X}^T)^a(\mathbf{Y})^{(1-a)}]$.

Određivanje parametara α prema PCA i PLS kriterijima u literaturi su razrađeni i zasnivaju se na primjeni dekompozicije na singularne vrijednosti (SVD), kada se parametri određuju na simultani način, ili na primjeni NIPALS algoritma, kada se parametri određuju na hijerarhijski način, tj. jedan po jedan vektor α_j [14, 7]. Međutim, određivanje parametara modela prema CR kriteriju (13) općenito je složen matematički problem. U literaturi su stoga predloženi različiti načini kako odrediti ove parametre [15], pa CR metoda nije u potpunosti jednoznačna. Postoji nekoliko verzija CR metode, zasnovanih na različitim pojednostavljenjima i primjeni različitih matematičkih »trikova«. Jedno od

pojednostavljenja se odnosi na definiciju međukorelacijske matrice kao:

$$(\mathbf{X}^T)^\gamma(\mathbf{Y}), \quad \text{za } \gamma = 0, \dots, \infty, \quad (14)$$

na čemu je zasnovana u literaturi češće susretana definicija CR kriterijske funkcije kao:

$$\max_{\alpha_j} \left\{ [\text{corr}^2(\mathbf{y}, \mathbf{X}\alpha_j)] [\text{var}(\mathbf{X}\alpha_j)]^\gamma \right\}. \quad (15)$$

Kontinuum parametri γ i a se pri tome odnose kao:

$$\gamma = \frac{a}{1-a}, \quad \text{odnosno: } a = \frac{\gamma}{1+\gamma}. \quad (16)$$

CR metoda je vrlo važna sa stajališta izgradnje modela s dobrim predikcijskim svojstvima jer, uz odabir dimenzije modela (u ovom slučaju dimenzije preslikavanja J), kroz odabir iznosa kontinuum parametra posjeduje dopunski mehanizam za utjecaj na preraspodjelu pomak-varijanaca u predikciji. Na ovaj način se procjena parametara α može u značajnoj mjeri prilagoditi svojstvima raspoloživog skupa podataka za izgradnju modela [16], što je posebno važno kada se proces modelira na temelju pogonkih podataka.

Uz linearnu regresiju na linearnim latentnim varijablama model (7) poprima oblik:

$$f_m(\mathbf{x}, \Theta) = \sum_{j=1}^J \beta_j \sum_{n=1}^N \alpha_{nj} x_n. \quad (17)$$

Uz procjenu parametara α prema kriteriju (10) slijedi PCR (*Principal Component Regression*) model, dok se uz kriterij (11) dobije PLSR (*Partial Least Squares Regression*) model. Kada se za procjenu parametara α koristi kriterij (12) kaže se da nema projekcije ulaza, jer rezultirajući model poprima svojstva MLR modela [13]. Primjena CR kriterija za procjenu parametara α rezultira u najpoptenijem modelu iz ove klase modela, pri čemu PCR, PLSR i MLR predstavljaju samo specijalne slučajeve CR metode. Ovi specijalni slučajevi CR metode sada se mogu pregledno izložiti, na način:

kont. param.	MLR	PLSR	PCR
γ	0	1	∞
a	0	0.5	1

4. SJEDINJENJE CR METODE I MLP MREŽE

Neuronske mreže imaju strukturu modela (7) te su dobra osnova za izgradnju modela zasnovanih na preslikavanju ulaza u latentno područje. Osim toga, što je još važnije, neuronske mreže su otvorene matematičke strukture koje se u pogledu procjene parametara (težina) mogu nadograđivati razli-

čitim kriterijima i tehnikama. Najčešće korištena funkcija za modeliranje nelinearnog ulazno-izlaznog preslikavanja ima oblik [6]:

$$f_m(\mathbf{x}, \Theta) = \sum_{j=1}^J \beta_j \psi_j \left(\sum_{n=1}^N \alpha_{nj} x_n \right). \quad (18)$$

Model (18) predstavlja poopćenje linearnog regresijskog modela (17) na nelinearno modeliranje. To je ujedno specijalni slučaj modela (7) i zadržava njegova univerzalna aproksimacijska svojstva. Međutim, model (7) dopušta primjenu šire klase metoda preslikavanja ulaza i obuhvaća različite tipove neuronskih mreža. Oblik modela (18) je izravno primjenjiv na MLP neuronsku mrežu s jednim skrivenim slojem.

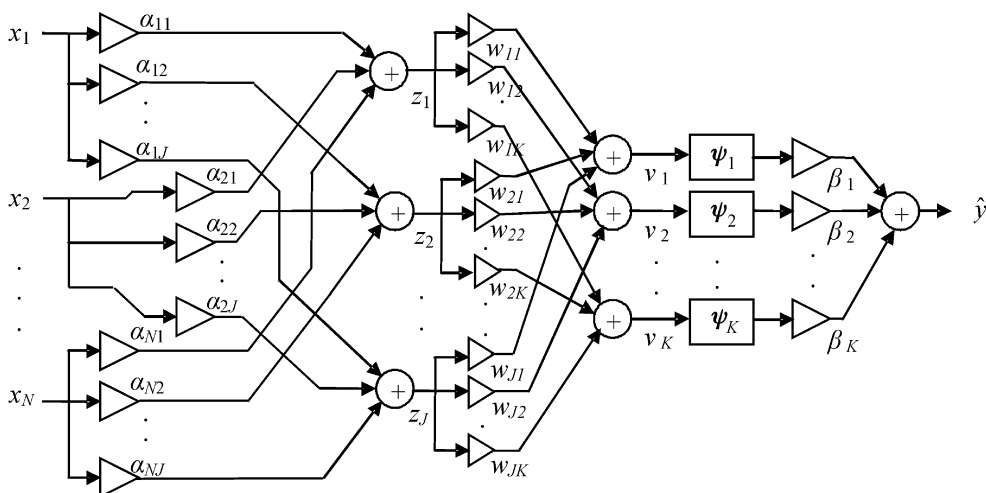
Sa stajališta izgradnje predikcijskog modela na temelju visokodimenzionalnog, koreliranog i onečišćenog skupa podataka MLP mreža ima bitne nedostatke, koji se u prvom redu očituju u potrebi za velikim omjerom između broja podataka za učenje i broja ulaza modela, te učenjem osjetljivim na početne uvjete i lokalne minimume. Razlog tome je njena nelinearna struktura, te činjenica da se podešavanje težina provodi kroz regresiju. U pogledu učenja na ovakvim podacima MLP mrežu je moguće unaprijediti kroz primjenu metoda zasnovanih na preslikavanju ulaza. Pri tome se koristi zasebni optimizacijski kriterij za procjenu ulaznih težina. Ovakva nadgradnja neuronske mreže realizira se kroz odgovarajuće proširenje uobičajenog postupka učenja, što rezultira hibridnim metodama modeliranja na podacima.

S druge strane, metode statističke multivarijantne analize nemaju spomenute probleme neuronskih mreža, ali ove tehnike u osnovi ne mogu modelirati nelinearne odnose među procesnim veličinama.

Komplementarnost između ovih dviju tehnika ukazuje na potencijalni dobitak kroz njihovu kombinaciju. Model (18) je pogodan za hibridizaciju s tehnikama multivarijantne analize, opisanim u prethodnom odjeljku, na način da se ove tehnike koriste za određivanje ulaznih težina mreže (α). Pri tome se ulazni prostor preslikava u potprostor varijabli na koje djeluju aktivacijske funkcije. Ovaj način učenja mreže u osnovi se može realizirati off-line (*batch*) učenjem, te predstavlja vrstu kombiniranog učenja u kojem se na odgovarajući način najprije određuju parametri α , a nakon toga se nekim od uobičajenih postupaka za učenje izlaznih težina mreže odrede regresijski parametri β .

Mreža s jednim nelinearnim skrivenim slojem (18) općenito može opisati svaki tip nelinearnosti. Međutim, kada se ulazne težine određuju prema kriteriju koji ne obuhvaća aktivacijske funkcije $\psi(\cdot)$ može nastupiti ograničenje u opisu nelinearnosti, jer se ne vodi računa o optimalnoj pobudenosti neurona. Na primjer, uz primjenu »tansig« aktivacijskih funkcija, mali iznos ulaznih težina pobuđuje neurone samo u linearnom dijelu te im je mala efikasnost u opisu ulazno-izlaznih nelinearnosti. S druge strane, pri velikim vrijednostima ulaznih težina neuroni dostižu zasićenje i postaju neupotrebivi za izgradnju modela. Ovaj se problem može izbjeći primjenom neuronske mreže s dva skrivena sloja, od kojih je prvi linearni, dimenzije J , a drugi nelinearni sloj, dimenzije K , kako je to prikazano na slici 2.

U prvom skrivenom sloju provodi se linearno preslikavanje ulaza, pri čemu se njegove težine određuju nekom od tehnika unutar CR-a. Drugi skriveni i izlazni sloj predstavljaju dio modela u kojem se provodi (nelinearna) regresija na latentnim varijablama (z), pa se težine drugog skrivenog sloja (w) i iz-



Sl. 2. Blokovska shema MISO NNCR modela s dva skrivena sloja, linearnim i nelinearnim

laznog sloja (β) određuju na standardni način, kroz minimizaciju izlazne pogreške modela (3), odnosno:

$$\min_{w, \beta} (\|y - \hat{y}\|^2). \quad (19)$$

Budući da je dimenziju drugog skrivenog sloja (K) moguće proizvoljno odrediti, jer ne ovisi o dimenziji ulaznog preslikavanja (J), ova struktura modela je fleksibilnija u opisu procesa, u odnosu na strukturu s jednim skrivenim slojem. Optimalna pobuđenost neurona je postignuta kroz podešavanje težina drugog skrivenog sloja (w) jer kriterij (19) obuhvaća aktivacijske funkcije. Međutim, i kada se koriste dva skrivena sloja, od kojih je prvi linearni sloj, jednostavnom primjenom linearne algebre problem se može svesti na problem izgradnje neuronske mreže s jednim nelinearnim skrivenim slojem. Razlog tome je što je model linearan sve do ulaza u aktivacijske funkcije, pa se varijable v (slika 2) mogu tretirati kao izlazi iz dijela modela u kojem se provodi (linearna) transformacija ulaznih veličina. Pri tome treba voditi računa da dimenzija preslikavanja ulaza nije jednaka broju neurona drugog skrivenog sloja (K), te da struktura prostora (vektorska baza) varijabli v nije definirana nekim kriterijem.

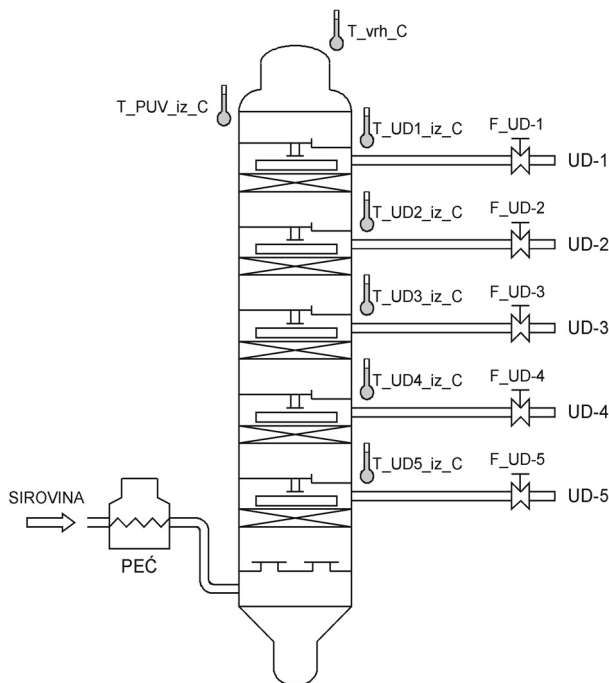
5. REZULTATI IZGRADNJE MODELA NA ODABRANOM PRIMJERU

Svojstva odabranih metoda za izgradnju modela procesa na temelju podataka analiziraju se na primjeru procesa destilacije nafte. Pri tome se na temelju raspoloživih mjerenja lako mjerljivih veličina estimiraju viskoznosti derivata postrojenja destilacijske kolone.

5.1. Opis procesa i načina formiranja skupova za izgradnju modela

Postrojenje za destilaciju nafte prikazano je na slici 3. Na ulaz postrojenja dovodi se sirova nafta koja se najprije predgrijava, a zatim vodi u destilacijsku kolonu. Ovdje se metodom vakuumske destilacije na različitim visinama postrojenja izdvaja nekoliko derivata koji predstavljaju izlazni proizvod. Važne karakteristike ovih derivata su: viskoznost, gustoća, temperatura isparavanja, palište, i druge. Navedene procesne veličine pripadaju kategoriji teško mjerljivih procesnih veličina, te im se iznos utvrđuje kroz laboratorijsku analizu.

Viskoznost je najvažniji pokazatelj kakvoće izlaznog proizvoda pa je važno omogućiti kontinuirani nadzor nad ovom veličinom. Na načelnoj shemi postrojenja viskoznosti derivata su označene svojim procesnim oznakama, od UD-1 do UD-5. Od lako mjerljivih procesnih veličina viskoznost derivata je najviše povezana s temperaturama i protocima na



Sl. 3. Pojednostavljena shema vakuumske destilacijske kolone

pojednim točkama postrojenja, te u manjoj mjeri i tlakovima nekih dijelova postrojenja, pa su ove lako mjerljive veličine potencijalni ulazi modela procesa za estimaciju viskoznosti. Za konačni odabir od pedesetak raspoloživih lako mjerljivih veličina pomogla je kvalitativna ocjena važnosti pojedinih lako mjerljivih veličina dana od strane tehnologa i operativnog osoblja postrojenja. Uz ove smjernice provedena je i provjera korelacije između predloženih lako mjerljivih veličina i veličina koje se estimiraju. Za ulaze modela procesa konačno je odabrano sedam temperatura (T) i pet protoka (F), koji su također naznačeni na načelnoj shemi postrojenja na slici 3. Kako se u radu estimiraju viskoznosti UD-1 i UD-3, modelira se dva procesa: INA-UD-1 i INA-UD-3.

Mjerni podaci o procesnim veličinama preuzeti su iz procesne baze podataka. Lako mjerljive veličine mjerene su svakih 5 minuta kroz period od oko 100 dana, dok su viskoznosti derivata određivane laboratorijskom analizom u približno istom razdoblju, ali s prosječnom periodom uzorkovanja od oko 3 sata. Na lako mjerljivim veličinama je provedena predobradba. Najprije su odstranjeni mjerni uzorci s grubim mjernim pogreškama, a zatim provedena filtracija niskopropusnim filtrom. Pri tome je dobiven skup od 804 mjerna uzorka s kompletnim podacima o ulazno-izlaznim veličinama. Prvih 655 uzoraka odabrani su za izgradnju modela procesa, a preostalih 149 za ispitivanje izgrađenih modela. Nakon ove predobradbe pristupilo se iden-

tifikaciji mjernih uzoraka s pogreškama (stršećim vrijednostima) čije vrijednosti ne izlaze iz očekivanog područja vrijednosti procesnih veličina. Budući da podaci s pogrešnim vrijednostima općenito imaju drugačiju korelacijsku strukturu, ovom vrlo složenom problemu pristupilo se na način da su se tražili uži vremenski intervali u kojima je korelacijska struktura podataka bitno drugačija od korelacijske strukture cijelog skupa podataka, tj. skupa »804«. Nakon većeg broja iteracija provjere podataka procijenjeno je da se podaci sa sumnjivim mjerenjima, s obzirom na redni broj mjernog uzorka, nalaze u intervalima: 91-180, 250-274, 461-470, 481-490, 600-604, 646-655, što obuhvaća 150 mjernih uzoraka. Na ovaj je način skup »655« podijeljen na skup »505«, u kojem se nalaze »čisti podaci«, i skup »150«, u kojem se nalaze podaci za koje se pretpostavlja da sadrže stršeće vrijednosti. U slučaju da ovakva predobradba ne daje zadovoljavajuće rezultate, mogu se primijeniti i druge metode detekcije i odstranjivanja mjernih uzoraka s pogreškama koje se mogu naći u literaturi [7, 4].

5.2. Izgrađeni modeli i postignuti rezultati

Na formiranim skupovima podataka za izgradnju modela izgrađeni su modeli procesa primjenom metoda opisanih u 3. i 4. odjeljku. S obzirom na namjenu, najvažnije svojstvo izgrađenih modela je njihova sposobnost predikcije. Vrijednovanje modela s ovog stajališta može se procijeniti na temelju estimacije izlaza na novim (ispitnim) podacima. Za pokazatelje kakvoće modela procesa stoga su odabrani srednja kvadratna pogreška (MSE) i koeficijent determinacije (R^2) na podacima za ispitivanje, te maksimalna postotna pogreška (MPErr) kao dopunski pokazatelj kakvoće modela. Srednja kvadratna pogreška modela na ispitnim podacima se definira kao:

$$MSE_t = \frac{1}{I_t} \mathbf{e}_t^T \mathbf{e}_t = \frac{1}{I_t} \sum_{i=1}^{I_t} e_t^2(i); \quad e_t = y_t - \hat{y}_t, \quad (20)$$

gdje je:

- e_t – pogreška modela na ispitnim podacima,
- y_t – podaci o izlaznoj veličini u ispitnim podacima,
- \hat{y} – izlaz modela procesa uz ispitne podatke dovedene na ulaz,
- I_t – broj ispitnih mjernih uzoraka.

Koeficijent determinacije predstavlja kvadrat korelacije između stvarne izlazne veličine i estimirane izlazne veličine [5], te je definiran kao:

$$R^2 = 1 - \frac{SSE}{SST} = 1 - \frac{\sum_i (y(i) - \hat{y}(i))^2}{\sum_i (y(i) - \bar{y})^2}, \quad 0 \leq R^2 \leq 1, \quad (21)$$

gdje je:

- SSE – suma kvadrata pogreške, odnosno rasipanje izlazne veličine oko njene estimacije,
- SST – ukupna suma kvadrata, odnosno rasipanje izlazne veličine oko njene srednje vrijednosti.

Budući da R^2 mjeri koliko se ukupne varijacije u podacima o izlaznoj veličini može objasniti modelom, ova statistika u dobroj mjeri pruža uvid u pouzdanost izgrađenog modela. Ako je modelom obuhvaćen mali postotak ukupnog rasipanja izlazne veličine ($R^2 \rightarrow 0$), odnosno ako je tzv. neobjašnjeno rasipanje preveliko, tada ili model nije dobar ili je slaba korelacija između ulaznih veličina i izlazne veličine. Određivanjem koeficijenta determinacije na podacima za ispitivanje (R^2 u tablici 1) može se dobiti dobar uvid u predikcijsku sposobnost modela.

Za izgradnju i ispitivanje modela procesa korišten je programski paket Matlab, naročito programske funkcije iz Neural Network Toolbox-a. Za određivanje parametara preslikavanja ulaza u latentno područje prema CR kriteriju u radu se koristi programski alat PLS-Toolbox [17], uz određenu modifikaciju funkcije »cr« [16], u kojem je funkcija kontinuuma regresije zasnovana na implementaciji CPR (*Continuum Power Regression*) metode [15]. Kako bi se smanjili računski zahtjevi i ubrzalo izračunavanje parametara, u CPR-u se provodi redukcija dimenzije problema unutar r -dimenzionalnog kanoničkog prostora matrice \mathbf{X} , uz primjenu ortogonalizacije matrice \mathbf{K} koja je formirana iz Krylove sekvence kovarijancne matrice $(\mathbf{X}\mathbf{X}^T)^r$ i vektora izlazne veličine \mathbf{y} . Pri tome r predstavlja rang matrice \mathbf{X} . Za izgradnju modela podaci su normirani usrednjavanjem ili standardizacijom, pri čemu se uz usrednjavanje podaci svode na jediničnu standardnu devijaciju. CPR metoda je u najvećem broju slučajeva davala nešto bolje rezultate na standardiziranim podacima.

Iz rezultata izgradnje linearnih modela (17), koji su sažeto prikazani u tablici 1, može se uočiti da MLR procjena parametara ni u jednom slučaju ne rezultira modelom s dobrim predikcijskim svojstvima. Samo se malo bolji rezultati dobiju ako se umjesto CPR metode (uz $\gamma = 0$) koristi uobičajeni način MLR procjene parametara [5], uz pseudoinvertiraju kovarijancne matrice ulaznih podataka. Parametri α PCR modela procijenjeni su prema kriteriju (10), korištenjem funkcije »pca«. Međutim, funkcija »cr« daje PCA preslikavanje ulaza već pri odabiru $\gamma > 4$. CPR algoritam implementiran u ovu funkciju stabilan je do vrijednosti kontinuuma parametra $\gamma = 8$. Na temelju rezultata danih u tablici 1 također se može uočiti da CR model postiže najbolju predikciju, te da optimalni smjerovi projekcije ulaza teže PCA smjerovima, naročito u slučaju INA-UD-1 procesa. Ova situacija je uvijek prisutna kada model ne može u dovoljnoj mjeri obuhvatiti

Tablica 1. Najbolji rezultati postignuti modelima obuhvaćenih kontinuum regresijom

Skup podataka za modeliranje	Pokazatelji kakvoće modela				dimen. preslik.	kontin. param.
	MSE _t	R _t ²	R ²	MPErr _t , %	J	γ
PCR model						
INA-UD1 655	0.00417	0.6840	0.8816	7.28	8	∞
INA-UD1 505	0.00486	0.6308	0.9448	12.2	8	∞
INA-UD1 150	0.00856	0.3518	0.6601	12.8	6	∞
INA-UD3 655	0.11484	0.3120	0.2958	8.95	8	∞
INA-UD3 505	0.10995	0.3413	0.3944	9.14	6	∞
INA-UD3 150	0.1517	0.0586	0.3199	10.3	8	∞
PLSR(CR) model						
INA-UD1 655	0.004977	0.6232	0.8970	8.82	5	1
INA-UD1 505	0.00544	0.5879	0.9430	12.5	7	1
INA-UD1 150	0.02341	-0.7720	0.7315	16.1	6	1
INA-UD3 655	0.08995	0.4611	0.3136	8.70	4	1
INA-UD3 505	0.093343	0.4408	0.4063	9.03	5	1
INA-UD3 150	0.3144	-0.8833	0.3805	12.8	4	1
MLR(CR) model						
INA-UD1 655	0.00924	0.3005	0.9198	13.1	-	0
INA-UD1 505	0.04609	-2.4888	0.9490	40.6	-	0
INA-UD1 150	0.04666	-2.5328	0.8975	22.2	-	0
INA-UD3 655	0.204	-0.2220	0.3568	10.8	-	0
INA-UD3 505	0.2955	-0.7701	0.4239	23.4	-	0
INA-UD3 150	0.604	-2.62	0.4518	15.9	-	0
CR model						
INA-UD1 655	0.004167	0.6845	0.8816	7.28	8	7
INA-UD1 505	0.00488	0.6307	0.9448	12.2	9	4
INA-UD1 150	0.006997	0.4703	0.6970	11.1	6	4.7
INA-UD3 655	0.078624	0.5290	0.3167	8.05	5	1.35
INA-UD3 505	0.086436	0.4822	0.4086	8.52	7	2.33
INA-UD3 150	0.10080	0.3961	0.2865	8.10	5	2.64

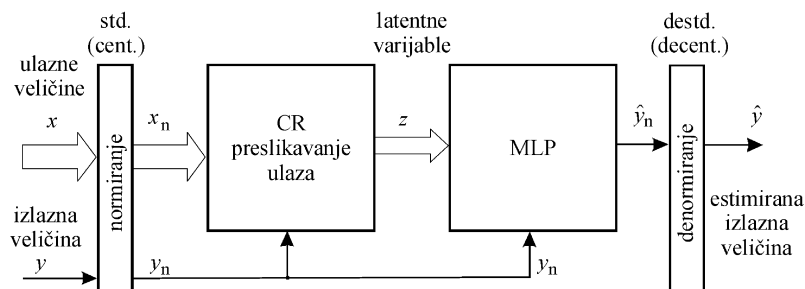
ulazno-izlazno vladanje procesa koji opisuje [7]. Razlog tome može biti slaba informativnost podataka za izgradnju modela, uslijed malog broja mjernih uzoraka i/ili male korelacije između ulaznih i izlaznih procesnih veličina i/ili visokog sadržaja stršećih vrijednosti u podacima. Veliko oslanjanje na ulazni prostor pri procjeni parametara uvijek zahtijeva veći broj latentnih veličina za dobru predikciju, odnosno veću dimenziju modela.

Razlog nemogućnosti (linearnog) modela u obuhvaćanju ulazno-izlaznog vladanja je i izražena nelinearnost procesa. Nizak iznos koeficijenta determinacije (R^2), te čak veća vrijednost ovog koeficijenta na ispitnim podacima, u dobroj mjeri ukazuju na veliku nelinearnost procesa INA-UD-3. Posljedica je da CR model rezultira velikom pogreškom te da ne može pružiti zadovoljavajuću predikciju izlazne veličine INA-UD-3 procesa. Modeli kojima R^2 na ispitnim podacima prelazi vrijednost 2/3 obećavajući su u pogledu predikcije, dok su modeli kojima je ovaj pokazatelj kakvoće ispod vrijednosti 1/3 potpuno beskorisni u ulozi prediktora. Stoga se za modeliranje INA-UD-3 procesa treba koristiti nelinearne modele.

Pri izgradnji nelinearnih modela, gdje se koriste iterativni numerički postupci procjene parametara,

radi postizanja dobrih predikcijskih svojstava modela poželjno je u fazu učenja uvesti međuvrjednovanje modela [6]. Stoga je, osim skupova za procjenu parametara i ispitivanje modela, potrebno raspolagati dopunskim skupovima podataka. Skupovi za međuvrjednovanje u ovom slučaju su formirani na način da su već pripremljeni skupovi za izgradnju modela na slučajan način dopunski podijeljeni, tako da je jedna trećina podataka odabrana za međuvrjednovanje, a preostale dvije trećine se koriste za procjenu parametara (težina) modela.

Za usporedbu s modelima izgrađenim odabranim metodama, izgrađen je i MLP model (18) INA procesa. Ovaj model je zasnovan na MLP neuronskoj mreži, realiziranoj s pomoću programskog alata Neural Network Toolbox. Pri tome su odabrane »tansig« aktivacijske funkcije i Levenberg-Marquardt (LM) algoritam učenja, uz početne vrijednosti parametara odabrane prema Nguyen-Widrow (NW) preporukama. Nakon 15 ponovljenih učenja, za dimenzije od 4 do 12 neurona u skrivenom sloju, uočeno je vrlo veliko rasipanje rezultata između pojedinih ponavljanja učenja, s često značajnim predikcijskim svojstvima modela ($R_t^2 \ll 1/3$). Ova prevelika osjetljivost mreže na početne vrijednosti parametara uzrokovana je u prvom redu vrlo složenim odnosima u podacima za učenje, čemu u najvećoj mjeri



Sl. 4. Učenje MLP mreže na CR latentnim varijablama

doprinosi prisutnost velikog broja stršećih vrijednosti. Iako je MLP model potencijalno sposobniji za opis ovih nelinearnih procesa, rezultati postignuti MLP modelom potpuno su usporedivi s rezultatima CR modela. Jasno je da se u ovom slučaju s jako velikim brojem ponavljanja učenja, uz različite početne vrijednosti parametara, mogu pronaći parametri mreže koji će rezultirati modelom boljih predikcijskih svojstava.

Kako bi se MLP neuronska mreža učinila efikasnijom u izgradnji modela procesa na temelju pogonskih podataka, ova tehnika se ujedinjuje s CR metodom, na način da mreža uči na latentnim varijablama dobivenim preslikavanjem ulaza prema CR kriteriju. Rezultirajući model je nazvan MLPpreCR modelom. Na slici 4 je shematski prikazan način realizacije modela ovog tipa.

Struktura izgrađenog MLPpreCR modela odgovara strukturi modela sa slike 2, uz »tansig« aktivacijske funkcije i jednaku dimenziju oba skrivena sloja. MLP mreža je strukturirana i učena na isti način kao u slučaju izgradnje MLP modela. Ispitivanjem izgrađenih modela uočeno je relativno malo rasipanje rezultata između pojedinih učenja MLP mreže na prethodno formiranim latentnim varijablama, pa su prikazani najbolji rezultati od samo 5 ponovljenih učenja mreže (slika 5). U promatranim primjerima izgradnje modela procesa MLPpreCR model uvijek postiže bolje rezultate u odnosu na MLP model. Ovo se naročito odnosi na slučaj modeliranja INA-UD-3 procesa. Razlog lakšem učenju MLP mreže na latentnim varijablama, nego na ulaznim veličinama, leži u tome što su ovako formirane latentne varijable međusobno nekorelirane, ali i u tome što je utjecaj stršećih vrijednosti i šuma iz ulaznog prostora u latentnom području značajno prigušen.

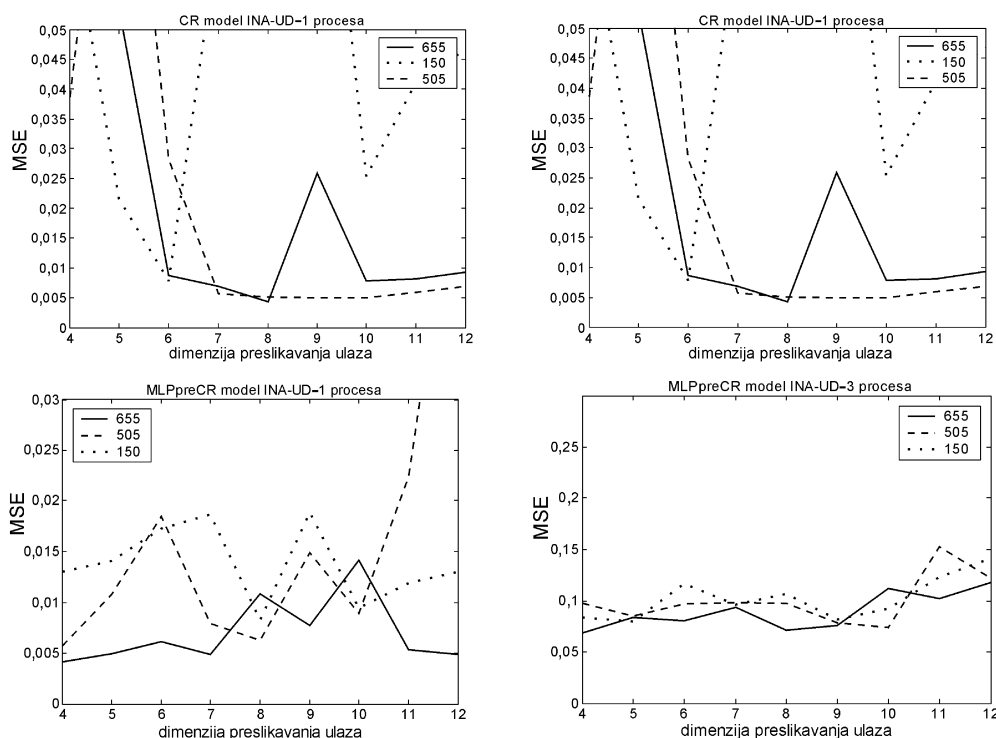
Iz prikaza rezultata na slici 5 uočava se da MLPpreCR model na malom broju latentnih varijabli postiže daleko bolju estimaciju izlazne veličine u odnosu na CR model. Dva su moguća razloga za to. Prvi je nelinearna struktura MLPpreCR modela, jer MLP mreža provodi nelinearnu regresiju na latentnim varijablama, uslijed čega može bolje obuhvatiti nelinearne ulazno-izlazne odnose u procesu

koji se modelira. Drugi razlog se može tražiti u mogućnosti da su se učenjem MLP mreže pronašli neki smjerovi projekcije ulaza izvan CR kriterija, tj. neki neortonormalni smjerovi, koji su rezultirali još informativnijim latentnim varijablama. Da je ovo moguće jasno je iz prikaza na slici 2. Naime, iako je prostor latentnih varijabli (z) formiran na temelju ortonormalnih smjerova projekcije ulaza, prostor v varijabli na ulazu u aktivacijske funkcije, nakon podešavanja ulaznih težina MLP mreže, ne mora imati to svojstvo.

U slučaju modeliranja INA-UD-3 procesa MLPpreCR model pokazuje najveću prednost u odnosu na sve ostale izgrađene modele. Činjenica da R^2 na ispitnim podacima gotovo dostiže vrijednost 0,6 (uz 4 latentne varijable), a maksimalna postotna pogreška pada na 7 %, značajan je pomak u pogledu kakvoće predikcije modela ovog procesa, izgrađenog na raspoloživim podacima.

Ovakvo unaprjeđenje MLP mreže, zasnovano na linearnom preslikavanju ulaza i nelinearnoj regresiji na latentnim varijablama, omogućava i funkcija »prePCA« koja se nalazi u Neural Network Toolbox-u. Dok funkcija »prePCA« preslikava ulazni prostor isključivo u smjerovima maksimalnih varijacija ulaznog prostora, preslikavanje CR metodom omogućava odabir smjera preslikavanja kroz cijeli kontinuum. Stoga primjena CR metode u modelu MLPpreCR predstavlja poopćenje funkcije »prePCA« i pruža dopunsku fleksibilnost u izgradnji modela ovakvog tipa.

Iako različiti modeli postižu jednaku srednju kvadratnu pogrešku na ispitnom skupu podataka, njihova točnost u estimaciji pojedinačnih vrijednosti izlazne veličine može se znatno razlikovati. Ova raspršenost izlaza modela ima veliko praktično značenje kada se model koristi kao estimator teško mjerljive procesne veličine. Veću glatkoću odziva modela moguće je postići dodavanjem niskopropusnog filtra na izlaz modela, pri čemu se maksimalna postotna pogreška u većini slučajeva smanjuje na polovicu svoje vrijednosti [16]. Ovo ukazuje da dodavanje dinamičke komponente u izgrađeni model procesa može značajno poboljšati kakvoću estimacije izlaza. Međutim, još nije razrađena me-



Sl. 5. Srednja kvadratna pogreška CR i MLPpreCR modela na ispitnim podacima

odologija kako iz pogonskih podataka s teško mjerljivim veličinama izlučiti informaciju o dinamičkom vladanju procesa.

Iako je pri formiranju skupa »505« na podacima provedena dugotrajna i složena analiza s ciljem otkrivanja mjernih uzoraka u kojima se nalaze pogreške, rezultati ispitivanja izgrađenih modela pokazali su da u konačnici nije dobiven puno informativniji skup za modeliranje u odnosu na osnovni skup »655«. Linearni modeli su bolje učili na skupu »505«, na što ukazuje i koeficijent determinacije (R^2), međutim, nelinearni modeli su »pronašli« dosta informacija o procesu i u skupu »150«. Prema tome, korištena metoda je rezultirala skupom »505« koji uglavnom sadrži mjerne uzorke s informacijom o linearnim ulazno-izlaznim odnosima u procesu, te skupom »150« u kojem su uzorci s mjernim pogreškama ali i oni što nose informaciju o nelinearnim ulazno-izlaznim odnosima. Budući da nije jednostavno razlučiti mjerne uzorke s pogreškama, od onih što nose informaciju o nelinearnosti procesa, upitno je da li bi druge metode za otkrivanje mjernih uzoraka s pogreškama dale bolje rezultate.

6. ZAKLJUČAK

Problemi procjene parametara modela proizlaze u prvom redu iz svojstava podataka za modeliranje. Izgradnja modela procesa s velikim brojem koreliranih ulaznih veličina na temelju pogonskih podata-

ka predstavlja poseban problem na području modeliranja procesa. Metode izgradnje modela zasnovane na preslikavanju ulaznog prostora u latentni prostor su metode koje i pod ovim uvjetima mogu rezultirati dobrom procjenom parametara. U ovim metodama model je podijeljen na dvije razine preslikavanja, što u osnovi omogućava bolje strukturiranje modela za određenu zadaću modeliranja, te zasebno podešavanje parametara ulaznog preslikavanja, čime se postiže efikasnija procjena parametara i time veća kakvoća modela.

Procjena parametara preslikavanja ulaza prema CR kriteriju dala je dobre rezultate, čak i kada se linearnim modelom opisuje nelinearni proces. CR model bi uz veći broj latentnih varijabli omogućio još točniju estimaciju izlazne veličine, međutim, ovako veliki broj parametara nije moguće dovoljno pouzdano procijeniti na temelju raspoloživih pogonskih podataka. Ovo je posebno izraženo u pogledu izgradnje modela INA-UD-3 procesa. Budući da su s linearnim modelima postignuti dobri rezultati uz CR procjenu parametara, poboljšanje učenja MLP neuronske mreže nastoji se postići kroz udruživanje s ovom metodom. Pri tome se provodi linearno preslikavanje ulaza prema CR kriteriju, a MLP mreža uči na latentnim varijablama dobivenim ovim preslikavanjem. Rezultirajući MLPpreCR model daje dobre rezultate u izgradnji modela procesa i na temelju manjeg skupa podataka, i uz prisutnost stršćih vrijednosti u podacima. Stoga se ovakav

način učenja MLP mreže predlaže za izgradnju predikcijskog modela procesa na pogonskim podacima.

Budući da je nelinearna regresija na latentnim varijablama dala bolje rezultate od linearne regresije, može se zaključiti da se i uz linearno preslikavanje ulaza u latentni prostor značajan dio informacije o nelinearnim ulazno-izlaznim odnosima procesa ipak prenese u latentni prostor. Međutim, za očekivati je da se primjenom pogodnog tipa nelinearnog preslikavanja ulaznog prostora mogu postići još informativnije latentne varijable, pa time i model procesa s boljim predikcijskim svojstvima.

LITERATURA

- [1] G. D. Gonzalez, **Soft Sensors for Processing Plants**. Proc. of the 2nd International Conference on Intelligent Processing and Manufacturing of Materials, IPMM'99, IEEE, Part vol. 1, pp. 59–69, 1999.
- [2] T. J. McAvoy, **Intelligent »Control« Applications in the Process Industries**. Annual Reviews in Control, 26, pp. 75–86, 2002.
- [3] D. Slišković, E. K. Nyarko, N. Perić, **Estimation of Difficult-to-Measure Process Variables Using Neural Networks – A Comparison of Simple MLP and RBF Neural Network Properties**. Proceedings of the 12th IEEE Mediterranean Electrotechnical Conference, Dubrovnik, pp. 387–390, 2004.
- [4] S. J. Qin, **Neural Networks for Intelligent Sensor and Control – Practical Issues and Some Solutions**. Chapter 8. in Neural Networks for Control, D. Elliott, Ed. Academic Press, 1996.
- [5] J. L. Devore, **Probability and Statistics for Engineering and the Sciences**. 3rd edition, Brooks/Cole Publishing Company, Pacific Grove, California, 1991.
- [6] S. Haykin, **Neural Networks – A Comprehensive Foundation**. 2nd edition, Prentice Hall, 1999.
- [7] H. Martens, T. Naes, **Multivariate Calibration**. 2nd edition, John Wiley & Sons, New York, 1991.
- [8] S. J. Qin, T. J. McAvoy, **A Data-Based Process Modeling Approach and Its Applications**. In reprints of the 3rd IFAC Dycord Symposium, pp. 321–326, 1992.
- [9] S. J. Qin, T. J. McAvoy, **Nonlinear PLS Modeling Using Neural Networks**. Computers and Chemical Engineering, 16 (4), pp. 379–391, 1992.
- [10] M. Stone, R. J. Brooks, **Continuum Regression: Cross-validated Sequentially Constructed Prediction Embracing Ordinary Least Squares, Partial Least Squares and Principal Components Regression**. Journal of the Royal Statistical Society, Ser. B, 52, pp. 237–269, 1990; Corrigendum in Ser. 54, pp. 906–907, 1992.
- [11] R. Brooks, M. Stone, **Joint Continuum Regression for Multiple Predictands**. Journal of the American Statistical Association, Vol. 89, No. 428, pp. 1374–1377, 1994.
- [12] ..., **Encyclopedia of Mathematics**. Kluwer Academic Publishers, 1997.
- [13] B. R. Bakshi, U. Utojo, **Unification of Neural and Statistical Modeling Methods that Combine Inputs by Linear Projection**. Computers and Chemical Engineering, Vol. 22, No. 12, pp. 1859–1878, 1998.
- [14] P. Geladi, B. R. Kowalski, **Partial Least-squares Regression: A Tutorial**. Analytica Chimica Acta, 185, pp. 1–17, 1986.
- [15] S. DeJong, B. M. Wise, N. L. Ricker, **Canonical Partial Least Squares and Continuum Power Regression**. Journal of Chemometrics, 15, pp. 85–100., 2001.
- [16] D. Slišković, **Procjena teško mjerljivih procesnih veličina na temelju pogonskih podataka**. Doktorska disertacija, Fakultet elektrotehnike i računarstva, Zagreb, 2005.
- [17] B. M. Wise, N. B. Gallagher, R. Bro, J. M. Shaver, **PLS_Toolbox – for Use with MATLAB®, User's Guide**. Eigenvector Research, 2003.

Continuum Regression in Process Modeling Based on Plant Data. Important process variables which give information about the final product quality cannot often be measured by a sensor but their value is determined based on laboratory analysis. In order to perform a continuous monitoring of a process variable and an efficient process control, it is necessary to estimate this difficult-to-measure process variable, i.e. to determine it on the basis of a mathematical model. However, to build an appropriate process model in many cases there are available only process measurement data stored in a process data base.

This paper gives appropriate methodology for process modeling based on plant data. Regression methods based on input space projection into a latent subspace are proposed to build a model. The paper investigates, in particular, properties of continuum regression (CR). As neural networks present a good basis for data based model building, possibility of hybridization of multilayer perceptron (MLP) neural network with CR method is additionally investigated. The aim of that is to use good properties of both methods and to avoid their weaknesses in process model building based on plant data. Advantages of the proposed methods for process model building as compared to the usually used regression methods are demonstrated by the modeling of crude oil distillation process based on the measuring data available.

Key words: process modeling, plant data, difficult-to-measure process variable estimation, projection into a latent space, continuum regression, neural networks

NASLOVI AUTORA

Dr. sc. Dražen Slišković
Elektrotehnički fakultet
Sveučilišta u Osijeku
K. Trpimira 2b, 31000 Osijek

Prof. dr. sc. Nedjeljko Perić
Fakultet elektrotehnike i računarstva
Sveučilišta u Zagrebu
Unska 3, 10000 Zagreb

Prof. dr. sc. Ivan Petrović
Fakultet elektrotehnike i računarstva
Sveučilišta u Zagrebu
Unska 3, 10000 Zagreb

E-mail: drazen.sliskovic@etfos.hr
nedjeljko.peric@fer.hr
ivan.petrovic@fer.hr

Primljeno: 2005-12-29