

# Praćenje i vođenje procesa kristalizacije primjenom procesne analitičke tehnologije

---

Duić, Tea

Master's thesis / Diplomski rad

2023

*Degree Grantor / Ustanova koja je dodijelila akademski / stručni stupanj:* **University of Zagreb, Faculty of Chemical Engineering and Technology / Sveučilište u Zagrebu, Fakultet kemijskog inženjerstva i tehnologije**

*Permanent link / Trajna poveznica:* <https://urn.nsk.hr/urn:nbn:hr:149:817809>

*Rights / Prava:* [In copyright](#)/[Zaštićeno autorskim pravom.](#)

*Download date / Datum preuzimanja:* **2024-07-13**



*Repository / Repozitorij:*

[Repository of Faculty of Chemical Engineering and Technology University of Zagreb](#)



SVEUČILIŠTE U ZAGREBU  
FAKULTET KEMIJSKOG INŽENJERSTVA I TEHNOLOGIJE  
SVEUČILIŠNI DIPLOMSKI STUDIJ

Tea Duić

# **DIPLOMSKI RAD**

Zagreb, rujan 2023.

SVEUČILIŠTE U ZAGREBU  
FAKULTET KEMIJSKOG INŽENJERSTVA I TEHNOLOGIJE  
SVEUČILIŠNI DIPLOMSKI STUDIJ KEMIJSKOG INŽENJERSTVA

Tea Duić

Praćenje i vođenje procesa kristalizacije primjenom procesne analitičke  
tehnologije

DIPLOMSKI RAD

Mentor rada: prof. dr. sc. Nenad Bolf

Izravni voditelj: Josip Sacher, mag. ing. cheming.

Članovi ispitnog povjerenstva:

prof. dr. sc. Nenad Bolf

prof. dr. sc. Jasna Prlić Kardum

doc. dr. sc. Željka Ujević Andrijić

Zagreb, rujan 2023.

*Zahvaljujem se prof. dr. sc. Nenadu Bolfu i mag. ing. cheming. Josipu Sacheru koji su me vodili u izradi ovog rada, na njihovom strpljenju i stručnoj pomoći.*

*Zahvaljujem se doktorandici Matei Gavran na svojoj pruženoj pomoći i savjetima te strpljenju tijekom provedbe eksperimenata.*

*Posebno se zahvaljujem svojoj obitelji, dečku i vjernim prijateljicama Dori, Nini i Marti na svojoj podršci i pomoći u teškim trenucima studiranja.*

## SAŽETAK

Kristalizacija je jedinična operacija od velike važnosti u brojnim industrijskim granama, kako u kemijskoj i prehrambenoj industriji, tako i u farmaceutskoj industriji. Kroz povijest farmaceutske industrije kristalizacija je bila najvažniji proces odvajanja i pročišćavanja. U današnje vrijeme velika pažnja posvećuje se vođenju procesa kristalizacije.

Procesna analitička tehnologija (PAT) je sustav za projektiranje, analizu i vođenje proizvodnih procesa praćenjem kritičnih atributa kvalitete i rada. PAT primjenjuje napredne analitičke instrumente pomoću kojih se dobivaju podaci za razvoj i primjenu kalibracijskih modela.

Cilj rada bio je razviti kalibracijske modele za praćenje koncentracije ceritiniba forme A u tetrahidrofuranu primjenom parcijalne regresijske metode najmanjih kvadrata i umjetne neuronske mreže. Izmjereni podaci su predobrađeni, definiran je broj latentnih varijabli za kalibraciju PLSR modela, odabrana je struktura neuronske mreže, vrsta prijenosne aktivacijske funkcije i trening funkcija za izvedbu modela umjetnih neuronskih mreža. Izvedeni su zaključci o valjanosti pojedinih modela za praćenje koncentracije ceritiniba forme A u tetrahidrofuranu na temelju kriterija  $R^2$  i RMSE.

### **Ključne riječi:**

kristalizacija, procesna analitička tehnologija, ATR-FTIR spektrometrija, PLSR, umjetna neuronska mreža, ceritinib forme A

## SUMMARY

Crystallisation is a unit operation of great importance in numerous industries - the chemical industry, the food industry and the pharmaceutical industry. Throughout the history of the pharmaceutical industry, crystallisation has been the most important separation and purification process. Nowadays, a lot of attention is paid to the crystallisation process control.

Process Analytical Technology (PAT) is a system for designing, analyzing and managing production processes by monitoring critical quality and performance attributes. PAT applies advanced analytical instruments that provide data for the development and application of calibration models.

The aim of the work was to develop calibration models to predict the concentration of ceritinib form A in tetrahydrofuran using partial least squares regression and an artificial neural network. The measured data were preprocessed and the number of latent variables for the calibration of the PLSR model was determined. The number of neurons in the hidden layer of the neural network structure was optimised, the type of activation function and training function was chosen and the algorithm for optimising the neural network model. Based on the performance criteria  $R^2$  and RMSE, conclusions were drawn about the validity of each model for monitoring the concentration output of ceritinib form A in tetrahydrofuran.

### **Keywords:**

crystallization, process analytical technology, ATR-FTIR spectrometry, PLSR, artificial neural network, ceritinib form A

# Sadržaj

|  |    |
|--|----|
| 1. UVOD  | 1  |
| 2. TEORIJSKI DIO   | 2  |
| 2.1. Kristalizacija  | 2  |
| 2.1.1. Topljivost  | 3  |
| 2.1.2. Metastabilna zona   | 6  |
| 2.1.3. Nukleacija i rast kristala                                  | 7  |
| 2.2. Kristali  | 12 |
| 2.2.1. Morfologija kristala  | 13 |
| 2.2.2. Polimorfija   | 13 |
| 2.3. Procesne analitičke tehnologije                               | 15 |
| 2.3.1. ATR-FTIR spektroskopija                                     | 17 |
| 2.4. Kemometrija   | 18 |
| 2.5. Strojno učenje  | 18 |
| 2.6. Razvoj modela strojnog učenja                                 | 19 |
| 2.6.1. Prikupljanje, predobrada i podjela podataka                 | 20 |
| 2.7. Analiza glavnih komponenata                                   | 21 |
| 2.8. Parcijalna regresija metodom najmanjih kvadrata               | 23 |
| 2.9. Umjetne neuronske mreže                                       | 26 |
| 2.9.1. Prijenosna funkcija   | 27 |
| 2.9.2. Postupak učenja neuronske mreže                             | 28 |
| 2.9.3. Broj iteracija i pretreniranost neuronske mreže             | 30 |
| 3. EKSPERIMENTALNI DIO   | 31 |
| 3.1. Materijali  | 31 |
| 3.2. Uređaji   | 32 |
| 3.3. Eksperimenti  | 33 |
| 3.3.1. Prikupljanje podataka                                       | 33 |
| 3.3.2. Eksperimentalno praćenje kristalizacije linearnim hlađenjem | 34 |
| 4. REZULTATI I RASPRAVA  | 35 |

|  |    |
|--|----|
| 4.1. Predobrada spektralnih podataka                       | 35 |
| 4.2. Razvoj kalibracijskog modela                          | 35 |
| 4.2.1. PLSR metoda   | 37 |
| 4.2.2. Umjetne neuronske mreže                             | 43 |
| 4.2.3. Praćenje procesa kristalizacije linearnim hlađenjem | 48 |
| 5. ZAKLJUČAK   | 49 |
| 6. LITERATURA  | 50 |



## 1. UVOD

Kristalizacija je od velike važnosti u brojnim industrijskim granama, kako u kemijskoj i prehrambenoj industriji, tako i u farmaceutskoj industriji. Glavni cilj provođenja kristalizacije je dobivanje čvrstog produkta visoke čistoće, uniformne veličine i točno definiranih svojstava, stoga je ključna za dobivanje proizvoda visoke kvalitete [1]. Kroz povijest farmaceutske industrije kristalizacija je bila najvažniji proces odvajanja i pročišćavanja. U današnje vrijeme je skrenuta velika pažnja u smjeru vođenja procesa kristalizacije što zahtijeva bolje razumijevanje osnova kristalizacije kao i radnih karakteristika opreme za kristalizaciju [2].

U farmaceutskoj industriji naglasak je na djelatnoj tvari (engl. *active pharmaceutical ingredient*, API). Najvažnije karakteristike API-ja su biokompatibilnost, odnosno biološka prihvatljivost u organizmu (nenanošenje štete organizmu, već ciljano djelovanje na pojedini dio organizma) i kvaliteta. Uz sve navedeno, neki od važnih pojmova za definiranje djelatne tvari su: topljivost, morfologija, raspodjela veličina čestica, polimorfija koji zajedno utječu na konačno djelovanje lijeka.

Uspoređujući druge separacijske metode i jedinične operacije, kristalizacija nije strogo definiran postupak, bar ne do te razine koja je postignuta za npr. destilaciju, ekstrakciju ili adsorpciju. To se brzo mijenja sa sve većim brojem istraživanja i automatizacijom procesa kristalizacije uz istovremeno postizanje visokih standarda, za što je potrebna i adekvatno projektirana oprema [1].

Procesna analitička tehnologija (engl. *process analytical technology*, PAT) definirana je kao sustav za projektiranje, analizu i vođenje procesa putem pravovremenih mjerenja (tj. tijekom samog procesa) kritičnih atributa kvalitete i performansi [3].

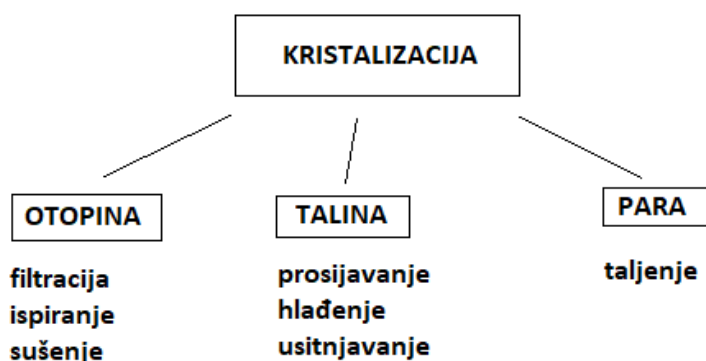
Brzi razvoj spektrometrijske tehnologije pratio je i razvoj odgovarajućih instrumenta za *in-situ* mjerenja u svrhu praćenja procesa kristalizacije. Prigušena totalna refleksija Fourier transformacijske infracrvene spektroskopije (eng. *attenuated total reflectance fourier transform infrared spectroscopy*, ATR-FTIR) je jedna od spektrometrijskih tehnika koja se primjenjuje za mjerenje koncentracije u otopini, a ključni korak je izrada kalibracijskoga modela.

U ovom diplomskom radu proveden je eksperiment na osnovi topljivosti ceritiniba forme A u tetrahidrofuranu. Iz izmjerenih podataka izveden je kalibracijski model za praćenje koncentracije djelatne tvari ceritiniba forme A.

## 2. TEORIJSKI DIO

### 2.1. Kristalizacija

Kristalizacija je toplinski separacijski proces za pročišćavanja. Može se definirati kao proces pri kojem dolazi do izdvajanja čvrste faze, odnosno formiranja kristala iz otopine, taline ili parne faze (slika 1). Kristali su čvrste tvari građene od atoma, iona ili molekula raspoređenih u periodički ponavljajući uzorak koji se proteže u tri dimenzije [4].



**Slika 1.** Načini provođenja kristalizacije [5]

Kako bi došlo do kristalizacije mora postojati pokretačka sila. Prezasićenost je pokretačka sila koja omogućuje kristalizaciju. Definira se razlika između trenutne koncentracije otopine i ravnotežne koncentracije i postiže se selektivnim uklanjanjem otapala ili promjenom topljivosti otopljene tvari u otopini (smanjenjem temperature ili promjenom sastava otapala). To je termodinamički nestabilno stanje, a do kristalizacije dolazi kako bi se sustav vratio u stanje ravnoteže. Jedna od ključnih faza procesa kristalizacije je proces nukleacije kojom nastaju centri kristalizacije nakon čega slijedi rast kristala [5].

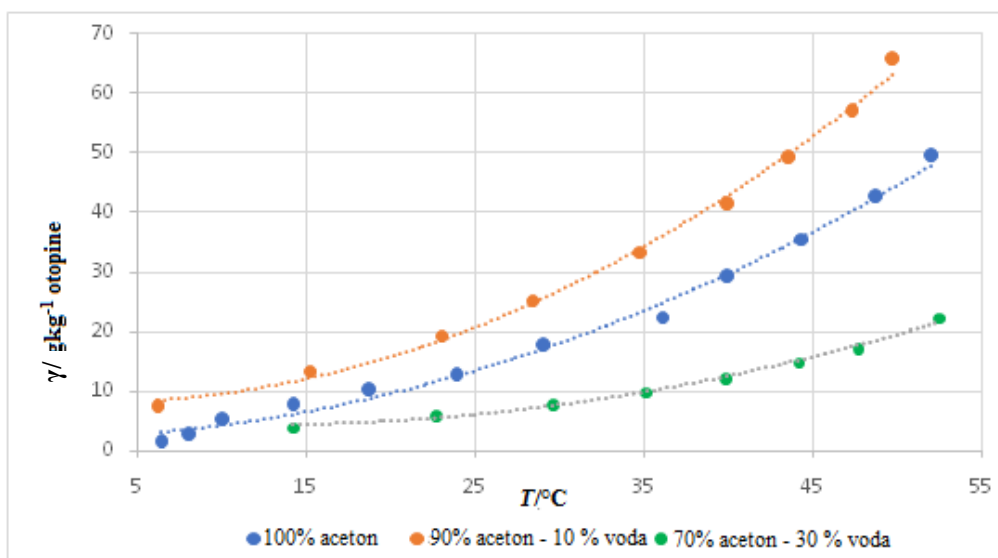
Gotovo svi industrijski procesi uključuju kristalizaciju iz otopine. Otopina je homogena smjesa dvaju ili više tvari koju čine otapalo i otopljena tvar. Da bi se procesom kristalizacije dobili čvrsti i suhi kristali potrebno je odvojiti kristale od kapljevite faze filtracijom i nakon toga ih osušiti. Kristalizacija iz otopine može se provesti na dva načina, bez pomoćne komponente – isparavanjem otapala, hlađenjem ili zaleđivanjem, ili s pomoćnom komponentom – isoljavanjem, kemijskom reakcijom, adukcijskom kristalizacijom ili raspršivanjem [5].

### 2.1.1. Topljivost

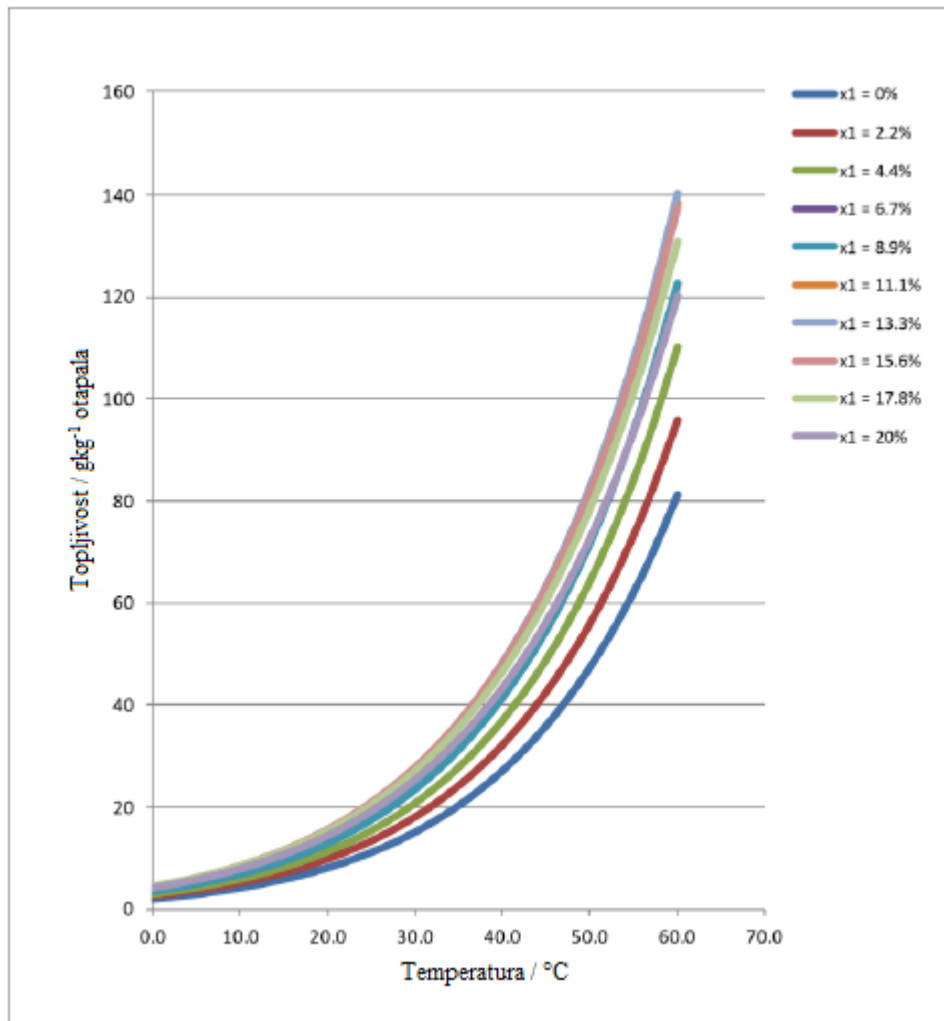
Otopina je homogena smjesa dvije ili više tvari koja se sastoji od otapala u suvišku i otopljene komponente. Sastav otopine izražava se koncentracijom otopljene tvari, najčešće kao maseni odnos ili maseni udio [5]. Prilikom procesa kristalizacije potrebno je poznavati topljivost pojedine tvari u pojedinom otapalu. Slikom 2 prikazana je topljivost pojedinih tvari pri čemu je grafički prikazana maksimalna masa pojedinih tvari koja se može otopiti u 100 g vode pri definiranoj temperaturi. Maksimalna količina tvari koja se može otopiti u nekom otapalu ovisi o prirodi te tvari i otapala, kao i o temperaturi. Kao što je vidljivo na slici 2, topljivost tvari s porastom temperature uglavnom raste.

Odnos između tvari i otapala bitan je za topljivost. Jake privlačne sile između tvari i otapala rezultiraju većom, dok slabe privlačne sile rezultiraju manjom topljivošću. Kao posljedica toga, polarne će se tvari otapati u polarnim otapalima, a nepolarne u nepolarnim. Topljivost se za neke sustave može objasniti i mogućnošću stvaranja vodikovih veza između tvari i otapala. Ako jedna komponenta djeluje kao donor, a druga kao akceptor vodikove veze tvar će biti topljiva u otapalu [5].

Količina otopljene tvari potrebna za dobivanje zasićene otopine pod danim uvjetima naziva se topljivost. Varijacija topljivosti prema temperaturi, pH razini, tlaku, prisutnosti aditiva i vrsti otapala je osnova za dizajn procesa kristalizacije (slike 2 i 3) Štoviše, fazni dijagram pruža korisne podatke za projektiranje procesa kristalizacije za dvokomponentni sustav.



**Slika 2.** Promjena topljivosti djelatne tvari s porastom temperature u različitim sastavima otapala [6]

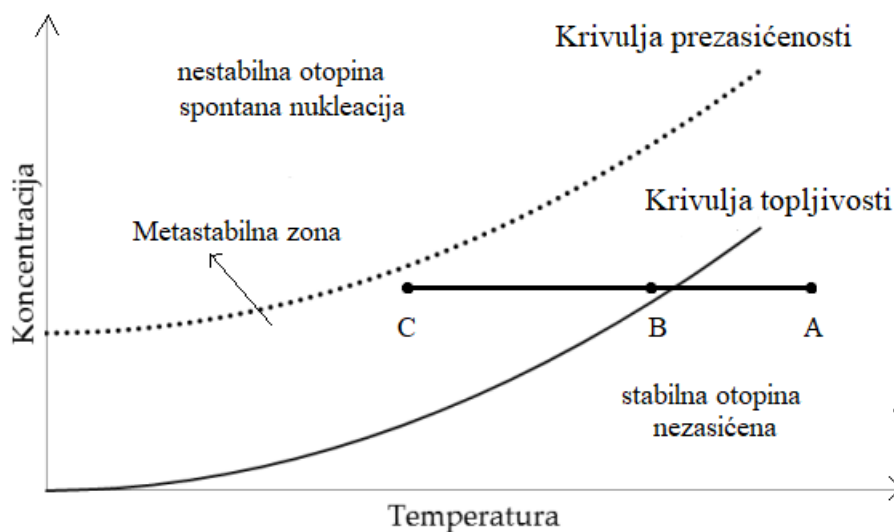


**Slika 3.** Ovisnost topljivosti djelatne tvari o temperaturi i masenom postotku vode u smjesi otapala [6]

Sastav otopine može se kvantitativno izraziti [7]:

- masenim, množinskim ili volumnim omjerom otopljene tvari prema otapalu,
- masenom, volumnom ili množinskom koncentracijom,
- masenim, množinskim ili volumnim udjelom i
- molalnošću.

Ovisnost koncentracije otopljene tvari o temperaturi prikazuje se dijagramom topljivosti, (slika 4). Donja krivulja predstavlja krivulju topljivosti, odnosno zasićenja ili ravnotežnu krivulju. Područje ispod krivulje topljivosti je stabilno područje nezasićene otopine u kojem neće doći do kristalizacije. Gornja krivulja predstavlja krivulju prezasićenosti iznad koje je područje nestabilne prezasićena otopina. Dovođenjem otopine u nestabilno područje dolazi do spontane nukleacije zbog dovoljno velike pokretačke sile. Područje između krivulje topljivosti i krivulje prezasićenosti naziva se metastabilno područje. Svaka otopina ima maksimalnu količinu tvari kojom može biti prezasićena (izvan granice topljivosti) prije nego što postane nestabilna. Na slici 4 prikazane su tri karakteristične točke, koje predstavljaju tijekom hlađenja otopine. Otopina definirana točkom A hladi se pri čemu se stanje otopine mijenja shodno liniji koja prolazi kroz točke B i C. Čim se prijeđe stanje zasićenja (točka B u dijagramu - metastabilno područje) postoji mala pokretačka sila. Ipak, do spontane nukleacije neće doći unutar metastabilnog područja, već tek nakon što se sustav ohladi na temperaturu nižu od one koja odgovara točki C odnosno kad otopina postane prezasićena [4, 5].



**Slika 4.** Prikaz metastabilne zone [4]

Zasićena je otopina u ravnoteži sa krutom fazom, odnosno kristalima, dok prezasićena otopina sadrži suvišak otopljene komponente. Kod prezasićene otopine je koncentracija tvari veća od koncentracije zasićenja [5].

Uvriježeno je da se prezasićenosti smatra pokretačkom silom za proces kristalizacije. Zapravo, prava pokretačka snaga je razlika između kemijskog potencijala otopljene tvari u kapljevitost i u čvrstoj fazi. Posebno je ovaj pristup važan u reakcijskoj kristalizaciji. Stvaranje kristala uvijek se odvija u nestabilnim otopinama (područje prezasićenosti) i do njega dolazi jer se sustav želi vratiti u stanje ravnoteže. Zbog činjenica da apsolutni kemijski potencijal nije lako izmjeriti, pokretačka sila kristalizacija se obično izražava kao koncentracija otopljene tvari koja je veća od njezine ravnotežne koncentracije, tj. topljivosti [4].

Unutar metastabilne zone mala je promjena koncentracije potrebna za promjenu stanja. Međutim, ako se malo prijeđe krivulja topljivosti neće uvijek doći do kristalizacije iako će s vremenom nastati određena količina kristala. S povećanjem koncentracije smanjuje se metastabilnost otopine, a stanje otopine se približava prezasićenoj otopini [5].

Prezasićenost otopine,  $\Delta c$ , izražava se kao razlika u koncentraciji otopine,  $c$ , i ravnotežne koncentracije,  $c^*$ :

$$\Delta c = c - c^* \quad (1)$$

ili kao relativna prezasićenost,  $S$ :

$$S = \frac{c}{c^*} \quad (2)$$

### 2.1.2. Metastabilna zona

Područje faznog dijagrama omeđeno krivuljom topljivosti i krivuljom prezasićenost naziva se metastabilna zona. To je područje metastabilne ravnoteže na kojem uslijed dugog zadržavanja u tom području postoji mala vjerojatnost da će doći do spontane nukleacije. Kod metastabilne zone važna je širina metastabilne zone jer utječe na karakteristike dobivenih kristala (oblik, veličina i raspodjela) i govori o stabilnosti otopine. Što je širina metastabilne zone veća, to je otopina stabilnija. Parametri koji utječu na širinu metastabilne zone su: vrijeme, brzina hlađenja ili isparavanja, intenzitet miješanja i potencijalne nečistoće [4].

Za promatranje utjecaja temperature na širinu metastabilne zone potrebno je definirati ravnotežnu temperaturu zasićenja,  $T^*$  kao temperaturu pri kojoj se sustav nalazi u stanju termodinamičke ravnoteže. Eksperimentalno je dokazano da je smanjenje širine metastabilne zone proporcionalno smanjenju maksimalne prezasićenosti otopine,  $S_{max}$ . Smanjenjem ravnotežne temperature prezasićenosti smanjuje se maksimalna prezasićenost otopine, a time i širina metastabilne zone [4,5].

Porastom intenziteta, odnosno brzine miješanja mijenjaju se hidrodinamički uvjeti za prijenos topline i tvari pri čemu se smanjuje širina metastabilne zone. S porastom brzine vrtnje miješala ostvaruju se povoljniji hidrodinamički uvjeti u kristalizadoru, drugim riječima smanjuju se otpori prijenosu količine gibanja, topline i tvari. Ako su otpori manji, veća će biti brzina prijenosa tvari i ugradnja strukturnih jedinica u kristalnu rešetku. Dakle, potrebna će biti manja prezasićenost otopine (pokretačka sila) za stvaranje stabilnog nukleusa. Minimalni otpor prijenosu tvari postiže se pri optimalnoj brzini miješanja, a daljnjim porastom brzine miješanja ne postižu se povoljniji hidrodinamički uvjeti [5].

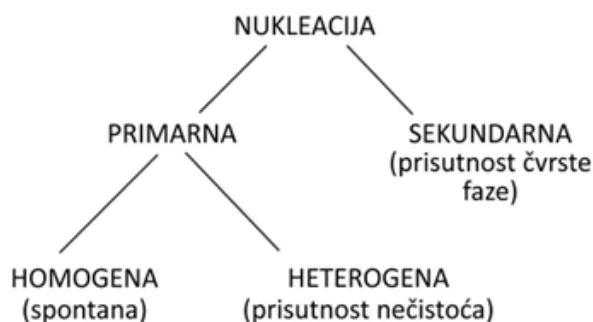
Veća brzina hlađenja dovodi do stvaranja većeg broja sitnijih kristala. Što je veća brzina hlađenja to je šira metastabilna zona, dok potencijalna prisutnost nečistoća dovodi do suženja metastabilne zone. Utjecaj nečistoća na proces kristalizacije je nepogodan [5].

### 2.1.3. Nukleacija i rast kristala

Kristalizacija iz otopina je složen proces koji se odvija u nekoliko faza. Prva faza je stvaranje prezasićene otopine jer se spontana pojava kristala može dogoditi samo kada je sustav u neravnotežnom prezasićenom stanju. U sljedećoj fazi molekule otopljene u otopini počinju se agregirati kako bi smanjile prezasićenost i pomaknule sustav prema ravnoteži [4]. Proces molekularne agregacije na kraju dovodi do stvaranja jezgri koje mogu djelovati kao centri kristalizacije. Nukleusi, stabilni centri nukleacije, mogu se definirati kao minimalna količina nove faze (kristala) sposobna za samostalno postojanje i spontani rast [5].

Nukleacija se smatra ključnim korakom u procesu kristalizacije. Tijekom nukleacije nastaju čestice malih dimenzija čije mjesto nastanka u sustavu nije moguće predvidjeti. Nakon nukleacije slijedi proces rasta kristala [5].

Nukleacija se dijeli na primarnu nukleaciju koja može biti homogena ili heterogena i na sekundarnu, što je prikazano na slici 5 [8].



Slika 5. Vrste nukleacije [8]

Homogena nukleacija je spontana nukleacija koja započinje u prezasićenoj otopini, u sustavu u kojem je prisutna samo jedna faza. Posljedica je sudara molekula i iona zbog njihove male kinetičke energije koja nije dovoljna da prevlada sile privlačenja kada su oni u neposrednoj blizini pri čemu nastaju nukleusi. Nakon nastanka primarnih nukleusa, rast nukleusa nastupa u vrlo kratkom vremenu u širokom rasponu temperatura istovremeno. Kristali se moraju oduprijeti ponovnom otapanju i orijentirati se u kristalnoj rešetki [4].

Klasična teorija nukleacije (engl. *classical nucleation theory*, CNT) je najjednostavnija i najčešća teorija koja opisuje proces nukleacije. Iako je izvorno izveden za kondenzaciju pare u kapljevini, također "po analogiji" objašnjava stvaranja kristala iz prezasićenih otopina (Mullin 1997) [4]. Prema ovoj teoriji, u metastabilnoj zoni dolazi do stvaranja prednukleacijskih klastera, koji se istovremeno stvaraju i raspadaju, te nastaju novi, dok ne dostignu određenu kritičnu veličinu. Klasteri koji postignu kritičnu veličinu postaju stabilni i nastavljaju spontano rasti, što rezultira stvaranjem kristalnih jezgri, odnosno nukleusa. Stabilni nukleus nastaje pri određenom stupnju prezasićenosti ukoliko je savladana energetska barijera tj. kritična prezasićenost,  $S^*$ . Kritična prezasićenost raste s porastom razlike kemijskog potencijala  $\Delta\mu$ , a smanjuje se s porastom termodinamičke temperature  $T$  što je vidljivo iz izraza (3) [4].

$$\ln S^* = \frac{\Delta\mu}{RT} \quad (3)$$

$$\Delta\mu = \mu_1 - \mu_2 \quad (4)$$

Termodinamički opis izvorno je krajem devetnaestog stoljeća razvio Gibbs, koji je definirao promjenu slobodne energije potrebnu za formiranje nukleusa ( $\Delta G$ ) kao zbroj promjene slobodne energije za faznu transformaciju ( $\Delta G_v$ ) i promjene slobodne energije za formiranje površine ( $\Delta G_s$ ). Za sferne klasterne polumjera  $r$ , prikazan je termodinamički izraz pod (3):

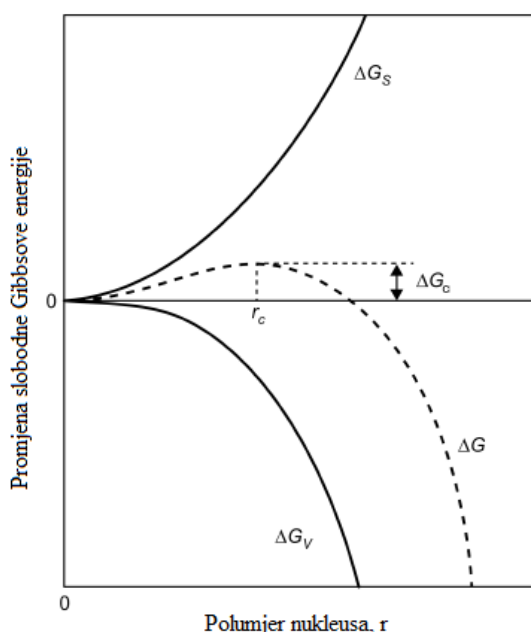
$$\Delta G = \Delta G_s + \Delta G_v = 4\pi r^2 \sigma + \frac{4}{3}\pi r^3 \Delta G_v \quad (5)$$

gdje je:  $4\pi r^2$  površina čvrste faze na granici kruto-tekuće,  $\sigma$  je površinska napetost (slobodna energija površine) i  $\frac{4}{3}\pi r^3$  je volumen sfere.



Kod kristalizacije iz otopina, drugi član ( $\Delta G_v$ ) opisuje spontanu tendenciju prezasićene otopine ka stvaranju kristala. Nasuprot tome, uvođenje međupovršine čvrsto-tekuće povećava slobodnu energiju za iznos proporcionalan površini klastera. Kao rezultat toga, rast klastera ovisi o kompetenciji između smanjenja  $\Delta G_v$ , koja pogoduje rastu, i povećanja  $\Delta G_s$ , koja pogoduje otapanju [8].

Prema slici 6 vidimo da je  $\Delta G_s > 0$  i da je prema izrazu (5) proporcionalna  $r^2$ , dok je  $\Delta G_v < 0$  i da je prema izrazu (5) proporcionalna  $r^3$ . Ovaj se koncept može jasnije vidjeti na slici 6, gdje su dva suprotna člana i ukupna promjena slobodne energije potrebne za formiranje nuklusa ( $\Delta G$ ) ucrtani u odnosu na polumjer nukleusa  $r$  [8].

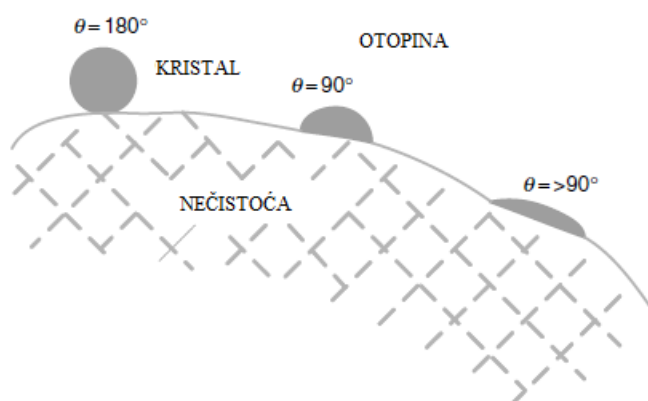


**Slika 6.** Grafički prikaz promjene slobodne Gibbsove energije i polumjera nukleusa [8]

Na slici 6 uočava se maksimum promjene slobodne Gibbsove energije,  $\Delta G_c$ , pri kritičnom polumjeru nukleusa  $r_c$ . Upravo u toj točki nastaje nova faze, odnosno dolazi do kristalizacije. Nestabilne čestice koje odgovaraju uvjetu da je  $r < r_c$  nazivaju se klasteri, dok one koje odgovaraju uvjetu  $r > r_c$  nazivaju se nukleusi. Ako klaster ima veličinu  $r = r_c$ , tada klaster postaje kritični nukleus i u ravnoteži je sa prezasićenom otopinom. Pri vrijednostima polumjera nukleusa  $r$  manjima od polumjera kritičnog nukleusa  $r_c$  dolazi do porasta Gibbsove energije nukleacije  $\Delta G$ . Razlog tome je stabilizacija zbog otapanja nestabilnih klastera jer u kompeticiji dvaju članova u izrazu (5) prevladava pozitivna vrijednost  $\Delta G_s$ . Pri vrijednostima polumjera nukleusa  $r$  većima od polumjera kritičnog nukleusa  $r_c$ , dolazi do smanjenja Gibbsove slobodne

energije nukleacije  $\Delta G$  ili stabilizacije zbog spontanog rasta nukleusa jer u konkurenciji dvaju članova u izrazu (5) prevladava negativna vrijednost  $\Delta G_v$  [8].

Stvaranje čvrste faze u prezasićenoj otopini u kojoj do kristalizacije dolazi zbog prisutnosti nečistoća definira se kao heterogenom nukleacijom. Sitne čestice djeluju kao centri nukleacije oko kojih kristal nastavlja rasti. Strana tvar u prezasićenoj otopini smanjuje energiju potrebnu za nukleaciju pa do nukleacije dolazi pri nižoj prezasićenosti nego kod homogene nukleacije [5]. Adsorpcijom iona ili molekula na površinu nečistoća može se prikazati mehanizam heterogene nukleacije. Bitnu ulogu ima površinska napetost između faza, te kontaktni kut  $\theta$  između faza [5,8].



**Slika 7.** Prikaz heterogene nukleacije [8]

Do sekundarne nukleacije dolazi u prisustvu kristala u prezasićenoj otopini. Do nukleacije dolazi pri nižim prezasićenostima od onih potrebnih za primarnu nukleaciju. Prema uzroku nastanka sekundarne nukleacije, ona se dijeli na: cijepljenu nukleaciju, nukleaciju zbog loma kristala, nukleaciju zbog strujanja i kontaktnu nukleaciju. Jedina željena i kontrolirana sekundarna nukleacija je cijepljena nukleacija (kontrola veličine i raspodjele veličina čestica). Dodani sitni kristali predstavljaju mjesta nukleacije uz uvjet da je njihova veličina veća od kritične [5].

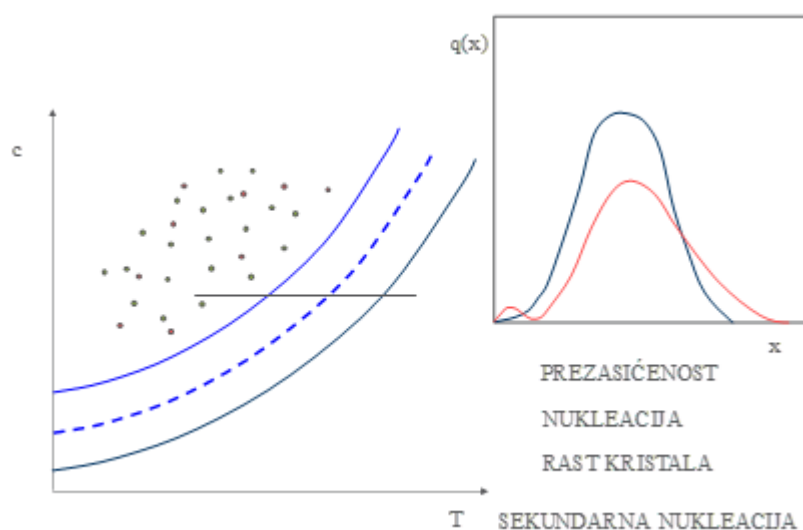
Pri velikim brzinama miješanja može doći do loma kristala, pri čemu nastale krhotine djeluju kao centri nukleacije te dolazi do nukleacije zbog loma kristala. Osim loma kristala zbog kontakta sa mješalom, u područjima velikih smičnih naprezanja u blizini lopatice miješala može doći do stvaranja nukleusa tada dolazi do nukleacije zbog strujanja [5].

Kontaktne nukleacije ne može izbjeći u kristalizatoru. Zbog miješanja kristali se sudaraju međusobno, sa stjenkom posude (kristalizatora) te s lopaticama miješala, tako da postoje tri vrste kontaktne nukleacije:

- kristal-kristal
- kristal-stjenka
- kristal miješalo [5]

Ovisno o veličini kristalizatora drugi tip kontaktne nukleacije će biti dominantan. Primjerice, u industrijskom kristalizatoru put koji prijeđe kristal između dva sudara s lopaticom daleko je veći nego u malom laboratorijskom kristalizatoru, što znači da će kontakt kristal-miješalo biti češći u malom kristalizatoru. Najintenzivniji su sudari između lopatice i kristala [5].

Kontroliranim hlađenjem otopine (u laboratoriju) i pri odgovarajućim hidrodinamičkim uvjetima moguće je provesti kristalizaciju tako da nakon primarne homogene nukleacije kristali samo nastavljaju rasti te da ne dolazi do sekundarne nukleacije. U tom će se slučaju dobiti unimodalna funkcija raspodjele kristala (s jednim maksimumom) kao što je prikazano na slici 8. Međutim ako iz nekog razloga dođe do sekundarne nukleacije, neće se sva raspoloživa pokretačka sila trošiti na rast primarno nastalih nukleusa već i na novonastale pa će nastajati višemodalne funkcije raspodjele veličina kristala (crvena krivulja). U industrijskim kristalizatorima česta je sekundarna nukleacija [5].



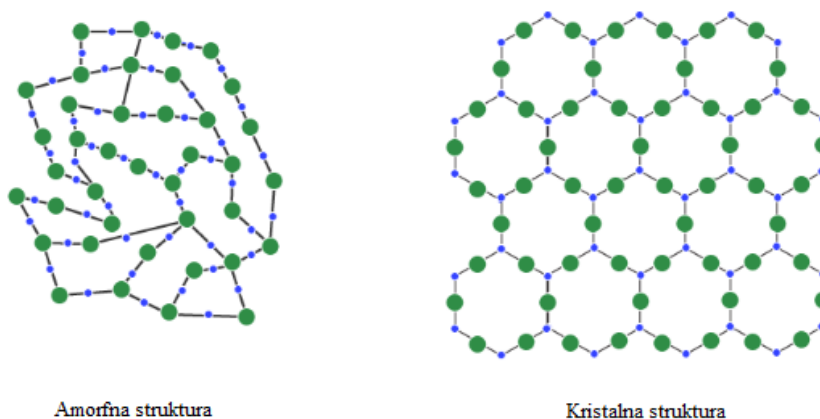
**Slika 8.** Nukleacija i raspodjela veličina čestica [5]

Posljednja faza procesa kristalizacije iz otopine je rast kristala, i slijedi nakon uspješnog formiranja nukleusa. Linearna brzina rasta kristala pokazuje promjenu karakteristične dimenzije kristala u vremenu i glavni je pokazatelj rasta kristala. Vanjski čimbenici kao što su: temperatura, nečistoće, prekomjerna zasićenost i vrsta otapala, te unutarnji čimbenici definirani kristalnom strukturom, kontroliraju rast kristala i diktiraju morfologiju. Oblik kristala i njihova veličina, kao i raspodjela veličina kristala utječu na mogućnost separacije nakon procesa kristalizacije, i govore nam o kvaliteti kristala [1,5].

## 2.2. Kristali

Prema sadašnjem stanju znanosti, sve tvari koje nas okružuju nalaze se u pet agregatnih stanja: kruto, kapljevito, tekuće, plazma i Bose-Einsteinov kondenzat. Pokretljivost i uređenost molekula i atoma je različita ovisno o pojedinom agregatnom stanju. Kristalima se svojstva mijenjaju ovisno o smjeru u kojem su mjerena, dok su kapljevine i plinovi izotropni (jednaka svojstva u svim smjerovima).

Kristali su krute tvari u kojima su strukturne jedinice raspoređene na periodički ponavljajući način u tri dimenzije. Treba imati u vidu da nisu sve krutine kristali već da postoje i amorfne odnosno nekristalinične krutine. Kristalna je struktura uređena, dok je amorfna neuređena. Za razliku od kristala, amorfne krutine također su izotropne.



**Slika 9.** Prikaz amorfne i kristalne strukture [9]

### 2.2.1. Morfologija kristala

Oblik kristala, također poznat kao morfologija ili habitus, određen je relativnom brzinom rasta njegovih različitih površina. Kristalni habitus odnosi se na vanjski izgled kristala. Kvantitativni opis kristala znači poznavanje prisutnih kristalnih površina, njihovih relativnih površina, duljina osi u tri smjera, kutova između strana i faktora oblika kristala. Faktori oblika su prikladan matematički način za opisivanje geometrije kristala. Ako je veličina kristala definirana u smislu karakteristične dimenzije  $L$ , mogu se definirati dva faktora oblika: faktor oblika volumena  $\alpha$  i faktor oblika površine  $\beta$  [4].

Prvi tip kristala naziva se igličastim jer ima samo jednu karakterističnu dimenziju. Druga vrsta kristala naziva se pločasti jer ima dvije karakteristične dimenzije. Treći tip kristala naziva se štapičast ili kockasti jer ima tri karakteristične dimenzije. Iz tih oblika kristala moguće je zaključiti o površini i relativnoj brzini rasta različitih kristalnih površina. Za igličaste kristale, površina za rast kristala prvenstveno je na dva kraja. Površina duž duljine igle ima mnogo sporiji rast. Za pločaste kristale, površina za rast kristala je ona na rubovima. Površina ploče ima mnogo sporiji rast. Za štapičaste ili kockaste kristale, sve kristalne površine rastu po usporedivim stopama [2].

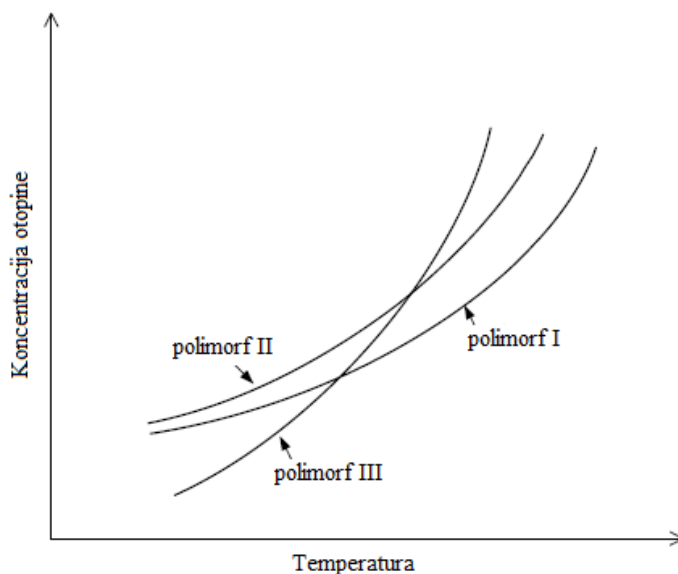
Važno je prepoznati da različiti kristalni oblici materijala ne moraju nužno imati različita fizikalna svojstva kao što su talište, topljivost i brzine otapanja. Morfologija kristala može se odrediti mikroskopom, elektronskim mikroskopom ili analizatorom slike. Kvantitativna karakterizacija krutih pomoćnih tvari temeljena na analizi pojedinačnih čestica i kvantificiranju oblika čestica objavljena je i pruža dodatne informacije [10].

### 2.2.2. Polimorfija

Sposobnost tvari, jednog spoja, da kristalizira u više kristalnih oblika naziva se polimorfija. Polimorfi imaju istu empirijsku formulu, a razlikuju se po kristalnim strukturama. Kristalni polimorfi pokazuju različita fizikalna svojstva, kao što su topljivost i talište, različita kemijska i mehanička svojstva idr. S termodinamičkog gledišta postoje dvije vrste polimorfije: enantiotropija ili monotropija. Kod monotropnih oblika jedan je polimorf stabilna čvrsta modifikacija u cijelom temperaturnom rasponu do tališta [1].

Slika 10 pokazuje krivulje topljivosti monotropnih i enantiotropnih polimorfa. Za monotropne polimorfe, polimorf I ima manju topljivost od polimorfa II u cijelom rasponu topljivosti. Za enantiotropne polimorfe, polimorf I ima veću topljivost od polimorfa III u nižem

dijelu raspona topljivosti, ali nižu topljivost od polimorfa III u višem dijelu raspona topljivosti [1].



**Slika 10.** Krivulje topljivosti monotropnih i enantiotropnih polimorfa [1]

Važnost otkrivanja najstabilnijih polimorfa, ako ne i svih polimorfa lijeka u ranoj fazi razvoja lijeka, daleko nadilazi kristalizaciju. Ako se stabilniji oblik otkrije u kasnoj fazi razvoja lijeka, svi prethodni razvojni rezultati, uključujući kristalizaciju, potrebno je ponovno procijeniti [1].

U prisutnosti polimorfa, razvoj procesa kristalizacije za dobivanje željenog kristalnog oblika pravi je izazov. Dobro razumijevanje ravnoteže kruto-tekuće pod različitim uvjetima - na primjer, temperatura ili smjesa otapala, je neophodno. Cijepljenje i vođenje prezasićenosti dva su kritična, ako ne i neophodna zahtjeva. Nije dovoljno iskristalizirati željeni polimorf, već je potrebno održati stabilnost željenog polimorfa tijekom slijeda operacija nakon procesa kristalizacije, uključujući filtraciju, sušenje i mljevenje [1].

Za enantiotropni sustav može doći do prelaska iz jednog polimorfnog oblika koji je stabilan na sobnoj temperaturi u drugi polimorfnu oblik koji je stabilan na višoj temperaturi ako je temperatura sušenja iznad temperature prijelaza [1].

### 2.3. Procesne analitičke tehnologije

Procesna analitička tehnologija (engl. *process analytical technology*, PAT) opisuje sustav koji projektira, prati i vodi proizvodne procese uzastopnim mjerenjem glavnih indikatora kvalitete i rada procesa [10]. Procesna analitička tehnologija definirana je kao sustav za projektiranje, analizu i kontrolu proizvodnih procesa putem pravovremenih mjerenja ključnih značajki kvalitete (engl. *critical quality attributes*, CQA). Mjerenja se mogu provoditi na sirovinama, međuproduktima i proizvodima, ali često su to ključni parametri procesa koji utječu na učinkovitost procesa i kvalitetu konačnog proizvoda procesa [11].

Pojam (i akronim) „procesna analitička tehnologija“ (PAT) uvela je američka agencija (engl. *The Food and Drug Administration*, FDA) kao inicijativu za bolje razumijevanje procesa farmaceutske proizvodnje, kako bi se povećala kvaliteta njihovih proizvoda. PAT tehnologija koristi se kako bi se u idealnom slučaju visoka kvaliteta proizvoda postigla već u fazi projektiranja proizvodnog procesa. Također je naglašena potreba za poboljšanim *on-line* nadzorom i metodama kontrole kako bi se održala visoka kvaliteta proizvoda tijekom proizvodnih operacija. U biofarmaceutskoj industriji PAT principi se usvajaju s velikom pažnjom zbog činjenice da su biofarmaceutici i njihovi proizvodni sustavi vrlo složeni [13].

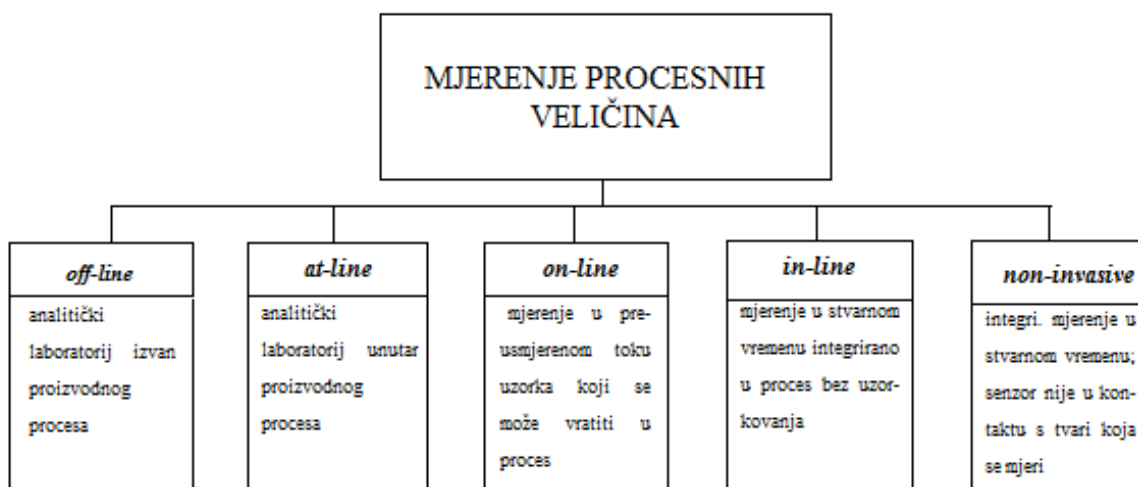
Dostupni su brojni PAT alati i tehnike koje se mogu koristiti za praćenje i vođenje procesa farmaceutske kristalizacije, za karakterizaciju kristalnih materijala i za bolje razumijevanje proizvoda i procesa za krajnju svrhu osiguravanja kvalitete i sigurnosti proizvoda [11,12].

Farmaceutska industrija usvojila je procesnu analitičku tehnologiju prije više od dva desetljeća, a glavnu primjenu nalazi u istraživanju i razvoju. Glavna svrha PAT-a je u boljem razumijevanju samog procesa u njegovoj ranijoj laboratorijskoj fazi, ali i kasnije tijekom razvoja i uvećanja [10]. Zadnjih desetak godina PAT se primjenjuje također i u široj akademskoj zajednici, gdje služi za kvalitetnije razumijevanje mehanizma procesa u prvoj fazi razvoja. PAT se zahvaljujući novim tehnologijama i novim primjenama postojećih tehnologija proširio na svaki segment proizvodnje lijekova u farmaceutskoj industriji, što je dovelo do uspostavljanja novih znanja neophodnih za unapređenje farmaceutika [12].

Primjena PAT-a doprinosi pouzdanosti uslijed uvećanja procesa, a za monitoring procesa u stvarnom vremenu u većini slučajeva nije potrebno pripremati uzorke, dok je obrada samih podataka puno brža u odnosu na standardne analitičke tehnike [12].

Provedba PAT-a korisna je prilikom poteškoća pri uzorkovanju koje nastaju zbog sigurnosti ili problema pri uzorkovanju, kada uzorak nije reprezentativan ili kad je nužno često

uzorkovanje. Za praćenje i vođenje procesa, neophodni su senzori koji pružaju informacije o ključnim procesnim varijablama, ali i kvaliteti produkta u stvarnom vremenu [11,12].



**Slika 11.** Načini mjerenja procesnih veličina [12]

Problemi koji se javljaju kod mjerenja *off-line* i *at-line* su vremenska zadržka od trenutka uzorkovanja do trenutka dobivanja mjerne informacije, statistička pogreška pri uzorkovanju, nereprezentativnost uzorka i pogreška pri uzorkovanju koja izaziva promjenu stanja tvari [12].

Napredak u području procesne analitičke tehnologije otvorio je mogućnost različitih načina mjerenja. Senzore koji se upotrebljavaju za mjerenje ključnih varijabli kao što su prezasićenost, veličina čestica, oblik i polimorfija produkata razvijeni su radi postizanja veće djelotvornosti procesa kristalizacije [12].

Sustav za mjerenje i vođenje uključuje: analitičke instrumente, alate za kemometrijsko modeliranje, softver za razvoj PAT metode, obradu podataka i integraciju u sustav te softver i algoritme za vođenje procesa. Za kvalitativnu ili kvantitativnu analizu zbog specifičnih interakcija svjetlosti s pojedinim funkcionalnim skupinama u spojevima sirovina, međuprodukata i produkata često se primjenjuju spektroskopske metode. Sonda uronjena u procesni medij omogućava mjerenje u stvarnom vremenu smanjujući mogućnost kontaminacije uzoraka. Spektralni podatak predstavlja izlazni signal spektroskopskih uređaja. Iz takvog podatka korisne informacije moguće je dobiti kemometrijskim/kalibracijskim modelom kojim se računaju trenutačne vrijednosti značajke. Mjerenje trenutačnih vrijednosti značajke omogućuju vođenje procesa u stvarnom vremenu [12].



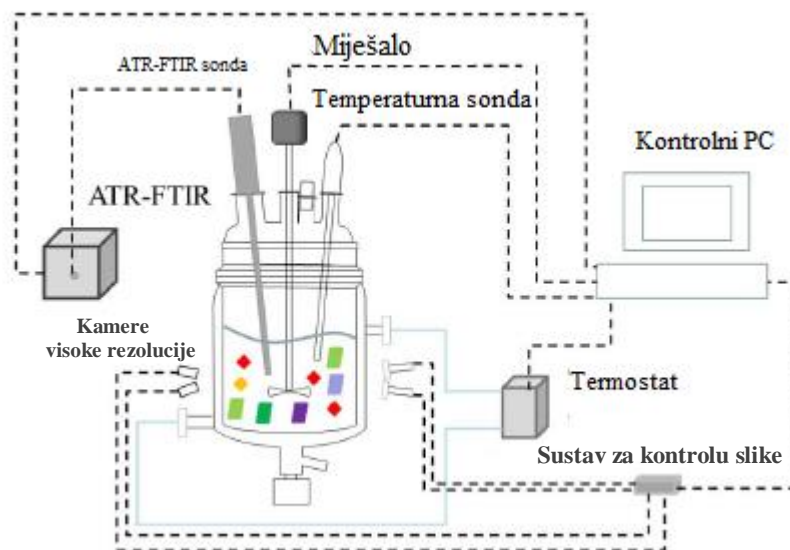
### 2.3.1. ATR-FTIR spektroskopija

Razvojem i istraživanjima u području prigušene totalne refleksije (ATR-a), slijedio je i razvoj instrumenata za mjerenje i monitoring procesa kristalizacije. Među njima najveći je značaj pokazala ATR-FTIR tehnologija koja se koristi za praćenje i vođenje procesa kristalizacije [14]. Prigušena totalna refleksija Fourier transformacijske infracrvene spektroskopije (eng. *attenuated total reflectance fourier transform infrared spectroscopy*, ATR-FTIR) je spektroskopska metoda kojom se dobiva IC spektar apsorpcije ili emisije tekućine, krutine ili plina. Kada IC zraka putuje od medija s visokim indeksom loma (npr. kristal cinkovog selenida) do medija s niskim indeksom loma (uzorak), određena se količina zračenja reflektira natrag u medij s niskim indeksom loma. Pod određenim kutom upada gotovo svi svjetlosni valovi se reflektiraju natrag. Taj se fenomen naziva totalna unutarnja refleksija [15].

Izraz infracrvena spektroskopija Fourierove transformacije (FTIR) potječe iz činjenice da je Fourierova transformacija (matematička obrada) potrebna za pretvaranje sirovih podataka u stvarni spektar [16].

Budući da je poznato da je prezasićenost pokretačka sila procesa kristalizacije, potrebno je primijeniti neku od metoda mjerenja koncentracije. Jedna od najrasprostranjenijih tehnologija za ovu namjenu je ATR-FTIR spektroskopija (Dunuwila et al., 1994.). Prednost ove mjerne metode je u tome što se omogućuje mjerenje koncentracije otopina unatoč prisutnosti krutih čestica [17].

Za korištenje ATR-FTIR spektrometrije za mjerenje koncentracije otopine ključni korak je razvoj kalibracijskog modela [18].



**Slika 12.** Shema praćenja procesa kristalizacije primjenom procesne analitičke tehnologije

[14]

## 2.4. Kemometrija

Kemometrija predstavlja interdisciplinarnu granu znanosti u kojoj se rabe metode regresijske statistike, primijenjene matematike i računalne znanosti u kemiji, biokemiji, medicini, biologiji i kemijskom inženjerstvu. Bavi se predviđanjem i optimiranjem pokusa i dobivanjem kemijske informacije analizom podataka. Čini ju obrada podataka dobivenih različitim analitičkim tehnikama radi dobivanja korisnih informacija [19].

Godine 1971. švedski znanstvenik Svante Wold nazvao je pojam "kemometri", što je na švedskom, a na engleskom ekvivalentno "chemometrics". Nedugo nakon toga već spomenuti Svante Wold i Bruce Kowalski profesor analitičke kemije na Sveučilištu Washington u Seattleu osnivaju Međunarodno društvo za kemometriju. Znanost o kemometriji opisana je kao interakcija određenih matematičkih i statističkih metoda u kemijskim mjernim procesima [20].

Od 1990-ih dolazi do naglog razvoja instrumentalnih tehnika, ponajviše zbog obilatih financijskih sredstava farmaceutske industrije koja je počela izdvajati za vlastita istraživanja. Instrumentalne tehnike postaju sve brže i produciraju sve više eksperimentalnih podataka. To je dalo novi preokret razvoju kemometrije usmjeravajući njezinu primjenu i na izvlačenja relevantnih informacija iz baze podataka [21].

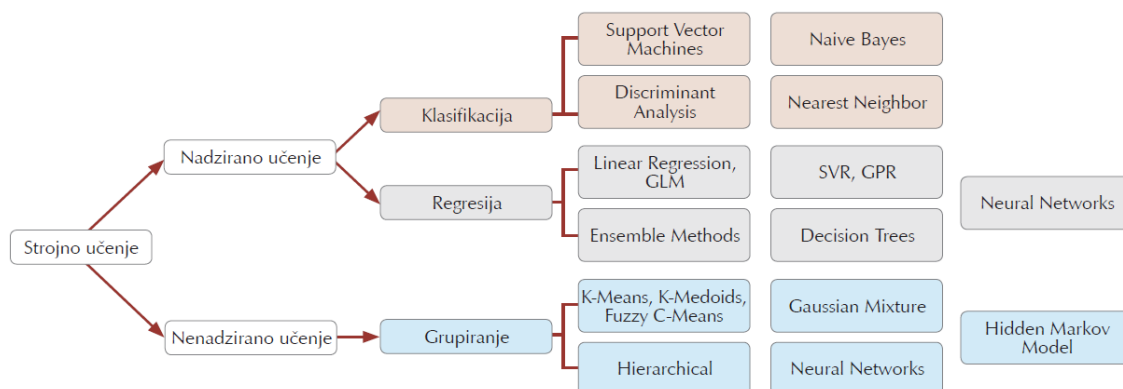
Kemometrijske tehnike imaju zadaću ukloniti šum iz podataka, kako bi preostale samo korisne informacije. Korisne informacije iz podataka primjenjuju se za predikciju i interpretaciju nekog svojstva uzoraka ili sustava. To je moguće ukoliko su podaci reprezentativni dovoljno da sadrže informaciju, te postoji zavisnost između informacija unutar podataka.

## 2.5. Strojno učenje

Umjetna inteligencija ( engl. *artificial intelligence*, AI) grana je računalne znanosti i njezin je cilj izrada pametnih strojeva i uređaja za obavljanje zadataka, rješavanje problema za koje je inače potrebna ljudska inteligencija [22].

Strojno učenje (engl. *machine learning*) je grana umjetne inteligencije koja "uči" računala onome što je ljudima prirodno. Algoritmi strojnog učenja "uče" informacije iz dostupnih podataka, ne koristeći teoretsku podlogu niti matematičke modele, već "uče iz iskustva". Porastom količine podataka u svijetu u različitim granama industrije i znanosti strojno učenje pronalazi važnu ulogu u više područja, a neka od njih su: računalne financije, medicina i biologija, obrada slika, industrijska proizvodnja itd. [23].

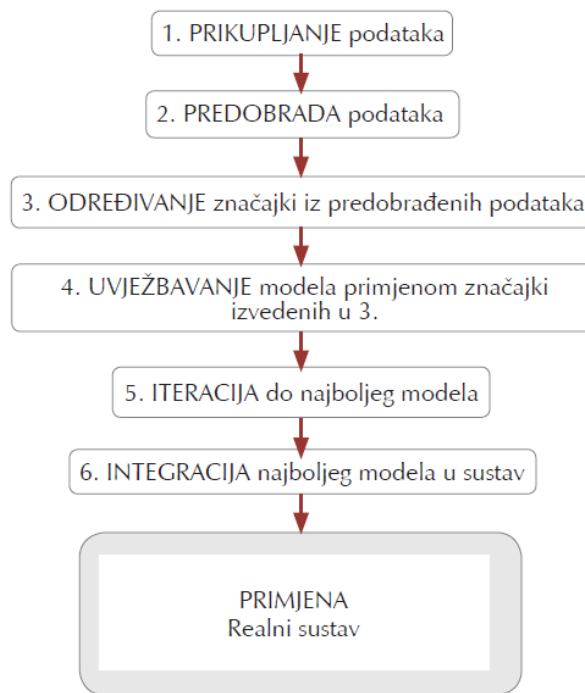
Slikom 13 prikazana je podjela strojnog učenja na dvije metode: nadzirano učenje (engl. *supervised learning*) i nenadzirano učenje (engl. *unsupervised learning*). Modeli nadziranog učenja primjenjuju se u postupcima klasifikacije i regresije. Ti modeli izvode se na temelju ulaznih i izlaznih podataka dobivenih eksperimentom ili koristeći već od prije poznate podatke, s ciljem izrade modela koji može predvidjeti buduće izlaze na temelju poznatih ili nekih budućih ulaznih podataka. Postupci klasifikacije koriste ulazne podatke razvrstane u različite kategorije, te se predviđaju diskretni odzivi. Razlikujemo više metoda regresije, najpoznatije su linearna regresija i ansambl metode. Regresija je postupak kojim se predviđa kontinuirani odziv poput koncentracije ili temperature u reaktoru, kvaliteta proizvoda, sastav otopine ili neke druge smjese itd. Modeli nenadzornog učenja temelje se na pronalasku skrivenih uzoraka i inherentne strukture u ulaznim podacima bez informacija o izlaznim podacima [23.]



**Slika 13.** Podjela metoda strojnog učenja [24]

## 2.6. Razvoj modela strojnog učenja

Postupci koji prethode razvoju modela strojnog učenja su: prikupljanje i podjela podataka, predobrada podataka, odabir modela i iteracija do najboljeg rezultata, integracija najboljeg modela u sustav te njegova primjena u realnom sustavu. Slika 14 prikazuje korake pri razvoju modela strojnog učenja.



**Slika 14.** Razvoj modela strojnog učenja [25]

### 2.6.1. Prikupljanje, predobrada i podjela podataka

Prvi korak u razvoju modela strojnog učenja je prikupljanje eksperimentalnih podataka koji mogu biti u različitim formatima i oblicima. Također je važno tim podacima, klasifikacija podataka ovisno što je za pojedini model potrebno. Ključno je poznavati teoretsku i matematičku pozadinu svakog modela kako bi znali što su varijable i ulazne veličine, a što su izlazi i što želimo previdjeti. Sirovi podaci nisu prikladni za direktnu primjenu u razvoju modela, već je takve podatke potrebno predobraditi.

Sirovi podaci zahtijevaju različite tehnike predobrade. Predobrada podataka odnosi se na metode za pronalaženje i uklanjanje ili zamjenu ekstremnim loših vrijednosti te generiranje podataka koji nedostaju. Filtriranje je postupak kojim se nastoji smanjiti mjerni šum prisutan kod mjernih pretvornika. Svrha filtriranja je "izgladiti" podatke te olakšati razvoj modela, tj. onemogućiti modeliranje šuma što rezultira nepotrebnom parametrizacijom modela. Filtriranje treba provoditi s mjerom jer u obrađenim podacima treba ostaviti dovoljno informacija. Neke od metoda filtriranja su: *moving average filter*, *Savitzky-Golay filter*, *local regression smoothing*. Metode grupiranja identificiraju karakteristike podataka po grupama [26].

Skaliranje i normiranje podataka provodi se kako bi se uklonio utjecaj različitih mjernih područja instruenta tijekom razvoja modela. Najčešće metode normalizacije su *Min-Max normalizacija* i *Z-score normalizacija* uz još brojne druge [26].

Nakon predobrade podataka potrebno ih je razdijeliti na skup za učenje (treniranje modela) i na skup za validaciju (provjeru modela). Skup za učenje mora obuhvatiti najveći raspon podataka kako bi se model što bolje istrenirao, i mogao predvidjeti vladanje unutar tog raspona vrijednosti. Skup za validaciju mora sadržavati modelu nepoznati skup vrijednosti kako bi se njime model validirao, odnosno provjerila sposobnost predikcije modela na nepoznatim podacima. Skup za učenje je veći od skup za validaciju, sadrži više podataka kako bi se omogućile sve dostupne informacije za učenje. Za konačnu provjeru modela uobičajeno je koristiti i skup za testiranje modela [27].

## 2.7. Analiza glavnih komponenta

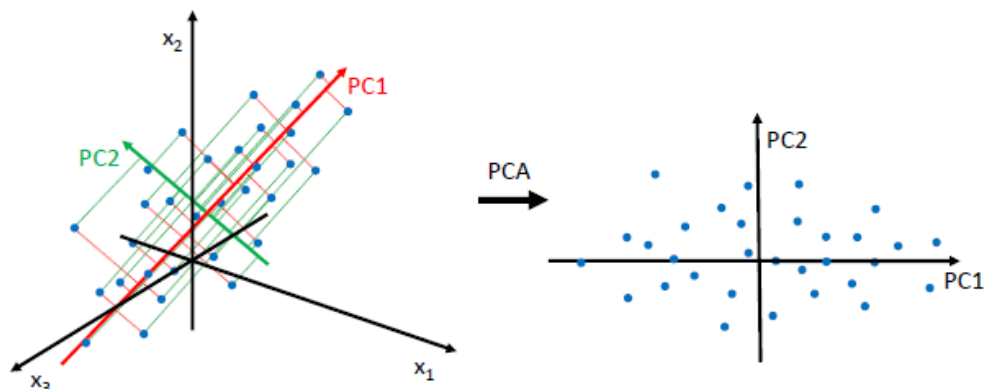
Analiza glavnih komponenti (engl. *principal component analysis*, PCA) je najstarija i najpoznatija tehnika multivarijatne (viševeličinske) analize podataka. Prvi ju je predstavio Pearson (1901.), a samostalno razvio Hotelling (1933.). Poput mnogih multivarijatnih analiza, nije bila široko korištena sve do pojave elektroničkih računala, ali sada je dobro ukorijenjena u gotovo svim statističkim računalnim paketima. Središnja ideja analize glavnih komponenti je smanjiti dimenzionalnost skupa podataka u kojem postoji veliki broj međusobno povezanih varijabli, zadržavajući što je više moguće varijacija prisutnih u skupu podataka. Smanjenje se postiže pretvorbom u novi skup varijabli, glavnih komponenti (engl. *principal components*, PCs), koje nisu u korelaciji i koje su poredane tako da prvih nekoliko zadržava većinu varijacija prisutnih u svim izvornim varijablama [28].

Ciljevi PCA su:

- (1) izdvojiti najvažnije informacije iz podataka;
- (2) komprimirati veličinu skupa podataka zadržavajući samo ove važne informacije;
- (3) pojednostaviti opis skupa podataka;
- (4) analizirati strukturu opažanja i varijabli [28].

Kako bi postigao te ciljeve, PCA izračunava nove varijable koje se nazivaju glavne komponente koje se dobivaju kao linearne kombinacije izvornih varijabli. Prva glavna komponenta ima najveću moguću varijancu. Druga glavna komponenta, PC2, ortogonalna je na PC1. Za dvodimenzionalne podatke, moguć je samo jedan smjer za PC2, onaj ortogonalan

na PC1. Broj glavnih komponenata koji se može izračunati ograničen je brojem varijabli u matrici  $\mathbf{X}$ . Svaka sljedeća glavna komponenta ortogonalna je na sve prethodne glavne komponente, i njezin smjer mora osigurati maksimum varijance ( slika 15) [29].



**Slika 15.** Princip analize glavnih komponenti. Skup od 3 varijable ( $x_1$ ,  $x_2$ ,  $x_3$ ) reducira se na skup od 2 glavne komponente (PC1 i PC2) koje objašnjavaju maksimalni dio varijance u podacima [30]

Matrice su označene velikim, uspravnim i podebljanim slovima (npr.  $\mathbf{X}$ ), vektori (stupci) malim uspravnim i (npr.  $x_i$ ), vektori (retci) malim, nakošenim i podebljanim (npr.  $\mathbf{x}_j$ ), a elementi i skalari malim nakošenim (npr.  $x_{j,i}$ ). Vektori i elementi iz iste matrice svi koriste isto slovo (npr.  $\mathbf{X}$ ,  $x_i$ ,  $\mathbf{x}_j$ ,  $x_{j,i}$ ). Operacija transponiranja označava se s T.

U PCA problem kolinearnosti između prediktorskih varijabli rješava se tako što se podatci reduciraju na manji skup od samo nekoliko latentnih varijabli koje najbolje opisuju izvorne, tj. objašnjavaju maksimalnu varijancu u podacima. To se postiže dekompozicijom  $\mathbf{X}$  ( $n \times p$ ) na matricu faktorskih bodova (engl. *scores*)  $\mathbf{T}$  ( $n \times l$ ) i matricu opterećenja (engl. *loadings*)  $\mathbf{P}$  ( $p \times l$ ), uz minimizaciju X-reziduala sadržanih u  $\mathbf{E}$  ( $n \times p$ ):

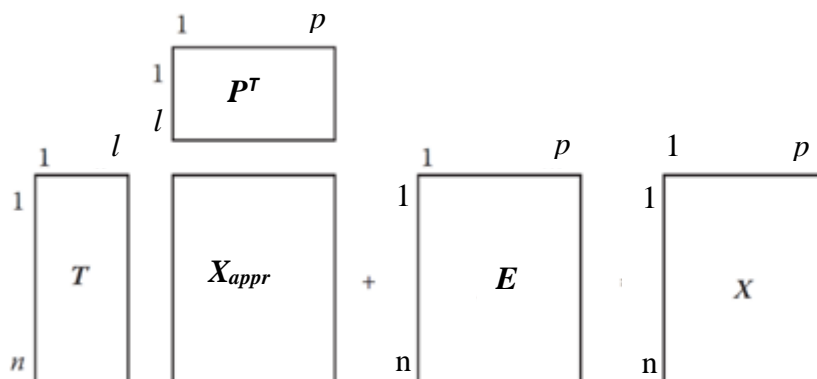
$$\mathbf{X} = \mathbf{TP}^T + \mathbf{E} = \sum_{i=1}^l t_i \mathbf{p}_i^T + \mathbf{E} \quad (6)$$

gdje su:

$\mathbf{T}$  matrica faktorskih bodova, broj redova jednak je broju u originalnoj matrici  $\mathbf{X}$

$\mathbf{P}^T$  matrica opterećenja, broj stupaca jednak je broju stupaca u originalnoj matrici  $\mathbf{X}$

$E$  matrica pogrešaka, jednakih je dimenzija kao i originalna matrica  $X$   
 $p_j^T$  definiran je kao težinski vektor (vektor opterećenja)  
 $t_i$  definiran je kao vektor faktorskih bodova (slika 16)



**Slika 16.** Približna rekonstrukcija,  $X_{appr}$ ,  $X$  matrice iz PCA matrice faktorskih bodova  $T$  i matrice opterećenja  $P$  koristeći broj komponenata  $a$ ,  $E$  je matrica pogrešaka [6]

Vektori opterećenja  $p$  određuju osi novog koordinatnog sustava, dok vektori faktorskih bodova  $t$  određuju projekcije izvornih varijabli na nove koordinate. Kažemo da  $P$  predstavlja glavne komponente, a  $T$  (latentne varijable) projekciju  $X$  na glavne komponente. Pri tom,  $t$  i  $p$  poredani su od onih koji objašnjavaju najveći udio varijance u  $X$  prema onima koji objašnjavaju najmanji. Desna strana jedn. 6 prikazuje dekompoziciju  $X$  kao zbroj matrica vektorskih produkata  $t_i p_j^T$ , što znači da je  $X$  linearna kombinacija glavnih komponenti [6, 28].

Zbog ortonormiranosti  $P$  i pretpostavke da su  $E$  mali, jedn. 6 može se formulirati u:

$$T = X P \tag{7}$$

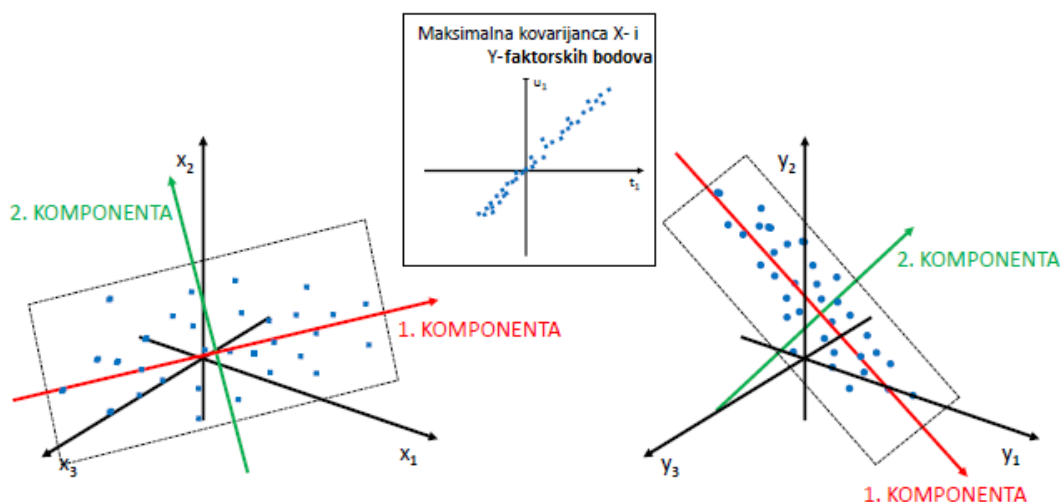
Ova inverzija omogućava izračunavanje  $T$  za bilo koje nove skupove  $X$ -varijabli, uz poznati  $P$ . Spomenuti  $P$  se izračunava iz kalibracijskih  $X$ -podataka koristeći jedn. 6, a potom je moguće izračunati  $T$  za novi skup  $X$ -podataka pomoću jedn. 7.

## 2.8. Parcijalna regresija metodom najmanjih kvadrata

Parcijalnu regresiju metodom najmanjih kvadrata (engl. *partial least squares regression*, PLSR) razvio je Herman Wold 1975. godine u području ekonometrije, a potom je uvedena kao metoda za analizu u kemometriji. PLSR je novija metoda koja generalizira i kombinira značajke

analize glavnih komponenti (PCA) i višestruku linearnu regresiju (MLR). Razlika između PCA i PLS-a je u tome što je PLS koristi značajke MLR-a. U principu koristi dva različita PCA modela, jedan za matricu podataka X, drugi za matricu odziva Y i pronalazi vezu između njih. Cilj PLSR metode je predvidjeti skup zavisnih varijabli iz skupa nezavisnih varijabli ili prediktora. Ovo se predviđanje postiže izdvajanjem iz prediktora skupa ortogonalnih faktora koji se nazivaju latentne varijable i imaju najbolju prediktivnu moć. Postoje dvije glavne varijante PLS regresije, PLS1 i PLS2 [31].

Ova tehnika je proširenje višestruke regresijske analize u kojoj se analiziraju učinci linearnih kombinacija nekoliko prediktora na varijablu odgovora (ili varijable višestrukog odgovora). PLS ima za cilj razviti linearni regresijski model povezivanjem X i Y preko latentnih varijabli, određenih na način da se maksimizira kovarijanca između X- i Y faktorskih bodova (Slika 17). Zbog toga omogućava analizu podataka s velikim brojem snažno koreliranih varijabli koje sadrže šum, uz modeliranje multivarijatnog odgovora Y. Pri tom, PLS za izračunavanje latentnih varijabli koristi informacije iz obje X i Y.



**Slika 17.** Princip regresije parcijalnih najmanjih kvadrata [30]

Povezanosti su uspostavljene s latentnim faktorima izdvojenim iz prediktorskih varijabli koje maksimiziraju objašnjenu varijancu u ovisnim varijablama. Ovi latentni čimbenici definirani su kao linearne kombinacije konstruirane između varijabli prediktora i odgovora, tako da je izvorna višedimenzionalnost smanjena na manji broj ortogonalnih čimbenika kako bi se otkrila struktura u odnosima između varijabli prediktora i između tih latentnih čimbenika i varijabli odgovora [31,32].



PLSR je posebno koristan kada:

- (1) je broj prediktorskih varijabli sličan ili veći od broja opažanja (tj. prekomjerno prilagođavanje). Ova situacija predstavlja ograničenje jer se regresijski koeficijenti ne mogu izračunati.
- (2) su prediktori visoko korelirani (tj. postoji jaka kolinearnost).

Ako su dane dvije centralne matrice varijable,  $\mathbf{X}$  i  $\mathbf{Y}$ , i to takve da im je očekivanje nula. Pretpostavke su da je  $\mathbf{X} \in \mathbb{R}^{n \times p}$  te da je  $\mathbf{Y} \in \mathbb{R}^{n \times q}$ , gdje  $p$  i  $q$  predstavljaju broj varijabli u svakom skupu, a  $n$  predstavlja veličinu (broj uzoraka, mjerenja). Kao što je već rečeno pomoću PLS metode  $\mathbf{X}$  i  $\mathbf{Y}$  je potrebno prikazati u novim koordinatama i najčešće se koristi za rješavanje problema linearne regresije [33].

Kako bi se uspjelo zadržati što više upotrebljivih informacija o skupovima  $\mathbf{X}$  i  $\mathbf{Y}$  korisne su tzv. latentne varijable kojima se reducira broj početnih varijabli svakog skupa na manji broj. S obzirom da su varijable često jako korelirane, ideja je kreirati mali broj novih varijabli koje imaju svojstvo da u nekom smislu najbolje predviđaju originalne varijable, ali za oba skupa (i  $\mathbf{X}$  i  $\mathbf{Y}$ ) istovremeno. Analiza parcijalnih najmanjih kvadrata pronalazi linearne kombinacije originalnih varijabli koje su najbolji linearni prediktori svih originalnih varijabli. Kreira se matrica čiji su stupci međusobno ortogonalni vektori, koji su dobiveni maksimiziranjem kovarijance između tih različitih skupova varijabli  $\mathbf{X}$  i  $\mathbf{Y}$ . Matrice  $\mathbf{X}$  i  $\mathbf{Y}$  se mogu prikazati u obliku:

$$\mathbf{X} = \mathbf{TP}^T + \mathbf{E} \quad (8)$$

$$\mathbf{Y} = \mathbf{UQ}^T + \mathbf{F} \quad (9)$$

gdje su  $\mathbf{T} \in \mathbb{R}^{n \times l}$ ,  $\mathbf{U} \in \mathbb{R}^{n \times k}$ ,  $\mathbf{P} \in \mathbb{R}^{p \times l}$ ,  $\mathbf{Q} \in \mathbb{R}^{q \times k}$ ,  $\mathbf{X} \in \mathbb{R}^{n \times p}$ ,  $\mathbf{Y} \in \mathbb{R}^{n \times q}$ ,  $\mathbf{E} \in \mathbb{R}^{n \times p}$ ,  $\mathbf{F} \in \mathbb{R}^{n \times q}$  [33].

Potrebno je odabrati matricu  $\mathbf{T} = [t_1, \dots, t_l]$  tako da su joj stupci međusobno ortogonalni i da mogu prikazati kao linearna kombinacija originalne matrice  $\mathbf{X}$  tako da vrijedi

$$\mathbf{T} = \mathbf{XW} \quad (10)$$

gdje je  $\mathbf{W} = [w_1, \dots, w_l]$ , matrica težinskih vektora.  $\mathbf{T}$  je istovremeno i dobar prediktor za  $\mathbf{Y}$ , te se zaključuje da između  $\mathbf{T}$  i  $\mathbf{U}$  iz jedn. 8 i 9 postoji linearna ovisnost prikazana:

$$\mathbf{U} = \mathbf{TB} + \mathbf{H} \quad (11)$$

gdje je  $\mathbf{B}$  dijagonalna matrica u kojoj je dijagonala od regresijskih težinskih vektora  $\mathbf{b}$ , a  $\mathbf{H}$  je matrica pogrešaka. Na sličan način kao što je prikazana matrica  $\mathbf{T}$ , možemo objasniti matricu  $\mathbf{U} = [\mathbf{u}_1, \dots, \mathbf{u}_k]$  tako da su joj stupci međusobno ortogonalni i da je linearna kombinacija matrice  $\mathbf{Y}$  tako da vrijedi

$$\mathbf{U} = \mathbf{Y}\mathbf{C} \quad (12)$$

gdje je  $\mathbf{C} = [c_1, \dots, c_k]$ , matrica težinskih vektora.

Potrebno je odrediti nizom jednostavnih operacija: (1) pretpostavka početnog vektora faktorskih bodova  $\mathbf{u}$  od  $\mathbf{Y}$ , (2) izračun težinskih vektora  $\mathbf{w}$  od  $\mathbf{X}$ , (3) izračun težinskih vektora  $\mathbf{c}$  od  $\mathbf{Y}$ , (4) izračun faktorskih bodova  $\mathbf{u}$  od  $\mathbf{Y}$  te regresijske težinske vektore  $\mathbf{b}$ . Provodi se iteracijski postupka do onog trena kada je:

$$\mathbf{X}_{1+i}(\mathbf{E}) = \mathbf{X}_1 - \mathbf{T}\mathbf{P}^T = 0 \quad (13)$$

Na taj način dobivene matrice  $\mathbf{T}$ ,  $\mathbf{C}$ ,  $\mathbf{B}$  služe za predikciju  $\mathbf{Y}$  prema izrazu:

$$\hat{\mathbf{Y}} = \mathbf{T}\mathbf{B}\mathbf{C}^T = \mathbf{X}\hat{\mathbf{B}} \quad (14)$$

## 2.9. Umjetne neuronske mreže

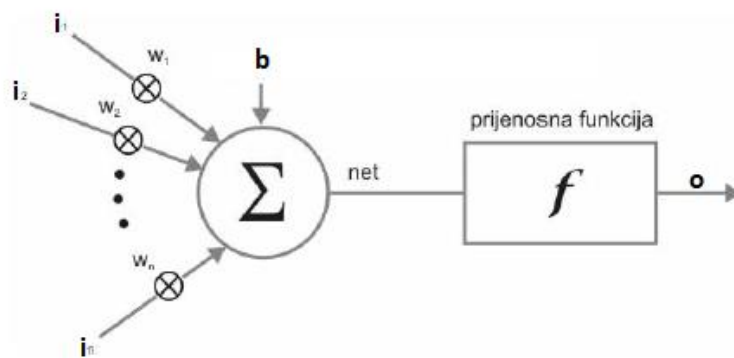
Ljudski mozak sastoji se od velikog broja neurona koji rade paralelno pri obradi informacija. Području umjetne inteligencije cilj je ostvariti imitaciju ljudskog mozga jer ljudski mozak pokazuje maksimalnu inteligenciju koju poznajemo, kao i kreativnost, sposobnost izbora ispravnih hipoteza i vođenje pomoću logičkih pravila. Učenje je jedan od bitnih aspekata ljudskog mozga, te se smatra da bi sustav koji bi bio sposoban učiti i oponašao rad ljudskog mozga bio upravo inteligentan. Područje računarstva u kojem brojni procesni elementi obrađuju podatke paralelno, kao što je to u ljudskom mozgu, naziva se *neuro-računarstvo*, a paradigma obrade podataka umjetnom neuronskom mrežom (engl. *artificial neural network*, ANN) [34].

Strukturu umjetne neuronske mreže čine: ulazni sloj, skriveni sloj i izlazni sloj. Ulazni sloj (engl. *input layer*) je prvi sloj, on prima informacije,  $i_n$  i proslijeđuje ih mreži. Skriveni sloj(evi) (engl. *hidden layer*) prima informacije iz ulaznog sloja i obrađuje ih, sumiranjem „otežanih“ ulaza i stvara vlastitu internu aktivaciju. Ako je zbroj otežanih signala veći od praga osjetljivosti neurona,  $T_j$  (engl. *threshold*), nelinearna aktivacijska funkcija  $f$  generira izlazni

signal neurona  $o$ . Izlazni signal neurona  $o$ , izlazi iz izlaznog sloja (engl. *output layer*) i daje rezultate.

Svaki neuron prima ulazne signale,  $i_n$  koji stižu kao izlazne vrijednosti iz drugih neurona ili dolaze izvan mreže. Svaki od tih ulaznih signala ima svoj težinski faktor  $w_n$ , on govori o jakosti veze između dva neurona. Što je on veći, jača je veza između dva neurona. Umnožak svakog pojedinog ulaza sa svojim težinskim faktorom daje „otežane“ ulaze. Takvi otežani ulazi nakon sumiranja daju rezultirajući signal, tzv. *aktivaciju* koja se predaje prijenosnoj funkciji  $f$ . Krajnji izlaz iz neurona može se proslijediti slijedećim neuronima ili postaje izlazni sloj mreže  $o$  (jedn. 15). Prijenos signala unutar neuronske mreže prikazan je na slici 18 [35].

$$o = f\left(\sum_{i=1}^n w_i i_i - T_j\right) = f(b) \quad (15)$$

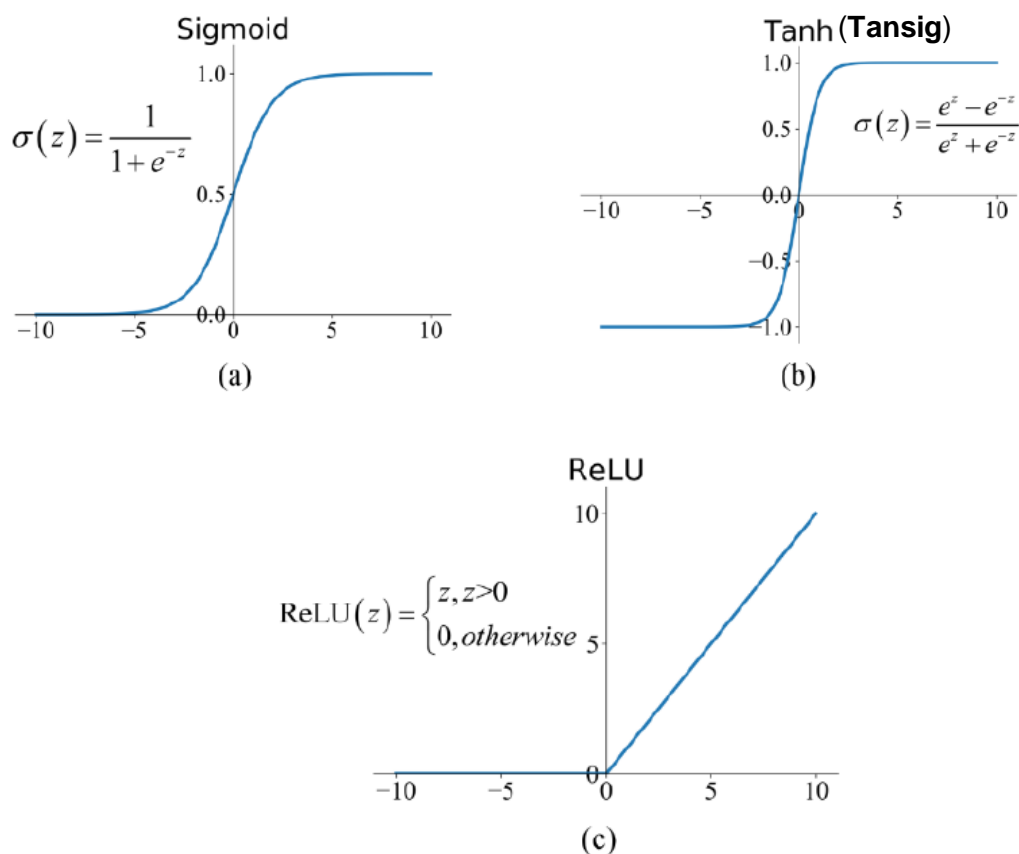


**Slika 18.** Struktura umjetne neuronske mreže [34]

### 2.9.1. Prijenosna funkcija

Prijenosna funkcija može biti linearna funkcija, iako je češće nelinearna. Najčešće korištene prijenosne funkcije su: sigmoidna, tangens-hiperbolna. One su kontinuirane funkcije i aproksimacija su skokomične funkcije kakva se nalazi i u biološkom neuronu, te ReLu funkcija (engl. *rectified linear units*). Sigmoidna funkcija uvodi nelinearnost u model pri čemu izlazne veličine poprimaju vrijednosti u intervalu  $[0,1]$  te nije centrirana oko nule. Tangens - hiperbolna funkcija poprima vrijednosti u intervalu  $[-1,1]$  i centrirana je oko nule. Prijenosne funkcije prikazane su na slici 19.

Bitan parametar u modelu neuronskih mreža je koeficijent učenja,  $\eta$  čija je uloga nametanje brzine promjene težinskih koeficijenata,  $w_n$  s obzirom na vrijednost pogreške,  $\varepsilon$ . Parametar je potrebno zadati jer o njemu ovisi brzina konvergencije [27].



**Slika 19.** Najčešće korištene prijenosne funkcije [36]

### 2.9.2. Postupak učenja neuronske mreže

Rad umjetne neuronske mreže odvija se u dva koraka. Prvi korak je faza učenja odnosno treniranja neuronske mreže, a zatim testiranje neuronske mreže. Učenje mreže je proces u kojemu se koeficijenti težina mijenjaju kao reakcija na podatke koji dolaze izvana i predstavljaju ulaze u neuronsku mrežu. Odnosno, učenje neuronske mreže je zapravo podešavanje težinskih koeficijenata tako da se dobije minimalna vrijednost pogreške  $\varepsilon$ , između izlaza dobivenih modelom i stvarne vrijednosti izlazne veličine. Postoje različiti algoritmi učenja mreža prema kojima se izračunava težinski koeficijent te različite paradigme učenja

definiraju odnos neuronske mreže prema okolini. Osnovne paradigme učenja neuronske mreže su: učenje pod nadzorom (postoji znanje o okolini u obliku parova ulaz-izlaz i mreža se uči sve dok ne nauči imitirati odnos ulaz-izlaz), učenje podrškom (princip pokušaja i pogreške, gdje se indeks mjere kvalitete učenja maksimizira) i učenje bez nadzora (mreža se optimizira, odnosno njezini parametri s obzirom mjeru znanja koju mreža mora zadovoljiti) [37].

*Algoritam učenja* je formula koja se upotrebljava za prilagođavanje težinskih koeficijenata  $w_n$  među neuronima. Najčešće su korištena ova četiri pravila [38]:

- Delta pravilo (Widrow/Hoff-ovo pravilo),
- Poopćeno Delta pravilo,
- Delta-Bar-Delta i Prošireno Delta-Bar-Delta pravilo,
- Kohonen-ovo pravilo (koristi se za nenadgledane mreže).

Delta pravilo učenja poznato je i kao pravilo najmanjih srednjih kvadrata, budući da ima za cilj minimizirati ciljnu funkciju određivanjem vrijednosti težinskih faktora. Cilj je minimizirati sumu kvadrata grešaka, gdje je greška definirana kao razlika između izračunatog i stvarnog (željenog) izlaza nekog neurona za dane ulazne podatke. Jednadžba za Delta pravilo je:

$$\Delta w_{ij} = \eta \cdot y_i \cdot \varepsilon_j \quad (16)$$

$$\Delta w_{ij} = w_{ij}^{nova} - w_{ij}^{stara} \quad (17)$$

$$\varepsilon_i = y_i - y_i' \quad (18)$$

gdje su:  $w_{ij}$  – neuron  $i$  povezan s neuronom  $j$ ,

$y_j$  - vrijednost izlaza izračunatog u neuronu  $j$ ,

$y_j'$  - vrijednost stvarnog željenog izlaza neurona  $i$ ,

$\eta$  – koeficijent učenja,

$\varepsilon_j$  - sirova greška [39]

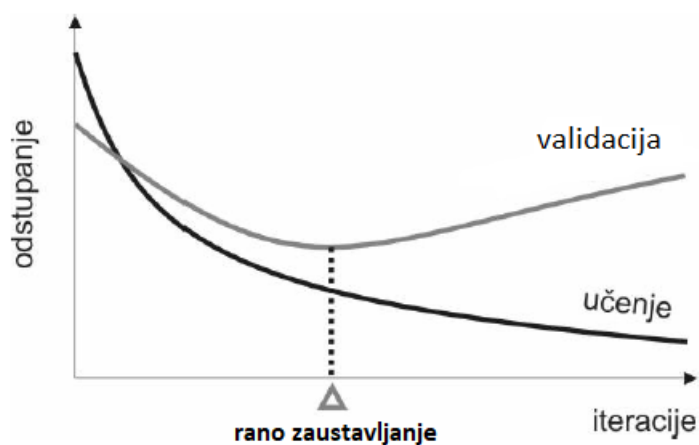
Novi koeficijent težine između neurona računa se tako da se stari koeficijent korigira za vrijednost greške ( $\varepsilon$ ) pomnožen s vrijednošću izlaza neurona  $j$  ( $y_j$ ), te sa koeficijentom učenja ( $\eta$ ).

Dva su problema ove vrste algoritma:

- Problem lokalnog minimuma pojavljuje se kada je najmanja greška funkcije pronađena samo za lokalno područje te je učenje zaustavljeno bez dostizanja globalnog minimuma.
- Problem pretreniranja pojavljuje se jer nije moguće unaprijed odrediti koliko dugo treba učiti mrežu da bi ona mogla naučeno znanje generalizirati na novim podacima [39].

### 2.9.3. Broj iteracija i pretreniranost neuronske mreže

Prilikom učenja neuronske mreže razlikujemo kao što je već rečeno skup za treniranje (učenje), skup za validaciju i skup za testiranje. Prvi skup služi za podešavanje težinskih koeficijenata i za treniranje neuronske mreže u užem smislu. Skup za validaciju služi kako bi se tokom učenja neuronske mreže provjeravao rad s trenutnim težinskim koeficijentima, kako bi se postupak učenja zaustavio u trenutku degradacije performanse mreže. Umjetnu neuronsku mrežu moguće je *pretrenirati*, odnosno da nakon određenog broja iteracija (koraka) mreža gubi svojstvo generalizacije i postaje stručna za obradu podataka iz skupa primjera za učenje. Međutim, kada se susretne sa podacima van tog skupa, te podatke obrađuje loše. Kako bi se spriječilo pretreniranje neuronske mreže koristi se tehnika ranog zaustavljanja (engl. *early stopping*) kao što je prikazano na slici 20 [27, 34]



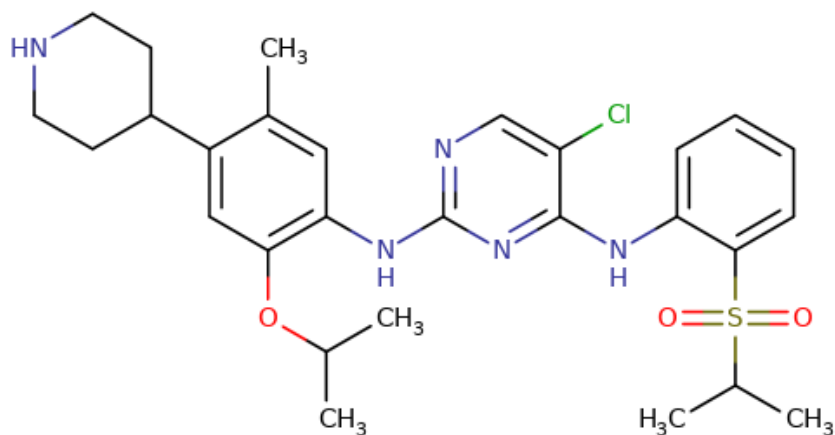
**Slika 20.** Tehnika ranog zaustavljanja [34]

### 3. EKSPERIMENTALNI DIO

#### 3.1. Materijali

Djelatna tvar za izradu kalibracijskih modela je ceritinib forme A (molekulska formula:  $C_{28}H_{36}ClN_5O_3S$ , molekulska masa: 558,1 g/mol) otopljen u tetrahidrofuranu (THF, molekulska formula:  $C_4H_8O$ , molekulska masa: 72,11 g/mol). Na slici 21 prikazana je struktura ceritiniba.

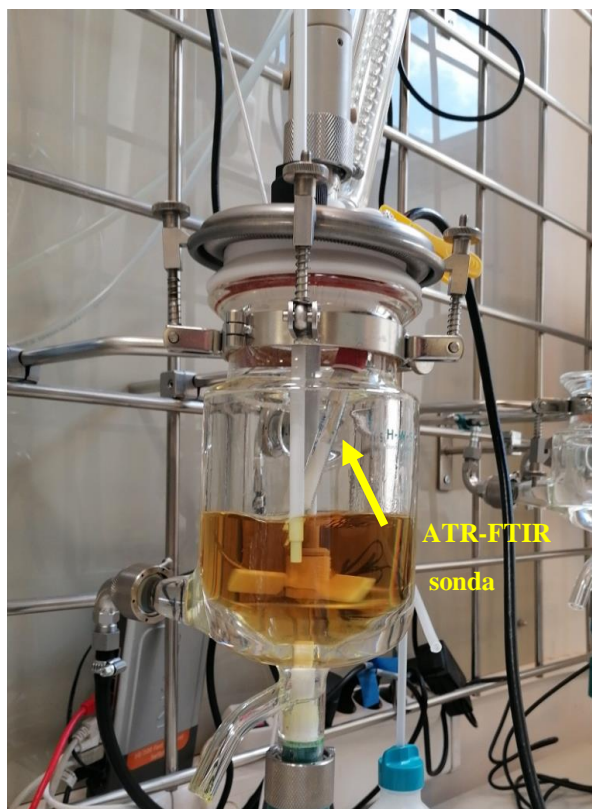
Ceritinib je vrsta ciljanog (pametnog) lijeka. Koristi se u terapiji tretiranja raka pluća koji se naziva rak nemalih stanica (engl. *non-small cell lung cancer*, NSCLC). Ceritinib se javlja u 3 različite polimorfne forme, A, B i C. IUPAC naziv je 5-chloro-2-*N*-(5-methyl-4-piperidin-4-yl-2-propan-2-yloxyphenyl)-4-*N*-(2-propan-2-ylsulfonylphenyl)pyrimidine-2,4-diamine.



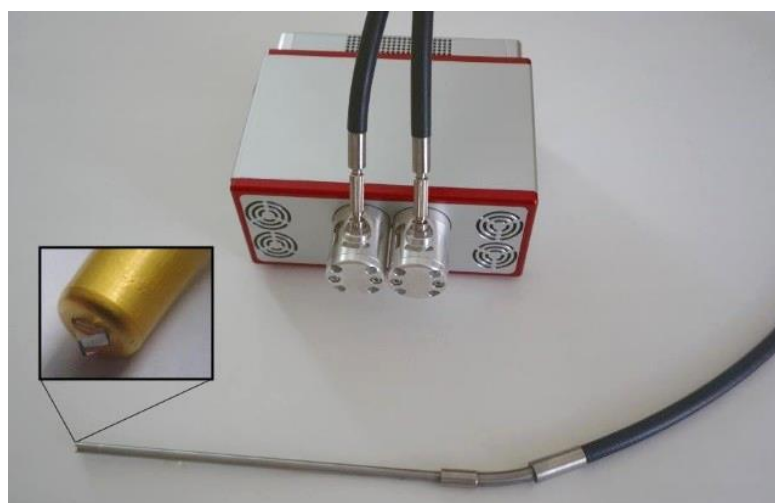
**Slika 21.** Ceritinib (5-chloro-2-*N*-(5-methyl-4-piperidin-4-yl-2-propan-2-yloxyphenyl)-4-*N*-(2-propan-2-ylsulfonylphenyl)pyrimidine-2,4-diamine)[23]

### 3.2. Uredaji

Aparatura za provedbu eksperimenta sastoji se od: staklenoga kristalizatora s plaštem, temperaturnoga osjetila Pt-100, povratnog hladila, termostata za regulaciju temperature te miješala pokretanog motorom (slika 22). Za prikupljanje spektralnih podataka upotrebljava se ATR-FTIR spektrometar (ARC Optics), čije je sonda (slika 23) uronjena u otopinu.



**Slika 22.** Kristalizator s plaštem, miješalom i ATR-FTIR sondom



**Slika 23.** ATR- FTIR spektrometar



### 3.3. Eksperimenti

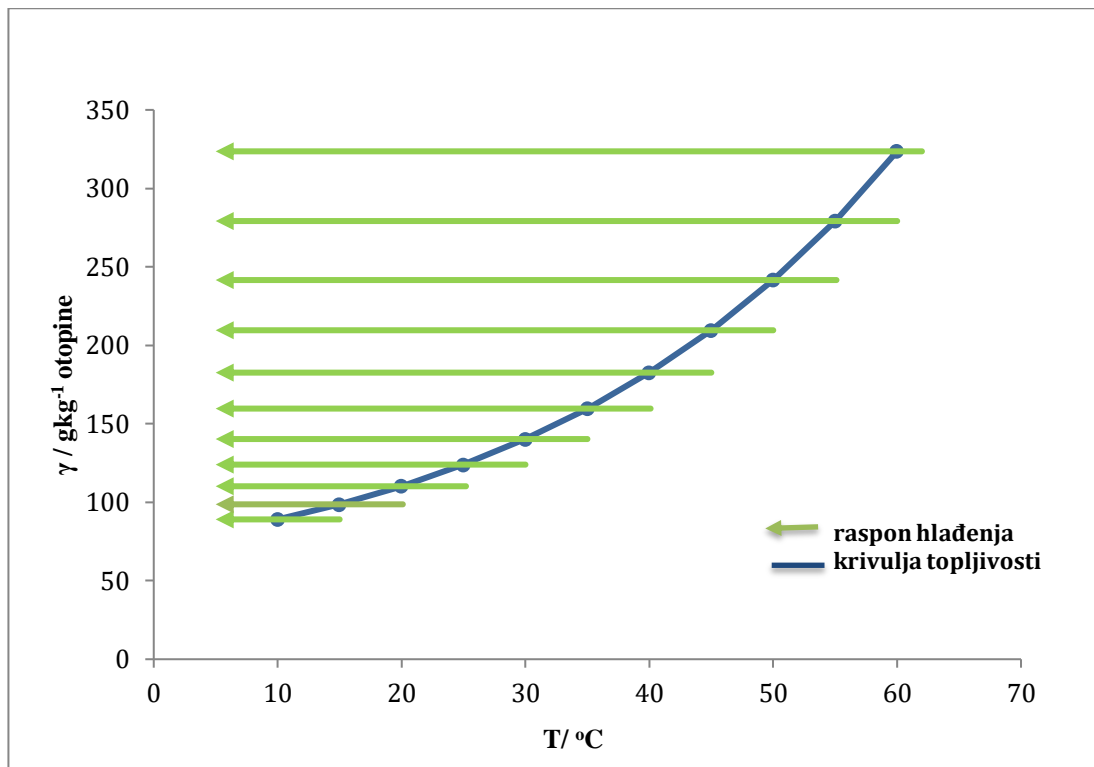
#### 3.3.1. Prikupljanje podataka

Na temelju krivulje topljivosti ceratiniba forme A za eksperiment je odabrano 11 koncentracija. Prvi korak bila je odvaga definiranih masa ceritiniba kako bi se u kristalizatoru pripremile otopine koncentracija: 89,1; 98,7; 110,2; 124,0; 140,3; 159,7; 182,6; 209,6; 241,6; 279,2 i 323,6 g/kg<sub>THF</sub>. U kristalizatoru su pripremljene otopine od najniže prema najvišoj koncentraciji dodatkom odgovarajućih odvaga kao što je prikazano u tablici 1.

**Tablica 1.** Odvage ceritiniba forme A za pripremu otopina potrebnih koncentracija

| <b>SUSTAV: ceritinib forme A + THF</b> |                |                        |
|--|----------------|------------------------|
| $\gamma$ , g/kg <sub>THF</sub>         | $\Delta m$ , g | $T_{\text{ravn}}$ , °C |
| <b>89,1</b>                            | 47,51          | 10                     |
| <b>98,7</b>                            | 5,08           | 15                     |
| <b>110,2</b>                           | 6,12           | 20                     |
| <b>124,0</b>                           | 7,30           | 25                     |
| <b>140,3</b>                           | 8,70           | 30                     |
| <b>159,7</b>                           | 10,31          | 35                     |
| <b>182,6</b>                           | 12,20          | 40                     |
| <b>209,6</b>                           | 14,40          | 45                     |
| <b>241,6</b>                           | 17,00          | 50                     |
| <b>279,2</b>                           | 20,10          | 55                     |
| <b>323,6</b>                           | 23,60          | 60                     |

Prikupljanje spektralnih podataka ceritiniba forme A u tetrahidrofuranu provedeno je ATR-FTIR spektroskopijom. Mjerenje kreće pri temperaturi višoj od temperature topljivosti za svaku koncentraciju (otopina u nezasićenom području) te se hladi 0,5 °C/min do 5°C. ATR-FTIR sonda je uronjena u otopinu, a uređaj kontinuirano prikuplja spektralne podatke koji se spremaju na računalo u Excel datoteku. Za očitane temperature u trenutku snimanja pri određenoj koncentraciji dobiva se spektar s rasponom valnoga broja od 836 do 5000 cm<sup>-1</sup>. Na slici 24 prikazana je krivulja topljivosti (plava linija), koncentracije za koje su provedena mjerenja označene su točkama. Zelenim linijama označen je temperaturni raspon za koji su snimljeni ATR-FTIR spektri.



**Slika 24.** ATR-FTIR snimanje otopina

### 3.3.2. Eksperimentalno praćenje kristalizacije linearnim hlađenjem

Pri temperaturi od 60°C moguće je otopiti 323 g ceritiniba forme A u 1kg THF-a. Hlađenjem otopine do 50°C dobivena je prezasićena otopina, te je u takvu otopinu dodana mala količina kristala (cjepiva). Nakon toga provedena je kristalizacija linearnim hlađenjem.

## 4. REZULTATI I RASPRAVA

### 4.1. Predobrada spektralnih podataka

Nakon prikupljanja „sirovih podataka“ podaci se predobrađuju. Predobrada obuhvaća korekciju bazne linije pomoću Matlab funkcije *msbackadj*, te dodatnu korekciju bazne linije funkcijom *baseline*. Normalizira se primjenom *matrix 2-norm* normalizacija:

$$\|X\|_2 = \sqrt{\lambda_{\max}(X^T X)} \quad (19)$$

$$\det(X^T X - \lambda I) = 0 \quad (20)$$

Izraz (20) služi za dobivanje svojstvenog vektora, pri čemu je:

$X^T$  - transponirana matrica,

$X$  – originalna matrica

$\lambda$  - vrijednosti svojstvenog vektora

$I$  – jedinična matrica; matrica identiteta

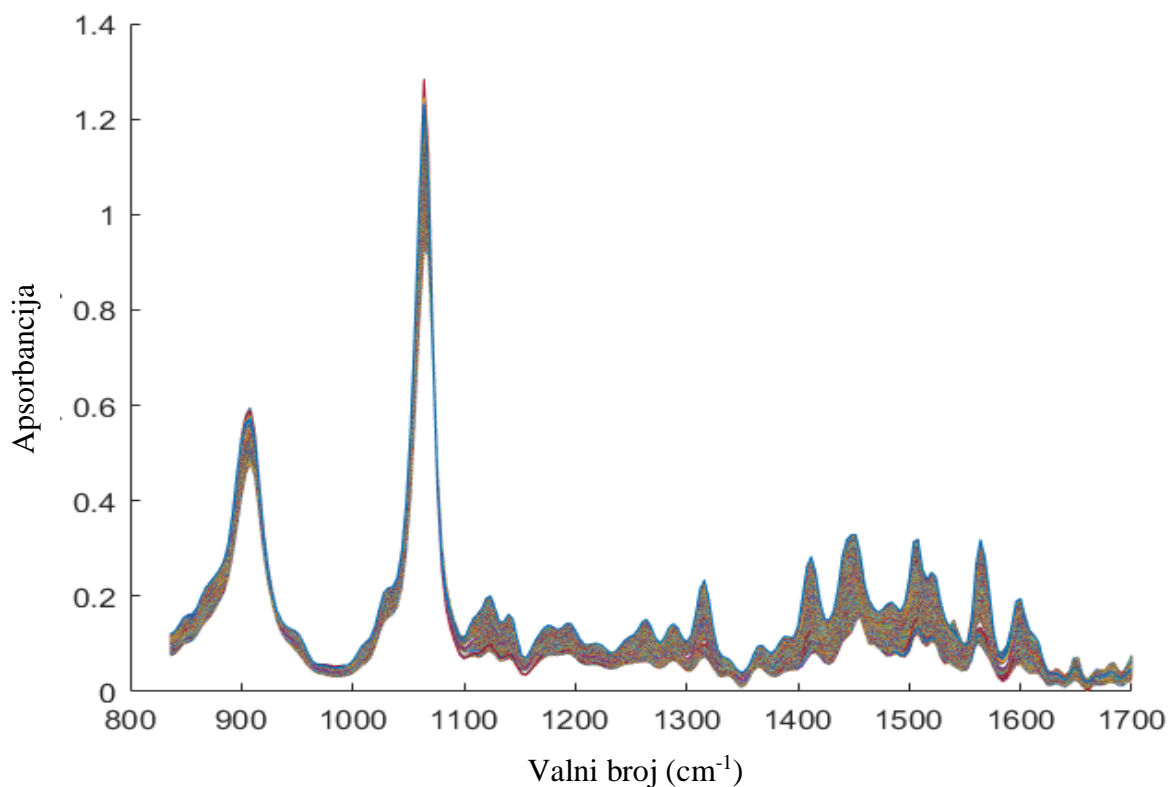
*Matrix 2-norm* je drugi korijen rezultata množenja najvećeg svojstvenog vektora ( $\lambda_{\max}$ ) s umnoškom transponirane ( $X^T$ ) i originalne matrice ( $X$ ). Daljnji korak u normalizaciji je desno matrično dijeljenje spektralnih podataka s normaliziranom vrijednošću. Nakon toga podeševa se bazna linija u  $y=0$  i za filtriranje podataka korišten je Savitzky-Golay filter.

### 4.2. Razvoj kalibracijskog modela

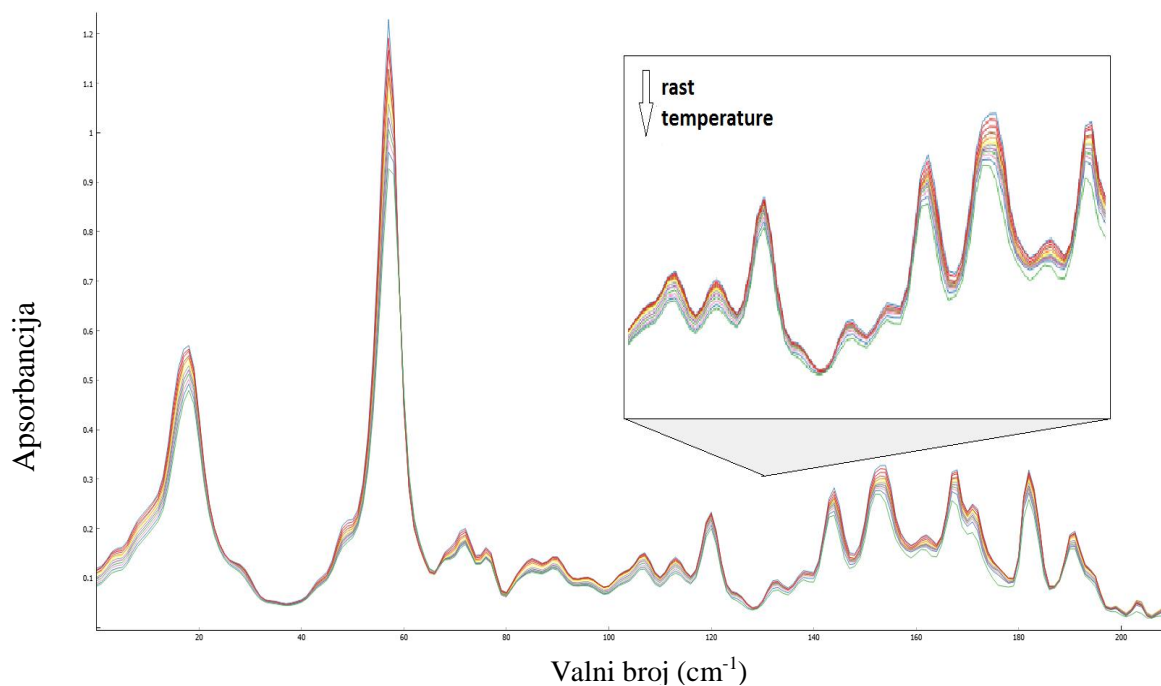
Prikupljeni i predobrađeni podaci upotrijebljeni su za razvoj kalibracijskih modela. Eksperimentalni skup podataka sastoji se od intenziteta vrpce ATR-FTIR spektra, temperature i koncentracije otopine. Metodama PLSR i umjetnom neuronskom mrežom predviđaju se koncentracije otopine. Iz cijelog skupa podataka izdvajaju se podaci koji će odgovarati  $X$  matrici koja sadrži spektre i podaci koji će odgovarati  $Y$  matrici koja sadrži koncentracije otopine ceritiniba forme A u THF-u.

Slikom 25 prikazani su svi ATR-FTIR spektri provedenog eksperimenta.  $X$  os označava valne brojeve, odnosno promjenu valne duljine reflektiranih fotona, dok su na  $y$  osi prikazana apsorbancija, odnosno intenziteti koji predstavljaju broj detektiranih reflektiranih fotona. Temperatura, kao i koncentracija, utječe na intenzitet ATR-FTIR spektra. Pri konstantnoj

koncentraciji apsorbancija se smanjuje porastom temperature. Za razliku od toga, pri konstantnoj temperaturi apsorbancija raste porastom koncentracije. Slikom 26 prikazana je ovisnost spektra masene koncentracije 323,6 g/kg<sub>THF</sub> o temperaturi.

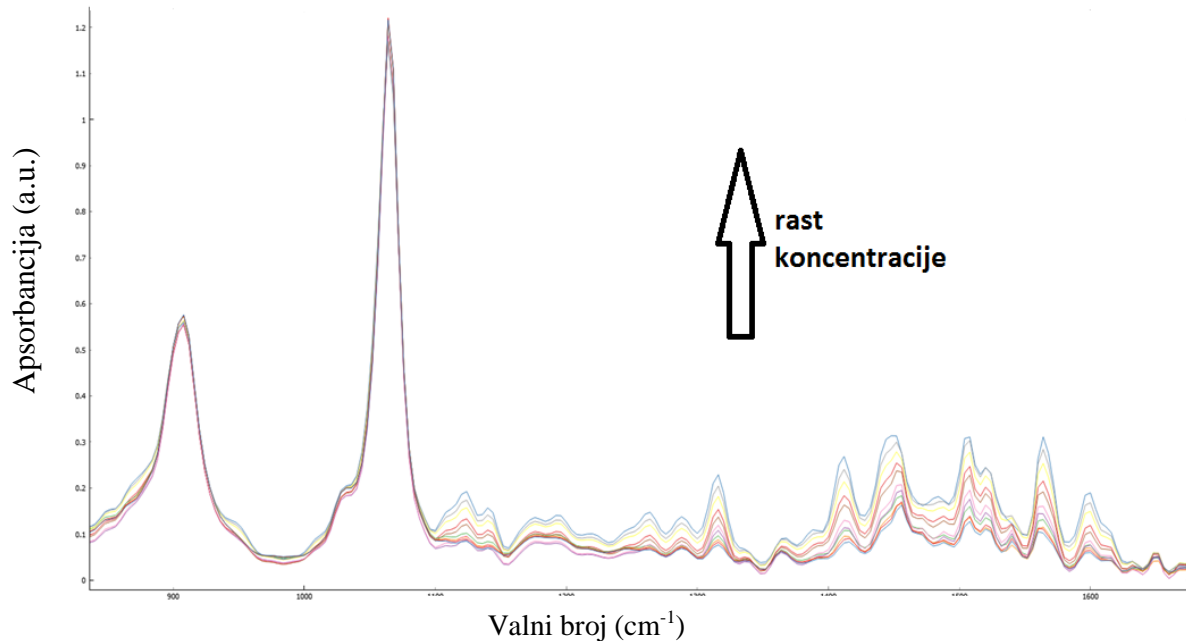


**Slika 25.** Spektri provedenog eksperimenta



**Slika 26.** Prikaz smanjenja apsorbancije s porastom temperature pri koncentraciji 323,6 g/kg<sub>THF</sub>

Na slici 27 prikazana je ovisnost spektara otopina o koncentraciji pri temperaturi od 15 °C.



**Slika 27.** Promjena apsorbancije s porastom koncentracije pri temperaturi od 15 °C

#### 4.2.1. PLSR metoda

PLSR metoda reducira dimenzionalnost podataka na manji broj latentnih varijabli kao što je objašnjeno u poglavlju 2.6. Kao kriterij valjanosti metode uzete su korijen srednje kvadratne pogreške (eng. *root mean squared error*, *RMSE*) i koeficijent korelacije (eng. *coefficient of determination*,  $R^2$ ). Koeficijent korelacije,  $R^2$  ocjenjuju podudarnost vladanja modela s vladanjem stvarnog procesa unutar postavljenih radnih uvjeta na neovisnom skupu podataka. On opisuje korelaciju između predviđanja i eksperimentalnih podataka. Ako teži 1 tada je veliki dio varijance obuhvaćen modelom. Korijen srednje kvadratne pogreške, RMSE mjeri prosječnu razliku između predviđenih vrijednosti modela i stvarnih vrijednosti. Matematički, to je standardna devijacija reziduala (pogreške predviđanja).

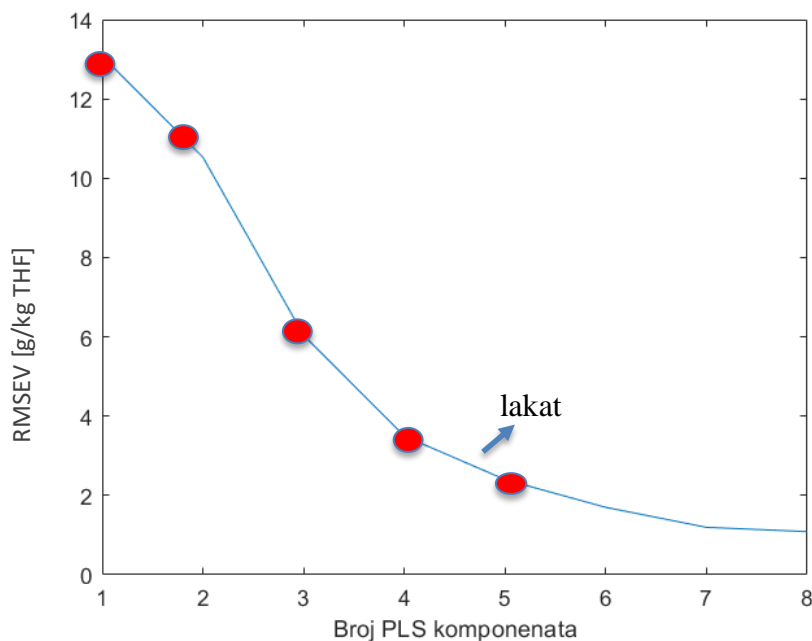
$$RMSE = \sqrt{\frac{\sum_{i=1}^n (y - \hat{y})^2}{n}} \quad (21)$$

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (22)$$

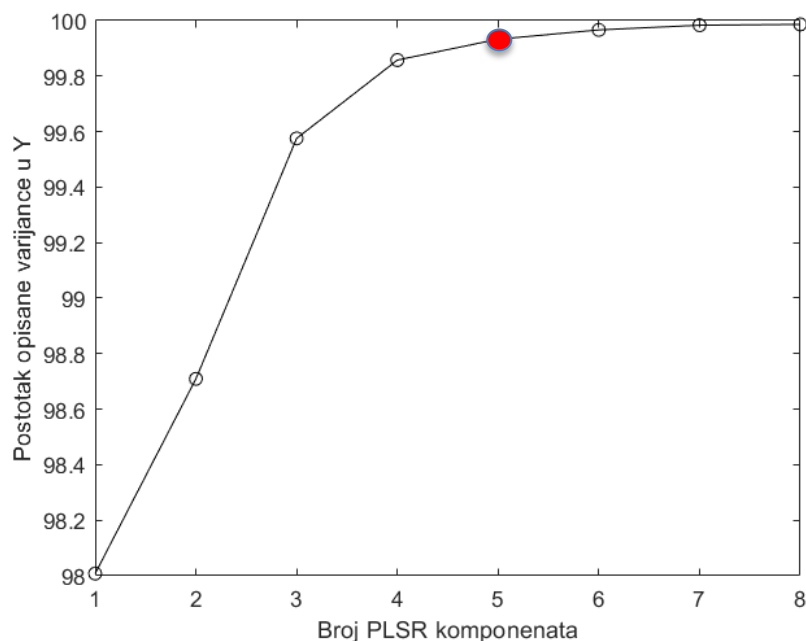
Pri čemu su:

- $\hat{y}_i$  – izlaz predviđen modelom
- $y_i$  – eksperimentalna vrijednost
- $\bar{y}$  – srednja vrijednost  $y$

Prije izrade PLSR modela, podaci su podijeljeni u tri skupa: za uvježbavanje, testiranje i validaciju kao što je prikazano u tablici 2. Skup za testiranje sadrži jednu koncentraciju za koju model testira. Od deset koncentracija koje preostanu nakon određivanja testne koncentracije, osam koncentracija je u skupu za uvježbavanje, dok su preostale dvije u skupu za validaciju. Izvedena su četiri slučaja, u svakome su koncentracije za test i validaciju odabrane proizvoljno. Iznimka je slučaj 1 pri čemu se dvije koncentracije izuzete i jedna odabrana kao testna. Od preostalih koncentracija dvije su uvrštene u skup za validaciju, a ostale u skup za uvježbavanje. Svaki slučaj sadrži skup za uvježbavanje, validaciju i test.



**Slika 28.** Odabir broja komponenata za PLSR metodu



**Slika 29.** Postotak opisane varijance za različiti broj komponenata

Kao optimalni broj komponenti uzima se onaj pri kojem se u grafičkom prikazu na slici 28 uočava smanjenje nagiba padajuće eksponencijalne krivulje, odnosno točka sa manjom promjenom postotka opisanosti varijance s brojem komponenata. Isti trend uočava se i na slici 29 gdje se za 5 komponenta opisuje već 99,93% varijance. Optimalan broj komponenata je 5, međutim radi istraživanja modela uzeto je u obzir još i 4 i 6 komponenata. Cilj je postići što bolju procjenu koncentracije, što bližu stvarnoj koncentraciji, a da model ne bude previše kompleksan. Isti broj komponenata uzet je za sva četiri slučaja iz tablice 2.

**Tablice 2.** Kombinacije koncentracija za izradu optimalnog PLSR modela

| Koncentracija g/kg <sub>THF</sub> | 1. slučaj  | 2. slučaj  | 3. slučaj  | 4. slučaj  |
|-----------------------------------|------------|------------|------------|------------|
| 89,1                              | trening    | trening    | trening    | trening    |
| 98,7                              | trening    | validacija | trening    | test       |
| 110,2                             | trening    | trening    | validacija | trening    |
| 124,0                             | trening    | trening    | trening    | trening    |
| 140,3                             | izuzeto    | trening    | trening    | trening    |
| 159,7                             | izuzeto    | test       | trening    | trening    |
| 182,6                             | trening    | trening    | test       | validacija |
| 209,6                             | validacija | trening    | trening    | validacija |
| 241,6                             | validacija | trening    | trening    | trening    |
| 279,2                             | test       | trening    | validacija | trening    |
| 323,6                             | trening    | validacija | trening    | trening    |

Tablica 3 prikazuje vrijednosti  $R^2$  i RMSEV (korijen srednje kvadratne pogreške za validacijski skup), kriterije valjanosti PLSR modela. Što je manja vrijednost RMSEV, a veća vrijednost  $R^2$  to bolje je podudaranje između procijenjenih vrijednosti i stvarnih vrijednosti koncentracija otopina.

**Tablica 3.** Vrijednosti  $R^2$  i RMSEV za pojedine slučajeve

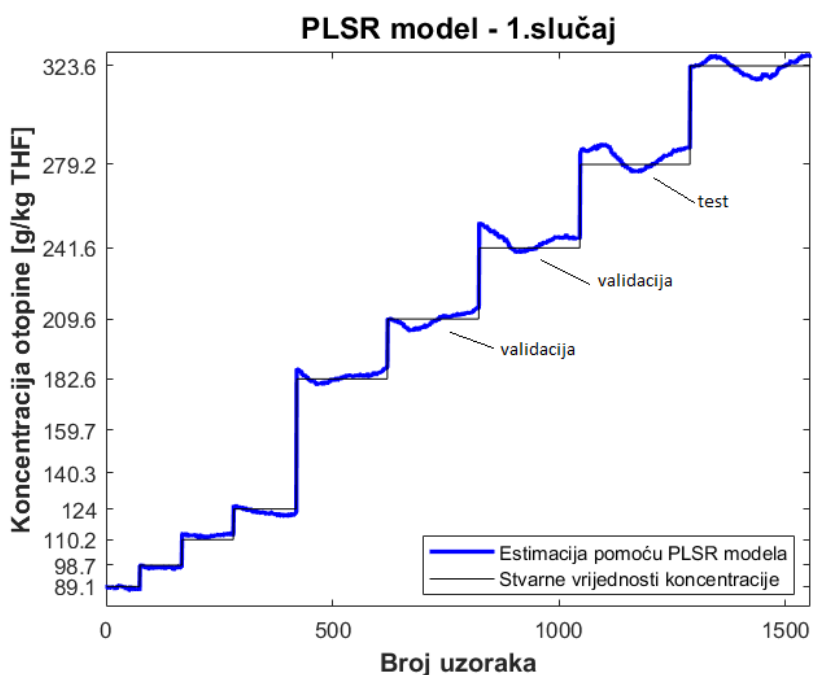
|           | Broj komponenata | $R^2$         | RMSEV         |
|-----------|------------------|---------------|---------------|
| 1. slučaj | 4                | 0,9973        | 4,0095        |
|           | <b>5</b>         | <b>0,9980</b> | <b>3,4670</b> |
|           | 6                | 0,9975        | 3,8861        |
| 2. slučaj | 4                | 0,9906        | 7,2199        |
|           | 5                | 0,9964        | 4,4845        |
|           | <b>6</b>         | <b>0,9978</b> | <b>3,5225</b> |
| 3. slučaj | 4                | 0,9955        | 4,9863        |
|           | 5                | 0,9953        | 5,0938        |
|           | <b>6</b>         | <b>0,9965</b> | <b>4,4211</b> |
| 4. slučaj | <b>4</b>         | <b>0,9950</b> | <b>5,2481</b> |
|           | 5                | 0,9911        | 7,0129        |
|           | 6                | 0,9920        | 6,6443        |

Prema tablici 3. u Matlabu je izrađeno dvanaest PLSR modela za estimaciju koncentracije ceritinba forme A u THF-u. Za sva 4 slučaja mijenjan je broj latentnih varijabli, odnosno broj komponenata od 4 do 6 kako bi se, nakon treniranja, odabrao model koji najbolje procjenjuje koncentraciju ceritiniba. Slikama 30., 31., 32. i 33. prikazan su 4 najbolja PLSR modela prema vrijednostima  $R^2$  i RMSE, svaki model za jedan slučaj i određen broj komponenata prema tablici 3.

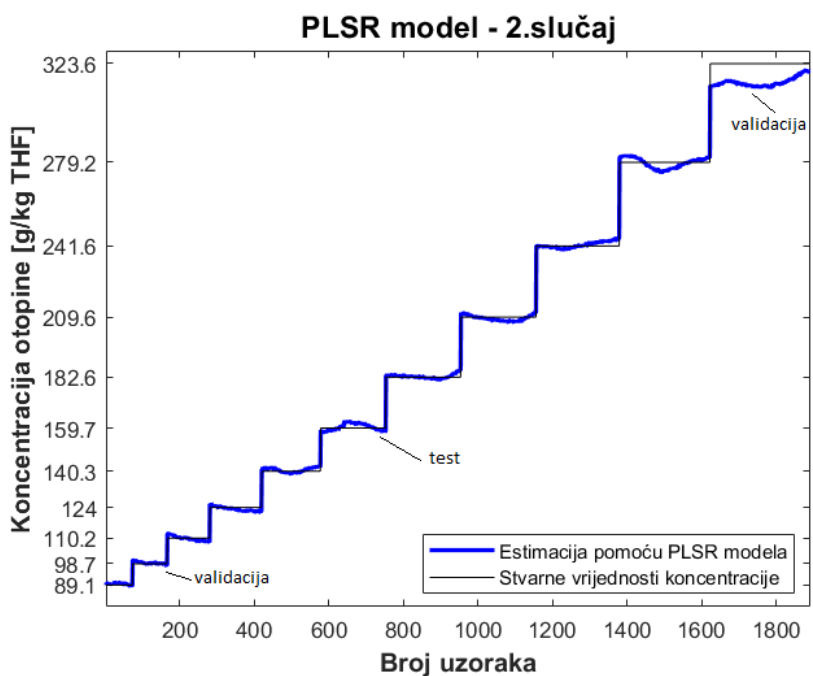
Slike 30 i 31 prikazuju najmanju RMSE i najveću  $R^2$  vrijednost što govori najboljoj mogućoj procjeni koncentracije ceritiniba u THF-u. Modeli nisu pretrenirani zbog načina na koji je proveden odabir broja komponenata. Na slici 30 tri su uzastopne koncentracije izuzete iz skupa za treniranje, te model daje zadovoljavajuće vrijednosti kriterija uspješnosti s obzirom na nepoznavanje triju uzastopnih koncentracija, dok na slici 31 primjećujemo pomak (engl. *offset*) pri ekstrapolaciji modela za najveću koncentraciju.



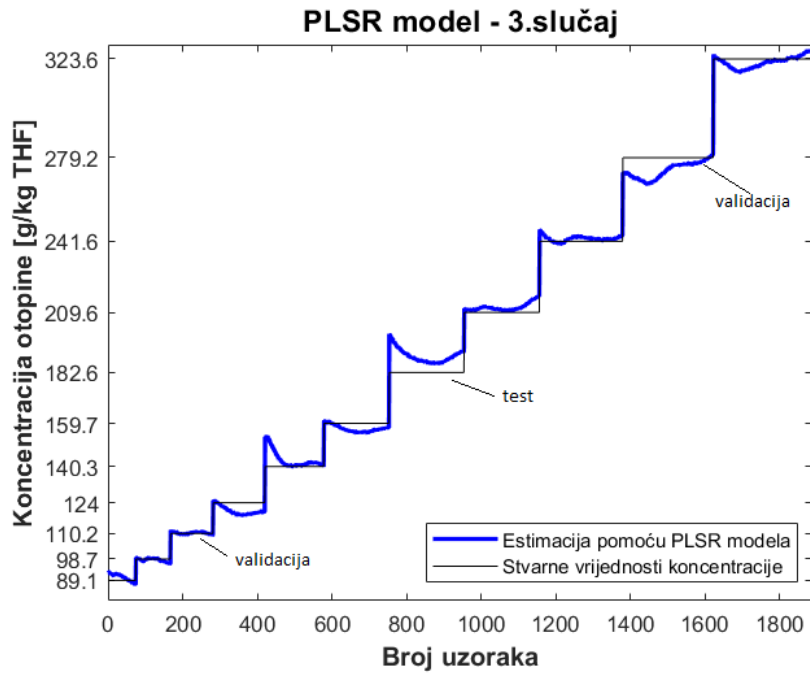
Modeli prikazani slikama 32 i 33 ne daju zadovoljavajuću procjenu koncentracija prema kriterijima valjanosti modela. Na grafičkim prikazima su izražene nelinearnostima u procjeni pri svakoj od tih koncentracija.



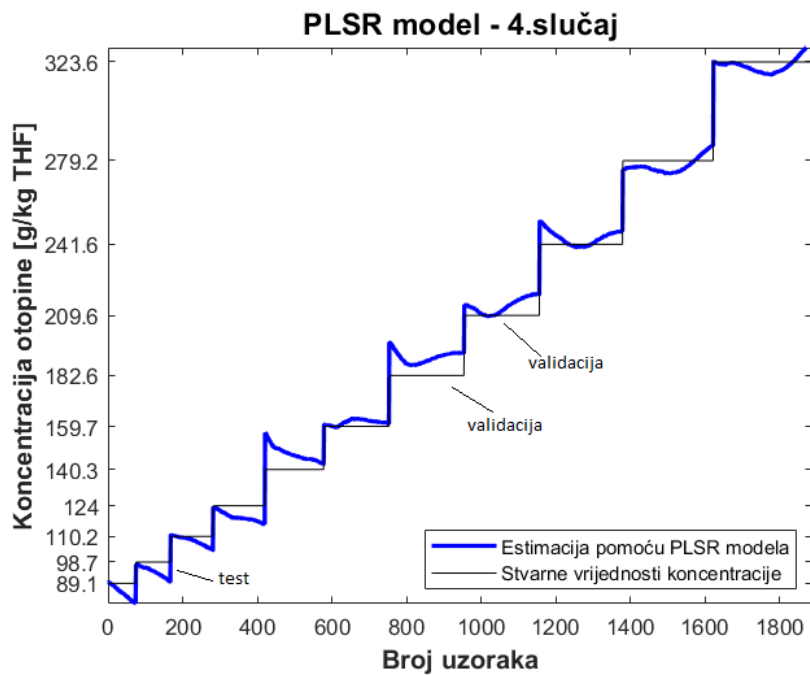
Slika 30. Prikaz PLSR modela za 1.slučaj s 5 komponenata



Slika 31. Prikaz PLSR modela za 2.slučaj s 6 komponenata



Slika 32. Prikaz PLSR modela za 3.slučaj s 6 komponenta



Slika 33. Prikaz PLSR modela za 4.slučaj s 4 komponente

#### 4.2.2. Umjetne neuronske mreže

Prilikom razvoja umjetnih neuronskih mreža u obzir su uzeti samo 1. i 2. slučaj iz tablice 3 sa svojim definiranim test skupom i skupom za validaciju. Razlog tome je u mogućnosti usporedbe najboljeg PLSR modela s najboljim modelom neuronskih mreža. Prilikom izrade modela neuronskih mreža preinaka je u tome što ne postoji validacijski skup već ono što je pri izradi PLSR modela bilo u skupu za test i za validaciju, sada postaje skup za testiranje. Skup za treniranje u oba slučaja ostaje isti što je ključno za usporedbu ovih dvaju metoda.

**Tablica 4.** Kombinacije koncentracija za izradu optimalnog NN modela

| Koncentracija g/kg <sub>THF</sub> | 1. slučaj | 2. slučaj |
|-----------------------------------|-----------|-----------|
| 89,1                              | trening   | trening   |
| 98,7                              | trening   | test      |
| 110,2                             | trening   | trening   |
| 124,0                             | trening   | trening   |
| 140,3                             | izuzeto   | trening   |
| 159,7                             | izuzeto   | test      |
| 182,6                             | trening   | trening   |
| 209,6                             | test      | trening   |
| 241,6                             | test      | trening   |
| 279,2                             | test      | trening   |
| 323,6                             | trening   | test      |

Kao kriterij valjanosti metode umjetnih neuronskih mreža upotrijebljene su vrijednosti RMSE i  $R^2$ . Promjenom hiperparametara nastoji se optimirati model koji daje što manju RMSE vrijednost i vrijednost  $R^2$  što bližu 1. U tablicama 5 do 8 prikazana je optimizacija parametara neuronske mreže. Pri tome je mijenjan broj neurona u skrivenom sloju (5, 8, 24 ili 32 neurona) i algoritam učenja. Broj epoha je 10, prijenosna aktivacijska funkcija je tangentska sigmoidna funkcija a broj skrivenih slojeva je 1.

**Tablica 5.** Optimizacija modela neuronske mreže s 5 neurona u skrivenom sloju

|                           | <b>Broj neurona u skrivenom sloju = 5, prijenosna aktivacijska fj. <i>tansig</i></b> |                      |                  |                      |
|---------------------------|--|----------------------|------------------|----------------------|
| <b>Algoritam učenja</b>   | <b>1. slučaj</b>   |                      | <b>2. slučaj</b> |                      |
|                           | <b>RMSE</b>  | <b>R<sup>2</sup></b> | <b>RMSE</b>      | <b>R<sup>2</sup></b> |
| Levenberg-Marquardt       | 5,1656   | 0,9956               | 2,9344           | 0,9984               |
| Bayesian Regularization   | 7,2014   | 0,9914               | 6,9021           | 0,9914               |
| Scaled Conjugate Gradient | 7,1443   | 0,9915               | 5,1343           | 0,9952               |

**Tablica 6.** Optimizacija modela neuronske mreže s 8 neurona u skrivenom sloju

|                           | <b>Broj neurona u skrivenom sloju = 8, prijenosna aktivacijska fj. <i>tansig</i></b> |                      |                  |                      |
|---------------------------|--|----------------------|------------------|----------------------|
| <b>Algoritam učenja</b>   | <b>1. slučaj</b>   |                      | <b>2. slučaj</b> |                      |
|                           | <b>RMSE</b>  | <b>R<sup>2</sup></b> | <b>RMSE</b>      | <b>R<sup>2</sup></b> |
| Levenberg-Marquardt       | <b>2,4203</b>  | <b>0,9990</b>        | <b>1,5888</b>    | <b>0,9995</b>        |
| Bayesian Regularization   | 5,5794   | 0,9948               | 5,7570           | 0,9940               |
| Scaled Conjugate Gradient | 27,9515  | 0,8700               | 7,3239           | 0,9903               |

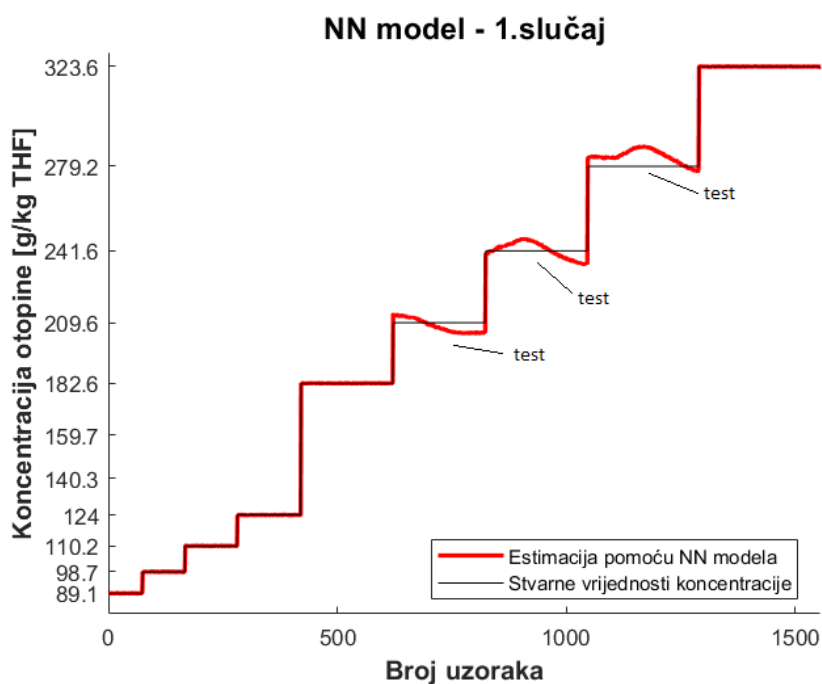
**Tablica 7.** Optimizacija modela neuronske mreže s 24 neurona u skrivenom sloju

|                           | <b>Broj neurona u skrivenom sloju = 8, prijenosna aktivacijska fj. <i>tansig</i></b> |                      |                  |                      |
|---------------------------|--|----------------------|------------------|----------------------|
| <b>Algoritam učenja</b>   | <b>1. slučaj</b>   |                      | <b>2. slučaj</b> |                      |
|                           | <b>RMSE</b>  | <b>R<sup>2</sup></b> | <b>RMSE</b>      | <b>R<sup>2</sup></b> |
| Levenberg-Marquardt       | 12,0368  | 0,9759               | 5,3395           | 0,9949               |
| Bayesian Regularization   | 3,2882   | 0,9982               | 4,1828           | 0,9968               |
| Scaled Conjugate Gradient | 23,0069  | 0,9119               | 18,1066          | 0,9409               |

**Tablica 8.** Optimizacija modela neuronske mreže s 32 neurona u skrivenom sloju

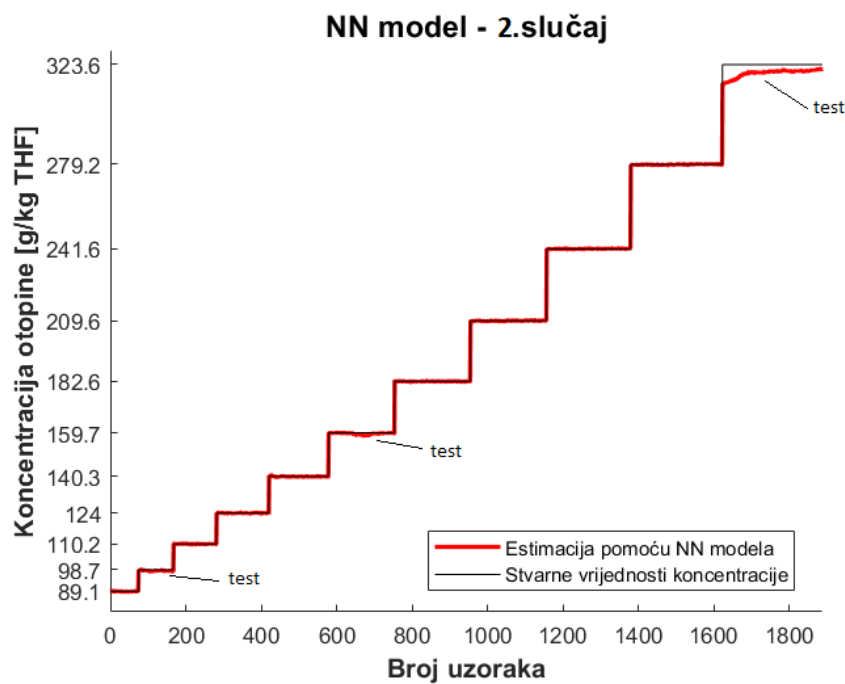
|                           | <b>Broj neurona u skrivenom sloju = 8, prijenosna aktivacijska fj. <i>tansig</i></b> |                      |                  |                      |
|---------------------------|--|----------------------|------------------|----------------------|
| <b>Algoritam učenja</b>   | <b>1. slučaj</b>   |                      | <b>2. slučaj</b> |                      |
|                           | <b>RMSE</b>  | <b>R<sup>2</sup></b> | <b>RMSE</b>      | <b>R<sup>2</sup></b> |
| Levenberg-Marquardt       | 10,5141  | 0,9816               | 6,3198           | 0,9928               |
| Bayesian Regularization   | 3,1560   | 0,9983               | 3,3981           | 0,9979               |
| Scaled Conjugate Gradient | 27,7042  | 0,8723               | 12,8531          | 0,9702               |

Slike 34 i 35 prikazuju najbolje modele neuronskih mreža za 1. i 2. slučaj što je vidljivo i iz tablice 6. Na slici 34 tri su uzastopne koncentracije izuzete iz skupa za treniranje, te model daje zadovoljavajuće vrijednosti kriterija uspješnosti s obzirom na nepoznavanje triju uzastopnih koncentracija. Prisutna je nelinearnost modela pri procjeni koncentracija testnog skupa.



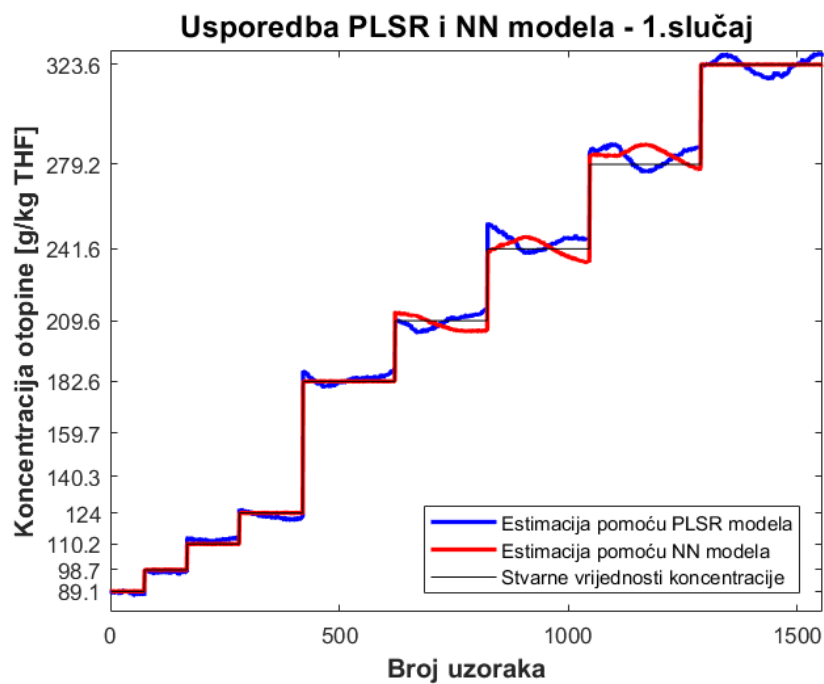
**Slika 34.** Prikaz modela neuronske mreže za 1. slučaj

Na slici 35 uočavamo podudaranje procijenjenih vrijednosti koncentracija sa stvarnim vrijednostima koncentracija, sve do zadnje koncentracije 323,6 g/kg THF-a. Zadnja koncentracija je najveća koncentracija u ovome eksperimentu i pri ekstrapolaciji vrijednosti koncentracije pomoću modela uočavamo pomak (engl. *offset*). To ukazuje na činjenicu da model neuronskih mreža u 2. slučaju dobro procjenjuje testne koncentracije koje su u rasponu koncentracija korištenih za treniranje modela, ali kada u obzir dolazi koncentracija izvan raspona koncentracija za treniranje, tu je procjena koncentracije lošija.

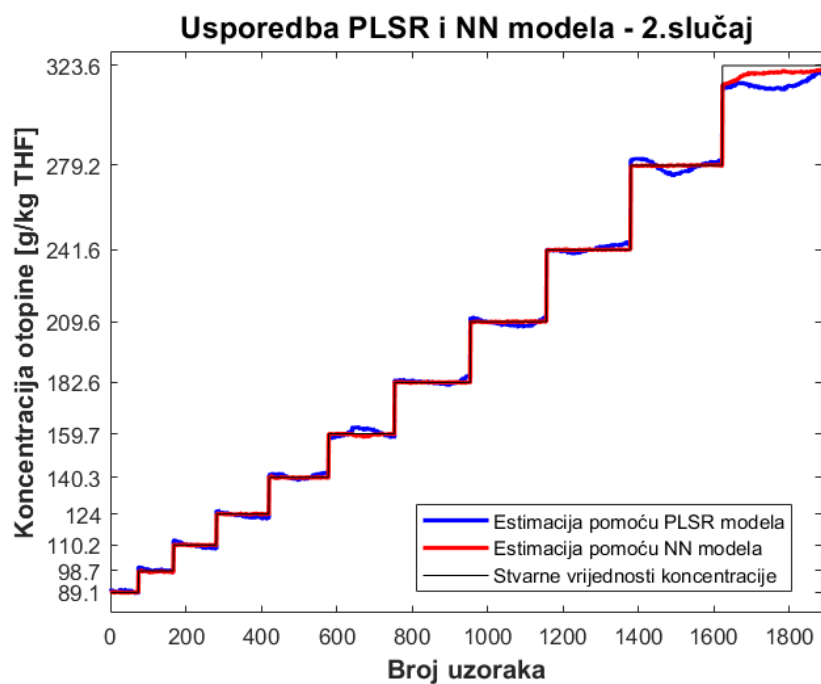


**Slika 35.** Prikaz modela neuronske mreže za 2.slučaj

Kada bismo uspoređivali PLSR modele i modele neuronskih mreža za 1. i 2. slučaj (slika 36 i slika 37) mogu se usporediti kriteriji uspješnosti. U oba slučaja neuronske mreže pokazuju se kao bolji izbor za procjenu koncentracije u ovom eksperimentu. To je najbolje vidljivo na slici 37 koja prikazuje najbolje modele za procjenu koncentracije i njihovu usporedbu. Primjećuje se da neuronske mreže bolje opisuju ovaj nelinearni sustav zbog prikazane veće linearnosti modela pri svim koncentracijama u eksperimentu. Na slici 37, za oba modela uočljiv je *offset*, međutim manji u modelu umjetne neuronske mreže nego što je to u PLSR modelu.



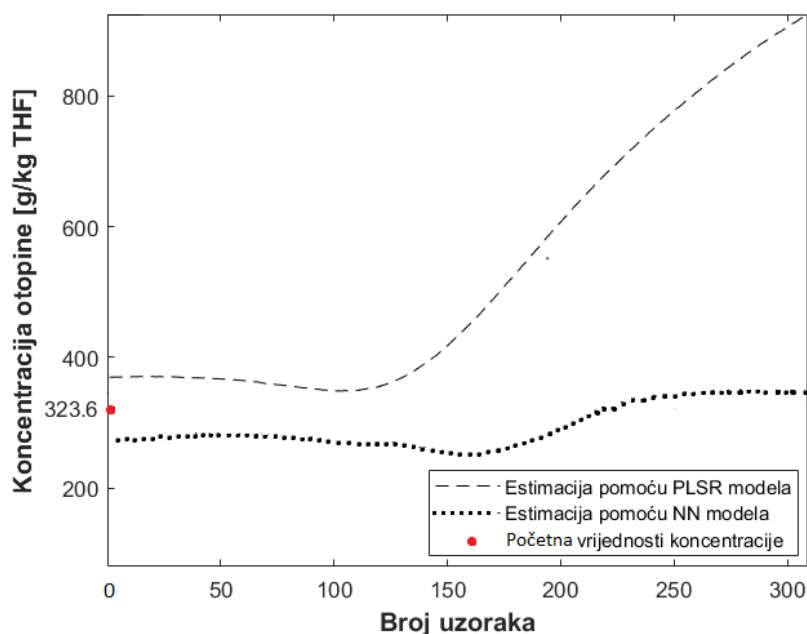
**Slika 36.** Usporedba PLSR modela i modela neuronske mreže za 1.slučaj



**Slika 37.** Usporedba PLSR modela i modela neuronske mreže za 2.slučaj

#### 4.2.3. Praćenje procesa kristalizacije linearnim hlađenjem

Prilikom provođenja dijela eksperimenta koji se odnosi na linearno hlađenje otopine koncentracije 323,6 g/kg u THF-u može se primijetiti da PLSR model kao i model umjetne neuronske mreže u jednom trenutku prikazuje porast koncentracije. U stvarnosti to nije tako jer se kristalizacijom tvari iz otopine koncentracija mora smanjivati (slika 38). Ono što je primijećeno je da postoje kristali na ATR-FTIR sondi te iz toga razloga modeli procjenjuju prividno veću koncentraciju od one stvarne.



**Slika 38.** Promjena predviđene koncentracije otopine pri kristalizaciji linearnim hlađenjem



## 5. ZAKLJUČAK

Cilj ovog rada bio je razviti model za praćenje koncentracije ceritiniba forme A u THF-u. Pri tome su primijenjene metode PLSR i umjetne neuronske mreže.

Prema dobivenim rezultatima može se zaključiti da model razvijen primjenom neuronskih mreža daje bolju procjenu koncentracije ceritiniba forme A u THF-u od PLSR modela. Linearna forma PLSR modela nešto je inferiornija jer relacije koje se opisuju karakterizira nelinearnost, stoga je model neuronske mreže dao bolje rezultate.

Prilikom praćenja procesa kristalizacije linearnim hlađenjem oba modela su, nakon određenog vremena pokazali porast koncentracije koji je nije karakterističan za proces kristalizacije. Kristalizacijom tvari koncentracija otopine treba se smanjivati, stoga porast koncentracije koju proračunavaju modeli ukazuje na problem pri mjerenje – formiranju kristala na ATR-FTIR sondi.

Kontinuirano praćenje i kontrola kvalitete farmaceutski aktivnih tvari u praksi bi bilo moguće primjenom procesne analitičke tehnologije (PAT). Ključan korak je razvoj kalibracijskog modela kao što su: model neuronskih mreža i PLSR model u ovome radu. Bez potrebe za zaustavljanjem procesa uz izvođenje *in-line* analiza omogućila bi se kvalitetnija proizvodnja i minimalizirala pogreška.

## 6. LITERATURA

- [1] S. Palosaari, M. Louhi-Kultanen, Z. Sha, 51 Industrial Crystallization, 2006, str. 1-23
- [2] H. H. Tung, E. L. Paul, M. Midler, J. A. McCauley, Crystallization of organic compounds: an industrial perspective, John Wiley and Sons, 2009, str. 1-166
- [3] J. Glassey, K. V. Gernaey, C. Clemens, T. W. Schulz, R. Oliveira, G. Striedner, C. F. Mandenius, Process analytical technology (PAT) for biopharmaceuticals, Biotechnology Journal, 2011, 6(4), 369
- [4] A. S. Myerson, D. Erdemir, A. Y. Lee, Handbook of industrial crystallization, Third Edition, Cambridge University Press, 2019, 1-187
- [5] A. Sander, Toplinsko procesno inženjerstvo, predavanja, Lekcija-Kristalizacija, FKIT, Zagreb, 2020-2021
- [6] M. Klier, Primjena procesne analitičke tehnologije za praćenje procesa kristalizacije, diplomski rad, FKIT, Sveučilište u Zagrebu, 2022., str. 38.
- [7] I. Filipović, S. Lipanović, Opća i anorganska kemija, Školska knjiga, Zagreb, 1985.
- [8] J. W. Mullin, Crystallization, Elsevier, 2001
- [9] <https://www.geeksforgeeks.org/difference-between-crystalline-and-amorphous-solids/> (5.6.2023)
- [10] S. Ahuja, S. Scypinski, (Eds.), Handbook of modern pharmaceutical analysis (Vol. 3), Academic press, 2001
- [11] L. L. Simon, H. Pataki, G. Marosi, F. Meemken, K. Hungerbühler, A. Baiker,... M. S. Chiu, Assessment of recent process analytical technology (PAT) trends: a multi author review, Organic Process Research & Development, 2015, 19(1), 3-62
- [12] N. Bolf, H. Dorić, Procesna analitička tehnologija, Kem. Ind. 67 (9-10) (2018) 466–469.
- [13] X. Y. Lawrence, R. A. Lionberger, A. S. Raw, R. D'Costa, H. Wu, A. S. Hussain, Applications of process analytical technology to crystallization processes, Advanced drug delivery reviews, 2004, 56(3), str. 349-369.
- [14] F. Zhang K. Du, L. Guo, Q. Xu, B. Shan, Comparative Study of Preprocessing on an ATR-FTIR Calibration Model for In Situ Monitoring of Solution Concentration in Cooling Crystallization, Chemical Engineering & Technology, 2021, 44(12), 2279-2289
- [15] A. Subramanian, L. Rodriguez-Saona, Fourier transform infrared (FTIR) spectroscopy, Infrared spectroscopy for food quality analysis and control, 2009, str. 145-178.
- [16] [https://en.wikipedia.org/wiki/Attenuated\\_total\\_reflectance](https://en.wikipedia.org/wiki/Attenuated_total_reflectance) (1.6.2023.)

- [17] R. Singh, R., Z. Yuan, Z. (Eds.), Process Systems Engineering for Pharmaceutical Manufacturing, 2018, Elsevier
- [18] F. Zhang, T. Liu, X. Z. Wang, J. Liu, X. Jiang, Comparative study on ATR-FTIR calibration models for monitoring solution concentration in cooling crystallization, Journal of Crystal Growth, 2017, 459, 50-55.
- [19] N. Kumar, A. Bansal, G. S. Sarma, R. K. Rawal, Chemometrics tools used in analytical chemistry: An overview, 2014, Talanta, 123, 186-199.
- [20] <https://en.wikipedia.org/wiki/Chemometrics> (5.6.2023)
- [21] Š. Ukić, Imenje i nazivlje: Nazivlje u kemometriji/kemometrici?, Kemija u industriji: Časopis kemičara i kemijskih inženjera Hrvatske, 2016, 65(3-4), 181-182.
- [22] <https://builtin.com/artificial-intelligence> (16.8.2023)
- [23] N. Bolf, Osvježimo znanje: Strojno učenje, Kemija u industriji: Časopis kemičara i kemijskih inženjera Hrvatske, 2021, 70.9-10: 591-593
- [24] <https://www.mathworks.com/solutions/machine-learning.html> (16.8.2023.)
- [25] N. Bolf, Strojno učenje, Metode umjetne inteligencije, Predavanje, FKIT, Zagreb, 2021-2022
- [26] <https://www.mathworks.com/help/matlab/preprocessing-data.html> (16.8.2023)
- [27] T. Herceg, Razvoj kalibracijskih modela za aktivne farmaceutske tvari, diplomski rad, FKIT, Sveučilište u Zagrebu, 2022
- [28] I. Jolliffe, Principal component analysis Encyclopedia of statistics in behavioral science, 2005, str.
- [29] H. Abdi, L. J. Williams, Principal component analysis, Wiley interdisciplinary reviews: computational statistics, 2(4), 2010, str. 433-459
- [30] N. Požar, Metode multivarijatne regresije za kemometrijsku analizu spektara u bliskom infracrvenom području, diplomski rad, Sveučilište u Rijeci, Odjel za biotehnologiju, 2019
- [31] H. Abdi, Partial least square regression (PLS regression), Encyclopedia for research methods for the social sciences, 2003, 6.4: 792-795
- [32] V. Vincenzo Esposito, G. Russolollo, Partial least squares algorithms and methods, Wiley Interdisciplinary Reviews: Computational Statistics, 2013, 5.1: 1-19.
- [33] T. Sente, Metoda parcijalnih najmanjih kvadrata: Regresijski model, diplomski rad, Sveučilište u Zagrebu, Prirodoslovno-matematički fakultet, 2016
- [34] B. D. Bašić, M. Čupić, J. Šnajder. Umjetne neuronske mreže, Zagreb: Fakultet elektrotehnike i računarstva, 2008.

- [35] N. Bolf, Ž. Ujević Andrijić, Umjetne neuronske mreže, Metode umjetne inteligencije u kemijskom inženjerstvu, Predavanje, FKIT, Zagreb, 2021-2022
- [36] R. Leppich, Pre-training of Deep Transformer Encoders for Time Series Representation Models, 2021.
- [37] E. Cetinić, Primjena različitih metoda strojnog učenja u problemu klasifikacije slikarskih djela prema autoru, Sveučilište u Zagrebu–fakultet elektrotehnike i računarstva, Zagreb, 2013.
- [38] Ž. Ujević Andrijić, Osvježimo znanje: Umjetne neuronske mreže, Kemija u industriji: Časopis kemičara i kemijskih inženjera Hrvatske, 2019., 68(5-6), 219-220.
- [39] M. Zekić-Sušac, Neuronske mreže, Sveučilište JJ Strossmayera u Osijeku, Ekonomski, 2017.