

# Primjena procesne analitičke tehnologije za praćenje procesa kristalizacije

---

**Klier, Monika**

**Master's thesis / Diplomski rad**

**2022**

*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:149:813935>

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

*Download date / Datum preuzimanja:* **2025-03-11**



*Repository / Repozitorij:*

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



SVEUČILIŠTE U ZAGREBU  
FAKULTET KEMIJSKOG INŽENJERSTVA I TEHNOLOGIJE  
SVEUČILIŠNI DIPLOMSKI STUDIJ

Monika Klier

DIPLOMSKI RAD

Zagreb, rujan 2022.

SVEUČILIŠTE U ZAGREBU  
FAKULTET KEMIJSKOG INŽENJERSTVA I TEHNOLOGIJE  
SVEUČILIŠNI DIPLOMSKI STUDIJ

Monika Klier

PRIMJENA PROCESNE ANALITIČKE TEHNOLOGIJE ZA PRAĆENJE PROCESA  
KRISTALIZACIJE

DIPLOMSKI RAD

Mentor: Prof. dr. sc. Nenad Bolf

Članovi ispitnog povjerenstva:

- 1) Prof. dr. sc. Nenad Bolf
- 2) Izv. prof. dr. sc. Krunoslav Žižek
- 3) Izv. prof. dr. sc. Igor Dejanović

Zagreb, rujan 2022.

## ***Zahvala***

Želim se zahvaliti mentoru prof. dr. sc. Nenadu Bolfu na vodstvu i pruženoj pomoći tijekom izrade ovog rada te na pruženoj prilici izrade ovog rada u sklopu projekta Napredno vođenje procesa kristalizacije - CrystAPC.

Zahvaljujem se dr. sc. Damiru Šahniću koji je vodio ovaj projekt sa strane Pliva/Teva TAPI R&D-a i na njegovoj stručnoj pomoći, strpljenju i prenesenom znanju tijekom provedbe projekta.

Zahvaljujem se doktorandici Matei Gavran na svoj pruženoj pomoći i savjetima te strpljenju tijekom provedbe eksperimenata i tijekom izrade rada.

Također se zahvaljujem asistentu Josipu Sacheru, asistentu Nikoli Rimcu i doc. dr. sc. Željki Ujević Andrijić na pomoći i savjetima.

Hvala mojoj obitelji na podršci tijekom studiranja te na njihovoj bezuvjetnoj ljubavi.

Hvala prijateljima na zajednički provedenom vremenu i lijepim trenucima tijekom studiranja.

## SAŽETAK RADA

Ovaj rad istražuje primjenu procesne analitičke tehnologije u svrhu praćenja procesa kristalizacije. Za predviđanje koncentracije otopine primjenjuje se Ramanova *in-line* spektroskopija pri čemu se na temelju trenutne informacije o koncentraciji otopine kontinuirano prati proces kristalizacije.

Eksperimentalno se određivala koncentracija djelatne tvari ceritiniba u sustavu otapala 90 % aceton - 10 % voda. Cilj je bio izraditi kalibracijski model za kontinuirani proračun koncentracije ceritiniba tijekom kristalizacije.

Za izradu kalibracijskog modela primjenile su se parcijalna regresija metodom najmanjih kvadrata i unaprijedna neuronska mreža. Kriteriji za ocjenu modela bili su kvadrat koeficijenta korelacije,  $R^2$  i korijen srednje kvadratne pogreške, RMSE.

Na temelju kriterija modeli za predviđanje koncentracije otopine i gustoće suspenzije razvijeni umjetnom neuronskom mrežom pokazali su se boljim od modela razvijenih parcijalnom regresijom metodom najmanjih kvadrata. Ovako razvijeni modeli mogu se primijeniti za procjenu koncentracije otopine u stvarnom vremenu.

### *Ključne riječi:*

Kristalizacija, djelatna tvar, procesna analitička tehnologija, umjetna neuronska mreža, parcijalna regresija metodom najmanjih kvadrata, kontinuirano praćenje procesa

## ABSTRACT

This thesis explores the application of Process Analytical Technology (PAT) to crystallization process monitoring. Raman in-line spectroscopy was utilized for prediction of solution concentration and slurry density. Continuous crystallization process monitoring is based on real-time calculation of the solution concentration.

The concentration of an Active Pharmaceutical Ingredient (API), Ceritinib was determined experimentally in a solution mixture of 90 % acetone and 10 % water. The goal was to build a calibration model for prediction of Ceritinib concentration during the crystallization process.

Methods used to build the calibration models include Partial Least Square Regression (PLSR) and an Artificial Feed Forward Neural Network (FFNN). Criteria for model assessment were the square of the Correlation Coefficient,  $R^2$  and the Root Mean Square Percentage Error, RMSPE.

The model built with an Artificial Feed Forward Neural Network (FFNN) proved better performance than the model built by Partial Least Square Regression (PLSR) for solution concentration and slurry density prediction. The models developed in this way can be applied to estimate the solution concentration in real time.

*Key words:*

crystallization, active pharmaceutical ingredient, process analytical technology, artificial neural network, Partial Least Square Regression, continuous process monitoring

# SADRŽAJ

Uvod.....	1
Teorijski dio .....	2
Kristalizacija .....	2
Procesna analitička tehnologija.....	11
Ramanova spektroskopija .....	13
Kemometrija .....	16
Viševeličinski sustavi.....	17
Analiza glavnih komponenata.....	18
Parcijalna regresija metodom najmanjih kvadrata .....	22
Umjetne neuronske mreže.....	25
Eksperimentalni dio .....	28
Djelatna tvar .....	28
Aparatura.....	29
Opis eksperimenta.....	31
Razvoj kalibracijskog modela.....	33
Rezultati i rasprava .....	37
Određivanje optimalnog sustava otapala .....	37
Izrada kalibracijskog modela .....	39
PLSR metoda .....	39
Unaprijedna neuronska mreža.....	47
Zaključak .....	55
Literatura.....	56

## Uvod

U procesu kristalizacije javlja se promjena faze iz kapljevitog u čvrsto čime se daljnom filtracijom i sušenjem dobivaju suhi kristali. Kristalizacija iz otopine ima veliku važnost u farmaceutskoj proizvodnji. Kristalna struktura najvažnije je svojstvo krute tvari koje diktira njena fizikalno-kemijska svojstva. Tvar koja kristalizira u više kristalnih oblika pokazuje polimorfizam.

Jedna od najbitnijih tvari koja se proizvodi kristalizacijom je djelatna tvar (eng. *Active Pharmaceutical Ingredient, API*). U proizvodnji djelatnih tvari važno je postići točno određenu polimorfiju budući da se fizikalno-kemijska svojstva razlikuju između različitih polimorfa.

Procesna analitička tehnologija (eng. *Process Analytical Technology, PAT*) primjenjuje se pri projektiranju, analizi i vođenju procesa s ciljem osiguranja visoke kvalitete konačnog proizvoda. Američka agencija za hranu i lijekove (eng. *The Food and Drug Administration, FDA*) izdala je dokument sa smjernicama vezanim za procesnu analitičku tehnologiju u svrhu inovativne farmaceutske proizvodnje i osiguranje kvalitete. Naglasak je na istraživanju, razvoju i proizvodnji koja nije samo empirijska nego se temelji na fenomenološkom razumijevanju procesa.

Razvoj procesne analitičke tehnologije bio je intenzivan u proteklom desetljeću. U sljedećim godinama predviđa se znatan rast radi sve veća potražnje u svim u segmentima kemijske, farmaceutske i biokemijske industrije, od istraživanja i razvoja do proizvodnje.

Noviji napredak na području procesne analitičke kemije omogućio je različite načine mjerenja. Tako je za veću djelotvornost u procesima kristalizacije potrebno razviti senzore za mjerenje ključnih procesnih varijabli kao što su prezasićenje, veličina čestica, oblik i polimorfija produkata. Razvoj kalibracijskih modela za predviđanje koncentracije otopine omogućava kontinuirano praćenje i vođenje procesa, a time je moguće mjeriti ključne procesne varijabli kao što je prezasićenje, jedne od ključnih tijekom procesa kristalizacije.

U ovom radu istraživao se utjecaj temperature, koncentracije djelatne tvari i gustoće suspenzije na Ramanov spektar. Iz izmjerenih podataka izveden je kalibracijski model za predviđanje koncentracije djelatne tvari ceritiniba.



# Teorijski dio

## Kristalizacija

Kristalizacija je separacijska tehnika i tehnika pročišćavanja koja služi za dobivanje širokog raspona različitih materijala. Kristalizacija se može definirati kao promjena faze tijekom koje se kristali formiraju iz otopine, taline ili pare. Otopina je homogena smjesa dviju ili više komponenata. Prilikom procesa kristalizacije, prije pojave kristala, homogenu smjesu tvore otapalo i otopljena tvar.[1]

Kristalizacija iz otopina je najbitnija u farmaceutskoj industriji. Da bi se završila priprava čistih suhih kristala potrebno ih je odvojiti od kapljevite faze. Ovaj postupak najčešće se provodi centrifugiranjem ili filtracijom te se nakon toga provodi sušenje kristala.[2]

Kristali su omeđeni glatkim plohama koje se na rubovima spajaju pod određenim kutovima koji su karakteristični za tvar koja kristalizira. Pravilan oblik kristala dolazi od njegove uređenosti, tj. ponavljajućih strukturnih jedinica, atoma ili molekula od kojih se sastoji unutrašnjost kristala. Ovakva struktura ponavljajućih jedinica razlikuje kristale od plinova i kapljevina, kao i od krutina u amorfnom stanju i uzrok je fizikalno-kemijskih svojstava kristala. Ova svojstva se mogu podešavati tijekom procesa kristalizacije.[3]

Još jedna bitna posljedica pravilnog slaganja molekula je pročišćavanje tijekom kristalizacije. U kristalnu rešetku ugrađuje se samo jedna vrsta molekula, dok se molekule kao što su nečistoće ne ugrađuju ili se ugrađuju minimalno. Do toga dolazi zbog geometrijski i energetski povoljnijih razloga.[3]

Mnogi organski i anorganski spojevi mogu postojati u različitim čvrstim formama. One mogu biti amorfne, tj. nesređene ili kristalne, tj. sređene strukture. Polimorfni su kristali iste molekulske formule čiji su atomi različito posloženi u prostoru. Hidrati su kristali u kojima je u kristalnu rešetku ugrađena molekula vode. Solvati su kristali u kojima su molekule otapala ugrađene u kristalnu rešetku u stehiometrijskom ili nestehiometrijskom omjeru, nekad se nazivaju i pseudopolimorfima.[4] Farmaceutski kokristali definirani su kao višekomponentni kristali u kojima je barem jedna komponenta molekularna i čvrsta na sobnoj temperaturi (koformer) i tvori supramolekularni sinton s molekularnom ili ionskom djelatnom tvari (eng. *Active Pharmaceutical Ingredient, API*).[5] Solvati i ko-kristali mogu postojati i u različitim polimorfnim formama. Kada je spoj kisel ili bazičan moguće je nastajanje soli s odgovarajućom bazom ili kiselinom.[4] Soli su, prema tome, ionske krutine nastale privlačenjem suprotno nabijenih iona.[5] Soli također mogu kristalizirati u različitim polimorfnim formama ili kao solvati(slika 1).[4]



**Slika 1.** Shematski prikaz različitih formi krutine[4]

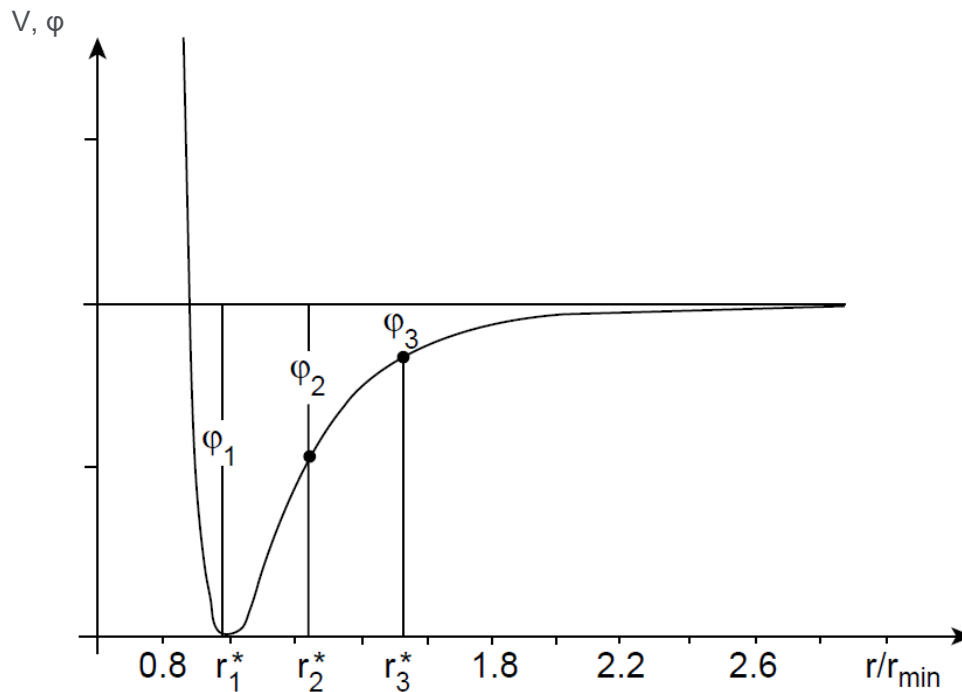
Razvojem molekularnih spojeva za napredne materijale i farmaceutike potreban je svestran pristup kontroli i karakterizaciji svojstava krutine. Za postizanje toga cilja potrebno je razumijevati kristalizaciju tih spojeva. Proizvodnja kristalnih i čvrstih tvari ima veliku primjenu što povećava interes za proces kristalizacije, otkrića novih polimorfnih formi, kontrolu polimorfizma, za formulacije i uvećanje, kao i za optimizaciju procesa kristalizacije u šaržnom i kontinuiranom načinu rada.[6]

Uvjeti tijekom procesa kristalizacije potrebni za određenu čistoću, iskorištenje i polimorfnu formu kristala najčešće se određuju eksperimentalno.[2]

Kristalizacija je fazna transformacija tijekom koje promjena slobodne Gibbsove energije mora biti manja od nule što je opisano jednačbom (1.1)

$$\Delta_{tr}G = \Delta_{tr}H - \Delta_{tr}S \quad (1.1)$$

Kristal će biti stabiliziran smanjenjem člana  $\Delta_{tr}H$ , koji je određen interakcijama između jedinica koje grade kristal. Sile koje djeluju mogu biti van der Waalove ili elektrostatke sile. U molekularnim kristalima važnu ulogu imaju i vodikove veze. Van der Waalove sile djeluju na malim udaljenostima, promjena potencijala opisana je jednačbom  $\varphi = \left(\frac{A}{r^{12}}\right) - \left(\frac{B}{r^6}\right)$ , gdje je  $r$  udaljenost između jedinki. Potencijal elektrostatskih sila mijenja se prema izrazu  $\varphi = \frac{C}{r}$ . [3]



**Slika 2.** Promjena potencijala s udaljenosti za van der Waalsovo privlačenje[3]

Promjena potencijala između 2 jedinice prikazana je na slici 2. S obzirom na brz pad potencijala s udaljenošću, energiju određuju interakcije s najbližom susjednom jedinicom. Kao posljedica toga, dozvoljene su samo određene simetrije u građevnim jedinicama kristala.[3]

Proces kristalizacije moguć je jedino u prezasićenom stanju otopine kada je koncentracija  $c$  veća od ravnotežne koncentracije,  $c^*$ . Zasićena otopina ima ravnotežnu koncentraciju,  $c^*$ , koja je u termodinamičkoj ravnoteži s čvrstom fazom na određenoj temperaturi. Brzina procesa kristalizacije ovisi o stupnju prezasićenosti. Prezasićenost,  $\Delta c$ , se izražava kao razlika u koncentraciji otopine,  $c$ , i ravnotežnoj koncentraciji,  $c^*$

$$\Delta C = C - C^* \text{ ili } \Delta c = c - c^* \quad (1.2)$$

$$c = C * \bar{M} \left[ \frac{\text{kmol kg}}{\text{m}^3 \text{ kmol}} \right] \quad (1.3)$$

ili kao relativna prezasićenost

$$S = \frac{C}{C^*} = \frac{c}{c^*} \text{ ili } \sigma = S - 1 \quad (1.4)$$

Razlika u kemijskom potencijalu fluida i kristala je kinetička pokretačka sila kristalizacije, te se može opisati odnosom između kemijskog potencijala,  $\mu$ , standardnog potencijala,  $\mu_0$ , aktiviteta,  $a$ , i entalpije kristalizacije,  $\Delta H_{CL}$ [7]

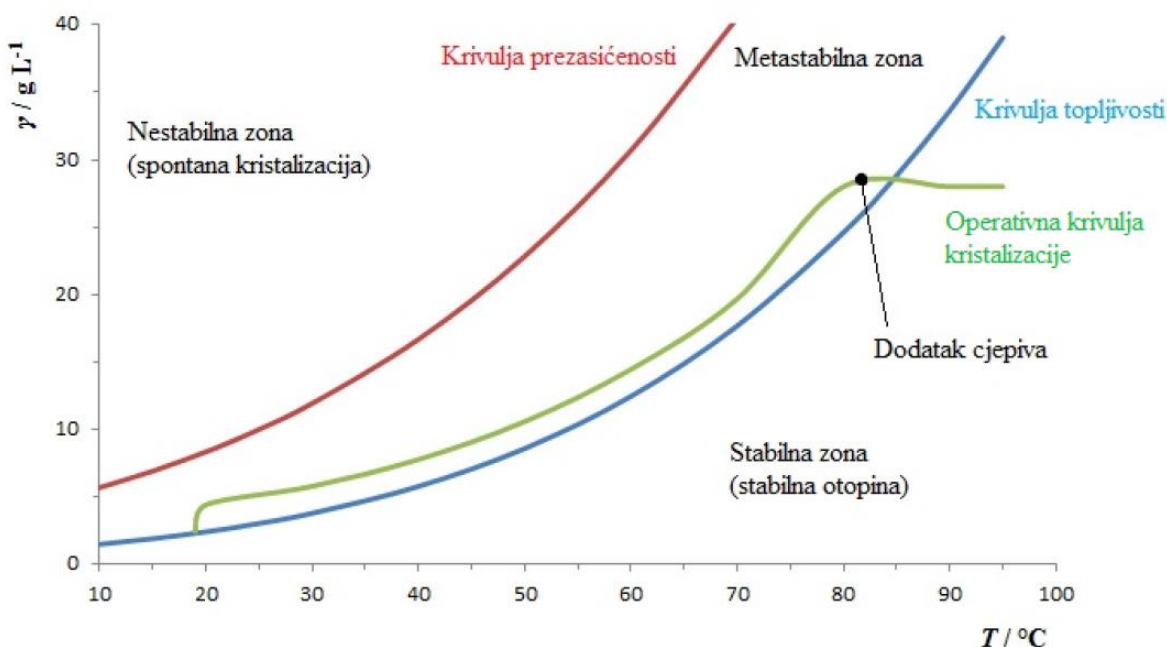
$$\mu = \mu_0 + RT \ln a \quad (1.5)$$

ili, za otopinu

$$\frac{\mu - \mu^*}{RT} = \ln \left( \frac{a}{a^*} \right) = \ln \left( \frac{\gamma c}{\gamma^* c^*} \right) = \ln S_a \quad (1.6)$$

gdje je  $\mu$  kemijski potencijal,  $c$  je koncentracija,  $\gamma$  je koeficijent aktiviteta,  $a$  je aktivitet,  $S$  je relativna prezasićenost,  $*$  predstavlja svojstvo sastava u zasićenom stanju. [7,1]

Postoje tri dobro definirana područja u kristalizaciji iz otopine. Prvo je stabilna, nezasićena zona u kojoj kristalizacija nije moguća. Druga je, metastabilna zona omeđena krivuljom topljivosti i krivuljom prezasićenja gdje je rast kristala moguć uz dodatak cjepiva koje predstavlja mjesta heterogene nukleacije. Širina metastabilne zone ukazuje na stabilnost otopine, što je šira metastabilna zona, otopina je stabilnija. Proces kristalizacije preferirano se vodi u ovoj zoni jer se ovdje može dobro kontrolirati podešavanjem uvjeta u sustavu, a moguće je spriječiti dodatnu neželjenu nukleaciju. Iza krivulje prezasićenja sustav je labilan i u tom području dolazi do spontane nukleacije (slika 3) [8]



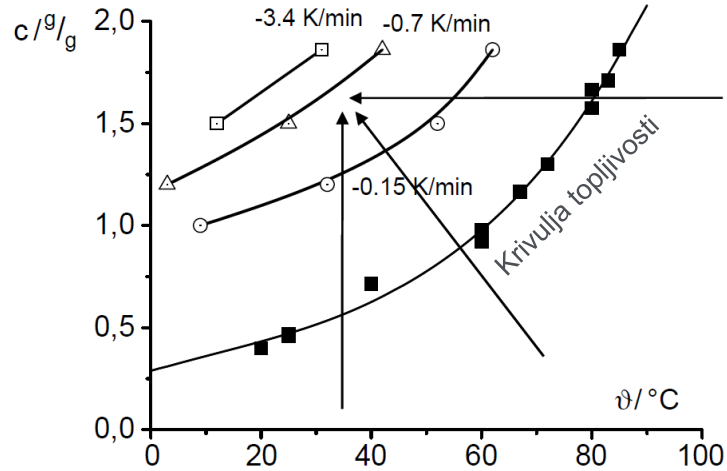
**Slika 3.** Ostwald-Miersov diagram na primjeru fumaratne soli API-a. Prikazana je i radna krivulja kristalizacije uz dodatak cjepiva i uz kontrolu stupnja prezasićenosti programiranim hlađenjem. [9]

Širina metastabilne zone (eng. *Metastable Zone Width, MSZW*) karakteristična je za skup uvjeta pri kojima se provodi kristalizacija.

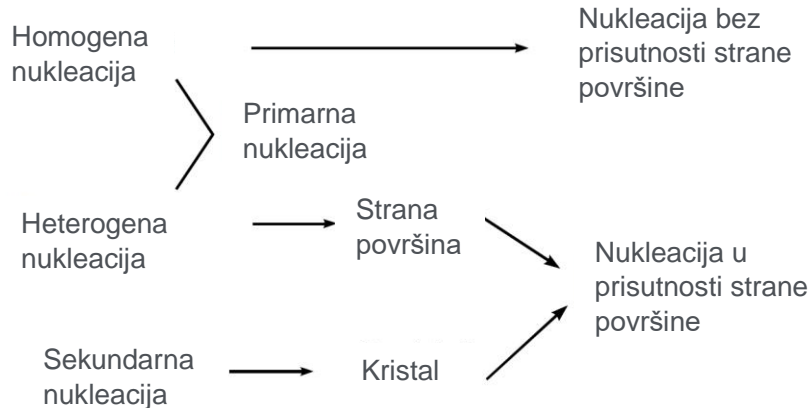
Najbrži pad prezasićenja događa se tijekom najsporijeg koraka u reakcijskom mehanizmu kristalizacije (eng. *rate determining step*). Tijekom kristalizacije u otopini najsporiji procesi mogu

biti prijenos tvari do mjesta rasta kristala, prevladavanje energetske barijere površine i odvođenje topline oslobođene kristalizacijom.[8]

Na širinu metastabilne zone najviše utječe brzina kojom se postiže prezasićenje otopine, tj. brzina hlađenja kod kristalizacije hlađenjem, brzina isparavanja otapala kod kristalizacije isparanjem i brzina dokapavanja antiotapala kod kristalizacije pomoću antiotapala. Širina metastabilne zone ovisi o brzini hlađenja, što je brže hlađenje, ona je veća. Jednako važan parametar je koncentracija aditiva ili nečistoća u otopini(slika 4). [3,10]



**Slika 4.** Širina metastabilne zone za sustav šećer-voda.[3]



**Slika 5.** Vrste nukleacije[3]

Homogena nukleacija događa se kada u sustavu postoji samo jedna faza, dok se heterogena nukleacija događa kada je u sustavu prisutna nečistoća, može se dogoditi i na stijenkama reaktora ili miješalu.[8]

Spontana nukleacija događa se kada u sustavu nema strane plohe na kojoj započinje kristalizacija, tada je ona homogena, a kada je prisutna strana površina kao što je npr. prašina, tada se naziva heterogena. Spontana nukleacija naziva se i primarna nukleacija. Sekundarna nukleacija događa se kada u sustavu već postoje kristali nastali primarnom nukleacijom(slika 5).[3] Spontana nukleacija događa se kada se dovoljan broj molekula niske kinetičke energije nađe na bliskoj udaljenosti te je tada sila privlačenja između molekula dovoljna da prevlada kinetičku energiju. Rast nukleusa događa se u vrlo kratkom vremenu u području visoke lokalne koncentracije.[8] Prilikom kristalizacije iz otopine nukleacija i rast kristala događaju se istovremeno u širokom rasponu temperatura.[2]Vrijeme od pojave prezasićene otopine i trenutka kada se detektiraju čestice krute faze zove se indukcijsko vrijeme nukleacije (eng. *induction period of nucleation*).[8]

Da bi se dobio izraz koji opisuje brzinu primarne nukleacije, potrebno je izračunati entalpiju formiranja nukleusa,  $\Delta_k G_r$ , radijusa,  $r$ . Dva člana doprinose entalpiji formiranja: entalpija kristalizacije i entalpija nužna za stvaranje nove površine (jednadžba 1.7)

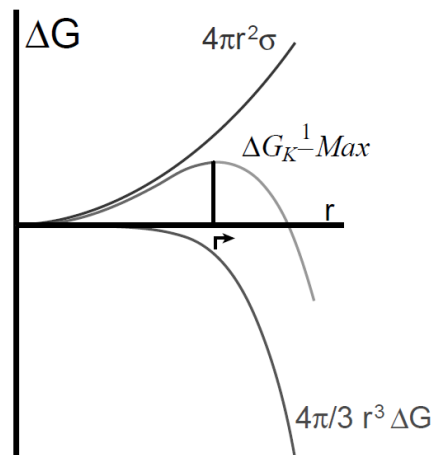
$$\Delta_k G_r = \frac{4\pi}{3} r^3 \Delta_v G - 4\pi r^2 \sigma \quad (1.7)$$

gdje prvi član opisuje promjenu faze (volumena), dok drugi član opisuje promjenu u površini.[3]

Na slici 6 vidljivo je da promjena na površini uzrokuje porast Gibbsove energije, a samim time predstavlja prepreku nastajanju nukleusa. Kada je nukleus vrlo mali, ova energetska prepreka onemogućava daljni rast kristala.[8]

Da bi nukleus rastao, promjena u slobodnoj Gibbsovoj energiji mora biti negativna. Na slici 6 vidljivo je da je to slučaj samo za  $r > r^*$ . Za nukleuse veličine  $r < r^*$ , dodavanje novog atoma ili molekule uzrokuje porast gibsove energije. Ako nukleus ima veličinu  $r = r^*$ , imat će vjerojatnost zadržavanja ili gubljenja atoma ili molekule, u ravnoteži je s prezasićenom otopinom i takvi nukleusi nazivaju se kritičnim nukleusima(slika 6).

Kada kritični nukleus ugradi još jednu jedinku, ugradnja te sljedeće jedinice je još povoljnija jer dovodi do stabilnosti nukleusa zbog smanjenja slobodne Gibbsove energije.[3]

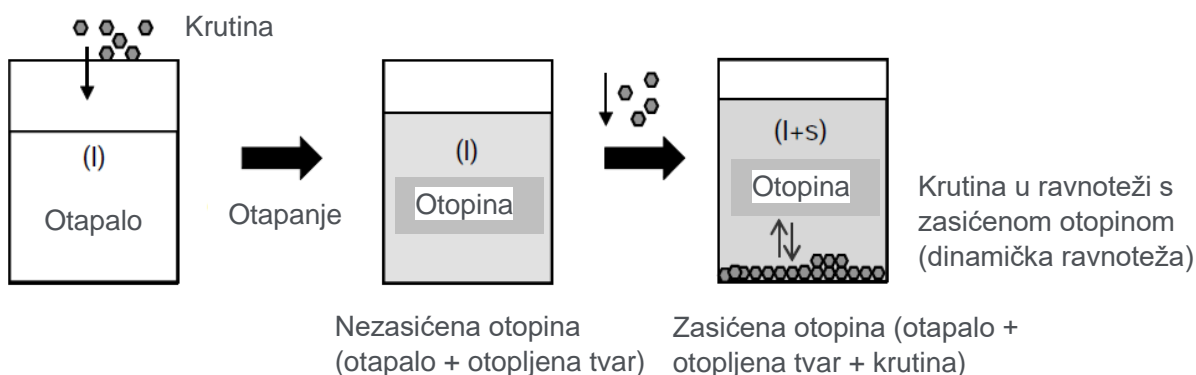


**Slika 6.** Promjena slobodne Gibbsove energije prilikom nukleacije[3]

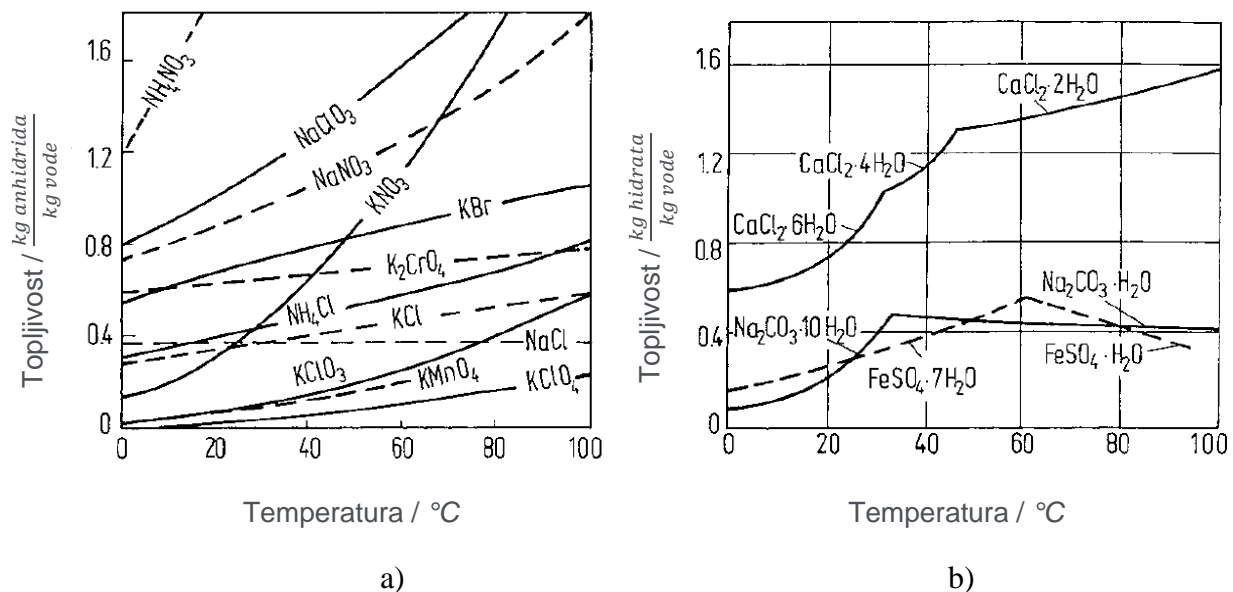
Nakon procesa nukleacija slijedi rast kristala koji uključuje rast kritičnih nukleusa, ponovno s uloženom energijom kako bi se savladala energetska barijera vezana za promjenu na površini.[8]

## Topljivost

Topljivost je maksimalna količina krutine koja se može otopiti u određenoj količini otapala na određenoj temperaturi i tlaku. Tada je otopina u ravnotežnom stanju s krutinom i smatra se zasićenom (slika 7). Ovisnost topljivosti o tlaku je vrlo mala. Krivulja topljivosti grafički je prikaz ovisnosti topljivosti o temperaturi za određeno otapalo ili sustav otapala. Topljivosti mogu značajno varirati, ovisno o tvari i sustavu i to do nekoliko reda veličina (slika 8). Topljivost kiselina, baza i njihovih soli značajno ovisi o pH otopine.[3]



**Slika 7.** Otapalo, nezasićena otopina i zasićena otopina[3]



**Slika 8.** a) Krivulje topljivosti anhidrida b) krivulje topljivosti hidrata[7]

## Polimorfizam

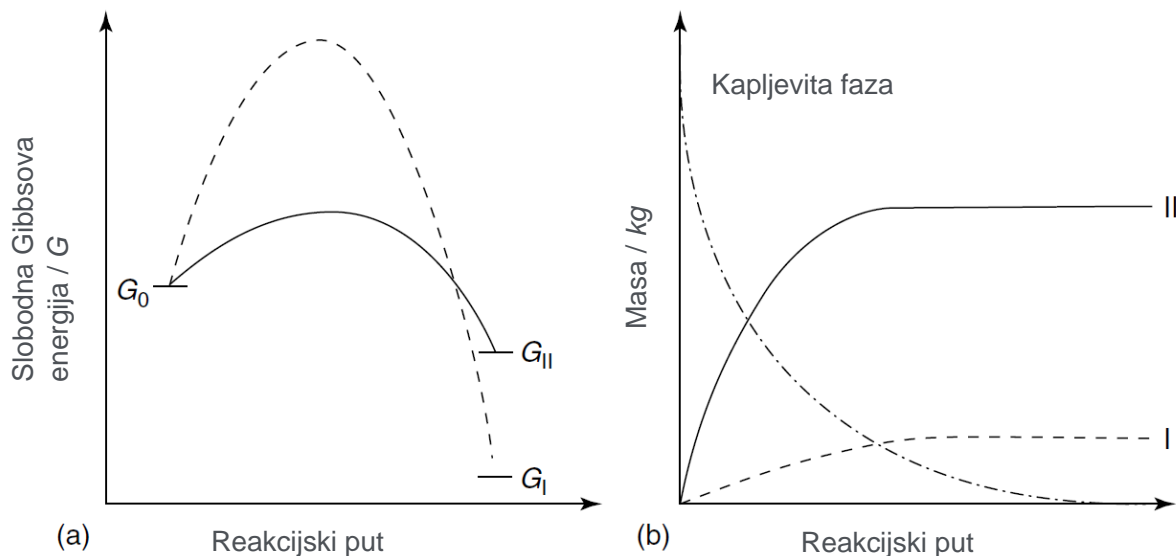
Određena tvar može kristalizirati u više kristalnih oblika, kaže se da pokazuje polimorfizam. Unutrašnja struktura, tj. raspored ponavljajućih jedinica u kristalu, razlikuju se od polimorfa do polimorfa, dok empirijska formula ostaje ista. [3,11]

Važnost održavanja uređenosti kristala prilikom procesa kristalizacije leži u različitim fizikalnim i kemijskim svojstvima različitih polimorfa koji uključuju:

- Kemijska i fizikalna svojstva: gustoću i indeks refrakcije, toplinsku i električnu vodljivost, higroskopnost, kemijski potencijal i slobodnu energiju, talište, toplinski kapacitet, tlak para, topljivost, toplinsku stabilnost, kemijsku i fotokemijsku reaktivnost;
- Kinetička svojstva: brzinu otapanja, kinetiku čvrstih tvari, stabilnost;
- Površinska svojstva: površinsku slobodnu energiju, ponašanje kristala, površinu, raspodjelu veličina kristala;
- Mehanička svojstva: tvrdoću, kompresibilnost, toplinsku širljivost [8,12]

Ako je tijekom procesa kristalizacije moguće formiranje dviju različitih polimorfnih formi (I i II), pri čemu je slobodna Gibbsova energija otopine na početku označena s  $G_0$ , a forma I je stabilnija od forme II ( $G_I < G_{II}$ ), moguće je formiranje manje stabilne forme zbog njezinog bržeg rasta. Na polimorfizam utječu nukleacija i rast kristalakoji su pod kinetičkim utjecajima. Time može nastati metastabilna faza koja teško prelazi u stabilniju zbog visoke energetske barijere.

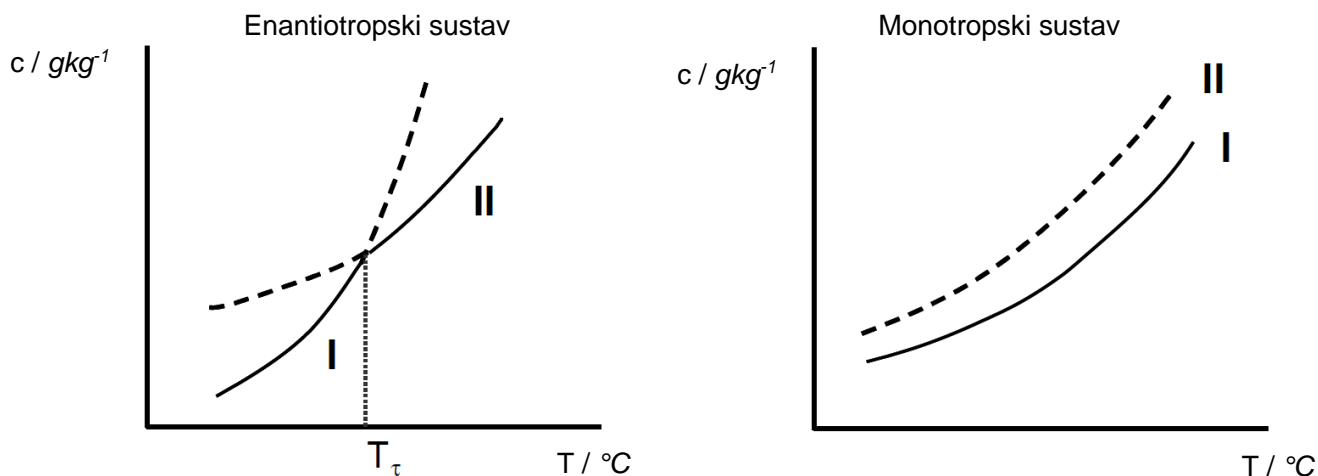
Na slici 9 prikazano je nastajanje manje stabilne polimorfne forme (II) ispred stabilnije polimorfne forme (I) zbog manje aktivacijske energije prilikom formacije forme II te istodobno postojanje obje forme u konačnoj smjesi kristala.[8]



**Slika 9.** Formiranje različitih polimorfnih formi[8]



Kristalizacija se odvija pod različitim utjecajima, koji se sastoje od termodinamičkih i kinetičkih faktora. Oni značajno ovise o uvjetima u kojima se kristalizacija odvija.[8] S obzirom na polimorfnu transformaciju, postoje dvije vrste sustava: monotropski i enantiotropski(slika 10). Transformacije polimornih formi su reverzibilne u enantiotropskom sustavu na temperature  $T_{\tau}$ . U monotropskom sustavu jedna forma je uvijek manje stabilna (metastabilna) u odnosu na drugu, krivulje topljivosti se ne sijeku. [5]



**Slika 10.** Krivulje topljivosti prilikom polimorfije [3]

### Ostwaldovo pravilo

U hidratima i solvatima broj molekula otapala po molekuli otopljene tvari može se razlikovati u polimorfima koji se razlikuju u stabilnosti.[7] Polimorf s najnižom slobodnom Gibbsovom energijom smatra se najstabilnijim, a forme s višom slobodnom Gibbsovom energijom od one koje posjeduje nastabilnija forma smatraju se metastabilnim.[11] U brzoj kristalizaciji ili precipitaciji isprva se može pojaviti metastabilna faza (bilo amorfna ili kristalna), zatim se taj intermedijar pretvara u drugi polimorf, hidrat ili solvat. Proces transformacije može biti vrlo spor, ponekad se može zaustaviti.

Različiti polimorfi tvari imaju različita fizikalna svojstva i topljivosti u otapalu na određenoj temperaturi. Tijekom kristalizacije ili precipitacije, prezasićenosti polimorfa razlikuju se, što rezultira da polimorf s najvećom prezasićenosti kristalizira prvi. Najčešće taj polimorf nije najstabilniji. Prema Ostwaldu, ako se stanje prezasićenja spontano ukloni, umjesto čvrste faze koja je u tim uvjetima najstabilnija, formirat će se manje stabilna faza. Kristalizaciju isprva kontrolira kinetika, što dovodi do formiranja manje stabilnog polimorfa.[7,13]

## Procesna analitička tehnologija

Procesna analitička tehnologija (eng. *Process Analytical Technology, PAT*) predstavlja sustav za projektiranje, praćenje i vođenje proizvodnih procesa kontinuiranim mjerenjem ključnih pokazatelja kvalitete i rada procesa. Mjeriti se mogu svojstva sirovina, međuprodukata i produkata, a najčešće se mjere ključne veličine koje utječu na djelotvornost procesa i kvalitetu produkta.[15] Američka agencija za hranu i lijekove (eng. *The Food and Drug Administration, FDA*), izdala je dokument sa smjernicama vezanim za procesnu analitičku tehnologiju.[2, 15]

Procesna analitička tehnologija znatno se razvija posljednjih godina. Razvoj laserske tehnologije, minijaturizacija optičkih i elektroničkih komponenata i primjena složenih kemometrijskih modela omogućili su primjenu novih serija instrumenata instaliranih u polju koji mogu veoma selektivno analizirati čak i raznovrsne smjese tvari.[16]

Primjena PAT mjerenja zahtijeva specijalizirane alate. Sustav za mjerenje i vođenje čine

- analitički instrumenti,
- alati za kemometrijsko modeliranje,
- softver za razvoj PAT metode, obradu podataka i integraciju u sustav,
- softver i algoritmi za vođenje procesa.

Za mjerenje se obično upotrebljavaju napredni analitički instrumenti koji pomoću svjetlosnog spektra kvalitativno i/ili kvantitativno karakteriziraju određeno svojstvo (npr. NIR, FTIR, Raman, masena i UV-Vis spektroskopija, FBRM)(tablica 1.).[14]

Iako kristalizacija ima značajnu ulogu u proizvodnji, donedavno se posvećivalo malo pažnje proučavanju fenomena koji utječu na process kristalizacije. Postalo je jasno da je razumijevanje i kontrola fenomena kristalizacije ključna, što je dovelo do razvoja naprednih tehnika za praćenje i kontrolu procesa kristalizacije.[8]

Da bi se unaprijedilo nadziranje postupka razvoja farmaceutskog proizvoda, potrebne su informacije o načinu na koji sustav reagira kada se promijene proizvodne varijable. [2] Tradicionalna *off-line* kontrola kvalitete oslanja na uzimanje uzoraka i transport na analizu u laboratorij za kvalitetu. Proći će nekoliko sati, a često i više, prije nego što rezultati postanu dostupni.[16] Poznavanje karakteristika različitih šarži i procesnih uvjeta pri kojima su one dobivene moguće je dobivanje proizvoda određene kvalitete. Posljednjih godina, analitičke metode i njihova primjena poboljšale su se do te mjere da su mjerenja moguća u stvarnom vremenu, čime omogućuje kontinuirano vođenje procesa. Procesi se mogu nadzirati *in-situ*, u stvarnom vremenu i/ili s povratnom vezom kako bi se osigurala tražena kvaliteta proizvoda tako se ne čeka se na laboratorijske analize, već se proizvodi (šarže) mogu distribuirati odmah po završetku proizvodnje. [2,16]

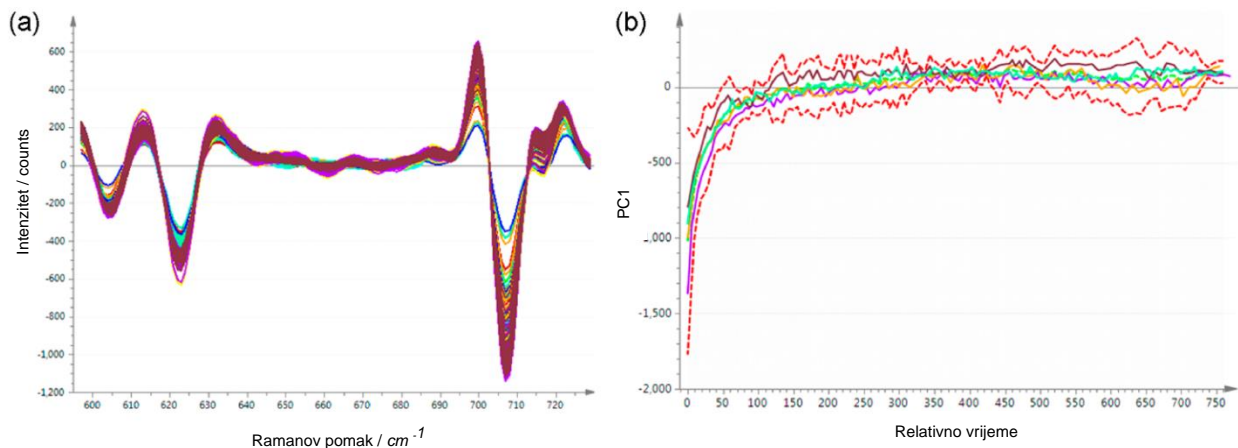
Upotreba PAT-a i strategija vođenja povratnom vezom omogućava dinamičko podešavanje radnih uvjeta, kako bi se kompenzirali poremećaji u sustavu i smanjile varijacije u proizvodu, čime se

poboljšava robustnost procesa.[15]

**Tablica 1.** Odabir PAT uređaja za mjerenje različitih veličina[16]

PLINOVI	KAPLJEVINE												KRUTINE
	boja	sastav	koncentracija	vodljivost	gustoća	dezinfekcija	metali nutrienti	otopljeni kisik	pH ORP	sumarni parametar	mutnoća	viskoznost	
	Photometer	Raman	Ultrasonic	Conductive	Resonance	Chlorine	Colorimetric	Amperomet	Class	Colorimetric	Absorption	Torque	
Spectrometer	IR-Absorption	Resonance	Inductive	Radiometric			Optical	Isfet		Scattering			
	sastav	koncentracija	gustoća	kisik	Napomene:							sastav	mokrina
	Raman	TDL-AS <sup>1)</sup>	MEMS <sup>2)</sup>	QF <sup>3)</sup>	<sup>1)</sup> TDL-AS = Tunable-diode laser absorption spectroscopy <sup>2)</sup> MEMS = Micro-mechanical electrical system <sup>3)</sup> QF = Quench fluorescence <sup>4)</sup> TDR = Time-domain reflectometry							Raman	TDR <sup>4)</sup>

Ramanova spektroskopija je *online* analitička tehnika za kvantitativno određivanje sastava smjese različitih polimorfnih formi.[15] Raman se također koristi za *in-situ* praćenje napredovanja različitih reakcija. Na slici 11a) prikazan je Ramanov spekter tijekom napredovanja kemijske reakcije, dok slika 11b) prikazuje trendove u šaržama za prvu glavnu komponentu (PC1).[17]



**Slika 11.** a) druga derivacija Ramanovog spektra b) prva glavna komponenta za različite šarže[17]

Mjerenje koncentracije otopine može se provesti uzorkovanjem i isparavanjem otapala te vaganjem suhog uzorka, no takav pristup je dug i neprikladan. Procesnom analitičkim tehnologijom mjere se svojstva sustava koja su u vezi s koncentracijom. Brzim razvojem spektroskopskih tehnika koriste se nedestruktivne metode mjerenja koncentracije otopine.

Izrada kalibracijskog modela primjenom Ramanovog spektra zahtjeva provedbu velikog broja eksperimenata, budući da je Ramanov spektralni odziv funkcija temperature, koncentracije otopljene tvari, gustoće suspenzije, vrste polimorfa i raspodjele veličine čestica (eng. *Particle Size*

*Distribution, PSD*).[15]

## Ramanova spektroskopija

Prilikom interakcije svjetla i materije fotoni se mogu absorbirati, reflektirati ili proći kroz nju. Ako je energija fotona jednaka energiji između osnovog i pobuđenog stanja molekule, foton će se vjerojatno apsorbirati, a molekula će preći u pobuđeno stanje. Ova promjena se mjeri apsorpcijskom spektroskopijom. No, moguće je da se foton i reflektira od tvari. Raspršeni fotoni se mogu pratiti sakupljanjem svjetla pod određenim kutom.

U Ramanovoj spektroskopiji svjetlost određene valne duljine pobuđuje molekulu. Mjeri se raspršeno zračenje pri čemu su određeni fotoni različite valne duljine od početnog zračenja. [26] [18]Energija fotona dana je Planckovom jednačinom.[19]

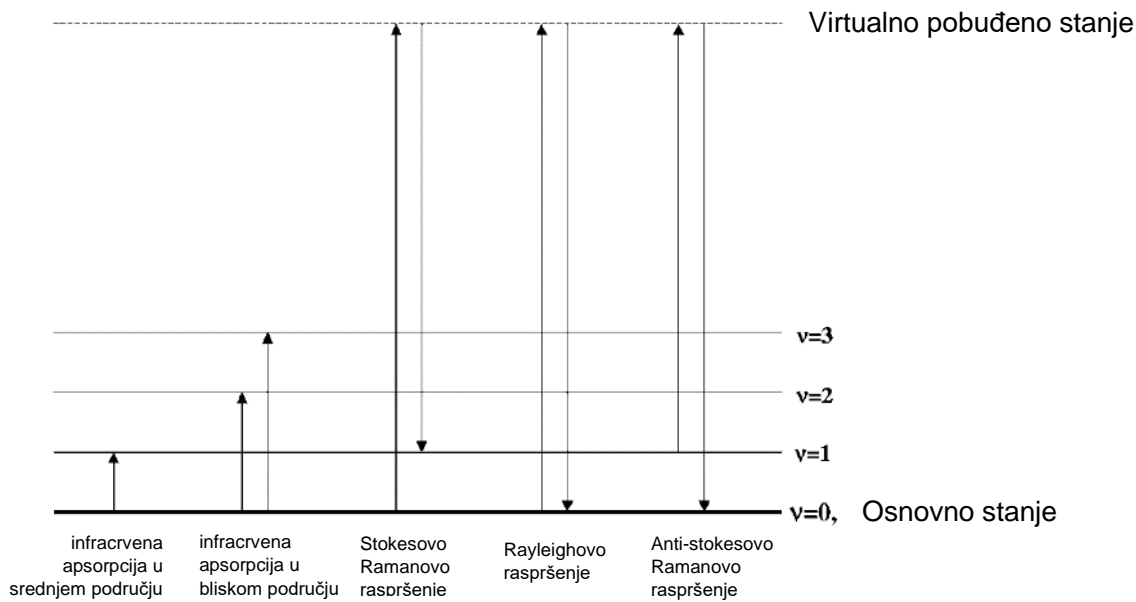
$$E = h\nu \quad (1.8)$$

Prijenos energije odvija se s 2 fotona:

- 1) prvi foton u interakciji s elektronskim oblakom molekule inducira dipolni moment i modificira vibracijsko energetska stanje elektrona koji prelazi u takozvano virtualno stanje
- 2) interakcija pod 1) izaziva emisiju drugog fotona[20]

Virtualno stanje nije stacionarno stanje sustava. Nije rješenje vremenski zavisne Schrödingerove jednačine te nema točno definiranu vrijednost energije. Jednostavnije rečeno, virtualno stanje je prijelazno stanje između početnog i konačnog stanja molekule.[21] Virtualno stanje se javlja kada laserska zraka dolazi u kontakt s molekulom i stvara virtualno stanje zbog promjene u polarizabilnosti.

Ramanova spektroskopija bazirana je na neelastičnom procesu raspršenja koji uključuje prijenos energije između ulaznog zračenja i molekule. Na malom dijelu ulaznog zračenja dolazi do pomaka frekvencije zračenja zbog interakcije s pobuđenim molekulama. Vibracijsko stanje molekula određuje koliko će se frekvencija zračenja pomaknuti i koliko će različitih pomaka postojati.[20]



**Slika 12.** Dijagram energetskih nivoa koji ukazuje na prijelaze između energetskih stanja za apsorpciju u infracrvenom području, Ramanov i Rayleighov proces raspršenja[4]

Prilikom Ramanovog raspršenja samo je jedan od  $10^6 - 10^8$  fotona je raspršen neelastično. Ostatak ovih fotona imaju istu valnu duljinu kao i ulazno zračenje.[20] Kod lasera velike snage često se javlja i fluorescencija.[18] Fluorescencija se može djelomično izbjeći mjereći laserom koji koristi svjetlost veće valne duljine. Intenzitet Ramanovog signala obrnuto je proporcionalan četvrtoj potenciji valne duljine. [4, 20]

Kada se promatra Ramanovo raspršenje, analizira se dipolni moment koji je induciran zbog električnog polja svjetlosti koja je u interakciji s molekulom

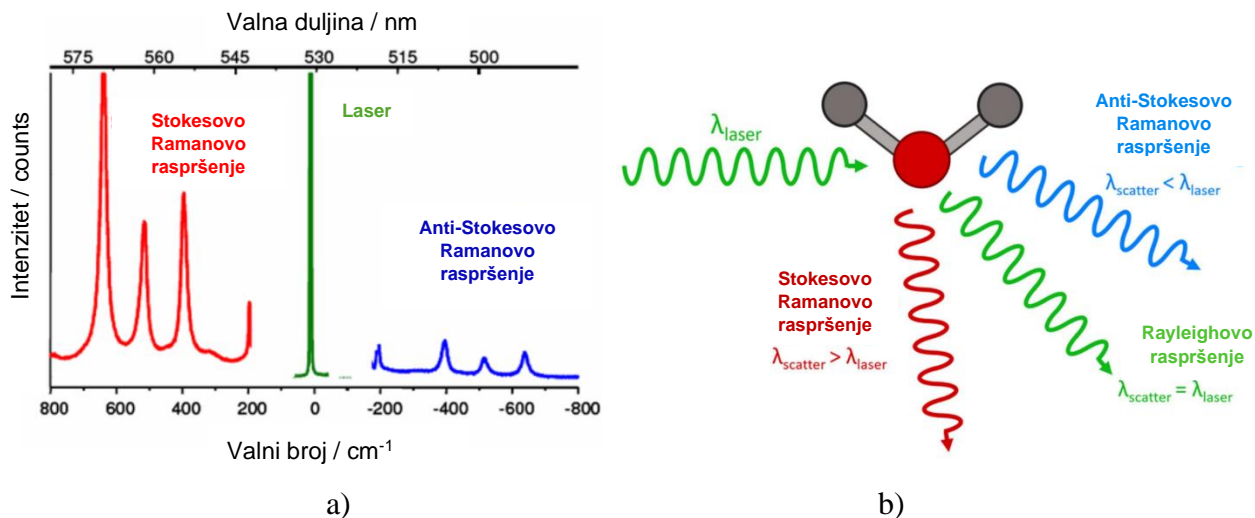
$$P = \alpha E \quad (1.9)$$

gdje je  $\alpha$  molekulska polarizibilnost.[22]

Da bi se izmjerilo Ramanovo zračenje, mora doći do promjene u polarizibilnosti molekule, tada se za nju kaže da je Raman aktivna. Kod infracrvene (IR) spektroskopije, dolazi do promjene u dipolnom momentu, tada je molekula IR aktivna.[4] Ove dvije tehnike su komplementarne, kada je signal jedne jak, druge je slab. Na primjer, istežanje karbonilne skupine,  $>C=O$  obično je popraćeno velikom promjenom u dipolnom momentu i time daje značajnu IR apsorpcijsku vrpcu, ali Ramanova vrpca je slabog intenziteta.

Ramanov spektar posljedica je neelastičnog raspršivanja monokromatske svjetlosti. Ako se monokromatskim svjetlom osvijetli uzorak, većina raspršene svjetlosti raspršena je elastično Rayleighovim raspršenjem u svim smjerovima, dok mali dio svjetlosti mijenja frekvenciju.[4] Kada je emitirani foton niže energije (veće valne duljine), od izvornog, radi se o Stokesovom raspršenju. Ako je više energije (manje valne duljine), govorimo o Anti-Stokesovom raspršenju. Raspršenje je prikazano na slici 13 pri čemu je valna duljina ulaznog laserskog zračenja označena s  $\lambda_{laser}$ , s  $\lambda_{scatter} > \lambda_{laser}$  je označeno Stokesovo raspršenje prilikom kojeg se energija prenosi na

molekulu, raspršeno zračenje ima veću valnu duljinu i manju energiju od ulaznog zračenja (eng. *Red-shifted Scattered Light*). S  $\lambda_{scatter} < \lambda_{laser}$  označeno je Anti-Stokesovo raspršenje, pri čemu se energija prenosi s molekule na ulazno zračenje, raspršeno zračenje sada ima manju valnu duljinu i veću energiju od ulaznog zračenja (eng. *Blue-shifted Scattered Light*). Za Rayleighovo raspršenje vrijedi  $\lambda_{scatter} = \lambda_{laser}$ . [19, 20]



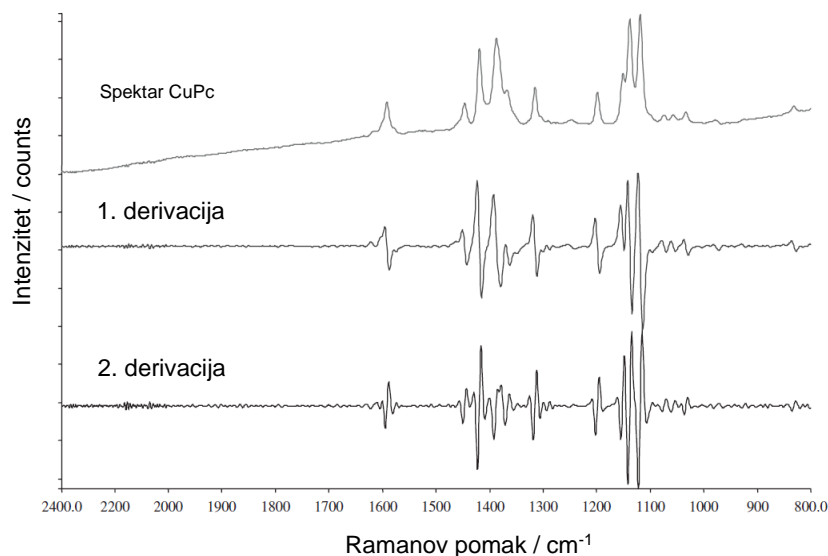
**Slika 13.** a) Stokesov and anti-Stokesov Ramanov spektar anataze (TiO<sub>2</sub>)[23] b) interakcija svjetla s molekulom[24]

Jedna od prednosti Ramana je njegova mala osjetljivost na molekule vode pa je lakše promatrati vodene sustave.

Glavne komponente Ramanovog spektrometra su: izvor monokromatske svjetlosti (laser), barijere između medija i instrumenta, filter koji uklanja Rayleighovo raspršenje, spektrograf koji razdvaja zračenje prema valnoj duljini, CCD detektor i sustava za prijenos informacija.[4, 20] Većina sonde konfigurirano je pod kutom od 180° koji se odnosi na uzorak i optike kolektora na uređaju. Pošto se pobudna i kolektorska optika u istom prostoru, Ramanova spektroskopija provodi se jednom sondom uronjenom u procesni medij.[20]

Najčešće je prikazan Stokesov spektar. Ovisno o literaturi, valni brojevi mogu biti prikazani uzlazno ili silazno na x osi. Svi pikovi Ramanovog spektra kreću od bazne linije (eng. *baseline*) i mjere se na linearnoj skali koja varira od instrumenta do instrumenta.

Postoji je puno programa za reduciranje ili uklanjanje nagiba iz dijelova spektra, što je korisno za uklanjanje fluorescencije kao što su Origin, The Unscrambler, itd. Ako se radi o kvantitativnom mjerenju, *background* nagib se uklanja derivacijom spektra. Na slici 14 vidljivo je da druga derivacija ima ravnu baznu liniju, ali identifikacija individualnih pikova može biti zahtjevna.[18]



**Slika 14.** Derivacije Ramanovog spektra[18]

## Kemometrija

Kemometrija je način obrade podataka uz pomoć raznih numeričkih postupaka u svrhu dobivanja korisnih informacija sadržanih u izmjenom skupu podataka. Cilj je iskoristiti informacije sadržane u podacima kako bi bolje razumjeli uzorke ili sustave u kojima su ti podaci izmjereni.[25] Kemometrija je interdisciplinarna grana znanosti koja uključuje viševeličinsku statistiku, matematičko modeliranje, računarstvo i analitičku kemiju.[26]

Pojam “kemometrija” prvi put se javlja 1970-ih godina. Naziv su prvi put uveli Svante Wold i Bruce R. Kowalski. Nova disciplina kemije bila je korisna prvenstveno analitičkim kemičarima, koji su razvijali nove metode kako bi zadovoljili rastuće potrebe znanstvenih istraživanja. Dva su razloga zašto se kemometrija tako brzo razvijala u to vrijeme. Prvi je velika količina podataka koja se prikupljala analitičkim instrumentima, a drugi je napredak u mikroelektronici u tom periodu. Od 1980-ih uvodi se u kurikulume diplomskih i poslijediplomskih studija na američkim i europskim sveučilištima.

Napredak u računarstvu i informacijskoj znanosti, statistici i primijenjenoj matematici omogućio je uvođenje novih elemenata u kemometriju. Neki od novih elemenata su umjetne neuronske mreže koje simuliraju prijenos signala u biološkim sustavima, genetički algoritmi i simulirano kaljenje (*eng. simulated annealing*).

Podaci sadrže kombinaciju korisne informacije i šuma (smetnji). Realni podaci sadrže veliku količinu šuma, a mogu biti u kompleksnom odnosu sa svojstvima uzoraka ili sustava. Kemometrijske tehnike analize čine i takve podatke iskoristivim.

Kemometrijske tehnike služe za

1. uklanjanje šuma iz podataka,
2. izvlačenje korisnih informacija iz podataka,
3. primjenu informacija za predikciju svojstava uzoraka.

Da bi navedeno bilo ostvareno, potrebno je zadovoljiti dva uvjeta

1. podaci moraju sadržavati informacije o uzorcima ili sustavu,
  2. funkcionalna ovisnost između informacija u podacima i svojstva uzoraka ili sustava.
- [25]

## Viševeličinski sustavi

Viševeličinski (multivarijatan) sustav je onaj u kojem se može napraviti više mjerenja na jednom uzorku, tj. mjerenje se provodi u više dimenzija. Tipičan primjer je spektroskopija gdje se mjere spektri na stotinama različitih valnih duljina u samo jednom uzorku.[27] Kada je moguće kalibrirati jednu komponentu u sustavu s jednim pikom, sustav se smatra univarijatan, no kada se na spektru očituje preklapanje pikova odziva više različitih tvari u sustavu, on se smatra multivarijatan.[28] Mnogo tradicionalnih pristupa su univarijatan, no na njima se propušta mnoštvo informacija.

Multivarijatan analiza centralni je dio kemometrije. Mnoge studije i mjerenja u laboratoriju uključuju mjerenje puno varijabli u jednom uzorku. Tako spektar mjeri na stotinama različitih valnih duljina. Tradicionalno su se koristile jedna ili dvije varijable, npr. visina pika na određenoj valnoj duljini, da bi se opisao uzorak. Ovdje se javlja problem odabira najboljih varijabli te su često u jednom piku pomiješane informacije koje dolaze od više komponenata iz uzorka.

Univarijatan sustav sastoji se od jedne varijable, podaci su zapisani u jednom stupcu (ili redu), dok su kod multivarijatanog sustava podaci zapisani u obliku matrice, pri čemu redovi predstavljaju broj izmjerenih uzoraka, a stupci valne brojeve ako se radi o spektroskopskim metodama. Ako se izmjeri intenzitet 1000 različitih valnih brojeva u 200 različitih uzoraka, dobiva se matrica  $200 \times 1000$  ( $m \times n$ ), gdje  $m$  označava broj redova, a  $n$  broj stupaca.[27]



## Analiza glavnih komponenata

Analiza glavnih komponenata (eng. *Principle Component Analysis, PCA*) smatra se „majkom svih metoda“ u multivarijantnoj analizi podataka. Cilj PCA je reducirati dimenzionalnost izmjerenih podataka. Može se smatrati kao metoda koja generira novi koordinatni sustav kojeg oblikuju glavne komponente (eng. *Principle Components, PCs*), koje su međusobno ortogonalne. U tom novom koordinatnom sustavu promatraju se samo glavne komponente koje sadrže najviše informacija.

Reduciranje dimenzija metodom PCA koristi se za

- vizualizaciju multivarijantnih podataka dijagramima raspršenja (eng. *Scatter Plot*),
- transformaciju visoko koreliranih  $x$  varijabli u manji broj nekoreliranih glavnih komponenata koje se zatim koriste u drugim metodama,
- separaciju relevantnih informacija (opisanih glavnim komponentama) od šuma,
- kombinaciju više varijabli koje opisuju kemijsko-tehnološki proces u jednu ili nekoliko “karakterističnih” varijabli[29]

Metoda PCA često se koristi kada su kada su broj glavnih komponenti ili njihova kvantitativna vrijednost u nekom odnosu s fizičkim procesom ili komponentom u sustavu. Također se koristi kao kvalitativna metoda za vizualizaciju i pojednostavljenje skupa podataka.

Svaka glavna komponenta opisana je s dvije informacije, faktorskim bodovima (eng. *scores*), i opterećenjima (eng. *loadings*).

PCA (slika 15) uključuje apstraktnu matematičku transformaciju originalne matrice  $X$  s eksperimentalnim podacima, koja ima oblik

$$X = TP' + E \quad (1.10)$$

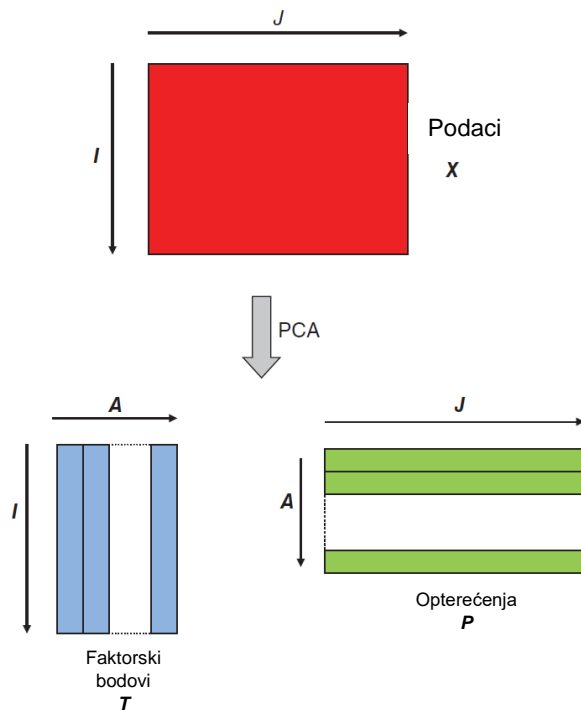
pri čemu su

$T$  matrica faktorskih bodova, broj redova jednak je broju u originalnoj matrici  $X$

$P$  matrica opterećenja, broj stupaca jednak je broju stupaca u originalnoj matrici  $X$

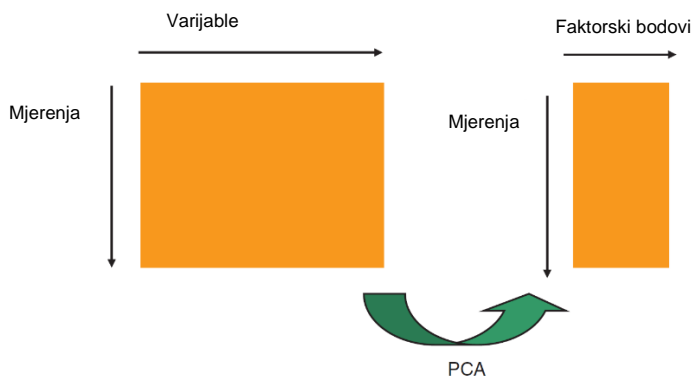
$E$  matrica pogrešaka, jednakih je dimenzija kao i originalna matrica  $X$

Svaka matrica faktorskih bodova sastoji se od stupčastih vektora, dok se svaka matrica opterećenja sastoji od rednih vektora. Ovi vektori se nazivaju svojstvenim vektorima (eng. *eigenvectors*).[27]



**Slika 15.** Analiza glavnih komponentata[21]

Primjenom PCA originalni broj varijabli reducira se na puno manji broj značajnih glavnih komponentata (slika 16). Novi, manji skup podataka lakše je interpretirati.[27]



**Slika 16.** PCA za reduciranje broja varijabli[27]

Pravac unutar prostora varijabli koji najbolje zadržava relativne udaljenosti između objekata je latentna varijabla koja ima maksimalnu varijancu. Ovaj pravac zove se prva glavna komponenta. Definiran je težinskim vektorom,  $p_1$  (slika 17)..

$$p_1 = p_1, p_2, \dots, p_m \quad (1.11)$$

gdje je  $m$  broj varijabli.

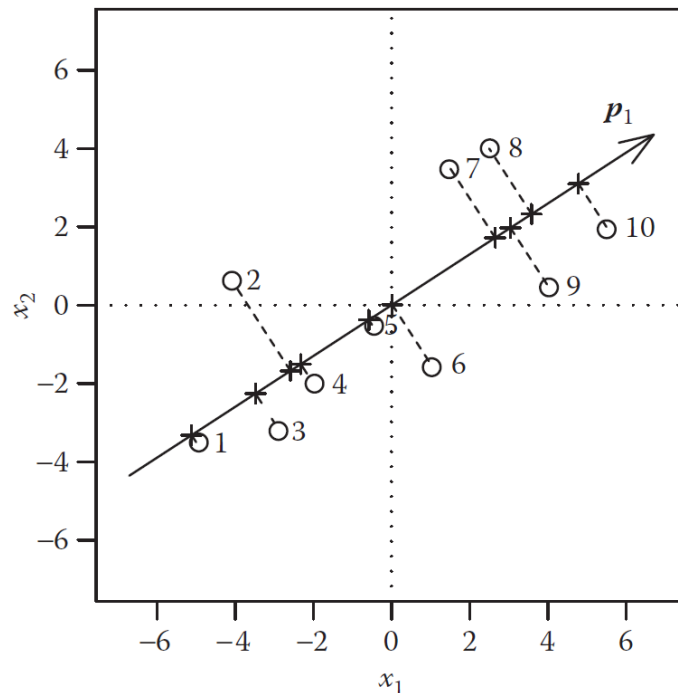
Faktorski bodovi linearne su kombinacije opterećenja i varijabli. Na primjer, za red  $i$ , definiran vektorom  $x_i$ , s elementima  $x_{i1}$  do  $x_{im}$ , faktorski bod  $t_{i1}$  za glavnu komponentu PC1 je

$$t_{i1} = x_{i1}p_1 + x_{i2}p_2 + \dots + x_{im}p_m = x_i^T p_1 \quad (1.12)$$

Za sve  $n$  redove u matrici  $X$ , vektor faktorskih bodova,  $t_1$  glavne komponente PC1 dobiva se

$$t_1 = X \cdot p_1 \quad (1.13)$$

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  $X$ . Svaka sljedeća glavna komponenta ortogonalna je na sve prethodne glavne komponente, i njezin smjer mora osigurati maksimum varijance.



**Slika 17.** Prva glavna komponenta (PC1) definirana je vektorom opterećenja  $p_1$ . Faktorski bodovi su ortogonalne projekcije podataka (točke 1-10) na vektor opterećenja[29]

Pošto su vektori opterećenja svih glavnih komponentata ortogonalni jedni na druge, kao što su i osi u originalnom koordinatnom sustavu, ova transformacija je rotacija koordinatnog sustava. Za ortogonalne vektore, skalarni produkt je nula, stoga za sve parove vektora opterećenja vrijedi

$$\mathbf{p}_j^T \cdot \mathbf{p}_k = 0 \quad j, k = 1, \dots, m \quad (1.14)$$

Varijance glavnih komponentata viših vrijednosti često su vrlo male ili nula, stoga se, ovisno o podacima, uzima nekoliko glavnih komponenti.

Svi vektori opterećenja zapisani su kao stupci u matrici opterećenja,  $\mathbf{P}$ , a svi vektori faktorskih bodova u matrici faktorskih bodova,  $\mathbf{T}$

$$\mathbf{T} = \mathbf{X} \cdot \mathbf{P} \quad (1.15)$$

Faktorski bodovi imaju matematičko svojstvo da su ortogonalni jedan na drugi. Pošto su često centrirani, par bilo kojih dvaju faktorskih bodova su nekorelirani, tj. korelacijski koeficijent im je nula. Ni jedna druga rotacija koordinatnog sustava nema ovakvo svojstvo.

$$\mathbf{t}_j^T \cdot \mathbf{t}_k = 0 \quad j, k = 1, \dots, m \quad (1.16)$$

Matrica  $\mathbf{X}$  može se rekonstruirati iz PCA matrice faktorskih bodova,  $\mathbf{T}$ . Uglavnom se koristi samo nekoliko glavnih komponentata. Ovo rezultira aproksimiranom  $\mathbf{X}$  matricom s smanjenim šumom,  $\mathbf{X}_{appr}$  (slika 18). Ako bi se koristile sve glavne komponente aproksimirana matrica bila bi jednaka originalnoj te bi matrica pogreške,  $\mathbf{E}$ , bila nula.[29]

$$\mathbf{X}_{appr} = \mathbf{T} \cdot \mathbf{P}^T \quad \mathbf{X} = \mathbf{T} \cdot \mathbf{P}^T + \mathbf{E} \quad \mathbf{E} = \mathbf{X} - \mathbf{X}_{appr} \quad (1.17)$$

The diagram illustrates the matrix dimensions for equation (1.17). It shows four matrices:  $\mathbf{P}^T$  (a row vector of size  $1 \times m$ ),  $\mathbf{T}$  (a matrix of size  $n \times a$ ),  $\mathbf{X}_{appr}$  (a matrix of size  $n \times m$ ),  $\mathbf{E}$  (a matrix of size  $n \times m$ ), and  $\mathbf{X}$  (a matrix of size  $n \times m$ ). The equation is represented as  $\mathbf{X}_{appr} + \mathbf{E} = \mathbf{X}$ .

**Slika 18.** Približna rekonstrukcija,  $\mathbf{X}_{appr}$ ,  $\mathbf{X}$  matrice iz PCA matrice faktorskih bodova  $\mathbf{T}$  i matrice opterećenja  $\mathbf{P}$  koristeći broj komponentata  $a$ ,  $\mathbf{E}$  je matrica pogrešaka[29]

## Parcijalna regresija metodom najmanjih kvadrata

Parcijalna regresija metodom najmanjih kvadrata (eng. *Partial Least Squares Regression, PLSR*) je metoda koja matricu podataka  $X$  stavlja u odnos s vektorom  $y$  ili matricom odziva  $Y$ . Originalnu PLS metodu razvio je 1975. statističar Herman Wold, dok je njegov sin Svante Wold uveo metodu u kemometriju.

Postoje 2 vrste PLS metode, ako se matrica podataka  $X$  stavlja u odnos s vektorom  $y$ , onda se naziva PLS1, ako se matricu podataka  $X$  stavlja u odnos s matricom odziva  $Y$  onda se naziva PLS2.

Jedan od zadataka je modelirati svojstvo  $y$  s jednom ili nekoliko varijabli  $x$ . Svojstvo se smatra karakteristikom sustava od velike važnosti, ali se često ne može odrediti direktno ili su troškovi visoki. Suprotno tome,  $x$  podaci lako su mjerljivi, ali nisu cilj istraživanja. Cilj je predvidjeti  $y$ , npr. koncentraciju, preko  $x$  podataka, npr. spektra.

Modeli koji imaju oblik  $y = f(x)$  ili  $y = f(x_1, x_2, \dots, x_m)$  mogu biti linearni ili nelinearni. Mogu biti formulirani kao relativno jednostavna jednadžba ili mogu biti u obliku algoritma, npr. u umjetnim neuronskim mrežama (eng. *Artificial Neural Networks, ANN*).

Regresija provedena PLS modelom koristi mali broj latentnih varijabli umjesto velikog skupa originalnih  $x$  varijabli.

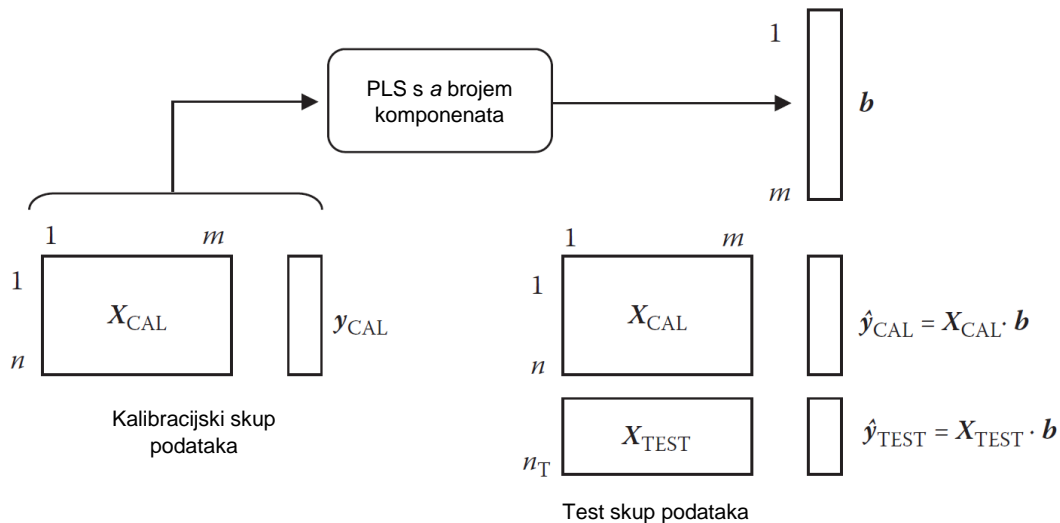
Tijek izrade modela PLS metodom uključuje transformaciju  $x$  varijabli u set linearnih latentnih varijabli koje se zatim koriste za regresiju s zavisnim  $y$  varijablama. PLS koristi komponente koje imaju vezu s vektorom  $y$  ili matricom  $Y$ .

Kriterij za latentne varijable u PLS metodi je maksimiziranje kovarijance između faktorskih bodova i podatka  $y$  (odnosno između faktorskih bodova u  $x$  prostoru i faktorskih bodova u  $y$  prostoru). Kovarijanca kombinira visoku varijancu  $X$  i visoku korelaciju sa svojstvom sustava koje se želi mjeriti.

Postupak regresije  $x$  varijabli s  $y$  varijablom:

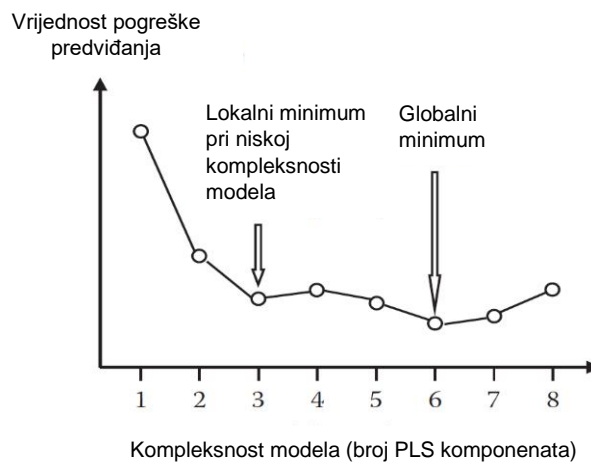
1. Prva PLS komponenta računa se kao latentna varijabla koja ima maksimalnu kovarijancu između faktorskih bodova i svojstva koje se modelira,  $y$ . Kovarijanca je kompromis između maksimalnog koeficijenta korelacije i maksimalne varijance;
2. Računanje matrice residuala,  $X_{RES}$  koja se dobiva oduzimanjem varijance prve PLS komponente od  $x$  varijabli u originalnoj  $X$  matrici;
3. Iz matrice reziduala računa se sljedeća komponenta, opet s maksimumom kovarijance između faktorskih bodova i vektora  $y$ . Treba uzeti u obzir da se parametri svake sljedeće komponente odnose na matricu residuala,  $X_{RES}$ , a ne na originalnu matricu  $X$ ;
4. Postupak se ponavlja sve dok više nema poboljšanja  $y$  svojstva.

Na slici 19 prikazana je metoda PLS za linearnu regresiju koja predviđa  $y$  iz varijabli  $x_1 \dots x_m$ , primjenjujući regresijske koeficijente  $b_1 \dots b_m$ . PLS model se izrađuje iz podataka za kalibraciju, dobiva se regresijski vektor  $b$ , koji se zatim primjenjuje na  $x$  varijable u skupu za kalibraciju, kao i u skupu za testiranje kako bi se predvidjela varijabla  $y$ . [29]



**Slika 19.** PLS metoda [29]

Kompleksnost modela, tj. broj latentnih varijabli određuje se križnom validacijom (eng. *cross validation, CV*). Često se uzima lokalni minimum kako bi se izbjeglo pretreniranje modela. (slika 20)



**Slika 20.** Određivanje optimalne kompleksnosti regresijskog modela. [29]

Kriterij za validaciju modela je korijen srednje kvadratne pogreške za križnu validaciju (eng. *Root Mean Squared Error for Cross Validation, RMSECV*).

$$RMSECV = \sqrt{\frac{1}{z} \sum_{i=1}^z (y_i - \hat{y}_i)^2} \quad (1.18)$$

gdje je  $y_i$  izmjerena (prava) vrijednost,  $\hat{y}_i$  je predviđena vrijednost, a  $z$  je broj predviđanja.

RMSECV se izračunava iz pogrešaka predviđanja provedenih u križnoj validaciji, npr. prilikom optimiziranja broja komponenata PLS modela.[29]

Cilj PLS regresije je naći linearnu vezu za

$$Y = XB + E \quad (1.19)$$

gdje je  $X$  matrica varijabli prediktora

$Y$  matrica varijabli odziva

$B$  matrica regresijskih koeficijenta

$E$  matrica pogrešaka

Ako se radi o PLS1 regresiji, imamo samo jednu varijablu odziva, vektor  $y$ . Problem se reducira na  $y = Xb + e$ .

U PLS metodi, iz originalne matrice  $X$  generira se matrica faktorskih bodova,  $T$  i matrica opterećenja,  $P$ . Iz matrice  $Y$ , koja sadrži vrijednosti nekog svojstva više različitih jedinki u sustavu, generira se matrica faktorskih bodova  $U$  i matrica opterećenja  $Q$ . U izrazu  $P^T$ ,  $T$  označava transponiranu matricu, što vrijedi i za ostale matrice s ekponentom  $T$ .

$$X = TP^T + E \quad (1.20)$$

$$Y = UQ^T + F \quad (1.21)$$

Cilj PLS metode je modelirati sve članove koji tvore  $X$  i  $Y$  matrice, tako da su matrice pogrešaka, matrica  $E$  (za  $X$ ) te matrica  $F$  (za  $Y$ ) približno nula. Razvijena je unutarnja veza koja povezuje faktorske bodove  $X$  matrice i  $Y$  matrice. Veza se poboljšava iteracijski mijenjajući matrice faktorskih bodova  $U$  i  $T$ .[26]

Veza između matrice faktorskih bodova  $T$  (za matricu  $X$ ) i  $U$  (za matricu  $Y$ ) dana je jednažbom (1.22), gdje je  $D$  dijagonalna matrica s elementima  $d_1 \dots d_a$ , i matricom pogrešaka  $H$ .

$$U = TD + H \quad (1.22)$$

Pošto za PLS1 faktorskih bodova nema, jednažba (1.22) reducira se na

$$\mathbf{y} = \mathbf{Td} + \mathbf{h} \quad (1.23)$$

Cilj PLS2 metode je maksimizirati kovarijancu između faktorskih bodova  $\mathbf{X}$  i  $\mathbf{Y}$  matrice, za PLS1 je to kovarianca između faktorskih bodova  $\mathbf{X}$  matrice i  $\mathbf{y}$  varijabli. Kovarijanca kombinira visoku varijancu matrice  $\mathbf{X}$  i visoku korelaciju između matrica  $\mathbf{X}$  i  $\mathbf{Y}$ . [29]

Jednom kada je model izveden, kombinacijom jednačbi dobiva se matrica regresijskih vektora,  $\mathbf{B}$ , svaki vektor za jednu komponentu u  $\mathbf{Y}$  matrici. [26]

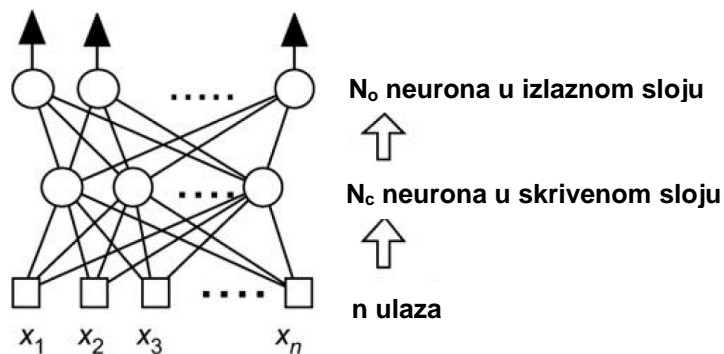
$$\hat{\mathbf{Y}} = \mathbf{X}\hat{\mathbf{B}} \quad (1.24)$$

## Umjetne neuronske mreže

Umjetne neuronske mreže (eng. *Artificial Neural Networks, ANN*) su matematički modeli koji se, između ostalog, rabe za procjenu izlaza viševeličinske funkcije za određene ulaze. [30] Njihova građa inspirirana je biološkim neuronskim mrežama. [31]

Sastavljene su od pojedinačnih elemenata, čvorova, koji su organizirani u slojeve, pa su slični neuronima i neuralnom sustavu u ljudskom mozgu. Svaki neuron povezuje se s drugim vezom koja ima asociranu vrijedost, tj. težinski koeficijent (eng. *weight*). Neuron je jednostavna procesna jedinica koja računa težinski zbroj ulaza filtriranih aktivacijskom funkcijom. Težinski koeficijent su slobodni parametri modela koji se podešavaju metodama treniranja i to na temelju informacija dostavljenih skupom podataka koji prikazuje ulaze i izlaze funkcije koja se procjenjuje.

Postoji velik broj struktura neuronskim mreža koje za modeliranje nelinearnih veza,. Često se primjenjuju unaprijedne neuronske mreže (eng. *Feed-Forward Neural Networks, FFNN*) čiji su neuroni organizirani u nekoliko slojeva, a neuroni iz jednog sloja povezani su s neuronima u sljedećem sloju (slika 21). [32] Informacije teku samo u jednom smjeru, od ulaza prema izlazu. Druga skupina neuronskim mreža su povratne neuronske mreže (eng. *Recurrent Neural Networks, RNN*). [30]



**Slika 21.** Unaprijedna mreža s  $n$  ulaza,  $N_c$  neurona u skrivenom sloju i  $N_o$  neurona u izlaznom sloju [30]



Strojno učenje je sposobnost sustava umjetne inteligencije da iz skupa podataka ekstrahira relevantne karakteristike, uzorke ili saznanja te zatim koristi to znanje da savlada nove probleme donošenjem odluka. Metode strojnog učenja mogu se uvježbati za prepoznavanje značajke u Ramanovom spektru i povežu ih s određenim jedinkama u sustavu.

Najčešći oblik strojnog/dubokog učenja je nadzirano učenje. Zove se nadzirano jer se treniranje modela provodi prezentiranjem više poznatih primjera različitih objekata, razreda ili izlaza izmjerenih u sličnim uvjetima. Model zatim uči ekstrahiranjem saznanja iz podataka i provodi generalizaciju klasifikacijskog problema. Kod Ramanove spektroskopije, novi Ramanovi spektri sekvencionalno se predstavljaju mreži, koja zatim predviđa odgovarajuću koncentraciju (izlaz).

Sve vrste neuronskih mreža imaju tendenciju pretreniranja. Mreža jednostavno nauči izlaze i ne stvara zaključke potrebne za predviđanje. Kako se to ne bi dogodilo potrebna je pažljivo izgrađena neuronska mreža, odabir ispravnih hiperparametara i validacija modela na nezavisnom skupu podataka.[33]

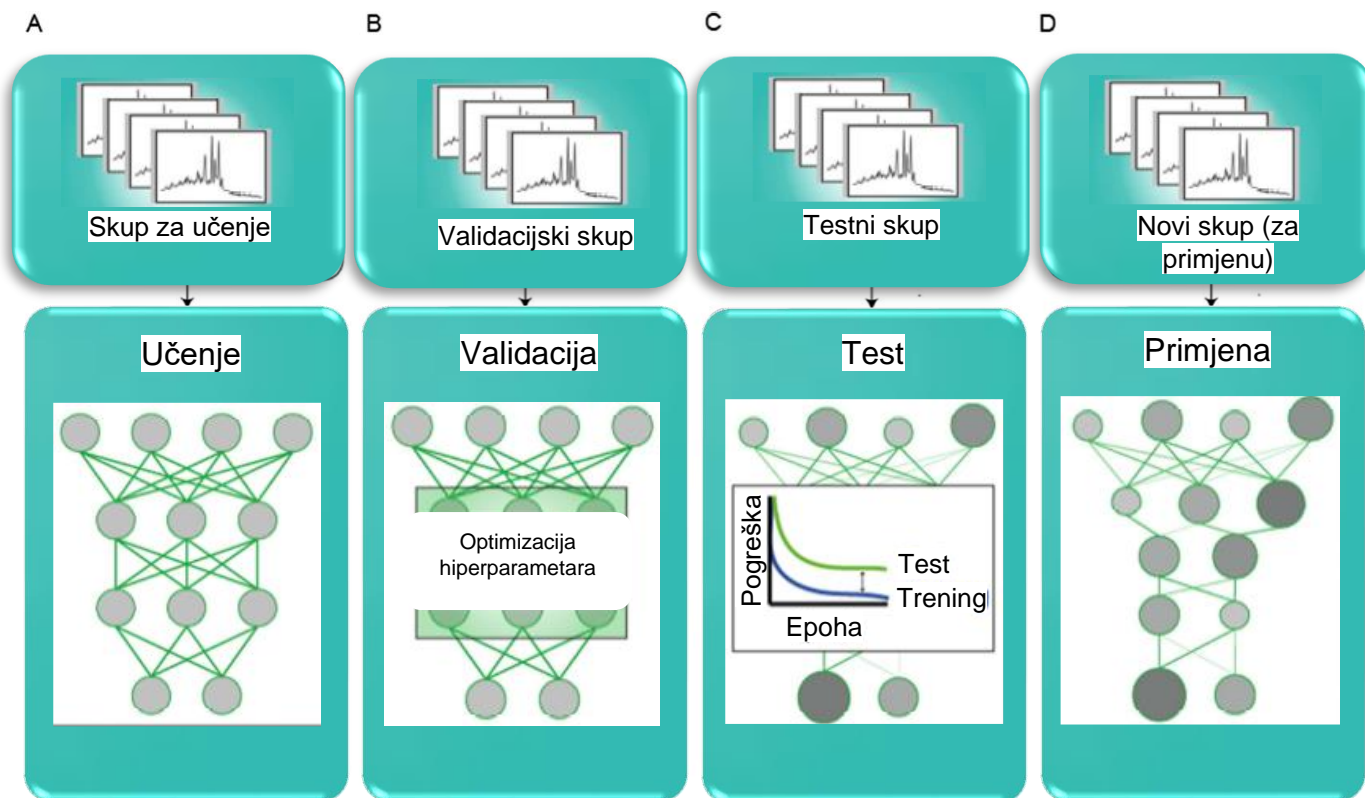
Za analizu Ramanovih spektara primjenjuju se mnogi algoritmi, a mogu se podijeliti na

- 1) metode temeljene na diskriminantnoj analizi kao što je linearna diskriminatna analiza (*Linear Discriminant Analysis, LDA*), metoda parcijalnih najmanjih kvadrata koju slijedi diskriminantna analiza (*Partial Least-Squares followed by a Discriminant Analysis, PLS-DA*), ili analiza diskriminantne funkcije (*Discriminant Function Analysis, DFA*)
- 2) modeli zasnovani na umjetnoj neuronskoj mreži (*Artificial Neural Network, ANN*), kao što višeslojni perceptron (*Multilayer Perceptron, MLP*), konvolucijska neuronska mreža (*Convolutional Neural Network, CNN*) ili algoritam strojnog vektora podrške (*Support Vector Machine, SVM*)
- 3) modeli temeljeni na regresijskoj analizi poput višestruke linearne regresije (*Multiple Linear Regression, MLR*), regresije glavne komponente (*Principal Components Regression, PCR*) ili metoda parcijalnih najmanjih kvadrata (*Partial Least Squares, PLS*)
- 4) modeli temeljeni na regresijskom stablu poput slučajnih šuma (*Random Forests, RF*), stabla klasifikacije i regresije (*Classification and Regression Trees, CART*)
- 5) evolucijski bazirani algoritmi tipa genetski algoritam (*Genetic Algorithm, GA*), genetsko programiranje i računanje (*Genetic Programming and Computing, GP i GC*), evolucijski algoritam (*Evolutionary Algorithm, EA*) i evolucijsko programiranje (*Evolutionary Programming, EP*)[33]

Kvaliteta podataka je kritična. Mora postojati ovisnost između informacija u podacima i svojstva uzoraka ili sustava koje se predviđa. Jedan od nedostataka modela je pretreniranost.[33]

## Učenje modela neuronske mreže

Nadzirano učenje zahtjeva skup podataka za učenje. Skup za validaciju koristi se prilikom odabira hiperparametara. Modeli se treniraju s različitim vrijednostima hiperparametara i provjerava se njihova točnost na skupu za validaciju. Testni skup podataka služi za nepristanu provjeru modela.[33]



Slika 22. Opća shema optimiziranja umjetne neuronske mreže[33]

Na slici 22 prikazana su tri nezavisna Ramanova spektralna skupa podataka (učenje, validiranje i testiranje). Neuroni su označeni sivim krugovima, dok su težinski koeficijenti označeni kao zelene linije koje povezuju neurone različitih slojeva mreže.

Multivarijantni modeli često slijede princip “linearni modeli za linearne sustave, nelinearni modeli za nelinearne sustave”. Linearni modeli su jednostavniji, bezirani su na dobro poznatim fizikalno-kemijskim zakonitostima te su pouzdani. Preferirano se koriste kada je sustav linearan.

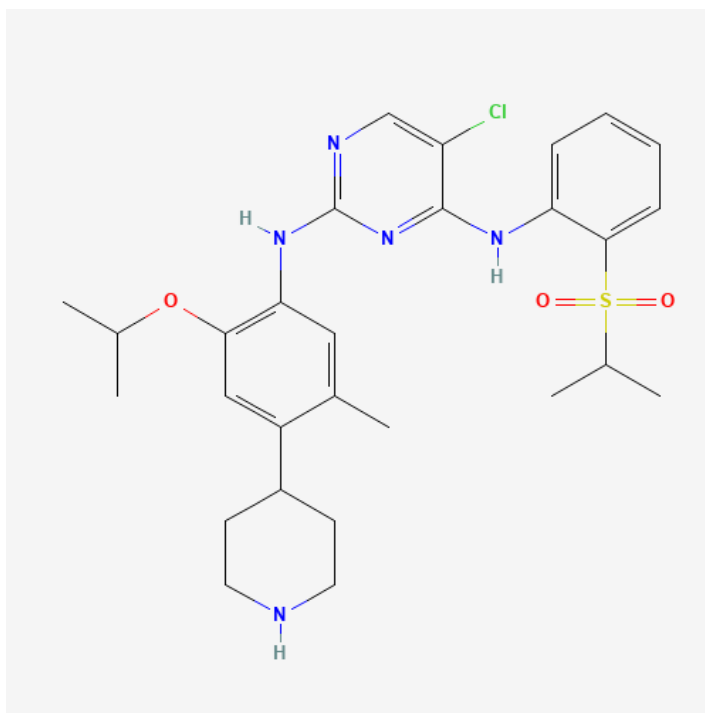
Ako se sumnja na prisutnost nelinearnosti, poželjno je početi s primjenom linearnog modela (PCR, PLS1, PLS2) i vidjeti da li je on zadovoljavajuć. Korisno je koristiti taj pristup jer je mnogo spektralnih signala u linearnoj vezi s koncentracijom. Ako takav model nije zadovoljavajuć, tek tada je poželjno preći na nelinearne modele.

Umjetne neuronske mreže se koriste za modeliranje nelinearnih odnosa između multivarijantnog signala i koncentracije analita (ceritiniba). U ovom kontekstu, mogu se promatrati kao kalibracijski modeli s određenim brojem promjenjivih parametara, koji opisuju nelinearne relacije između podataka.

## Ekperimentalni dio

### Djelatna tvar

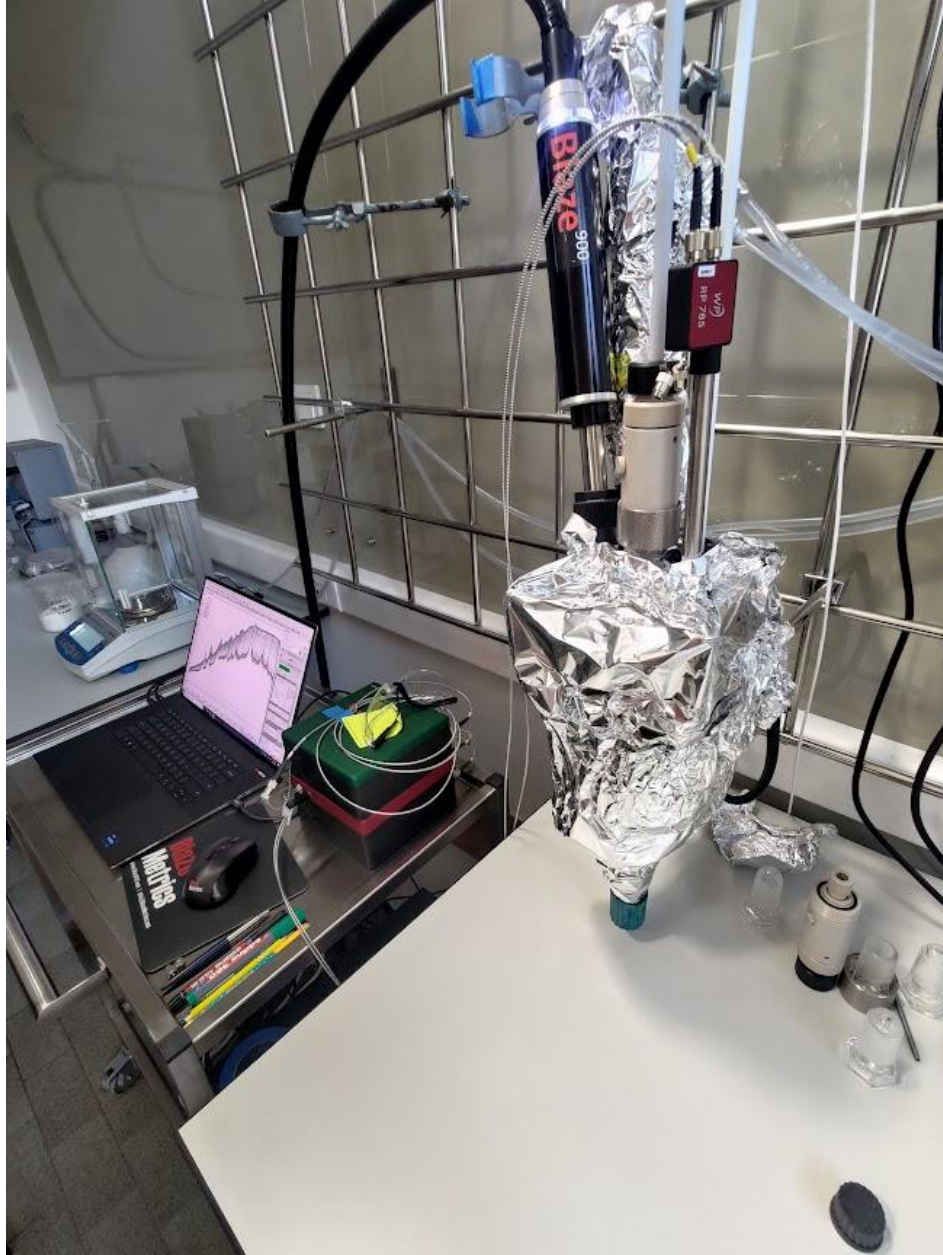
Kalibracijski model se razvijao za djelatnu tvar ceritinib forme A. Ceritinib se javlja u 3 različite forme, A, B i C. IUPAC naziv je 5-chloro-2-*N*-(5-methyl-4-piperidin-4-yl-2-propan-2-yloxyphenyl)-4-*N*-(2-propan-2-ylsulfonylphenyl)pyrimidine-2,4-diamine (slika 25).



**Slika 25.** 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) [36]

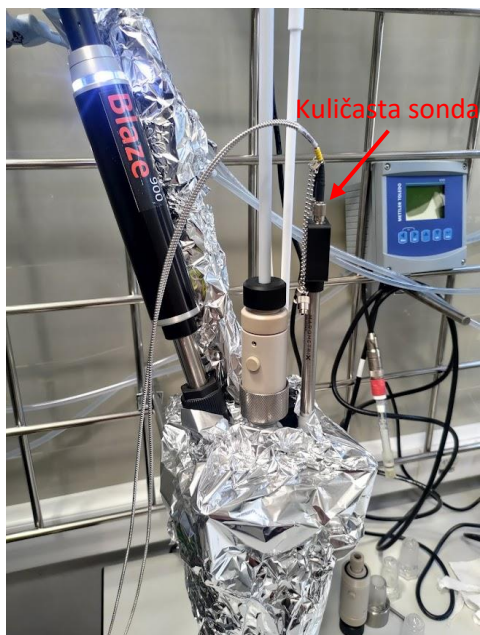
## Aparatura

Aparatura za provedbu hlađenja na slici 26 sastoji se od 1 L staklenog reaktora s plaštom, povratnog hladila, termostata za grijanje/hlađenje putem plašta reaktora, temperaturnog osjetila i miješalice s motorom.



**Slika 26.** Laboratorijski sustav za provedbu eksperimenata

Za mjerenje Ramanovih spektara primjenjuje se spektrometar (*Wasatch Photonics*) s izvorom zračenja valne duljine 785 nm i softverski paket (*Panorama*) te kugličasta sonda (eng. *ball-probe*) (*Wasatch Photonics*) (slika 27).



Slika 27. Kugličasta sonda(eng. *ball-probe*)

### Aparatura za određivanje sustava otapala

Aparatura za određivanje sustava otapala u kojem je najveća topljivost ceritiniba forme A sastojala se od 250 mL reaktora s plaštem za medij, termostata koji služi za grijanje/hlađenje reaktora, temperaturnog osjetila Pt-100, termostata, računala za upravljanje termostatom i miješalice s motorom (slika 28).



Slika 28. Aparatura za određivanje topljivosti

## Opis eksperimenta

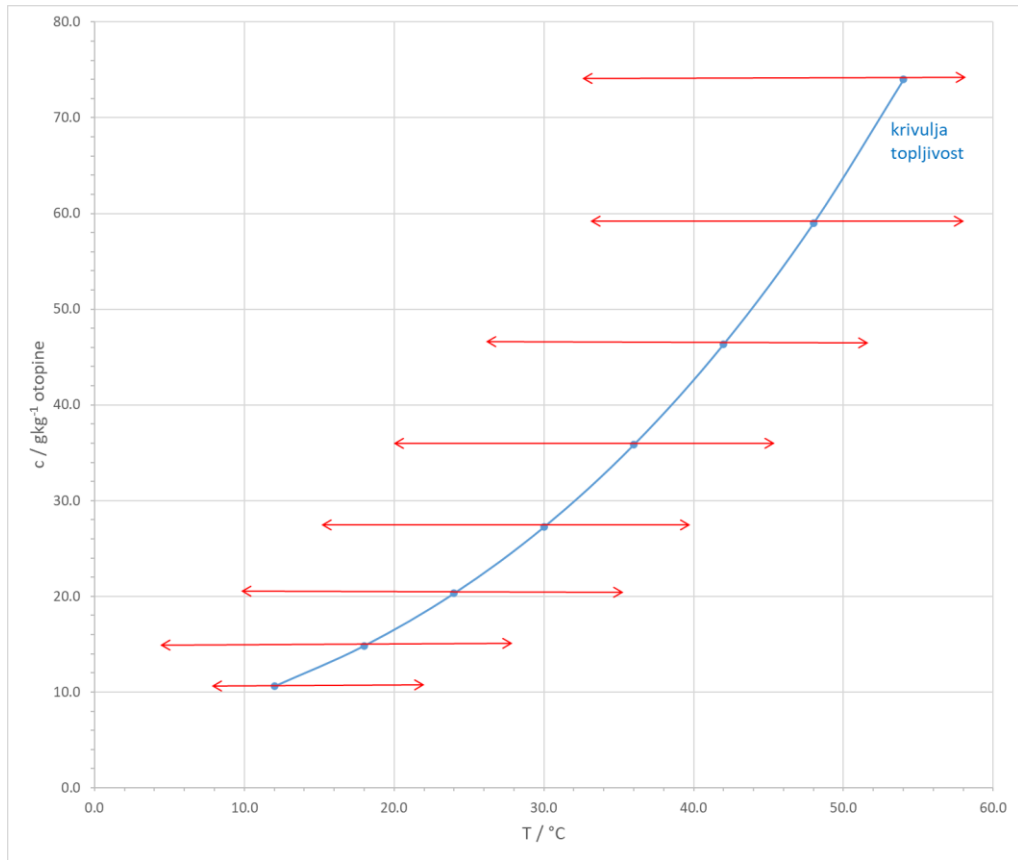
### 1) Određivanje kristalizacijskog sustava

- a. Kako bi se mogao izraditi kalibracijski model potrebni su dovoljno veliki intenziteti vrpca na Ramanovom spektru. Što je veća koncentracija otopljene tvari, to je veći intenzitet vrpca na Ramanovom spektru koje su karakteristične za nju. Eksperimentalno je određena krivulja topljivosti za više različitih omjera sustava otapala aceton-voda. Voda je u ovom slučaju antiotapalo.
- b. Eksperimentalno su određene topljivosti ceritiniba forme A u sljedećim sustavima otapala: 100 % aceton, 90 % aceton - 10 % voda i 70 % aceton - 30 % voda.
- c. Na temelju eksperimentalnih podataka topljivosti za prethodno navedene sustave, u programu Dynochem moguće je proračunati topljivosti za druge omjere otapala acetona i vode. Za provedbu daljnjih eksperimenta odabran je sustav 90 % aceton - 10 % voda.

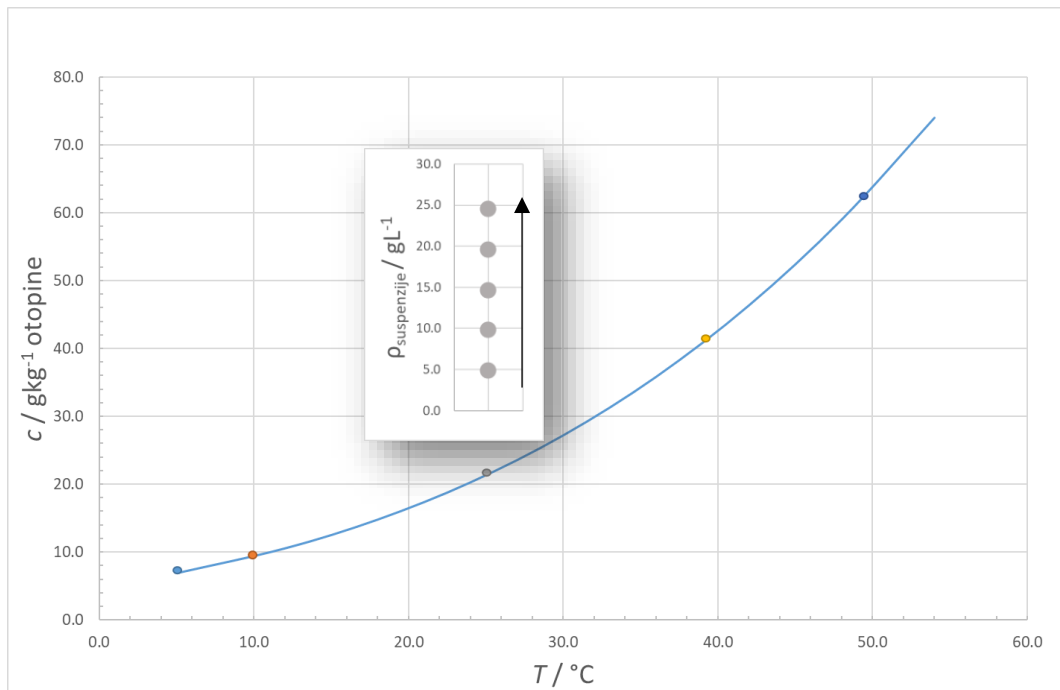
### 2) Provedba kalibracijskog eksperimenta

Kalibracijski eksperiment provodi se u sljedećim koracima:

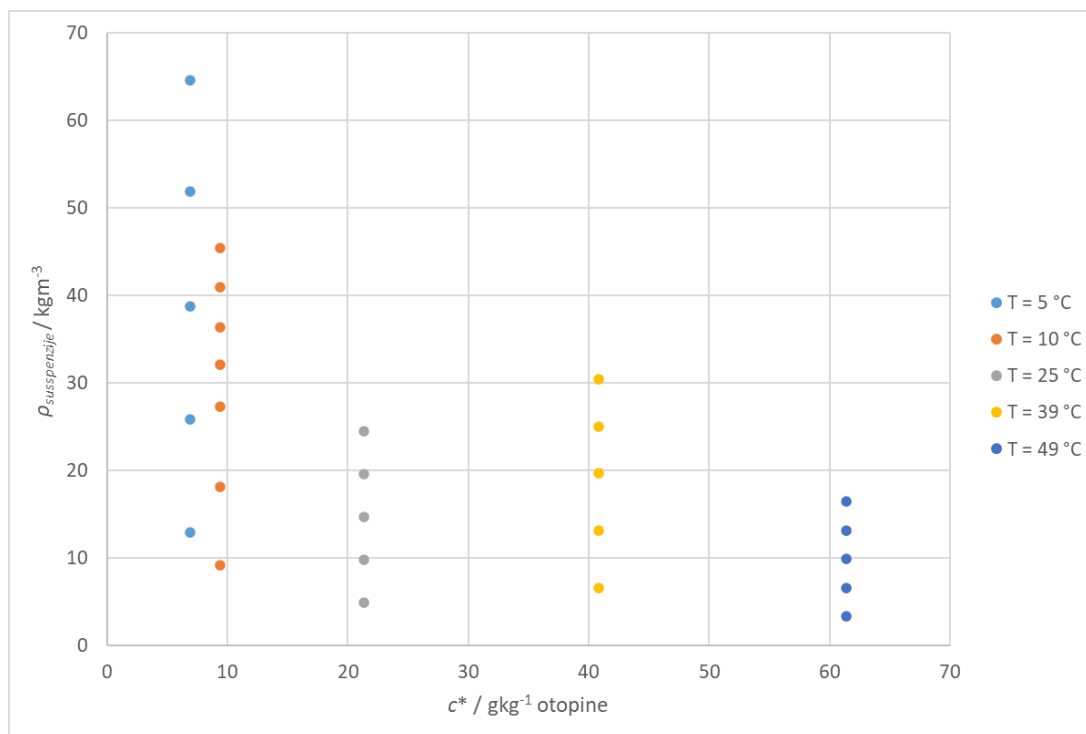
- a) Mjerenje ovisnosti Ramanovog spektra o temperaturi otapala.  
Mjerenje ovisnosti Ramanovog spektra o temperaturi pri različitim koncentracijama otopine ceritiniba. Mjerenje kreće od više temperature gdje je otopina nezasićena te se hladi određenom brzinom te time prelazi u metastabilno područje. Mjerenje se prekida neposredno prije krivulje prezasićenja, a ponavlja se za niz otopina različitih koncentracija. Na slici 29 prikazana je krivulja topljivosti, koncentracije za koje su provedena mjerenja označene su plavim točkama. Crvenom strelicom označen je temperaturni raspon za koji su snimani ramanovi spektri.
- b) Mjerenje ovisnosti Ramanovog spektra o gustoći suspenzije ceritiniba forme A.  
Provedeni eksperimenti prikazani su na slici 30. U određeni volumen otapala se pri određenoj temperaturi doda određena masa kristala da bi se postigla zasićena otopina. Na slici 30 točkama na krivulji topljivosti različitim bojama označene su zasićene otopine u koje su se zatim dodavali kristali. Doda se jedan obrok kristala i snimi se Ramanov spektar. Zatim se doda sljedeći obrok čime se povećava gustoća suspenzije i opet snimi Ramanov spektar. Dodavanje obroka kristala, tj. povećanje gustoće suspenzije označeno je sivim točkama na slici 30. Na slici 31 na y osi prikazane su sve gustoće suspenzije otopine za koje je snimljen Ramanov spektar, dok je na x osi označena ravnotežna koncentracija otopine. Za vrijeme provedbe snimanja ramanovih spektara pri jednoj ravnotežnoj koncentraciji temperatura se održava konstantnom.



Slika 29. Eksperimenti snimanja otopina



Slika 30. Eksperimenti snimanja suspenzija



**Slika 31.** Gustoće suspenzija za određenu zasićenu otopinu pri konstantnoj temperaturi

### 3) Razvoj modela

Prikupljeni podaci upotrijebljeni su za razvoj kalibracijskih modela. Kalibracijski model za predviđanje gustoće suspenzije razvijen je nezavisno od kalibracijskog modela za predviđanje koncentracije otopine.

## Razvoj kalibracijskog modela

Eksperimentalno prikupljen skup podataka sastoji se od relativnih intenziteta vrpci Ramanovog spektra i temperature. Koncentracije otopine i gustoće suspenzije proračunate su za određeni volumen otapala.

Metodama PLSR i neuronskom mrežom, predviđaju se koncentracije otopine i gustoće suspenzije. Kalibracijski modeli razvijeni su odvojeno za koncentraciju i gustoću suspenzije.

Ukupni skup podataka sastoji se od relativnih intenziteta vrpci Ramanovog spektra, temperature, koncentracije otopina i gustoće suspenzija. Iz ukupnog skupa podataka za izradu kalibracijskog modela za predviđanje koncentracije otopine odvajaju se podaci za Y matricu koja sadrži koncentracije otopine ceritiniba i X matricu koja sadrži Ramanove spektre i temperature (tablica 2). Za izradu kalibracijskog modela za predviđanje gustoće suspenzije Y matrica umjesto

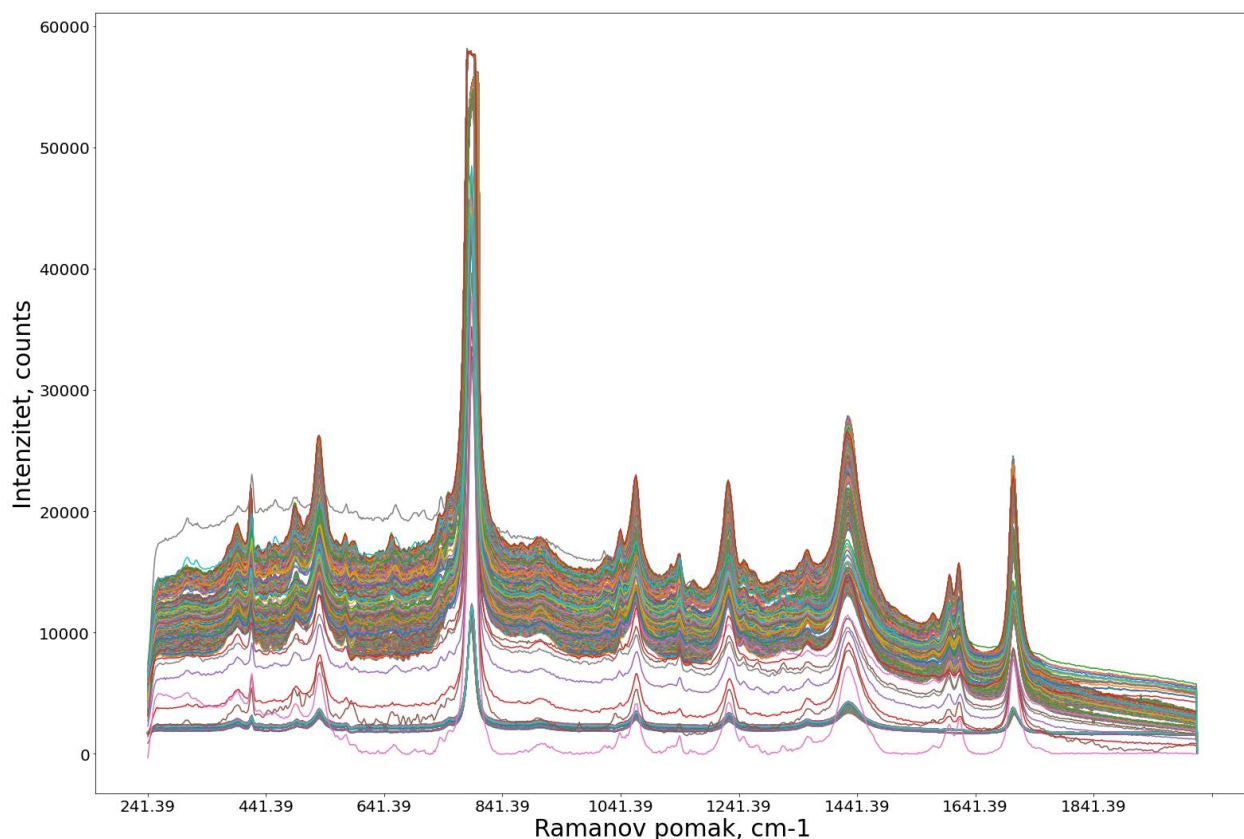


koncentracija otopina sadrži gustoće suspenzija.

Ukupni skup podataka dijeli se na skup za učenje i test skup podataka. Skup za učenje služi za učenje modela. Testni skup podataka koriste se za nepristranu ocjenu modela. Ulazi modela su Ramanovi spektri i temperature, dok su izlazi modela predviđene koncentracije otopine i gustoće suspenzije.

**Tablica 2.** X i Y matrice za kalibracijski model otopine

Eksperiment	Y						
	c / g/kg solution	X					
		Ramanov pomak / cm <sup>-1</sup>					T / °C
		241.39	243.39	...	2013.39	2015.39	
1	10,8	1706,00	1772,21	...	1573,68	1574,92	21,9
	...	...	...	...	...	...	...
	10,8	1749,00	1809,60	...	1579,25	1580,55	7,9
2	15,1	4595,00	5351,65	...	2866,31	2870,12	28,7
	...	...	...	...	...	...	...
	15,1	4121,00	4635,22	...	2897,02	2852,80	5,1
...	...	...	...	...	...	...	...
8	74,6	3167,00	3967,66	...	1800,06	1813,42	56,9
	...	...	...	...	...	...	...
	74,6	3498,00	4412,33	...	2172,17	2178,46	31,8
9	7,1	1413,00	1812,65	...	672,67	697,86	5,0
	7,1	2268,67	2856,12	...	1222,31	1198,90	5,0
...	...	...	...	...	...	...	...
13	62,9	6470,00	7191,54	...	5203,19	5172,00	49,0
	62,9	6497,33	7321,14	...	5272,48	5283,17	49,0

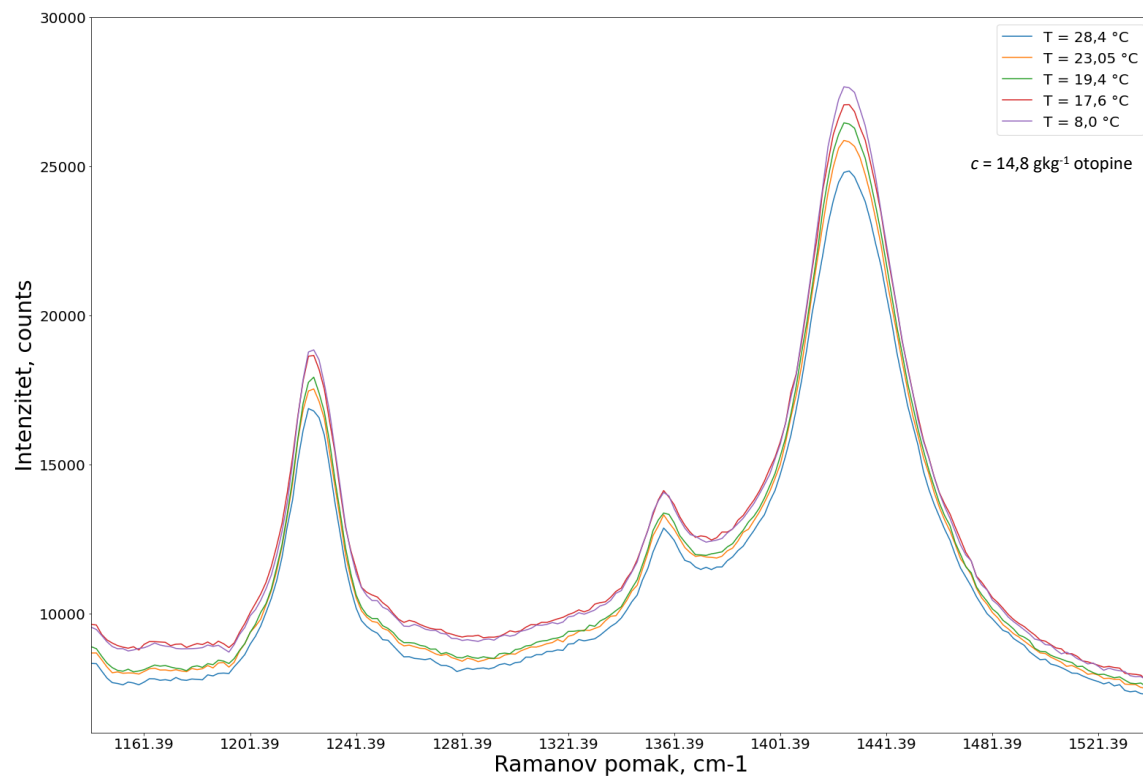


**Slika 32.** Ramanovi spektri

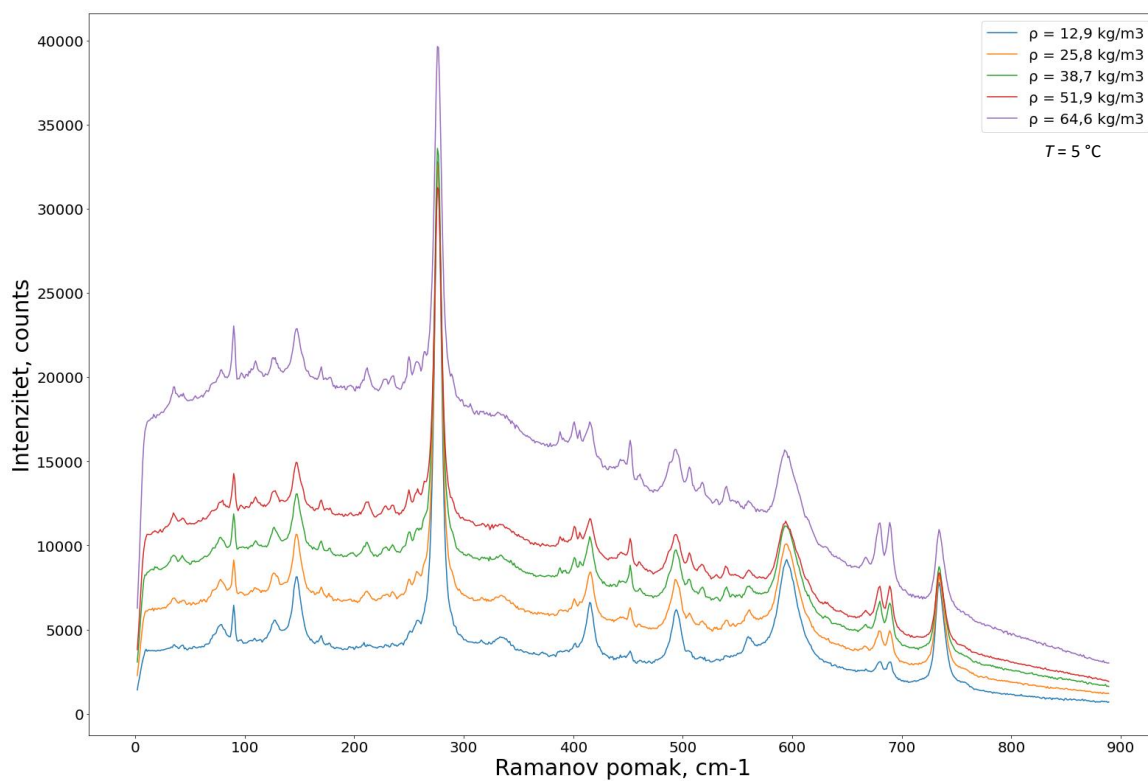
Na slici 32 prikazane su vrpce Ramanovih spektara svih provedenih eksperimenata. Na x osi označen je Ramanov pomak koji označava promjenu u valnoj duljini reflektiranih fotona. Iznos i broj pomaka određuje vrsta tvari koja je mjeri. Na y osi označen je intenzitet koji predstavlja detektirani broj reflektivnih fotona. Uočava se pomak u intenzitetu ramanovog spektra s promjenom temperature, koncentracije i gustoće otopine. Detaljniji prikaz ovisnosti ramanovog spektra o navedenim varijablama prikazan je na slikama 33 i 34.

Na slici 33 prikazana je ovisnost Ramanovog spektra otopine koncentracije 14,8 g/kg otopine o temperaturi. Porastom temperature raste intenzitet pojedinih vrpca. Intenzitet svih vrpca ne raste jednako što ovisi o tvari koja se mjeri. Pošto dolazi do preklapanja određenih vrpca otapala i otopljenje tvari nije moguće razlučiti da li porast intenziteta dolazi od vrpce otapala ili otopljene tvari.

Na slici 34 prikazana je ovisnost Ramanovog spektra suspenzije o gustoći suspenzije. Temperatura je konstantna i iznosi 39,0 °C. Porastom gustoće suspenzije raste intenzitet vrpca Ramanovog spektra. Ovaj spektar se razlikuje od onog na slici 32 jer se u sustavu nalaze i kristali ceritiniba forme A koji također imaju svoje karakteristične vrpce Ramanovog spektra. Stoga dolazi do preklapanja vrpca otapala, otopljene tvari i kristala.



**Slika 33.** Ovisnost Ramanovog spektra otopine o temperaturi



**Slika 34.** Ovisnost Ramanovog spektra o gustoći suspenzije

Za razvoj kalibracijskih modela primjenjene su parcijalna regresija metodom najmanjih kvadrata (eng. *Partial Least Squares Regression, PLSR*) i unaprijedna neuronska mreža (eng. *Feed Forward Neural Network, FFNN*).

PLSR metoda reducira dimenzionalnost podataka na mali broj latentnih varijabli čiji se broj određuje križnom validacijom računajući vrijednosti srednje kvadratne pogreške za križnu validaciju (eng. *Root Mean Squared Error for Cross Validation, RMSECV*) i kvadrat koeficijenta korelacije (eng. *Coefficient of Determination,  $R_{CV}^2$* ).

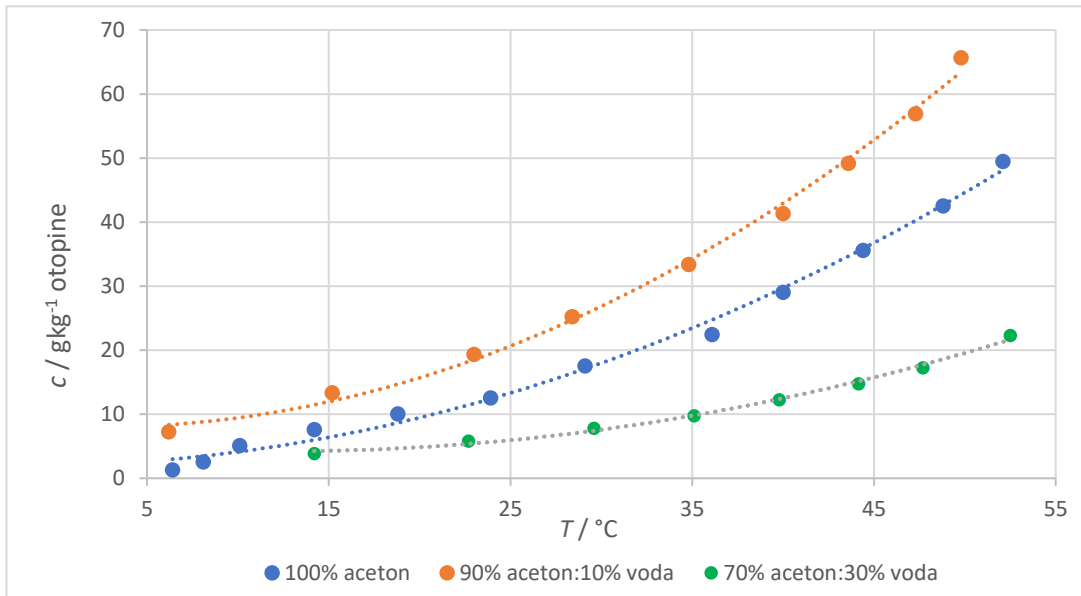
## Rezultati i rasprava

### Određivanje optimalnog sustava otapala

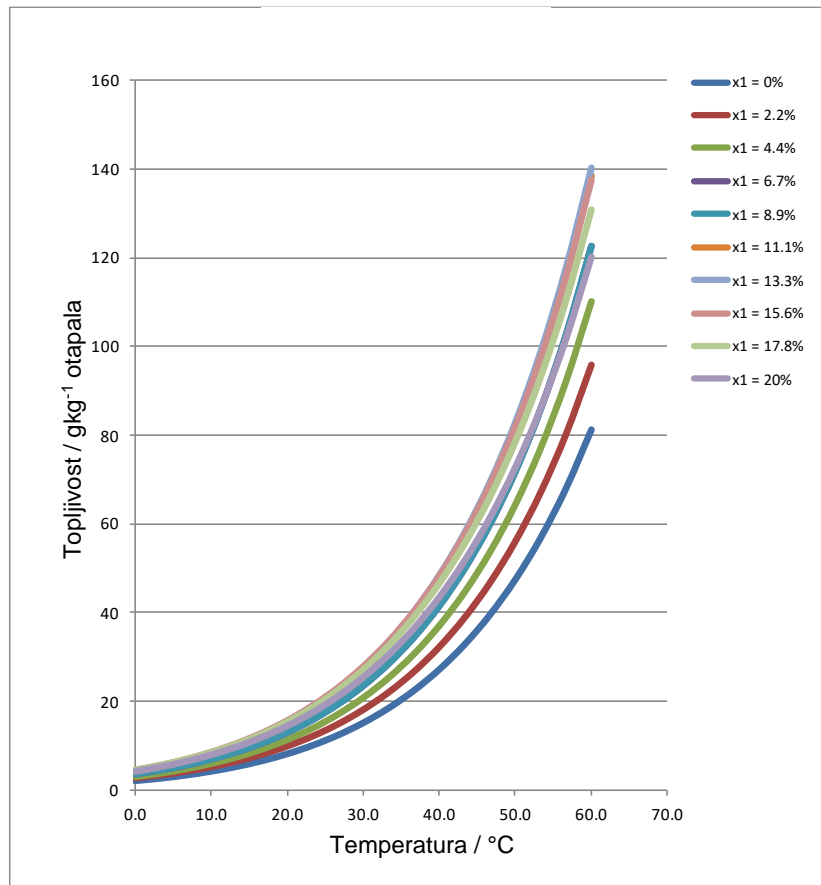
Podaci snimljeni u eksperimentu određivanja topljivosti ceritiniba forme A u različitim sustavima otapala aceton-voda prikazani su u tablici 3. Određene su krivulje topljivosti ceritiniba forme A za sustave otapala: 100 % aceton, 90 % aceton - 10 % voda i 70 % aceton - 30 % voda (slika 35).

**Tablica 3.** Eksperimentalni podaci za određivanje krivulje topljivosti ceritiniba forme A u različitim sustavima otapala aceton-voda

100 % aceton			90 % aceton - 10 % voda			70 % aceton - 30 % voda		
$m_{kum}/g$	$T/°C$	$c /gkg^{-1}$ otopine	$m_{kum}/g$	$T/°C$	$c /gkg^{-1}$ otopine	$m_{kum}/g$	$T/°C$	$c /gkg^{-1}$ otopine
0,10	6,4	1,3	0,60	6,2	7,3	0,33	14,2	3,9
0,20	8,1	2,6	1,10	15,2	13,4	0,50	22,7	5,8
0,41	10,1	5,1	1,61	23,0	19,4	0,67	29,6	7,8
0,61	14,2	7,6	2,11	28,4	25,3	0,84	35,1	9,8
0,81	18,8	10,1	2,81	34,8	33,4	1,06	39,8	12,2
1,01	23,9	12,6	3,51	40,0	41,4	1,28	44,1	14,8
1,42	29,1	17,5	4,21	43,6	49,2	1,50	47,7	17,2
1,82	36,1	22,4	4,91	47,3	56,9	1,95	52,5	22,3
2,37	40,0	29,1	5,72	49,8	65,7			
2,93	44,4	35,6						
3,52	48,8	42,6						
4,13	52,1	49,5						



**Slika 35.** Krivulje topljivosti za sustav aceton - voda



**Slika 36.** Ovisnost topljivosti ceritiniba forme A o masenom udjelu vode za sustav otapala aceton-voda

Na slici 36 prikazana je ovisnost topljivosti ceritiniba forme A o masenom udjelu vode za sustav otapa aceton-voda. Voda je antiotapalo pa je njezin maseni udio označen s X1 u tablici 4. Maksimum topljivosti je između 11,1 % i 13,3 % masenog udjela vode (antiotapalo) u sustavu otapala aceton-voda. Na temelju zadovoljavajuće topljivosti i jednostavnosti izračuna i provedbe daljnjih eksperimenata odabran je sustav 90 % aceton - 10 % voda kao sustav otapala.

U programu *Dynochem* proračunati su koeficijenti jednadže koja opisuje eksperimentalne podatke za sustav 90 % aceton - 10 % voda (tablica 4). Dobivenim izrazom dalje se računaju mase ceritiniba forme A potrebne za provedbu eksperimenata.

**Tablica 4.** Koeficijenti jednadžbe

$c = \exp(\ln A5 - B5/(R \cdot T) - C5/(R \cdot T)^2 - D5 \cdot \exp(X1))$							
lnA5	B5	C5	D5	R	antiotapalo	otapalo	otopljena tvar
-	kJ/mol	(kJ/mol) <sup>2</sup>	-	kJ/(molK)	kg/m <sup>3</sup>	kg/m <sup>3</sup>	kg/m <sup>3</sup>
7,8	-13,4	62,5	0	$8,3 \cdot 10^{-3}$	998	793	1250

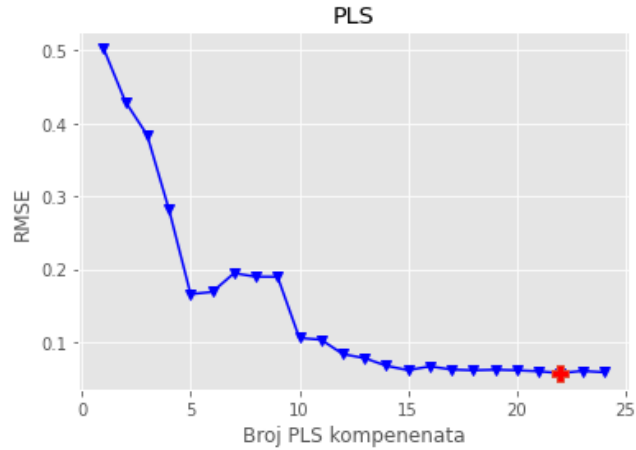
## Izrada kalibracijskog modela

### PLSR metoda

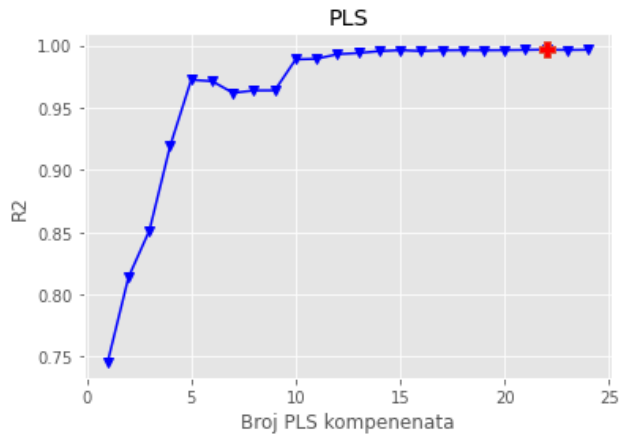
Provedena je 10-fold križna validacija. Skup za učenje podijeli se na deset jednakih dijelova, pri čemu se podaci u svaki biraju nasumično. Devet dijelova se koristi za izradu modela, dok se jedan dio ostavlja za validaciju. Ovaj postupak provodi se deset puta, s tim da se svaki put novi dio ostavlja kao validacijski, a model se gradi s preostalih devet dijelova. Uzima se prosječna vrijednost korijena srednje kvadratne pogreške i kvadrata koeficijenta korelacije kao vrijednosti korijena srednje kvadratne pogreške za križnu validaciju, RMSECV i kvadrat koeficijenta korelacije,  $R^2_{CV}$ .

Na slikama 37 i 38 prikazana je ovisnost korijena srednje kvadratne pogreške za križnu validaciju, RMSECV i maksimalan kvadrat koeficijenta korelacije,  $R^2_{CV}$  za prvih 25 latentnih varijabli PLSR metode modela za koncentraciju otopine.

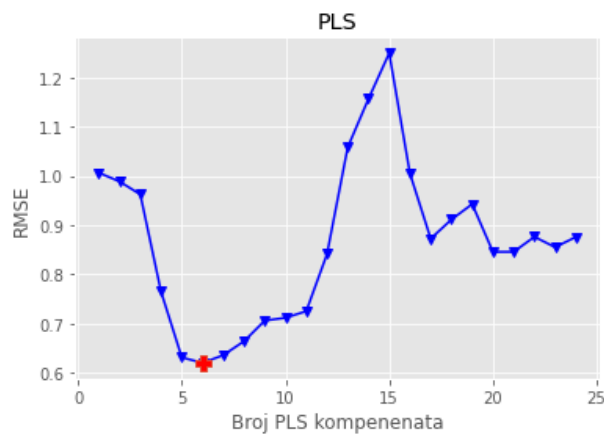
Na slikama 39 i 40 prikazana je ovisnost korijena srednje kvadratne pogreške za križnu validaciju, RMSECV i maksimalan kvadrat koeficijenta korelacije,  $R^2_{CV}$  za prvih 25 latentnih varijabli PLSR metode modela za gustoću suspenzije.



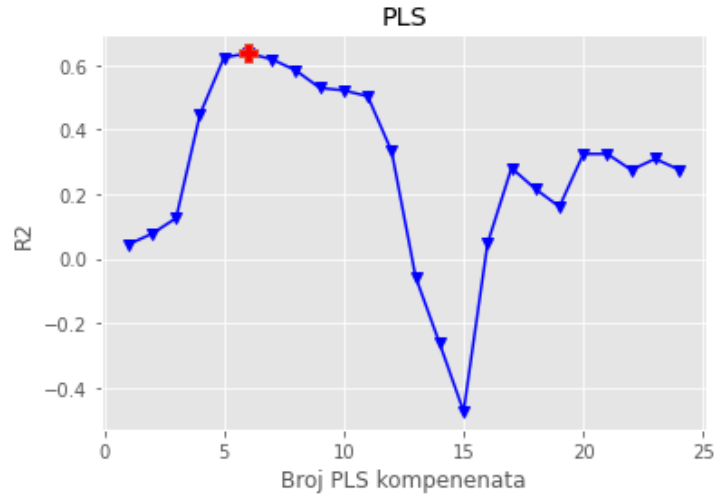
**Slika 37.** Ovisnost RMSECV o broju PLS kompenenata (latentnih varijabli) za kalibracijski model koncentracije



**Slika 38.** Ovisnost  $R^2_{CV}$  o broju PLS kompenenata (latentnih varijabli) za kalibracijski model koncentracije



**Slika 39.** Ovisnost RMSECV o broju PLS kompenenata (latentnih varijabli) za kalibracijski model gustoće



**Slika 40.** Ovisnost  $R^2_{CV}$  o broju PLS kompenenata (latentnih varijabli) za kalibracijski model gustoće

Za kalibracijski model koncentracije minimum korijena srednje kvadratne pogreške za križnu validaciju, RMSECV nalazi se pri 22 PLS kompenenata (latentnih varijabli). Maksimum kvadrata koeficijenta korelacije,  $R^2_{CV}$  je također pri 22 PLS komponente. Minimum RMSECV i maksimum  $R^2_{CV}$  potvrđuju da su 22 latentne varijable najbolji izbor za PLS metodu (tablica 5).

Za kalibracijski model gustoće suspenzije minimum korijena srednje kvadratne pogreške za križnu validaciju, RMSECV nalazi se pri 6 PLS kompenenata (latentnih varijabli). Maksimum kvadrata koeficijenta korelacije,  $R^2_{CV}$  je također pri 6 PLS komponenti. Minimum RMSECV i maksimum  $R^2_{CV}$  potvrđuju da je 6 latentnih varijabli najbolji izbor za PLS metodu (tablica 5).

**Tablica 5.** Vrijednosti RMSECV i  $R^2_{CV}$  za 22 latentih varijabli za kalibracijski model koncentracije i za 6 latentnih varijabli za kalibracijski model gustoće

Pokazatelj	Kalibracijski model za koncentraciju otopine	Kalibracijski model za gustoću suspenzije
$R^2_{CV}$	0,99	0,75
RMSECV	0,06	0,51



Različitim metodama predobrade podataka moguće je smanjiti RMSE, tj. povećati točnost modela. U tablici 6 prikazana je ovisnost RMSE testnog skupa podataka o vrsti predobrade podataka.

**Tablica 6.** Ovisnost RMSE testnog skupa podataka o vrsti predobrade podataka

Predobrada	RMSE modela za koncentraciju / gkg <sup>-1</sup>	RMSE modela za gustoću suspenzije / gL <sup>-1</sup>
Bez predobrade	1,62	0,43
Korekcija bazne linije	0,08	0,45
Filtriranje Savitzky – Golay filtrom	0,09	0,43
“Min-max” normalizacija	0,04	0,34
Uklanjanje dijela spektra (260-285 cm <sup>-1</sup> )	0,11	0,42
Korekcija bazne linije + filtriranje Savitzky – Golay filtrom	0,09	0,46
Korekcija bazne linije + filtriranje Savitzky – Golay filtrom + “Min-max” normalizacija	<b>0,04</b>	0,32
Korekcija bazne linije + “Min-max” normalizacija	0,05	0,30
Druga derivacija	0,10	0,52
MSC	0,04	<b>0,26</b>
Odabir dijela spektra karakterističnog za kristal	-	0,49

Nad podacima Y matrice tj. y vektora provedena je "Z-score" normalizacija (standardizacija) dok je nad podacima X matrice provedena kombinacija različitih predobrada podataka.

Minimum RMSE testnog skupa pokazuje da je preobrada podataka koja se sastoji od korekcije bazne linije, filtriranje podataka Savitzky-Golay filtrom (duljina prozora 15, red polinoma 3) te “Min-max” normalizacije najbolji izbor za X matricu. Istu vrijednost RMSE pokazuju i predobrada podataka samo s “Min-max” normalizacijom te MSC predobrada. Od navedenih predobrada odabrana je ona koja se sastoji od korekcije bazne linije, filtriranja podataka Savitzky-Golay filtrom (duljina prozora 15, red polinoma 3) i “Min-max” normalizacije, iako bi se mogle odabrati bilo koje od navedenih kombinacija predobrade koje pokazuju minimalan RMSE.

Jedna od metoda je izrezivanje dijela spektra (260 - 285 cm<sup>-1</sup>) kako bi se uklonila vrpca koja je karakteristična za metanol.

Moguće je i odabrati samo dio spektra gdje se nalaze vrpce koje su karakteristične za kristal. To su redom sljedeća područja valnih brojeva: 343,70 – 351,86, 457,66 – 460,87, 617,87- 620,80, 643,75 – 647,27, 753,39 – 759,67 i 777,04 – 781,38  $\text{cm}^{-1}$ . Ova vrsta predobrade rabi se samo pri izradi modela za gustoću suspenzije.

Druga derivacija spektralnih podataka pokazuje bolje odvajanje vrpce spektra te je moguće točnije određivanje njihovog maksimuma.

MSC (eng. *Multiplicative Scatter Correction*) korigira efekt raspršenja (eng. *scatter effect*). Nastoji ispraviti sve spektre tako da su oni što bliže referentnom spektru, koji je u ovom slučaju prosjek skupa podataka.

Standardizacijom se od svake vrijednosti oduzima prosječna. Taj iznos dijeli sa standardnom devijacijom skupa podataka.

$$x' = \frac{x - \bar{x}}{\sigma_x}$$

"Min-max" normalizacija skalira podatke između vrijednosti 0 i 1 kako bi se izbjegao veliki doprinos brojčano velikih veličina.

$$x' = \frac{x - x_{min}}{x_{max} - x_{min}}$$

Kvadrat koeficijenta korelacije (eng. *Coefficient of Determination*,  $R^2$ ) i korijen srednje kvadratne pogreške (eng. *Root Mean Squared Error*, *RMSE*) kriteriji su za ocjenu modela.

Kvadrat koeficijenta korelacije je broj između 0 i 1 koji pokazuje koliko dobro model predviđa izlaz tj. zavisnu varijablu. Što je vrijednosti veća 1 to je bolje predviđanje modela.

Korijen srednje kvadratne pogreške koristi se kao mjera slučajne pogreške modela pri predviđanju izlaza. Što je manji korijen srednje kvadratne pogreške to je bolje predviđanje modela. Mjera je standardne devijacije pogrešaka (residuala).

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

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

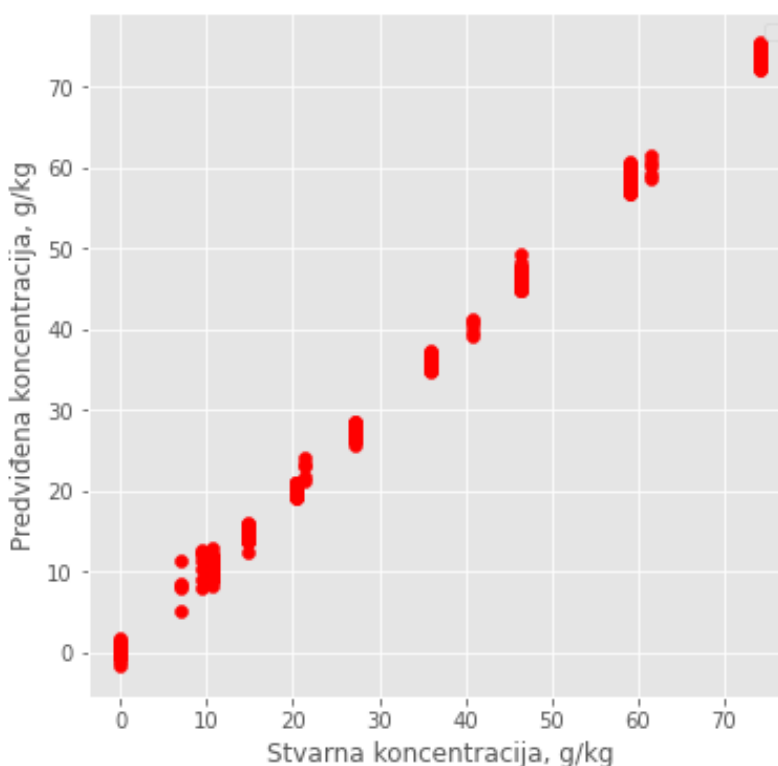
$\hat{y}_i$  – izlaz predviđen modelom  
 $y_i$  – eksperimentalna vrijednost  
 $\bar{y}$  – srednja vrijednost y

Ocjene PLS modela prikazane su u tablici 7.

**Tablica 7.** Vrijednosti korijena srednje kvadratne pogreške u postocima, RMSE i  $R^2$  za testni skup podataka za PLS modele

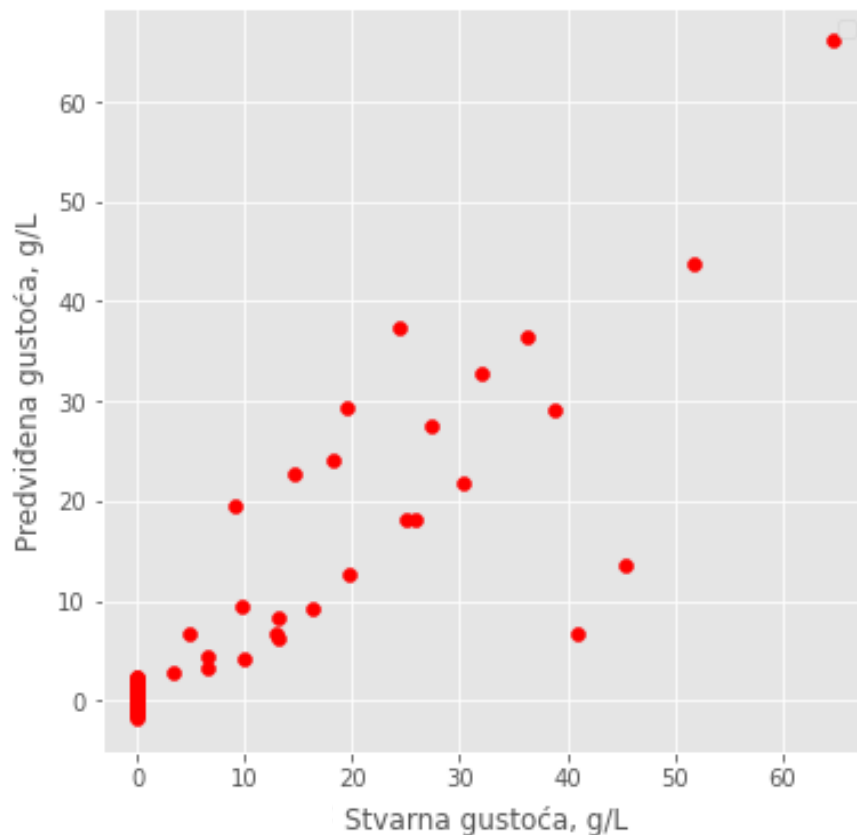
Pokazatelj	Kalibracijski model za koncentraciju otopine	Kalibracijski model za gustoću suspenzije
$R^2$	0,99	0,90
RMSE %	1,32 %	1,62 %

Na slici 41 prikazane su predviđene koncentracije PLS modela. Na x osi označena je stvarna koncentracija otopine, do je na y osi označena predviđena koncentracije otopine. Može se uočiti da model ne predviđa idealno koncentracije otopine, pošto su predviđanja koncentracije otopine raspršena po y osi. Što je bolje predviđanje modela to su predviđene koncentracije otopina “zbijene” u jednu točku.



**Slika 41.** Predviđene koncentracije otopine PLS modelom

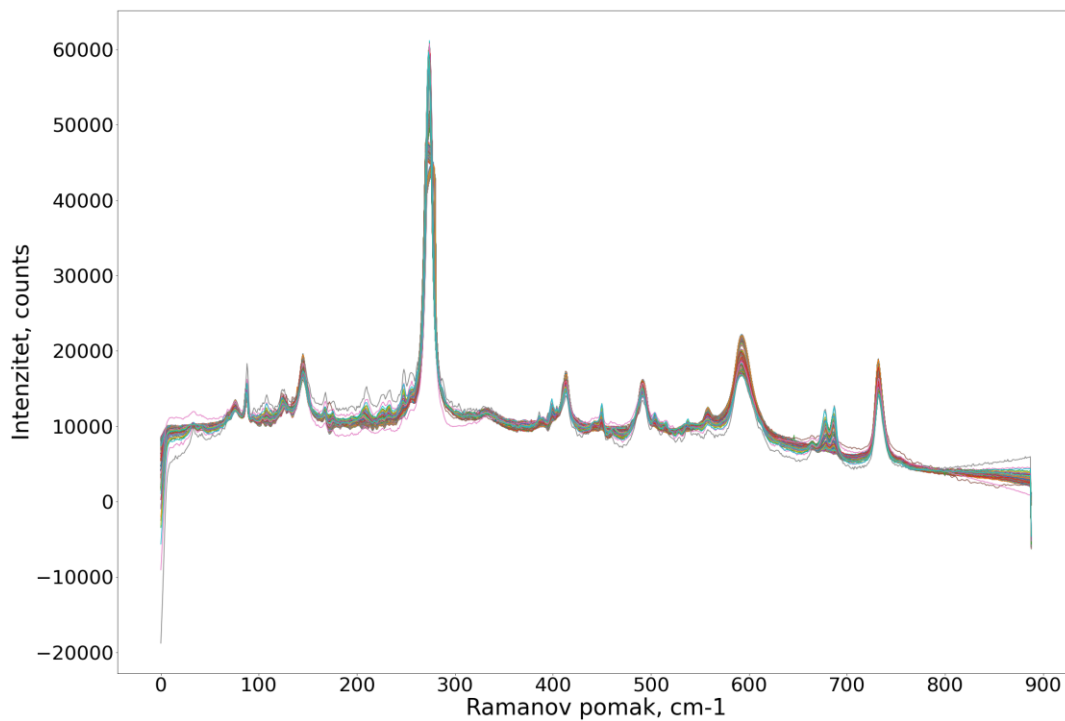
Na slici 42 prikazane su predviđene gustoće suspenzije PLS modela. Mreža lošije predviđa gustoće suspenzije kada su u njoj prisutni kristali. Na x osi označena je stvarna gustoća do je na y osi označena predviđena gustoća suspenzije.



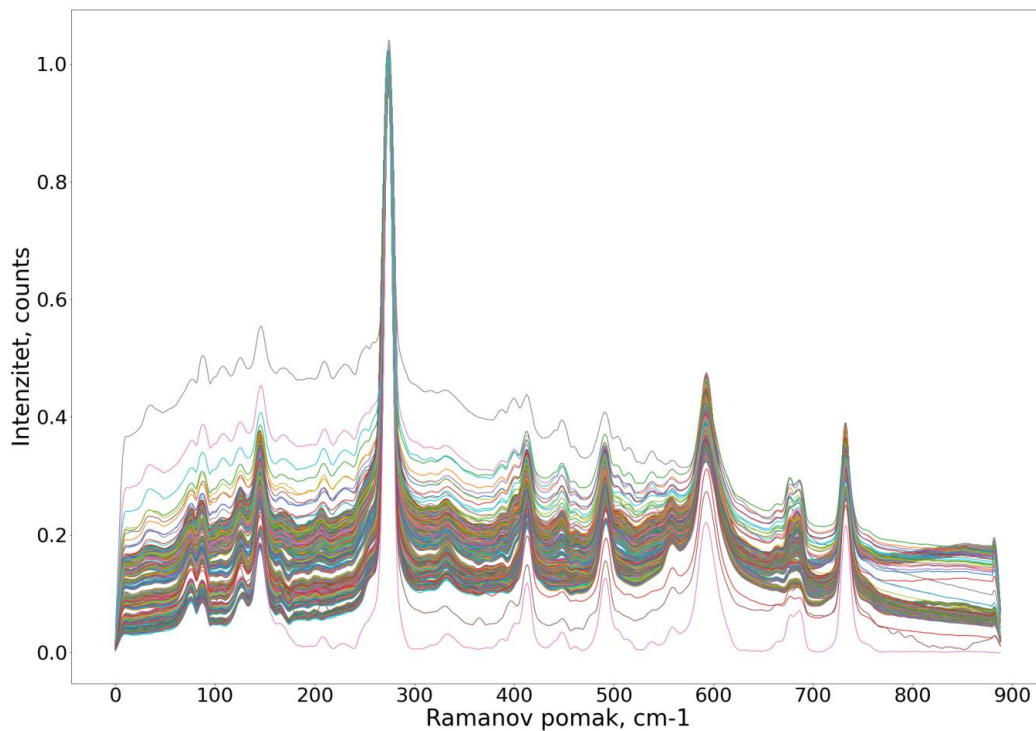
**Slika 42.** Predviđene gustoće suspenzije PLS modelom

Na slici 43 prikazani su podaci X matrice s MSC predobradom podataka. Korigiran je efekt raspršenja na y osi. Spektri su korigirani tako da su oni što bliže referentnom spektru, koji je u ovom slučaju prosjek skupa podataka.

Na slici 44 prikazani podaci X matrice s preobradom podataka koja se sastoji korekcije bazne linije, filtriranje podataka Savitzky-Golay filtrom te „Min-max” normalizacijom. Korekcijom bazne linije uklonjen nagib bazne linije, Svitzky-Golay filtrom (duljina prozora 15, red polinoma 3) uklonjen je šum koji je bio prisutan u podacima, dok je „Min-max“ normalizacijom y os skalirana na vrijednosti 0-1.



**Slika 43.** Podaci X matrice s MSC preobradom podataka



**Slika 44.** Podaci X matrice s preobradom podataka korekcije bazne linije, filtriranje podataka Savitzky-Golay filtrom te “Min-max” normalizacijom

## Unaprijedna neuronska mreža

Promjenom hiperparametara moguće je smanjiti vrijednost RMSE, tj. povećati točnost modela. Potrebno je naći kombinaciju hiperparametara za koji je RMSE minimalan. Potrebno je mijenjati jedan, eventualno dva hiperparametra dok se svi ostali drže konstantnima te izračunati RMSE (tablica 8). Tijekom procesa optimizacije neuronske mreže korišten je ADAM (eng. *Adaptive Moment Estimation*) optimizator jer se pokazao bržim od ostalih, broj epoha iznosi 6000, a funkcija pogreške je korijen srednje kvadratne pogreške.

**Tablica 8.** Optimizacija hiperparametara neuronske mreže

Broj skrivenih slojeva	Broj neurona u prvom skrivenom sloju	Broj neurona u drugom skrivenom sloju	Broj neurona u trećem skrivenom sloju	Veličina podskupa	Prijenosna funkcija	RMSE model za koncentraciju / gkg <sup>-1</sup>	RMSE model za gustoću suspenzije/ gL <sup>-1</sup>
1	5	-	-	32	tanh ReLu sigmoid	0,05 0,14 0,04	0,61 1,40 0,88
	10	-	-	32	tanh ReLu sigmoid	0,09 0,24 0,04	0,51 0,97 0,89
	20	-	-	32	tanh ReLu sigmoid	0,17 0,04 0,08	0,17 0,30 0,92
	40	-	-	32	tanh ReLu sigmoid	0,12 0,05 0,03	0,47 0,14 0,87
2	5	2	-	32	tanh ReLu sigmoid	0,06 0,11 0,02	1,01 0,67 1,60
	10	7	-	64	tanh ReLu sigmoid	0,06 0,11 0,06	0,10 0,74 0,10
	20	14	-	64	tanh ReLu sigmoid	0,07 0,04 0,02	0,67 0,12 0,29
	40	28	-	128	tanh ReLu sigmoid	0,28 <b>0,02</b> 0,10	<b>0,08</b> 0,13 0,30
3	5	3	2	128	tanh ReLu sigmoid	0,09 0,98 0,35	1,10 0,46 1,50
	10	5	2	64	tanh	0,04	1,20

					ReLu sigmoid	0,09 0,23	0,31 0,85
	20	14	9	32	tanh ReLu sigmoid	0,21 0,06 0,15	0,43 0,68 0,94
	40	28	15	32	tanh ReLu sigmoid	0,04 0,04 0,04	0,22 0,22 0,91

Više različitih struktura neuronskih mreža daju minimum RMSE za testni skup za predviđanje koncentracije otopine. Odabrana je neuronska mreža s jednim ulaznim i jednim izlaznim slojem te 2 skrivena sloja. Broj neurona u prvom sloju je 40, broj neurona u drugom sloju je 28, veličina podskupa je 128, dok je prijenosna funkcija ReLu. Ova kombinacija hiperparametara odabrana je zbog minimalne vrijednosti RMSE koji daje ova mreža.

Za mrežu koja predviđa gustoću suspenzije hiperparametri koji daju minimum RMSE za testni skup podataka neuronske mreže su 2 skrivena sloja, broj neurona u prvom sloju je 40, broj neurona u drugom sloju je 28, veličina podskupa je 128, dok je prijenosna funkcija tangens hiperbolična.

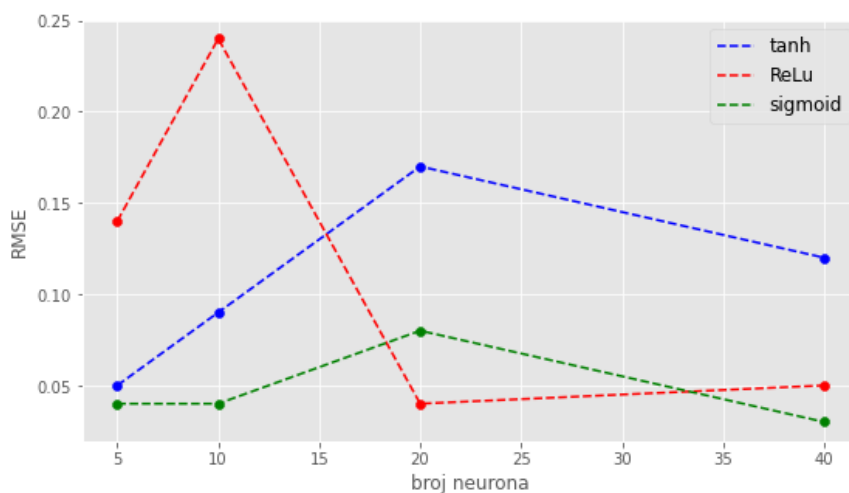
U tablici 9 prikazani su hiperparametri neuronske mreže za predviđanje koncentracije otopine i neuronske mreže za predviđanje gustoće suspenzije.

**Tablica 9.** Hiperparametri neuronske mreže

Parametar		Mreža za predviđanje koncentracije otopine	Mreža za predviđanje gustoće suspenzije
Sloj	Ulazni	1	1
	Skriveni	2	2
	Izlazni	1	1
Broj neurona u sloju	Prvi skriveni	40	40
	Drugi skriveni	28	28
	Izlazni	1	1
Aktivacijska funkcija	Prvi skriveni	ReLu	tanh
	Drugi skriveni	ReLu	tanh
	Izlazni	Linearna	Linearna
Funkcija pogreške		Srednja kvadratna pogreška	Srednja kvadratna pogreška
Optimizator		ADAM	ADAM
Veličina podskupa		128	128
Broj epoha		6000	6000

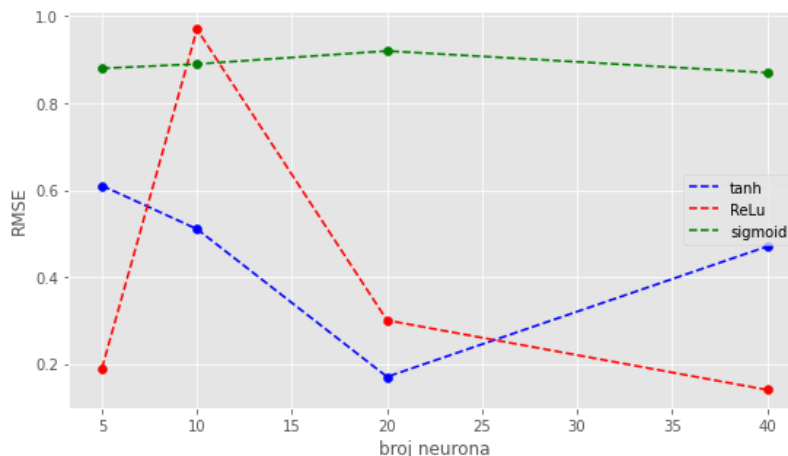
Na slici 45 prikazana je ovisnost RMSE testnog skupa podataka o broju neurona za različite prijenosne funkcije u prvom skrivenom sloju za mrežu koja predviđa koncentraciju otopine. Model s ReLu prijenosnom funkcijom daje veliku vrijednost RMSE za mali broj neurona u skrivenom sloju, dok RMSE vrijednost opada povećanjem broja neurona. Minimum RMSE postiže se za 25 neurona u prvom skrivenom sloju. Modeli s tangens hiperboličnom i sigmoidnom funkcijom imaju mali iznos RMSE za mali broj neurona, RMSE raste za obe funkcije povećanjem broja neurona i zatim opada daljnjim povećanjem broja neurona.

Na slici 46 prikazana je ovisnost RMSE testnog skupa podataka o broju neurona za različite prijenosne funkcije u prvom skrivenom sloju za mrežu koja predviđa gustoću suspenzije. Model s ReLu prijenosnom funkcijom daje veliku vrijednost RMSE za mali broj neurona u skrivenom sloju, te RMSE vrijednost opada povećanjem broja neurona. Model s 20 neurona daje minimum RMSE za tangens hiperboličnu funkciju. Model sa sigmoidnom funkcijom ne pokazuje veliku promjenu u RMSE s promjenom broja neurona.



**Slika 45.** Ovisnost RMSE testnog skupa podataka o broju neurona mreže za predviđanje koncentracije za tri različite funkcije





**Slika 46 .** Ovisnost RMSE testnog skupa podataka o broju neurona mreže za predviđanje gustoće za tri različite funkcije

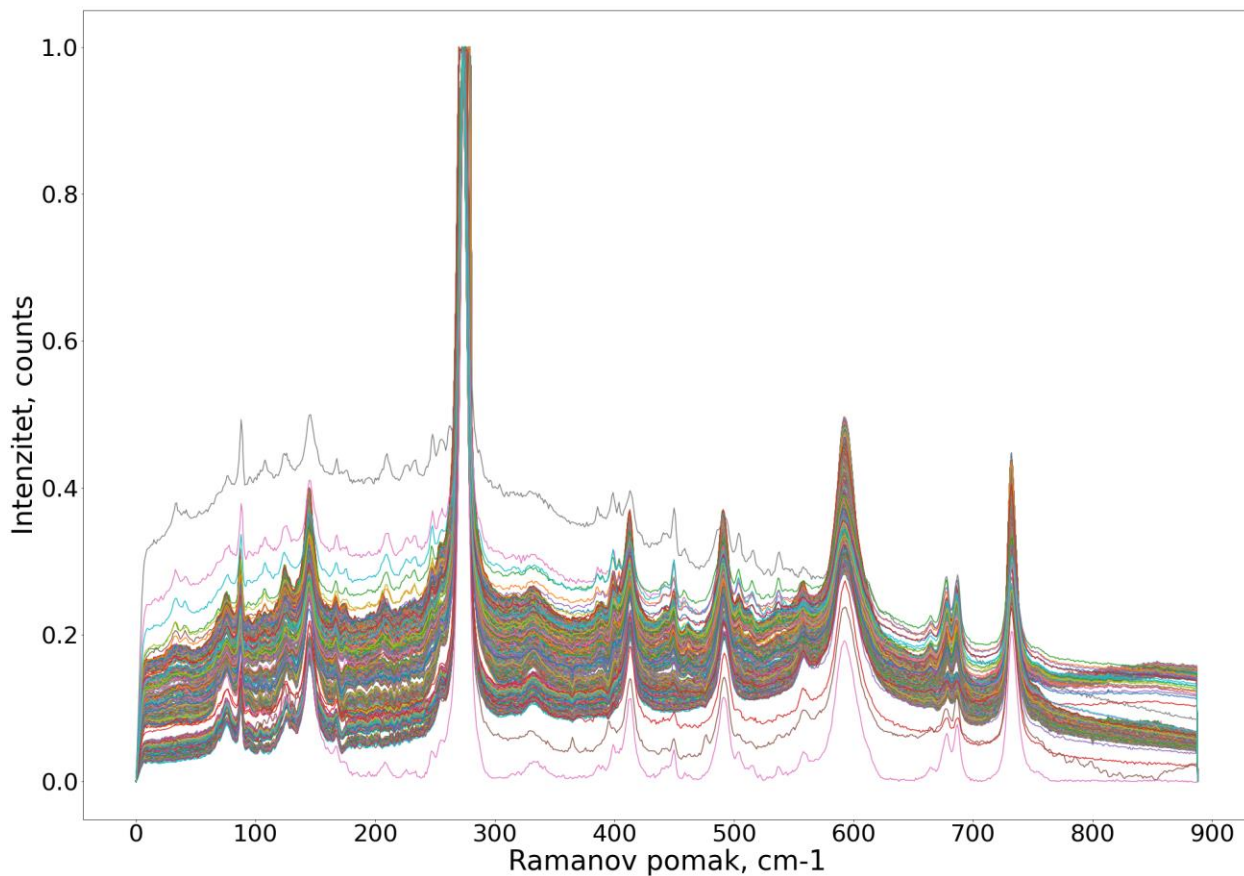
Predobrada podataka može znatno povećati točnost modela. Različite metode predobrade podataka spomenute su u poglavlju o izradi PLS modela. U tablici 10 prikazana je ovisnost RMSE testnog skupa podataka o vrsti predobrade podataka.

**Tablica 10.** Ovisnost RMSE testnog skupa podataka o vrsti predobrade podataka za neuronsku mrežu

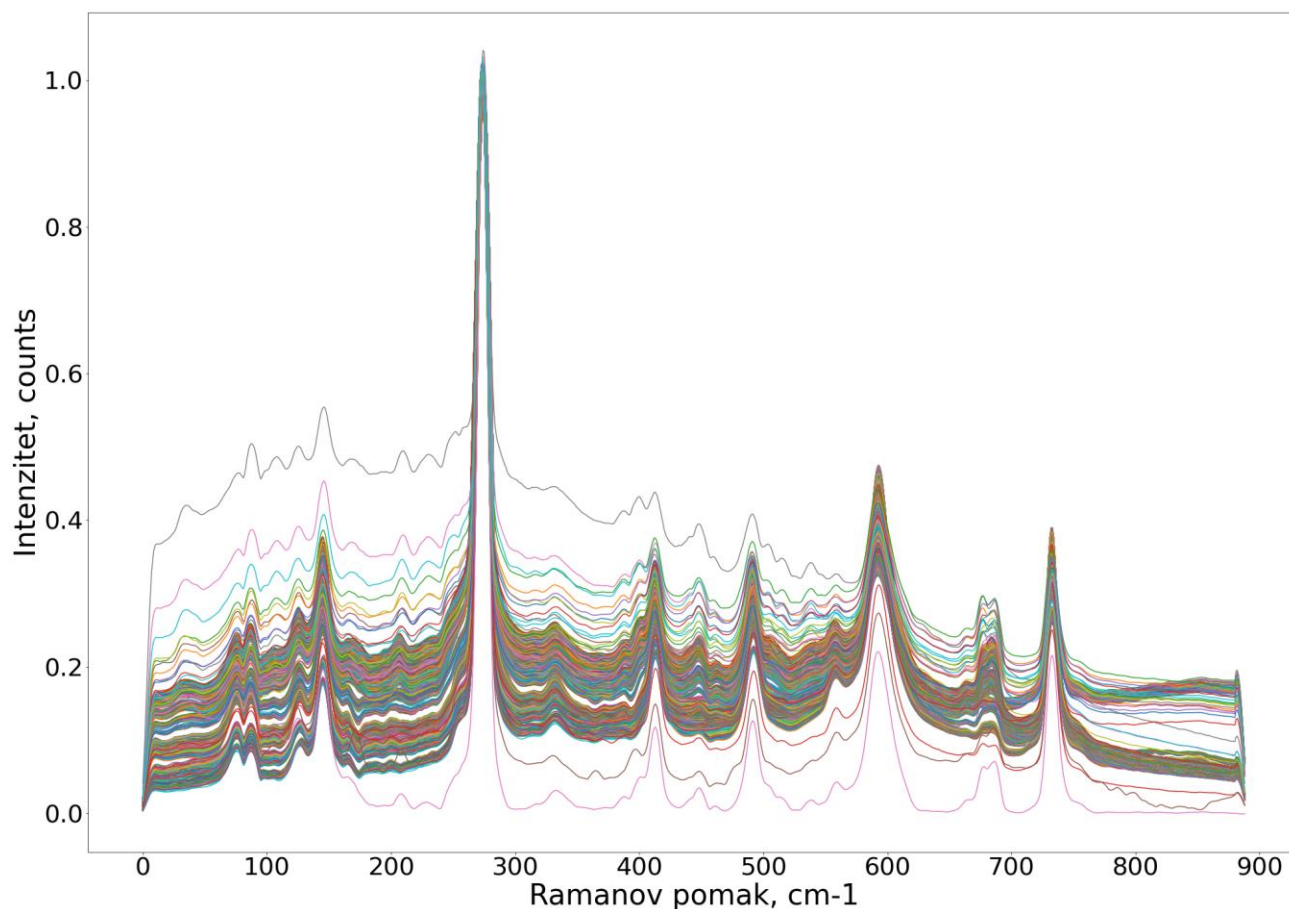
Predobrada	RMSE modela za koncentraciju otopine / gkg <sup>-1</sup>	RMSE modela za gustoću suspenzije / gL <sup>-1</sup>
Bez predobrade	0,45	0,83
Korekcija bazne linije	0,64	0,42
Filtriranje Savitzky – Golay filtrom (window length 15, polyorder 3)	0,56	0,83
“Min-max” normalizacija	0,05	0,36
Uklanjanje dijela spektra (260-285 cm <sup>-1</sup> )	0,52	0,38
Korekcija bazne linije + filtriranje Savitzky – Golay filtrom	0,65	0,56
Korekcija bazne linije + filtriranje Savitzky – Golay filtrom + “Min-max” normalizacija	0,16	<b>0,08</b>
Korekcija bazne linije + “Min-max” normalizacija	<b>0,03</b>	0,54
Druga derivacija	0,15	0,61
PCA	0,130	0,25
MSC	0,09	0,98
Korištenje samo vrpce karakterističnih za kristal	-	1,07

Na slici 47 prikazani su obrađeni podaci X matrice upotrebom korekcije bazne linije i Savitzky-Golay filtra. Može se uočiti da je uklonjen nagib bazne linije i ona je ravnija nego na slici 32. Savitzky-Golay filterom (duljina prozora 15, red polinoma 3) uklonjen je šum koji je bio prisutan u podacima, spektri izgledaju „izglađenije“ nego neobrađeni podaci prikazani na slici 32.

Na slici 48 prikazani su obrađeni podaci X matrice upotrebom korekcije bazne linije i Savitzky-Golay filtra te „Min-max“ normalizacijom. Korekcijom bazne linije uklonjen nagib bazne linije, Savitzky-Golay filterom (duljina prozora 15, red polinoma 3) uklonjen je šum koji je bio prisutan u podacima, dok je „Min-max“ normalizacijom y os skalirana na vrijednosti 0-1.



**Slika 47.** Podaci X matrice s korekcijom bazne linije i “Min-max” normalizacijom



**Slika 48.** Obradeni podaci X matrice upotrebom korekcije bazne linije, Savitzky-Golay filtra i „Min-max“ normalizacijom

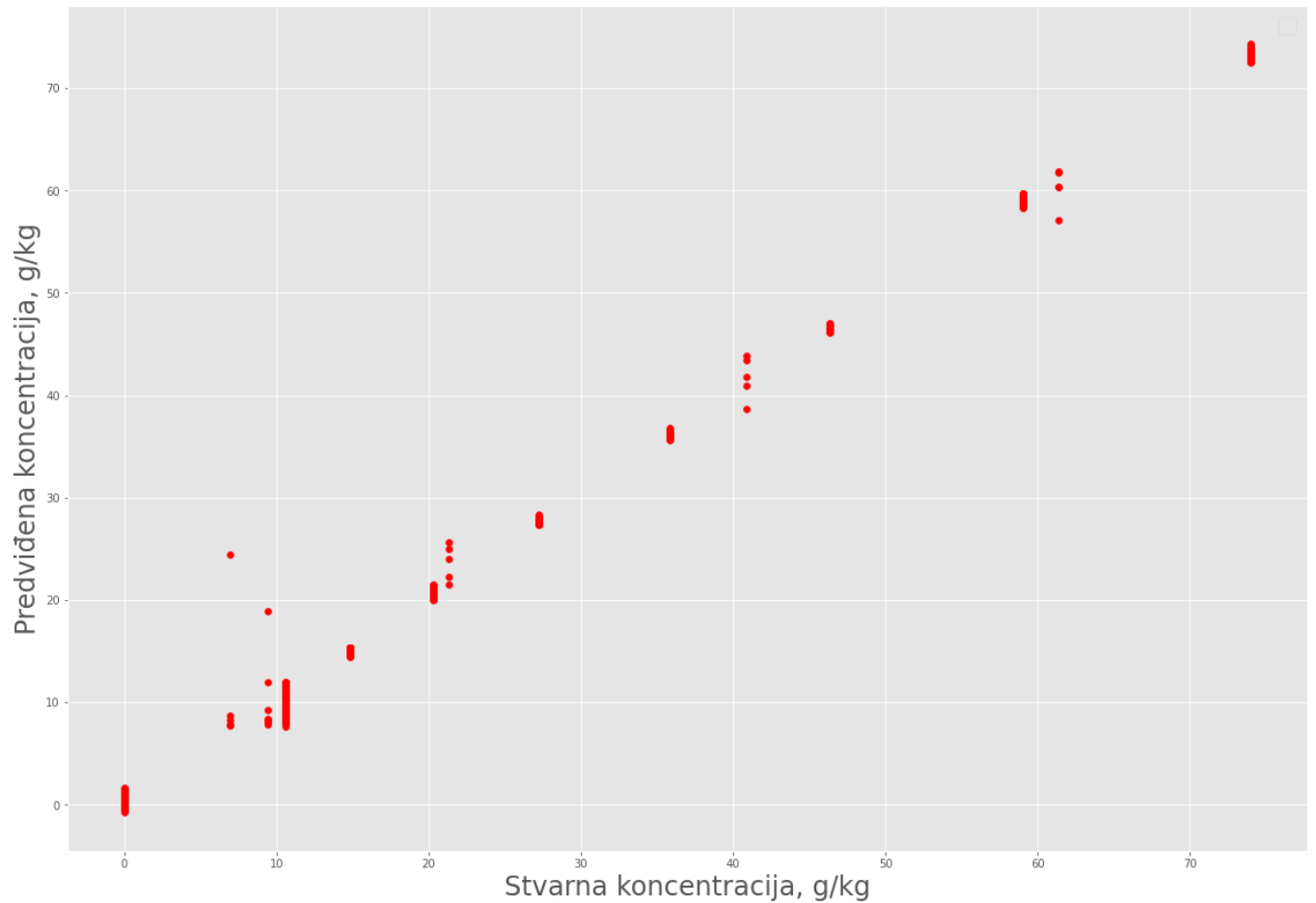
U tablici 11 prikazani su kvadrat koeficijenta korelacije,  $R^2$  i korijen srednje kvadratne pogreške, RMSE za test skup podataka za dvije neuronske mreže od kojih prva predviđa koncentracije otopine a druga predviđa gustoće suspenzije.

Vrijednost RMSE od 0,17 % u ovom slučaju treba uzeti sa zadržkom jer se na slici 50 može uočiti da model jako dobro predviđa gustoće suspenzije kada one iznose nula, a griješi kada one imaju vrijednost veću od nula, tj. kada su u sustavu prisutni kristali. Pošto je u skupu podataka izmjereno puno Ramanovih spektara otopine, a malo ramanovih spektara suspenzije, statistika modela pokazuje malu pogrešku, ali grafički se može uočiti da model ne predviđa dobro gustoće suspenzija.

**Tablica 11.** kvadrat koeficijenta korelacije,  $R^2$  i korijen srednje kvadratne pogreške, RMSE za test skup podataka za kalibracijske modele

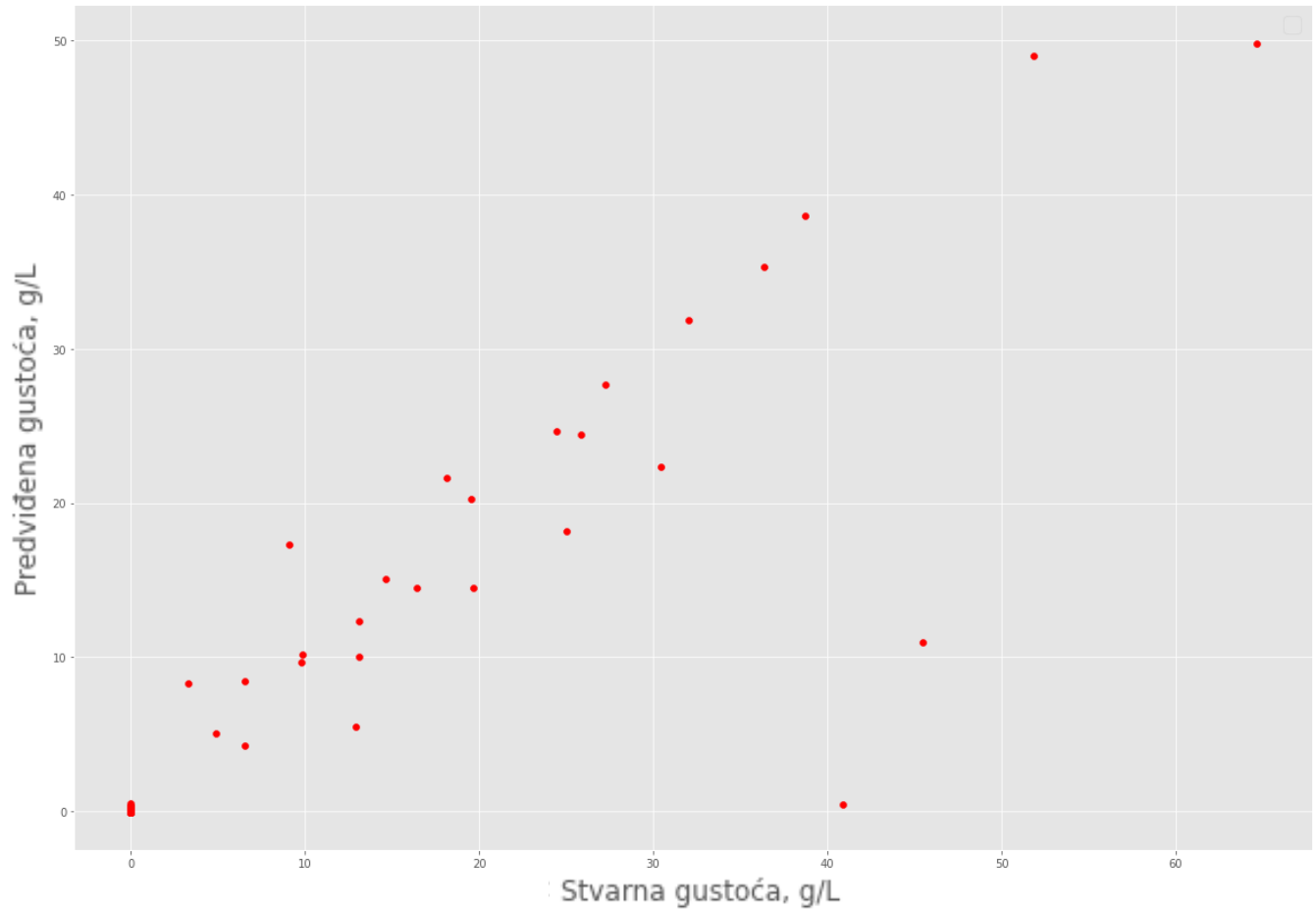
Pokazatelj	Kalibracijski model za koncentraciju otopine	Kalibracijski model za gustoću suspenzije
$R^2$	0,99	0,98
RMSE %	1,04 %	0,17 %

Na slici 49 prikazani su izlazi neuronske mreže koja predviđa koncentraciju otopine. Mreže lošije predviđa koncentracije otopina kada je u sustavu prisutan kristal, tj. kada je izmjeren ramanov spektar suspenzija. Predviđanja modela označena su crvenim točkama. Što su one raspršenije po y osi, to model lošije predviđa tu koncentraciju.



**Slika 49.** Izlazi neuronske mreže modela za koncentraciju otopine

Na slici 50 prikazani su izlazi neuronske mreže koja predviđa gustoću suspenzije. Predviđanja modela označena su crvenim točkama. Može se uočiti da model vrlo dobro predviđa gustoću suspenzije koja je jednaka nuli, tj. kada u sustavu nisu prisutni kristali. Lošije predviđa gustoće suspenzije različite od nule, tj. kada je u sustavu prisutan kristal.



**Slika 50.** Izlazi neuronske mreže modela za gustoću suspenzije

## Zaključak

U ovom radu razvijeni su kalibracijski modeli za predviđanje koncentracije ceritiniba i gustoće suspenzije ceritiniba forme A u sustavu otapala 90 % aceton - 10 % voda.

Eksperimentalno su određene topljivosti ceritiniba forme A u sustavima otapala s različitim omjerima acetona i vode. Na temelju eksperimentalnih podataka moguće proračunati topljivosti za ostale omjere otapala te odabrati sustav za daljnu provedbu eksperimenata.

Poznavanje koncentracije otopine u realnom vremenu bitno je prilikom provedbe procesa kristalizacije pri konstantnom prezasićenju. Kalibracijski model daje informaciju o koncentraciji otopine u stvarnom vremenu što omogućava kontinuirano vođenje procesa kristalizacije.

Model razvijen umjetnom unaprijednom neuronskom mrežom pokazao se bolji od modela razvijenog parcijalnom regresijom metodom najmanjih kvadrata za predviđanje koncentracije otopine, kao i za predviđanje gustoće suspenzije.

## Literatura

- [1] A. S. Myerson, D. Erdemir, A. Y. Lee, Handbook of industrial crystallization, Second Edition, Elsevier Science & Technology Books, 2001., str. 16
- [2] A. J. Hickey, D. Ganderton, Pharmaceutical Process Engineering, Second Edition, Informa Healthcare, New York, 2016., str. 92,202
- [3] W. Beckmann, Crystallization: Basic Concepts and Industrial Applications, Wiley-VCH Verlag GmbH & Co. , Weinheim, 2013. str. 8,16-21,53-58
- [4] Hilfiker, Polymorphism: in the Pharmaceutical Industry, Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim, 2006., str 1,97-102
- [5] D. R. Weyna, Crystal Engineering of Multiple Component Crystal Forms of Active Pharmaceutical Ingredients, Phd dissertation, University of South Florida 2011., str. 9-20
- [6] H. Herrmann, H. Bucksch, Process Understanding, Dictionary Geotechnical Engineering/Wörterbuch GeoTechnik., Springer, Berlin, Heidelberg, 2014.
- [7] A. Mersmann, Crystallization Technology Handbook, Second Edition, Marcel Dekker, Inc. ,New York, 2001. str. 5-19
- [8] I.Houson, Process Understanding: For Scale-Up and Manufacture of Active Ingredients, Wiley-VCH, Weinheim,2011, str. 200-208
- [9] D. Šahnić, Kvantitativno praćenje reakcije sinteze i kristalizacije omeprazola ramanovom spektroskopijom in-line, doktorska disertacija, Sveučilište u Zagrebu, 2017., str. 20
- [10] S.Z. Mohd Noor, D.M. Camacho, C. Yun Ma, T. Mahmud, Effect of Crystallization Conditions on the Metastable Zone Width and Nucleation Kinetics of *p*-Aminobenzoic Acid in Ethanol. Chem. Eng. Technol., 43 (2020) 1105-1114.
- [11] P. Augustijns, M.E. Brewster, Solvent Systems and Their Selection in Pharmaceutics and Biopharmaceutics, u R. T. Borchardt, C. R. Middaugh (ur) Biotechnology: Pharmaceutical Aspects, Springer, New York, 2007., str. 16-19
- [12] H. Alatalo, Supersaturation Controlled Crystallization , Graduate Theses and Dissertations, (2010), str 34.
- [13] A.T. Aina, In situ monitoring of pharmaceutical crystallisation, PhD thesis, University of Nottingham. 2012., str. 8
- [14] N. Bolf, H. Dorić, Procesna analitička tehnologija, Kem. Ind. 67 (9-10) (2018) 466–469.
- [15] Y. Wu, A Study of The Deep Learning-based Monitoring and Efficient Numerical Modeling Methodologies for Crystallization Processes, Doktorska disertacija, The University of Western Ontario, 2021., str. 6-9

- [16] N. Bolf, Novosti i napredak u procesnoj analitičkoj tehnologiji, *Kem. Ind.* 70 (7-8) (2021) 476–479.
- [17] A. Chanda, A.M. Daly, D.A. Foley, M. A. LaPack, S.Mukherjee, J. D. Orr, G. L. Reid, D. R. Thompson, H. W. Ward, Industry Perspectives on Process Analytical Technology: Tools and Applications in API Development, *Org. Process Res. Dev.* (2015), 19, 1, 63–83
- [18] E. Smith, G. Dent, *Modern Raman Spectroscopy: A practical Approach*, John Wiley & Sons Ltd, West Sussex, 2005., str. 3,56
- [19] G. Turrell, J. Corset, *Raman Microscopy: Development and Applications*, Elsevier Ltd. 1996, Malta. str.1,201
- [20] K. A. Bakeev, *Process Analytical Technology: Spectroscopic Tools and Implementation Strategies for the Chemical and Pharmaceutical Industries*, Blackwell Publishing, Oxford, 2005., str. 97-102,141-144,200-202
- [21] D.A.Long, *The Raman Effect: A Unified Treatment of the Theory of Raman Scattering by Molecules*, John Wiley & Sons Ltd, West Sussex, 2002., str. 55
- [22] H.A. Szymanski, *Raman Spectroscopy: Theory and Practice*, Plenum Press, New York, 1970. str. 3
- [23] N.Tarcea, T. Frosch, P. Roesch, M. Hilchenbach, T. Stuffer, S. Hofer, H. Thiele, R. Hochleitner, J. Popp, *Raman Spectroscopy: A Powerful Tool for in situ Planetary Science*, *Space Sci Rev* (2008) 135: 281–292
- [24] [https://www.gmp.ch/pdf/Raman\\_what\\_is.pdf](https://www.gmp.ch/pdf/Raman_what_is.pdf) (pristup 30.5.2022.)
- [25] R.Kramer, *Chemometric Techniques for Quantitative Analysis*, Marcel Dekker, New York, 1998., str. 1-3
- [26] P. Gemperline, *Guide to Chemometrics, Second Edition*, Taylor & Francis, Boca Raton, 2006., str. 2,148-149
- [27] R. Brereton, *Chemometrics: data driven extraction for science, Second Edition*, John Wiley & Sons Ltd., Hoboken, NJ, 2018, str. 163-174
- [28] F.T. Chau, Y.Z Liang, J. Gao, X.G. Shao, *Chemometrics: From Basics to Wavelet Transform*, John Wiley & Sons Inc., Hoboken, 2004. str. 4-8
- [29] K. Varmuza, P. Filmoser, *Introduction to Multivariate Statistical Analysis in Chemometrics*, Taylor & Francis, Boca Raton, 2009.
- [30] G. Dreyfus, *Neural networks: Methodology and applications*, *Neural Networks: Methodology and Applications*, Springer, Berlin, 2005, str. 3-4
- [31] K. Mehrotra, C. K. Mohan, S. Ranka, *Elements of Artificial Neural Networks*, A Brandford



Book, 1996., str. 1

[32] N. González-Viveros, P. Gómez-Gil, J. Castro-Ramos, H.H. Cerecedo-Núñez, On the estimation of sugars concentrations using Raman spectroscopy and artificial neural networks. *Food chemistry*, 352 (2021) 129375., str 2

[33] F.Lussier, V. Thibault, B. Charron, G.Q. Wallace, J.-F. Masson, Deep learning and artificial intelligence methods for Raman and surface-enhanced Raman scattering, *Trends in Analytical Chemistry*. 124 (2020) 2-6.

[34] A. C. Olivieri, *Introduction to multivariate calibration: A practical approach*, Springer, Cham, 2018., str. 212-214

[35] F. Chollet, *Deep learning with Python*, Manning Publications Co., New York, 2018.,str. 58

[36] <https://pubchem.ncbi.nlm.nih.gov/compound/Ceritinib#section=2D-Structure> (pristup 27.8.2022.)