

UNIVERZITET U BEOGRADU

FIZIČKI FAKULTET

Lea I. Lenhardt

**PARALELNA FAKTORSKA ANALIZA
FLUORESCENTNIH SVOJSTAVA
VIŠEKOMPONENTNIH SISTEMA**

doktorska disertacija

Beograd, 2014

UNIVERSITY OF BELGRADE

FACULTY OF PHYSICS

Lea I. Lenhardt

**PARALLEL FACTOR ANALYSIS OF THE
FLUORESCENT PROPERTIES OF
MULTICOMPONENT SYSTEMS**

Doctoral Dissertation

Belgrade, 2014

Članovi komisije

1. *Prof. Dr Miroslav Dramićanin,*

Naučni savetnik Instituta za nuklearne nauke Vinča,
redovni profesor Fizičkog fakulteta Univerziteta u Beogradu

2. *Prof. Dr Ljubiša Zeković,*

redovni profesor Fizičkog fakulteta Univerziteta u Beogradu.

3. *Prof. Dr Ivan Belča,*

vanredni profesor Fizičkog fakulteta Univerziteta u Beogradu.

4. *Prof. Dr Milorad Kuraica,*

redovni profesor Fizičkog fakulteta Univerziteta u Beogradu.

Datum odbrane _____

(dan, mesec, godina)

Doktorska disertacija urađena je pod rukovodstvom mentora prof. dr Miroslava D. Dramićanina, naučnog savetnika INN „Vinča“ i redovnog profesora Fizičkog fakulteta Univerziteta u Beogradu i prof. dr Ljubiše Zekovića redovnog profesora Fizičkog fakulteta Univerziteta u Beogradu. Iskreno se zahvaljujem profesoru Dramićaninu na nesebičnoj podršci i velikoj pomoći koju mi je pružio tokom postdiplomskih studija, kao i na velikom doprinosu svojim sugestijama i idejama prilikom realizacije doktorske disertacije u svim njenim fazama. Posebnu zahvalnost dugujem i profesoru Zekoviću, na veoma korisnim savetima i podršci.

Zahvaljujem se prof. dr Rasmus Bro-u, redovnom profesoru Fakulteta prirodnih nauka Univerziteta u Kopenhagenu, na podršci, velikoj pomoći i stručnim savetima prilikom izrade ove disertacije.

Posebno se zahvaljujem dr Ivani Zeković, istraživaču saradniku INN „Vinča“, na neprocenjivoj pomoći, stručnim savetima i podršci pruženoj tokom izrade disertacije.

Takođe bih želela da se zahvalim svim kolegama koji su na bilo koji način doprineli procesu izrade ove teze.

Na kraju bih želela da se zahvalim svojoj porodici, prijateljima i Acku na nesebičnoj pomoći i podršci.

Lea Lenhardt

PARALELNA FAKTORSKA ANALIZA FLUORESCENTNIH SVOJSTAVA VIŠEKOMPONENTNIH SISTEMA

Sažetak

Kombinacija optičkih spektroskopskih tehnika i odgovarajućih metoda za analizu i modeliranje spektralnih podataka se pokazala kao veoma obećavajuća tehnika za karakterizaciju i analizu organskih i neorganskih kompleksnih sistema. Razumevanje fizičkih fenomena koji se javljaju u kompleksnim sistemima predstavlja bitan preduslov za primenu spektroskopskih tehnika u različitim oblastima nauke. U ovoj tezi su izmerene fluorescentne karakteristike četiri kompleksna sistema: tri organska (rastvor vode i aminokiselina, med i tkivo dojke) i jedan neorganski (kompleksni sistem fosfora na bazi retkih zemalja). Dobijeni fluorescentni podaci su modelirani primenom paralelne faktorske analize. Izgrađeni modeli su omogućili definisanje fluorescirajućih komponenti (fluorofora) prisutnih u sistemima, pri čemu je model kao rezultat dao čiste spektre svake pojedinačne fluorofore, kao i njihove relativne koncentracije u svakom od uzoraka sistema. U neorganskom sistemu i rastvoru aminokiselina broj fluorofora i njihove koncentracije su unapred bile poznate i na taj način je testirana uspešnost modela kao i njegova sposobnost predikcije prisustva i koncentracija fluorofora u nepoznatim uzorcima. Pokazano je da je dobijeni model veoma uspešan, sa odstupanjem od samo 0.036% od stvarne koncentracije kod neorganskog sistema. Za sistem meda, model je pokazao da je u sistemu prisutno 6 fluorofora i dobijeni su čisti spektri svake od njih kao i njihove relativne koncentracije u svakom pojedinačnom uzorku. Dalje analize spektara meda su ukazale da je na osnovu koncentracija fluorofora moguće odrediti botaničko poreklo meda. Spektri meda su zatim analizirani metodom glavnih komponentata kao i diskriminantnom analizom parcijalno najmanjih kvadrata. Rezultati su pokazali veliku uspešnost primene metode fluorescentne spektroskopije za klasifikaciju i autentifikaciju uzoraka meda. Proračunom modela tkiva dojke je ustanovljeno da postoje 4 dominantne fluorofore. Utvrđeno je da je na osnovu njihovih koncentracija moguće sa tačnošću od 100% odrediti vrstu promena na tkivu (benigno ili maligno). Rezultati su potom poređeni sa rezultatima analize potpunih vektora primenjenih na neobrađene spektre, na osnovu kojih je pokazano da se sa istim podacima ne može dobiti podjednako uspešan model

(greška klasifikacije 35.71%). Međutim, pokazano je da se primenom drugog režima merenja spektara i određenom aritmetičkom manipulacijom podataka mogu dobiti modeli istog kvaliteta. Dobijeni rezultati u ovoj tezi su pokazali da fluorescentna spektroskopija u kombinaciji sa multivarijantnim analizama ima veliki potencijal u karakterizaciji, modeliranju i analizi višekomponentnih sistema.

Ključne reči: višekomponentni sistemi, fluorescentna spektroskopija, paralelna faktorska analiza, analiza glavnih komponentata

Naučna oblast: Fizika

Uža naučna oblast: Primenjena fizika

UDK broj: 543.42:535.33(043.3)

PARALLEL FACTOR ANALYSIS OF THE FLUORESCENT PROPERTIES OF MULTICOMPONENT SYSTEMS

Abstract

The combination of optical spectroscopic techniques and appropriate methods for analysis and modeling of spectral data proved to be a very promising technique for the characterization and analysis of organic and inorganic complex systems. Understanding of physical phenomena that occur in complex systems is an essential prerequisite for the application of spectroscopic techniques in various fields of science. Fluorescence characteristics of four complex systems: three organic (solution of water and amino acids, honey and breast tissue) and one inorganic (a complex system of phosphorus-based rare earth) were measured in this thesis. The obtained fluorescence data were modeled using parallel factor analysis. The constructed models have enabled definition of the fluorescing components (fluorophores) present in the system, wherein the model as a result gave pure spectra of the individual fluorophore as well as their relative concentrations in each sample of the system. In the inorganic system and solution of amino acids, the number of fluorophores and their concentrations were known in advance, and thus the performance of the model was tested as well as its ability to predict the presence and concentrations of fluorophores in the unknown samples. It was shown that the resulting model is very successful, with a deviation of only 0.036% of the actual concentration in inorganic systems. For the system of honey, the model showed that 6 fluorophores are present in the system and their pure spectra were obtained, as well as their relative concentrations in each sample. Further analysis of the spectra of honey have demonstrated that based on the concentrations of the fluorophores the botanical origin of honey can be determined. Spectra of honey were then analyzed using principal component analysis and discriminant partial least squares. Results have shown great success in application of fluorescence spectroscopy methods for classification and authentication of honey samples. Based on calculated model of the breast tissue it was determined that there are four dominant fluorophores. It was established that based on their concentrations it possible to determine the type of change in the tissue (benign or malignant) with an accuracy of 100%. The results were then compared with the results of support vector machine analysis applied to the raw spectra, from which it was shown that using the same data successful model can not be obtained (classification error 35.71%). However, it has been shown that with

application of measuring spectra in another mode and using specific arithmetic manipulation of the data, models of the same quality may be obtained. The results obtained in this thesis have shown that fluorescence spectroscopy in combination with multivariate analysis has great potential in the characterization, modeling and analysis of multi-component systems.

Key words: multi-component systems, fluorescent spectroscopy, parallel factor analysis, principal component analysis

Scientific field: Physics

Major in: Applied Physics

UDC number: 543.42:535.33(043.3)

SADRŽAJ

1. UVOD	1
2. MULTIVARIJANTNA ANALIZA	3
2.1 BILINEARNI MODELI – PCA I PLS	3
2.1.1 PCA	3
2.1.2 PLS	5
2.1.3 Diskriminantna analiza parcijalno najmanjih kvadrata – PLS-DA	5
2.1.4 Primena PCA i PLS metoda u fluorescentnoj spektroskopiji	6
2.2 METODA ZA ANALIZU VIŠEDIMENZIONALNIH PODATAKA - PARAFAC	6
2.2.1 Trodimenzionalni PARAFAC model	8
2.2.2 PARAFAC model u notaciji Kronecker proizvoda	14
2.2.3 PARAFAC model u tenzorskoj notaciji	14
2.2.4 PARAFAC model u vektorskoj i Kronecker notaciji	14
2.2.5 PARAFAC model u notaciji Khatri-Rao proizvoda	15
2.2.6 Rotaciona sloboda i unikatnost trodimenzionalnih PARAFAC modela	17
2.3 PARAFAC ALGORITAM	18
2.3.1 Tehnike optimizacije	19
2.3.2 PARAFAC algoritam naizmenično najmanjih kvadrata	22
2.3.3 Inicijalizacija PARAFAC – ALS	25
2.3.4 Proračun skorova novih uzoraka	26
2.3.5 Tretiranje matrica kojima nedostaju podaci	28
2.3.6 Uslovljavanje nenegativnosti	29
2.3.7 Validacija i dijagnostika PARAFAC modela	31
2.3.8 Test-set i unakrsna validacija	33
2.3.9 Odabir broja komponenti modela	36
2.3.10 Grafik svojstvenih vrednosti faktora	36
2.3.11 Split-half analiza	38
2.3.12 Analiza ostatka	39
2.3.13 Predprocesiranje podataka	41
2.3.14 Vizuelizacija rezultata	42
2.4 METODA POTPORNIH VEKTORA	43
3. FLUORESCENTNA SPEKTROSKOPIJA	48
3.1 ELEKTROMAGNETNO ZRAČENJE I NJEGOVA INTERAKCIJA SA MOLEKULIMA	48
3.1.1 Interakcija elektromagnetnog zračenja i molekula	49
3.1.2 Molarni koeficijent apsorpcije	50
3.1.3 Apsorpcija i energetska stanja	51
3.1.4 Tipovi elektronskih prelaza kod molekula	53
3.2 LUMINESCENCIJA	54
3.2.1 Interna konverzija	55
3.2.2 Fluorescencija	56
3.2.3 Unutarsistemski prelazi	58
3.2.4 Fosforescencija i neradijativna relaksacija	58
3.2.5 Eksterna konverzija i gašenje usled sudara	58
3.2.6 Kvantni prinos i intenzitet fotoluminescencije	59
3.2.7 Ekscitacioni i emisioni spektri	59
3.3 LUMINESCENTNI MATERIJALI	60
3.3.1 Organski luminescentni uzorci - fluorofore	60
3.3.2 Neorganski luminescentni materijali	65
3.4 UREĐAJ ZA MERENJE FLUORESCENCIJE	73

3.4.1 Izvor elektromagnetnog zračenja	74
3.4.2 Monohromator	76
2.4.3 Detektor	78
3.4.4 Merenje luminescentnih spektara i vremena života pobuđenog stanja	80
4. MATERIJALI I METODE.....	85
4.1 PRIPREMA UZORAKA.....	85
4.1.1 Rastvor vode i aminokiselina	85
4.1.2 Med	86
4.1.3 Tkivo dojke.....	86
4.1.4 Neorganski materijali.....	87
4.2 MERENJE FLUORESCENTNIH SPEKTARA	88
4.2.1 Ekscitaciono-emisione matrice (EEM)	91
4.2.2 Sinhroni fluorescentni spektri	92
4.2.3 Emisioni spektri i vreme života	92
4.3 ANALIZA I MODELIRANJE DOBIJENIH SPEKTARA.....	93
4.3.1 Priprema podataka za analizu	93
4.3.2 Proračun i analiza PARAFAC modela.....	96
4.3.3 Proračun i analiza PCA modela	97
4.3.4 Proračun i analiza PLS-DA modela.....	98
4.3.5 Proračun i analiza SVM modela	99
4.3.6 Proračun zapremine spektralnih domena	99
5. REZULTATI I DISKUSIJA.....	101
5.1 RASTVOR VODE I AMINOKISELINA	102
5.2 MED	106
5.2.1 Analiza EEM-a.....	106
5.2.2 Analiza sinhronih fluorescentnih spektara	114
5.2.3 Analiza infracrvenih spektara	120
5.3 KANCER DOJKE.....	122
5.3.1 PARAFAC analiza EEM-a.....	122
5.3.2 SVM analiza EEM-a i sinhronih spektara.....	127
5.4 KOMPLEKSNI SISTEMI FOSFORA NA BAZI RETKIH ZEMALJA	132
6. ZAKLJUČAK.....	137
7. LITERATURA.....	141
8. PRILOG	150
PRILOG A	150
A.1 Kronecker, Hadamard i Khatri-Rao proizvod.....	150
A.2 Koncept linearnosti modela.....	153
A.3 Rang i k-Rang matrice.....	154
A.4 Singularna dekompozicija matrice	156
8. BIOGRAFIJA.....	158
IZJAVA O AUTORSTVU.....	160
IZJAVA O ISTOVETNOSTI ŠTAMPANE I ELEKTRONSKE VERZIJE DOKTORSKOG RADA	161
IZJAVA O KORIŠĆENJU	162

1. Uvod

Razumevanje i modeliranje fizičkih procesa i interakcija u kompleksnim sistemima predstavlja osnovni preduslov za praktičnu primenu spektroskopskih metoda u raznim oblastima nauke poput fizike, biologije, medicine, biohemije itd. Kompleksni sistemi predstavljaju veliki izazov za ispitivanje i analizu usled prisustva više komponentata u sistemu kao i njihove međusobne interakcije. Iz tog razloga svaki kompleksan sistem ima svoj specifični „fingerprint“ spektar zbog prisustva više komponentata koje utiču na njegove optičke osobine. Upravo ova karakteristika predstavlja osnovu građenja jedinstvenog modela kompleksnog sistema. Uređaji koji se danas koriste za spektroskopska merenja su sofisticirani i precizni, i daju kvalitetne rezultate merenja (velika osetljivost, visoka rezolucija, itd.). Međutim, njihova analiza i interpretacija predstavlja svojevrsan izazov u cilju efikasne ekstrakcije najznačajnijih informacija iz velike grupe podataka.

Poslednjih decenija optičke spektroskopske tehnike su pokazale da imaju ogroman potencijal primene u različitim oblastima nauke. Važno je istaći da su pomenute metode brze i imaju potencijal da budu neinvazivne, a samim tim i veću mogućnost primene u odnosu na standardne metode. Takođe, ekspanzija primene ovih tehnika je uslovila razvoj odgovarajućih statističkih i matematičkih alata za obradu dobijenih spektroskopskih podataka. To je dovelo do povećanja primene spektroskopskih metoda u raznim naučnim oblastima poput: industrije [He *et al.* 2003], farmacije [Kauppinen *et al.* 1993], ispitivanja hrane [Munck *et al.* 1998], zaštite životne sredine [Silverman 1993] i biohemije [Plugge and van der Vlies 1993].

Jedna od najrazvijenijih i dosta primenjivanih tehnika za optičku karakterizaciju kompleksnih sistema je fluorescentna spektroskopija. Prednost ove metode su pored brzine i pouzdanosti, visoka senzitivnost [Li *et al.* 2003]. Veliku primenu ova metoda je našla u oblasti medicine, biologije, prehrambene industrije itd. [Karoui *et al.* 2007a, Lenhardt *et al.* 2014, Zeković *et al.* 2012, Alfano *et al.* 1987, Dramićanin *et al.* 2005, Dramićanin *et al.* 2012, Sterenborg *et al.* 2012, Lenhardt *et al.* 2013, Zeković *et al.* 2014]. Uočeno je da postoji linearna veza između signala fluorescencije uzorka i

koncentracije određene supstance u njemu. Ovaj fenomen je omogućio pravljenje predikcionih modela na osnovu promene amplitude signala. Međutim, merenjem samo jednog fluorescentnog spektra pokazano je da se ne obuhvata celokupna kompleksnost sistema jer često postoji više fenomena koji utiču na signal. Iz tog razloga se kod višekomponentnih sistema primenjuju složenije fluorescentne spektroskopske tehnike, ekscitaciono-emisione matrice (EEM) i sinhrona fluorescentna spektroskopija, sa ciljem da se detektuje najveći mogući broj spektralnih karakteristika ispitivanog sistema. Analizom dobijenih EEM odgovarajućim multivarijantnim metodama moguće je izdvojiti uticaj pojedinačnih supstanci unutar merenog sistema, odrediti njihovu relativnu koncentraciju, izdvojiti pojedinačne spektre svake supstance koja fluorescira u posmatranom sistemu itd. Multivarijantna tehnika koja se najčešće koristi u ove svrhe je paralelna faktorska analiza (eng. PARAllel FACtor Analysis - PARAFAC). Oblasti u kojima je ova metoda imala veliki doprinos su biologija i medicina. Razlog ovome je autofluorescencija bioloških uzoraka o kojoj će kasnije biti više reči. PARAFAC analiza EEM fluorescentnih spektara se pokazala uspešnom na raznim poljima primene [Bro 1997] kao što je klasifikacija i karakterizacija Sherry sirćeta [Callejón *et al.* 2012], analiza šećera [Bro 1999], analiza jogurta [Christensen *et al.* 2005] itd.

U ovoj disertaciji će biti detaljno objašnjena primena paralelne faktorske analize (PARAFAC) za modeliranje dobijenih spektroskopskih podataka četiri kompleksna sistema: rastvor vode i aminokiselina, med, kancer dojke i kompleksni sistem fosfora na bazi retkih zemalja. Osim navedene analize fluorescentni spektri će biti analizirani i drugim statističkim metodama: analiza glavnih komponentata, diskriminantna analiza parcijalno najmanjih kvadrata i analiza potpornih vektora. Rezultati ovih analiza će biti upoređivani sa rezultatima PARAFAC metode i istraživaće se koje metode imaju veću uspešnost u analizi spektralnih podataka kompleksnih sistema. Ideja ovog rada je bila ispitivanje potencijala fluorescentne spektroskopije za analizu, karakterizaciju i modeliranje kompleksnih sistema.

2. Multivarijantna analiza

Ova teza se bavi problematikom iz oblasti fluorescentne spektroskopije i analize i modeliranja podataka dobijenih tom metodom. Karakteristično za ovu metodu jeste mnogo veći broj varijabli u odnosu na broj merenih uzoraka, slučaj u kome nam je onemogućena primena standardnih statističkih metoda za obradu podataka. Umesto njih neophodno je koristiti metode specijalno razvijene za obradu i analizu višedimenzionalnih podataka. Multivarijantna analiza se bazira na činjenici da je za opisivanje kompleksnih problema neophodno više parametara. Stoga se korišćenjem i kombinovanjem više varijabli može izvući više informacija vezanih za posmatrani sistem.

Kada je u pitanju spektroskopija, umesto snimanja samo jednog ili više pikova koji nas zanimaju, snimaju se celi spektri ili površine koji se zatim koriste za evaluaciju određenog sistema. Spektroskopski podaci dobijeni na taj način se uglavnom sastoje od stotina varijabli, koje su generalno visoko korelisane. Klasične statističke metode ne mogu efikasno obrađivati ovaj tip podataka, dok metode multivarijantne analize pristupaju ovakvim podacima ekstrakcijom linerno nezavisnih latentnih varijabli iz originalnog seta međusobno korelisanih podataka. Dve osnovne metode koje se primenjuju u praksi su: metoda glavnih komponentata (eng. Principal Component Analysis (PCA)) i regresija metodom najmanjih kvadrata (eng. Partial Least Squares (PLS) regression). PCA i PLS su bilinearni modeli bazirani na linearnoj dekompoziciji originalnih podataka u novi set glavnih komponenti ili latentnih varijabli, respektivno (koncept linearnosti modela je detaljnije objašnjen u prilogu A.2). Ove tehnike su se pokazale kao jako robusne, pri čemu omogućavaju lakšu detekciju netipičnih uzoraka (eng. Outlier) i dobro izlaze na kraj sa mogućim šumovima i artefaktima u podacima [Bro 2003.].

2.1 Bilinearni modeli – PCA i PLS

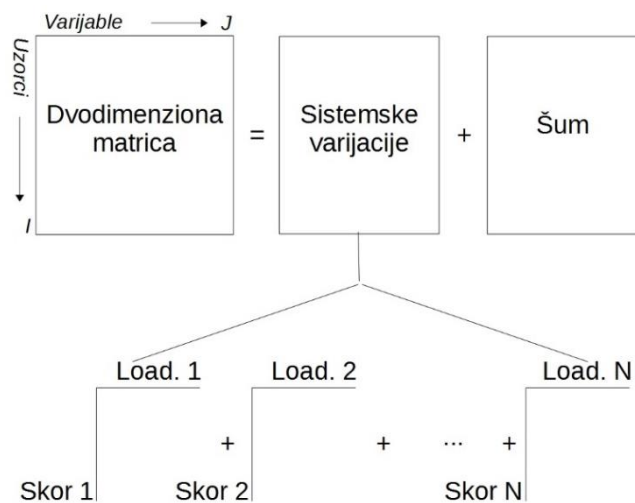
2.1.1 PCA

Ova metoda je osnovni alat za multivarijantnu i eksploratornu analizu podataka koja je u samom začetku bila razvijena za primenu u psihometriji početkom dvadesetog veka, a tek mnogo kasnije počela je da se primenjuje na druge vrste podataka.

Osnovni princip PCA metode se može sumirati sledećom formulom

$$\mathbf{X} = \mathbf{S} \cdot \mathbf{L}^T + \mathbf{E} \quad (1.1)$$

gde je \mathbf{X} ulazna matrica sa originalnim izmerenim podacima (svaki red odgovara merenom uzorku, dok svaka kolona odgovara merenom parametru odnosno varijabli) koja se dekomponuje na dve matrice: skor matricu \mathbf{S} (eng. score matrix) i *loading* matricu \mathbf{L} (eng. loading matrix) pri čemu one zajedno izražavaju glavne varijacije u podacima, ostavljajući sve nesistemske doprinose u ostatku \mathbf{E} (slika 2.1). Na ovaj način podaci iz matrice \mathbf{X} su transformisani u set međusobno ortogonalnih komponenti. Proizvodi vrednosti svakog skora i vektora opterećenja se nazivaju glavnim komponentama (eng. Principal Components (PC's)). Vrednost skora predstavlja poziciju uzoraka za svaku odgovarajuću glavnu komponentu. Vektori opterećenja se mogu posmatrati kao novi koordinatni prostor varijabli, odnosno kao pravci svake glavne komponente na osnovu originalnih ulaznih parametara. Kada su spektralni podaci u pitanju, u idealnom slučaju skor vrednosti predstavljaju koncentracije određenih supstanci u uzorku dok vektor opterećenja predstavlja spektralni profil posmatranog sistema.



Slika 2.1 – Grafički prikaz PCA analize.

Izdvajanjem samo bitnih varijacija u novi set glavnih komponenata dolazi do bitnog smanjenja dimenzionalnosti početnih podataka, pri čemu se uklanjaju razni šumovi i tretiraju postojeće kolinearnosti prisutne u podacima. PCA modeli i dalje nose informacije visokodimenzionalnih početnih podataka, ali u komprimovanom formatu. Stoga, cilj linearne dekompozicije u multivarijantnoj analizi je istovremena analiza svih prikupljenih podataka, što omogućava obradu velikih setova podataka i sagledavanje uticaja svih parametara istovremeno. Zavisno od ishoda analize može se lako smanjiti dimenzionalnost početnih podataka radi lakše evaluacije rezultata i boljeg fokusa na bitne informacije u sistemu.

2.1.2 PLS

Drugi osnovni alat u multivarijantnoj analizi je PLS regresija. Ova metoda se zasniva na dekompoziciji ulazne matrice \mathbf{X} sličnoj dekompoziciji primenjenoj kod PCA metode. Međutim, za razliku od PCA, PLS modeli su razvijeni za korelaciju dva seta podataka, \mathbf{X} i y , što je opisano sledećom jednačinom

$$y = \mathbf{X} \cdot b + e \quad (1.2)$$

\mathbf{X} matrica sadrži više izmerenih parametara, b nosi informaciju o regresionim koeficijentima izračunatih u kalibracionoj rutini, dok y predstavlja jednu ili više varijabli ili kvalitativnih parametara, za koje želimo da budu prediktovani na osnovu matrice \mathbf{X} . Ovaj metod regresije se sastoji iz dva bloka i simultano se odvija analiza dva seta podataka pri čemu se dekompozicija \mathbf{X} matrice odvija uzimajući sve vreme u obzir y . Ova analiza se najčešće koristi u spektroskopiji za korelaciju spektra sa određenim fizičkim ili hemijskim parametrom, poput koncentracije neke određene supstance u uzorku.

2.1.3 Diskriminantna analiza parcijalno najmanjih kvadrata – PLS-DA

Diskriminantna analiza parcijalno najmanjih kvadrata (eng. Partial Least Squares Discriminant Analysis – PLS-DA) predstavlja metodu za građenje linearnih

diskriminantnih (klasifikacionih) modela. Ova tehnika je bazirana na PLS regresiji i pokazala se kao korisna za primenu kod visoko dimenzionalnih podataka. Kao što je prethodno opisano PLS model transformiše početni set podataka u manji set koji se sastoji od usrednjenih, linearnih, latentnih parametara (eng. Latent variable – LV). U slučaju PLS-DA modeliranja, klasna pripadnost svakog uzorka se uzima kao zavisna varijabla koju želimo da model predvidi. Ovakvi modeli se uglavnom koriste za klasifikaciju novih nepoznatih uzoraka. Za validaciju PLS-DA modela se najčešće koriste različite vrste unakrsne validacije ili test setovi. U ovoj tezi je bio korišćen metod preklapanja (eng. venetian blinds), koji odabira svaki s -ti uzorak iz početnog seta podataka i deli ga na s podsetova tako da svaki uzorak bude jednom izostavljen iz test seta.

2.1.4 Primena PCA i PLS metoda u fluorescentnoj spektroskopiji

Multivarijantne analize su našle široku primenu u oblasti fluorescentne spektroskopije za građenje bilinearnih modela u cilju evaluacije merenih spektara. Ovakvi modeli se često koriste za klasifikaciju, karakterizaciju i predikciju kvalitativnih i procesnih parametara raznih sistema. Oblast u kojoj se najčešće koriste jeste prehrambena industrija i medicina [Dufour & Riublanc 1997, Dufour *et al.* 2000, Dufour *et al.* 2001, Herbert *et al.* 1999, Mazerolles *et al.* 2001]. Razlog ovome je postojanje fenomena autofluorescencije bioloških uzoraka. Jedna od osnovnih karakteristika bioloških sistema jeste njihova kompleksnost i iz tog razloga je nekad nemoguće dobro opisati sistem bilinearnim modelom već je neophodno uzeti u obzir više dimenzija. U tom slučaju se koriste metode za modeliranje i analizu višedimenzionalnih podataka.

2.2 Metoda za analizu višedimenzionalnih podataka - PARAFAC

Opisani bilinearni modeli se koriste za analizu podataka organizovanih u dvodimenzionalne matrice, gde najčešće redovi predstavljaju uzorke, a parametri kolone. Pojam analize višedimenzionalnih podataka se odnosi na multivarijantnu analizu koja se vrši na matricama dimenzionalnosti većoj od dva. U daljem opisu analize višedimenzionalnih podataka fokus će biti na primeni paralelne faktorske

analize (eng. Parallel Factor Analysis – PARAFAC) na podatke dobijene merenjem fluorescentnih spektara organizovanih u tri dimenzije, trodimenzionalne matrice, koje sadrže kao informaciju u jednom slučaju intenzitet fluorescencije u funkciji uzorka, talasne dužine ekscitacije i talasne dužine emisije i u drugom u funkciji uzorka, talasne dužine emisije i vremena života.

Višedimenzionalna analiza podrazumeva analizu višedimenzionalnih podataka. Kada se izmeri EEM fluorescentnog spektra za više uzoraka, ovi podaci se mogu organizovati u trodimenzionalnu matricu veličine $I \times J \times K$, gde I predstavlja broj izmerenih uzoraka, a J i K broj talasnih dužina na kojima je izmerena emisija i ekscitacija, respektivno. Za obradu ove vrste podataka neophodna je primena višedimenzionalne analize. Problematika višedimenzionalne analize nije vezana samo za algoritam i dobijanje dobrog modela već postoje još neka tehnička pitanja na početku i tokom same analize koja se moraju rešiti. Prvo podaci moraju biti dobro izmereni da bi dalja analiza imala smisla, a zatim treba da se reše problemi "outlier"-a (uzoraka koji su izdvojeni od ostatka grupe), podataka koji nedostaju i prisutnost raznih smetnji u samim podacima (šum, rasejanje, artefakti). Kada su podaci dobro pripremljeni i obradjeni za samu analizu najbitnija stvar jeste definisanje problematike odnosno fenomena koji treba da se analizira.

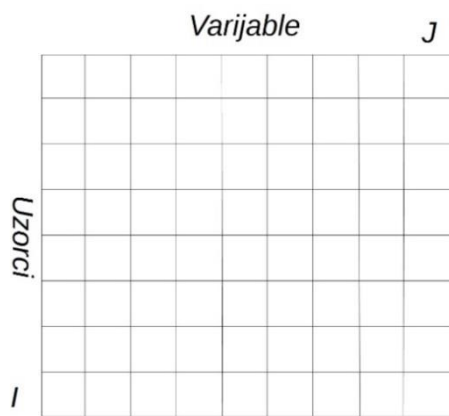
Radi boljeg razumevanja ideje i nastanka višedimenzionalne analize ukratko je u daljem tekstu dat istorijski pregled razvoja ovih metoda. Naučnik Raymond Cattell je objavio prve radove koji su se bavili metodama analize višedimenzionalnih podataka [Cattell 1944, Cattell 1952]. Turstonov (eng. Thurstone) princip parsimonije navodi da je potrebno naći jednostavnu strukturu koja opisuje podatke u matrici ili njenu korelacijsku matricu uz pomoć faktora [Thurstone 1935]. Za simultanu analizu više matrica odjednom, Cattell je predložio princip "paralelnih proporcionalnih profila" [Cattell 1944]. Princip "paralelnih proporcionalnih profila" govori da treba naći set zajedničkih faktora koji se mogu fitovati (prilagoditi) sa različitim dimenzionalnim težinskim faktorima koji odgovaraju za više matrica istovremeno. U principu to predstavlja isti problem kao i traženje zajedničkog seta faktora za trodimenzionalnu matricu. Da citiramo Cattell-a:

„Čini se da princip parsimonije ne treba da pita "Koji je najjednostavniji set faktora koji reprodukuju ovu korelacionu matricu?" već "Koji set faktora je najviše parsimon istovremeno za tu i druge matrice posmatrane zajedno?"

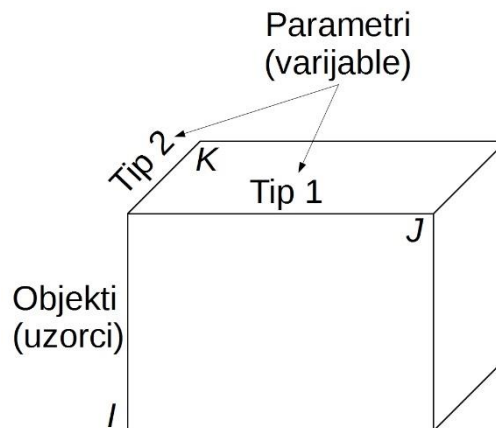
Stoga cilj višedimenzionalne analize podrazumeva pronalaženje najmanje različitih faktora za sve matrice posmatrane zajedno. Jedan od najznačajnijih Cattell-ovih radova je rad u kome definiše višedimenzionalne matrice [Cattell 1952]. PARAFAC model prvi put se spominje u radu naučnika Harshman-a 1970. godine gde je prvi put predstavljen ovaj trodimenzionalni model [Harshman 1970]. Osnovna ideja iza tog modela je bila da se koriste isti faktori za opisivanje varijacija u više matrica istovremeno, ali sa različitim težinskim koeficijentima specifičnim za svaku matricu. To je upravo i bila osnovna ideja "paralelnih proporcionalnih profila" koje je Cattell predložio. PARAFAC [Carroll 1970, Harshman 1970] metoda je takođe kao i PCA razvijena za rešavanje problema u oblasti psihometrije i tek mnogo kasnije je adaptirana kao metoda multivarijantne analize za analizu spektroskopskih podataka [Geladi 1989, Ross 1991, Russell 1988, Smilde 1991]. Međutim krajem sedamdesetih godina 20. veka razvijane su druge i slične metode faktorske analize, koje su bile korišćene za evaluaciju trodimenzionalnih podataka [Ho 1978, Marchiarullo 1982, Sanchez 1986].

2.2.1 Trodimenzionalni PARAFAC model

U cilju razumevanja ove metode, prvo će detaljnije biti objašnjeno na koji način su trodimenzionalne matrice organizovane i koji se novi termini uvode u tom slučaju. Kako bi se najlakše opisale trodimenzionalne matrice prvo treba poći od dvodimenzionalnih. Dvodimenzionalni podaci se tipično predstavljaju dvodimenzionalnom matricom gde kolone predstavljaju izmerene parametre, dok redovi odgovaraju merenim uzorcima (slika 2.2).



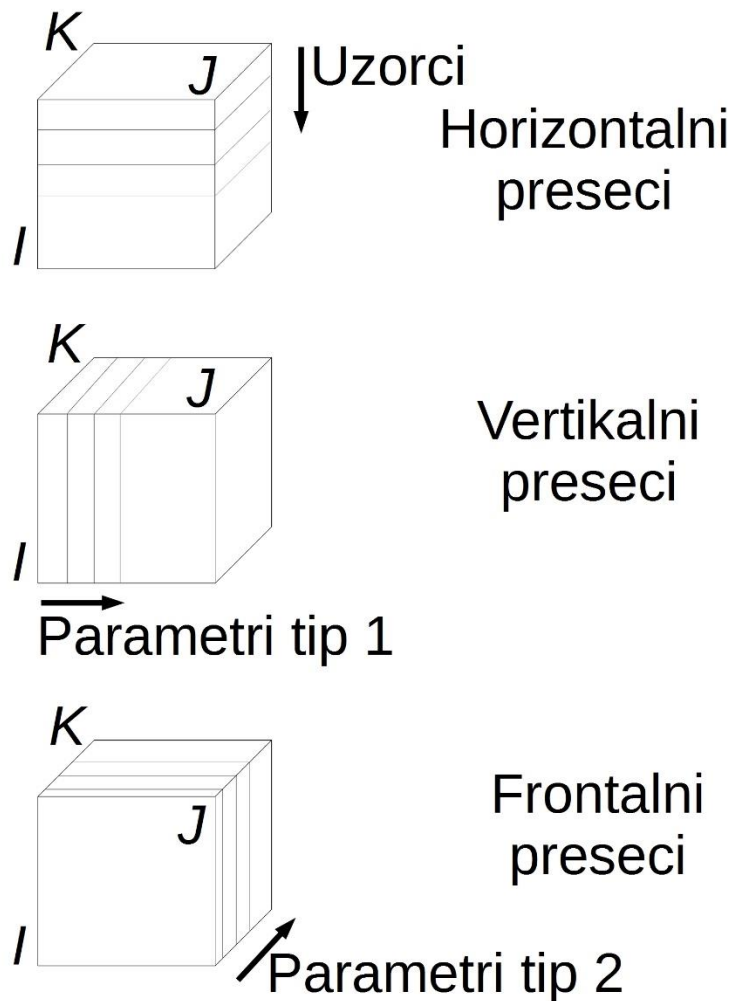
Slika 2.2 – Podaci organizovani u dvodimenzionalnu matricu.



Slika 2.3 – Podaci organizovani u trodimenzionu matricu.

Trodimenzionalni podaci se predstavljaju u obliku trodimenzionalne matrice (oblik kvadra) (slika 2.3). U ovom slučaju redovi i kolone se "menjaju" terminom *slices* (preseki) koji se odnosi na *slice*-ove nastale presecanjem matrice u različitim pravcima, što je grafički prikazano na slici 2.4. Svaki horizontalni *slice* odgovara merenom uzorku, a svaki vertikalni i frontalni *slice* prikazuje podatke za određeni izmereni parametar prvog i drugog tipa, respektivno (u slučaju podataka fluorescencije vertikalni bi odgovarali određenim talasnim dužinama emisije, a frontalni talasnim dužinama ekscitacije). Ove matrice ne moraju biti uređene po predstavljenom pravilu, ali je to postala praksa radi lakše analize i interpretacije rezultata.

PARAFAC se može smatrati ekstenzijom PCA metode na podatke višeg reda. Ona podrazumeva dekompoziciju početne matrice na trilinearne komponente tri *loading* vektora, često opisanih kao jedan skor vektor i dva *loading* vektora (u slučaju više dimenzionalnih podataka dekompozicija je proširena na kvadrolinearnost, kvintilinearnost itd.). Deo početne matrice koji nije opisan komponentama predstavlja ostatak. Kod savršenog modela suma komponentata (model) objašnjava sve sistemske varijacije u matrici i ostavlja samo šum u ostatku.



Slika 2.4 – Preseci trodimenzionalne matrice.

PARAFAC model se može predstaviti pomoću više tipova notacija. Iako će u daljem objašnjenju biti korišćena opšte poznata i uobičajena notacija, za opis ovog modela se takođe mogu koristiti i neke manje uobičajene notacije poput Kronecker, Khatri-Rao i tenzorskog proizvoda [Prilog A.1].

PARAFAC model se ovde uvodi generalizacijom singularne dekompozicije matrice [Prilog A.4]. Dvodimenzionalni model matrice \mathbf{X} ($I \times J$), sa elementima matrice x_{ij} , na osnovu dekompozicije singularnom vrednošću sveden na R komponentata predstavljen je u zbirnoj notaciji

$$x_{ij} = \sum_{r=1}^R a_{ir} g_{rr} b_{jr} + e_{ij}; i = 1, \dots, I; j = 1, \dots, J \quad (1.3)$$

gde su \mathbf{A} ($I \times R$), sa elementima a_{ir} , i \mathbf{B} ($J \times R$), sa elementima b_{jr} , obe ortogonalne matrice, a $\mathbf{G} = \text{diag}(g_{11}, \dots, g_{RR})$ je dijagonalna matrica odnosno singularna vrednost ili *core* matrica koja sadrži R najvećih singularnih vrednosti matrice \mathbf{X} u opadajućem redosledu, dok e_{ij} su ostatci. Prethodna jednačina predstavljena u matričnoj notaciji postaje

$$\mathbf{X} = \mathbf{AGB}' + \mathbf{E} \quad (1.4)$$

gde \mathbf{E} sadrži elemente e_{ij} . Vrednosti g_{rr} se mogu apsorbovati u a_{ir} ili b_{jr} ili obe. To dovodi do sledeće formulacije modela

$$\mathbf{X} = \mathbf{AB}' + \mathbf{E} \quad (1.5)$$

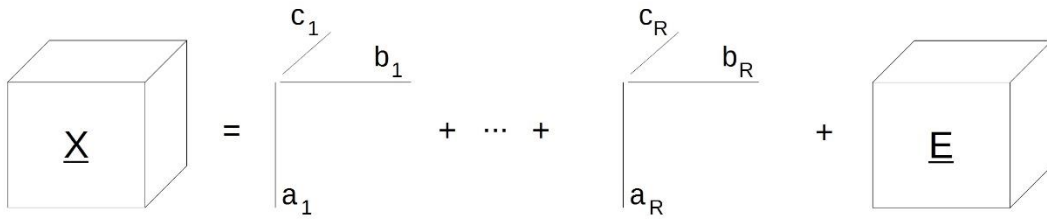
gde su nazivi matrica ostali isti iako matrice \mathbf{A} i \mathbf{B} u jednačini 1.4 i 1.5 neće biti iste zbog dodavanja uticaja g_{rr} vrednosti jednoj od njih. Ovaj model se takođe može napisati i na sledeći način

$$x_{ij} = \sum_{r=1}^R a_{ir} b_{jr} + e_{ij} \quad (1.6)$$

Za trodimenzionalnu matricu $\underline{\mathbf{X}}$ ($I \times J \times K$) sa elementima x_{ijk} , jednačina 1.6 se može generalizovati za PARAFAC model

$$x_{ijk} = \sum_{r=1}^R a_{ir} b_{jr} c_{kr} + e_{ijk} \quad (1.7)$$

gde je R broj komponenta odnosno faktora korišćenih za građenje PARAFAC modela, dok je e_{ijk} ostatak koji sadrži u sebi sve nerazjašnjene varijacije. Grafički prikaz modela je dat na slici 2.5.



Slika 2.5 – Grafički prikaz PARAFAC modela.

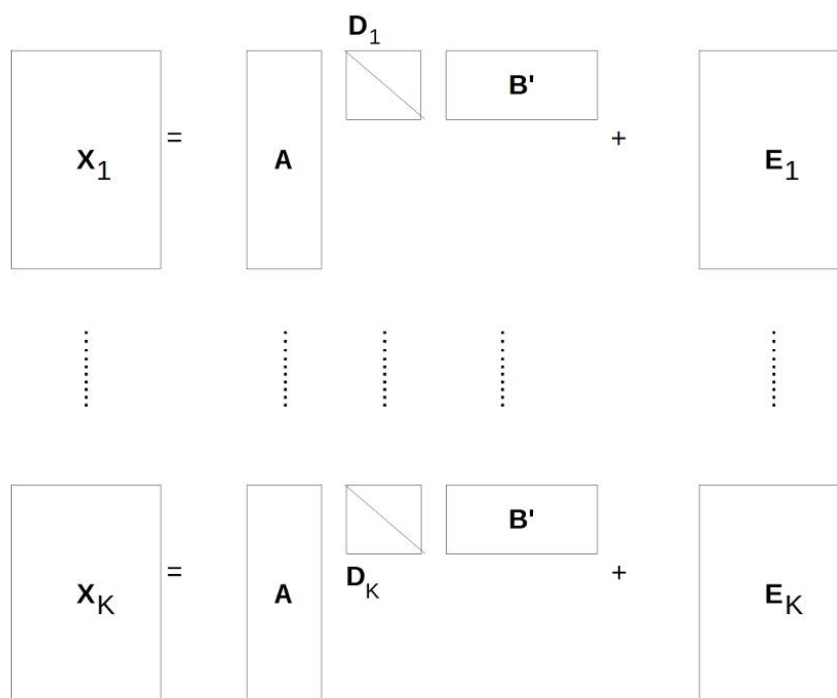
Model predstavljen jednačinom 1.7 je trilinearni model: sa dva fiksirana seta parametara (na primer a i b), dok je x_{ijk} izražen kao linearna funkcija preostalih parametara (na primer c seta parametara). Model u jednačini 1.7 može biti izražen i na sledeći način ako se uvede skalirajući parametar g_{rrr} :

$$x_{ijk} = \sum_{r=1}^R g_{rrr} a_{ir} b_{jr} c_{kr} + e_{ijk} \quad (1.8)$$

Ako elemente a_{ir} "spakujemo" u matricu \mathbf{A} ($I \times R$), elemente b_{jr} u \mathbf{B} ($J \times R$) i c_{jr} u \mathbf{C} ($K \times R$) onda jednačina 1.7 može da se predstavi u funkciji matrice \mathbf{X}_k , gde je \mathbf{X}_k k -ti ($I \times J$) slice trodimenzionalne matrice $\underline{\mathbf{X}}$ ($I \times J \times K$) što znači da elementi matrice \mathbf{X}_k su x_{ijk} ($i = 1, \dots, I; j = 1, \dots, J$ za svako k):

$$\mathbf{X}_k = \mathbf{A} \mathbf{D}_k \mathbf{B}' + \mathbf{E}_k = c_{k1} \mathbf{a}_1 \mathbf{b}'_1 + \dots + c_{kR} \mathbf{a}_R \mathbf{b}'_R + \mathbf{E}_k \quad (1.9)$$

gde je \mathbf{D}_k dijagonalna matrica sa elementima k -tog reda matrice \mathbf{C} na njenoj dijagonali (elementi c_{k1}, \dots, c_{kR}); $\mathbf{a}_r, \mathbf{b}_r$ su r -te kolone matrica \mathbf{A} i \mathbf{B} , respektivno, a ostatak \mathbf{E}_k ($I \times J$) je definisan slično matrici \mathbf{X}_k . Dakle, svako \mathbf{X}_k se modeluje korišćenjem istih komponenti matrica \mathbf{A} i \mathbf{B} , ali sa različitim težinskim faktorima koji su deo matrice \mathbf{D}_k . Grafički prikaz jednačine 1.9 je prikazan na slici 2.6.



Slika 2.6 – Grafički prikaz jednačine (1.9).

Za sve \mathbf{X}_k *slice*-ove komponente \mathbf{a}_r i \mathbf{b}_r ostaju iste, samo su njihovi težinski faktori d_{k1}, \dots, d_{kR} različiti. Stoga se može zaključiti da su svi \mathbf{X}_k *slice*-ovi modelovani paralelnim i proporcionalnim profilima $d_{k1}\mathbf{a}_1\mathbf{b}'_1, \dots, d_{kR}\mathbf{a}_R\mathbf{b}'_R$. Ovo omogućava opisivanje modela primenom notacije simultanih komponenata. Postoje tri ekvivalentna načina za opis PARAFAC modela ovom notacijom zbog simetrije samog modela (1.9). Ove jednačine opisuju frontalni, horizontalni i vertikalni *slice*, respektivno:

$$\begin{aligned} \mathbf{X}_k &= \mathbf{A}\mathbf{D}_k(\mathbf{C})\mathbf{B}', k = 1, \dots, K \\ \mathbf{X}_i &= \mathbf{B}\mathbf{D}_i(\mathbf{A})\mathbf{C}', i = 1, \dots, I \\ \mathbf{X}_j &= \mathbf{A}\mathbf{D}_j(\mathbf{B})\mathbf{C}', j = 1, \dots, J \end{aligned} \quad (1.10)$$

gde dijagonalne matrice $\mathbf{D}_k(\mathbf{C})$, $\mathbf{D}_i(\mathbf{A})$ i $\mathbf{D}_j(\mathbf{B})$ su operatori dobijeni ekstrakcijom k -tog, i -tog i j -tog reda iz odgovarajuće loading matrice (\mathbf{C} , \mathbf{A} i \mathbf{B} , respektivno).

Postoje i drugi načini i notacije za opis PARAFAC modela koji se u nekim slučajevima koriste kada je jednostavnije prikazati problem na načine koji su dati u daljem tekstu (definicije Kronecker, Hadamard i Khatri-Rao proizvoda su date u prilogu A.1).

2.2.2 PARAFAC model u notaciji Kronecker proizvoda

Jednačinu (1.9) je takođe moguće napisati i u notaciji Kronecker proizvoda koristeći da je $\mathbf{X} = [\mathbf{X}_1 \mathbf{X}_2 \dots \mathbf{X}_K]$; $\mathbf{E} = [\mathbf{E}_1 \mathbf{E}_2 \dots \mathbf{E}_K]$ i $\mathbf{D} = [\mathbf{D}_1 \mathbf{D}_2 \dots \mathbf{D}_K]$:

$$\mathbf{X} = \mathbf{A}\mathbf{D}(\mathbf{I} \otimes \mathbf{B}') + \mathbf{E} \quad (1.11)$$

ili korišćenjem superdijagonalne matrice $\mathbf{I}(R \times R \times R)$ svedene na matricu $\mathbf{H}(R \times R^2)$

$$\mathbf{X} = \mathbf{A}\mathbf{H}(\mathbf{C}' \otimes \mathbf{B}') + \mathbf{E} \quad (1.12)$$

pri čemu za obe jednačine postoje dve ekvivalentne alternative zbog simetrične strukture PARAFAC modela.

2.2.3 PARAFAC model u tenzorskoj notaciji

Korišćenjem tenzorskog proizvoda PARAFAC model se može opisati na sledeći način

$$\underline{\mathbf{X}} = \sum_{r=1}^R \mathbf{a}_r \Delta \mathbf{b}_r \Delta \mathbf{c}_r + \underline{\mathbf{E}} \quad (1.13)$$

gde su \mathbf{a}_r , \mathbf{b}_r i \mathbf{c}_r r -te kolone matrica \mathbf{A} , \mathbf{B} i \mathbf{C} , respektivno.

2.2.4 PARAFAC model u vektorskoj i Kronecker notaciji

PARAFAC model se može opisati i pomoću vektorskog operatora i Kronecker proizvoda. To je dato sledećom jednačinom

$$\text{vec } \underline{\mathbf{X}} = \sum_{r=1}^R \mathbf{a}_r \otimes \mathbf{b}_r \otimes \mathbf{c}_r + \text{vec } \underline{\mathbf{E}} \quad (1.14)$$

gde je $\underline{\mathbf{X}}$ trodimenzionalna matrica veličine $(I \times J \times K)$. $\text{Vec } \underline{\mathbf{X}} = \text{vec } \mathbf{X}$ gde je \mathbf{X} dobijeno svođenjem matrice $\underline{\mathbf{X}}$ u dvodimenzionalnu matricu.

2.2.5 PARAFAC model u notaciji Khatri-Rao proizvoda

Poslednji način na koji se može opisati PARAFAC model je dat Khatri-Rao proizvodom.

$$\mathbf{X} = \mathbf{A}(\mathbf{C} \odot \mathbf{B})' + \mathbf{E} \quad (1.15)$$

gde su \mathbf{X} i \mathbf{E} definisani na isti način kao i u jednačini (1.11).

Khatri-Rao proizvod je koristan za više-linearne modele. Ako krenemo od bilinearnog dvodimenzionalnog modela

$$\mathbf{X} = \mathbf{A}\mathbf{B}' \quad (1.16)$$

Ovaj model se može vektorisati i onda se dobija

$$\text{vec } \mathbf{X} = \text{vec } (\mathbf{A}\mathbf{B}') \quad (1.17)$$

Što jednostavno znači da su matrice preuređene u vektore. Na primer, $\text{vec } \mathbf{X}$ je IJ vektor dobijen kao

$$\begin{bmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \\ \vdots \\ \mathbf{x}_J \end{bmatrix} \quad (1.18)$$

gde je \mathbf{x}_j j -ta kolona matrice \mathbf{X} . U skladu sa tim, bilinearni deo ovog modela se može napisati na sledeći način

$$\begin{aligned} \text{vec}(\mathbf{AB}') &= \\ \text{vec}(\mathbf{a}_1\mathbf{b}'_1) + \text{vec}(\mathbf{a}_2\mathbf{b}'_2) + \dots + \text{vec}(\mathbf{a}_R\mathbf{b}'_R) &= \\ \begin{bmatrix} b_{11}\mathbf{a}_1 \\ b_{21}\mathbf{a}_1 \\ \vdots \\ b_{J1}\mathbf{a}_1 \end{bmatrix} + \begin{bmatrix} b_{12}\mathbf{a}_2 \\ b_{22}\mathbf{a}_2 \\ \vdots \\ b_{J2}\mathbf{a}_2 \end{bmatrix} + \dots + \begin{bmatrix} b_{1R}\mathbf{a}_R \\ b_{2R}\mathbf{a}_R \\ \vdots \\ b_{JR}\mathbf{a}_R \end{bmatrix} & \quad (1.19) \end{aligned}$$

Stoga svaka bilinearna komponenta $\text{vec}(\mathbf{a}_r\mathbf{b}'_r)$ se može napisati kao definisani vektor koji je u funkciji samo od \mathbf{a}_r i \mathbf{b}_r . Ovaj vektor predstavlja Kronecker-ov proizvod \mathbf{b}_r i \mathbf{a}_r . Khatri-Rao proizvod dve matrice, \mathbf{A} i \mathbf{B} , sa istim brojem kolona je definisan na sledeći način

$$\mathbf{B} \odot \mathbf{A} = \begin{bmatrix} b_{11}\mathbf{a}_1 & \dots & b_{1R}\mathbf{a}_R \\ b_{21}\mathbf{a}_1 & \dots & b_{2R}\mathbf{a}_R \\ \vdots & \dots & \vdots \\ b_{J1}\mathbf{a}_1 & \dots & b_{JR}\mathbf{a}_R \end{bmatrix} = [\mathbf{b}_1 \otimes \mathbf{a}_1 \dots \mathbf{b}_R \otimes \mathbf{a}_R] \quad (1.20)$$

Onda se model iz jednačine (1.17) može napisati na sledeći način

$$\text{vec } \mathbf{X} = (\mathbf{B} \odot \mathbf{A})\mathbf{1} \quad (1.21)$$

gde je $\mathbf{1}$ R -vektor sačinjen od jedinica. Ovo je princip korišćen u jednačini (1.15) izražavajući istovremeno drugi i treći mod u dvodimenzionalnoj matrici.

PARAFAC model je specifičan po tome što daje jedinstveno rešenje, odnosno izračunate \mathbf{A} , \mathbf{B} i \mathbf{C} matrice ne mogu biti promenjene bez promene ostatka (ne postoji rotaciona sloboda). Ova osobina se drugačije naziva i osobina unutrašnjih osa zato što sa PARAFAC-om se ne nalaze samo jedinstveni podprostori (kao kod PCA analize) već i jedinstvene orijentacije bazičnog vektora.

Parametri u matricama \mathbf{A} , \mathbf{B} i \mathbf{C} mogu biti izračunati uz pomoć različitih algoritama. Bitno je istaći da za razliku od PCA gde se komponente računaju jedna po jedna, kod PARAFAC modela faktori se moraju računati simultano. Razlog tome je što

komponente PARAFAC modela nisu međusobno ortogonalne i samim tim zavise jedna od druge.

2.2.6 Rotaciona sloboda i unikatnost trodimenzionalnih PARAFAC modela

Unikatnost modela je najbitnija osobina PARAFAC modela. Jedinstven model je model kod koga su svi parametri izračunati pod definisanim premisama odnosno uslovima, kao što je struktura i dodatna ograničenja (ortogonalnost, unimodalnost). Uslov da model matrice bude bilinearan nije dovoljan za dobijanje parametara modela na jedinstven način. Međutim, ako se koriste neka dodatna ograničenja model može biti jedinstven odnosno unikat. Mnoge primene PARAFAC modela se baziraju upravo na ovoj osobini, poput kalibracije drugog reda i razlaganja spektralnih podataka [Leurgans & Ross 1992, Sanchez & Kowalski 1988]. Zato je bitno posvetiti pažnju ovoj temi.

Pojam jedinstvenosti se koristi u prethodno pomenutom smislu, odnosno trivijalna nejedinstvenost usled skaliranja i permutacije matrica komponenti se ne uzima u obzir. Ovakva vrsta neodređenosti se uglavnom rešava po principu rešavanja problema. Pretpostavimo da je matrica $\underline{\mathbf{X}}$ modelirana R- komponentnim PARAFAC modelom, sa matricama komponenti \mathbf{A} , \mathbf{B} , \mathbf{C} sa elementima a_{ir} , b_{jr} i c_{kr} , respektivno. Stoga, elementi x_{ijk} matrice $\underline{\mathbf{X}}$ mogu biti napisani na sledeći način

$$x_{ijk} = \sum_{r=1}^R a_{ir} b_{jr} c_{kr} + e_{ijk} \quad (1.22)$$

gde je e_{ijk} element matrice ostatka $\underline{\mathbf{E}}$ i jednačina (1.22) važi za $i = 1, \dots, I$; $j = 1, \dots, J$; $k = 1, \dots, K$. Sada pretpostavimo da postoje matrice $\tilde{\mathbf{A}}$, $\tilde{\mathbf{B}}$, $\tilde{\mathbf{C}}$ sa elementima \tilde{a}_{ir} , \tilde{b}_{jr} , \tilde{c}_{kr} respektivno tako da važi

$$x_{ijk} = \sum_{r=1}^R \tilde{a}_{ir} \tilde{b}_{jr} \tilde{c}_{kr} + e_{ijk} \quad (1.23)$$

pri čemu su x_{ijk} i e_{ijk} identični odgovarajućim vrednostima u jednačini (1.22), u skladu sa time osobina jedinstvenosti modela tvrdi da \mathbf{A} , \mathbf{B} , \mathbf{C} je jednako $\tilde{\mathbf{A}}$, $\tilde{\mathbf{B}}$, $\tilde{\mathbf{C}}$ respektivno,

osim razlika usled skaliranja i permutacija. Drugim rečima, nije moguće rotirati matrice PARAFAC komponenti bez promene samog modela. Treba naglasiti da se osobina jedinstvenosti odnosi na izračunate parametre modela. Ne postoji garancija da PARAFAC model sadrži u sebi "pravo" rešenje. Međutim, ako je model dobro okarakterisan sa tačnim brojem komponentata i sama PARAFAC trilinearna struktura približno validna, u tom slučaju će rešenje PARAFAC analize dati dobru procenu parametara.

Dovoljan uslov da PARAFAC model da jedinstvenu procenu parametara je dao Kruskal [Kruskal 1989] koji su kasnije proširili Sidiropoulos i Bro [Sidiropoulos & Bro 2000]. Uslov je sledeći

$$k_A + k_B + k_C \geq 2R + 2 \quad (1.24)$$

gde k_A , k_B i k_C su k -rangovi matrica komponenti **A**, **B** i **C**, respektivno; R je broj komponentata PARAFAC modela. k -rang matrice se definiše kao veličina najvećeg podskupa svih podskupova kolona te matrice koji uvek imaju pun rang. Takođe važi da je jednokomponentni model uvek jedinstven, u tom slučaju se zanemaruje jednačina (1.24). Kruskal-ovi uslovi ne moraju uvek biti ispunjeni osim kada je vrednost R manja od četiri [Ten Berge & Sidiropoulos 2001]. Uslovi za jedinstvenost PARAFAC modela koji imaju više od tri komponente još uvek nisu poznati. Kada je uslov iz jednačine (1.24) ispunjen, PARAFAC model daje jedinstvenu procenu parametara. Takođe, ako je na pravi način definisan, PARAFAC model može da da procenu osnovnih fizičkih parametara koji se nalaze skriveni među početnim podacima.

2.3 PARAFAC algoritam

U ovom delu će biti detaljno opisani algoritmi za računanje parametara kod višedimenzionalnih modela. Prvo će biti opisane tehnike optimizacije na primeru fitovanja dvodimenzionalnog modela. Nakon opisa algoritma biće objašnjeno kako se koriste fitovani modeli za nove podatke. Praktična primena toga bi mogla podrazumevati predviđanje vrednosti određenih parametara (kalibracija) ili pronalaženje "outlier"-a (statistička kontrola procesa). Podaci su uglavnom organizovani na takav

način da jedan pravac matrice odgovara opservacijama/uzorcima, dok druga dva pravca odgovaraju varijablama. Kada se izračunaju *loading* matrice prethodno fitovanog modela može se pristupiti izračunavanju *score*-ova novih uzoraka.

2.3.1 Tehnike optimizacije

Postoji više načina za procenu odnosno kalkulaciju parametara modela. Traženje parametara je optimizacioni problem i u nekim slučajevima rešenje direktnog izračunavanja može postojati. Ukoliko to nije slučaj mora se koristiti iterativni algoritam. Dva alata koji se najviše koriste za fitovanje modela kod višedimenzionalne analize su rešenja na osnovu naizmenično najmanjih kvadrata (eng. alternating least square – ALS) i svojstvene vrednosti (eng. eigenvalue).

Uzmimo u razmatranje dvodimenzionalni problem fitovanja bilinearnog modela ranga R za datu matricu \mathbf{X} ($I \times J$). Dakle, parametri \mathbf{A} ($I \times R$) i \mathbf{B} ($J \times R$) se dobijaju tražeći rešenje najmanjeg kvadrata

$$\min_{\mathbf{A}, \mathbf{B}} \|\mathbf{X} - \mathbf{AB}'\|^2 \quad (1.25)$$

Idealan slučaj predstavlja situaciju kada postoji direktno rešenje zatvorene forme. U tom slučaju kompleksnost algoritma proračuna je unapred poznata i sam algoritam je lakši za optimizaciju. Ukoliko takvo rešenje nije dostupno za dati slučaj onda se moraju koristiti iterativni algoritmi. Oni rade po principu postepenog poboljšavanja trenutne procene parametara sve dok ne dođu do momenta kada dalje poboljšanje više nije moguće. Loša strana ovog pristupa se ogleda u tome što kompleksnost može da varira od situacije do situacije [Gill *et al.* 1981].

Za prethodno predstavljeni dvodimenzionalni problem, dekompozicija singularnom vrednošću matrice \mathbf{X} ili dekompozicija svojstvenom vrednošću matrice $\mathbf{X}'\mathbf{X}$ može dati direktno rešenje. Kada je trodimenzionalni problem u pitanju slična rešenja takođe postoje, ali za razliku od dvodimenzionalnog problema gde se dobija rešenje najmanjih kvadrata, u ovom slučaju se dobija uglavnom aproksimativno rešenje.

Iterativni algoritmi takođe postoje za fitovanje bilinearnog modela iz jednačine (1.25). Opšte poznati iterativni algoritam za fitovanje prethodno prikazanog modela su nelinearna iterativna metoda parcijalno najmanjih kvadrata (eng. Non-linear iterative partial least squares - NIPALS) ili nelinearna iterativna metoda najmanjih kvadrata (eng. Non-linear iterative least square – NILES) koje je razvio Wold [Wold 1966, Wold 1975]. NIPALS algoritam je baziran na principima naizmenično najmanjih kvadrata. Kod naizmenično najmanjih kvadrata parametri se dele na nekoliko setova. Svaki set parametara se uslovno računa korišćenjem najmanjih kvadrata na osnovu trenutnih vrednosti ostalih parametara. Računanje odnosno procena parametara se iterativno ponavlja sve do trenutka kada više nema greške. Razlog deljenja parametara u više grupa je mogućnost korišćenja jednostavnijih algoritama za procenu parametara.

Uzmimo u obzir bilinearni model

$$\hat{\mathbf{X}} = \mathbf{A}\mathbf{B}' \quad (1.26)$$

Ako je poznata inicijalna procena \mathbf{A} , označena sa $\tilde{\mathbf{A}}$, onda se procena \mathbf{B} na osnovu $\tilde{\mathbf{A}}$ svodi na regresioni problem sa sledećim rešenjem

$$\tilde{\mathbf{B}} = \mathbf{X}'(\tilde{\mathbf{A}}^+)' \quad (1.27)$$

Jednačina (1.27) ne rešava celokupni problem, ali poboljšava bilo koju trenutnu procenu parametra \mathbf{B} . Nova i bolja procena \mathbf{A} sa datim $\tilde{\mathbf{B}}$ se zatim može odrediti sledećom jednačinom

$$\tilde{\mathbf{A}} = \mathbf{X}(\tilde{\mathbf{B}}^+)' \quad (1.28)$$

na osnovu čega se dalje može dobiti bolja procena \mathbf{B} itd. Osnovna ideja algoritma naizmenično najmanjih kvadrata je: redefinisati osnovni problem (1.25) u uglavnom linearne podprobleme koji su laki za rešavanje. Zatim iterativno rešavati ove probleme dok ne dođe do konvergencije. Kako su sve procene parametara procene najmanjih kvadrata, takav algoritam može samo da poboljša fitovanje modela ili da ostane nepromenjen kad konvergira. Stoga se propratna vrednost funkcije gubitka ne povećava i u praksi je pokazano da se uglavnom smanjuje. Kako je ovo ograničen problem

(funkcija gubitka ne može biti manja od nule), uglavnom je ispraćen konvergiranjem funkcije. Iako konvergiranje u globalnom minimumu nije zagarantovano, konvergencija je praktična, privlačna i jedan od razloga sve veće primene naizmenično najmanjih kvadrata [de Leeuw et al. 1976]. Još jedna od prednosti algoritma naizmenično najmanjih kvadrata je jednostavnost proračuna sa podeljenim grupama nasuprot algoritmu koji simultano radi na kompletnom problemu. Takođe, postoji puno dostupnih varijacija osnovne teorije regresije najmanjih kvadrata za rešavanje problema ograničenih parametara i nedostajućih vrednosti. Mana ove metode u nekim slučajevima može biti relativno puno vremena koje je potrebno da algoritam konvergira.

Uzmimo trodimenzionalnu matricu \mathbf{X} i generički model

$$\hat{\mathbf{X}} = \mathbf{M}(\mathbf{A}, \mathbf{B}, \mathbf{C}) \quad (1.29)$$

gde je \mathbf{M} model matrice \mathbf{X} , u funkciji parametara \mathbf{A} , \mathbf{B} i \mathbf{C} pri čemu su veličine kao i podaci sadržani u \mathbf{X} . Broj setova parametara može da varira od modela do modela. Kod PARAFAC-a postoje tri matrice, \mathbf{A} , \mathbf{B} i \mathbf{C} , dok kod bilinearnog modela postoje samo dve matrice. Za procenu parametara modela \mathbf{A} , \mathbf{B} i \mathbf{C} algoritam naizmenično najmanjih kvadrata može biti formulisan na sledeći način

1. Inicijalizacija parametara
2. $\tilde{\mathbf{A}} = \underset{\mathbf{S}}{\operatorname{argmin}} \|\mathbf{X} - \mathbf{M}(\mathbf{S}, \mathbf{B}, \mathbf{C})\|^2$
3. $\tilde{\mathbf{B}} = \underset{\mathbf{T}}{\operatorname{argmin}} \|\mathbf{X} - \mathbf{M}(\tilde{\mathbf{A}}, \mathbf{T}, \mathbf{C})\|^2$
4. $\tilde{\mathbf{C}} = \underset{\mathbf{U}}{\operatorname{argmin}} \|\mathbf{X} - \mathbf{M}(\tilde{\mathbf{A}}, \tilde{\mathbf{B}}, \mathbf{U})\|^2$
5. Vraćanje na korak dva dok algoritam ne konvergira

Izraz $\tilde{\mathbf{A}} = \underset{\mathbf{S}}{\operatorname{argmin}} \|\mathbf{X} - \mathbf{M}(\mathbf{S}, \mathbf{B}, \mathbf{C})\|^2$ pokazuje da argument \mathbf{S} , pri tom minimizirajući funkciju gubitka $\|\mathbf{X} - \mathbf{M}(\mathbf{S}, \mathbf{B}, \mathbf{C})\|^2$, daje optimalno ažuriranje matrice \mathbf{A} . U prethodnom primeru sa bilinearnim modelom $\mathbf{M}(\mathbf{A}, \mathbf{B}) = \mathbf{A}\mathbf{B}'$, ažuriranje parametara dato jednačinama (1.27) i (1.28) obezbeđuje neophodne korake u algoritmu naizmenično najmanjih kvadrata.

Inicijalizacija parametara i provera da li je algoritam konvergirao će dalje biti detaljno opisana za PARAFAC model.

2.3.2 PARAFAC algoritam naizmenično najmanjih kvadrata

Algoritmi za PARAFAC model se najčešće baziraju na metodi naizmenično najmanjih kvadrata. Prednost ove metode je što je jednostavna za implementaciju, lako se inkorporiraju ograničenja modela i garantuje konvergenciju. Međutim, nekad ume da bude i spora.

Opšti trodimenzionalni PARAFAC model je definisan na sledeći način

$$\mathbf{X}_{(I \times JK)} = \mathbf{A}(\mathbf{C} \odot \mathbf{B})' + \mathbf{E} \quad (1.30)$$

dok je odgovarajuća funkcija gubitka najmanjih kvadrata data sa

$$\min_{\mathbf{A}, \mathbf{B}, \mathbf{C}} \|\mathbf{X} - \mathbf{A}(\mathbf{C} \odot \mathbf{B})'\|^2 \quad (1.31)$$

U cilju fitovanja modela uz primenu naizmenično najmanjih kvadrata neophodno je ažurirati \mathbf{A} uz dato \mathbf{B} i \mathbf{C} ; \mathbf{B} uz dato \mathbf{A} i \mathbf{C} ; \mathbf{C} uz dato \mathbf{A} i \mathbf{B} . Zbog simetrije samog modela, ažuriranje jednog moda (pravca), na primer \mathbf{A} , u suštini je isto ažuriranje bilo koji pravac samo što se uloge različitih *loading* matrica menjaju. Za procenu \mathbf{A} , uslovljene sa \mathbf{B} i \mathbf{C} , može se formulirati sledeći optimizacioni problem

$$\min_{\mathbf{A}} \|\mathbf{X} - \mathbf{AZ}'\|^2 \quad (1.32)$$

pri čemu je \mathbf{Z} jednako $\mathbf{C} \odot \mathbf{B} (= [\mathbf{D}_1 \mathbf{B}' \ \mathbf{D}_2 \mathbf{B}' \ \dots \ \mathbf{D}_k \mathbf{B}']')$ dok dijagonalna matrica \mathbf{D}_k sadrži k -ti red \mathbf{C} matrice na svojoj dijagonali. Model u obliku dvodimenzionalne matrice, koja je matematički ekvivalent trodimenzionalnoj matrici, predstavlja bilinearan model za \mathbf{A} i \mathbf{Z} ($= \mathbf{C} \odot \mathbf{B}$). Stoga, nalaženje optimalnog \mathbf{A} se svodi na regresioni korak gde se vrši regresija podataka na $\mathbf{C} \odot \mathbf{B}$. To znači da za dato \mathbf{B} i \mathbf{C} , problem se svodi na dvodimenzionalni problem pronalaženja najmanjih kvadrata optimalnog \mathbf{A} u modelu $\mathbf{X} = \mathbf{AZ}' + \mathbf{E}$. Rešenje ovog problema, kada \mathbf{Z} ima puni rang kolona, jednostavno sledi

$$\mathbf{A} = \mathbf{X}(\mathbf{Z}')^+ = \mathbf{XZ}(\mathbf{Z}'\mathbf{Z})^{-1} \quad (1.33)$$

Zbog simetrije problema sledi da \mathbf{B} i \mathbf{C} mogu biti ažurirani na sličan način. Iz datih jednačina može se formulirati sledeći algoritam za matricu \mathbf{X} dimenzije $I \times J \times K$ sa traženom dimenzijom R .

1. Inicijalizacija \mathbf{B} i \mathbf{C}

2. $\mathbf{Z} = (\mathbf{C} \odot \mathbf{B})$

$$\mathbf{A} = \mathbf{X}_{(I \times JK)} \mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}$$

3. $\mathbf{Z} = (\mathbf{C} \odot \mathbf{A})$

$$\mathbf{B} = \mathbf{X}_{(J \times IK)} \mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}$$

4. $\mathbf{Z} = (\mathbf{B} \odot \mathbf{A})$

$$\mathbf{C} = \mathbf{X}_{(K \times JI)} \mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}$$

5. Vraćanje na korak 1 dok relativna promena u fitu modela nije dovoljno mala

Računanje \mathbf{Z} u konkretno prethodnom problemu može da bude jako "skupo" u smislu izračunavanja za velike matrice. Takođe, matrica sa podacima mora konstantno da se preuređuje ili da se čuva u tri verzije ($\mathbf{X}_{(I \times JK)}$, $\mathbf{X}_{(J \times IK)}$ i $\mathbf{X}_{(K \times JI)}$) što zahteva jako veliku memoriju. Harshman i Carroll i Chang [Carroll & Chang 1970, Harshman 1970] su primetili da je jednostavnija šema ažuriranja moguća zbog same strukture problema. Algoritam može biti formulisan u funkciji frontalnih *slice*-ova, \mathbf{X}_k , što ne zahteva preuređivanje matrica. Algoritam proračunava male matrice \mathbf{XZ} i $\mathbf{Z}'\mathbf{Z}$ direktno iz \mathbf{B} , \mathbf{C} i \mathbf{X} . Ovo je zgodno jer je "jeftinije" za izračunavanje i zahteva manje memorije. Za ažuriranje matrice \mathbf{A} , matrica \mathbf{X} ($I \times JK$) je jednaka $[\mathbf{X}_1 \mathbf{X}_2 \dots \mathbf{X}_K]$. Iz toga sledi

$$\begin{aligned} \mathbf{XZ} &= [\mathbf{X}_1 \mathbf{X}_2 \dots \mathbf{X}_K] [\mathbf{D}_1 \mathbf{B}' \mathbf{D}_2 \mathbf{B}' \dots \mathbf{D}_K \mathbf{B}'] \\ &= \mathbf{X}_1 \mathbf{B} \mathbf{D}_1 + \mathbf{X}_2 \mathbf{B} \mathbf{D}_2 + \dots + \mathbf{X}_K \mathbf{B} \mathbf{D}_K \end{aligned} \quad (1.34)$$

i dalje sledi iz osobina Khatri-Rao proizvoda

$$\mathbf{Z}'\mathbf{Z} = (\mathbf{C} \odot \mathbf{B})' (\mathbf{C} \odot \mathbf{B}) = (\mathbf{C}'\mathbf{C}) * (\mathbf{B}'\mathbf{B}) \quad (1.35)$$

Stoga, ažuriranje \mathbf{A} se može napisati na sledeći način

$$\mathbf{A} = \mathbf{XZ}(\mathbf{Z}'\mathbf{Z})^{-1} = \left(\sum_{k=1}^K \mathbf{X}_k \mathbf{B}\mathbf{D}_k \right) \{(\mathbf{C}'\mathbf{C}) * (\mathbf{B}'\mathbf{B})\}^{-1} \quad (1.36)$$

Za druga dva moda, slične prečice algoritma mogu biti razvijene i uobičajeno je da se sva tri podproblema izraze uz pomoć frontalnih *slice*-ova matrice \mathbf{X} na način kao što je prethodno bilo prikazano [Kiers & Krijnen 1991]. Efikasniji algoritam za PARAFAC model je dat u daljem tekstu. Ovaj algoritam je manje očigledan od prethodnog, ali je efikasniji po pitanju računanja. Tri koraka za ažuriranje svakog moda nisu identična. Razlog ovom očiglednom odsustvu simetrije algoritma je taj što su sva ažuriranja bazirana na frontalnim *slice*-ovima podataka (\mathbf{X}_k). Sam model je i dalje simetričan i ekvivalentan modelu iz prethodnog algoritma.

Algoritam za matricu \mathbf{X} dimenzije $I \times J \times K$ sa traženom dimenzijom R , uz korišćenje frontalnih *slice*-ova

1. Inicijalizacija \mathbf{B} i \mathbf{C}

$$2. \mathbf{A} = (\sum_{k=1}^K \mathbf{X}_k \mathbf{B}\mathbf{D}_k) \{(\mathbf{C}'\mathbf{C}) * (\mathbf{B}'\mathbf{B})\}^{-1}$$

$$3. \mathbf{B} = (\sum_{k=1}^K \mathbf{X}'_k \mathbf{A}\mathbf{D}_k) \{(\mathbf{C}'\mathbf{C}) * (\mathbf{A}'\mathbf{A})\}^{-1}$$

$$4. \mathbf{d}_k = \{(\mathbf{B}'\mathbf{B}) * (\mathbf{A}'\mathbf{A})\}^{-1} \{(\mathbf{A}'\mathbf{X}_k \mathbf{B} * \mathbf{I})\} \mathbf{1}, k = 1, \dots, K$$

5. Vraćanje na korak 2 sve dok relativna promena u fitu modela nije dovoljno mala

U koraku četiri, \mathbf{d}_k je k -ti red matrice \mathbf{C} , \mathbf{I} je matrica identiteta dimenzije $R \times R$, a $\mathbf{1}$ je vektor dimenzije R sačinjen od jedinica. Izraz $\{(\mathbf{A}'\mathbf{X}_k \mathbf{B} * \mathbf{I})\} \mathbf{1}$ predstavlja vektor sa R kolona čiji su članovi matrice dijagonala matrice dobijene izrazom $\mathbf{A}'\mathbf{X}_k \mathbf{B}$.

Teorijski, konvergencija u fitu modela ne znači neophodno da je došlo i do konvergencije parametara, ali u praksi je to uglavnom slučaj. Uzimajući to u obzir, kriterijum za zaustavljanje algoritma je uglavnom mala relativna promena u fitu modela. Na primer algoritam može da se zaustavi ukoliko je poboljšanje u fitu modela između

dve iteracije manja od 0.0001%. Pri modeliraju jako velikih matrica sam proračun može dosta dugo da traje ukoliko se računa fitovanje modela posle svake iteracije. Ako je algoritam monotono konvergentan, poput algoritma naizmenično najmanjih kvadrata, onda je moguće koristiti druge jednostavnije mere konvergencije algoritma. Jedan od primera takvih mera konvergencije jeste praćenje relativne promene parametara umesto samog modela. Iako ova promena nije ekvivalent promeni u fitu modela, ona će konvergirati simultano sa fitom i samim tim mala promena u parametrima će ukazivati na konvergenciju [Mitchell & Burdick 1994]. Ukoliko su podaci takvi da rešenje ima dvofaktorsku degeneraciju, odnosno procena parametara nikada ne konvergira iako fit konvergira, onda se ručno unosi određen broj iteracija jer bi u suprotnom algoritam imao beskonačan broj iteracija.

2.3.3 Inicijalizacija PARAFAC – ALS

Kao i sa bilo kojom optimizacijom, dobre početne procene parametara mogu da ubrzaju PARAFAC – ALS algoritam i da umanje rizik konvergiranja u lokalnom minimumu. Dobra procena se karakteriše kao ona koja direktnije dovodi do globalnog minimuma. Nasumična inicijalizacija ne daje uvek dobru prvu procenu [Burdick et al. 1990, Harshman & Lundy 1984, Li & Gemperline 1993, Sanchez & Kowalski 1990]. Kod težih slučajeva mogu se javiti dva problema ako inicijalizacija PARAFAC – ALS algoritma nije adekvatna: algoritam daje rešenje u lokalnom minimumu ili proračun zahteva puno vremena. Oba pomenuta problema se najčešće javljaju u dva slučaja: kada su neki *loading* vektori visoko korelisani i kada model nije u potpunosti identifikovan. U prvom slučaju to dovodi do situacije da su regresioni problemi koji se rešavaju tokom iterativnog fitovanja loše uslovljeni. Ovo dovodi do situacije da se fitovanje modela jako sporo odvija. U drugom slučaju neidentifikovani modeli se pojavljuju kada k -rang uslov nije zadovoljen (jednačina (1.24)). Do ovoga može doći ukoliko neki parametar (supstanca) u novom nepoznatom uzorku nije identifikovan, odnosno ako ga nema u uzorcima na osnovu kojih je model građen. Čak i u ovom slučaju je moguće izgraditi model, ali je proračun vremenski zahtevan jer je orijentacija nekih *loading* vektora proizvoljna.

Rešenje prethodno navedenih problema, osim u slučaju degeneracije, jeste korišćenje što bolje inicijalizacije koja se drugačije naziva racionalni start. PARAFAC racionalni start se uglavnom bazira na približnom rešenju dobijenom direktnom trilinearnom dekompozicijom. Racionalni start ima tu prednost da je najčešće vrlo blizu pravog rešenja što štedi puno vremena za proračun. Problem sa racionalnim startom je procena da li dobijeno rešenje konvergira u globalnom minimumu, jer ukoliko se fituje više puta iz iste početne pozicije uvek će se po definiciji dobiti identični parametri. Sličan pristup racionalnom startu je polu racionalni start. On se uglavnom bazira na singularnim vektorima. Singularni vektori približno obuhvata pravi podprostor što je dovoljno da se izbegnu navedeni problemi u mnogo situacija. Iako racionalni i poluracionalni start uglavnom dobro funkcionišu nekad se može desiti da dovedu do lokalnog minimuma. Tada postoji treća mogućnost odnosno nasumični start. Ako se koristi više različitih nasumičnih startova može se videti da li će se negde naći lokalni minimum. Drugim rečima ako su fitovi svih modela sa nasumičnim startom isti onda je to indikacija da je algoritam dobro proračunao parametre, a ukoliko to nije slučaj onda ili postoji lokalni minimum ili je algoritam prerano završio proračun.

2.3.4 Proračun skorova novih uzoraka

Kada je model fitovan i treba da se odrede skorovi novih uzoraka, osnovni problem koji treba rešiti je višestruki linearni regresioni problem. PARAFAC model je dat jednačinom

$$\mathbf{X}_{(I \times JK)} = \mathbf{A}(\mathbf{C} \odot \mathbf{B})' + \mathbf{E} \quad (1.37)$$

za trodimenzionalnu matricu transformisanu u dvodimenzionalnu. Novi uzorak će biti definisan matricom \mathbf{X}_n dimenzije $J \times K$. Oblikovanjem matrice \mathbf{X}_n u skladu sa prethodnim modelom uvodi se vektor dimenzije $1 \times JK$ ($\text{vec } \mathbf{X}_n$) pri čemu model novog uzorka uz korišćenje izračunatih *loading*-a iz prethodnog modela se može definisati na sledeći način

$$(\text{vec } \mathbf{X}_n)' = \mathbf{a}'(\mathbf{C} \odot \mathbf{B})' + \mathbf{E}_n \quad (1.38)$$

gde vektor \mathbf{a} ($R \times 1$) sadrži skor vrednosti novog uzorka (odgovara jednom redu matrice \mathbf{A}) gde su \mathbf{B} i \mathbf{C} već proračunati u prethodnom modelu. Na slici 2.7 je dat grafički prikaz modela novog uzorka.

Matrični prikaz

$$\begin{array}{c} \text{parallelogram } \mathbf{X}_{\text{nov}} \\ \text{parallelogram } \mathbf{C} \\ \text{rectangle } \mathbf{B} \\ \text{vector } \mathbf{a} \end{array} =$$

Vektorski prikaz

$$\begin{array}{c} \text{vector } (\text{vec } \mathbf{X}_{\text{nov}})' \\ \text{rectangle } \mathbf{D}_1 \mathbf{B}' \\ \text{rectangle } \mathbf{D}_k \mathbf{B}' \\ \text{vector } \mathbf{a} \end{array} =$$

Slika 2.7 – Grafički prikaz modela novog uzorka.

Kada se prvi mod modela odnosi na uzorke onda je procena skorova uzorka, čiji se podaci nalaze u matrici \mathbf{X}_n ($J \times K$), definisana u obliku problema najmanjih kvadrata na sledeći način

$$\min_{\mathbf{a}} \|\text{vec } \mathbf{X}_n - (\mathbf{C} \odot \mathbf{B})\mathbf{a}\|^2 \quad (1.39)$$

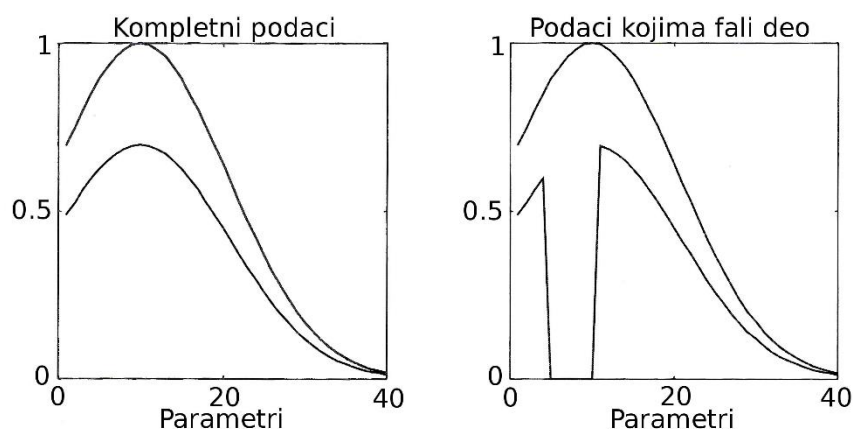
gde su \mathbf{B} i \mathbf{C} dati prethodno izračunatim modelom. Dakle \mathbf{a} se može definisati na sledeći način

$$\mathbf{a} = (\mathbf{C} \odot \mathbf{B})^+ \text{vec } \mathbf{X}_n \quad (1.40)$$

Slični izrazi se mogu izvesti ukoliko se traže parametri drugog ili trećeg moda (pri čemu treba uzeti u obzir da se podrazumeva da su druga dva moda nepromenljiva za sve uzorke).

2.3.5 Tretiranje matrica kojima nedostaju podaci

Prvo treba shvatiti da se podaci koji nedostaju u matrici ne mogu kompenzovati, ali ukoliko su podaci suvišni to onda ne predstavlja problem za analizu. Ukoliko se sve vrednosti podese na vrednost nula to može dovesti do netačnih rezultata. Razmišljanje koje je pogrešno jeste da vrednost nula kojom se menjaju podaci koji nedostaju znači „ništa“ i da ne utiče na rezultat modela. Kod modeliranja to nije slučaj. Kao primer uzmimo u obzir situaciju predstavljenu na slici 2.8.



Slika 2.8 – Primer spektara sa svim podacima i u slučaju kada podaci nedostaju.

Na slici 2.8 u prvom slučaju imamo matricu koja ima sve podatke i u tom slučaju to predstavlja problem sa jednim komponentom jer postoji samo jedan spektralni fenomen. U drugom slučaju je dat isti spektar, ali kod njega fale podaci kod spektra manjeg inteziteta. U principu i u ovom slučaju to mora biti problem sa jednom komponentom. Međutim, vrednosti koje nedostaju su bile zamenjene nulom i očigledno je da je nemoguće modelirati podatke sa vrednostima nula jednokomponentnim modelom. Najbolji način za modeliranje matrica kojima nedostaju podaci jeste pravljenje modela samo na osnovu podataka koji postoje, a to znači da se vrši optimizacija funkcije

gubitka za podatke koji postoje. Stoga se funkcija gubitka za bilo koji model nekompletnih podataka može opisati na sledeći način

$$\min_{\mathbf{M}} \|\mathbf{W} * (\mathbf{X} - \mathbf{M})\|^2 \quad (1.41)$$

gde matrica \mathbf{X} sadrži podatke, \mathbf{M} je model, a $*$ označava Hadamardov (pojedinačno element po element matrice) proizvod. Struktura i ograničenja matrice \mathbf{M} su definisani određenim modelom, na primer \mathbf{AB}' u slučaju bilinearnog modela. Matrica \mathbf{W} sadrži težinske faktore koji imaju vrednost ili jedan ukoliko odgovara elementu početne matrice koji postoji, ili nula ukoliko odgovara elementu početne matrice koji nedostaje. Logičan način modeliranja podataka koji nedostaju je uz primenu težinske regresije sa težinskim faktorima u obliku binarne matrice [Kiers 1997, Paatero 1997]. Ovaj pristup funkcioniše za sve modele koji imaju dobro definisane celokupne kriterijume optimizacije, kao što je slučaj kod PARAFAC modela.

2.3.6 Uslovljavanje nenegativnosti

Prethodno opisan algoritam dozvoljava da *loading* matrice sadrže i pozitivne i negativne vrednosti. To je matematički korektno, ali nije uvek u skladu sa prethodnim znanjem vezanim za merene uzorke. Nekada predznanje o uzorku uslovljava “nenegativnost” za neke modove modela, na primer koncentracije ne mogu imati negativne vrednosti, kao što ni pravilno izmereni spektri uglavnom ne mogu biti negativni. Ovo je vrlo bitna informacija koja se može nekad uneti direktno u algoritam. Samim tim se može reći da fitovanje modela kojem je uslovljena nenegativnost za određene modove može biti jako praktično u nekim slučajevima. Takođe, Lawton i Sylvestre [Lawton & Sylvestre 1971] su pretpostavili da korišćenje nenegativnosti bitno umanjuje prostor u kome treba izračunati parametre.

Nenegativnost je manje ograničen problem, gde je svaki parametar ograničen samo time da ima vrednost veću ili jednaku nuli. Takav problem se može efikasno rešiti

algoritmom aktivnog seta [Gill et al. 1980]. Dalje će biti detaljno objašnjeno na koji način se taj algoritam implementira u algoritam naizmenično najmanjih kvadrata.

Uzmimo u obzir PARAFAC model matrice \mathbf{X} veličine $I \times J \times K$. Model se najčešće fituje minimiziranjem funkcije gubitka

$$\min_{\mathbf{A}, \mathbf{B}, \mathbf{C}} \|\mathbf{X} - \mathbf{A}(\mathbf{C} \odot \mathbf{B})'\|^2 \quad (1.42)$$

Ukoliko je prethodno poznato da parametri modela moraju biti nenegativni (na primer ukoliko predstavljaju relativnu koncentraciju), onda je moguće fitovati ograničenu verziju tog modela gde je uslovljena nenegativnost. Ovaj problem se može definisati na sledeći način

$$\min_{\mathbf{A}, \mathbf{B}, \mathbf{C}} \|\mathbf{X} - \mathbf{A}(\mathbf{C} \odot \mathbf{B})'\|^2 \quad a_{ir} \geq 0, b_{jr} \geq 0, c_{kr} \geq 0 \text{ za } i = 1, \dots, I; j = 1, \dots, J; k = 1, \dots, K; r = 1, \dots, R. \quad (1.43)$$

Elementi a_{ir} predstavljaju elemente *loading* matrice \mathbf{A} prvog moda veličine $I \times R$, dok su b_{jr} i c_{kr} elementi matrice \mathbf{B} i \mathbf{C} , respektivno. Razlika između neuslovljenog i uslovljenog problema je samo ograničenje da svi parametri moraju biti nenegativni. Algoritam za fitovanje modela može biti podeljen uz primenu naizmenično najmanjih kvadrata vrlo slično kao i kod neuslovljenog modela. *Loading* matrice \mathbf{A} , \mathbf{B} i \mathbf{C} se ažuriraju iterativno dok su drugi parametri fiksni. Ažuriranje na primer parametara prvog moda je predstavljeno sledećom jednačinom

$$\min_{\mathbf{A}} \|\mathbf{X} - \mathbf{A}(\mathbf{C} \odot \mathbf{B})'\|^2 \quad a_{ir} \geq 0 \text{ za } i = 1, \dots, I; r = 1, \dots, R. \quad (1.44)$$

Pošto je ažuriranje parametara za dato i (dati red matrice \mathbf{A}) nezavisno od drugih parametara u matrici \mathbf{A} , onda se svaki red može odvojeno ažurirati dajući globalno rešenje najmanjih kvadrata za jednačinu (1.44). Stoga, umesto razmatranja problema datog jednačinom (1.44) dovoljno je razmatrati sledeći problem

$$\min_{\mathbf{a}_i} \|\text{vec } \mathbf{X}_i - (\mathbf{C} \odot \mathbf{B}) \mathbf{a}_i\|^2 \quad a_{ir} \geq 0 \text{ za } f = 1, \dots, F. \quad (1.45)$$

gde je $\text{vec } \mathbf{X}_i$ i -ti horizontalni *slice* (uzorak i) reorganizovan u vektor. Vektor \mathbf{a}_i sadrži parametre u i -tom redu matrice \mathbf{A} . Rešavanjem ovog problema za sve vrednosti i ($i = 1, \dots, I$) dobija se ažuriranje najmanjim kvadratima za matricu \mathbf{A} . Ovaj problem je ekvivalentan problemu standardne višestruke linearne regresije i predstavlja osnovu jednog od najviše korišćenih algoritama za izračunavanje parametara ograničenih nenegativnošću.

2.3.7 Validacija i dijagnostika PARAFAC modela

Analiza podataka je često bazirana na evaluaciji fitovanih modela. Ovi modeli nam daju aproksimaciju i interpolaciju realnih sistema koji se posmatraju. Često postoji veliki izbor modela i njihov kvalitet zavisi od raznih faktora. Razlog tome je taj što matematički i statistički modeli koji se baziraju na realnim podacima imaju karakteristike koje zavise ne samo od samog sistema koji se posmatra već i od uzorkovanja, karakteristika modela, algoritma itd. Validacija modela se može posmatrati kao analize gde se ispituje da li se iz modela mogu izvesti validni zaključci. Pri validaciji se proverava na primer da li su varijacije u podacima predstavljene na najjednostavniji način, da li model efikasno vrši predikcije, da li je na model uticao neki određen uzorak ili parametar u smislu da model više opisuje varijacije tog jednog uzorka ili karakteristike, da li je algoritam konvergirao itd. Harshman [Harshman 1984] deli validaciju na četiri dela:

- teorijska podesnost;
- tačnost proračuna;
- statistička pouzdanost;
- validnost objašnjenja.

Teorijska podesnost modela se odnosi na izbor modela. Da li je posmatrana struktura po prirodi bilinearna ili trilinearna? Da li je neophodno predprocesiranje ili linearna transformacija? To su samo neka od pitanja koja su specifična za ovaj deo validacije.

Tačnost proračuna se uglavnom bavi problemom da li je algoritam konvergirao u željenom globalnom minimumu ili neželjenom lokalnom minimumu i zašto je do toga došlo. Takođe je bitna greška zaokruživanja, kao i reproduktivnost proračuna. To sve predstavlja probleme algoritma.

Statistička pouzdanost je vezana za podesnost pretpostavke distribucije, stabilnost rešenja, izbor dimenzionalnosti parametara modela itd. Statističku pouzdanost je često jako teško kvantifikovati u slučaju analize realnih podataka (zbog malog broja uzoraka ili lošeg predznanja vezanog za posmatrani sistem).

Validnost objašnjenja se odnosi na procenu koliko model dobro reflektuje stvarni fenomen koji se istražuje odnosno podesnost modela. Da li ostaci ukazuju da postoji još sistemske varijacije? Da li rezultati potvrđuju zaključke izvučene iz modela?

Bitan deo validacije je primena dijagnostike za karakterizaciju kvaliteta modela. Većina dijagnostika modela se mogu grafički prikazati, što ih čini lakšim za istraživanje i pregled. Harshman [Harshman 1984] deli dijagnostiku na četiri nivoa: nula-fit (eng. zero-fit) dijagnostika, jedan-fit (eng. one-fit) dijagnostika, mnogo-fit (eng. many-fit) dijagnostika (jedan set podataka) i mnogo-fit (eng. many-fit) dijagnostika (više setova podataka).

Zero-fit dijagnostike su one vezane za podatke pre nego što je model bio fitovan. One podrazumevaju proveru očiglednih *outlier*-a (uzoraka koji se na neki način izdvajaju od grupe merenih uzoraka) i pouzdanosti podataka.

One-fit dijagnostike su one koje se koriste za fitovanje određenog modela i koje su korisne za validaciju tog modela. Pri tome se proverava statistička doslednost modela. Dijagnostike se mogu podeliti na one koje se odnose na parametre i one koje se bave ostacima.

Many-fit dijagnostike (jedan set podataka) su korisne za poređenje ili procenu modela sa različitim početnim uslovima ili različite forme modela. I u ovom slučaju se one mogu podeliti na dijagnostike parametara i dijagnostike ostataka.

Many-fit dijagnostike (više setova podataka) se koriste kada imamo više setova podataka koji opisuju isti problem ili kada je set toliko veliki da se može podeliti na više manjih setova podataka. U prvom slučaju je moguće primeniti test-set validaciju, dok u

drugom slučaju se može koristiti unakrsna validacija, raspolovljena (eng. split-half) analiza itd.

Svi prethodno opisani delovi validacije i dijagnostike u praksi se svode na rešavanje tri specifična problema:

- unakrsna validacija i test-set validacija;
- odabir broja komponenti modela;
- analiza uticaja i ostataka.

2.3.8 Test-set i unakrsna validacija

Test-set i unakrsna validacija predstavljaju standardne metode koje se koriste u različite svrhe pri validaciji. One nam omogućavaju da dobijemo realnije ostatke modela od onih koji su dobijeni pri proračunu modela. Stoga su test-set i unakrsna validacija veoma korisne kada se ostaci koriste za određivanje *outlier*-a, za dobijanje procenta objašnjene varijacije u modelu itd. U nekim slučajevima ostaci modela umeju da budu različiti od ostataka koji će se dobiti kada se model primeni na nove podatke, iz čega se može zaključiti da ponekad struktura i veličina ostataka modela mogu dovesti do pogrešnih zaključaka i rezultata. Test-set i unakrsna validacija mogu da umanje ovaj problem pružajući ostatke koji tačnije reflektuju ostatke koje je moguće dobiti na novim podacima. Ostaci dobijeni na taj način se dalje mogu koristiti za više stvari: određivanje broja komponenta modela, procenjivanje predikcione sposobnosti modela, upoređivanje unakrsno validiranih i fitovanih ostataka radi utvrđivanja optimalnosti modela i analiza ostataka. Može se zaključiti da validirani ostaci daju objektivniju informaciju u pogledu buduće primene modela.

Test-set validacija podrazumeva fitovanje trenutnog modela na osnovu novih podataka. Ostaci se zatim dobijaju jednostavnim oduzimanjem modela novih podataka od početnih podataka. Ova validacija simulira buduću primenu ovog modela u praksi na novim podacima. Ukoliko je test set dobijen nezavisno od kalibracionih podataka onda je test set validacija definitivna validacija modela jer se novi podaci mogu posmatrati

kao izvučeni set uzoraka iz populacije uzoraka nekih budućih merenja. Bitno je takođe napomenuti da mali broj test uzoraka može da da nepouzdan rezultate.

Jedan od problema koji se javljaju kod test-set validacije u praksi je taj da dobijeni ostaci nisu nezavisni od podataka korišćenih za fitovanje modela. Uzmimo jednostavan primer dvodimenzionalnog PCA modela. Ukoliko je *loading* matrica \mathbf{P} ($J \times R$) dobijena modeliranjem kalibracionog seta podataka dok se PCA model testira test set uzorcima ili izostavljenim uzorcima tokom unakrsne validacije, onda se ostaci dobijaju tako što se prvo računaju skorovi za nove podatke \mathbf{X}_{test} na sledeći način

$$\mathbf{T}_{\text{test}} = \mathbf{X}_{\text{test}}\mathbf{P} \quad (1.46)$$

sa pretpostavkom da je \mathbf{X}_{test} bio na pravi način predprocesiran (slično kalibracionim podacima). Zbir kvadrata ostatka za test set je dat sledećom jednačinom

$$\|\mathbf{X}_{\text{test}} - \hat{\mathbf{M}}\|^2 = \|\mathbf{X}_{\text{test}} - \mathbf{T}_{\text{test}}\mathbf{P}'\|^2 = \|\mathbf{X}_{\text{test}} - \mathbf{X}_{\text{test}}\mathbf{P}\mathbf{P}'\|^2 \quad (1.47)$$

Može se jasno videti da vrednosti modela ($\mathbf{X}_{\text{test}}\mathbf{P}\mathbf{P}'$) nisu nezavisne od podataka (\mathbf{X}_{test}) i samim tim „overfitovanje“ može bit problem čak i za test set podataka. Stoga da bi se poredili različiti modeli na osnovu zbira kvadrata računatog na ovaj način, neophodno je uneti neke korekcije stepena slobode [Martens & Næs 1989]. Kako oni nisu jednostavni za definisanje, pogotovo za višedimenzionalne modele, najčešće se koriste umerene aproksimacije.

Alternativa test-set validacije je unakrsna validacija. Unakrsna validacija se može posmatrati kao simulacija test-set validacije gde se test-set validacija ponavlja koristeći svaki put drugačije uzorke ili parametre u test setu. U skladu sa tim, unakrsna validacija je posebno korisna ukoliko je broj uzoraka ograničen. Takođe određene vrste unakrsne validacije su korisne za validaciju modela jer prethodno pomenuti problem zavisnih ostataka je eliminisan samom konstrukcijom unakrsne validacije.

Unakrsna validacija predstavlja metodu internog odabira uzorka za testiranje i za kalibraciju. Osnovna ideja unakrsne validacije je vrlo jednostavna i sastoji se iz pet koraka:

- 1) Izostaviti jedan deo podataka (uzoraka).
- 2) Izgraditi model bez tih podataka.
- 3) Koristiti dobijeni model za predikciju parametara izostavljenih uzoraka.
- 4) Izračunati odgovarajuću grešku ostatka.
- 5) Ponoviti ovu proceduru tako da svaki uzorak bude jednom izostavljen i sumirati sve kvadratne greške predikcije (eng. Predicted residual sum of squares – PRESS).

U slučaju kada se unakrsna validacija koristi za određivanje broja komponenti modela, bira se broj komponenti za koji je vrednost PRESS-a najmanja. U praksi, minimalne PRESS vrednosti nisu uvek prisutne ili je ponekad minimum samo malo bolji od jednostavnijeg modela. U takvim slučajevima je mnogo teže odrediti broj komponentata i tada se dodatno moraju uzeti u obzir: predznanje o sistemu, drugi rezultati modela i celokupni cilj analize. Kada treba izabrati koje vrednosti (uzorke) treba izostaviti pri validaciji, treba imati u vidu da će različiti pristupi dovesti do različitih rezultata. Na primer ukoliko se izostave samo replikati dobiće se greška ponovljivosti, a ukoliko se izostave uzorci iz novog eksperimenta iz druge laboratorije dobiće se greška reproduktivnosti.

PRESS se koristi za dobijanje procene varijanse ili kao pomoć pri odabiru broja komponentata modela. PRESS će biti u sledećem obliku

$$\text{PRESS} = \sum_{i=1}^I (y_i - \hat{y}_{(i)})^2 \quad (1.48)$$

gde član $\hat{y}_{(i)}$ je predikcija od y_i uz korišćenje modela koji je izgrađen bez korišćenja i -tog objekta pri proceni modela. Ovaj objekat može biti i segment nekoliko individualnih objekata što se tada naziva segmentna unakrsna validacija. Daljim proširivanjem ove

segmentacije, segmenti mogu biti nezavisni test setovi, u tom slučaju unakrsna validacija će dati rezultate slične rezultatima test-set validacije.

2.3.9 Odabir broja komponenti modela

Jedna od najbitnijih stavki kod PARAFAC analize jeste odabir odgovarajućeg broja komponenti modela. Odgovarajuća dimenzionalnost modela nije samo funkcija podataka već i funkcija konteksta i cilja same analize. Stoga, prikladan PARAFAC model za ispitivanje seta podataka može da ima broj komponenta različit od PARAFAC modela čiji se skorovi kasnije koriste za regresioni model.

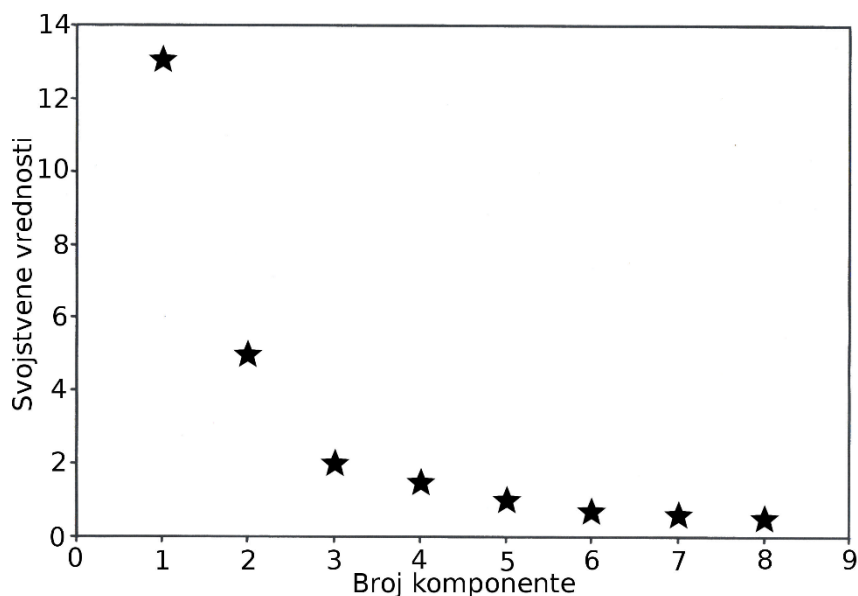
Postoje razni alati koji se koriste za odabir broja komponenta. Neki se baziraju na statističkim pretpostavkama i testiranju značajnosti, dok se neki baziraju na empirijskim pravilima. Generalno se ne treba oslanjati na samo jedno pravilo za određivanje dimenzionalnosti. Svako pravilo ima svoje uslove i nijedni realni podaci neće ispuniti u potpunosti sve njih. Za određivanje broja komponenta se najčešće koriste sledeći alati: grafik svojstvenih vrednosti faktora (eng. scree plot), *split-half* analiza, unakrsna validacija i vizuelna provera ostataka. Takođe predznanje o fenomenu koji se krije u podacima često može pomoći pri proceni modela.

2.3.10 Grafik svojstvenih vrednosti faktora

U daljem tekstu biće opisano na koji način se grafik svojstvenih vrednosti faktora koristi za određivanje broja komponenta. Radi jednostavnosti objašnjenja biće opisano na koji način se ovom metodom određuje broj komponenta PCA modela, a zatim će biti prikazano na koji način se to modifikuje za PARAFAC model.

Najčešće korišćena analiza za određivanje broja komponenta PCA modela jeste grafik svojstvenih vrednosti faktora [Cattell 1966]. Grafik daje svojstvene vrednosti unakrsnog proizvoda matrica sortiranih po veličini u funkciji broja komponenta. Prva svojstvena vrednost je jednaka zbiru kvadrata objašnjenih prvom komponentom, druga vrednost je jednaka zbiru kvadrata objašnjenih drugom komponentom itd. Suma svih svojstvenih

vrednosti je jednaka tragu unakrsnog proizvoda matrice. Singularne vrednosti dobijene dekompozicijom singularnom vrednošću podataka takođe mogu biti korišćene umesto svojstvenih vrednosti. Singularne vrednosti su jednake kvadratnom korenu odgovarajućih svojstvenih vrednosti. Svojstvene vrednosti se prikazuju na vertikalnoj osi grafika u funkciji broja komponenata na horizontalnoj osi (slika 2.9). Grafik svojstvenih vrednosti faktora kod PCA se takođe može prikazati kumulativno jer svojstvene vrednosti su nezavisne i sabiraju se do 100% zbira kvadrata. Određena granična vrednost na grafiku se koristi za određivanje komponenata koje su previše male da bi se koristile. Kada se vrši samo istraživanje podataka dovoljno je da se izabere granična vrednost od 80% objašnjenih varijacija kod podataka sa šumom, ali kada je kvantitativna namena u pitanju onda je korisno imati složeniju determinaciju odgovarajuće granične vrednosti. Uglavnom je broj komponenata odabran na onoj graničnoj vrednosti gde grafik počinje linearno da opada.



Slika 2.9 – Primer grafika svojstvenih vrednosti faktora.

Sličan grafik se može dobiti i za PARAFAC model prikazivanjem zbira kvadrata individualnih komponenata. Međutim u ovom slučaju nije moguće napraviti direktno kumulativni grafik jer varijanse pojedinačnih komponenti nisu aditivne. Može se zaključiti da ovaj grafik nije direktno koristan za PARAFAC modele. S druge strane,

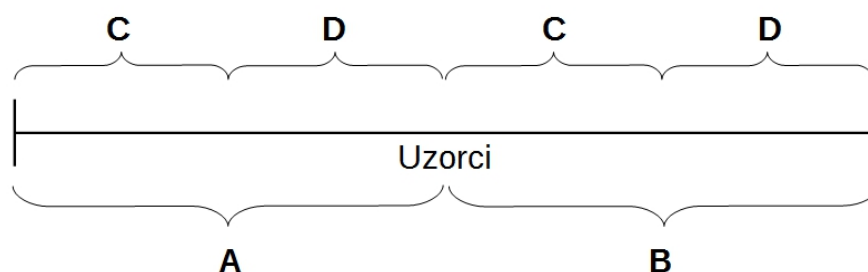
kumulativni grafik za PARAFAC modele može bit konstruisan prikazivanjem objašnjenih ili ostataka zbira kvadrata za jednokomponentni model, dvokomponentni model itd. Takav grafik će dati sličnu informaciju kao grafik svojstvenih vrednosti, sa izuzetkom što će se faktori menjati za svaki model jer PARAFAC nije sekvencijalno fitovan. Osnovni princip interpretacije grafika ostaje isti: broj komponenata koji će se koristiti se bira kao broj komponenata kod kojeg pad varijacije ostataka na grafiku postaje linearan.

2.3.11 Split-half analiza

Kod *split-half* analize različiti podsetovi podataka se nezavisno analiziraju. Zbog jedinstvenosti PARAFAC modela, kada je odabran odgovarajuću broj komponenata moraju se dobiti isti rezultati odnosno isti *loadings*-i za bilo koji podset podataka istog sistema. Da bi se ocenilo da li su dva modela ista moraju se obratiti pažnja na neodređenosti trilinearnog modela koje se mogu javiti: redosled i razmera mogu se menjati ukoliko nisu fiksirane algoritmom. Ukoliko model ima previše ili premalo komponenata, parametri modela će se razlikovati kada se model gradi na osnovu različitih setova podataka. Ukoliko je model stabilan u smislu *split-half* analize onda je to očigledna indikacija da je model realan, odnosno da realno oslikava varijacije sistema, a ne karakteristike samo jednog određenog uzorka. S druge strane ukoliko neke komponente nisu stabilne, to ukazuje da one ili reflektuju neki šum ili je pretpostavljen pogrešan broj komponenata. Takav model nije validan.

Kada se radi *split-half* analiza mora se odlučiti po kom će se modu deliti podaci. Deljenje se mora vršiti u onom modu gde postoji dovoljan broj razdvojenih parametara/objekata, na taj način da svaki podset obuhvati ceo domen i da svi parametri mogu biti izračunati iz individualnih podsetova podataka. Najčešće se u praksi dele podaci po modu koji oslikava uzorke. Način podele podataka na dva podseta je prikazan na slici 2.10. Podset A čini prva polovina uzoraka, a B druga polovina. Pošto se slučajno može dogoditi da jedan od ovih podsetova ne sadrži informaciju o svim latentnim fenomenima u sistemu, generišu se još dva seta C i D kako bi se osiguralo da se u svim podsetovima nalaze sve varijacije sistema. Set C čini prvu polovinu uzoraka setova A i

B, a D čine druga polovina A i B. Model se računa na osnovu svakog od ovih setova i ako se rezultati setova A i B ili C i D replikuju, onda je tačnost rezultata na taj način empirijski potvrđena. U idealnom sličaju treba da se dobiju isti rezultati za A, B, C i D. Ukoliko to nije ispunjeno, dovoljan je uslov da se dobije isti rezultat samo za jedan par podsetova (A/B ili C/D) jer je malo verovatno da rezultati budu isti za jedan par ako su uzorci nezavisni i rešenje nasumično. Razlog dobijanja različitih rezultata za jedan par podsetova može da bude previše mali broj uzoraka. U tom slučaju se može desiti da nije moguće podeliti četiri različita seta podataka, a da pri tome svaki od njih sadrži informaciju o svim bitnim fenomenima sistema.



Slika 2.10 – Primer podele seta uzoraka za *split-half* analizu.

2.3.12 Analiza ostatka

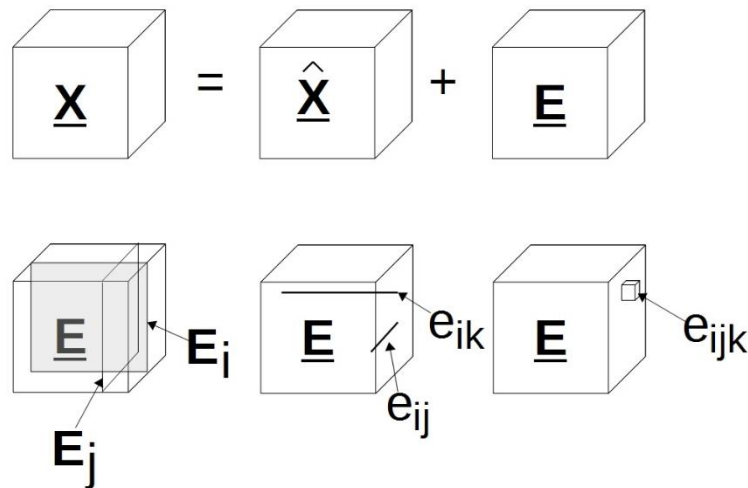
Svaki trodimenzionalni PARAFAC model se može predstaviti sledećom formulom:

$$\underline{\mathbf{X}} = \underline{\hat{\mathbf{X}}} + \underline{\mathbf{E}} \quad (1.49)$$

Glavni cilj analize jeste dobijanje $\underline{\hat{\mathbf{X}}}$, ali ostatak $\underline{\mathbf{E}}$ takođe može da da bitne indikacije o kvalitetu modela. Ono što se normalno očekuje da se nađe u ostatku jeste šum ili neke distribucije koje su simetričnog oblika, što se može verifikovati grafičkim prikazom ostataka ili uz primenu dijagnostičkih alata.

Činjenica da je $\underline{\mathbf{E}}$ trodimenzionalna matrica veličine $I \times J \times K$ kao i matrica $\underline{\mathbf{X}}$, daje nam mogućnost da ekstrahujemo podmatrice svakog uzorka ili parametra i da posmatramo i upoređujemo za svaki od njih zasebno promenu varijacija u ostatku. To je prikazano na

slici 2.11. Ove podmatrice se mogu sumirati na različite načine. Prva podmatrica je sama $\underline{\mathbf{E}}$ matrica, dok se zbir kvadrata matrice $\underline{\mathbf{E}}$ koristi za grafik svojstvenih vrednosti faktora, što je prethodno bilo opisano. Matrica $\underline{\mathbf{E}}$ takođe se može podeliti na matrice \mathbf{E}_i ($i = 1, \dots, I$), \mathbf{E}_j ($j = 1, \dots, J$) i \mathbf{E}_k ($k = 1, \dots, K$). Ove matrice se mogu ispitivati koristeći njihov zbir kvadrata, ili ponekad njihove varijanse, standardne devijacije itd. Kao rezultat zbira kvadrata ovih matrica se dobijaju tri vektora vrednosti \mathbf{s}_I ($I \times 1$), \mathbf{s}_J ($J \times 1$) i \mathbf{s}_K ($K \times 1$) gde i -ti element vektora \mathbf{s}_I predstavlja zbir kvadrata od \mathbf{E}_i itd. Ovi vektori se mogu posmatrati na različite načine, ali ono našta se treba fokusirati jesu *outlier*-i, zaostale strukture ili artefakti. Slika 2.12 prikazuje praktičnu interpretaciju takvih vektora.

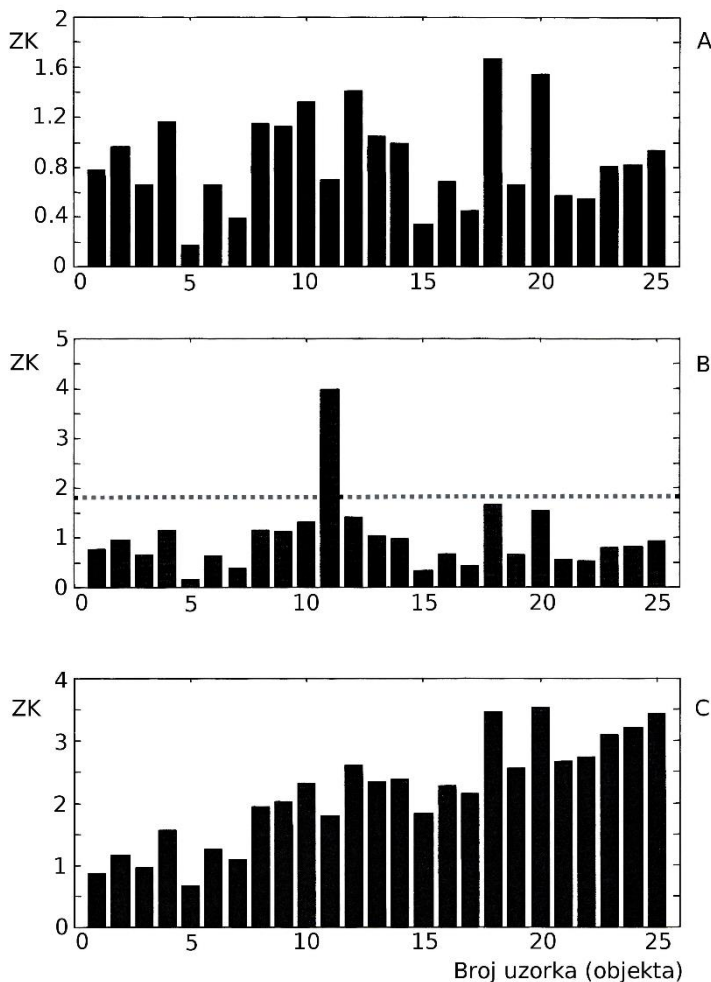


Slika 2.11 – Model matrice $\underline{\mathbf{X}}$ i njena odgovarajuća matrica ostatka $\underline{\mathbf{E}}$ i njene podmatrice.

Ukoliko $\underline{\mathbf{X}}$ ima na primer I uzoraka, J talasnih dužina emisije i K talasnih dužina ekscitacije, ispitivanje \mathbf{s}_I će indikovati ukoliko postoje *outlier*-i ili neki trend među uzorcima. Na isti način se može videti ako postoje *outlier*-i među talasnim dužinama emisije i ekscitacije ispitivanjem \mathbf{s}_J i \mathbf{s}_K , respektivno. Kada su spektri u pitanju određeni diskontinuiteti u nekim delovima vektora \mathbf{s}_J i \mathbf{s}_K ukazuju na neke hemijske ili fizičke artefakte.

Zbir kvadrata podelemenata od matrice $\underline{\mathbf{E}}$ mogu biti izračunati i za redove, kolone i tube: \mathbf{e}_{ij} , \mathbf{e}_{jk} i \mathbf{e}_{ki} (slika 2.11). Zbir kvadrata svih redova \mathbf{e}_{jk} formira matricu \mathbf{S}_{JK} dimenzije

($J \times K$). Kada J i K imaju velike vrednosti, što je slučaj kod fluorescentnih spektara, neophodno je nacrtati konturni grafik matrice S_{JK} . U tom slučaju se na grafiku mogu videti površine koje liče na EEM koje predstavljaju zbir kvadrata ostatka za svaki par talasne dužine ekscitacije i emisije. Ispitivanjem ovog grafika može se videti da li je neka struktura, koja je bitna za fenomen koji posmatramo, ostala u ostatku i nije obuhvaćena modelom. Ukoliko model dobro opisuje posmatrani sistem u ovim graficima ne bi smelo da ima nekih preostalih struktura (pikova).



Slika 2.12 – Različiti primeri zbira kvadrata (ZK) ostataka. (A) Zbir kvadrata preseka koji odgovaraju objektima matrice \underline{E} pokazuju očekivane varijacije. (B) U ovom slučaju se uočava jako veliko odstupanje kod uzorka broj 11, što ukazuje na mogućnost da je taj uzorak *outlier*. (C) Na poslednjem grafiku se vidi da u ostatku postoji određeni trend među uzorcima koji sam model nije obuhvatio.

2.3.13 Predprocesiranje podataka

Pre svakog proračuna PARAFAC modela uglavnom je neophodno pripremiti podatke za dalju analizu odnosno predprocesirati. Predprocesiranje podrazumeva centriranje podataka srednjom vrednošću i skaliranje. Centriranje podataka srednjom vrednošću

podrazumeva nalaženje srednje vrednosti matrice i oduzimanje te vrednosti od svih članova matrice čime se dobija da srednja vrednost matrice bude nula. Na ovaj način se uklanjaju odstupanja u podacima ukoliko postoje. Skaliranje se vrši u slučaju kada su parametri različitih mernih jedinica i neophodno je skaliranje kako bi njihov uticaj na varijacije u modelu bio podjednak. Ukoliko neki parametar ima velike varijacije model će pretpostaviti da je on najbitniji za opis sistema. Pošto kod fluorescentnih spektara ne postoje odstupanja u podacima (vrednost nula u spektru znači nula koncentraciju te supstance) i svi parametri (talasne dužine) su istih mernih jedinica (nanometri) onda nema potrebe za skaliranjem i centriranjem podataka. Iz tog razloga skaliranje i centriranje neće biti dalje detaljno objašnjeni.

Jedina neophodna obrada fluorescentnih spektara pre pristupanja PARAFAC analizi jeste uklanjanje Rejljevog rasejanja prvog i drugog reda iz razloga što ono ne opisuje varijacije sistema. Rasejanje se iz spektara uklanja tako što se na mestima u matrici gde se javlja rasejanje vrednost intenziteta menja vrednošću NaN (eng. Not a Number), odnosno podešava se da ta vrednost nedostaje, dok se delovi spektra ispod Rejljevog rasejanja prvog reda i iznad Rejljevog rasejanja drugog reda podešavaju na vrednost nula, ukoliko u tim delovima ne postoje pikovi koji su bitni za model.

2.3.14 Vizuelizacija rezultata

Za predstavljanje rezultata PARAFAC modela mogu se koristiti razni grafički prikazi. Svi grafici imaju za cilj da pomognu korisniku da interpretira svoje rezultate. Vrsta grafika koji će se koristiti puno zavisi i od vrste podataka koji se modeliraju i problema koji treba da se reši. U daljem tekstu će biti opisani grafici koji se koriste za prikaz i interpretaciju modela na osnovu fluorescentnih spektara.

Stubasti dijagram prikazuje skalarne vrednosti vektora u obliku stubova različitih dužina. Pozicije stubova ukazuju na poziciju u vektoru. Stubasti dijagrami se koriste za prikaz zbira kvadrata objašnjenih varijacija za svaku komponentu, prikaz zbira kvadrata ostataka u funkciji uzoraka itd.

Linijski dijagram grafik vektora gde na horizontalnoj osi je data pozicija u vektoru, često predstavlja kontinualni parametar (npr. spektar), dok je na vertikalnoj osi data skalarna vrednost u vektoru. Ove tačke mogu biti spojene pravim ili krivim linijama. Ovakvi dijagrami se koriste za prikaz spektara, ali i za prikaz *loading* vektora dobijenih na osnovu njih.

Grafik rasutosti (eng. Scatter plot) predstavlja grafički prikaz dva vektora iste veličine prikazanih jedan nasuprot drugog, pri čemu se skalarne vrednosti vektora koriste kao koordinate tačaka. Ovakvi grafici se koriste za prikaz dva loading vektora, dva skor vektora itd. Na njima se mogu jasno uočiti *outlier*-i, grupisanja i trendovi među uzorcima.

2.4 Metoda potpornih vektora

Metoda potpornih vektora (eng. Support Vector Machine – SVM) je multivarijantna tehnika koja se koristi za građenje klasifikacionih i regresionih modela. Prednost ove metode se ogleda u tome što ne zahteva veliki broj uzoraka za obučavanje modela i nije osetljiva na prisustvo “outlier”-a [Burges 1998]. Koncept klasifikacije primenom SVM metode podrazumeva transformaciju početnog prostora podataka u prostor veće dimenzije kako bi klase bilo moguće linearno razdvojiti. Koristeći početne podatke kao koordinate uzoraka, SVM pronalazi hiper ravan u transformisanom prostoru koja deli prostor na taj način da što veći broj uzoraka koji pripadaju istoj klasi budu sa iste strane ravni pri tome maksimizirajući udaljenost svake klase od hiper ravni. Optimalna hiper ravan umanjuje rizik pogrešnog klasifikovanja test uzoraka.

Kao što je prethodno navedeno osnovni princip SVM metode je određivanje optimalne hiper ravni za separaciju, koja maksimizira rastojanje između dve klase u multidimenzionalnom prostoru.

Pretpostavimo da imamo linearno razdvojiv set podataka S koji se sastoji od m objekata, gde svaki n -dimenzionalni objekat \mathbf{x} je definisan sa n koordinata ($\mathbf{x} = (x_1, x_2, \dots, x_n)$),

gde je svako x_i realan broj). Bilo koja hiper ravan u S prostoru može biti opisana na sledeći način

$$\{\mathbf{x} \in S \mid \mathbf{w} \cdot \mathbf{x} + b = 0\}, \quad \mathbf{w} \in S, b \in R, \quad (1.50)$$

gde je proizvod definisan izrazom

$$\mathbf{w} \cdot \mathbf{x} = \sum_{i=1}^n w_i x_i. \quad (1.51)$$

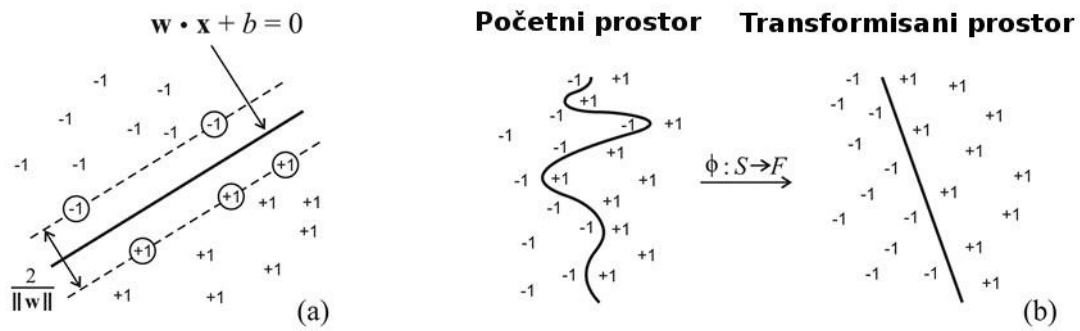
Uzimimo da svaki objekat \mathbf{x}_j , $j = (1, 2, \dots, m)$, pripada klasi $y_j \in \{-1, +1\}$, zadatak je u tom slučaju naći hiper ravan koja deli prostor S na takav način da svi objekti koji pripadaju različitim klasama leže u prostoru sa različitih strana hiper ravni. Drugim rečima, treba naći par (\mathbf{w}, b) koja zadovoljava sledeću jednačinu

$$y_i(\mathbf{w} \cdot \mathbf{x}_i + b) \geq +1, \quad (1.52)$$

za svako $i=1$ do n . Tada jednačina hiper ravni koja razdvaja objekte (slika 2.13) postaje

$$\mathbf{w} \cdot \mathbf{x} + b = 0. \quad (1.53)$$

U slučaju kada su klase linearno razdvojive, što je trenutna pretpostavka, postoji bar jedna ili više hiper ravni (\mathbf{w}, b) koje zadovoljavaju prethodni uslov. Svaka takva hiper ravan predstavlja klasifikator koji tačno razdvaja sve objekte. Među njima SVM pokušava da nađe jedan objekat koji ima najveće rastojanje od najbliže tačke ravni i na taj način maksimizira odstojanje između klasa za bolju generalizaciju modela. Ova vrsta hiper ravni se naziva optimalna separatišuća hiper ravan (eng. optimal separating hyperplane – OSH).



Slika 2.13 – (a) Optimalna separatišuća hiper ravan; (b) Mapiranje linearno nerazdvojivih podataka u prostor veće dimenzionalnosti omogućava njihovo linerno razdvajanje.

Ako uzmemo u obzir da odstojanje između objekata različitih klasa je $2/\|\mathbf{w}\|$, drugi zadatak je da se među svim mogućim hiper ravnima nađe jedna koja ima najveću vrednost $2/\|\mathbf{w}\|$, što je ekvivalentno minimumu $\|\mathbf{w}\|^2/2$. Time problem optimizacije postaje

$$\min \frac{\|\mathbf{w}\|^2}{2} \tag{1.54}$$

sa ograničenjem

$$y_i(\mathbf{w} \cdot \mathbf{x}_i + b) - 1 \geq 0, \quad i = 1, \dots, m. \tag{1.55}$$

što predstavlja minimizaciju kvadratne funkcije sa uslovljenom linearnošću.

Ovaj problem može biti rešen korišćenjem klasične metode Lagranžovih množioca u formulaciji dualnog problema:

$$\max L(\mathbf{w}, b, \Lambda) = \sum_{i=1}^m \lambda_i - \frac{1}{2} \sum_{i=1}^m \sum_{j=1}^m \lambda_i \lambda_j y_i y_j \mathbf{x}_i \cdot \mathbf{x}_j, \tag{1.56}$$

sa ograničenjem

$$\sum_{i=1}^m \lambda_i y_i = 0, \quad (1.57)$$

gde su $\Lambda = (\lambda_1, \lambda_2, \dots, \lambda_m)$ Lagranžovi množiocci, $\lambda_i \geq 0$, i predstavljaju rešenje optimizacionog problema. Objekti SVM modela koji imaju $\lambda_i > 0$ predstavljaju potporne vektore, dok objekti koji imaju $\lambda_i = 0$ nisu bitni i ne doprinose SVM modelu.

Parametri (\mathbf{w}_o, b_o) OSH-a se mogu izračunati na sledeći način

$$\mathbf{w}_o = \sum_{i=1}^m \lambda_i y_i \mathbf{x}_i = \sum_{i=1}^{Ns} \lambda_i y_i \mathbf{s}_i. \quad (1.58)$$

i b_o predstavlja srednju vrednost dobijenih vrednosti b za sve objekte potpornih vektora

$$b = y_j - \sum_{i=1}^m \lambda_i y_i \mathbf{x}_i \cdot \mathbf{x}_j = y_j - \sum_{i=1}^{Ns} \lambda_i y_i \cdot \mathbf{s}_i \cdot \mathbf{x}_j. \quad (1.59)$$

Ovde, \mathbf{s} predstavlja potporne vektore, a Ns je njihov broj.

Bilo koji novi objekat može biti klasifikovan uz primenu izraza $\mathbf{w}_o \cdot \mathbf{x} + b$. Međutim, takođe je moguće klasifikovati objekte bez računanja \mathbf{w}_o . U ovom slučaju mogu se koristiti potporni vektori iz seta za obučavanje i odgovarajući Lagranžovi množiocci za pronalaženje funkcije odluke $f(\mathbf{x}_{nov})$ za klasifikaciju novog objekta \mathbf{x}_{nov} :

$$f(\mathbf{x}_{nov}) = \text{sgn} \left(\sum_{i=1}^m \lambda_i y_i \mathbf{x}_i \cdot \mathbf{x}_{nov} + b \right) = \text{sgn} \left(\sum_{i=1}^{Ns} \lambda_i y_i \mathbf{s}_i \cdot \mathbf{x}_{nov} + b \right). \quad (1.60)$$

Većina objekata u setu podataka nisu linerno razdvojivi. U tom slučaju neophodno je mapirati podatke u neki drugi prostor skalarnog proizvoda veće dimenzionalnosti koristeći funkciju $\Phi(\mathbf{x})$, $\Phi: S \rightarrow F$, tako da se postigne pouzdana linearna separacija. Tada algoritam za obuku zavisi samo od podataka koji ulaze u skalarni proizvod F , odnosno funkcija oblika $\Phi(\mathbf{x}_i) \cdot \Phi(\mathbf{x}_j)$:

$$f(\mathbf{x}_{new}) = \text{sgn}\left(\sum_{i=1}^m \lambda_i y_i \Phi(\mathbf{x}_i) \cdot \Phi(\mathbf{x}_{new}) + b\right). \quad (1.61)$$

Međutim, pošto je F visoko dimenzionalno za ovaj proizvod je potrebno jako puno vremena da bi se proračunao. Srećom, proračun u visoko dimenzionalnom prostoru se može prevazići korišćenjem jednog trika, tzv. kernel trik [Ivancuic 2007], koji omogućava proračun skalarnog proizvoda u inicijalnom prostoru podataka:

$$K(\mathbf{x}_i, \mathbf{x}_j) = \Phi(\mathbf{x}_i) \cdot \Phi(\mathbf{x}_j). \quad (1.62)$$

Na ovaj način se dobija SVM model koji se nalazi u višedimenzionalnom prostoru i vrši proračun u skoro istom vremenskom intervalu kao i kod nemapiranih podataka:

$$f(\mathbf{x}_{new}) = \text{sgn}\left(\sum_{i=1}^{N_s} \lambda_i y_i K(s_i, \mathbf{x}_{new}) + b\right). \quad (1.63)$$

Izbor kernel funkcije je ograničen nekim uslovima, poput Mercer-ovog uslova [Mercer 1909], koji ovde neće biti detaljno opisivani. Neki od najpopularnijih kernela korišćenih za proračun SVM modela su [Ivancuic 2007]:

- linearni kernel: $K(\mathbf{x}_i, \mathbf{x}_j) = \mathbf{x}_i \cdot \mathbf{x}_j$;
- polinomski kernel: $K(\mathbf{x}_i, \mathbf{x}_j) = (1 + \mathbf{x}_i \cdot \mathbf{x}_j)^d$;
- neuronski kernel: $K(\mathbf{x}_i, \mathbf{x}_j) = \tanh(a\mathbf{x}_i \cdot \mathbf{x}_j + b)$;
- Gausov RBF (eng. radial basis function) kernel:

$$K(\mathbf{x}_i, \mathbf{x}_j) = \exp(-\|\mathbf{x}_i - \mathbf{x}_j\|^2 / 2\sigma^2).$$

3. Fluorescentna spektroskopija

Fluorescentna spektroskopija predstavlja analitičku metodu velike senzitivnosti i specifičnosti. Ova metoda ima veliki potencijal kao nedestruktivna metoda za karakterizaciju bioloških i neorganskih uzoraka. Pokazano je da je ova tehnika od 100 do 1000 puta senzitivnija od drugih spektrofotometrijskih metoda [Strasburg & Ludescher 1995]. Jedan od razloga visoke senzitivnosti ove metode je taj što većina molekula ne fluorescira usled čega je uticaj fluorescencije okoline vrlo mali i lako se mogu detektovati određeni molekuli.

U ovom poglavlju biće detaljno objašnjeni principi i teorija fluorescencije kao i opis tehnike merenja fluorescentnih spektara.

3.1 Elektromagnetno zračenje i njegova interakcija sa molekulima

Mnoge metode koje služe za ispitivanje molekula koriste elektromagnetno zračenje. Elektromagnetno zračenje može biti generisano u velikom opsegu talasnih dužina. Različite talasne dužine imaju različite energije i samim tim interaguju drugačije sa molekulima. Visoko energetski fotoni imaju dovoljno veliku energiju da raskinu kovalentne veze, mada, da bi do toga došlo moraju biti ispunjeni određeni uslovi. Fotoni malih energija imaju previše malu energiju da bi uticali na kovalentne interakcije i oni mogu samo da predaju svoju energiju molekulima bez menjanja molekularne strukture. Interakcija elektromagnetnog zračenja sa molekulima je jako kompleksna. Razumevanje ovih interakcija zahteva poznavanje nekih elemenata kvantne mehanike.

Prilikom interakcije elektromagnetnog zračenja sa materijom jedan deo svetlosti će se reflektovati o površinu materijala, deo će se propustiti dok će se jedan deo prostirati kroz materijal. Prilikom prostiranja svetlosti kroz materijal mogu se javiti tri fenomena: refrakcija, rasejanje i apsorpcija. Prilikom refrakcije dolazi do savijanja svetlosnih zraka na površini materije jer se talasi prostiru manjom brzinom nego u slobodnom prostoru. Usled rasejanja dolazi do promene pravca svetlosti i smanjenja njenog intenziteta pri

interakciji sa materijom. ApSORpcija predstavlja pojavu kod koje se pod uticajem dejstva elektromagnetnog zračenja elektron pobuđuje iz osnovnog u više energetske stanje.

3.1.1 Interakcija elektromagnetnog zračenja i molekula

Elektromagnetno zračenje se sastoji od vremenski promenljivog električnog polja i vremenski promenljivog magnetnog polja. Za većinu elektronskih prelaza uticaj magnetnog polja se može zanemariti. Kad je spektroskopija u pitanju, talasna dužina zračenja je velika u odnosu na veličinu hromofora i iz tog razloga je uglavnom nepotrebno uzimati u obzir varijaciju u jačini polja koja utiče na molekul. Ako električno polje ima amplitudu \mathbf{E}_0 onda je uticaj polja na molekul

$$\mathbf{E}(t) = \mathbf{E}_0 e^{i\omega t} \quad (1.64)$$

Član ω predstavlja kružnu frekvenciju svetlosti ($\omega = 2\pi\nu = \frac{2\pi c}{\lambda}$). Kako je električno polje vremenski zavisno onda se mora koristiti vremenski zavisna Šredingerova jednačina. Pošto efekat potencijala može biti zanemaren, ostaje nam vremenski nezavisan Hamiltonijan i operator potencijala. Operator bitan za interakciju sa svetlošću je operator električnog dipola $\underline{\mu}$.

$$\underline{\mu} = \sum_i e_i \underline{\mathbf{r}}_i \quad (1.65)$$

gde je e_i pojedinačno naelektrisanje, a $\underline{\mathbf{r}}_i$ pozicija operatora. Frank-Kondonov princip kaže da se pomeranja jezgra mogu zanemariti jer svetlost interaguje sa molekulima vrlo kratak vremenski period. Ovo predstavlja modifikovanu Born-Openhajmerovu aproksimaciju. Stoga se u obzir uzimaju samo elektroni. Električno polje svetlosti utiče na talasnu funkciju stanja a tako što dovodi do promene stanja a u stanje približno stanju b . Klasično važi da je indukovani dipolni momenat $\mu_{\text{ind}} = \underline{\alpha} \cdot \mathbf{E}$ gde je $\underline{\alpha}$ operator polarizabilnosti molekula. Kod interakcije svetlosti sa molekulom, integral $\langle \Psi_b | \underline{\mu} | \Psi_a \rangle$ predstavlja dipol indukovano svetlošću.

3.1.2 Molarni koeficijent apsorpcije

Brzina promene apsorpcije svetlosti od strane sistema je definisana sledećom jednačinom

$$\frac{dP_b}{dt} = B_{ab}I(\nu) \quad (1.66)$$

gde je B_{ab} koeficijent verovatnoće prelaza po jedinici gustine energije zračenja, a $I(\nu)$ je gustina energije na frekvenciji ν . Gustina energije svetlosti je $I(\nu) = |E_0|^2/4\pi$ i definiše se kao funkcija broja prisutnih fotona na toj frekvenciji, jer svaki elektron doprinosi jačini električnog polja. Koeficijent verovatnoće prelaza B_{ab} (Ajnštajnov koeficijent za stimulisanu apsorpciju) je:

$$B_{ab} = \frac{2}{3} \frac{\pi}{\hbar^2} \left| \langle \Psi_b | \underline{\mu} | \Psi_a \rangle \right|^2 \quad (1.67)$$

Brzina kojom svetlost predaje energiju sistemu zavisi od verovatnoće apsorpcionih prelaza, brzine emisionih prelaza iste frekvencije, energije po prelazu ($h\nu = E_b - E_a$) i od N_a i N_b koji predstavljaju broj molekula po 1 cm^3 u stanjima a i b .

$$\frac{-dI(\nu)}{dt} = h\nu(N_a B_{ab} - N_b B_{ba})I(\nu) \quad (1.68)$$

Apsorpcija svetlosti kroz uzorak debljine dl sa koncentracijom c je data izrazom

$$\frac{-dI}{I} = c\varepsilon' dl \quad (1.69)$$

Integraljenjem od I_0 do I i od 0 do l

$$\int_{I_0}^I \frac{-dI}{I} = \int_0^l c\varepsilon' dl \quad (1.70)$$

dobija se

$$\ln\left(\frac{I_0}{I}\right) = c\varepsilon'l \quad (1.71)$$

Izražavanjem logaritmom sa osnovom deset dobija se sledeći oblik prehodnog izraza

$$\log\left(\frac{I_0}{I}\right) = c\varepsilon l \quad (1.72)$$

gde je ε molarni koeficijent apsorpcije, a l je dužina apsorpcionog puta (debljina apsorpcionog medijuma) u centimetrima. Menjanjem leve strane izraza apsorbancom A dobija se izraz $A(\lambda) = c\varepsilon(\lambda)l$, koji predstavlja standardni oblik Lamber-Berovog zakona.

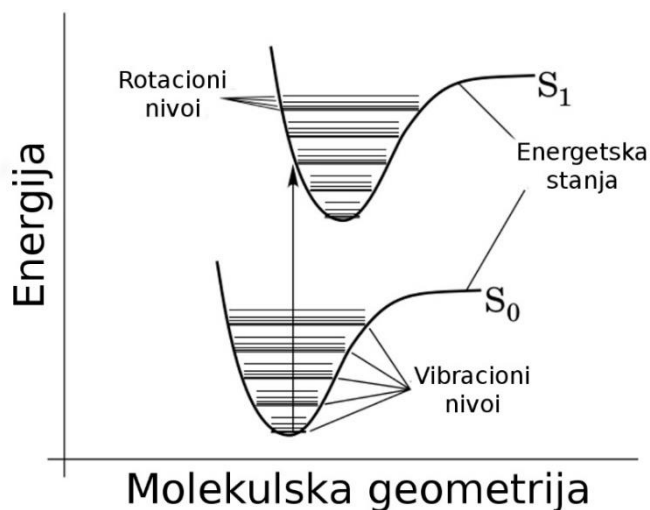
Ako pretpostavimo da molekuli u rastvoru ne interaguju, onda je koeficijent ε konstanta za dati molekul u rastvoru na datoj talasnoj dužini. Ukoliko molekuli interaguju međusobno na takav način da dolazi do promene energetske nivoa prelaza, onda će se koeficijent ekstinkcije menjati zavisno od koncentracije.

3.1.3 Apsorpcija i energetska stanja

Elektromagnetno zračenje interaguje sa molekulima samo na određenim energijama. Da bi električno polje indukovalo dipol, razlika u energiji između dva stanja mora biti jednaka energiji koju foton sadrži i orijentacija vektora električnog polja mora se poklapati sa orijentacijom molekula.

Stanje molekula se može opisati energetskim dijagramom prikazanom na slici 3.1. Energetski dijagram prikazuje dva elektronska stanja, pri čemu se svaki elektronski nivo sastoji od više vibracionih i rotacionih nivoa. Razlika energije između jednog rotacionog nivoa i sledećeg je ~ 4 kJ/mol, što spada u opseg termalne energije molekula. Međutim, energije vibracionih nivoa su razdvojene za oko 40 kJ/mol, a elektronska stanja su u razmaku od oko 150 do 450 kJ/mol (zavisno od molekula). Pošto su ove

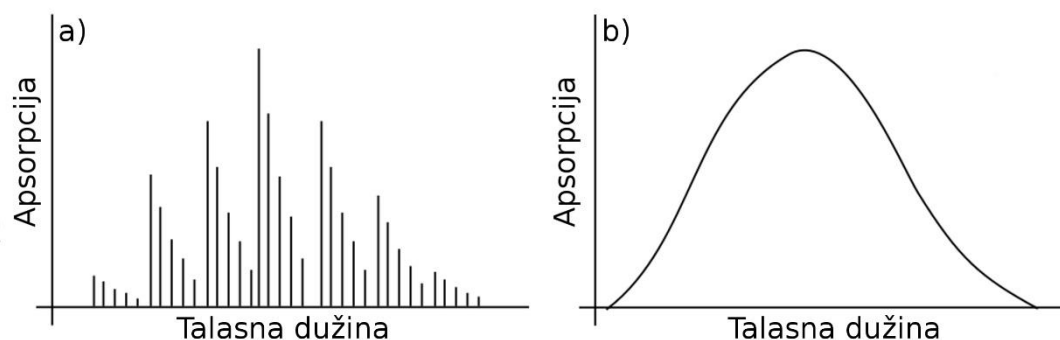
energije daleko iznad energije termalnog opsega, ukoliko nema eksterno dovedene energije biće popunjeni samo najniži vibracioni nivoi najnižih elektronskih nivoa.



Slika 3.1 – Energetski dijagram molekula.

Svako energetske stanje u kome molekul može da bude je opisano talasnom funkcijom. Najniže energetske stanje, osnovno stanje, ima talasnu funkciju Ψ_0 . Druga stanja u okviru osnovnog elektronskog stanja su opisana sa $\Psi_{0,v,r}$, gde v i r označavaju određene vibracione i rotacione nivoe.

Interakcija molekula sa elektromagnetnim zračenjem u ultraljubičastoj ili vidljivoj oblasti indukuje prelaze sa jednog na drugo elektronsko stanje. Verovatnoća prelaza se može izračunati uz pomoć integrala preklapanja za dva stanja. Za elektronske prelaze integral preklapanja je $\langle \Psi_{1,y,r} | \underline{\mu} | \Psi_0 \rangle$. Zbog operatora $\underline{\mu}$ integral preklapanja može imati vrednost različitu od nule. Do ovoga će doći samo ukoliko je energija fotona jednaka razlici energija dva stanja Ψ_0 i $\Psi_{1,y,r}$. Čak i tada verovatnoća za taj prelaz je manja od jedan ukoliko vektor električnog polja indukuje dipol u molekulu sa istom orijentacijom kao i električno polje fotona koji se približava. U principu, apsorpcioni spektar molekula, intenzitet apsorpcije u funkciji frekvencije ili talasne dužine svetlosti, treba da bude skup oštih pikova (linija), gde svaka linija odgovara prelasku sa jednog energetskog stanja na drugo (slika 3.2a). Visina svake apsorpcione linije je direktno vezana za verovatnoću prelaza.



Slika 3.2 – Izgled apsorpcionog spektra a) teorijski predviđen i b) izmeren za rastvor.

Za mnoge molekule Doplerov efekat i drugi uticaji okoline u rastvoru dovode do toga da se kao rezultat dobije relativno široki pik (traka) apsorpcije sa širinom većom od 50 nm (slika 3.2b). Celokupna apsorpciona traka odgovara jednom elektronskom prelazu, koji se širi zbog vibracionih i rotacionih prelaza.

3.1.4 Tipovi elektronskih prelaza kod molekula

Elektronski prelaz predstavlja prelaz elektrona, pri apsorpciji fotona, iz jedne orbitale molekula u osnovnom stanju na upražnjenu orbitalu. Na taj način molekul prelazi u pobuđeno stanje. Radi lakšeg daljeg objašnjenja, ukratko ćemo se podsetiti koji tipovi orbitala postoje. σ orbitala se može formirati od dve s atomske orbitale ili jedne s i jedne p atomske orbitale ili dve p atomske orbitale. Veza koja se na ovaj način ostvaruje naziva se σ veza. π orbitala se formira preklapanjem dve p atomske orbitale i dobijena veza se naziva π veza. Apsorpcijom fotona odgovarajuće energije omogućava se prelazak elektrona sa π orbitale na antivezujuću orbitalu π^* . Takav prelaz se označava sa $\pi \rightarrow \pi^*$. Molekul može da poseduje i nevezujuće elektrone na heteroatomima i odgovarajuće molekulske orbitale se nazivaju n orbitalama. Prelazak nevezujućeg elektrona na antivezujuću orbitalu se označava sa $n \rightarrow \pi^*$. Redosled energija elektronskih prelaza se može prikazati na sledeći način

$$n \rightarrow \pi^* < \pi \rightarrow \pi^* < n \rightarrow \sigma^* < \sigma \rightarrow \pi^* < \sigma \rightarrow \sigma^*$$

Kada jedan od dva elektrona suprotnih spinova pređe na molekulska orbitalu veće energije njegov spin u principu ostaje nepromenjen tako da njegov ukupan spinski kvantni broj S ostaje jednak nuli. Pošto multipletnost osnovnog i pobuđenog stanja ($M=2S+1$) ima vrednost jedan, ova stanja se nazivaju singletna stanja (osnovno stanje se označava sa S_0 , a pobuđena sa S_1, S_2, \dots). Odgovarajući prelazi se nazivaju singlet-singlet prelazi. Pri prelazu elektrona može doći do konverzije singletnog stanja u stanje gde dolazi do promene spina elektrona. Kako tada postoje dva elektrona paralelnih spinova, ukupni spinski kvantni broj je 1, a multipletnost 3. Takvo stanje se naziva tripletno stanje jer odgovara trima stanjima istih energija. Po Hundovom pravilu tripletno stanje ima manju energiju od singletnog stanja iste konfiguracije.

3.2 Luminescencija

Luminescencija predstavlja fenomen emisije ultraljubičastih, vidljivih ili infracrvenih fotona iz energetski pobuđene materije. Pojam luminescencije (lat. *lumen* – svetlost) je prvi uveo fizičar Eilhardt Wiedemann 1888. godine u cilju opisivanja svih fenomena svetlosti koji nisu u potpunosti uslovljeni povećanjem temperature. Iz tog razloga se luminescencija drugačije naziva i „hladna svetlost“. Zavisno od načina pobuđivanja luminescencija se može podeliti na više tipova koji su predstavljeni u tabeli 3.1.

Tabela 3.1 Različiti tipovi luminescencije

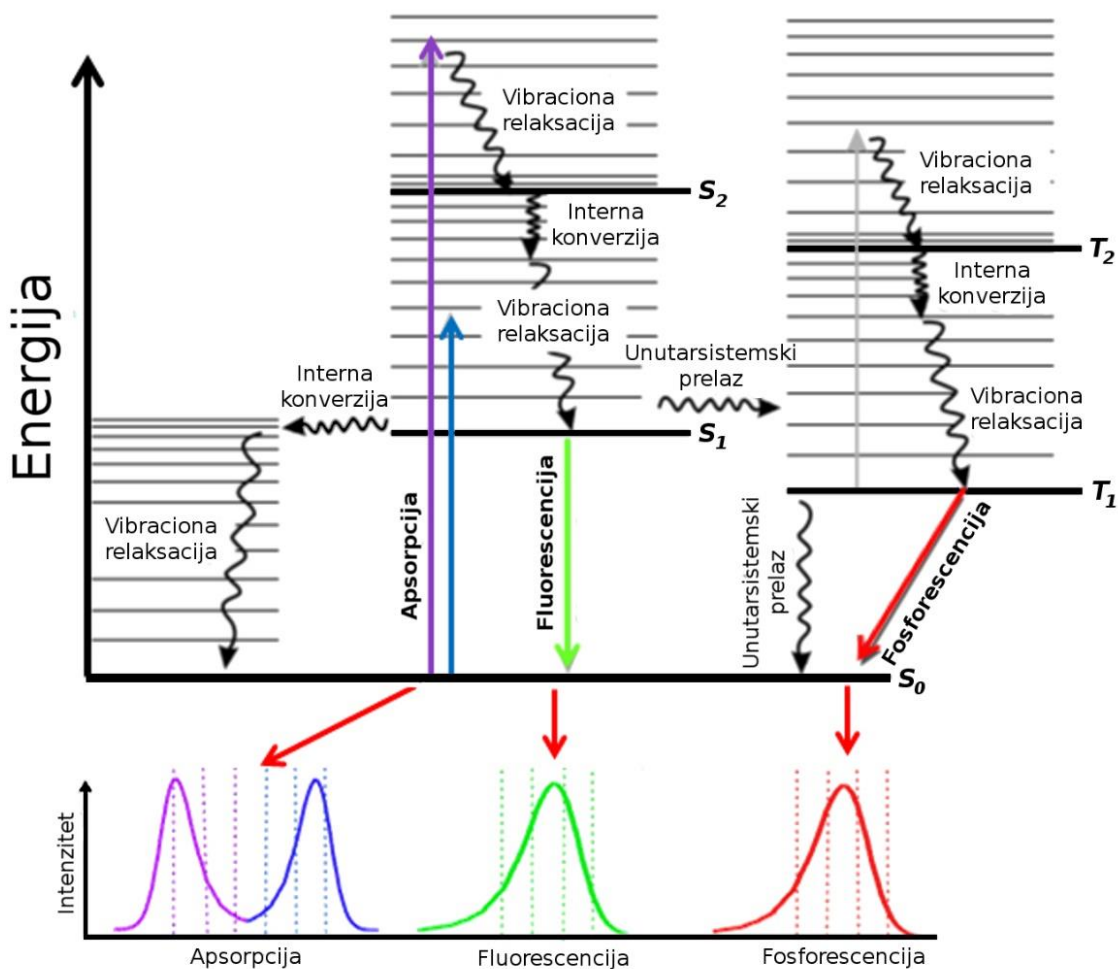
<i>Fenomen</i>	<i>Način pobuđivanja</i>
Fotoluminescencija	Apsorpcija svetlosti (fotona)
Radioluminescencija	Jonizujuće zračenje (x-zraci, α , β , γ)
Katodoluminescencija	Katodni zraci
Elektroluminescencija	Električno polje
Termoluminescencija	Toplota (zagrevanjem)
Hemiluminescencija	Hemijski procesi
Bioluminescencija	Biohemijski procesi
Triboluminescencija	Trenje ili elektrostatičke sile
Sonoluminescencija	Ultrazvuk

Zavisno od vremena provedenog u pobuđenom stanju, fotoluminescenciju, fenomen kod kojeg dolazi do prelaza elektrona u pobuđeno stanje usled apsorpcije elektromagnetnog zračenja, možemo podeliti na fluorescenciju ($\sim 10^{-9}$ s) i fosforescenciju (10^{-4} - 10^2 s). Radi lakšeg objašnjenja i razumevanja ovih procesa koristi se Jablonski dijagram (slika 3.3). On opisuje sve procese do kojih dolazi pri apsorpciji elektromagnetnog zračenja: apsorpcija fotona, interna konverzija, fluorescencija, unutarsistemski prelaz, fosforescencija, odložena fluorescencija i triplet-triplet tranzicija.

Osnovna singletna stanja se označavaju sa S_0 (osnovno stanje), S_1 , S_2 ,... dok se tripletna označavaju sa T_1 , T_2 ,... Vibracioni nivoi se vezuju za svako elektronsko stanje. Treba napomenuti da je apsorpcija jako brza ($\sim 10^{-5}$ s) u odnosu na druge procese, što prema Frank-Kondonovom principu podrazumeva da ne dolazi do pomeranja jezgra. Vertikalne strelice koje odgovaraju apsorpciji kreću od nultog vibracionog energetskog nivoa S_0 jer se većina molekula nalazi na tom nivou na sobnoj temperaturi. Apsorpcijom fotona molekul se dovodi na neki vibracioni nivo pobuđenog stanja S_1 , S_2 ,... Kada se molekul nađe u tom stanju postoji više procesa koji se mogu desiti pri vraćanju u osnovno stanje, što će dalje biti opisano.

3.2.1 Interna konverzija

Interna konverzija predstavlja neradijativni prelaz između dva elektronska stanja iste multipletnosti. Ovaj proces je ispraćen vibracionom relaksacijom na najniži vibracioni nivo elektronskog stanja. Pri ovom procesu višak vibracione energije se predaje okolini. Ukoliko se molekul pobudi na energetski nivo viši od najnižeg vibracionog nivoa prvog elektronskog stanja (S_1) onda se vibracionom relaksacijom pobuđeni molekul dovodi na nulti vibracioni nivo S_1 singletnog stanja. Ovaj proces traje između 10^{-13} – 10^{-11} s. Interna konverzija između S_1 i S_0 je moguća, ali vrlo malo verovatna i manje efikasna zbog mnogo veće energetske razlike nego kod konverzije između S_1 i S_2 stanja (što je manja energetska barijera veća je efikasnost interne konverzije).

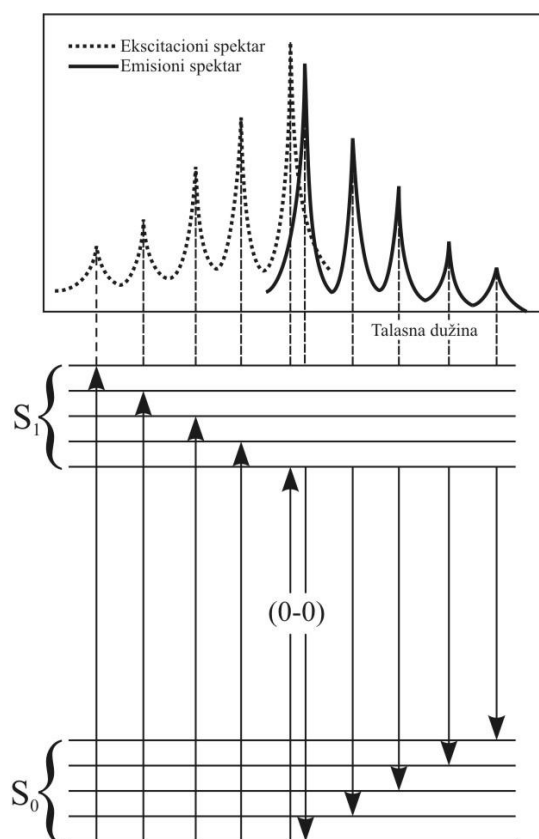


Slika 3.3 – Jablonski dijagram.

3.2.2 Fluorescencija

Fluorescencija je fenomen emisije fotona praćen $S_1 \rightarrow S_0$ relaksacijom. Kako je radijativni prelaz $S_1 \rightarrow S_0$ dozvoljen po spinu, odnosno stanja su iste multipletnosti, ovaj proces se odvija jako brzo tj. nakon $\sim 10^{-9}$ s. Međutim, pre nego što dođe do emisije fotona ili nekog drugog deekscitacionog procesa, pobuđeni molekuli ostaju u S_1 stanju neko vreme. Zavisno od tipa molekula ono može biti reda veličine od nekoliko desetina pikosekundi do nekoliko stotina nanosekundi. To podrazumeva da će se posle pobuđivanja molekula neko određeno vreme intenzitet fluorescencije eksponencijalno smanjivati što određuje vreme života nekog molekula u S_1 pobuđenom stanju.

0-0 prelaz je uglavnom isti za apsorpciju i fluorescenciju, međutim zbog gubitka energije u pobuđenom stanju usled vibracione relaksacije, talasna dužina fluorescencije će biti veća odnosno imaće manju energiju od apsorpcije. Prema Stoksovom pravilu talasna dužina fluorescencije uvek mora biti veća od talasne dužine apsorpcije. U većini slučajeva apsorpcioni spektar se delom preklapa sa fluorescentnim spektrom, odnosno deo svetlosti se emituje na manjoj talasnoj dužini od apsorbovane svetlosti što je grafički prikazano na slici 3.4. Iako ova pojava deluje u suprotnosti zakonu očuvanja energije, ona se javlja zbog činjenice da se i na sobnoj temperaturi nalazi mali broj molekula kako na višim vibracionim nivoima tako i u pobuđenom stanju. Na niskim temperaturama uglavnom ne dolazi do pojave ovog defekta. Fluorescentni spektar često liči na apsorpcioni zbog sličnosti vibracionih nivoa u osnovnom i pobuđenom stanju. Razlika između položaja maksimuma apsorpcionog i fluorescentnog spektra naziva se Stoksov pomeraj.



Slika 3.4 – Grafički prikaz emisionog i ekscitacionog spektra.

3.2.3 Unutarsistemski prelazi

Unutarsistemski prelaz predstavlja neradijativni prelaz između dva vibraciona nivoa istih energija koji pripadaju elektronskim stanjima različite multipletnosti. Na primer, pobuđeni molekul na 0 vibracionom nivou S_1 stanja može da pređe na vibracioni nivo iste energije T_n tripletnog stanja. Vibracionom relaksacijom molekul se dovodi na najniži vibracioni nivo T_1 stanja. Prelazi između stanja različite multipletnosti su zabranjeni prelazi, ali ukoliko je spin-orbitalno sprezanje dovoljno veliko ovi prelazi će biti omogućeni. Verovatnoća unutarsistemskih prelaza zavisi od odgovarajućih singletnih i tripletnih stanja.

3.2.4 Fosforescencija i neradijativna relaksacija

Pri prelasku molekula sa T_1 tripletnog stanja na osnovno stanje mogu se javiti dva procesa: radijativni proces - fosforescencija i neradijativna relaksacija. Kod rastvora na sobnoj temperaturi uvek je veća verovatnoća da dođe do neradijativne relaksacije. Pošto je ovaj prelaz zabranjen, konstanta brzine radijacije je samim tim jako mala. Tokom ovako sporih procesa brojni sudari molekula u rastvoru favorizuju unutarsistemske prelaze i vibracionu relaksaciju u S_1 stanje. Nasuprot tome na niskim temperaturama i kod rigidnih medijuma može doći do pojave fosforescencije. Vreme života tripletnog stanja pod ovim uslovima može biti dovoljno dugo da se fosforescencija može posmatrati u trajanju od nekoliko sekundi, a u nekim slučajevima čak i nekoliko minuta ili više. Fosforescentni spektri se javljaju na većim talasnim dužinama od fluorescentnih jer je energija najnižeg vibracionog nivoa tripletnog stanja T_1 niža od najnižeg nivoa stanja S_1 .

3.2.5 Eksterna konverzija i gašenje usled sudara

Eksterna konverzija predstavlja neradijativan proces koji omogućava molekulu da se vrati u osnovno stanje kao rezultat sudara sa molekulima u rastvoru. Gašenje usled sudara je povezano sa eksternom konverzijom, jer kod gašenja dolazi do oduzimanja viška energije pobuđenog stanja od strane molekula u rastvoru. Iako je gašenje proces

drugog reda, velike koncentracije agensa koji gasi omogućavaju analize pseudo prvog reda. Brzina procesa zavisi od konstante brzine i koncentracije agensa koji gasi [Q].

3.2.6 Kvantni prinos i intenzitet fotoluminescencije

Kvantni prinos fluorescencije Φ_F predstavlja deo pobuđenih molekula koji se vrata u osnovno stanje S_0 uz emisiju fluorescentnih fotona.

$$\Phi_F = \frac{N_e}{N_a} = \frac{k_r}{k_r + k_{nr}} = \frac{k_F}{k_F + k_{ic} + k_{ec} + k_{is} + k_{pd} + k_q[Q]} = k_r \tau_S \quad (1.73)$$

k_r – konstanta brzine radijativnih prelaza

k_{nr} – konstanta brzine neradijativnih prelaza

k_F – konstanta brzine fluorescencije

k_{ic} – konstanta brzine interne konverzije

k_{ec} – konstanta brzine eksterne konverzije

k_{is} – konstanta brzine unutarsistemskog prelaza

k_{pd} – konstanta brzine fotodekompozicije

k_q – konstanta brzine gašenja

τ_S – vreme života pobuđenog stanja S_1

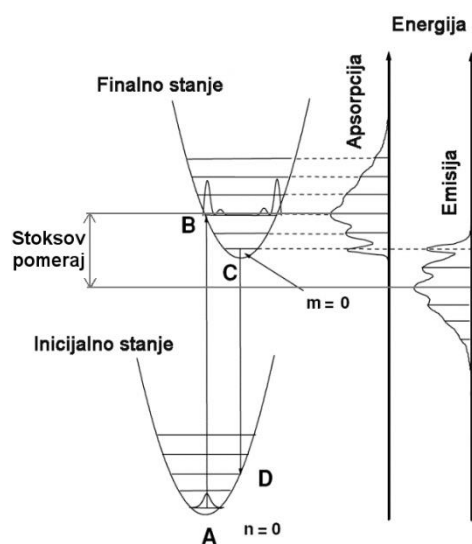
Intenzitet fotoluminescencije (W_l) predstavlja količinu emitovane energije iz materije po jedinici vremena i definisana je sledećim izrazom

$$W_l = \int_0^{\infty} W_l(\nu) d\nu \quad (1.74)$$

3.2.7 Ekscitacioni i emisijski spektri

Postoje dve vrste fotoluminescentnih spektara koji se mogu posmatrati: ekscitacioni i emisijski spektri. Ekscitacioni spektar prikazuje promenu intenziteta fotoluminescentne

emisije na određenoj talasnoj dužini ili frekvenciji pri različitim talasnim dužinama ekscitacije. Emisioni spektar prikazuje promenu intenziteta fotoluminescentne emisije za određeni opseg talasnih dužina za fiksnu talasnu dužini ekscitacije. Usled Stoksovog pomeraja emisioni spektar će biti pomeren ka višim talasnim dužinama u odnosu na ekscitacioni spektar što se može videti na konfiguracionom koordinatnom dijagramu na slici 3.5. Linije na slici predstavljaju elektronske prelaze u konfiguraciono-koordinatnom modelu u skladu sa Kondonovim principom, o kojem je ranije bilo reči. Takođe je prikazan izgled apsorpcionog spektra koji potiče od $n = 0 \rightarrow m = 0, 1, 2, \dots$ prelaza i emisionog spektra koji potiče od $m = 0 \rightarrow n = 0, 1, 2, \dots$ prelaza. Maksimumi apsorpcije i emisije se javljaju na mestu prelaza označenog strelicom AB i CD, respektivno. U skladu sa Stoksovim pomerajem, vidi se da se maksimum emisije nalazi na manjoj energiji od maksimuma apsorpcije, čija razlika odgovara Stoksovoj energiji.



Slika 3.5 – Konfiguraciono-koordinatni dijagram prelaza između dva stanja.

3.3 Luminescentni materijali

3.3.1 Organski luminescentni uzorci - fluorofore

Fluorescentni molekuli i fluorescentni funkcionalni delovi unutar većeg molekula se nazivaju fluoroforima. Fluorofore se mogu podeliti na endogene i egzogene. Endogene fluorofore su one koje se prirodno nalaze u uzorku. Fluorescencija koja potiče od endogenih fluorofora naziva se autofluorescencija. Egzogene fluorofore se dodaju

uzorku koji ne pokazuje željene spektralne karakteristike. Kod proteina dominantna fluorofora je indolna grupa triptofana. Indol apsorbuje na talasnoj dužini oko 280 nm i emituje blizu 340 nm. U tabeli 3.2 je dat pregled endogenih fluorofora koje se najčešće nalaze u biološkim uzorcima. Može se videti da se ekscitacioni maksimumi nalaze u opsegu od 250-450 nm (UV/VIS spektralni region), dok su emisijski maksimumi između 280 i 700 nm (UV/VIS/NIR spektralni region).

Tabela 3.2 Ekscitacioni i emisijski maksimumi endogenih fluorofora u biološkim uzorcima

<i>Endogene fluorofore</i>	<i>Ekscitacioni maksimum [nm]</i>	<i>Emisijski maksimum [nm]</i>
Triptofan	280	350
Tirozin	275	300
Fenilalanin	258	284
Vitamin A (retinol)	346	480
Vitamin B2 (riboflavin)	270 (382, 448)	518
Vitamin B6 (piridoksin)	328	393
Vitamin E (α -Tocopherol)	298	326
NADH	344	465
NADPH	336	464
FAD	450	535
Hlorofil A	428	663
Hematoporfirin	396	614
Porfirini	400-450	630,690
Kolagen	325	400,405
Elastin	290,325	340,400

3.3.1.1 Faktori koji utiču na organsku fluorescenciju

Fluorescentne karakteristike molekula mogu se menjati zavisno od uticaja različitih faktora. Koncentracija, zamućenost, lokalno molekularno okruženje, pH i temperatura

utiču na fluorescenciju zbog pojave fenomena poput gašenja fluorescencije, efekta unutrašnjeg filtera, rasejanja itd.

Gašenje fluorescencije odnosno smanjenje intenziteta fluorescencije se javlja kao posledica međumolekulskih interakcija unutar samog fluorescirajućeg molekula ili sa drugim molekulima u matrici uzorka. Gašenje može biti statičko ili dinamičko. Statičko gašenje se javlja kada fluorofora formira nefluorescirajući kompleks sa molekulom koji gasi fluorescenciju što onemogućuje dovođenje fluorofore u pobuđeno stanje. S druge strane dinamičko gašenje se javlja pri difuziji ili molekularnim sudarima. Do njega dolazi usled deaktivacije pobuđenog stanja pri kontaktu fluorofore sa drugim molekulom u matrici. Poznato je da kiseonik, teški metali, organske i neorganske nitro komponente kao i međumolekularne interakcije indukuju dinamičko gašenje. Molekul se vraća u osnovno stanje bez fluorescentne emisije i bez hemijske modifikacije. Pri visokim temperaturama takođe može doći do pojave dinamičkog gašenja jer pri povećanju brzina molekula veće su šanse da dođe do molekularnih sudara.

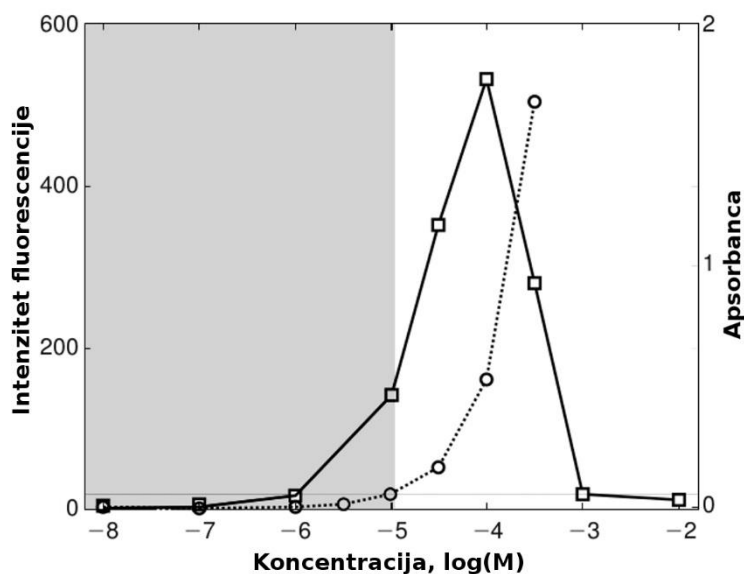
Drugi tip mehanizma koji smanjuje mereni intenzitet fluorescencije je tzv. *efekat unutrašnjeg filtera*. On se javlja usled apsorpcije ekscitovane ili emitovane svetlosti od strane fluorofora ili hromofora u uzorku. Efekat unutrašnjeg filtera se dešava kad hromofora reapsorbuje emitovanu fluorescenciju ili kad nefluorescirajuća hromofora apsorbuje deo ekscitovane svetlosti. Kao rezultat se javlja smanjen intenzitet fluorescencije ili izobličenje spektara. Problem koncentracionog gašenja i efekta unutrašnjeg filtera se može rešiti razblaživanjem uzorka ili smanjivanjem dužine optičkog puta. Međutim razblaživanje uzorka smanjuje koncentraciju bitnih fluorofora i menja molekularnu interakciju u uzorku.

Intenzitet fluorescencije i *koncentracija* imaju linearnu zavisnost, ali koncentracija mora biti u određenom opsegu da ne bi došlo do pojave artefakta poput gašenja i efekata unutrašnjih filtera. Za transparentne rastvore sa apsorbancom manjom od 0.05, intenzitet fluorescencije je linearno povezan sa koncentracijom fluorofora [Lakowicz 1999]. U tom slučaju intenzitet fluorescencije I_f , zavisi od intenziteta pobuđujuće

svetlosti I_0 , molarnog koeficijenta ekstinkcije ε , kvantni prinos fluorescencije Φ_F , dužine optičkog puta l i molarne koncentracije fluorofore c

$$I_f = 2.3\Phi_F I_0 \varepsilon c l \quad (1.75)$$

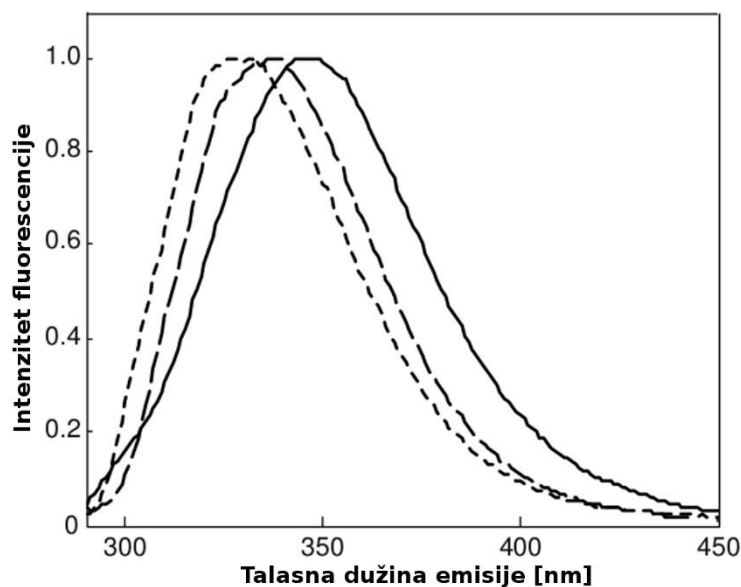
Kvantni prinos fluorescencije zajedno sa molarnim koeficijentnom ekstinkcije predstavlja meru efektivnosti fluorofore. Na slici 3.6 je prikazana zavisnost intenziteta fluorescencije i koncentracije čistog rastvora triptofana. Može se uočiti da za koncentracije triptofana do 10^{-4} M intenzitet fluorescencije raste sa koncentracijom, a za veće koncentracije počinje da opada usled pojave gašenja i unutrašnjih filtera. Takođe se primećuje da tek kada koncentracija dostigne vrednost 10^{-5} M apsorbanca je 0.05 (ova granica je prikazana horizontalnom linijom na grafiku). Samo za koncentracije ispod ove granice važi izraz dat u jednačini (1.75).



Slika 3.6 – Grafički prikaz zavisnosti intenziteta fluorescencije i koncentracije čistog rastvora triptofana.

Osim koncentracije na intenzitet fluorescencije utiču i *svojstva okoline* u kojoj se molekul nalazi poput pH vrednosti, polarnosti, temperature itd. Na primer, zavisno od izloženosti triptofana okolini u proteinu on će pokazati drugačije fluorescentne karakteristike. Kod nekih proteina on se nalazi duboko unutar proteina i na njega u tom

slučaju ne utiče mnogo okolina, dok kod nekih proteina se nalazi na površini i tada reaguje na svaku promenu okoline u rastvoru. Na slici 3.7 je prikazan primer emisionog spektra merenog na proteinima mleka: α -lactalbumin, β -lactoglobulin i kasein. Kod kaseina triptofan je najviše izložen rastvoru, koji je u ovom slučaju voda. Na grafiku se uočava da kasein ima emisionu maksimum na najvećoj talasnoj dužini. Do ove pojave dolazi jer je voda hidrofilni rastvor i to pomera emisioni maksimum triptofana, te u zavisnosti od stepena izloženosti taj pomeraj može biti veći ili manji. Povećanje hidrofobnosti okoline fluorofore dovodi do relaksacije pobuđenog stanja na niže vibracione nivoe pre emisije fluorescencije usled čega se dobija emisija manje energije od onih dobijenih iz polarnijih rastvora. Ove razlike u spektrima omogućavaju izučavanje promena konformacija proteina kao i denaturaciju i međusobnu interakciju proteina sa drugim supstancama u uzorku.



Slika 3.7 – Prikaz emisionih spektara proteina mleka: α -lactalbumin (linija sa kratkim povlakama), β -lactoglobulin (linija sa dugim povlakama) i kasein (puna linija).

Rasejanje svetlosti predstavlja fenomen koji može da remeti mereni fluorescentni signal. Pojava svetlosti je posledica prisustva malih čestica u uzorku koju detektor vidi kao lažnu emitovanu svetlost. Rasejanje remeti merenje kada je Stoksov pomeraj mali i rasejanje se poklapa sa emisijom supstance. Postoje dve glavne vrste rasejanja: elastično Rejljevo i neelastično Ramanovo. Rejljevo rasejanje prvog reda se javlja na talasnoj

dužini jednakoj talasnoj dužini ekscitacije, dok se Rejljevo rasejanje drugog reda javlja na talasnoj dužini duplo većoj od talasne dužine ekscitacije. Kod Rejljevog rasejanja ne postoji elektronska interakcija između svetlosti i materije i zato se ono uvek javlja na istim talasnim dužinama i ima istu energiju. Intenzitet Rejljevog rasejanja je obrnuto srazmeran četvrtom stepenu talasne dužine. Stoga se ovaj efekat može ublažiti korišćenjem većih talasnih dužina pobude. Neelastično Ramanovo rasejanje je mnogo slabije od Rejljevog i uglavnom se ne vidi ukoliko je visok intenzitet fluorescencije. Ramanovo rasejanje se javlja na većim talasnim dužinama od talasne dužine ekscitacije sa konstantnom razlikom u frekvenciji u odnosu na pik Rejljevog rasejanja. Do ove pojave dolazi usled vibracionih interakcija sa rastvorom. Pik Ramanovog rasejanja za vodu se javlja na 3600 cm^{-1} manjem talasnom broju od svetlosti ekscitacije [Lakowicz 1999]. Bitno je obratiti pažnju na Ramanovo rasejanje jer može da se poklopi sa fluorescentnim signalom, pogotovo za niske intenzitete fluorescencije. Ramanovo rasejanje se može ukloniti oduzimanjem fluorescentnog signala čistog rastvora od signala merenog uzorka [Jiji & Booksh 2000].

3.3.2 Neorganski luminescentni materijali

Zavisno od mehanizma elektronskih prelaza usled kojih dolazi do emisije svetlosti, kod čvrstih neorganskih materijala mogu se razlikovati dve vrste fotoluminescencije: unutrašnja i spoljašnja fotoluminescencija. Unutrašnja fotoluminescencija se dalje može podeliti na međuzonsku, ekscitonsku i rekombinacionu fotoluminescenciju.

Najčešće primenjivan način za dobijanje luminescentnih neorganskih materijala jeste unos aktivatorskih jona u sam materijal. Aktivatori mogu biti atomi ili joni koji se nalaze u višku kada se materijal samo-aktivira, zatim mogu biti neke intersticijske ili supstitucijske nečistoće ili neke druge vrste defekta u kristalnoj rešetki. Ovako dobijeni materijali koji emituju svetlost nazivaju se fosfori.

Zavisno od aktivatora koji se koriste i elektronskog prelaza, fosfori se dele na sledeći način:

- 1) $1s \leftrightarrow 2p$ prelaz.

- 2) $ns^2 \leftrightarrow nsnp$ prelaz – u ovu grupu spadaju joni Tl⁺ tipa (Ga⁺, In⁺, Tl⁺, Ge²⁺, Sn²⁺, Pb²⁺, Bi³⁺, Cu⁻, Ag⁻, Au⁻ itd.).
- 3) $3d^{10} \leftrightarrow 3d^9 4s$ prelaz – poput Cu⁺, Ag⁺ i Au⁺.
- 4) $3d^n \leftrightarrow 3d^n$, $4d^n \leftrightarrow 4d^n$ prelaz – ovu grupu čine prvi i drugi red prelaznih metala.
- 5) $4f^n \leftrightarrow 4f^n$, $5f^n \leftrightarrow 5f^n$ prelaz – spadaju joni retkih zemalja i joni aktinoida.
- 6) $4f^n \leftrightarrow 4f^{n-1}5d$ prelaz – kao što su Ce³⁺, Pr³⁺, Sm²⁺, Eu²⁺, Tm²⁺ i Yb²⁺.
- 7) Prelaz između prazne orbitale katjona i p elektrona anjona ili prenos naelektrisanja. Poput međumolekulskih prelaza u kompleksima kao što su VO₄³⁻, WO₄²⁻ i MoO₄²⁻.
- 8) $\pi \leftrightarrow \pi$ i $n \leftrightarrow \pi^*$ prelaz – organski molekuli koji imaju π elektrone

3.3.2.1 Retke zemlje – elektronska konfiguracija jona i njihova interakcija

Pod retkim zemljama se podrazumevaju 15 elemenata III A grupe periodnog sistema elemenata koji se nazivaju lantanoidi, od lantana do lutecijuma. Elektronska konfiguracija lantanoida je data opštom formulom $4f^n 5s^3 5p^6 6s^2$ pri čemu n ima vrednosti od 0 do 14. Kako su 4f elektroni valentni i imaju najmanju energiju veze u atomu oni su odgovorni za optičke karakteristike ovih elemenata. Uklanjanjem 6s elektrona dobijaju se joni retkih zemalja pri čemu 4f orbitale unutar popunjenih 5s i 5p ljuski ostaju optički aktivne. Joni koji imaju luminescentne osobine i karakteristične energetske nivoe su joni od Ce³⁺ do Lu³⁺ sa delimično popunjenim 4f orbitalama, za razliku od Sc³⁺, Y³⁺, La³⁺, Lu³⁺ jona koji ne pokazuju luminescentne osobine jer nemaju slobodne 4f elektrone. Pri inkorporiranju jona retkih zemalja u rešetku matrice, na njih počinje da deluje kristalno polje koje ga okružuje. Usled dejstva ovog polja na inkorporirani jon doći će do pojave potencijala koji predstavlja perturbaciju na energetsko stanje slobodnog jona. Uz primenu perturbacionog računa dobija se ukupno razdvajanje energetskih nivoa i parametri kristalnog polja. Hamiltonijan kristalnog polja je dat sledećim izrazom:

$$\hat{H}_{CF} = \sum_{i=1}^N eV(r_i, \theta_i, \varphi_i) \quad (1.76)$$

gde je $eV(r_i, \theta_i, \varphi_i)$ potencijalna energija i -tog valentnog elektrona u kristalnom polju. U izrazu (1.76) je u obzir uzeta interakcija valentnih elektrona dopantnih jona sa elektrostatičkim poljem koje ga okružuje.

Kod atoma koji imaju više elektronskih stanja postoje tri tipa interakcije:

- 1) Spin-spin interakcija,
- 2) orbitno-orbitna interakcija i
- 3) spin-orbitna interakcija.

Kada se jon dopanta inkorporira u kristalnu rešetku energetske nivoi slobodnog jona retkih zemalja se razdvajaju na više optičkih energetske nivoa (termova) koji su definisani kvantnim brojevima L , S i J . Za 4f orbitalu orbitalni kvantni broj L ima vrednost 3, što znači da ima 7 orbitala tog tipa (-3, -2, -1, 0, +1, +2, +3). Raspored elektrona u tim orbitalama će uvek biti takav da spinski moment impulsa ($S=\sum s$) bude maksimalan. Kvantni broj J se dobija kombinacijom spinskog orbitalnog broja S i orbitalnog momenta L ($J=\sum l$) što je definisano sledećim jednačinama:

- kada je broj 4f elektrona manji od 7 onda važi $J = L - S$
- kada je broj 4f elektrona veći od 7 onda važi $J = L + S$

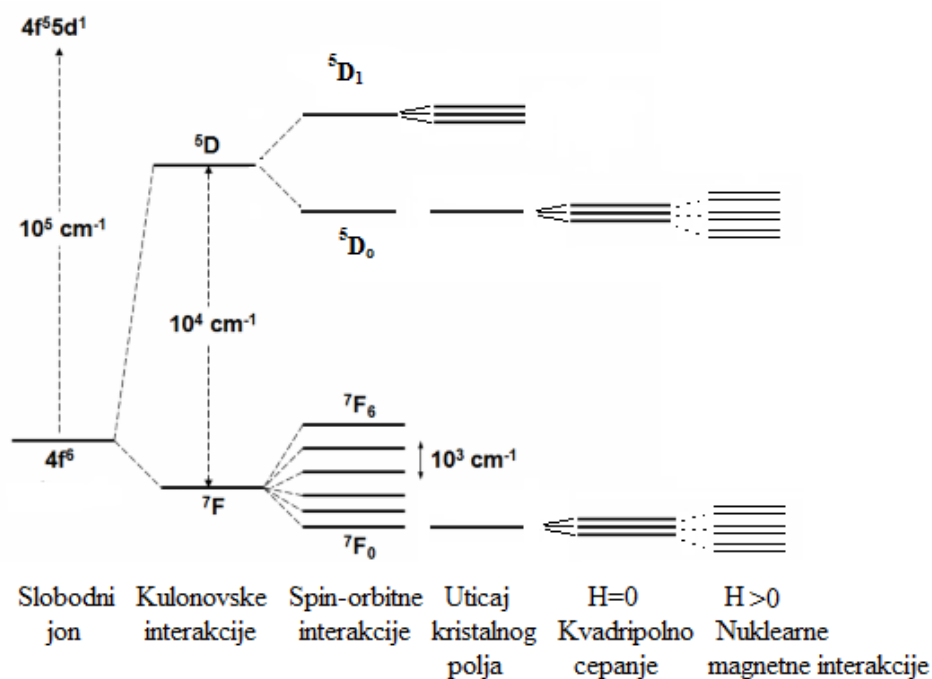
pri čemu L predstavlja S, P, D, F, G, H, I, K, L, M, ..., koji odgovaraju vrednostima $L = 1, 2, 3, 4, 5, 6, 7, 8, 9$, respektivno. Optički termovi se u skladu sa Rasel-Sanders-ovim modelom opisuju kao stanje nastalo sprezanjem orbitalnih i spinskih orbitala (LS sprega) pri čemu se za njihovo obeležavanje koriste simboli uz pomoć kojih se definišu vrednosti kvantnih brojeva $^M L_J$. M predstavlja multipletnost terma i definisan je na sledeći način:

- za $L \geq S$ važi $M = 2S + 1$
- za $L < S$ važi $M = 2L + 1$

Za nivoe bliske osnovnim stanjima važi da je sprezanje usled spin-orbitalne interakcije malo dok je za pobuđena stanja sa sličnim J brojevima veliko. Efekat sprezanja ima veliki uticaj na verovatnoću optičkih prelaza, a jako mali uticaj na energije nivoa.

Na slici 3.8 je prikazano cepanje energetskih nivoa na primeru trovalentnog jona europijuma. Kvantni brojevi za Eu^{3+} jon, koji ima konfiguraciju $4f^6$, su

- $S = \sum s = 6 \times \frac{1}{2} = 3$
- $L = \sum l = |-3 - 2 - 1 + 0 + 1 + 2| = 3$
- $L \geq S$ ($3 \geq 3$) $M = 2 \times 3 + 1 = 7$
- $L \geq S$ ($3 \geq 3$) $J = L + S, L + (S-1), L + (S-2), \dots, L+1, L, L-1, \dots, L - (S-2), L - (S-1), L - S$; $J = 0, 1, 2, 3, 4, 5, 6$



Slika 3.8 – Cepanje energetskih nivoa trovalentnog europijuma.

Uz pomoć datih kvantnih brojeva dobija se da je osnovni term za Eu^{3+} $7F_0$. Ono što se uočava na slici 3.8 jeste da pod uticajem kristalnog polja ne dolazi do cepanja osnovnog stanja ($M=2 \times 0 + 1 = 1$) pri čemu se nivo $7F_1$ cepa na tri podnivoa ($M=3$). Na prikazanom primeru je dat jon Eu^{3+} inkorporiran u matricu LaF_3 , gde je usled kulonovskih, spin-orbitalnih, interakcija kristalnog polja i nuklearnih magnetnih interakcija došlo do

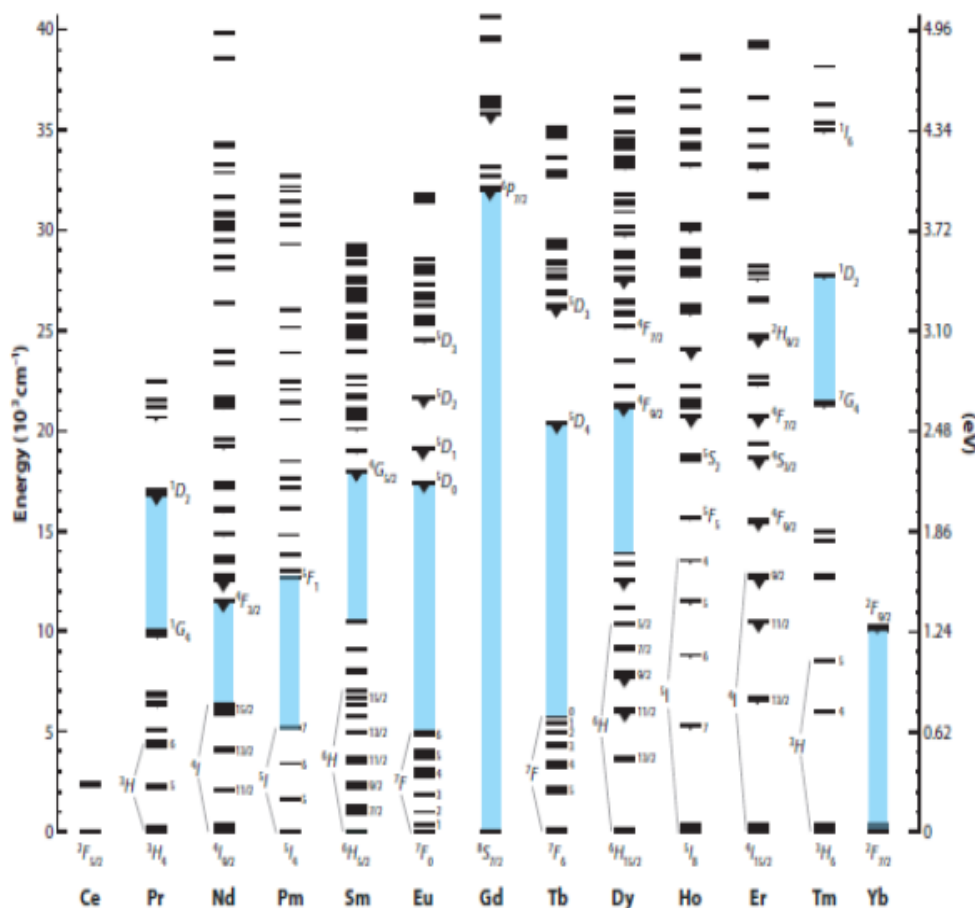
cepanja energetskih nivoa. Uticaj ovih interakcija opada po istom redosledu kao što je prethodno navedeno. U tabeli 3.3 su prikazane elektronske konfiguracije jona trovalentnih retkih zemalja.

Tabela 3.3 Elektronske konfiguracije jona trovalentnih retkih zemalja.

Z	Jon	m _l							S	L	J	Osnovni term
		-3	-2	-1	0	1	2	3	Σs	Σl	Σ(L+S)	
21	Sc ³⁺								0	0	0	¹ S ₀
39	Y ³⁺								0	0	0	¹ S ₀
57	La ³⁺								0	0	0	¹ S ₀
58	Ce ³⁺	↑							1/2	3	5/2	² F _{5/2}
59	Pr ³⁺	↑	↑						1	3	4	³ H ₄
60	Nd ³⁺	↑	↑	↑					3/2	6	9/2	⁴ I _{9/2}
61	Pm ³⁺	↑	↑	↑	↑				2	6	4	⁵ I ₄
62	Sm ³⁺	↑	↑	↑	↑	↑			5/2	5	5/2	⁶ H _{5/2}
63	Eu ³⁺	↑	↑	↑	↑	↑	↑		3	3	0	⁷ F ₀
64	Gd ³⁺	↑	↑	↑	↑	↑	↑	↑	7/2	0	7/2	⁸ S _{7/2}
65	Tb ³⁺	↑↓	↑	↑	↑	↑	↑	↑	3	3	6	⁷ F ₆
66	Dy ³⁺	↑↓	↑↓	↑	↑	↑	↑	↑	5/2	5	15/2	⁶ H _{15/2}
67	Ho ³⁺	↑↓	↑↓	↑↓	↑	↑	↑	↑	2	6	8	⁵ I ₈
68	Er ³⁺	↑↓	↑↓	↑↓	↑↓	↑	↑	↑	3/2	6	15/2	⁴ I _{15/2}
69	Tm ³⁺	↑↓	↑↓	↑↓	↑↓	↑↓	↑	↑	1	5	6	³ H ₆
70	Yb ³⁺	↑↓	↑↓	↑↓	↑↓	↑↓	↑↓	↑	1/2	3	7/2	² F _{7/2}
71	Lu	↑↓	↑↓	↑↓	↑↓	↑↓	↑↓	↑↓	0	0	0	¹ S ₀

Dieke i saradnici su na osnovu optičkih spektara pojedinačnih jona inkorporiranih u kristalu LaCl₃ eksperimentalno utvrdili energetske nivoe 4f elektrona. Dieke-ov dijagram je prikazan na slici 3.9. Svaki nivo obeležen brojem J se dalje cepa na više podnivoa pod uticajem kristalnog polja. Vrednost broja J i simetrija kristalnog polja direktno utiču na broj podnivoa na koje će se nivo podeliti. Maksimalan broj podnivoa je definisan sa dva izraza: $2J + 1$, ukoliko J ima celobrojnu vrednost, a ukoliko je J polovina celog broja onda važi $J+1/2$. Nivoi obeleženi polukrugom predstavljaju nivoe

sa kojih se dešava emisija i ovaj proces postaje dominantan usled zavisnosti da sa povećanjem energetske razlike između nivoa opada brzina neradijativne emisije fotona. Kako maksimalne varijacije u energetske nivoima usled različitosti matrica ne prelaze nekoliko stotina cm^{-1} ovaj dijagram se može primeniti na skoro sve matrice.

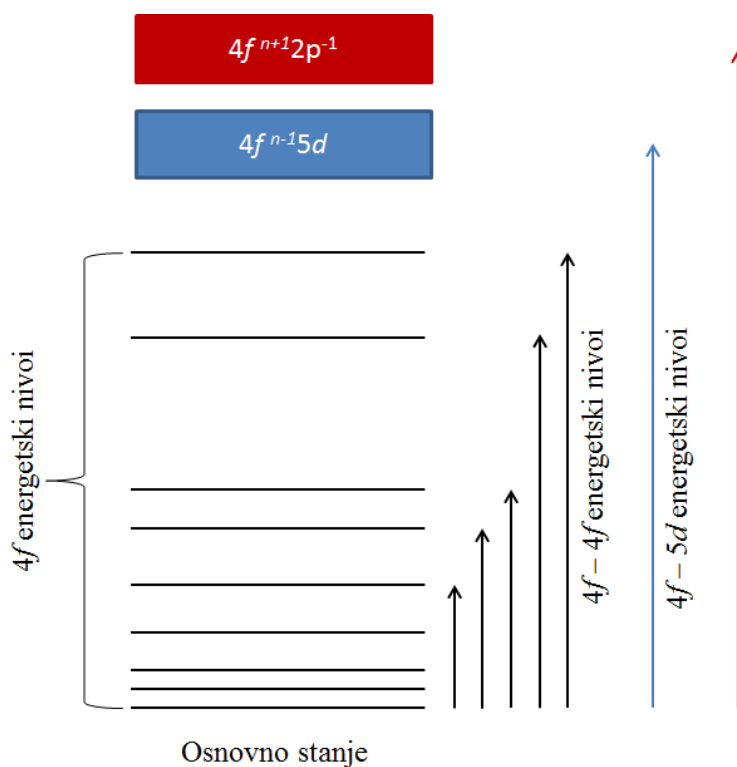


Slika 3.9 – Dieke-ov dijagram energetske nivoa.

3.3.2.2 Dozvoljeni i nedozvoljeni prelazi, selekciona pravila i intenziteti elektronskih prelaza

Kod oksida retkih zemalja dolazi do pojave tri vrste elektronskih prelaza: $4f-4f$, $4f-5d$ i prelaz prenosom naelektrisanja (eng. charge transfer). Na energetskom dijagramu oksida retkih zemalja prikazanom na slici 3.10 se jasno vidi šta se dešava sa energetskim nivoima pri apsorpciji energije.

Po pravilu parnosti prelazi $4f - 4f$, kod kojih dolazi do prelaza elektrona između različitih energetskih nivoa $4f$ orbitala u okviru istog jona retke zemlje, su zabranjeni. Pravilo parnosti nalaže da do elektronskih prelaza ne može doći ukoliko su energetski nivoi iste parnosti. U slučaju oksida retkih zemalja javlja se izuzetak od ovog pravila usled uticaja kristalnog polja i postojanja poremećaja poput vibracije elektrona. Kako je uticaj vibracija jako mali, glavnu odgovornost za ublažavanje pravila parnosti snosi uticaj kristalnog polja. Ovaj uticaj se javlja zbog okupacije jona retkih zemalja na mestu u rešetki koje nema inverznu simetriju (C_2). Rezultujući apsorpcioni pikovi ovih prelaza su uski i slabog intenziteta.



Slika 3.10 – Elektronski prelazi $4f - 4f$, $4f - 5d$ i prelazi prenosom naelektrisanja.

Kod $4f - 5d$ prelaza dolazi do premeštanja jednog $4f$ elektrona na nivo veće energije u $5d$ orbitali unutar istog jona retke zemlje. Takvi prelazi se označavaju kao $4f^n \rightarrow 4f^{n-1}5d$. Ovakvi prelazi su dozvoljeni po pravilu parnosti i imaju široke apsorpcione spektre visokog intenziteta.

Pri prelasku elektrona iz 2p orbitale okolnog jona O^{2-} na viši energetska nivo u 4f orbitali jona retke zemlje dolazi do pojave prelaza prenosom naelektrisanja i kao takvi se označavaju sa $4f^n \rightarrow 4f^{n+1}2p^{-1}$. Ovi prelazi su takođe dozvoljeni i imaju široke apsorpcijske pikove visokog intenziteta.

Radijativni prelazi odnosno emisija kod oksida retkih zemalja, do kojih dolazi po apsorpciji energije, su karakteristični za 4f – 4f prelaze. Ukoliko se elektroni iz pobuđenog stanja vrate na niži energetska nivo ili osnovno stanje bez emisije svetlosti, onda se to naziva neradijativnim prelazom. Još jedan od fenomena koji se može javiti jeste višefotonski prelaz do čije pojave može doći ukoliko je energetska razlika manja ili jednaka od vrednosti oko 5 puta veće od najviše vibracione frekvencije rešetke domaćina.

3.3.2.3 Definicija luminescentnih aktivatora i matrice domaćina

Svaki luminescentni materijal nastaje inkorporiranjem luminescentnog centra odnosno dopanta u matricu domaćina odnosno materijal domaćina (eng. host material). Postoje različite vrste matrice koje mogu biti organske i neorganske (češće se koriste), a to su:

- oksidi (Y_2O_3 , Gd_2O_3 , Lu_2O_3 , ZrO_2),
- mešani oksidi ($(Y_{1-x}Gd_x)_2O_3$, $(Y_{1-x}La_x)_2O_3$, $ZnO-SiO_2$),
- sulfidi (Y_2O_2S),
- fluoridi ($NaYF_4$, $KMgF_3$, $LiYbF_4$, YF_3 , $NaLuF_4$ itd.),
- silikati (Ln_2SiO_5 , $Ln_2Si_2O_7$),
- fosfati ($KCaPO_4$, $NaCaPO_4$, $BaCaPO_4$, $SrMg_2(PO_4)_2$) itd.

Zavisno od efekata koji žele da se postignu vrlo je bitno izabrati odgovarajuću matricu. Optička svojstva dobijenog materijala će puno zavisiti od interakcije matrice sa jonom dopanta kao i od svojstva same matrice. Matrica će diktirati relativnu prostornu poziciju jona dopanata, njihovu međusobnu udaljenost i vrstu anjona koji ih okružuju. Ukoliko postoje defekti u matrici može doći do povećanja multifononske relaksacije metastabilnih nivoa i smanjenja intenziteta emisije. Dobri kandidati za matricu domaćina su matrice na bazi Na^+ , Ca^{2+} i Y^{3+} jer su njihovi jonski radijusi slični

veličinama jona lantanoida usled čega se smanjuje napon u kristalnoj rešetki i stvaranje defekata.

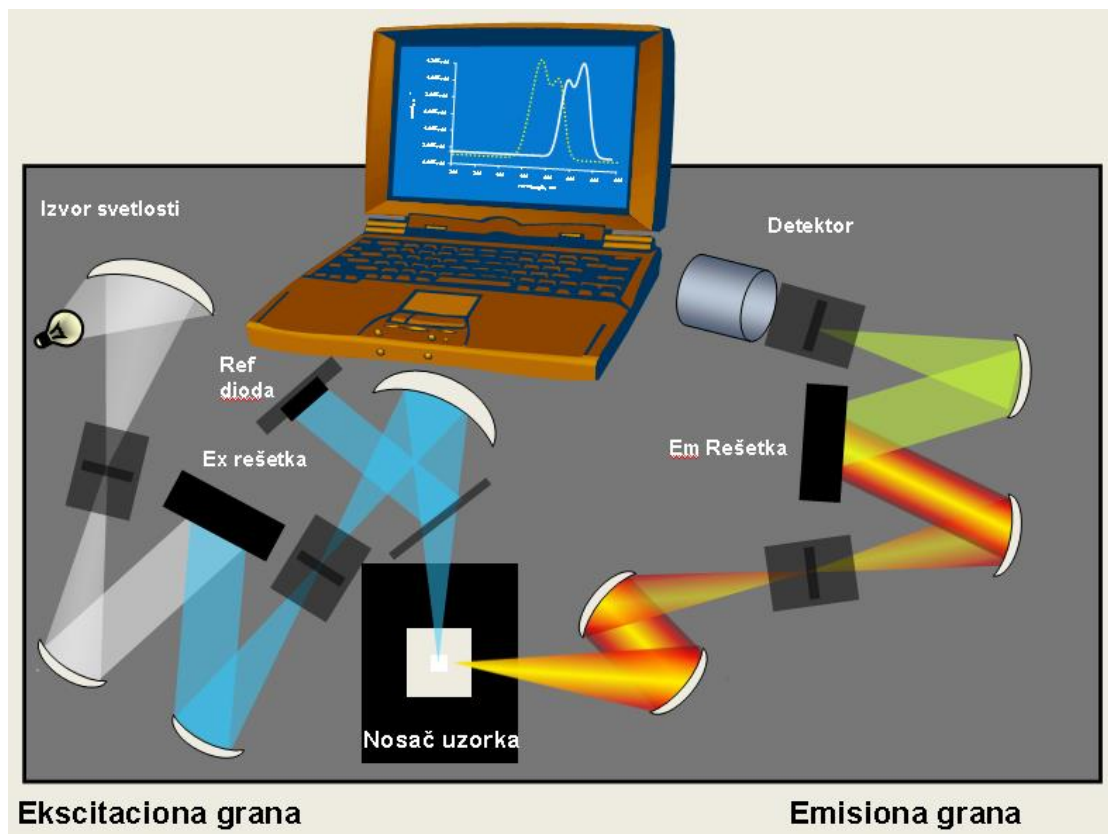
Joni sa kojih se emituje svetlost odnosno luminescentni centri u matrici nazivaju se aktivatorskim centrima ukoliko potiču od aktivatora. Emisione i apsorpcione karakteristike materijala će zavisiti direktno od vrste jona koji je ugrađen u matricu. Dobijeni spektri će davati određene boje. Ovakvi joni dopanta se unose u jako malim količinama u matricu domaćina pri čemu uglavnom zamenjuju neke jone domaćina. Pregled nekih karakterističnih luminescentnih centara i boja njihove emisije su date u tabeli 3.4.

Tabela 3.4 Boje emisija jona lantanoida

Joni	Boje emisije	Joni	Boje emisije
Ce³⁺	Ultraljubičasta, Zelena	Tb³⁺	Zelena
Pr³⁺	Plava, Zelena, Crvena	Dy³⁺	Žuta
Nd³⁺	Bliska infracrvena	Ho³⁺	Plava, Zelena
Sm³⁺	Roze	Er³⁺	Zelena, Infracrvena
Eu³⁺	Crvena	Tm³⁺	Plava
Gd³⁺	Ultraljubičasta	Yb³⁺	Plava, Bliska infracrvena

3.4 Uređaj za merenje fluorescencije

Za merenje fluorescentnih spektara koriste se spektrofotometri. Ova metoda se uglavnom koristi za karakterizaciju organskih i neorganskih materijala. Spektrofotometar predstavlja vrlo jednostavan uređaj čija je šema prikazana na slici 3.11. Osnovne komponente ovog uređaja su: izvor elektromagnetnog zračenja (svetlosti), monohromator i detektor. U sastavne delove takođe spadaju i nosač uzorka (bilo tečnih ili čvrstih) i kompjuter za registrovanje merenog signala.



Slika 3.11 – Šematski prikaz spektrofotometra.

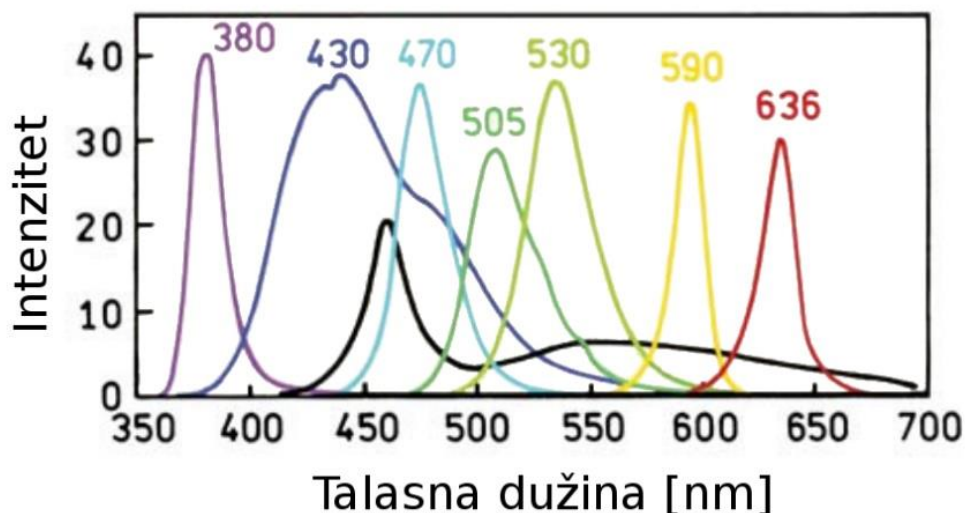
3.4.1 Izvor elektromagnetnog zračenja

Za pobuđivanje elektronskih prelaza mogu se koristiti različiti izvori svetlosti poput kontinualnih ili impulsnih lasera, lampe, emitujuće LE diode (*eng.* LE-light emitting) itd. Vrsta izvora pobude zavisi od fotoluminescentne metode. Kako je intenzitet luminescencije proporcionalan intenzitetu svetlosti ekscitacije, neophodno je da se svetlost generiše iz stabilnog visoko-energetskog izvora. Kao najčešći izvor u spektrofotometrima se koristi ksenonska lampa. Ona predstavlja kontinualni izvor koji emituje zračenje u širokom opsegu talasnih dužina (190 nm – 1000 nm), sa nekoliko oštrih linija koje se javljaju oko 450 nm i iznad 800 nm, i samim tim ima najširu primenu. Ksenonske lampe emituju kontinualnu svetlost kao rezultat rekombinacije elektrona sa jonizovanim atomima ksenona. Ovi joni se generišu pri sudarima atoma ksenona sa slobodnim elektronima koji se nalaze u lampi. Potpuna separacija elektrona od atoma rezultuje kontinualnom emisijom. Atomi ksenona koji su u pobuđenom stanju ali nisu jonizovani kao rezultat imaju uske emisione linije umesto širokih traka. Pikovi

koji se javljaju na 450 nm su upravo rezultat tih pobuđenih stanja. Dobra strana upotrebe ksenonske lampe jeste njena dugotrajnost, kratko vreme zagrevanja i osobina da ne zagreva previše sam instrument. Danas kompaktni spektrofotometri često koriste impulsne ksenonske lampe. Razlog tome je što impulsna ksenonska lampa daje bolje strukturiran izlaz od kontinualne lampe. Impulsna ksenonska lampa troši manje energije i generiše manje toplote.

Osim ksenonskih lampi često se koriste i živine lampe koje su primer linijskog izvora. Ove lampe su jačeg intenziteta od ksenonskih, ali je intenzitet skoncentrisan u linijama. One se najčešće primenjuju kada linije ekscitacije odgovaraju talasnim dužinama ekscitacije uzorka koji se meri.

LE diode (*eng.* LE-light emitting) – LED izvor svetlosti je našao primenu u novije vreme u spektrofotometriji. LED izvori se mogu naći u širokom opsegu talasnih dužina. U tom slučaju, za dobijanje izvora širokog opsega mogu se koristiti više LED izvora istovremeno. Prednosti LED-a su dug vek trajanja i niska potrošnja energije.

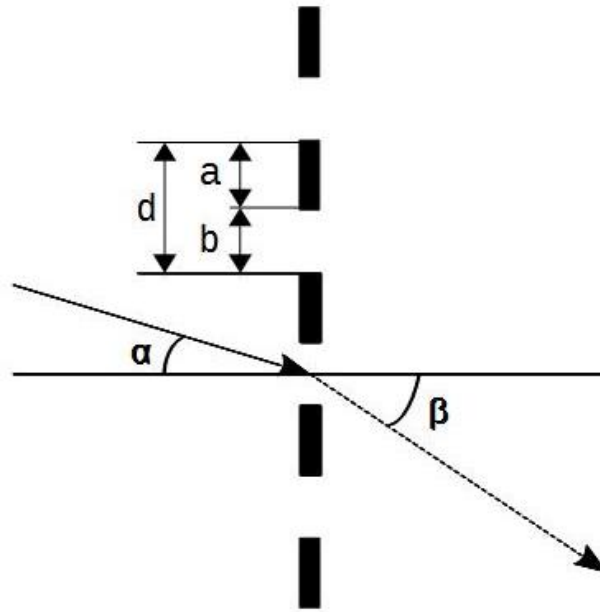


Slika 3.12 – Spektralni izlaz LE dioda, pri čemu crna linija predstavlja izlaz bele LE diode.

Još jedan od izvora svetlosti su laserske diode. Za razliku od LED-a, laserske diode emituju monohromatsku svetlost i dostupne su u opsegu talasnih dužina od 405 do 1500 nm (slika 3.12). Prednost laserskih dioda je u tome što je izlaz jednostavan za fokusiranje i manipulaciju. Za razliku od raznih lampi, izlaz LED-a i laserskih dioda može biti i pulsni i modulisan. Amplituda LED-a može biti modulisana čak i do 100 MHz, dok laserske diode mogu biti modulisane i do nekoliko GHz.

3.4.2 Monohromator

Monohromator predstavlja deo spektrofotometra koji je odgovoran za razlaganje polihromatske svetlosti u monohromatsku. On se sastoji iz tri dela: ulaznog razreza (eng. slit), disperzionog elementa i izlaznog razreza. Za disperziju se može koristiti prizma ili difrakciona rešetka. Disperzioni element koji se najčešće koristi je difrakciona rešetka. Specifikacija performansi monohromatora podrazumeva disperziju, efektivnost i nivo lutajuće svetlosti. Disperzija je uglavnom data u nm/mm. Monohromator treba da ima nizak nivo lutajuće svetlosti kako bi se izbegli problemi usled rasejanja ili zalutale svetlosti. Pod lutajućom svetlosti se podrazumeva svetlost propuštena od strane monohromatora, ali izvan željenog opsega talasnih dužina. Rezolucija snimanja će zavistiti od širine razreza monohromatora. Razrezi su uski prorezi kroz koje svetlost ulazi i izlazi iz monohromatora, respektivno. Širine razreza su najčešće podesive (od 0.005 – 1 mm) i monohromator najčešće ima i ulazni i izlazni razrez. Bitna karakteristika razreza jeste da ivice budu paralelne i pravolinijske kako bi se dobila simetrična spektralna linija pravilnog oblika. Razrezi mogu imati samo jednu ili obe pomerljive ivice, pri čemu u slučaju kada su obe pomerljive ne dolazi do pomeranja centra spektralnih linija pri promeni širine ivice. Intenzitet svetlosti koji prođe kroz razrez je uglavnom srazmeran kvadratu širine razreza. Širi razrezi rezultuju većim intenzitetom signala, ali takođe i većim odnosom signal-šum. Manji razrezi daju bolju rezoluciju, ali samim tim i manji intenzitet signala. Stoga zavisno od materijala koji se meri moraju se izabrati i odgovarajuće širine razreza.



Slika 3.13 – Šematski prikaz transmisione difrakcione rešetke.

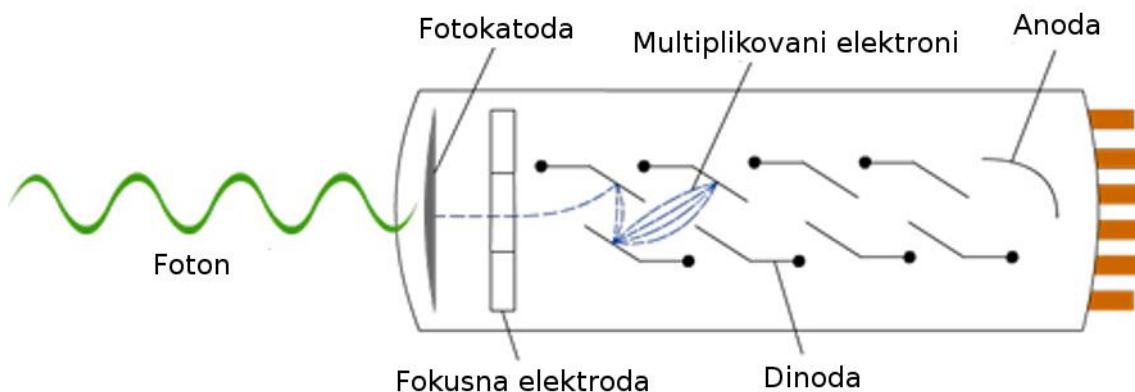
Difrakciona rešetka je optička komponenta sa periodičnom strukturom koja se sastoji od bliskih zarezata (a) između kojih se nalaze zatamnjeni delovi (b) kroz koje svetlost ne prolazi (transmisiona rešetka). Period rešetke se definiše kao konstanta jednaka zbiru rastojanja a i b (slika 3.13). Prorezi na rešetki mogu biti zamenjeni reflektujućim slojem i onda se takva rešetka naziva refleksijska, koja se najčešće i koristi u praksi. Moć razlaganja rešetke zavisi od gustine zarezata, odnosno što je veća gustina bolja je moć razlaganja. Kod kvalitetnih rešetki zarezata se urezuju holografskim postupkom na glatkoj staklenoj površini presvučenom reflektivnim slojem. Nesavršenosti rešetke su izvor lutajuće svetlosti transmisije monohromatora i pojave nepostojećih signala sa rešetke. Moć razlaganja rešetke je definisana sledećim izrazom:

$$R = \frac{\lambda}{d\lambda} = Nm \quad (1.77)$$

Gde je N ukupan broj zarezata na rešetki, a m difrakcioni ili spektralni red rešetke ($m=1, 2, 3, \dots$). Jednačina (1.77) upravo potvrđuje da moć razlaganja zavisi samo od broja zarezata i difrakcionog reda rešetke i da se sa većim brojem zarezata dobija i veća rezolucija.

2.4.3 Detektor

Kao detektor u spektrofotometru se može koristiti bilo koji fotosenzitivan uređaj. Najčešće korišćeni detektori u praksi su fotomultiplikator i CCD (eng. Charge Coupled Device) kamera.



Slika 3.14 – Šematski prikaz fotomultiplikatora.

Fotomultiplikator je detektor koji je jako osjetljiv na elektromagnetno zračenje u ultraljubičastom, vidljivom i infracrvenom opsegu. On pojačava struju proizvedenu od strane upadne svetlosti i do 100 miliona puta, što omogućava detekciju pojedinačnih fotona čak i kada je intenzitet svetlosti jako mali. Na slici 3.14 je prikazana šema fotomultiplikatora. Može se uočiti da osnovni delovi fotomultiplikatora čine fotokatode odnosno nekoliko dinoda i anode. Proces pojačavanja u fotomultiplikatoru se odvija na sledeći način:

- Elektromagnetno zračenje pada na fotokatodu i usled fotoefekta se izbijaju primarni elektroni koji počinju da se kreću ka prvoj dinodi.
- Svaka dinoda se nalazi na većem naponu (za oko 100 V) od prethodne. Grupa primarnih elektrona, nastala dolaskom inicijalnih fotona, se kreće ka prvoj dinodi pri čemu bivaju ubrzani električnim poljem.
- Pri udaru u dinodu izbija se više elektrona manje energije koja im je dovoljna da dalje izbijaju nove elektrone procesom sekundarne emisije.

- Ovaj proces se ponavlja dalje na svakoj dinodi i kao rezultat se dobija veliki broj elektrona koji stiže do anode i time omogućava lako detektovanje čak i slabog signala.

Problem koji se može javiti pri korišćenju fotomultiplikatora jeste šum i struja mraka koji se javljaju usled termoelektronske emisije. Hlađenjem fotomultiplikatora smanjuje se i termoelektronska emisija, a time i prateći fenomeni.

CCD (eng. Charge Coupled Device) detektor je uređaj koji može istovremeno da meri intenzitet svetlosti u više tačaka u ravni i konvertuje ga u digitalni signal koji se snima na nekom uređaju (kompjuteru). Prednost ovog detektora je što omogućava istovremeno snimanje spektra u širem spektralnom intervalu i iz tog razloga se ne koristi izlazni razrez. Ova vrsta detektora ima izuzetnu senzitivnost linerni dinamički opseg. Rezolucija snimanja spektra će zavisiti od broja mernih tačaka (piksela) po jedinici dužine senzora, odnosno što je taj broj veći i rezolucija snimanja će biti bolja. Standardan CCD detektor ima minimum 10^6 piksela. Karakteristično za CCD čip je njegov oblik matrice, u smislu da svaki red matrice snima po jedan spektar i kao rezultat daje usrednjenu vrednost spektara svih redova CCD čipa. Svaki piksel se ponaša kao jedan akumulirajući detektor gde se naelektrisanje akumulira proporcionalno količini svetlosti kojoj je izložen. Naelektrisanje svakog piksela može biti očitano u svakom trenutku i pri tome se dobija dvodimenzionalna slika. U praksi se koristi i kamera sa CCD senzorom (ICCD) koja može da radi u dva režima: impulsnom i kontinualnom. U cilju smanjenja struje mraka, a time i šuma, CCD senzor se može hladiti na -20°C uz primenu Peltijeovog elementa. Još jedan od bitnih delova kamere jeste i tzv. sistem za pojačavanje intenziteta slike koji ima funkciju za pojačavanje signala i kao brzi elektronski zatvarač koji se koristi u impulsnom režimu. Mana CCD senzora je efekat koji se javlja kada je jedan piksel izložen velikom intenzitetu svetlosti, pa time dolazi do lažnog povećanja izmerenog intenziteta na susednim pikselima iako do njih stiže svetlost relativno malog intenziteta. Ovaj efekat se naziva "preslušavanje". Stoga se može zaključiti da će linija detektovana fotomultiplikatorom imati manju širinu od one detektovane CCD senzorom. Zbog svojih malih dimenzija ova vrsta detektora se često koristi kod prenosnih spektrofotometara.

3.4.4 Merenje luminescentnih spektara i vremena života pobuđenog stanja

Kod većine spektrofotometara moguće je meriti i emisione i ekscitacione luminescentne spektre. Emisioni spektar podrazumeva merenje emisije za određeni opseg talasnih dužina pri čemu je talasna dužina ekscitacije konstantna. Ekscitacioni spektar predstavlja intenzitet luminescencije u funkciji talasne dužine ekscitacije pri fiksnoj talasnoj dužini emisije. U cilju detekcije najvećeg mogućeg broja spektralnih karakteristika posmatranog sistema koristi se složenije fluorescentne tehnike: ekscitaciono-emisiona matrica (EEM) i sinhrona fluorescentna spektroskopija. EEM predstavlja trodimenzionalni dijagram koji prikazuje intenzitet luminescencije u funkciji energije ekscitacije i emisije. Ovakav dijagram se dobija snimanjem više emisionih spektara za određene vrednosti talasne dužine ekscitacije i zatim se dobijeni linijski spektri slažu u matricu gde redovi i kolone odgovaraju talasnim dužinama emisije i ekscitacije, respektivno. EEM se pokazala kao jako korisna tehnika kod ispitivanja višekomponentnih sistema zbog velikog broja informacija koje se na ovaj način prikupe. Sinhroni fluorescentni spektri se dobijaju simultanim snimanjem ekscitacionih i emisionih spektara pri čemu se interval razlike talasnih dužina ($\Delta\lambda$) ili frekvencije između ekscitacije i emisije održava konstantnim.

Merenje ekscitacionih i emisionih spektara se u glavnom vrši u kontinualnom režimu, odnosno kad se za pobuđivanje koristi lampa koja u idealnom slučaju emituje konstantan broj fotona po jedinici vremena. Iako idealna kontinualna lampa ne postoji, može se smatrati da je broj pobuđenih elektrona stalan jer je konstanta brzine deekscitacije za nekoliko redova veličine manja od konstante brzine pobuđivanja elektrona fotonom. U novije vreme se osim kontinualnog režima koristi i vremenski režim odnosno merenje vremenski razloženih spektara. U tom slučaju se za pobuđivanje koriste impulsni laseri, a za detekciju jako brzi detektori. U tom slučaju je moguće u toku snimanja menjati vreme između detektovanja signala i na taj način dobiti različite spektre, zavisno od toga šta se traži. U slučaju kompleksnih sistema vremenski razloženi spektri će se razlikovati po različitim karakteristikama emisije supstanci (njihovih molekula, jona ili atoma) koje se nalaze u smeši.

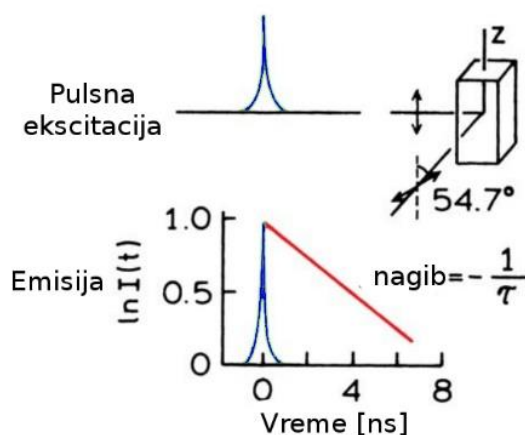
Merenje vremena života pobuđenog stanja podrazumeva merenje prosečnog vremena koje molekuli provedu u pobuđenom stanju pre povratka u osnovno stanje. Vreme života fluorescencije je oko 10 ns. Kako u fluorescentnim merenjima mi ne posmatramo broj pobuđenih molekula (n) već samo promenu intenziteta fluorescencije ($I(t)$), koji je proporcionalan broju n , matematički izraz koji opisuje vreme života odnosno vreme koje je potrebno da intenzitet fluorescencije opadne za $1/e$ od početne vrednosti je

$$I_t = I_0 \exp\left(-\frac{t}{\tau}\right) \quad (1.78)$$

gde je I_0 intenzitet u trenutku $t = 0$ s, a τ vreme života prelaza. Kao što je prethodno naznačeno vreme života je prosečno vreme koje molekuli provedu u pobuđenom stanju. To se naglašava iz razloga što u kompleksnim slučajevima kada postoji više prelaza ili prenosa energije u sistemu izraz (1.78) neće biti tačan i može da odstupa od eksponencijalne funkcije. Iz tog razloga se uvodi efektivno vreme života emisije koje usrednjava vreme t po intenzitetu emisije

$$\tau_{ef} = \frac{\int_0^{\infty} t I_t dt}{\int_0^{\infty} I_t dt} \quad (1.79)$$

Gde je I_t intenzitet fluorescencije u trenutku t za $0 < t < t_m$; $t_m \gg \tau_{ef}$.

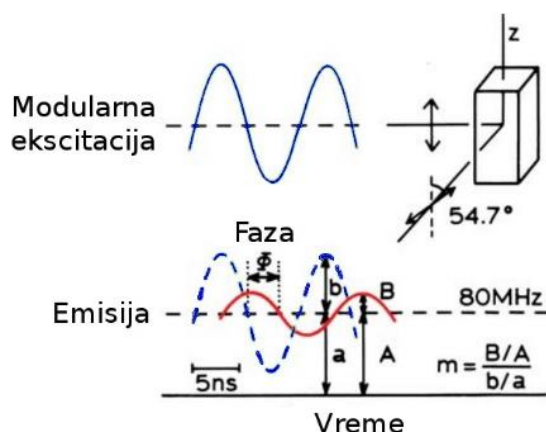


Slika 3.15 – Merenje vremena života u vremenskom domenu.

Za merenje vremena života pobuđenog stanja koriste se dve metode: merenje u vremenskom domenu i merenje u frekventnom domenu. Na slici 3.15 je grafički prikazano merenje vremena života pobuđenog stanja u vremenskom domenu, gde se uzorak pobuđuje pulsom svetlošću. Puls treba da bude što je moguće uži i ako je moguće mnogo kraći od vremena života τ uzorka koji se meri. Vremenski zavisani intenzitet fluorescencije se meri odmah nakon pobuđivanja pulsom svetlošću i vreme života τ se računa na osnovu krive sa grafika koji prikazuje zavisnost $\log I(t)$ i t ili vremena za koje se intenzitet smanji za $1/e$ od početne vrednosti intenziteta. Na slici 3.16 je prikazana alternativna metoda za merenje vremena života pobuđenog stanja. U ovom slučaju se uzorak pobuđuje svetlošću moduliranog intenziteta koja je najčešće sinusnog oblika. Intenzitet pobuđivačke svetlosti varira na visokim frekvencijama od oko 100MHz, tako da se recipročna vrednost frekvencije može porediti sa recipročnom vrednošću vremena života τ . Kada se uzorak pobudi na ovaj način emisija je prinuđena da isprati istu modulaciju frekvencije. Vreme života uzorka uzrokuje kašnjenje emisije u odnosu na ekscitaciju koje je prikazano kao fazni pomak (Φ) u desno koji se koristi za računanje vremena života. U skladu sa tim vreme života se u tom slučaju dobija uz pomoć sledećeg izraza:

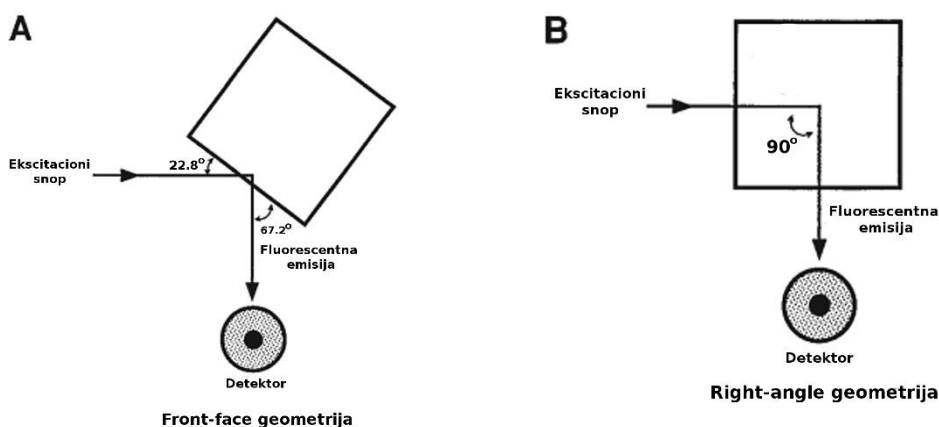
$$\tau = (\tan \Phi) / (2\pi f_{mod}) \quad (1.80)$$

gde je f_{mod} frekvencija modulacije ekscitacije. Merenja vremena života pobuđenog stanja u frekventnom modu su se jako lepo pokazala u slučaju kada su vrednosti vremena života reda veličine u opsegu $\mu s - ms$, dok se za kraća vremena života više koriste merenja u vremenskom domenu. Merenje vremena života ima široki opseg primene kako za organske tako i za neorganske materijale. Na taj način se mogu dobiti neke dodatne informacije koje nije moguće dobiti standardnim merenjima spektara. Na primer ova vrsta merenja se često koristi u slučaju kada u uzorku postoje dva molekula sa vrlo sličnim apsorpcionim i emisionim spektrima koji se preklapaju, ali čija su vremena života pobuđenog stanja različita. Na taj način merenjem spektara i vremena života zajedno moguće je razdvojiti odgovarajuće čiste spektre.



Slika 3.16 – Merenje vremena života u frekventnom domenu.

Postoje dve geometrije luminescentnih merenja koje se mogu koristiti: geometrija pravog ugla (eng. right-angle) i *front-face* geometrija (slika 3.17).



Slika 3.17 – Grafički prikaz front-face (A) i right-angle (B) geometrije.

Kada se koristi geometrija pravog ugla, emisija se detektuje pod pravim uglom u odnosu na pobudni zrak pri čemu je zrak fokusiran u centru kivete. Loša strana ove geometrije je u tome što na primer kod visoko apsorbujućih rastvora sva apsorpcija će se desiti blizu prednje površine kivete, sa vrlo malim pobuđivanjem koje stigne do samog centra što rezultuje jako slabim intenzitetom emisije. U suštini kod geometrije pravog ugla može doći do pojave samoapsorpcije kada se jako apsorbujući rastvor ponaša kao unutrašnji filter. Ovakav efekat se eliminiše *front-face* geometrijom kod koje upadni zrak zaklapa ugao od 22.8° ili 67.2° sa prednjom površinom kivete zavisno od

orijentacije adaptera, pri čemu je zrak fokusiran na prednju površinu uzorka. Na ovaj način se minimizira reflektovana i rasejana svetlost. *Front-face* geometrija se najviše koristi kod jako koncentrisanih, neprozirnih rastvora i čvrstih uzoraka poput krvi, živih ćelija, polimernih filmova ili fosfora.

4. Materijali i metode

U ovoj disertaciji mereno je i analizirano četiri sistema kompleksnih fluorescentnih karakteristika (sadrže više od jedne fluorofore) od čega su tri organska sistema i jedan neorganski:

1. Rastvor vode i dve aminokiseline (triptofan i tirozin) u različitim odnosima koncentracija.
2. Uzorci meda različitog botaničkog porekla (lipa, livada, bagrem, suncokret i falsifikati).
3. Uzorci tkiva kancera dojke sa malignim promenama i uzorci normalnog tkiva dojke.
4. Tri neorganska jedinjenja (Gd_2O_3 , $Gd_2Ti_2O_7$ i $GdVO_4$ dopirani europijumom (Eu^{3+})) čista i u mešavinama sa druga dva jedinjenja u različitim odnosima koncentracija.

4.1 Priprema uzoraka

4.1.1 Rastvor vode i aminokiselina

Za prvi sistem bilo je pripremljeno 6 uzoraka za dalju analizu. Priprema uzoraka je podrazumevala rastvaranje dve aminokiseline, triptofana i tirozina, u vodi u šest različitih odnosa njihovih koncentracija. Tačne količine aminokiselina za svaki uzorak su date u tabeli 4.1. Date količine aminokiselina su bile rastvarane u 200 ml destilovane vode na sobnoj temperaturi. Pošto su aminokiseline lako rastvorne u vodi na sobnoj temperaturi, nije bilo potrebno dodatno zagrevanje jedinjenja. Dobijeni rastvor je bio sipan u kvarcne kivete zapremine 3 ml, kvadratnog poprečnog preseka dimenzije 10×10 mm, a zatim je kiveta postavljena u poseban držač za kivete u spektrofotometru.

Tabela 4.1 Količina rastvorenih aminokiselina za svaki uzorak.

<i>Br. uzorka</i>	<i>Triptofan [mg]</i>	<i>Tirozin [mg]</i>
1.	0	150
2.	60	0
3.	30	180
4.	40	110
5.	40	150
6.	55	100

4.1.2 Med

Za drugi sistem bilo je sakupljeno 117 uzoraka meda (12 lipa, 30 livade, 47 bagrema, 14 suncokreta i 14 falsifikata) uz pomoć saveza pčelarskih organizacija Srbije (SPOS; www.spos.info). Svi uzorci su pre merenja bili zagrejeni na temperaturi od 40°C radi lakšeg sipanja uzoraka u kvarcne kivete za merenje. Osim zagrevanja, pre merenja uzorci meda nisu bili ni na koji drugi način tretirani. Svaki uzorak je bio sipan u kvarcnu kivetu zapremine 3 ml, kvadratnog poprečnog preseka dimenzije 10 × 10 mm, a zatim je kiveta postavljena u poseban držač za kivete u spektrofotometru.

4.1.3 Tkivo dojke

Uzorci kancera (8 uzoraka) i normalnog (10 uzoraka) tkiva dojke su bili sakupljeni nakon operacije 18 pacijenata u Institutu za Onkologiju i Radilogiju Srbije. Svi uzorci su bili histopatološki ispitani i zatim odloženi na -80°C do početka merenja. Po odmrzavanju uzoraka, skalpelom su bili odsečeni delovi tkiva veličine od 0.2 × 0.5 × 0.5 cm do 0.3 × 1.0 × 1.5 cm. Tako isečen uzorak je zatim stavljan između dve staklene pločice koje su učvršćene, a zatim je pločica postavljena na specijalan držač u spektrofotometru.

4.1.4 Neorganski materijali

Tri neorganska jedinjenja $Gd_2O_3 :Eu^{3+}$, $Gd_2Ti_2O_7 :Eu^{3+}$ i $GdVO_4 :Eu^{3+}$ su bila sintetisana kompleksnim hemijskim metodama: metoda polimerno-kompleksnog rastvora [Andrić *et al.* 2008], metoda precipitacije [Jovanović *et al.* 2013] i metoda mešovito metal limunska kiselina kompleksa [Ćulubrk *et al.* 2014]. Navedene metode neće biti detaljno opisane u ovom radu jer nisu predmet ispitivanja ove teze, a čija metodologija je detaljno opisana u navedenim naučnim radovima. Za merenje je bilo pripremljeno 7 uzoraka za kalibraciju i 21 uzorak za testiranje modela u kojima su se ova jedinjenja nalazila čista ili su bila međusobno mešana u različitim odnosima (tabela 4.2).

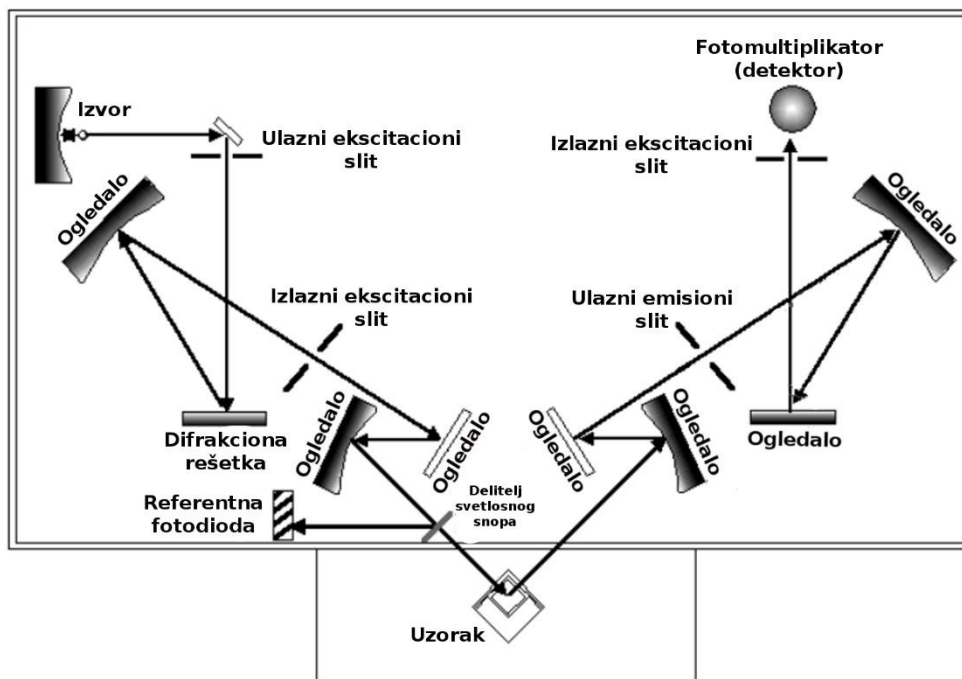
Tabela 4.2 Odnosi koncentracija tri neorganska jedinjenja za svaki uzorak.

	Broj uzorka	$Gd_2O_3 :Eu^{3+}$ [%]	$Gd_2Ti_2O_7 :Eu^{3+}$ [%]	$GdVO_4 :Eu^{3+}$ [%]
Set za kalibraciju	1.	26	0	74
	2.	43	20	37
	3.	43	57	0
	4.	72	0	28
	5.	0	90	10
	6.	0	70	30
	7.	0	80	20
Set za testiranje	1.	97	2.5	0.5
	2.	97	2	1
	3.	65	35	0
	4.	65	0	35
	5.	80	20	0
	6.	80	0	20
	7.	93	7	0
	8.	94	0	6
	9.	98	0	2
	10.	99.7	0	0.3
	11.	99.9	0.1	0
	12.	99.9	0	0.1
	13.	99	1	0

14.	99	0	1
15.	0	65	35
16.	0	80	20
17.	0	92	8
18.	0	98	2
19.	0	99.5	0.5
20.	0	99.9	0.1
21.	0	99	1

4.2 Merenje fluorescentnih spektara

Merenja fluorescentnih spektara su bila izvedena na dva spektrofotometra: Perkin Elmer LS45 i Horiba JobinYvon FL-221. Šeme ova dva uređaja su prikazane na slikama 4.1 i 4.2, respektivno.



Slika 4.1 – Šema spektrofotometra Perkin Elmer LS45.

Kao izvor pobude u spektrofotometru Perkin Elmer LS45 koristi se impulsna ksenonska lampa dok je detektor fotomultiplikatorska cev Hamamatsu R928. Ovaj

uređaj takođe ima i dva monohromatora, ekscitacioni i emisioni, čiji su opsezi rada od 200 do 800 nm. Princip rada je sledeći:

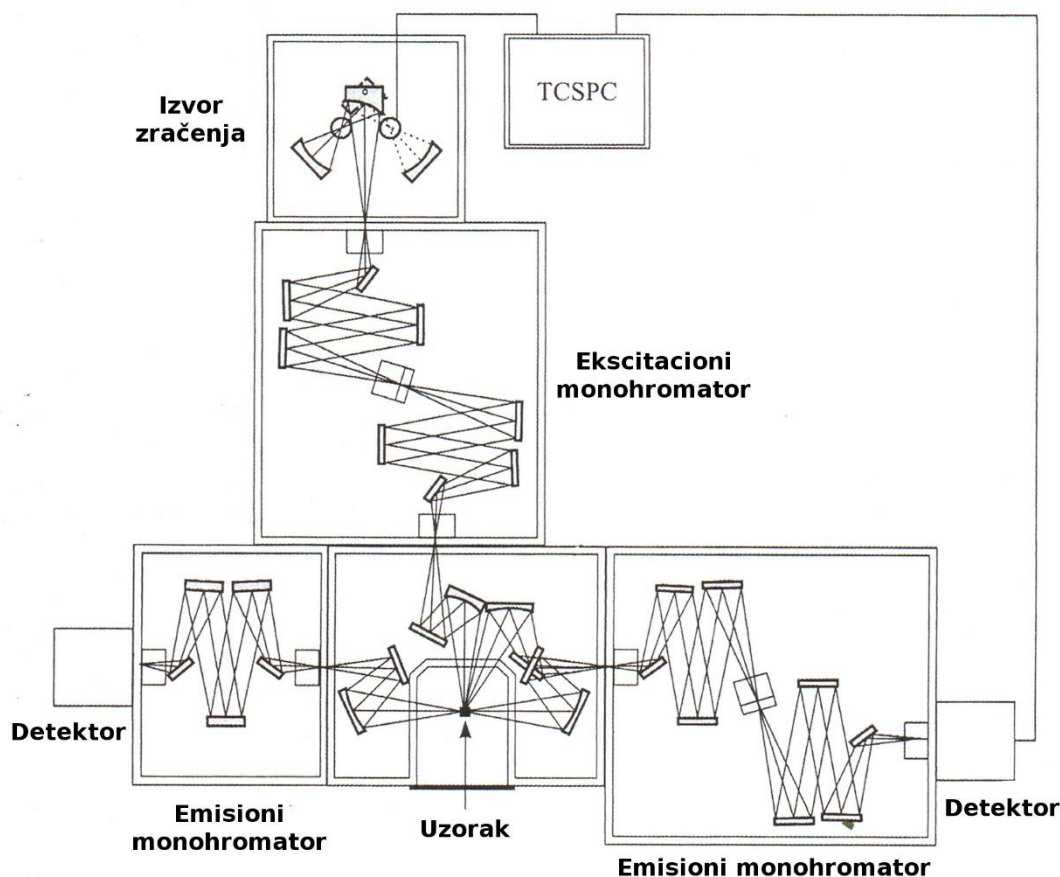
- Od lampe se emituje svetlost i uz pomoć sistema ogledala usmerava ka ulaznom razrezu monohromatora.
- Polihromatska svetlost koja uđe u monohromator se usmerava na difrakcionu rešetku gde se kao rezultat interferencije dobija monohromatska svetlost željene talasne dužine koja se dalje usmerava ka izlaznom razrezu monohromatora.
- Monohromatska svetlost zatim stiže do delitelja svetlosnog snopa gde se jedan deo svetlosti usmerava na referentni uzorak, koji za cilj ima automatsku korekciju intenziteta pobudne svetlosti, dok se drugi deo svetlosti usmerava ka uzorku koji je postavljen na poseban držač.
- Svetlost emitovana sa uzorka se dalje sakuplja na ogledalo koji ima poziciju pod uglom od 90° , a zatim usmerava na drugi monohromator (emisioni).
- Na kraju svetlost željene talasne dužine koja je izašla iz monohromatora usmerava se i stiže do detektora.
- Detektovan signal se zatim registruje na računaru i prikazuje izmereni spektar.

U spektrofotometru Horiba JobinYvon FL-221 se nalaze dve lampe koje se koriste kao izvor pobude. Ksenonska lampa 450-W se koristi kao izvor kod emisionih merenja dok se Xe-Hg pulsna lampa primenjuje za merenje vremenski razlučenih luminescentnih spektara i vremena života pobuđenog stanja. Za razliku od prethodnog uređaja ovaj ima dva emisiona monohromatora pri čemu je jedan dupli monohromator i pokriva opseg talasnih dužina od 200 do 850 nm, dok je drugi jednostruki monohromator i pokriva opseg od 800 do 1600 nm. Detektor korišćen za merenja za potrebe ove teze je fotomultiplikatorska cev TBX-04-D PMT, dok osim tog detektora ovaj uređaj poseduje i IR detektor koji se koristi za merenja u infracrvenom delu spektra. Princip rada ovog uređaja je:

- Svetlost izvora pobude se emituje i uz pomoć sistema ogledala usmerava ka ekscitacionom monohromatoru.
- Svetlost prolazi kroz dupli ekscitacioni monohromator, odnosno dva monohromatora vezana u red, pri čemu se dobija monohromatska svetlost

određene talasne dužine koja se zatim usmerava ka izlaznom razrezu monohromatora.

- Monohromatska svetlost se zatim usmerava ka uzorku i pobuđuje ga.
- Svetlost emitovana od uzorka zatim biva usmerena kaulazno razrezu duplog emisionog monohromatora koji se nalazi pod uglom od 90° u odnosu na osu ekscitacionog monohromatora. Ovom geometrijom se minimalizuju smetnje koje potiču od propuštene i rasejane ekscitacione svetlosti.
- Svetlost prolazi kroz emisioni monohromator, pri čemu se na izlaznom razrezu dobija monohromatski svetlosni signal određene talasne dužine koji stiže do detektora.
- Detektovani signal se prenosi na računar i na ekranu se prikazuje izmereni luminescentni spektar.



Slika 4.2 – Šema spektrofotometra Horiba JobinYvon FL-221.

Osim merenja ekscitacionih i emisionih spektara na ovom uređaju je moguće meriti i vreme života pobuđenog stanja. Ova merenja se vrše uz pomoć stanice sa podacima sa integrisanim TCSPC softverom (eng. Time Correlated Single Photon Counting). Pri ovim merenjima koristi se impulsna pobudna svetlost (Xe-Hg lampa) pri čemu je poželjno da impuls lampe bude kraći od vremena života merenog uzorka. Pri ovom merenju uzorak se pobuđuje serijom ponovljenih impulsa tokom čega se meri broj fotona koji se detektuje u mernim kanalima, što predstavlja vreme dolaska fotona na detektor, i na taj način se dobija emisija uzorka. Vreme života se na kraju računa na osnovu krive fluorescencije dobijene serijom pobudnih impulsa.

Za uzorke organskih sistema bile su merene ekscitaciono-emisione matrice, dok je za uzorke neorganskog sistema meren emisioni spektar i vreme života pobuđenog stanja. Sva merenja su bila izvedena u *front-face* geometriji merenja.

4.2.1 Ekscitaciono-emisione matrice (EEM)

EEM uzoraka sva tri organska sistema su bile merene na Perkin Elmer LS45 spektrofotometru. Za pobudu je bila korišćena impulsna ksenonska lampa dok je za detekciju bio korišćen detektor Hamamatsu R928. Ulazni i izlazni razrezi su bili 10 nm, pošto je njihova širina kod ovog uređaja fiksirana i nije je moguće podešavati. Primenom referentne fotofiode svi izmereni spektri su bili automatski korigovani na intenzitet pobude. Uslovi merenja za svaki sistem pojedinačno su bili sledeći:

- 1) Amino kiseline – opseg ekscitacije 240-500 nm sa korakom 5 nm, opseg emisije 270-500 nm sa korakom 0.5 nm, brzina snimanja u oba opsega je bila 1000 nm/min. Dobijena EEM je bila veličine 462×53 , odnosno bilo je snimljeno 53 emisiona spektra za 462 talasne dužine emisije.
- 2) Med – za ovaj sistem su bile izmerene EEM za 84 uzoraka pod sledećim uslovima merenja: opseg ekscitacije 240-495 nm sa korakom 5 nm, opseg emisije 270-640 nm sa korakom 0.5 nm, brzina snimanja u oba opsega je bila 200 nm/min. Dobijena EEM je bila veličine 741×52 .

- 3) Dojka – kod ovog sistema bila su merena dva regiona. Region I: opseg ekscitacije 335-400 nm sa korakom 5 nm, opseg emisije 430-645.5 nm sa korakom 0.5 nm, brzina snimanja u oba opsega je bila 150 nm/min. Dobijena EEM je bila veličine 432×14 . Region II: opseg ekscitacije 400-470 nm sa korakom 5 nm, opseg emisije 500-625 nm sa korakom 0.5 nm, brzina snimanja u oba opsega je bila 150 nm/min. Dobijena EEM je bila veličine 251×15 .

4.2.2 Sinhroni fluorescentni spektri

Sinhroni fluorescentni spektri uzoraka sva sistema meda i tkiva dojke su bile merene na Perkin Elmer LS45 spektrofotometru. Spektri su bili snimani uz održavanje konstantnog rastojanja između talasne dužine ekscitacije i emisije. Korišćeni detektor i izvor pobude su bili isti kao i za merenje EEM-a. Razrezi su bili 10 nm, a spektri su bili automatski korigovani na intenzitet pobude uz primenu referentne fotodiode. Uslovi merenja su bili sledeći:

- 1) Med – sinhroni spektri su bili izmereni za 109 uzoraka sa opsegom ekscitacije od 240-500 nm sa korakom 0.5 nm i sinhronim intervalom u opsegu od 30-300 nm sa korakom 5 nm (55 intervala). Kao rezultat merenja bilo je dobijeno 55 matrica dimenzija 109×521 .
- 2) Dojka - opseg ekscitacije od 330-650 nm sa korakom 0.5 nm i sinhronim intervalom u opsegu od 30-120 nm sa korakom 5 nm. Kao rezultat je dobijeno 19 matrica dimenzija 18×641 .

4.2.3 Emisioni spektri i vreme života

Za merenje emisionih spektara i vremena života uzoraka kompleksnog sistema fosfora na bazi retkih zemalja bio je korišćen Horiba JobinYvon FL-221 spektrofotometar. Za merenja emisionih spektara kao pobuda je korišćena 450-W ksenonska lampa, dok je za merenja vremena života korišćena Xe-Hg. Emisioni opseg merenja je bio od 570 do 650 nm u koraku od 0.1 nm, dok je talasna dužina ekscitacije bila 466 nm. Vreme života

pobuđenog stanja je bilo mereno nakon ekscitacije uzorka na 466 nm i zatim je dobijena kriva iz koje se računa vreme života merenog uzorka. Dobijeni spektri su bili dimenzije 801×101 , gde je bio meren intenzitet emisije za 801 talasnu dužinu emisije i za vreme od 0 do 10 ms sa korakom od 0.1 ms (101 vrednost).

4.3 Analiza i modeliranje dobijenih spektara

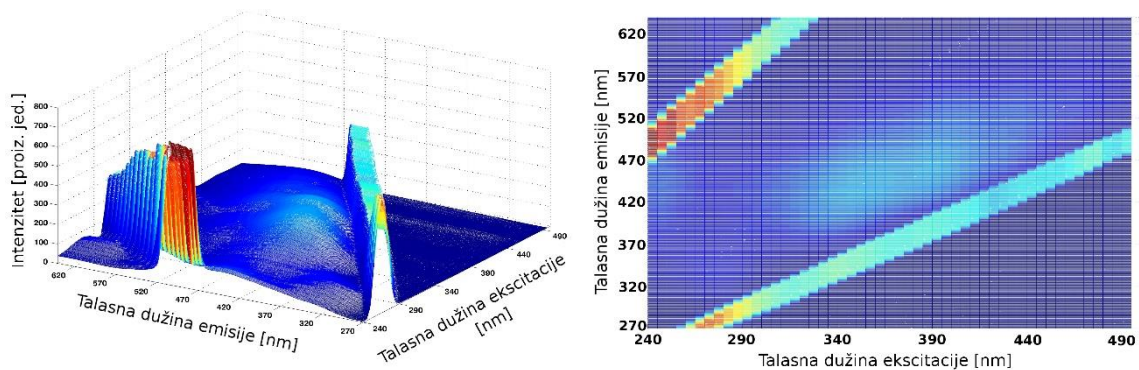
Deo obrade EEM spektara se sastojao iz dva dela: 1) adekvatna priprema odnosno predprocesiranje dobijenih EEM i 2) analiza i građenje PARAFAC modela merenih sistema kao i dalja analiza rezultata modela PLS-DA i SVM metodom. Analiza sinhronih spektara je obuhvatala takođe dva koraka: 1) predprocesiranje spektara i 2) analiza spektara primenom PCA, PLS-DA i SVM metoda. Za analizu i modeliranje spektara za potrebe ove teze bili su korišćeni programski paketi Solo verzija 6.5.4 (Eigenvector Inc., Chelan, WA, USA) i Matlab R2012a (uz dodatak N-way Toolbox-a, besplatno dostupnog na sajtu www.models.kvl.dk).

4.3.1 Priprema podataka za analizu

Kako bi osigurali tačnost dalje analize i modeliranja spektara neophodna je odgovarajuća priprema podataka za PARAFAC analizu. Da bi se podaci pripremili prvo ih je neophodno organizovati na odgovarajući način i to formiranjem trodimenzionalnih matrica za svaki sistem:

- 1) Podaci aminokiselina su bili organizovani u set podataka (eng. data set) koji je u sebi sadržao matricu dimenzije $6 \times 462 \times 53$ (broj uzoraka \times broj talasnih dužina emisije \times broj talasnih dužina ekscitacije) i informaciju o talasnoj dužini za svaki parametar. Kao posebna matrica bila je uneta i informacija o tačnim koncentracijama aminokiselina za svaki uzorak koja je bila neophodna za kasniju kalibraciju modela.

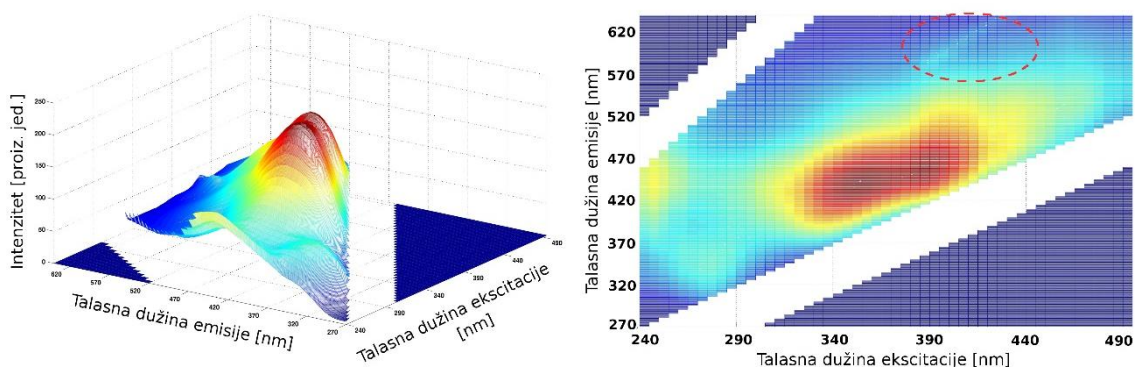
- 2) Podaci meda su bili organizovani u set podataka sa matricom dimenzije $84 \times 741 \times 52$ i informacijom o klasnoj pripadnosti svakog uzorka i talasnoj dužini kojoj svaki parametar odgovara.
- 3) Za podatke kancera dojke formiran je set podataka sa matricom veličine $18 \times 432 \times 14$ i informacijom o klasnoj pripadnosti svakog uzorka i talasnoj dužini kojoj svaki parametar odgovara.
- 4) Podaci kompleksnog sistema fosfora na bazi retkih zemalja su bili organizovani u dva seta podataka sa matricom dimenzije $7 \times 801 \times 101$ (broj uzoraka \times broj talasnih dužina emisije \times broj vremenskih mernih tačaka) za kalibracioni set i $21 \times 801 \times 101$ za test set i informacijom o tome kojem vremenskom trenutku i kojoj talasnoj dužini emisije svaki član matrice odnosno parametar odgovara.



Slika 4.3 – Trodimenzionalni (levo) i konturni (desno) grafik neobrađene EEM jednog uzorka meda.

Sinhroni spektri su već bili organizovani na odgovarajući način za dalju analizu tako da nisu zahtevali dodatnu doradu u tom smislu. Priprema podataka se uglavnom sastoji od centriranja podataka srednjom vrednošću, skaliranja i uklanjanja artefakta u podacima ukoliko postoje. Kao što je ranije bilo rečeno (odjeljak 2.3.13) kod EEM fluorescentnih spektara ne postoji potreba za centriranjem i skaliranjem podataka, dok se kod sinhronih spektara vrši centriranje radi lakše interpretacije rezultata PCA modela. Jedini problem koji se često javlja kod fluorescentnih podataka jeste pojava raznih artefakta poput rasejanja ili nekih drugih pikova koji nisu rezultat luminescentnih karakteristika supstanci u uzorku. Rasejanje se kod sinhronih spektara retko javlja jer se te talasne

dužine često ne obuhvate merenjem. Ovakve pojave se uglavnom lako uočavaju kod EEM jer se oni vide kao pikovi koji su ravni i konstantnog intenziteta. Na slici 4.3 je prikazan trodimenzionalni i konturni grafik neobrađene EEM jednog izmerenog uzorka meda. Može se primetiti očigledno prisustvo Rejljevog rasejanja prvog i drugog reda koje je mnogo većeg intenziteta od fluorescencije samog uzorka. Po uklanjanju Rejljevog rasejanja na način čiji je postupak bio objašnjen u prethodnom poglavlju dobija se EEM bez rasejanja gde se jasnije može videti izgled spektra merenog uzorka (slika 4.4). U ovom trenutku je takođe neophodno videti da li slučajno postoji još neka nepravilnost u spektru koja se nije mogla videti dok je rasejanje bilo prisutno. Na konturnom grafiku na slici 4.4 se uočava još jedan pik na talasnoj dužini ekscitacije od 400 do 430 nm i emisije od 580 do 630 nm (zaokruženo crvenom bojom na grafiku) oblika ravne linije koji se javljao na istom mestu i istog intenziteta kod svih uzoraka što može da bude vezano za sam uređaj, pa je i taj deo spektra bio odsečen jer u tom opsegu se u svakom slučaju nije javljao ni jedan drugi pik bitan za posmatrani sistem. Uzorci meda i aminokiselina su bili na isti način obrađivani, u smislu uklanjanja rasejanja i prethodno spomenutog artefakta, za koji se na kraju zaključilo da se javlja na svim EEM spektrima na toj poziciji merenim na spektrofotometru Perkin Elmer LS45. Kako se uzorci dojke nisu merili na talasnim dužinama gde se javlja Rejljevo rasejanje, kod njih nije bilo neophodno uklanjati te delove već samo navedeni artefakt. Celokupna priprema podataka je bila izvedena u programskom paketu Matlab. Kod sinhronih spektara ovaj artefakt se nije pojavljivao.



Slika 4.4 - Trodimenzionalni (levo) i konturni (desno) grafik EEM jednog uzorka meda sa uklonjenim Rejljevim rasejanjem.

4.3.2 Proračun i analiza PARAFAC modela

Posle odgovarajuće pripreme podataka može se pristupiti građenju PARAFAC modela na osnovu izmerenih i obrađenih EEM-a. Svaki proračun modela se sastoji iz više koraka. Prvo na osnovu prikupljenih informacija o merenim sistemima treba pretpostaviti broj komponenata modela, koji odgovara broju fluorofora koje se nalaze u posmatranom sistemu. Takođe je bitno napomenuti da nekad iako možda ima više fluorofora u sistemu je moguće da neke od njih nemaju veliki doprinos ili nisu dominantne u odnosu na druge tako da se one u tom slučaju zanemaruju u modelu. Kako je postupak proračuna PARAFAC modela vrlo kompleksan on neće biti detaljno opisan u daljem tekstu već će biti samo ukratko objašnjeno koji su parametri neophodni za proračun modela.

Parametri odnosno podaci koji su neophodni i koji se unose u kod pri modeliranju jednog sistema su: matrica X (podaci koje želimo da modeliramo), pretpostavljeni broj komponenata u sistemu, različite opcije modela, ograničenja itd. Matrica X predstavlja n-dimenzionalnu matricu koju želimo da modeliramo i za koju se podrazumeva da ima varijacije samo u jednoj dimenziji odnosno modu, dok su druge za sve uzorke iste. Na primer kod EEM to znači da su nam za sve uzorke jednog sistema ekscitacioni i emisijski spektri pojedinačnih fluorofora isti, a da se samo koncentracija tih fluorofora za svaki uzorak variraju. Opcije nam daju mogućnost ubacivanja nekih dodatnih parametara u proračunu poput: kriterijuma konvergencije – za koju vrednost relativne promene u fitu modela želimo da se algoritam zaustavi pri čemu je standardno podešeno da ta vrednost bude 10^{-6} ; metode inicijalizacije; načina prikaza rezultata; maksimalnog broja iteracija itd. Ograničenje nam omogućava da postavimo ograničenje svakom modu pojedinačno poput nenegativnosti, unimodalnosti ili ortogonalnosti moda. Za sve naše modele bilo je korišćeno ograničenje nenegativnosti pošto nijedan od modova (koncentracija, emisija, ekscitacija i vreme života) ne mogu imati negativnu vrednost. Sve opcije su bile ostavljene na standardnim podešavanjima. Broj komponenata za modele meda i kancera dojke su bili pretpostavljeni na osnovu prikupljenih podataka i bilo je izgrađeno više modela sa različitim brojem komponenata, zatim na osnovu validacije (split-half analiza i leave-one-out metoda odeljak) i interpretacije modela je

bilo odlučeno koji model najbolje opisuje posmatrani sistem. Za aminokiseline i kompleksni sistem fosfora na bazi retkih zemalja broj komponenata je bio već poznat jer su oba bila sintetisana kod nas u laboratoriji. U slučaju aminokiselina model je bio validiran na osnovu kalibracije.

Kada je dobijen odgovarajući model za svaki od kompleksnih sistema može se pristupiti njihovoj analizi. Kao rezultat svakog modela dobijamo tri grafika, odnosno za svaki mod po jedan grafik. Za dva fiksna moda dobijamo prikaz fenomena odnosno model koji je jedinstven za taj sistem, dok treći prikazuje vrednosti skorova za svaki uzorak odnosno varijacije među uzorcima tog sistema. Zavisno od sistema koji analiziramo pratimo različite trendove koji se javljaju među uzorcima i čime su oni uslovljeni. Na ovaj način možemo oceniti signifikantnost modela. Kod sistema meda i kancera dojke moguće je ispratiti kako promene koncentracija određenih fluoroforma utiču na pojedine trendove koji se javljaju među uzorcima. Kod aminokiselina i neorganskog sistema se može posmatrati kako se sa promenom intenziteta menja i koncentracija i uz koliku tačnost se može odrediti prisutnost i koncentracija neke fluorofore u nepoznatom uzorku na osnovu njegovog luminescentnog spektra ili emisionog spektra i vremena života, respektivno.

4.3.3 Proračun i analiza PCA modela

Postupak proračuna PCA modela je mnogo manje zahtevan od PARAFAC analize. Kod ovog modela se vrši redukcija dimenzionalnosti podataka pri čemu prvih nekoliko glavnih komponenata sadrže procentualno najviše informacija o varijabilnosti među uzorcima sistema. Cilj ovog modela je da sa što manjim brojem glavnih komponenata opiše procentualno što više varijabilnosti prisutnih u sistemu. Kod spektralnih podataka primenom PCA analize je moguće opisati sve karakteristike sistema sa jako malim brojem glavnih komponenata, nekad i sa samo dve. Prednost ove analize je pojednostavljena vizuelizacija spektralnih podataka, gde se uz pomoć grafika prve dve komponente može analizirati da li postoje neki trendovi među uzorcima i mogu se na jednostavan način interpretirati rezultati. Sam proračun podrazumeva unos centriranih podataka u program i građenje modela. Odabir broja komponenata modela se vrši na osnovu procenta obuhvaćene kumulativne varijabilnosti koju glavne komponente

opisuju, gde se najčešće bira broj komponentata čija je kumulativna varijabilnost preko 90%. Broj komponentata se ne mora uvek birati na osnovu te vrednosti, ali to prevashodno zavisi od podataka koji se modeliraju. Kada se odabere željeni broj glavnih komponentata, njihovi skorovi se mogu koristiti kao parametri sistema za dalju analizu. Dobijeni rezultati se zatim analiziraju uz pomoć grafika rasutosti skorova glavnih komponentata kao i grafika *loadings*-a glavnih komponentata. Na osnovu grafika rasutosti mogu se uočiti grupisanja i međusobni odnosi uzoraka. Analizom grafika *loadings*-a mogu se videti pozicije u spektrima gde se nalaze najveće varijacije među spektrima uzoraka. Te pozicije na spektrima se uočavaju kao pozicije maksimuma i minimuma krive.

4.3.4 Proračun i analiza PLS-DA modela

Za građenje PLS-DA modela u program se unose centrirane matrice sa podacima organizovane na način da svaki red odgovara uzorcima, a kolone parametrima. Ovaj model ne zahteva dimenzionalno redukovane podatke jer on sam vrši redukciju i računa nove latentne varijable na osnovu kojih vrši klasifikaciju uzoraka. U početnim podacima dodatno se moraju definisati i klasne pripadnosti svakog uzorka kako bi model mogao da se obuči za klasifikaciju. Ovaj model vrši klasifikaciju tako što pravi više modela gde svaki od njih razdvaja jednu klasu od svih drugih klasa. Na taj način ukoliko imamo 5 grupa ovaj model će se sastojati od 5 manjih modela i pri validaciji modela biće dobijeno 5 grešaka modela, odnosno dobiće se greška klasifikacije za svaku grupu ponaosob. I kod ovog modela neophodno je odabrati broj latentnih varijabli koje će biti korišćene za klasifikaciju. Uvek se teži tome da taj broj bude što manji kako model ne bi bio predefinisani i nestabilan. Broj varijabli se u ovom slučaju bira na osnovu grešaka klasifikacije modela i grafika praga klasifikacije. U ovoj analizi se razmatraju dve greške modela: kalibraciona i unakrsno validirana. Kalibraciona greška predstavlja grešku koju model pretpostavlja dok unakrsno validirana greška je dobijena realnim testiranjem modela nepoznatim uzorcima. Kod stabilnog i optimalnog modela ove dve greške bi trebale da budu istih ili što sličnijih vrednosti. Sa povećanjem ili smanjenjem optimalnog broja latentnih varijabli ove dve greške će se sve više varirati. U nekim slučajevima ove greške mogu biti iste, ali da sam model bude nestabilan usled

predefinisanosti, tada se proverava stabilnost modela analizom grafika praga klasifikacije. Programski paket kao jedan od rezultata analize daje grafički prikaz ovih grešaka za različite brojeve latentnih varijabli, uz pomoć kojeg se na jednostavniji način može uočiti koji broj najviše odgovara modelu. Grafik praga klasifikacije pokazuje senzitivnost i specifičnost modela u funkciji praga klasifikacije. Sa ovog grafika se može videti stabilnost dobijenog modela za svaku grupu ponaosob. Stabilan PLS-DA model mora imati isti ili sličan procenjen i unakrsno validiran odgovor, što se na ovom grafiku može videti ukoliko se procenjena i validirana kriva poklapaju.

4.3.5 Proračun i analiza SVM modela

Postupak proračuna SVM modela je vrlo sličan prethodno opisanom postupku za PLS-DA model. U sam program se unose podaci u obliku matrice gde važi da su redovi uzorci, a kolone parametri na osnovu kojih se gradi model. Za ovaj proračun podaci ne moraju biti centrirani već se mogu neobrađeni ubaciti u program. Kada se podaci unesu neophodno je definisati klase svakog uzorka u matrici. Nakon toga se podešavaju parametri modela, odnosno tip kernela koji će biti korišćen za obuku modela, tip kompresije podataka ukoliko želimo da redukujemo dimenziju ulaznih podataka i broj setova na koji će podaci biti deljeni radi obuke i testiranja. Za analize u ovom radu bio je korišćen linearan kernel pri čemu podaci nisu bili komprimovani. Ulazni podaci su bili deljeni na dva seta podataka, jedan za obuku i jedan za testiranje dobijenog modela. Kao rezultat na kraju se dobijaju greške klasifikacije modela koje zatim analiziraju.

4.3.6 Proračun zapremine spektralnih domena

Zapremine spektralnih domena (zapremine ispod površine luminescentnog intenziteta) su bile u ovom radu računane numerički. Spektralni domen od interesa (x - y površina $\rightarrow \lambda_{exc}-\Delta\lambda$ domen) predstavlja se mrežom diskretnih tačaka (x_i, y_j) , od kojih svaka ima jedinstvenu vrednost uzetu iz spektralnog domena: $x_i = x_{min} + i\Delta x$; $y_j = y_{min} + j\Delta y$, $\Delta x = (x_{max} - x_{min})/N$, $\Delta y = (y_{max} - y_{min})/M$, gde 'max' i 'min' označavaju gornju i donju

spektralnu granicu, a N i M su faktori diskretizacije. Zatim se zapremine spektralnih domena računaju primenom sledeće jednačine:

$$\int_{y_{\min}}^{y_{\max}} \int_{x_{\min}}^{x_{\max}} I(x, y) dx dy \cong (\Delta x \Delta y) \sum_{j=1}^M \sum_{i=1}^N I(x_i, y_j). \quad (1.81)$$

5. Rezultati i diskusija

Primena PARAFAC analize za modeliranje fluorescentnih EEM-a je doživelo veliku ekspanziju u poslednjih nekoliko godina [Callejón *et al.* 2012, Bro 1999, Christensen *et al.* 2005] pogotovo u analizi prehrambenih proizvoda. Kombinacija ovih metoda je pokazala veliki potencijal za primenu u kvalitativnoj analizi kako prehrambenih tako i bioloških uzoraka na brz i jednostavan način. Uprkos svojim prednostima ona se još uvek nedovoljno primenjuje iz razloga što dobijanje dobrog i pouzdanog modela jednog kompleksnog sistema zahteva veliki broj uzoraka i jako puno proračuna i podešavanja dok se ne odrede optimalni parametri za model određenog sistema. U ovoj tezi su pravljene modeli za četiri različita kompleksna sistema pri čemu se uzimao u obzir i krajnji cilj odnosno praktična primena dobijenog modela. Kod modeliranja prvog sistema, rastvora vode i dve aminokiseline, cilj je bio da se pokaže na jednostavnijem primeru na koji način se spektri modeliraju PARAFAC metodom i šta se dobija kao rezultat jednog takvog proračuna. Takođe je želja bila da se pokaže praktična primena jednog takvog modela i na koji način on funkcioniše. Druga tri sistema koja su modelirana u okviru ove teze se na osnovu naših saznanja iz stručne literature po prvi put modeliraju i iz tog razloga su zahtevala jako puno proračuna kako bi se odredili optimalni uslovi modela. Sistem meda je bio najkompleksniji od sva tri, a njegovo modeliranje je imalo za cilj da se prouči koje fluorofore imaju najveći uticaj na botaničko poreklo i kvalitet meda. S obzirom na kompleksnost i visoku cenu standardnih metoda za analizu meda, postoji velika potražnja za metodom koja je brza, jeftina i ne zahteva specijalnu obuku operatera. Osim modeliranja i analize EEM spektara uzoraka meda, bili su analizirani i njihovi sinhroni fluorescentni spektri metodom glavnih komponenta i PLS-DA metodom, kako bi se uporedili rezultati dva različita pristupa i ustanovile njihove prednosti i mane. Modeliranje zdravog i malignog tkiva dojke je imalo za cilj određivanje tačnog broja fluorofora prisutnih u tkivu i analizu varijacije njihovih koncentracija u malignom i zdravom tkivu. Današnji trendovi u medicini teže razvijanju novih neinvazivnih, brzih i pouzdanih metoda za praćenje promena u humanim tkivima. Ova metoda ima veliki potencijal kao metoda koja ispunjava te uslove i može se koristiti kao „screening“ test za razne vrste bolesti jer svaka patologija u organizmu utiče na promenu fluorescencije kako tkiva tako i telesnih

tečnosti. Osim trodimenzionalnog modela tkiva dojke, takođe je bio ispitivan i drugi pristup ovim podacima. U tom slučaju su bili računati parametri iz sinhronih fluorescentnih spektara i EEM-a koji su zatim korišćeni za obučavanje SVM algoritma. Dobijeni rezultati su nakon toga bili poređeni sa rezultatima SVM algoritma obučavanog na osnovu parametara PARAFAC modela u cilju određivanja ulaznih parametara koji omogućavaju izgradju najuspešnijeg modela. Poslednji sistem koji je jedini bio neorganski je bio modeliran kako bi se iz mešavine tri neorganska fluorescentna materijala odredila prisutnost i tačna koncentracija tih supstanci na osnovu njihovih emisionih spektara i vremena života. Neorganski luminescentni centri imaju veoma karakteristične emisione spektre, međutim neki od njihovih pikova se javljaju na jako bliskim pozicijama i time dolazi do njihovog preklapanja. Iz tog razloga je u takvim situacijama teško odrediti koje supstance i u kojoj količini se nalaze u uzorku. Uz primenu PARAFAC modela ovaj problem se vrlo jednostavno rešava i daje mogućnost karakterizacije datog materijala uz primenu samo jedne metode, na osnovu čijih se spektara može ustanoviti da li je materijal čist ili su prisutne još neke supstance u materijalu. Bitno je naglasiti da je u ovoj tezi po prvi put pri proračunu PARAFAC modela bila korišćena vremenska zavisnost kao treća dimenzija ulazne matrice kompleksnog sistema.

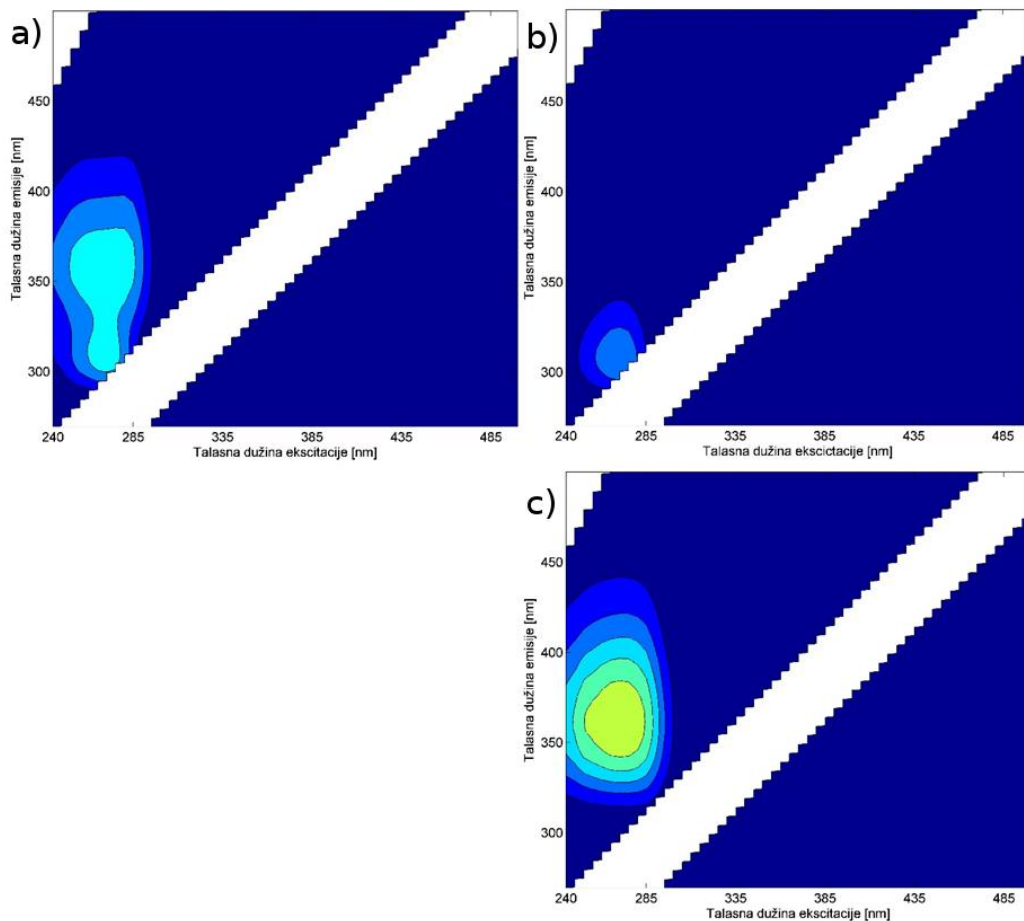
5.1 Rastvor vode i aminokiselina

Kao što je prethodno bilo navedeno za uzorke kompleksnog sistema rastvora vode i aminokiselina bile su izmerene ekscitaciono-emisione matrice. Na slici 5.1 su prikazani konturni grafici izmerenih EEM-a za tri uzorka na kojima je uklonjeno Rejljevo rasejanje prvog i drugog reda koje se na graficima vidi kao bela traka.

Na slici 5.1 a) je prikazan spektar za mešavinu koja sadrži obe aminokiseline dok su pod b) i c) prikazani spektri za čist tirozin i čist triptofan, respektivno. Ono što se primećuje na ovim spektrima jeste da je intenzitet fluorescencije triptofana mnogo jači od fluorescencije tirozina iako je koncentracija triptofana manja od koncentracije tirozina u rastvoru. Ovaj fenomen je već poznat i često se u biološkim uzorcima fluorescencija tirozina ne vidi ili je jako slaba usled mnogo dominantnije fluorescencije

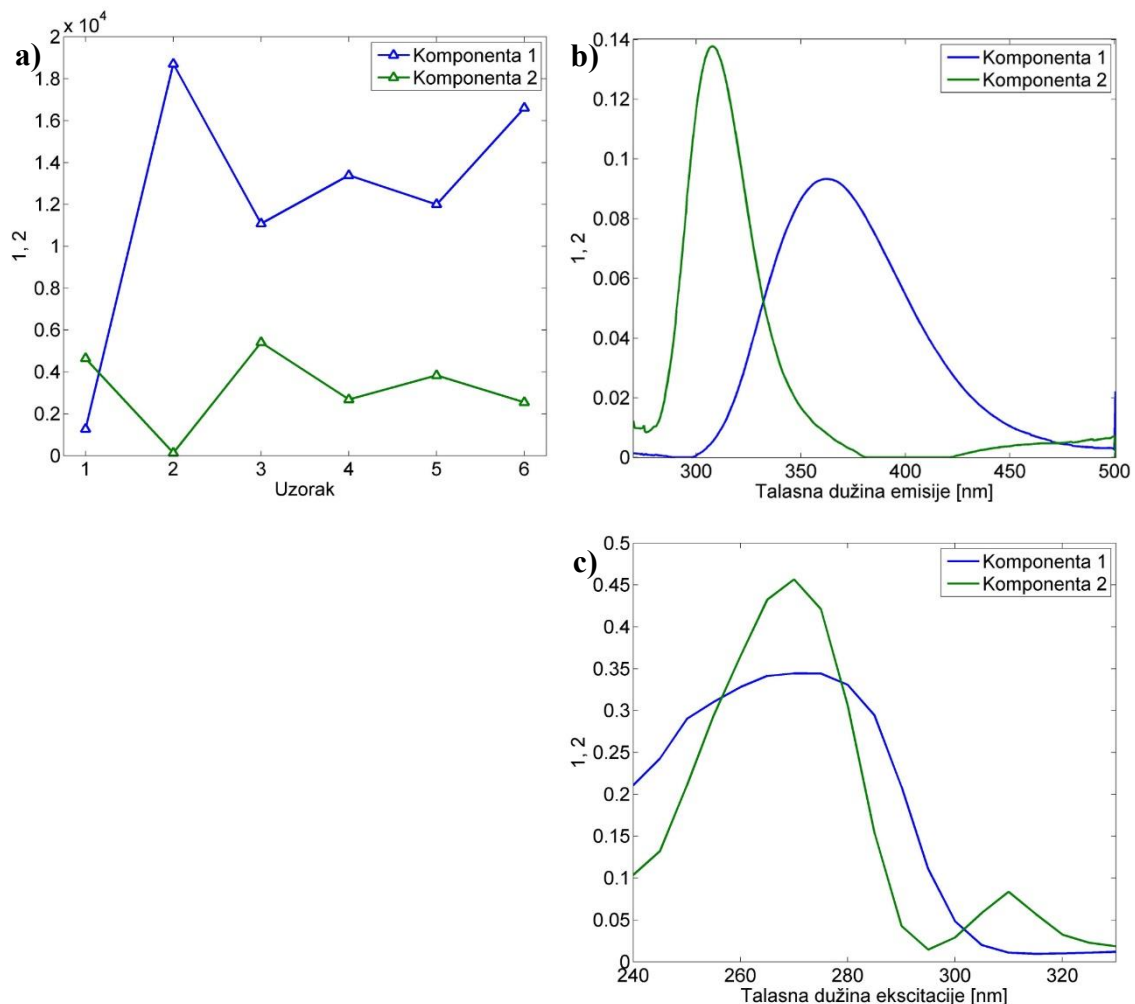
triptofana. Iz tog razloga je u pripremi uzoraka o tome vođeno računa i bile su stavljanje veće koncentracije tirozina kako bi i on došao do izražaja.

Cilj merenja i modeliranja ovog kompleksnog sistema sa samo dve fluorofore je bio da se prikaže koliko pouzdano može da se izmodelira jedan kompleksan sistem i sa kolikom tačnošću mogu da se odrede koncentracije obe fluorofore u svim uzorcima ako se kalibriše koncentracijom u samo jednom uzorku. Problem na koji se često nailazi pri modeliranju realnih bioloških sistema je taj da ne možemo uvek odrediti koncentracije fluorofora u uzorku standardnim hemijskim metodama pa je stoga i kalibracija i potvrda datog modela mnogo teža. Ali kako postoje druge metode validacije modela, ako dobijemo dobar model on nam uvek kao izlaz daje relativne koncentracije fluorofora u uzorcima, pa tako u situacijama gde se koncentracija nekih fluorofora ne može odrediti standardnim metodama ova metoda može da nađe svoju primenu.



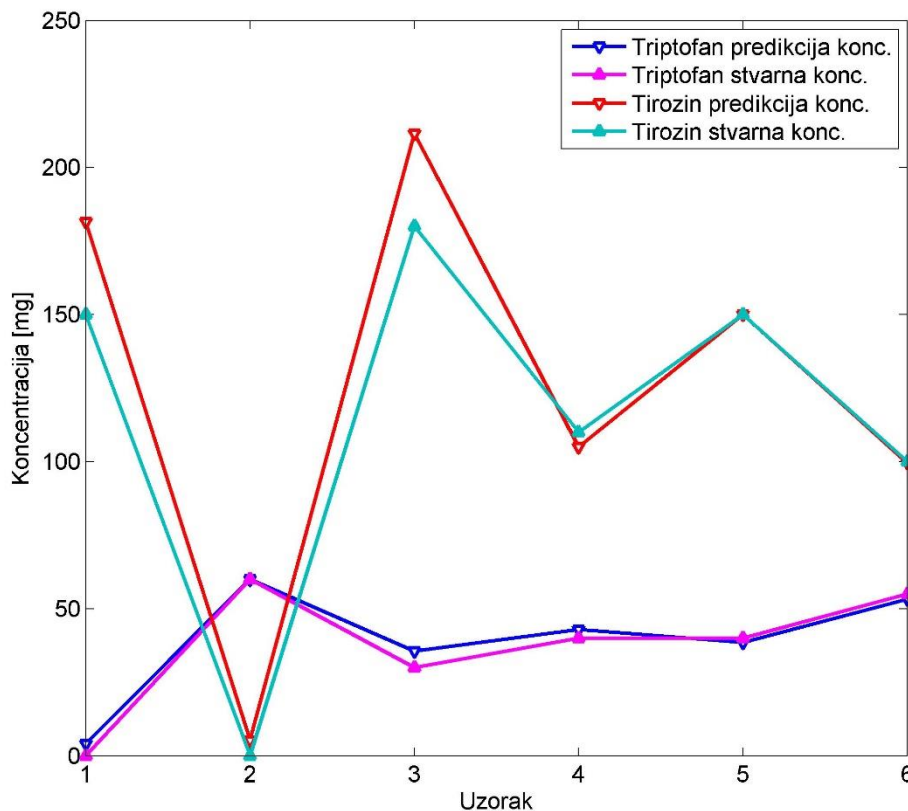
Slika 5.1 – Konturni grafici EEM-a tri uzorka rastvora vode i aminokiselina: a) mešavina sa obe aminokiseline, b) čist tirozin i c) čist triptofan.

Proračun PARAFAC modela za ovaj sistem je bio olakšan time što nam je broj komponenata bio poznat. Na slici 5.2 je prikazan PARAFAC model (sva tri moda) dobijen za ovaj sistem. Grafik prvog moda PARAFAC modela nam prikazuje relativne koncentracije obe aminokiseline za svaki uzorak, odnosno skorove svakog uzorka. Na graficima triptofan odgovara komponenti 1 dok komponenti 2 odgovara tirozin. Grafik drugog i trećeg moda modela prikazuje emisioni i ekscitacioni spektar, respektivno, za svaku komponentu. Ovi spektri su izvučeni kao osnovni fenomen ovog sistema i oni su isti za sve uzorke dok je koncentracija jedina koja varira među uzorcima sistema. Ukoliko bi uzeli skor jednog uzorka i pomnožili sa čistim spektrima u drugom i trećem modu dobili bi početni izmereni spektar.



Slika 5.2 – Prikaz PARAFAC modela na osnovu EEM-a rastvora vode i aminokiselina u: a) prvom, b) drugom i c) trećem modu.

Ono što takođe možemo da vidimo na graficima drugog i trećeg moda jesu emisijski i ekscitacioni maksimumi za obe komponente. Kao što je rečeno ranije, maksimumi zavisno od okoline mogu da budu malo pomereni odnosno da variraju, ali uglavnom se javljaju u jednom malom opsegu talasnih dužina. Ukoliko nam komponente sistema nisu poznate, onda na osnovu pozicije maksimuma pikova možemo da odredimo koje su fluorofore u nekom sistemu prisutne. Ovde vidimo da triptofan i tirozin imaju maksimum emisije na 365 i 310 nm, dok su im ekscitacioni maksimumi na 275 i 270 nm, respektivno, što približno odgovara pozicijama njihovih maksimuma datim u tabeli 3.2.



Slika 5.3 – Stvarne i modelom pretpostavljene koncentracije triptofana i tirozina u svim merenim uzorcima.

Da bi validirali dati model bila je izvršena kalibracija modela, tako što je modelu data poznata koncentracija fluorofora u jednom uzorku i zatim su dobijene pretpostavke modela za koncentracije fluorofora u svim drugim uzorcima na osnovu relativnih koncentracija koje je model izračunao. Na slici 5.3 je dat grafik na kome su prikazane

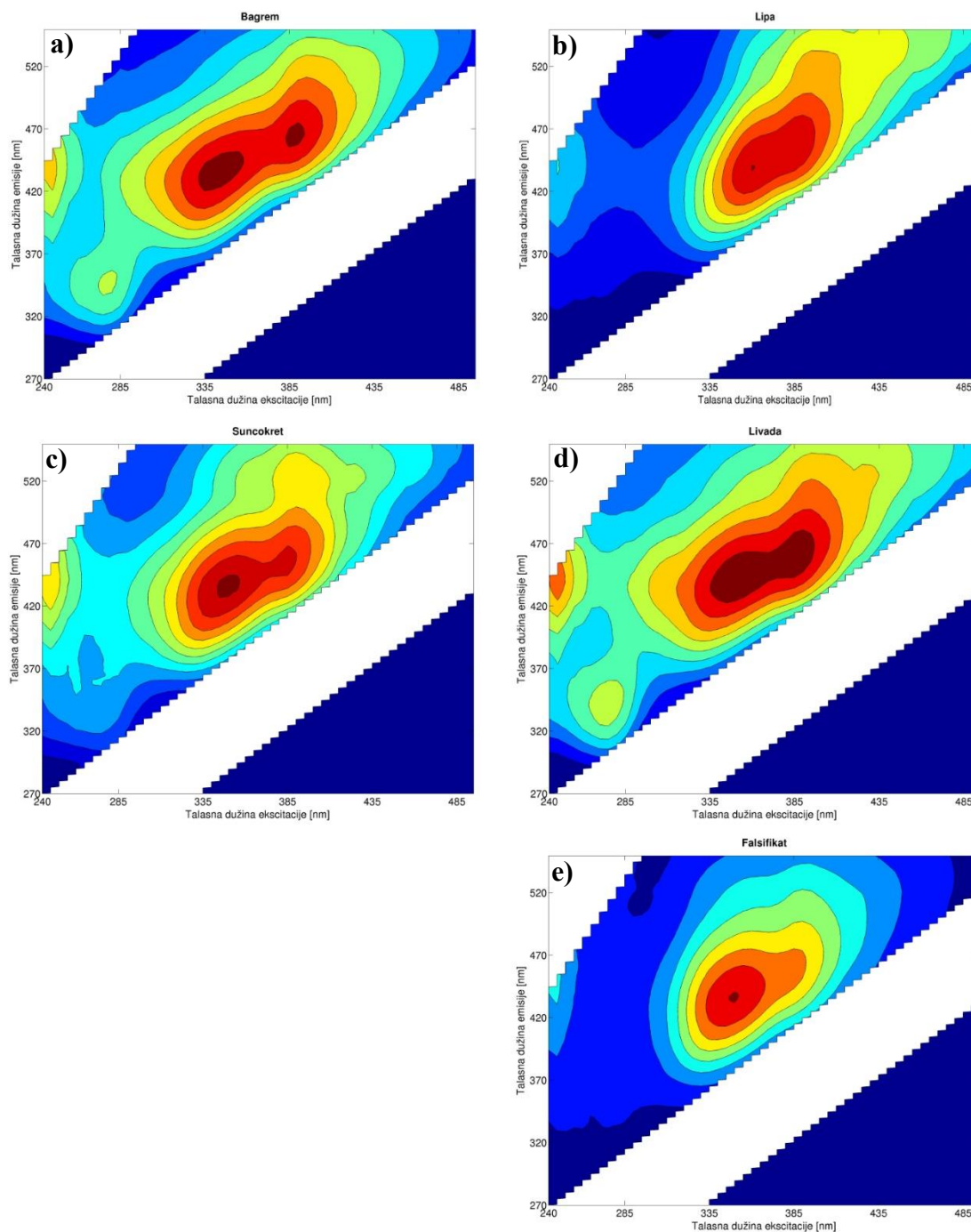
stvarne i modelom predviđene koncentracije obe fluorofore u sistemu. Primećujemo da su odstupanja od stvarne koncentracije relativno mala i da su veća u slučaju tirozina. Naime, odstupanja koncentracije u ovom konkretnom slučaju ne zavise toliko od modela već od preciznosti same pripreme uzoraka. Kako tačnost merenja kada su miligrami u pitanju može malo da varira i pošto su male količine supstanci bile rastvarane u velikoj količini vode to je moglo na kraju da rezultuje malo drugačijim odnosom koncentracija. Najtačnija provera validnosti modela bi bila da su rastvori bili pripremljeni, a zatim je određivana koncentracija aminokiselina u rastvoru uz primenu neke standardne hemijske metode. Međutim, usled nemogućnosti u datom trenutku da se na taj način izvede merenje, odlučeno je da se pokuša sa ovim pristupom. Iako ovaj postupak nije bio idealan, ipak je pokazano da model lepo opisuje posmatrani sistem i da sa visokom preciznošću predviđa koncentracije fluorofora u uzorcima.

5.2 Med

5.2.1 Analiza EEM-a

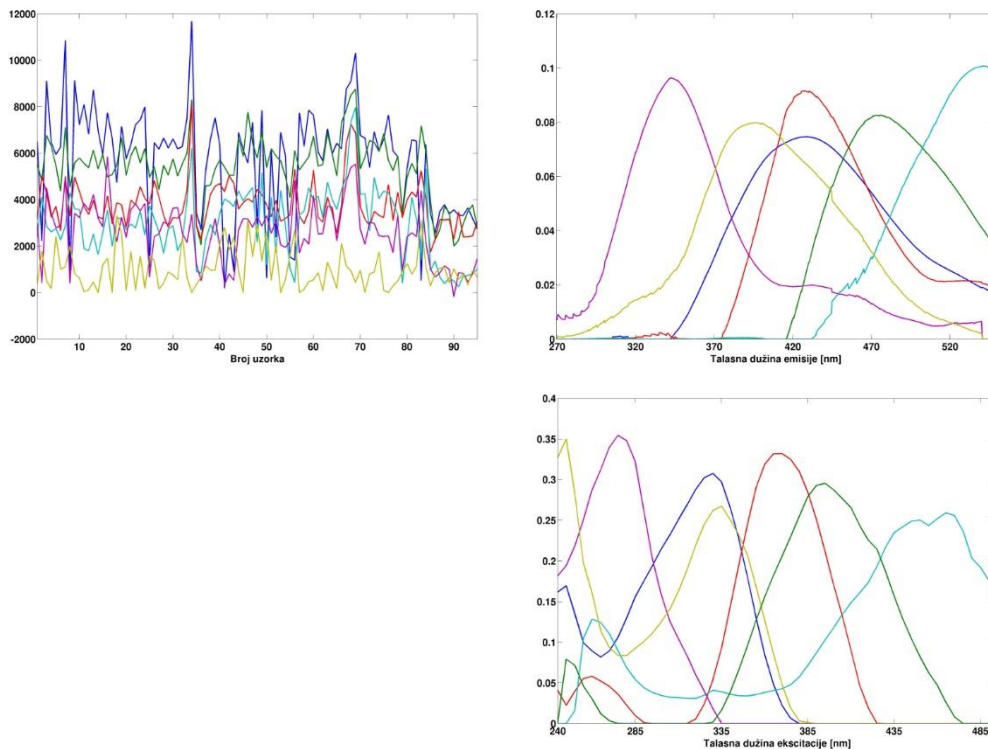
Na slici 5.4 su prikazani konturni grafici izmerenih EEM-a sa uklonjenim Rejljevskim rasejanjem za pet uzoraka koji predstavljaju po jedan tip meda: a) bagrem, b) lipa, c) suncokret, d) livada i e) falsifikat.

Na konturnim graficima prikazanim na slici 5.4 se uočava da se fluorescentne karakteristike uzorka iz različitih grupa botaničkog porekla meda razlikuju. Na primer, livadski, bagremov i suncokretov med imaju veoma izražene pikove u delu spektra od 260-290 nm talasne dužine ekscitacije i između 330 i 360 nm talasne dužine emisije, pri čemu livadski i bagremov med imaju jedan pik izraženiji od suncokretovog meda na emisiji od oko 350 nm i ekscitaciji od oko 270 nm. Za razliku od njih, falsifikat i lipov med u pomenutoj oblasti imaju pikove jako slabog intenziteta. Fluorescencija u toj oblasti potiče od aromatičnih aminokiselina i pretpostavlja se da na osnovu njihovih koncentracija može utvrditi botaničko poreklo meda.



Slika 5.4 – Konturni grafici EEM-a 5 uzoraka meda različitih tipova: a) bagrem, b) lipa, c) suncokret, d) livada i e) falsifikat.

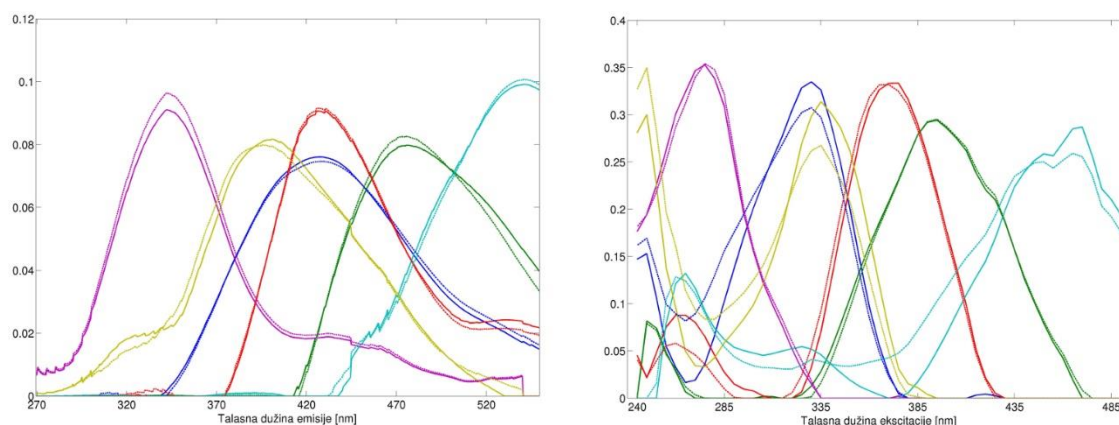
Na osnovu ekscitaciono-emisionih matrica bio je urađen proračun 10 modela (broj komponenta 1-10) i na osnovu *split-half* analize i analize ostataka, model koji je dao najbolje rezultate je model sa 6 komponenta sa ograničenjem nenegativnosti u sva tri moda. Na slici 5.5 su prikazana sva tri moda PARAFAC modela meda.



Slika 5.5 – Prikaz PARAFAC modela na osnovu EEM-a uzoraka meda za: a) prvi, b) drugi i c) treći mod.

Drugi i treći mod prikazanog PARAFAC modela meda nam prikazuju emisione i ekscitacione spektre 6 fluorofora prisutnih u medu. Kao što je ranije navedeno fluorescentni spektri fluorofora su jako osetljivi na okolinu u kojoj se nalaze, stoga pozicija i oblik spektra zavise od: količine vlage koju sadrži određena biljka, viskoznosti i pH vrednosti okoline u kojoj se fluorofora nalazi. Prva komponenta (plava) ima ekscitacioni maksimum na 330 nm i emisioni maksimum na 430 nm. Prema rezultatima Rodrigez-Delgado i saradnika [Rodriguez-Delgado *et al.*] fenolične komponente imaju maksimum ekscitacije u opsegu između 265 i 335 nm i maksimum emisije od 358 do 426 nm. U skladu sa tim prva komponenta bi mogla da odgovara fenoličnim supstancama prisutnim u medu. Druga (zelena) i treća (crvena) komponenta sa ekscitacionim maksimumima na 395 i 370 nm i emisionim na 473 i 426 nm najviše odgovaraju produktima Maillard-ove reakcije poput furozina i hidrosimetilfurfurala (HMF). Kulmyrzaev i Dufour [Kulmyrzaev & Dufour 2008] su pokazali da ove dve fluorofore koje se prirodno nalaze u medu imaju svoj emisioni maksimum na 440 nm za furozin i 425 nm za HMF sa ekscitacijom na oko 360 nm, što odgovara drugoj i trećoj

komponenti. Četvrta komponenta sa ekscitacionim maksimumom na 470 nm i emisionim maksimumom na 540 nm može poticati od fluorescencije riboflavina čije su standardne vrednosti ekscitacionih i emisionih maksimuma date u tabeli 3.2. Pik sa maksimumom ekscitacije na 275 nm i maksimumom emisije na 343 odgovara petoj (ljubičastoj) komponenti, čija pozicija najviše odgovara aromatičnim aminokiselinama. Poslednja šesta (svetlo zelana) komponenta sa ekscitacionim/emisionim maksimumom od 395/473 nm takođe može poticati od fenoličnih supstanci.

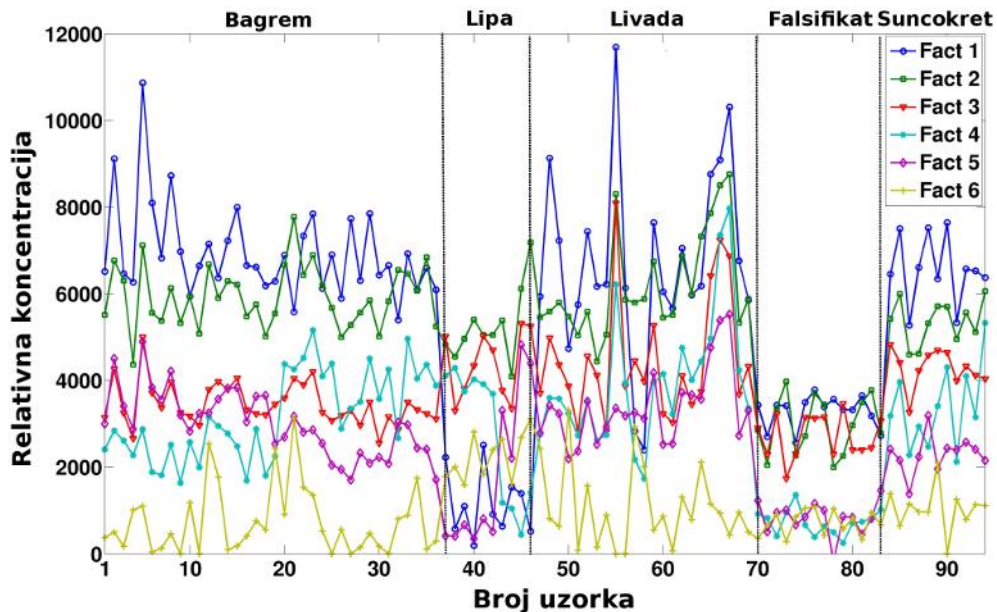


Slika 5.6 – Rezultati *split-half* analize.

Na slici 5.6 je prikazan rezultat *split-half* analize dobijenog modela, gde su uzorci bili podeljeni na dve grupe i na osnovu njih su bila izgrađena dva šestokomponentna modela sa istim početnim uslovima. Na grafiku se vide emisioni i ekscitacioni spektri oba modela i ono što se uočava je da ekscitacioni i emisioni spektri svih komponenata su skoro identični i da su njihova odstupanja jako mala. Na ovaj način nam je potvrđeno da je pravi broj komponenata za model bio odabran i da model precizno opisuje fenomene u sistemu.

U cilju ispitivanja koje fluorofore najviše variraju među različitim tipovima meda bili su analizirani koncentracioni nivoi fluorofora specifični za određene vrste meda. Koncentracioni mod PARAFAC modela je dat na slici 5.7 pri čemu su uzorci raspoređeni na osnovu njihovog botaničkog porekla radi lakše analize i interpretacije. Uočava se da uzorci lipe i falsifikata imaju manje koncentracije prve komponente u odnosu na ostale grupe. Takođe se primećuje da koncentracije druge i treće

komponente, koje odgovaraju produktima Maillard-ove reakcije su manje za uzorke falsifikata i bagrema, respektivno. Kako bi ove razlike koncentracija kvantifikovali i lakše analizirali u tabeli 5.1 su date srednje vrednosti i varijacije koncentracija šest komponenata izračunatih za svaku botaničku grupu ponaosob.



Slika 5.7 – Relativne koncentracije 6 komponenata za sve uzorke meda.

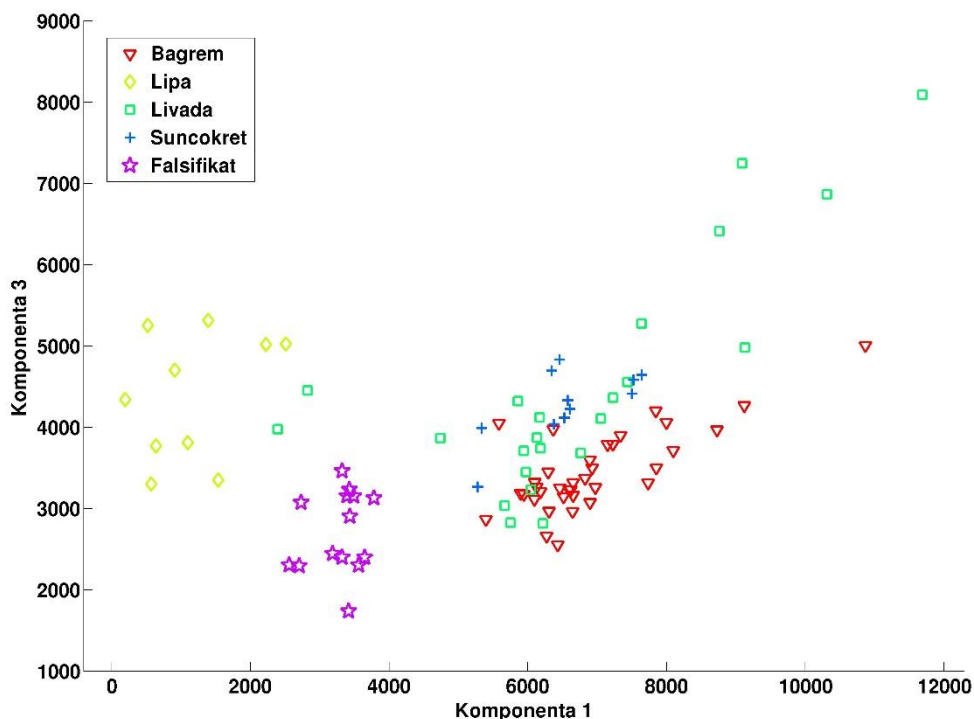
Iz tabele 5.1 može se uočiti sledeće: (a) Najveća razlika između uzoraka bagremovog meda i uzoraka dve botaničke grupe koji su njemu najbliži (suncokret i livada) je u koncentraciji treće komponente; (b) Koncentracioni nivo prve komponente predstavlja najveću razliku između uzoraka lipovog meda i ostalih grupa; (c) Uzorci falsifikata meda se najviše razlikuju od ostalih grupa na osnovu koncentracionih nivoa svih komponenata i imaju najmanje varijacije u koncentracijama unutar same grupe. Na osnovu ovih zaključaka prepostavili smo da će najbolja vizuelna razdvajanja među grupama biti postignuta ukoliko se na grafiku rasutosti prikažu relativne koncentracije prve i treće komponente za svaki uzorak, koje odgovaraju fenoličnoj komponenti i produktu Maillard-ove reakcije (slika 5.8).

Tabela 5.1 Srednja vrednost (SV) i varijacija (Var.) relativnih koncentracija 6 komponenata PARAFAC modela.

		<i>Bagrem</i>	<i>Lipa</i>	<i>Livada</i>	<i>Suncokret</i>	<i>Falsifikat</i>
Komp. 1	SV	6.93	1.15	6.73	6.55	3.28
	Var.	1.14	0.57	4.46	0.62	0.13
Komp. 2	SV	5.93	5.25	6.19	5.36	2.97
	Var.	0.51	0.74	1.41	0.25	0.44
Komp. 3	SV	3.45	4.38	4.47	4.28	2.71
	Var.	0.24	0.60	2.01	0.19	0.25
Komp. 4	SV	3.17	2.78	3.96	3.38	0.72
	Var.	1.02	2.39	2.31	0.94	0.08
Komp. 5	SV	3.00	1.78	3.38	2.30	0.83
	Var.	0.58	3.12	0.77	0.19	0.15
Komp. 6	SV	0.73	2.24	1.06	1.04	0.71
	Var.	0.62	0.30	0.87	0.26	0.08

Na grafiku se uočava da se sve klase osim uzoraka livadskog meda vrlo lepo grupišu. Ovakav trend je i očekivan s obzirom na to da livadski med u sebi sadrži polene više biljaka, dok ostali sadrže u većem procentu polen samo jedne biljne vrste. Stoga je normalno da se uzorci livadskog meda prostiru šire po grafiku dok su ostale grupe bolje separatisane. Ono što se na ovom grafiku može uočiti je da najveća razlika između različitih tipova meda leži u različitim nivoima koncentracija jedne od fenoličnih

komponenti i produktu Maillard-ove reakcije karakterističnim za svaku botaničku vrstu meda.



Slika 5.8 – Grafik rasutosti dobijen na osnovu relativne koncentracije prve i treće komponente PARAFAC modela za svaki uzorak.

Ako malo bolje pogledamo dati grafik uočava se da lipov med ima najmanje količine komponente 1, pa zatim falsifikati imaju malo veću koncentraciju ove fluorofore dok livadski, bagremov i suncokretov med sadrže veoma slične količine ove supstance. Međutim, razlika između ove tri grupe leži u različitim koncentracijama treće komponente gde suncokret i livada očigledno sadrže veću količinu treće fluorofore. Veoma bitan rezultat ove analize je da se falsifikat može vrlo lako detektovati na osnovu koncentracija navedene dve komponente modela čime se može olakšati autentifikacija meda.

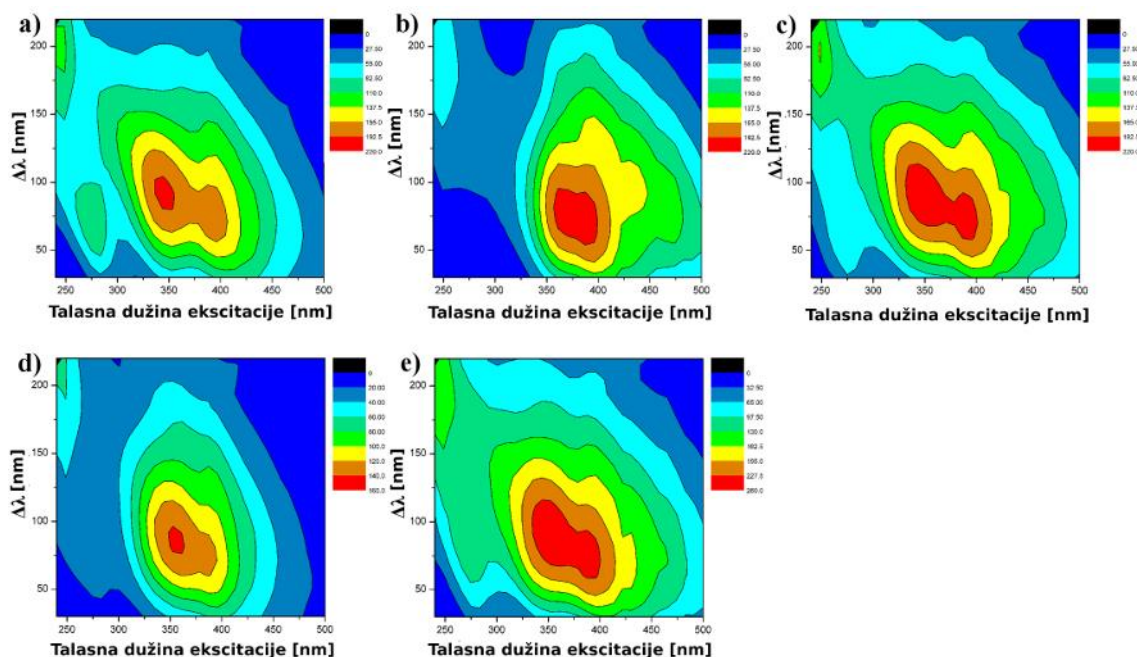
Na osnovu relativnih koncentracija uzoraka meda bila su izgrađena dva PLS-DA modela čije su kalibracione i unakrsno validirane greške klasifikacije date u tabeli 5.2. Prvi model je bio obučavan da otkrije da li je med prirodan ili falsifikat (model za

autentifikaciju). Drugi model je imao zadatak da klasifikuje uzorke na osnovu njihovog botaničkog porekla. Oba modela su bila validirana na osnovu 4 parametara: senzitivnosti, specifičnosti, kalibracione i validacione greške. Najbolji model za autentifikaciju je bio dobijen korišćenjem samo jedne latentne varijable, koji je dao iste vrednosti greške kao i specifičnost i senzitivnost kalibrisanog i validiranog modela (tabela 5.2). Ovim je potvrđeno da je dati model optimalan. On je dao odlične rezultate sa maksimalnom senzitivnošću i specifičnošću i greškom od 0%. Najbolji model za određivanje botaničkog porekla je bio izgrađen sa četiri latentne varijable. Model je pokazao odličnu senzitivnost i specifičnost u slučaju klasifikacije bagrema i lipe, i jako dobre rezultate za livadski i suncokretov med. Iako su greške bile malo veće kod klasifikacije livade i suncokreta, rezultati su bili bolji i greške manje od dobijenih grešaka klasifikacije meda u do sada objavljenim naučnim radovima ((Karoui et al., 2007; Ruoff et al., 2005; Ruoff et al., 2006).

Tabela 5.2 – Kalibrisana (kal.) i unakrsno validirana (CV) senzitivnost, specifičnost i klasifikaciona greške PLS-DA modela dobijenog na osnovu relativnih koncentracija fluorofora u uzorcima meda.

	Model za autentifikaciju		Model za određivanje botaničkog porekla			
	<i>Privodan</i>	<i>Falsifikat</i>	<i>Bagrem</i>	<i>Lipa</i>	<i>Livada</i>	<i>Suncokret</i>
Senzitivnost kal.	1	1	0.935	1	0.733	0.8
Senzitivnost CV	1	1	0.913	1	0.8	0.8
Specifičnost kal.	1	1	0.893	0.989	0.875	0.793
Specifičnost CV	1	1	0.875	0.989	0.833	0.805
Klas. greška kal. [%]	0	0	8.6	0.5	19.5	20.3
Klas. greška CV [%]	0	0	10.5	0.5	18.3	19.7

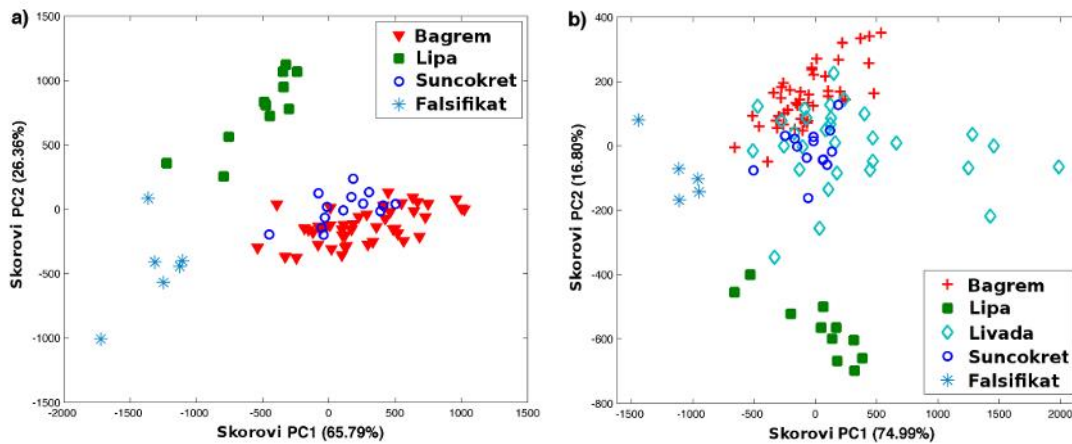
5.2.2 Analiza sinhronih fluorescentnih spektara



Slika 5.9 – Konturni sinhroni fluorescentni spektri uzoraka meda: a) bagrem, b) lipa, c) suncokret, d) falsifikat i e) livada.

Na slici 5.9 su prikazani konturni grafici sinhronih spektara uzoraka meda predstavnika svake od pet botaničkih vrsta ispitivanog meda (bagrem, lipa, suncokret, livada i falsifikat). Emisija na konturnim graficima odražava specifičnost endogenih fluorofora i njihovo mikrokruženje u medu. Glavne razlike između različitih botaničkih grupa meda kod sinhronih spektara se mogu uočiti u nekoliko spektralnih regiona. Prvi region obuhvata opseg talasne dužine ekscitacije od 240-270 nm i sinhronog intervala od 150-220 nm. Fluorescencija u ovom regionu se može javiti usled prisustva fenoličnih komponenata u medu. Ovi molekuli kao što je ranije bilo navedeno imaju maksimum ekscitacije i emisije između 265 i 335 nm i emisije od 358 do 426, respektivno. Emisija aromatičnih aminokiselina se može uočiti u spektralnom regionu koji odgovara talasnoj dužini ekscitacije od 250 do 300 nm i sinhronom intervalu od 40 do 100 nm. Pokazano je da emisijski spektar koji potiče od aromatičnih aminokiselina se može koristiti kao „fingerprint“ za određivanje botaničkog porekla uzorka meda [Karoui et al. 2007a]. Uočava se da emisija triptofana kod uzoraka lipovog meda ima manje intenzivnu emisiju od drugih botaničkih grupa. Treći spektralni region ima raspon talasne dužine

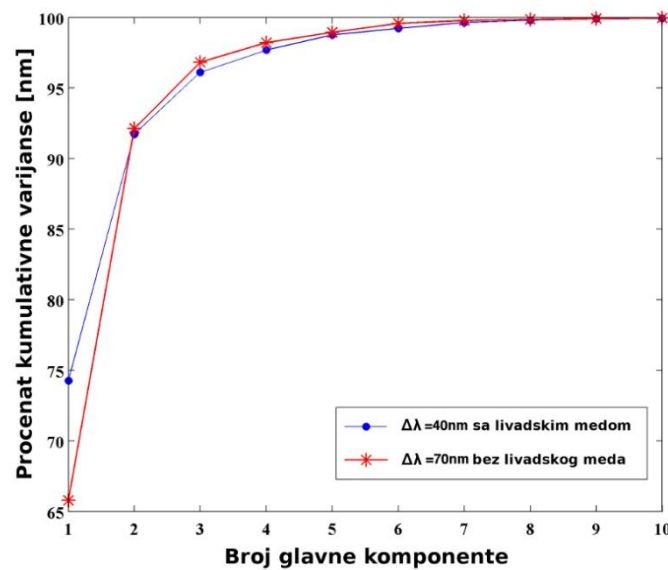
ekscitacije od 340-410 nm i sinhroni interval od 40-120 nm. Ovaj region može odgovarati fluorescenciji produkata Maillard-ove reakcije poput furozina i hidroksimetilfurfurala (HMF) ili neke druge fenolične komponente koja je prisutna u medu, čiji su specifični ekscitacioni i emisijski maksimumi bili ranije navedeni u tekstu.



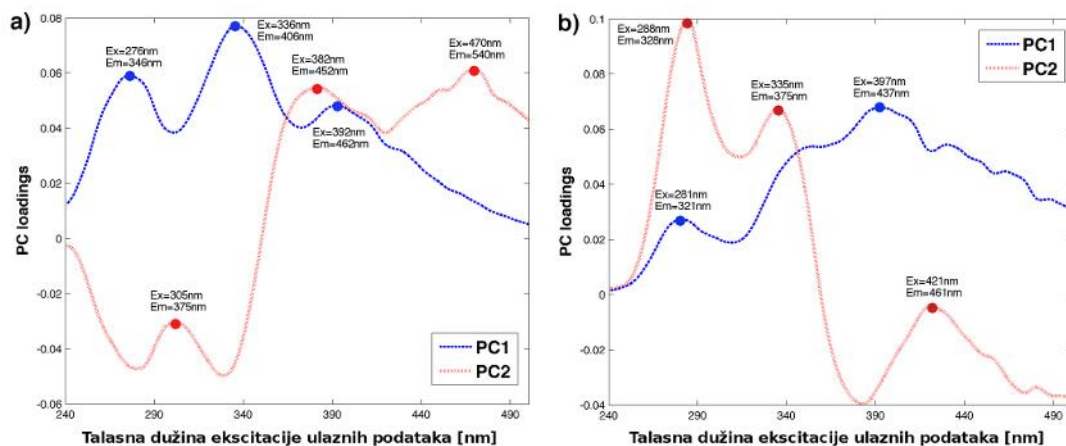
Slika 5.10 – Grafici rasutosti skorova glavnih komponentata PCA analize podataka sinhronih spektara za: a) sinhroni interval $\Delta\lambda = 70$ nm i b) sinhroni interval $\Delta\lambda = 40$ nm.

Na osnovu sinhronih spektara bili su izračunati bilinearni PCA modeli u cilju ispitivanja varijacija između različitih botaničkih grupa meda. Ovom metodom se ispitivalo da li grupe pokazuju različite spektralne karakteristike i bila je omogućena vizuelizacija podataka. Evaluacija grafika rasutosti skorova PCA komponentata je ukazala na to da kod spektara merenih sa manjim sinhronim intervalom dolazi do preklapanja nekih grupa. Najbolja vizuelna separacija među monofloralnim grupama meda (vrste meda kod kojih je udeo polena jedne vrste biljke u medu veći od 45%) je bila postignuta za vrednost sinhronog intervala $\Delta\lambda = 70$ nm (slika 5.10 a), dok je zajedno sa polifloralnom grupom najbolje razdvajanje bilo postignuto za vrednost $\Delta\lambda = 40$ nm (slika 5.10 b). Na graficima se uočava delimično preklapanje livadskog meda i bagrema i suncokreta, kao i preklapanje bagrema i suncokreta, na grafiku monofloralnih vrsta meda. Ova pojava se može objasniti time što su na graficima prikazane samo prve dve izračunate glavne komponente PCA modela, dok ostali PCA skorovi sadrže još dodatnih informacija koje mogu doprineti boljem razdvajanju grupa. Na slici 5.11 je prikazan grafik varijanse PCA modela koji opisuje koliko procenta varijacija opisuje prvih deset komponentata

PCA modela izračunatih na osnovu spektara za $\Delta\lambda = 70$ nm i $\Delta\lambda = 40$ nm. U oba slučaja prve dve komponente opisuju više od 90% varijacija u spektrima, dok prve tri opisuju već više od 95%, što ukazuje na visoko efikasnu redukciju dimenzionalnosti podataka uz primenu PCA analize. Na oba grafika je takođe evidentno razdvajanje uzoraka falsifikata od prirodnog meda. Sva ova zapažanja nam ukazuju na mogućnost izgradnje dobrog klasifikacionog modela korišćenjem sinhronih spektara.



Slika 5.11 – Grafik varijanse za prvih deset glavnih komponenta dobijenih PCA analizom sinhronih spektara za sinhronne intervale $\Delta\lambda = 70$ nm i $\Delta\lambda = 40$ nm.



Slika 5.12 – “Loadings”-i glavnih komponenta PCA analize za spektre sa: a) sinhronim intervalom $\Delta\lambda = 70$ nm i b) sinhronim intervalom $\Delta\lambda = 40$ nm.

Na graficima na slici 5.12 su prikazani “loadings”-i PCA model za prve dve glavne komponente dobijene na osnovu podataka za $\Delta\lambda = 70$ nm i $\Delta\lambda = 40$ nm. PCA “loadings”-i nam pokazuju koliki uticaj ima svaki parametar na datu komponentu, odnosno u ovom slučaju koliki doprinos ima spektar za svaku talasnu dužinu ekscitacije. Na ovaj način se može videti na kojim talasnim dužinama ekscitacije se javljaju najveće varijacije spektara kod uzoraka. U ovom slučaju maksimumi (pikovi) i minimumi kriva na grafiku (PC2 je ortogonalno na PC1) se nalaze na talasnim dužinama ekscitacije koje odgovaraju pozicijama pikova sinhronih spektara prikazanih na slici 5.9. Uočava se da za opseg talasne dužine ekscitacije od 270 – 290 nm i 330 – 340 nm vrednosti “loading”-a su najveće, odnosno na toj poziciji su prisutne najveće varijacije, što nam ukazuje na to da fluorescencija triptofana i fenoličnih komponenti predstavlja najveću razliku između uzoraka meda različitog botaničkog porekla.

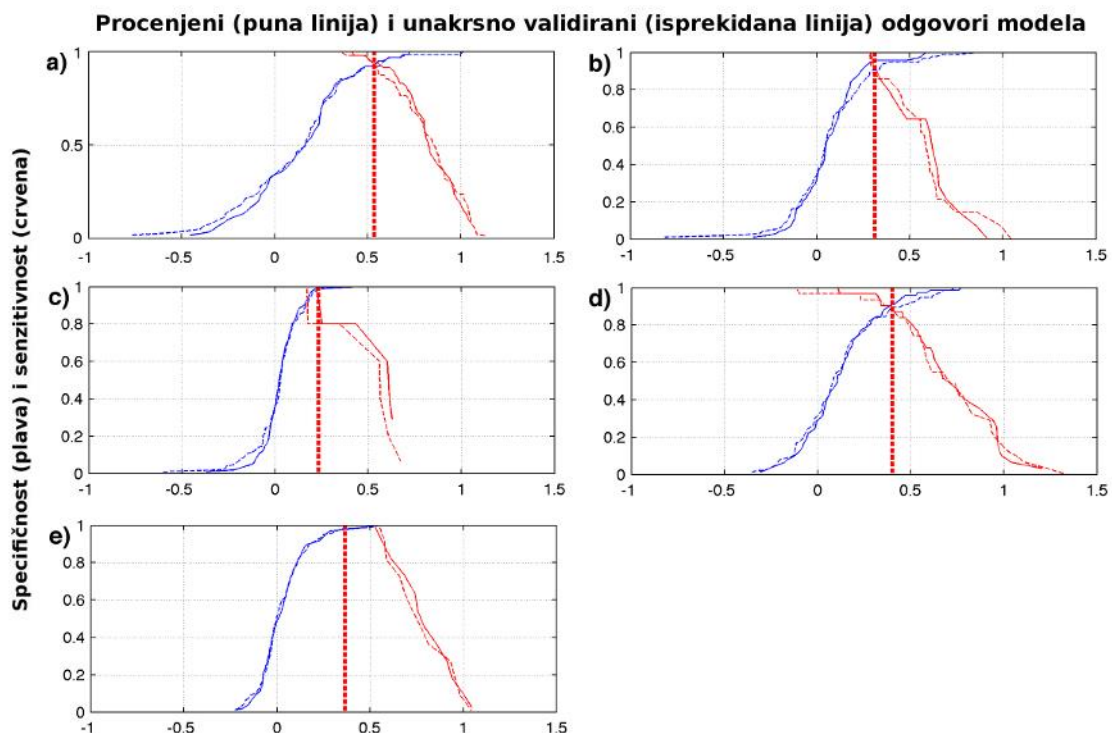
Na osnovu sinhronih spektara bili su izgrađeni PLS-DA klasifikacioni modeli za dva slučaja: u prvom su korišćeni svi uzorci, dok su u drugom bili uzeti u obzir samo uzorci monofloralnih vrsta i falsifikati. 55 PLS-DA modela je bilo izgrađeno kako bi se na osnovu greške klasifikacije utvrdilo za koji sinhroni interval je greška najmanja odnosno na osnovu kog spektra se postižu najbolja razdvajanja.

Tabela 5.3 – Kalibrisana (Cal.) i unakrsno validirana (CV) greška klasifikacije PLS-DA modela izgrađenog na osnovu spektara za $\Delta\lambda = 160$ i 215 nm.

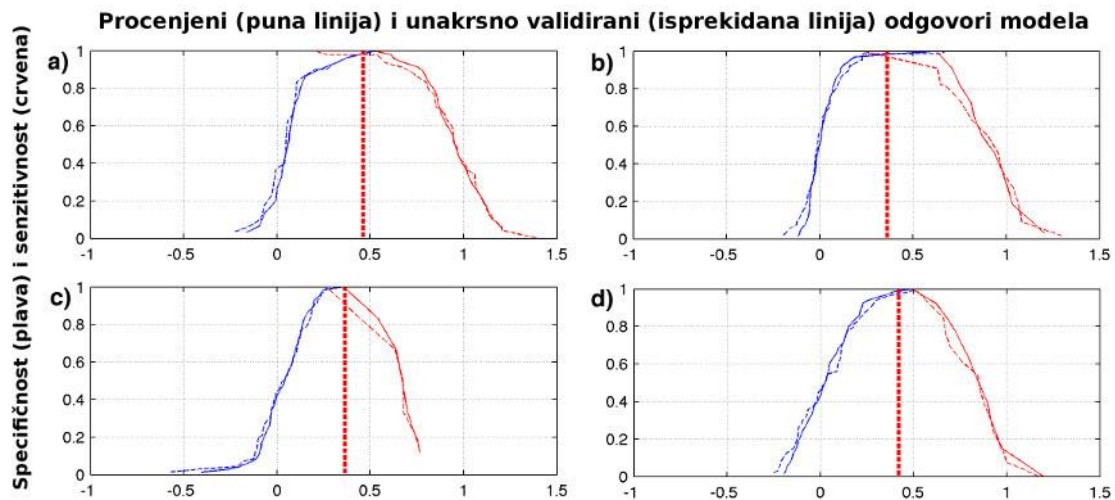
Klas. greška	Uzorci meda uključujući livadski med za $\Delta\lambda = 160$ nm		Uzorci meda monofloralnih grupa za $\Delta\lambda = 215$ nm	
	Kal. [%]	CV [%]	Kal. [%]	CV [%]
Bagrem	5.6	5.6	0	0
Falsifikat	0.9	2.4	0	0
Lipa	1.02	1.53	0.7	0.7
Suncokret	2.1	2.6	0	0
Livadski	11.7	11.7	-	-

U slučaju modela koji je izgrađen na osnovu svih uzoraka, najbolji model je bio dobijen na osnovu spektara za sinhroni interval $\Delta\lambda = 160$ nm. Kao što je bilo i očekivano model je imao najveću grešku pri klasifikaciji livadskog meda, 11.7% (tabela 5.3). Ova pojava se smatra normalnom pošto je pomenuti med sačinjen od polena više vrsta biljaka i

često se u našem geografskom regionu može naći polen bagrema i suncokreta. Stoga se na taj način može objasniti zašto se neki uzorci livade klasifikuju kao bagrem i suncokret i obrnuto. Uzimajući ovo u obzir može se reći da je ova greška relativno mala za livadski med. Iako su uzorci livadskog meda uticali na uspešnost modela te greške su bile mnogo manje od grešaka klasifikacionih modela dobijenih na osnovu pojedinačnih ekscitacionih i emisionih spektara [Ruoff et al. 2005, Ruoff et al. 2006]. Na slici 5.13 je prikazan grafik praga klasifikacije (eng. treshold plot) ovog modela koji vrši klasifikaciju na osnovu osam komponenata. Puna i isprekidana linija predstavljaju procenjen i unakrsno validiran odgovor modela, respektivno. Vertikalna linija pokazuje vrednost praga klasifikacije za koju se postiže najbolja senzitivnost i specifičnost modela, koju PLS-DA algoritam sam odabira. Sa ovog grafika je evidentno da su vrednosti i specifičnosti i senzitivnosti jako velike (veće od 0.9) za sve klase, što ukazuje na to da je model uspešan i da ima malu grešku klasifikacije.



Slika 5.13 – Grafik praga klasifikacije PLS-DA modela za spektre sa sinhronim intervalom $\Delta\lambda = 160$ nm za: a) bagrem, b) suncokret, c) falsifikat, d) livadski i e) lipu.



Slika 5.14 – Grafik praga klasifikacije PLS-DA modela za spektre sa sinhronim intervalom $\Delta\lambda = 215$ nm za: a) bagrem, b) lipu, c) falsifikat i d) suncokret.

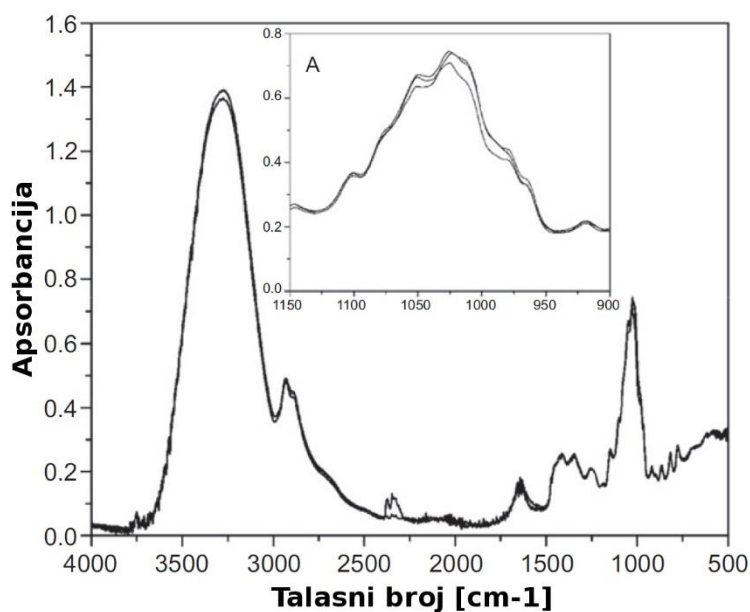
U drugom slučaju model na osnovu sinhronih spektara za $\Delta\lambda = 215$ nm se pokazao najuspešnijim. Ovaj model je vršio klasifikaciju na osnovu sedam komponenata sa tačnošću od 99.825% i izuzetnim grafikom praga klasifikacije (slika 5.14). Sa grafika praga klasifikacije može se videti da su procenjen i validiran odgovor modela skoro identični, što ukazuje na visoku stabilnost i kvalitet modela. Takođe bagremov i suncokretov med imaju skoro idealne odgovore u modelu sa visokom senzitivnošću i specifičnošću (preko 0.99) kod oba dela zbog njihovog vrlo jedinstvenog sinhronog spektra. Kao što je bilo i očekivano greške za ovaj model su bile manje nego u prvom slučaju.

PLS-DA modeli građeni na osnovu PARAFAC parametara i sinhronih spektara su pokazali odlične rezultate za klasifikaciju botaničkog porekla meda. Utvrđivanje autentičnosti meda je bilo 100% uspešno u oba slučajaja. Pokazano je da su se modeli na bazi sinhronih spektara malo bolje pokazali u klasifikaciji botaničkog porekla meda kod svih klasa osim kod lipovog gde je greška bila 1.53% dok je za drugi model bila 0.5%. Prednost modela baziranog na sinhronim podacima leži u tome što je potrebno izmeriti samo jedan spektar kako bi se klasifikovao uzorak, dok je za drugi model neophodno izmeriti celu EEM. Sinhroni spektri za razliku od emisionih mere drugi pravac spektra i na taj način u jednom snimanju obuhvataju istovremeno više bitnih fluorescentnih karakteristike za koje bi inače bilo neophodno meriti više emisionih spektara. Nasuprot

tome EEM meda sadrži sve informacije o sistemu i pokazala se korisnom za karakterizaciju uzoraka meda uz primenu PARAFAC metode.

5.2.3 Analiza infracrvenih spektara

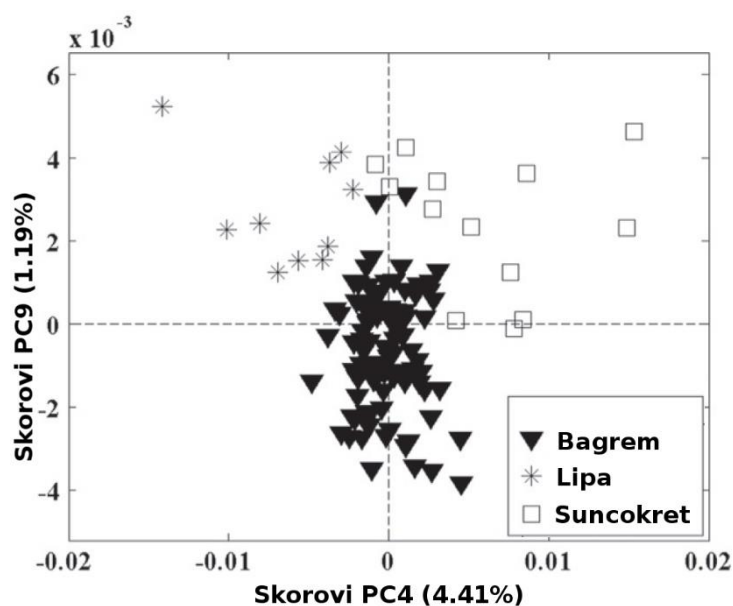
U cilju potvrde uspešnosti dobijenih rezultata uzorci meda su bili ispitivani i infracrvenom spektroskopijom. Infracrvena spektroskopija Furijeovom transformacijom (eng. Fourier transform infrared spectroscopy – FTIR) predstavlja metodu koja se poslednjih godina često koristi kao standardna metoda za ispitivanje hrane. Iz tog razloga su bili mereni i infracrveni spektri meda uz primenu tehnike prigušene totalne refleksije (eng. attenuated total reflectance – ATR)(Lenhardt et al. 2014) radi poređenja uspešnosti pomenutih metoda za klasifikaciju meda na osnovu botaničkog porekla. Za merenje je bio korišćen Thermo Nicolet 380 FTIR spektrofotometar sa dodatkom za ATR.



Slika 5.15 – FTIR-ATR spektri uzoraka različitih vrsta meda (A: uvećanje regiona između 900 i 1150 cm^{-1}).

Na slici 5.15 su prikazani FTIR-ATR spektri uzoraka meda različitog botaničkog porekla (bagrem, suncokret i lipa). Najveće razlike među spektrima se uočavaju u spektralnom regionu između 900 i 1050 cm^{-1} . Pokazano je da apsorpcija u tom regionu

potiče od monosaharida prisutnih u medu poput fruktoze i glukoze, i disaharida poput sukroze [Persano Oddo & Piro 2004]. Pikovi u spektralnom regionu između 1470 i 1150 cm^{-1} se javljaju zbog savijajućeg moda C-C-H, C-O-H i O-C-H grupa. Na poziciji oko 1060 – 1020 cm^{-1} javlja se pik kao rezultat vibracija O-H grupa, dok je pik između 3300 i 2820 cm^{-1} prisutan zbog vibracija istezanja O-H- i C-H- grupa koje potiču od šećera prisutnih u medu [Pataca et al 2007].



Slika 5.16 – Grafik rasutosti skorova PCA analize za različite vrste meda.

Za analizu spektara bio je korišćen opseg spektra između 3718 i 631 cm^{-1} kako bi se izbacio šum iz spektara. Prvo je dimenzija spektralnih podataka bila redukovana primenom PCA analize na 22 komponente. Grafik rasutosti PCA modela (prikazana četvrta i deveta glavna komponenta) je dat na slici 5.16. Na grafiku se zapaža da se uzorci istog botaničkog porekla grupišu i da svaka klasa ima jedinstvene spektralne karakteristike. Na osnovu toga može se pristupiti građenju klasifikacionog modela. U ovom slučaju bila je primenjena SVM diskriminantna analiza koja je za građenje modela koristila prethodno izračunate 22 glavne komponente. Nakon proračuna i unakrsne validacije modela bile su dobijene kalibracione i validacione greške modela (tabela 5.4). Za unakrsnu validaciju korišćena je metoda podele podataka gde su podaci bili podeljeni na dva seta, jedan za obuku SVM-a i jedan za testiranje modela. Iz tabele se vidi da model ima uspešnost od 98.6%, odnosno grešku klasifikacije od 1.4%.

Tabela 5.4 – Kalibrisana i unakrsno validirana greška SVM modela

Klasifikaciona greška	<i>Bagrem</i>	<i>Lipa</i>	<i>Suncokret</i>
Kalibraciona [%]	0	0	0
Unakrsno validirana [%]	0.7	2.4	1.1

Rezultati klasifikacije na osnovu infracrvenih spektara su pokazali uspešnost sličnu ili malo bolju od modela dobijenih na osnovu fluorescentnih spektara. Prednost fluorescentne spektroskopije u odnosu na infracrvena spektroskopska merenja je u količini informacija koju je moguće ekstrahovati iz dobijenih fluorescentnih spektara. Time je moguće jednom metodom, na osnovu samo jednog merenja, kompletno okarakterisati jedan uzorak kompleksnog sistema.

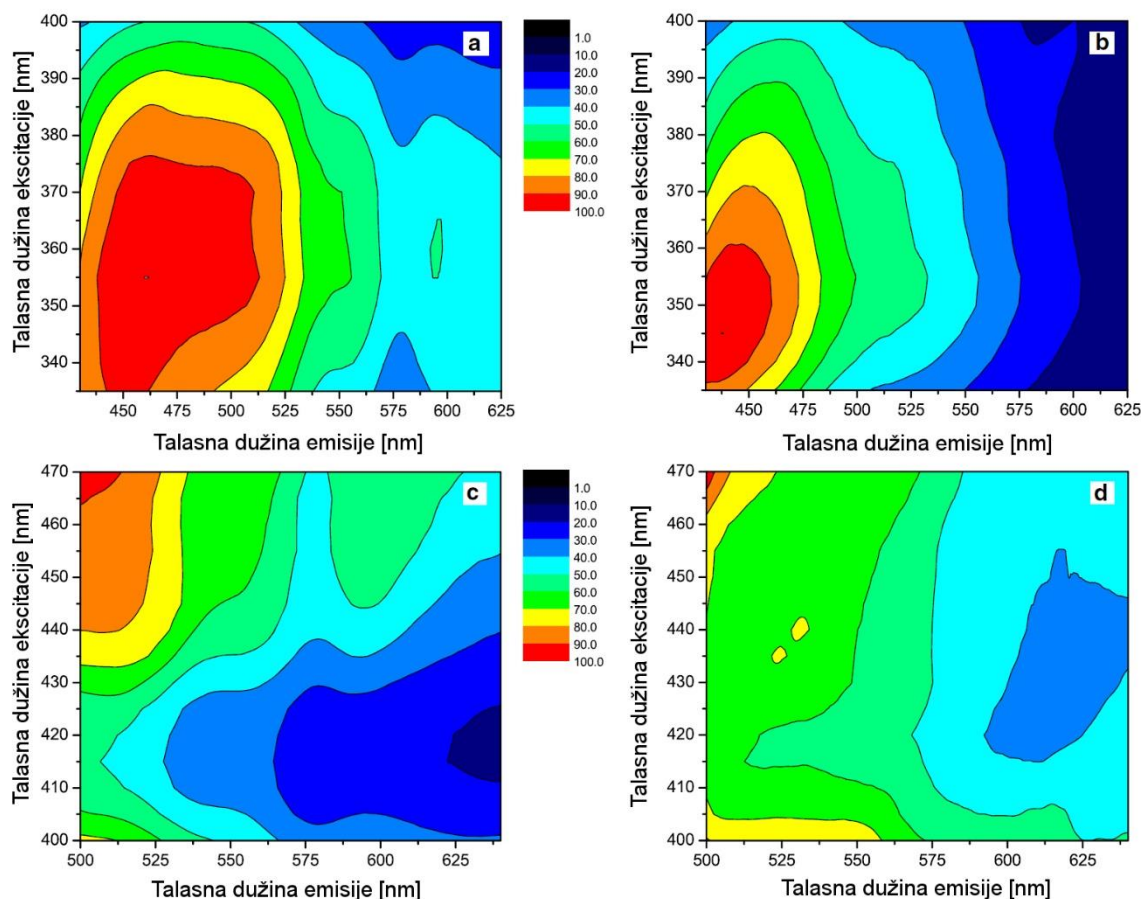
Primena fluorescentne spektroskopije zajedno sa odgovarajućim multivarijantnim metodama bi u mnogome olakšala analiziranje i karakterizaciju uzoraka meda. Standardne analize meda koje se danas koriste su jako komplikovane i zahtevaju puno vremena i novca da bi se izvele na odgovarajući način. Ukoliko bi se predstavljen PARAFAC model meda kalibrisao stvarnim koncentracijama prisutnih fluorofora u medu, omogućila bi se daleko lakša i jednostavnija analiza meda. Zbog veličine molekula šećera u medu, jako je teško odrediti standardnim hemijskim metodama (na pr. tačna hromatografija) tačne koncentracije malih molekula kao što su triptofan i druge fluorofore, kojih u medu ima relativno malo, ali čija je fluorescencija dominantna, i na osnovu koje se može vrlo jednostavno proveravati kvalitet i vrsta meda.

5.3 Kancer dojke

5.3.1 PARAFAC analiza EEM-a

Na slici 5.17 su prikazane izmerene ekscitaciono-emisione matrice dva regiona uzoraka tkiva dojke, pri čemu su oba merena na uzorku sa malignim promenama i na zdravom tkivu dojke. Drugi mereni region nije bio korišćen za PARAFAC analizu već samo za obučavanje SVM modela o kojem će kasnije biti reči. U drugom regionu mogu se uočiti

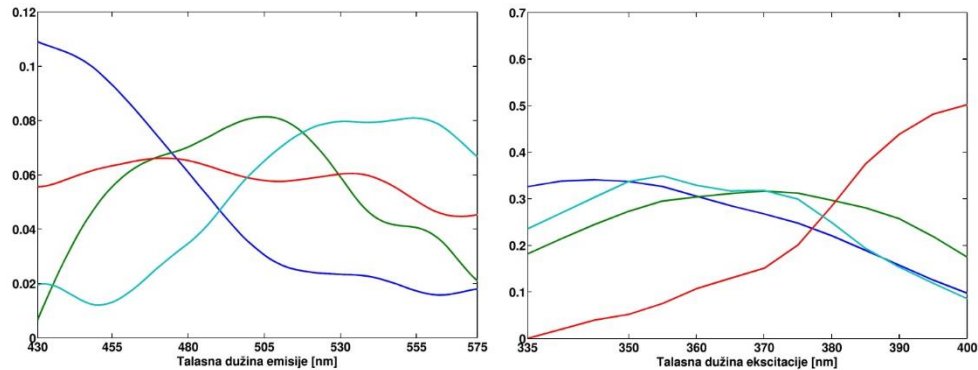
samo krajevi pikova pa nije bilo moguće izgraditi model na osnovu tako dobijenih spektara, ali je njihov doprinos korišćen za klasifikacioni model.



Slika 5.17 – Konturni grafici EEM-a prvog i drugog merenog regiona zdravog tkiva dojke (a) i (c) i kancera dojke (b) i (d), respektivno.

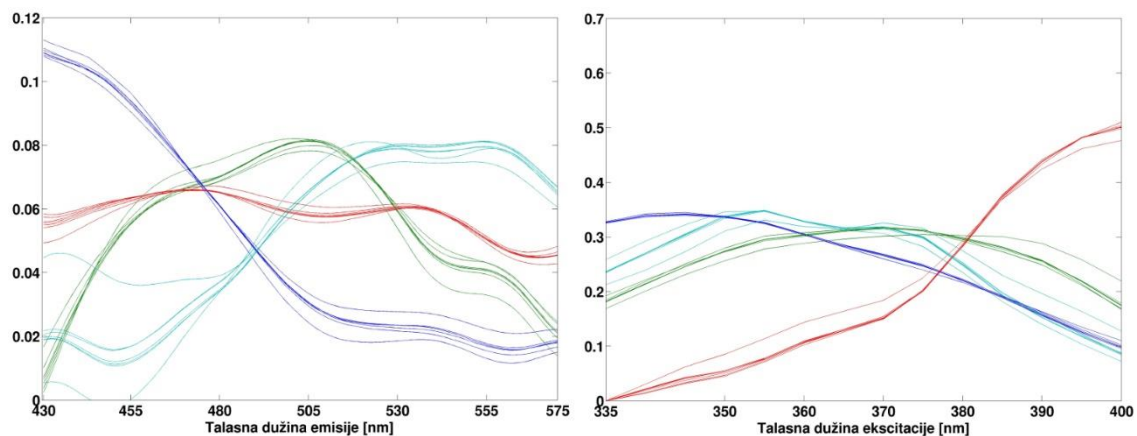
Na konturnim graficima prvog regiona se uočava jedan široki pik, koji predstavlja rezultat doprinosa više fluorofora koje se nalaze u tkivu dojke. Može se primetiti da je kod malignog tkiva taj pik širi što ukazuje na veću koncentraciju jedne ili više fluorofora koje se nalaze u delu spektra sa talasnom dužinom ekscitacije u opsegu od 370 do 395 nm i emisijom od 430 do 480 nm. Kako bi se dalje lakše analizirale te razlike bili su proračunati modeli sa opsegom komponenti od 1 do 6, pri čemu je posle analize ostataka i validacije pokazano da četvorokomponentni model najbolje opisuje ovaj sistem. Za proračun PARAFAC modela korišćen je samo deo spektra prvog regiona u opsegu ekscitacije od 430 do 575 nm i emisije od 335 do 400 nm. U delu spektra koji

se nije koristio nalazio se artefakt u obliku linije koji nije posledica fluorescentnih osobina uzorka tkiva i svojim prisustvom u spektru bi pravio zabunu u modelu.



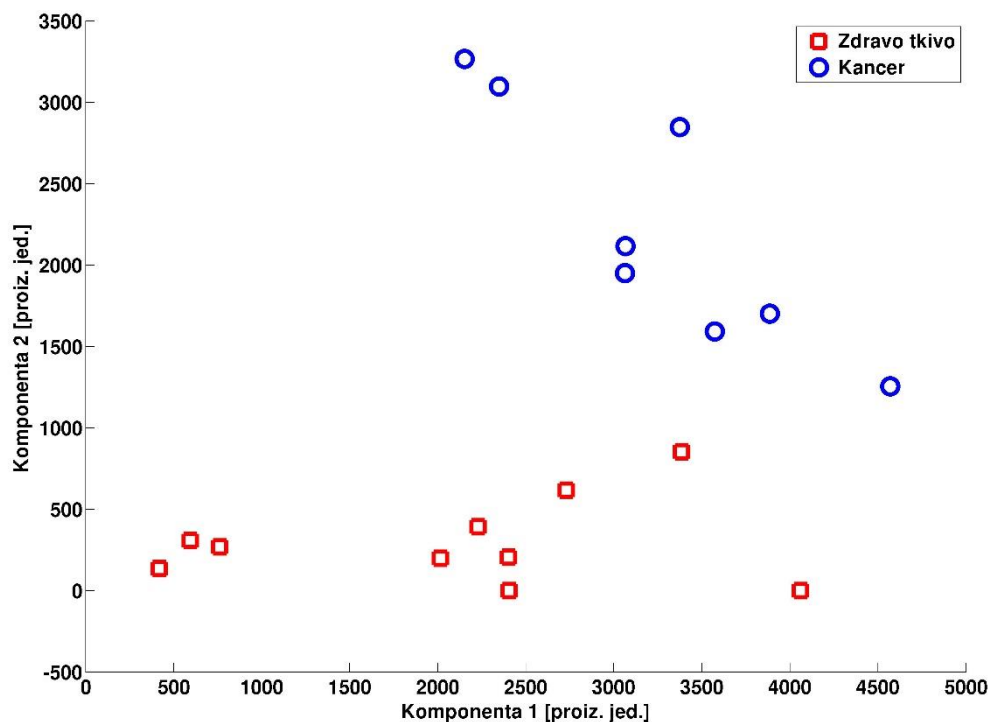
Slika 5.18 – Drugi i treći mod četvorokomponentnog PARAFAC modela tkiva dojke.

Na slici 5.18 su prikazani drugi i treći mod datog PARAFAC modela sa četiri komponente. Prva komponenta označena plavom bojom na grafiku ima ekscitacioni maksimum na 345 nm i emisioni na 430 nm. Ova komponenta bi mogla da odgovara kolagenu i elastinu pošto je njihov ekscitacioni/emisioni maksimum na 325/405 nm. Druga komponenta (zelena) i četvrta (azurno plava) sa ekscitacionim maksimumom na 370 i 355 nm i emisionim na 505 i 554 nm, respektivno odgovaraju oblasti u kojoj fluoresciraju lipidi. Trećoj komponenti sa emisijom na 472 nm i ekscitacijom na 400 nm bi najverovatnije mogao da odgovara NADH. Pošto je set uzoraka jako mali u ovom slučaju validacija split-half analizom ne bi dala realne rezultate, pa je iz tog razloga kao validacija primenjena *leave-one-out* metoda. Kod ove validacije bilo je izgrađeno 18 modela pri čemu je svaki model građen tako što je bio izostavljan po jedan uzorak. Na kraju su drugi i treći modovi izračunatih modela bili svi zajedno nacrtani na graficima kako bi se videle varijacije među modelima. Dobijena dva grafika za drugi i treći mod su prikazani na slici 5.19. Na grafiku se jasno vidi da su varijacije među modelima jako male i što je najbitnije pozicije pikova su iste za sve modele što potvrđuje da je odabran odgovarajući model sa tačno pretpostavljenim brojem komponenti. Ovakav rezultat je jako dobar s obzirom na mali broj uzoraka i može da posluži kao dobra početna pretpostavka u daljim ispitivanjima ovog sistema.



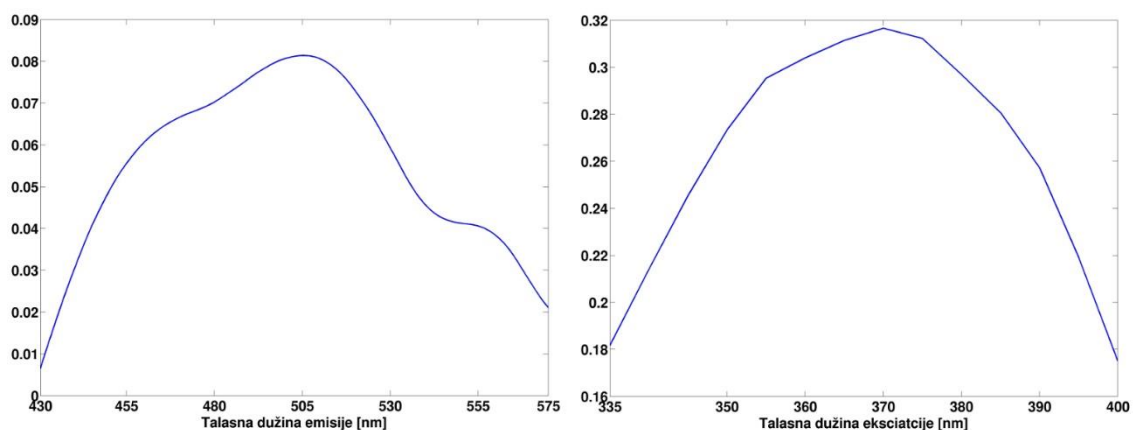
Slika 5.19 – Drugi i treći mod 18 PARAFAC modela dobijenih pri validaciji.

Kako bi se lakše analizirale razlike u koncentracijama ovih fluorofora u uzorcima korišćen je grafik rasutosti. Slika 5.20 prikazuje grafik rasutosti gde je svaki uzorak definisan relativnom koncentracijom prve i druge komponente.



Slika 5.20 – Grafik rasutosti na osnovu relativnih koncentracija prve i druge komponente PARAFAC modela tkiva dojke.

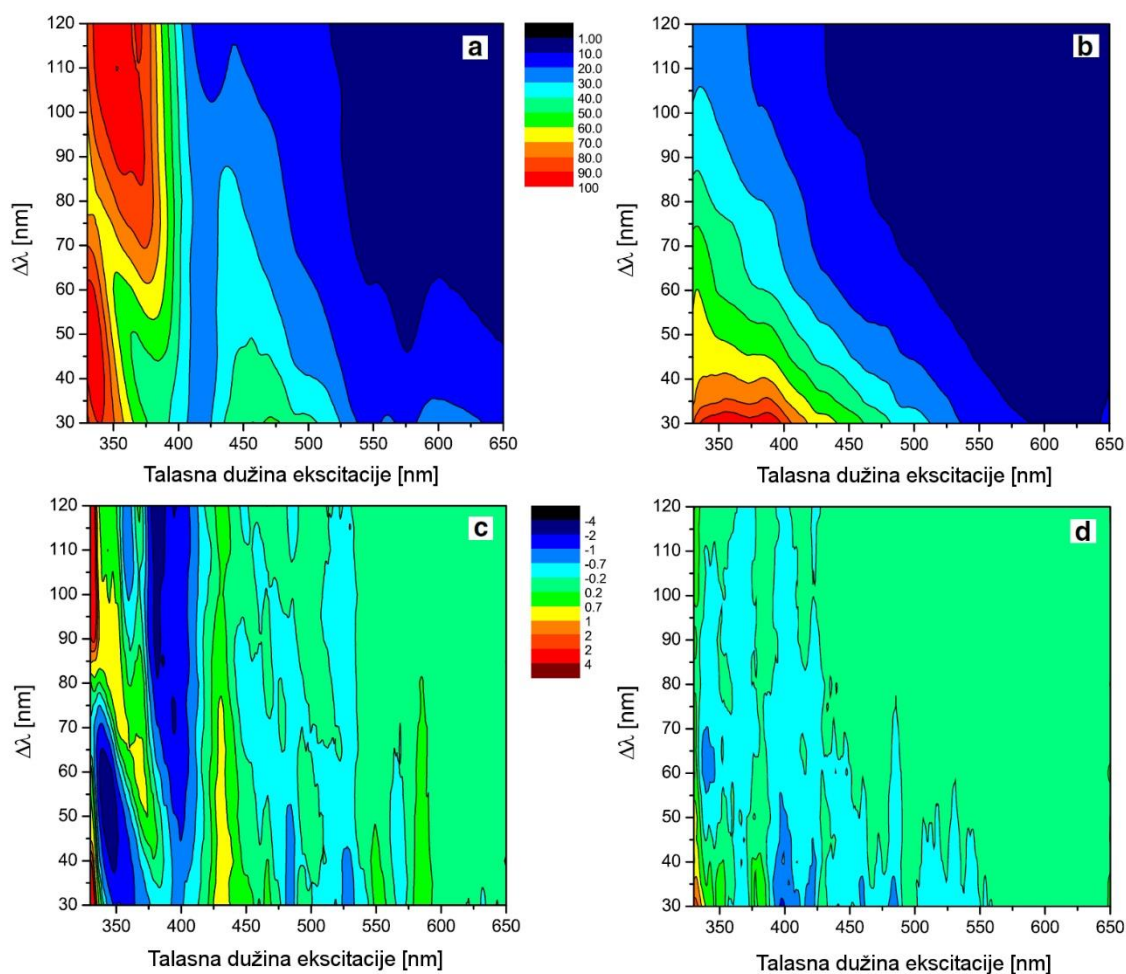
Na grafiku se uočava jasna granica između uzoraka kancera i zdravog tkiva. Parametar po kojem se oni razdvajaju jeste relativna koncentracija druge komponente, odnosno u uzorcima malignog tkiva je koncentracija ove fluorofore veća nego kod zdravog tkiva. Osim prve dve komponente bili su analizirani i grafici sa trećom i četvrtom komponentom, međutim kod svih se pokazalo da su sve bitne razlike u tkivu definisane koncentracijom druge komponente modela. Ekscitacioni i emisioni spektar ove fluorofore posebno je izdvojen na slici 5.21 gde se jasno vide ekscitacioni i emisioni maksimumi na 370 i 505 nm, respektivno. Kao što je ranije rečeno, pretpostavka je da ova fluorofora odgovara lipidima, mada postoji mogućnost da ova fluorofora može da bude i neka druga koja se zavisno od uslova okoline može javiti u toj oblasti poput FAD-a, NADPH i NADH. U našem slučaju nije moguće precizirati tačno koja supstanca u tkivu dojke fluorescira na toj poziciji jer kako je ranije navedeno fluorescentne karakteristike molekula se mogu menjati zavisno od okoline u kojoj se nalaze i njihove pozicije maksimuma mogu varirati i do 50 nm. U radovima koji se bave kancerom dojke utvrđeno je da u tkivu tumora dojke dolazi do povećanja koncentracije koenzima NADH i NADPH [Uppal 2003, Georgakoudi 2002]. U skladu sa tim može se reći da ova fluorofora verovatno odgovara nekom od ta dva koenzima čiji su ekscitacioni i emisioni maksimumi na 351, 336 nm i 460, 464 nm, respektivno. Na osnovu dobijenog modela može se jednostavno pratiti promena koncentracije druge komponente i na brz i jednostavan način odrediti da li postoje maligne promene u tkivu dojke.



Slika 5.21 – Emisioni i ekscitacioni spektar druge komponente dobijen PARAFAC modelom tkiva dojke.

5.3.2 SVM analiza EEM-a i sinhronih spektara

Konturni sinhroni spektri dojke su prikazani na slici 5.22. Kao i kod EEM spektara dojke (odjeljak 5.3.1) na spektrima se u istim regionima uočava fluorescentni doprinos kolagena, NADH, elastina i lipida.



Slika 5.22 – Konturni sinhroni spektri a) zdravog i b) malignog tkiva dojke i prvi izvos sinhronih spektara zdravog i malignog tkiva c) i d), respektivno.

Građenje SVM klasifikacionog modela je bilo izvedeno korišćenjem podataka ekstrahovanih iz određenih spektralnih regiona EEM i sinhronih spektara za svaki pojedinačan uzorak. Kod EEM spektara, komponente su bile računane kao zapremine ispod površine spektra primenom formule u pet podjeljenih regiona:

- 1) talasna dužina ekscitacije od 335 do 400 nm i talasna dužina emisije od 460 do 530 nm – označen sa EEM-1;
- 2) talasna dužina ekscitacije od 350 do 385 nm i talasna dužina emisije od 580 do 625 nm – označen sa EEM-2;
- 3) talasna dužina ekscitacije od 440 do 470 nm i talasna dužina emisije od 500 do 525 nm – označen sa EEM-3;
- 4) talasna dužina ekscitacije od 450 do 470 nm i talasna dužina emisije od 590 do 615 nm – označen sa EEM-4;
- 5) talasna dužina ekscitacije od 400 do 440 nm i talasna dužina emisije od 590 do 615 nm – označen sa EEM-5.

U slučaju sinhronih spektara, komponente izvučene iz spektralnih regiona u obliku zapremine ispod površine spektra i srednje vrednosti prvog izvoda sinhronih spektara.

Za nepromenjene sinhronne spektre bila su analizirana tri regiona:

- 1) talasna dužina ekscitacije od 330 do 400 nm i sinhroni interval od 65 do 120 nm – označen sa SFS-1;
- 2) talasna dužina ekscitacije od 330 do 425 nm i sinhroni interval od 30 do 55 nm – označen sa SFS-2;
- 3) talasna dužina ekscitacije od 430 do 530 nm i sinhroni interval od 30 do 64 nm – označen sa SFS-3.

Za prvi izvod sinhronog spektra bili su analizirani seldeći regiona:

- 1) talasna dužina ekscitacije od 340 do 375 nm i sinhroni interval od 30 do 65 nm – označen sa FDS-1;
- 2) talasna dužina ekscitacije od 380 do 420 nm i sinhroni interval od 65 do 120 nm – označen sa FDS-2;
- 3) talasna dužina ekscitacije od 425 do 440 nm i sinhroni interval od 30 do 85 nm – označen sa FDS-3;

Osim ovih komponenata različite aritmetičke kombinacije spektralnih komponenata su bile takođe izračunate i korišćene kao ulazni podaci za građenje SVM klasifikacionog modela (tabela 5.5).

Tabela 5.5 Kombinacije spektralnih komponenata korišćene kao ulazni podaci za obuku SVM modela za klasifikaciju malignih i benignih promena na tkivu dojke.

Oznaka	Formula	Oznaka	Formula
Etot	$EEM-1 + EEM-2 + EEM-3 + EEM-4 + EEM-5$	RT2	$SFS-2 / Stot$
E12	$EEM-1 / EEM-2$	RT3	$SFS-3 / Stot$
E13	$EEM-1 / EEM-3$	RT12	$(SFS-1 + SFS-2) / Stot$
E14	$EEM-1 / EEM-4$	RT13	$(SFS-1 + SFS-3) / Stot$
E15	$EEM-1 / EEM-5$	RT23	$(SFS-2 + SFS-3) / Stot$
E23	$EEM-2 / EEM-3$	PR112	$SFS-1 / (SFS-1 + SFS-2)$
E24	$EEM-2 / EEM-4$	PR113	$SFS-1 / (SFS-1 + SFS-3)$
E25	$EEM-2 / EEM-5$	PR223	$SFS-2 / (SFS-2 + SFS-3)$
E34	$EEM-3 / EEM-4$	Ftot	$FDS-1 + FDS-2 + FDS-3$
E35	$EEM-3 / EEM-5$	D1	$(FDS-1 - Ftot) / FDS-1$
E45	$EEM-4 / EEM-5$	D2	$(FDS-2 - Ftot) / FDS-2$
Stot	$SFS-1 + SFS-2 + SFS-3$	D3	$(FDS-3 - Ftot) / FDS-3$
S12	$SFS-1 / SFS-2$	D12	$(FDS-1 - FDS-2) / FDS-1$
S13	$SFS-1 / SFS-3$	D13	$(FDS-1 - FDS-3) / FDS-1$
S23	$SFS-2 / SFS-3$	D23	$(FDS-2 - FDS-3) / FDS-2$
RT1	$SFS-1 / Stot$		

U cilju građenja i testiranja modela, podaci su bili podeljeni u dva jednaka seta podataka: set za obuku i set za testiranje modela. U cilju dobijanja što boljeg modela, broj uzoraka je bio uvećan sa izmerenih 18 na 42. Kako bi se uvećao broj uzoraka generisani su novi setovi podataka (12 novih uzoraka normalnog i 12 malignog tkiva) na osnovu srednje vrednosti i standardne devijacije parametara prikazanih u tabeli 5.5. Zatim je SVM model bio obučavan za različite kombinacije 42 ulazna parametra koji opisuju 42 uzorka (11 spektralnih komponenti i 31 kombinacija spektralnih komponenata – tabela 5.5). Bitno je naglasiti da su sve kombinacije bile izračunate samo na osnovu EEM ili sinhronih spektara. U oba slučaja je za obuku SVM modela bio korišćen linearni kernel. Rezultati klasifikacionih modela za različite ulazne podatke su dati u tabeli 5.6.

Tabela 5.6 Predikcija SVM modela naspram rezultata histopatologije svih uzoraka tkiva za različite kombinacije ulaznih podataka.

Kombinacija parametara:	Rezultati testiranja SVM modela				
Sve kombinacije parametara na osnovu EEM		Benigno	Maligno	<i>Senzitivnost</i>	66.67%
	Benigno	15	6	<i>Specifičnost</i>	62.50%
	Maligno	9	12	<i>Greška</i>	35.71%
SFS-1, SFS-2, SFS-3, FDS-1, FDS-2, FDS-3		Benigno	Maligno	<i>Senzitivnost</i>	100%
	Benigno	20	0	<i>Specifičnost</i>	95.23%
	Maligno	1	21	<i>Greška</i>	2.38%
SFS-1, SFS-2, SFS-3, RT-12, RT-13, RT-23		Benigno	Maligno	<i>Senzitivnost</i>	100%
	Benigno	20	0	<i>Specifičnost</i>	95.23%
	Maligno	1	21	<i>Greška</i>	2.38%
FDS-1, FDS-2, FDS-3		Benigno	Maligno	<i>Senzitivnost</i>	100%
	Benigno	20	0	<i>Specifičnost</i>	95.23%
	Maligno	1	21	<i>Greška</i>	2.38%
FDS-1, FDS-2, FDS-3, RT12, RT13, RT23		Benigno	Maligno	<i>Senzitivnost</i>	100%
	Benigno	20	0	<i>Specifičnost</i>	95.23%
	Maligno	1	21	<i>Greška</i>	2.38%
SFS-1, SFS-2, SFS-3		Benigno	Maligno	<i>Senzitivnost</i>	76.19%
	Benigno	18	5	<i>Specifičnost</i>	85.71%
	Maligno	3	16	<i>Greška</i>	21.43%
SFS-1, SFS-2, SFS-3, FDS-1, FDS-2, FDS-3, RT12, RT13, RT23		Benigno	Maligno	<i>Senzitivnost</i>	95.24%
	Benigno	20	1	<i>Specifičnost</i>	95.24%
	Maligno	1	20	<i>Greška</i>	4.76%
D1, D2, D3, D12, D13, D23		Benigno	Maligno	<i>Senzitivnost</i>	100%
	Benigno	21	0	<i>Specifičnost</i>	100%
	Maligno	0	21	<i>Greška</i>	0%
RT12, RT23, SFS-3, FDS-1		Benigno	Maligno	<i>Senzitivnost</i>	100%
	Benigno	21	0	<i>Specifičnost</i>	100%
	Maligno	0	21	<i>Greška</i>	0%
FDS-3, P112, P113, D2, D3		Benigno	Maligno	<i>Senzitivnost</i>	100%
	Benigno	21	0	<i>Specifičnost</i>	100%
	Maligno	0	21	<i>Greška</i>	0%
FDS-3, P112, P113, RT23, RT2		Benigno	Maligno	<i>Senzitivnost</i>	100%
	Benigno	21	0	<i>Specifičnost</i>	100%
	Maligno	0	21	<i>Greška</i>	0%
FDS-3, P112, P113, RT23		Benigno	Maligno	<i>Senzitivnost</i>	100%
	Benigno	21	0	<i>Specifičnost</i>	100%
	Maligno	0	21	<i>Greška</i>	0%
FDS-3, P112, P113		Benigno	Maligno	<i>Senzitivnost</i>	100%
	Benigno	21	0	<i>Specifičnost</i>	100%
	Maligno	0	21	<i>Greška</i>	0%
SFS-1, SFS-2, SFS-3, FDS-1,		Benigno	Maligno	<i>Senzitivnost</i>	100%
	Benigno	21	0	<i>Specifičnost</i>	100%

FDS-2, FDS-3, RT1, RT2, RT3	Maligno	0	21	<i>Greška</i>	0%
FDS-1, FDS-2, FDS-3, D12, D13, D23		Benigno	Maligno	<i>Senzitivnost</i>	100%
	Benigno	21	0	<i>Specifičnost</i>	100%
	Maligno	0	21	<i>Greška</i>	0%
D12, D13, D23		Benigno	Maligno	<i>Senzitivnost</i>	100%
	Benigno	20	0	<i>Specifičnost</i>	95.23%
	Maligno	1	21	<i>Greška</i>	2.38%
FDS-3, PR-113		Benigno	Maligno	<i>Senzitivnost</i>	100%
	Benigno	21	0	<i>Specifičnost</i>	100%
	Maligno	0	21	<i>Greška</i>	0%

Generalno, pokazano je da SVM model obučen na osnovu fluorescentnih podatka može klasifikovati promene na tkivu dojke. U slučaju EEM podataka uspešnost modela je bila manja, sa senzitivnošću od 66.67% i specifičnošću od 62.50%, za slučaj kada su bile korišćene sve kombinacije parametara zajedno za obuku algoritma. Nasuprot tome devet SVM modela obučavanih na osnovu sinhronih spektara su bili mnogo uspešnji, sa specifičnošću i senzitivnošću od čak 100%. Od ovih modela dva su kao ulazne podatke koristili samo dva i tri parametra, FDS-3 i RT113 za prvi model i SFS-1, SFS-2 i SFS-3 za drugi model. Nalaženje uspešne kombinacije ulaznih podataka sa malim brojem parametara ima veliki značaj jer se time smanjuje opseg merenja spektara i pojednostavljuje sama metoda.

Tabela 5.7 Greška SVM klasifikacionog modela dobijenog na osnovu četiri relativne koncentracije fluorofora, dobijene PARAFAC modelom, prisutnih u tkivu dojke.

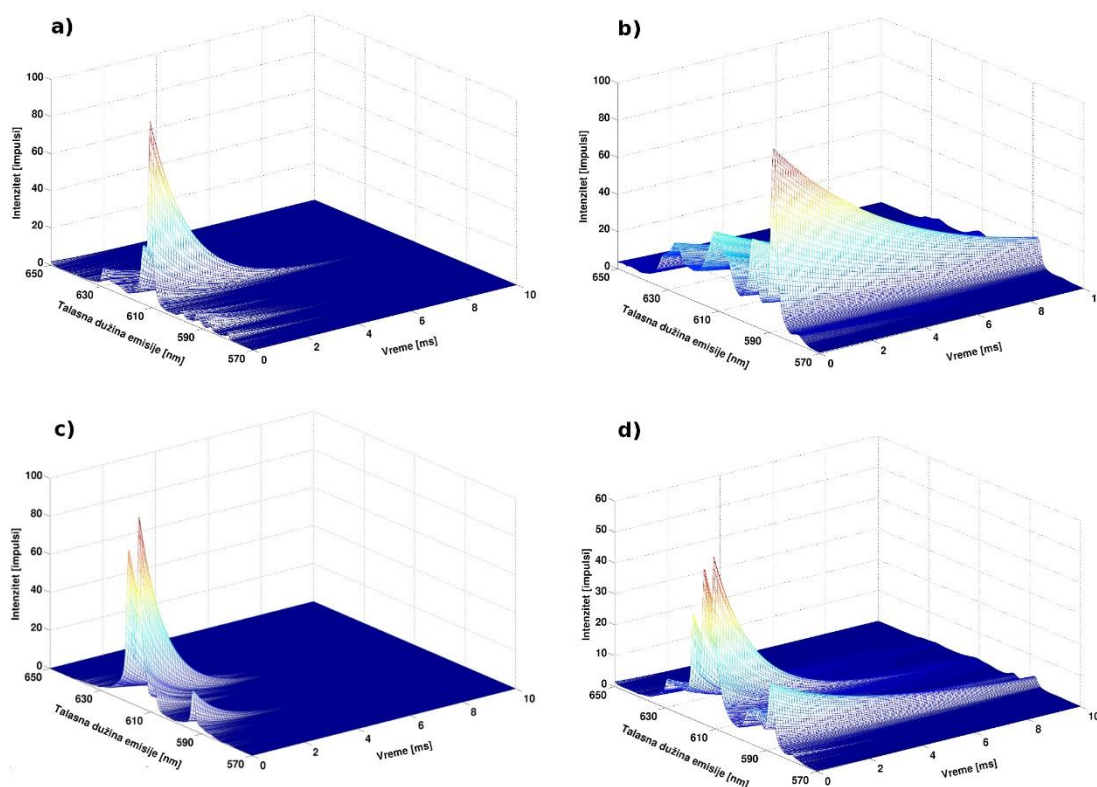
	Benigno	Maligno	<i>Senzitivnost</i>	100%
Benigno	21	0	<i>Specifičnost</i>	100%
Maligno	0	21	<i>Greška</i>	0%

Često obuka SVM modela na osnovu velikog broja parametara, koji opisuju svaki uzorak, može da rezultuje dobijanjem lošijeg klasifikacionog modela. Iz tog razloga je dodatno bio obučen još jedan SVM model na osnovu relativnih koncentracija 4 fluorofore prisutne u tkivu dojke dobijenih PARAFAC analizom. Ova četiri parametra objedinjuju informacije o najbitnijim varijacijama među uzorcima jednog sistema što je suštinski bitno za građenje dobrog kalsifikacionog modela. U tabeli 5.7 su date greške

dobijenog modela. Iz tabele se vidi da je uspešnost modela na osnovu pomenuta četiri parametra bila 100%, odnosno da je greška klasifikacije bila 0%.

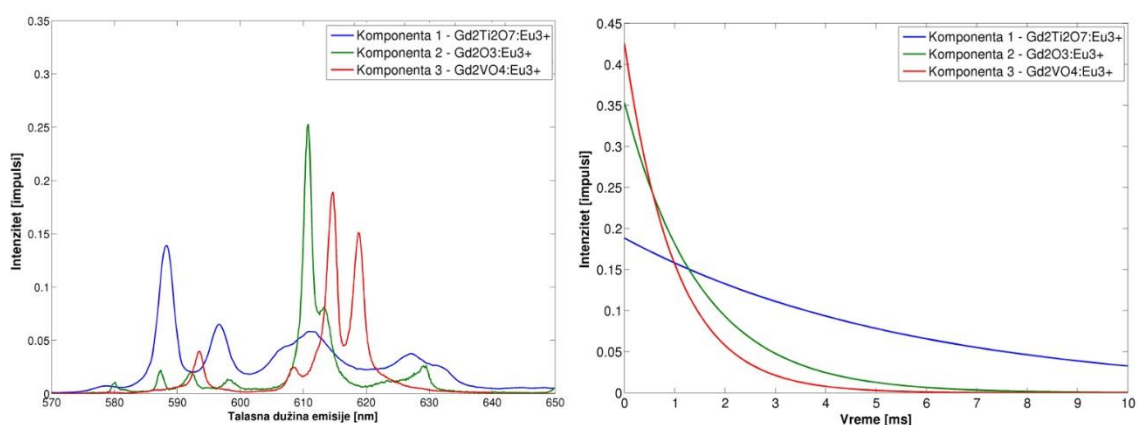
Dobijeni rezultati su pokazali, da je na osnovu fluorescentnih spektara, moguće okarakterisati tkivo dojke i klasifikovati uzorke, na osnovu vrste promena na tkivu (benigno i maligno), sa uspešnošću od 100%. Iako su se obe tehnike merenja pokazale uspešnim za datu klasifikaciju, pokazano je da je kod EEM spektara za izgradnju dobrog modela umesto celih spektara neophodno koristiti relativne koncentracije fluorofora prisutnih u tkivu dojke. Na taj način se model ne opterećuje informacijama prisutnim u spektrima koje su karakteristika sistema ili predstavljaju šum u spektrima, koje ne doprinose građenju klasifikacionog modela.

5.4 Kompleksni sistemi fosfora na bazi retkih zemalja



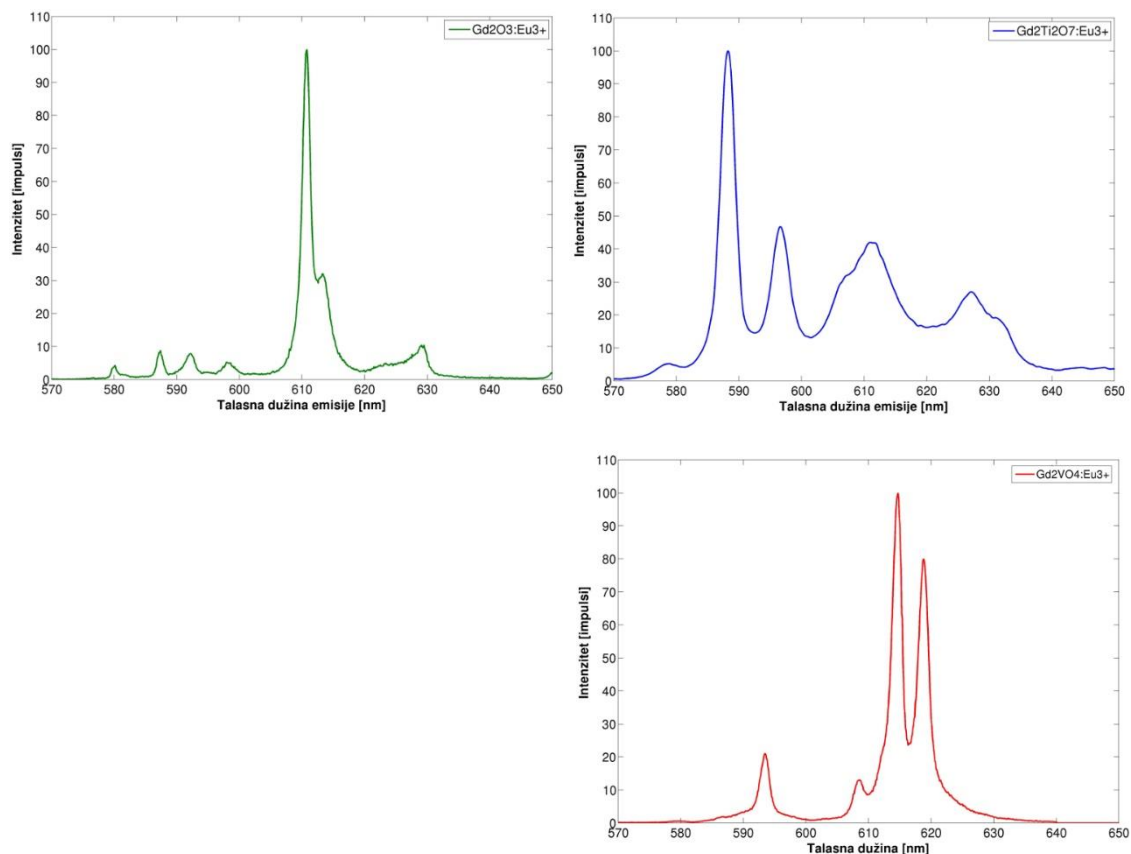
Slika 5.23 – Prikaz trodimenzionalnih podataka za a) $\text{Gd}_2\text{O}_3:\text{Eu}^{3+}$, b) $\text{Gd}_2\text{Ti}_2\text{O}_7:\text{Eu}^{3+}$, c) $\text{GdVO}_4:\text{Eu}^{3+}$ i d) mešavinu sva tri jedinjenja.

Na slici 5.23 su prikazani trodimenzionalni grafici čije x, y i z ose odgovaraju talasnoj dužini emisije, vremenu života i intenzitetu emisije, respektivno. Na graficima su prikazana tri spektra koja sadrže svaka od tri jedinjenja ponaosob ($\text{Gd}_2\text{O}_3:\text{Eu}^{3+}$, $\text{Gd}_2\text{Ti}_2\text{O}_7:\text{Eu}^{3+}$ i $\text{GdVO}_4:\text{Eu}^{3+}$) i jedan spektar mešavine sve tri supstance zajedno. Na spektrima pojedinačnih supstanci vide se jasni pikovi karakterističnih energetskih prelaza. Kada se pogleda grafik mešavine, tu već dolazi do preklapanja spektara i vremena života i samim tim je teško razdvojiti spektre čistih supstanci i odrediti njihove odnose koncentracija. Cilj je takođe bio ustanoviti koja je granica detekcije ove metode za prisutnost veoma malih koncentracija drugih supstanci (reda do 0.1%) u uzorku čistog materijala.



Slika 5.24 – Grafik drugog i trećeg moda PARAFAC modela za neorganski sistem.

Kako bi se rešio dati problem bio je proračunat trokomponentni PARAFAC model na osnovu merenih spektara za sedam uzoraka, gde dodatna validacija za određivanje broja komponenti nije bila potrebna jer je broj bio unapred poznat. Na slici 5.24 su prikazani svi modovi dobijenog modela, gde prvi mod prikazuje promenu koncentracija pojedinačnih luminescentnih supstanci u svakom merenom uzorku, drugi mod prikazuje emisione spektre pojedinačnih luminescentnih jedinjenja, dok treći pokazuje njihove karakteristične krive vremena života pobuđenog stanja. Na grafiku drugog moda se vidi da je model jasno izvukao karakteristične spektre pojedinačnih supstanci. Kako bi se uporedio rezultat modela sa realnim spektrima na slici 5.25 su prikazani snimljeni spektri čistih jedinjenja. Uočava se da je model ekstrahovao veoma precizno čiste spektre iz kompleksnih jedinjenja i da su verodostojni izmerenim spektrima.



Slika 5.25 – Izmereni emisijski spektri pojedinačnih jedinjenja.

Kako bi se potvrdila pouzdanost modela pri određivanju prisutnosti i koncentracije neke od date tri supstance u nepoznatim uzorcima, korišćeni su izmereni spektri 21 uzorka pri čemu je njima u Matlab-u na kraju dodat i šum od 5% intenziteta kako bi se ispitalo ponašanje modela u teškoj situaciji pošto uslovi merenja nisu uvek idealni i može se desiti da se u spektru nađe šum koji može otežati dalju analizu. Za nepoznati 21 uzorak traženo je da model odredi koncentracije supstanci u svakom pojedinačnom uzorku pri čemu je modelu rečeno da pri tome koristi prethodno izračunati drugi i treći mod kao polazni parametar koji su bili dobijeni na osnovu „idealnih“ spektara.

U tabeli 5.8 su prikazane stvarne koncentracije i koncentracije koje je model proračunao na osnovu spektara početnih 7 uzoraka. Uočava se da su odstupanja u koncentraciji jako mala i da je najveće odstupanje kod GdVO₄:Eu³⁺ gde je kod trećeg uzorka greška predviđanja bila 0.036%. Ovo se može smatrati jako malom greškom i ukazuje na to da model predviđa koncentracije sa jako visokom tačnošću i da je pouzdan čak i kod

spektara sa šumom iako je bio obučavan na „idealnim“ spektrima. Bitno je napomenuti da su dati rezultati pokazali da se može smatrati da ova metoda ima granicu detekcije od reda 0.1%, jer ako se obrati pažnja na dobijene predviđene koncentracije može se primetiti da ukoliko se one zaokruže na prvu decimalu dobija se upravo realna koncentracija svakog jedinjenja u uzorku. Time se može reći da za određivanje prisutnosti drugih jedinjenja u nekom materijalu ova metoda može dati pouzdane rezultate za koncentracije do reda 0.1%.

Tabela 5.8 Stvarne koncentracije i koncentracije luminescentnih supstanci dobijene na osnovu PARAFAC modela

Broj uzorka	$Gd_2O_3:Eu^{3+}$		$Gd_2Ti_2O_7:Eu^{3+}$		$GdVO_4:Eu^{3+}$	
	Stvarna konc. [%]	Predviđena konc. [%]	Stvarna konc. [%]	Predviđena konc. [%]	Stvarna konc. [%]	Predviđena konc. [%]
1.	97	97.001	2.5	2.5	0.5	0.534
2.	97	96.999	2	2	1	1.034
3.	65	65.012	35	34.99	0	0.015
4.	65	64.999	0	0.002	35	35.001
5.	80	80.006	20	19.994	0	0.024
6.	80	79.999	0	0.001	20	20.015
7.	93	93.001	7	6.998	0	0.032
8.	94	93.999	0	0.001	6	6.03
9.	98	97.999	0	0	2	2.034
10.	99.7	99.699	0	0	0.3	0.336
11.	99.9	99.899	0.1	0.1	0	0.036
12.	99.9	99.899	0	0	0.1	0.136
13.	99	98.999	1	1	0	0.036
14.	99	98.999	0	0	1	1.035
15.	0	0.023	65	64.984	35	34.96
16.	0	0.029	80	79.978	20	19.966
17.	0	0.033	92	91.974	8	7.971
18.	0	0.035	98	97.972	2	1.974
19.	0	0.036	99.5	99.472	0.5	0.475

20.	0	0.036	99.9	99.872	0.1	0.075
21.	0	0.035	99	98.972	1	0.975

Dobijeni model je pokazao potencijal za primenu ove vrste analize podataka odnosno modeliranja ne samo za ekscitaciono-emisione matrice fluorescentnih spektara već i kod druge vrste višedimenzionalnih podataka gde jedna dimenzija u sistemu varira dok su druge fiksne i opisuju fenomen jedinstven za posmatrani sistem.

6. Zaključak

Predmet istraživanja ove doktorske disertacije je fluorescentna spektroskopija, odnosno analiza fluorescentnih spektara višekomponentnih sistema, kao i analiza i modeliranje podataka dobijenih tom metodom. Cilj ove teze je bio primena paralelne faktorske analize i drugih multivarijantnih metoda, na EEM i sinhronu fluorescentne spektre više kompleksnih sistema, u cilju modeliranja i analize bioloških i neorganskih sistema. Razumevanje fizičkih fenomena u kompleksnim sistemima nam daje mogućnost primene ove metode za karakterizaciju organskih i neorganskih materijala u različitim oblastima nauke. PARAFAC modeli kompleksnih sistema omogućavaju da se na osnovu fluorescentnih spektara nepoznatog uzorka odredi prisutnost određenih fluorofora u njemu i njihove relativne koncentracije. Sa odgovarajućom kalibracijom PARAFAC modela mogu se određivati i stvarne koncentracije pojedinačnih fluorofora u uzorcima kompleksnog sistema i na taj način se izbeći standardne hemijske metode. Osim modela za karakterizaciju uzoraka, na osnovu fluorescentnih spektara je moguće graditi različite klasifikacione modela primenom različitih multivarijantnih analiza (PCA, PLS-DA, SVM itd.). Klasifikacioni modeli mogu biti korisni u prehrambenoj industriji i u medicini. Ovakvi modeli se mogu koristiti za autentifikaciju i klasifikaciju različitih prehrambenih proizvoda, kao i za klasifikaciju benignih i malignih promena kod bioloških uzoraka.

Pouzdanost i prediktabilnost PARAFAC modela je bila analizirana na primeru sistema vode i dve aminokiseline. Dati model je pokazao da sa velikom tačnošću mogu da se odrede koncentracije prisutnih aminokiselina u rastvoru. Osim toga model je kao rezultat iz spektara kompleksnih sistema izvukao čiste ekscitacione i emisione spektre pojedinačnih fluorofora, na osnovu čijih se pozicija moglo okarakterisati kojoj fluorofori odgovaraju. U datom primeru ta informacija nije bila od velike važnosti, međutim za kompleksne sisteme gde su fluorofore nepoznate takva vrsta informacije ima veliku ulogu u karakterizaciji posmatranog sistema.

Analizom šestokomponentnog PARAFAC modela dobijenog na osnovu EEM-a uzoraka meda dobijene su tačne pozicije ekscitacionih i emisionih maksimuma šest fluorofora koje imaju najveći doprinos fluorescentnim karakteristikama meda. Pretpostavljeno je da te pozicije odgovaraju: aromatičnoj aminokiselini – triptofanu, riboflavinu (vitamin B2), fenoličnim komponentama i produktima Maillard-ove reakcije (HMF, furozin). Uočeno je da botaničko poreklo meda direktno zavisi od koncentracije produkta Maillard-ove reakcije i fenoličnih komponenti u uzorku. Klasa lipovog meda se mogla jasno razdvojiti od ostalih grupa manjom koncentracijom fluorofore sa ekscitacionim i emisionim maksimumom na 330/428 nm dok su ostale grupe mogle da se razdvoje različitim doprinosima druge komponente sa maksimumom na 370/426 nm. Za uzorke falsifikata meda se pokazalo specifičnim da sadrže manje količine prve i treće komponente od prirodnog meda. Klasifikacioni modeli su potvrdili da se na osnovu fluorescentnih spektara meda može odrediti njegovo botaničko poreklo i autentičnost. Model izgrađen na osnovu sinhronih spektara se pokazao uspešniji sa srednjom greškom klasifikacije od 5.35%. Model za utvrđivanje autentičnosti meda je imao 100% uspešnost pri klasifikaciji. Takođe je bilo potvrđeno da su klasifikacioni modeli na osnovu infracrvenih spektara meda podjednako uspešni.

Dobijeni PARAFAC model za uzorke tkiva dojke je razdvojio doprinos 4 fluorofore u merenim uzorcima. Na osnovu rezultata modela dobijene su pozicije pikova tih fluorofora i pretpostavljeno je da prva komponenta odgovara kolagenu i elastinu, druga i četvrta lipidima i treća koenzimu NADH. Na osnovu grafika rasutosti primećeno je da je najveća razlika između tkiva tumora i zdravog tkiva u koncentraciji druge komponente koja je imala ekscitacioni maksimum na 370 nm, a emisioni na 505 nm. Iako je bilo pretpostavljeno da ova pozicija odgovara lipidima, u skladu sa rezultatima iz naučnih radova gde je pokazano da je u tkivu tumora povećana koncentracija koenzima NADH i NADPH, na kraju je zaključeno da ova fluorofora najverovatnije odgovara nekom od koenzima. Na osnovu parametara sinhronih spektara i relativnih koncentracija 4 fluorofore dobijenih PARAFAC analizom, bili su dobijeni SVM modeli sa validiranom greškom od 0%, za određivanje uzoraka sa malignim promenama.

Modeliranje neorganskog sistema je pokazalo uspešnu primenu PARAFAC analize ne samo na EEM fluorescentne spektre već i na druge višedimenzionalne podatke, u ovom slučaju emisione spektre i vremena života pobuđenog stanja u funkciji intenziteta. Model izgrađen na osnovu sedam uzoraka mešavine tri luminescentne supstance je pokazao veliku preciznost pri određivanju koncentracije 21 uzorka nepoznatih modelu. Spektrima ovih test uzoraka je bio dodat i šum radi dodatne provere pouzdanosti modela. I pored toga, model je sa visokom preciznošću pretpostavio koncentracije čak do reda 0.1% pri čemu je najveće odstupanje od stvarne koncentracije bilo 0.034%. Ovim je bilo pokazano da se uz primenu ovog modela mogu uspešno razdvajati i spektri neorganskih materijala u slučajevima kada dolazi do njihovog preklapanja čime je otežano karakterisanje datog materijala. Takođe ovaj model ima veliki potencijal za primenu ove tehnike pri proveru čistoće nekog materijala prilikom karakterizacije, pri čemu je neophodno samo jedno brzo merenje, bez potrebe za drugim metodama karakterizacije koje se standardno koriste u ove svrhe. Ovakva tehnika bi mogla da nađe primenu u proveru kvaliteta u industrijskoj proizvodnji ovakvih supstanci, pošto se može jednostavno primeniti direktno na proizvodnoj traci.

Rezultati ove teze su pokazali veliki potencijal primene multivarijantnih metoda (PARAFAC, PCA, PLS-DA i SVM) i fluorescentne spektroskopije u karakterizaciji i klasifikaciji organskih i neorganskih materijala. PARAFAC metoda je pokazala jako dobre rezultate kada je reč o preciznosti i pouzdanosti izgrađenih modela za date kompleksne sisteme. Primena multivarijantnih metoda na fluorescentne spektre se pokazala izuzetno uspešnom za klasifikaciju i autentifikaciju biloških uzoraka. Na osnovu prikazanih rezultata ovog rada može se zaključiti da kombinacija multivarijantnih metoda i fluorescentne spektroskopije ima veliki potencijal u budućnosti kao brza, veoma osetljiva, nedestruktivna i jeftina metoda za primenu u različitim oblastima nauke i industrije.

Dalja istraživanja će biti fokusirana na modeliranje novih kompleksnih sistema i sakupljanje većeg broja uzoraka radi dobijanja što boljih modela i mogućoj

standardizaciji ove metode. Osim toga težiće se i ka tome da se PARAFAC analiza olakša i prilagodi za korišćenje istraživačima iz različitih oblasti nauke. Metoda prilagođena na taj način neće zahtevati od korisnika bilo kakvo predznanje iz statistike i kompleksnog matričnog računa. Na taj način bi se ubrzala ekspanzija primene ove metode i dostupnost većem broju istraživača.

7. Literatura

Alfano R.R., Tang G.C., Pradhan A., Lam W., Choy D.S.J. & Opher E. Fluorescence spectra from cancerous and normal human breast and lung tissues. *IEEE J Quantum Electron* (1987) 23: 1806–1811.

Andersen C.M. & Mortensen G. Fluorescence spectroscopy: a rapid tool for analyzing dairy products. *J Agric Food Chem* (2008) 56: 720–729.

Andrić Ž., Dramićanin M.D., Mitrić M., Jokanović V., Bessière A. & Viana B. Polymer complex solution synthesis of $(Y_xGd_{1-x})_2O_3:Eu^{3+}$ nanopowders. *Optical Materials* (2008) 30: 1023-1027.

Aubourg S.P. Recent advances in assessment of marine lipid oxidation by using fluorescence. *J Am Oil Chem Soc* (1999) 76: 409–419.

Becker E.M., Christensen J., Frederiksen C.S. & Haugaard V.K. Front-face fluorescence spectroscopy and chemometrics in analysis of yogurt: rapid analysis of riboflavin. *J Dairy Sci* (2003) 86: 2508–2515.

Bro R. Exploratory study of sugar production using fluorescence spectroscopy and multi-way analysis. *Chemom. Intell. Lab. Syst.* (1999) 46: 133-147.

Bro R. PARAFAC: Tutorial & applications, *Chemom. Intell. Lab. Syst.* (1997) 38: 149-171.

Bro R. Multivariate calibration - What is in chemometrics for the analytical chemist? *Anal. Chim. Acta* (2003) 500 (1-2): 185-194.

Burdick D.S., Tu X.M., McGown L.B. & Millican D.W. Resolution of multicomponent fluorescent mixtures by analysis of excitation-emission-frequency array. *J. Chemom.* (1990) 4: 15-28.

Burges C.J.C. A tutorial on support vectro machines for pattern recognition. *Data Min Knowl Disc* (1998) 2: 121-167.

- Callejón R.M., Amigo J.M., Pairo E., Garmón S., Ocaña J.A. & Morales M.L. Classification of Sherry vinegars by combining multidimensional fluorescence, parafac and different classification approaches. *Talanta* (2012) 88: 456-462.
- Carroll J. D. & Chang, J. Analysis of individual differences in multidimensional scaling via an N-way generalization of 'Eckart- Young' decomposition. *Psychometrika* (1970) 35: 283-319.
- Cattell R. B. 'Parallel proportional profiles' and other principles for determining the choice of factors by rotation. *Psychometrika* (1944) 9: 267-283.
- Cattell R. B. The three basic factor-analytic research designs-their interrelations and derivatives. *Psychological Bulletin* (1952) 49: 499-520.
- Cattell R. B. The scree test for the number of factors. *Multivariate Behavioral Research* (1966) 1 (2): 245-276.
- Christensen J., Becker E.M. & Frederiksen C.S. Fluorescence spectroscopy and PARAFAC in the analysis of yogurt. *Chemomet. Intell Lab. Syst.* (2005) 75: 201–208.
- Ćulubrk S., Antić Ž., Marinović-Cincović M., Ahrenkiel P.S. & Dramićanin M.D. Synthesis and luminescent properties of rare-earth (Sm^{3+} and Eu^{3+}) doped $\text{Gd}_2\text{Ti}_2\text{O}_7$ pyrochlore nanopowders. *Optical Materials* (2014) DOI:10.1016/j.optmat.2014.08.001
- Dramićanin T., Dramićanin M.D., Jokanović V., Nikolić-Vukosavljević D. & Dimitrijević B. Three-dimensional total synchronous luminescence spectroscopy criteria for discrimination between normal and malignant breast tissues. *Photochem Photobiol* (2005) 81: 1554–1558.
- Dramićanin T., Lenhardt L., Zeković I. & Dramićanin M.D. Support Vector Machine on Fluorescence Landscapes for Breast Cancer Diagnostics. *J Fluoresc.*, (2012) 22: 1281-1289.
- Dufour E., Devaux M.F., Fortier P. & Herbert S. Delineation of the structure of soft cheeses at the molecular level by fluorescence spectroscopy - relationship with texture. *International Dairy Journal* (2001) 11 (4-7): 465-473.

- Dufour E., Mazerolles G., Devaux M. F., Duboz G., Duployer M. H. & Riou M. N. Phase transition of triglycerides during semi-hard cheese ripening. *International Dairy Journal* (2000) 10 (1-2): 81-93.
- Dufour E. & Riaublanc A. Potentiality of spectroscopic methods for the characterisation of dairy products. I. Front-face fluorescence study of raw, heated and homogenised milks. *Lait* (1997) 77 (6): 657- 670.
- Geladi P. Analysis of multi-way (multi-mode) data. *Chemometric and Intelligent Laboratory Systems* (1989) 7: 11-30.
- Georgakoudi I., Jacobson B.C., Muller M.G., Sheets E.E., Badizadegan K., Carr-Locke D.L., Crum C.P., Boone C.W., Dasari R.R., Van Dam J. & Feld M.S. NAD(P)H and collagen as in vivo quantitative fluorescent biomarkers of epithelial precancerous changes. *Cancer Res* (2002) 62(3): 682-687.
- Gill P.E., Murray W. & Wright M.H. Practical optimization. Academic Press Inc. (London) Limited. (1981)
- Harshman R.A. & Lundy M.E. Data preprocessing and the extended PARAFAC model, in "Research methods for Multimode data analysis". (Eds. H.G. Law, C.W. Snyder, J.A. Hattie and R.P. McDonald) Praeger, New York, 1984.
- Harshman, R. A. Foundations of the PARAFAC procedure: Models and conditions for an "explanatory" multi-modal factor analysis. *UCLA Working Papers in Phonetics (Ann Arbor: University Microfilms)* (1970) 16: 1-84.
- Harshman R. A. "How can I know if it's 'real'?" A catalog of diagnostics for use with three-mode factor analysis and multidimensional scaling. In H. G. Law, C. W. Snyder, Jr., J. Hattie, & R. P. McDonald (Eds.), (1984) Research methods for multimode data analysis (pp. 566-591). New York: Praeger.
- He L.M., Kear-Padilla L.L., Lieberman S.H. & Andrews J.M. Rapid in situ determination of total oil concentration in water using ultraviolet fluorescence and light scattering coupled with artificial neural networks. *Analytica Chimica Acta* (2003) 478: 245-258.

- Herbert S., Mouhous R.N., Devaux M.F., Riaublanc A., Bouchet B., Gallant D.J. & Dufour E. Monitoring the identity and the structure of soft cheeses by fluorescence spectroscopy. *Lait* (2000) 80: 621–634.
- Herbert S., Riaublanc A., Bouchet B., Gallant D.J. & Dufour E. Fluorescence spectroscopy investigation of acid-or rennet-induced coagulation of milk. *Journal of Dairy Science* (1999) 82 (10): 2056-2062.
- Ho C.N., Christian G.D., Davidson E.R. Application of Method of Rank Annihilation to Quantitative-Analyses of Multicomponent Fluorescence Data from Video Fluorometer. *Analytical Chemistry* (1978) 50 (8): 1108-1113.
- Ivancuic O. Applications of Support Vector Machines in Chemistry. In Reviews in Computational Chemistry, Lipkowitz KB Cundari TR (eds), Wiley-VCH, Weinheim, (2007) 23: 291-400.
- Jiji R.D. & Booksh K.S. Mitigation of Rayleigh and Raman spectral interferences in multi-way calibration of excitation-emission matrix fluorescence spectra. *Anal. Chem.* (2000) **72**: 718-725.
- Jovanović D.J., Antić Ž., Krsmanović R.M., Mitrić M., Đorđević V., Bártová B. & Dramićanin M.D. Annealing effects on the microstructure and photoluminescence of Eu^{3+} -doped GdVO_4 powders. *Optical Materials* (2013) 35 (10): 1797-1804.
- Kamath S.D. & Mahato K.K. Optical pathology using oral tissue fluorescence spectra: classification by principal component analysis and k-means nearest neighbor analysis. *J Biomed Opt* (2007) 12: 014028
- Karoui R., Dufour E., Bosset J.O. & De Baerdemaeker J. The use of front face fluorescence spectroscopy to classify the botanical origin of honey samples produced in Switzerland. *Food Chem* (2007a) 101: 314–323.
- Karoui R., Dufour E. & De Baerdemaeker J. Front face fluorescence spectroscopy coupled with chemometric tools for monitoring the oxidation of semi-hard cheeses throughout ripening. *Food Chem* (2007b) 101: 1305–1314.
- Kauppinen R.A., Williams S.R., Busza A.L. & Van Bruggen N. Applications of magnetic resonance spectroscopy and diffusion-weighted imaging to the study of brain biochemistry and pathology. *Trends in Neuroscience* (1993) 16(3): 88-95.

- Kiers H.A.L. & Krijnen W.P. An efficient algorithm for PARAFAC of three-way data with large numbers of observation units. *Psychometrika* (1991) 56: 147-152.
- Kiers H.A.L. Weighted least squares fitting using iterative ordinary least squares algorithms. *Psychometrika* (1997) 62: 251-266.
- Kruskal J.B. Rank, decomposition, and uniqueness for 3-way and N-way arrays, In "Multiway data analyses", (Eds. R. Coppi and S. Bolasco), Elsevier Science Pub. (North-Holland), (1989).
- Kulmyrzaev A.A. & Dufour E. Determination of lactulose and furosine in milk using front-face fluorescence spectroscopy. *Lait* (2002) 82 (6): 725-735.
- Kulmyrzaev A.A., Levieux D. & Dufour E. Front-face fluorescence spectroscopy allows the characterization of mild heat treatments. *Lab. Syst.* (2005) 75: 201-208.
- Lackowicz J.R. Principles of fluorescence spectroscopy, 2nd ed., Kluwer Academic/Plenum Publishers, New York, (1999).
- Lawton W.H & Sylvestre E.A. Self modeling curve resolution. *Technometr* (1971) 13: 617-633.
- Lenhardt L., Zeković I., Dramićanin T., Dramićanin M.D. & Bro R. Determination of the botanical origin of honey by front-face synchronous fluorescence spectroscopy. *Appl. Spectrosc.* (2014) 68 (5): 557-563.
- Lenhardt L., Zeković I., Dramićanin T. & Dramićanin M.D. Artificial neural network for processing fluorescence spectroscopy data in skin cancer diagnostics. *Phys Scripta* (2013) T157: 014057.
- Lenhardt L., Zeković I., Dramićanin T., Tešić Ž., Milojković-Opsenica D. & Dramićanin M.D. Authentication of the botanical origin of unifloral honey by infrared spectroscopy coupled with support vector machine algorithm. *Phys Scripta* (2014) T162: 014042.
- Leurgans S. & Ross R.T. Multilinear models in spectroscopy. *Statist. Sci.* (1992) 7: 289-319.
- Li J.S., Wang H., Zhang X. & Zhang H.S. Spectrofluorometric determination of total amount of nitrite and nitrate in biological sample with new fluorescent probe 1,3,5,7-

- tetramethyl-8-(3',4'-diaminophenyl)- difluoroboradiaza-s-indacene. *Talanta*. (2003) 61 (6): 797-802.
- Li S. and Gemperline P.J. Eliminating complex eigenvectors and eigenvalues in multiway analyses using the direct trilinear decomposition method. *J. Chemom.* (1993) 7: 77-88.
- Marchiarullo M.A. & Ross R.T. Factor-analysis of chlorophyll fluorescence in photosynthetic systems. *Biophysical Journal* (1982) 37 (2): A233.
- Martens H. & Næs T., *Multivariate Calibration*, John Wiley & Sons, Chichester, UK, (1989).
- Mazerolles G., Devaux M.F., Duboz G., Duployer M.H., Riou N.M. & Dufour E. Infrared and fluorescence spectroscopy for monitoring protein structure and interaction changes during cheese ripening. *Lait* (2001) 81 (4): 509-527.
- Mercer J. Functions of positive and negative type and their connection with the theory of integral equations. *Phil Trans Roy Soc London A* (1909) 209:415-446.
- Mitchell B.C. & Burdick D.S. Slowly converging PARAFAC sequences: Swamps and two-factor degeneracies. *J. Chemom.* (1994) 8: 155-168.
- Munck L., Norgaard L., Engelsen S.B., Bro R. & Andersson C.A. Chemometrics in food science – a demonstration of the feasibility of a highly exploratory, inductive evaluation strategy of fundamental scientific significance. *Chemometrics and Intelligent Laboratory Systems* (1998) 44 (1-2): 31-60.
- Paatero P. A weighted non-negative least squares algorithm for three-way "PARAFAC" factor analysis. *Chemometrics and Intelligent Laboratory Systems* (1997) 38: 223-242.
- Pataca L.C.M., Borges Neto W., Marcucci C.M. & Poppi J.R. Determination of apparent reducing sugars, moisture and acidity in honey by attenuated total reflectance-fourier transform infrared spectrometry. *Talanta* (2007) 71: 1926–1931.
- Persano Oddo L. & Piro R. Main European unifloral honeys: descriptive sheets *Apidologie* (2004) 35: 38–81.

- Plugge W. & Van Der Vlies C. Near-infrared spectroscopy as an alternative to assess compliance of ampicillin trihydrate with compendial specifications. *Journal of Pharmaceutical and Biomedical Analysis* (1993) 11 (6): 435-442.
- Ramanujam N. Fluorescence spectroscopy in vivo. In: Meyers RA (ed) Encyclopedia of analytical chemistry. John Willey & Sons, Ltd., Chichester, New York, pp 20–56, (2000).
- Rodriguez-Delgado M.A., Malovana S., Perez J.P., Borges T. & Garcia Montelongo F.J. Separation of phenolic compounds by high-performance liquid chromatography with absorbance and fluorometric detection. *J. Chromatogr. A.* (2001) 912 (2): 249–257.
- Ross R.T., Lee C.H., Davis C.M., Ezzeddine B.M., Fayyad E.A. & Leurgans S.E. Resolution of the fluorescence-spectra of plant pigment-complexes using trilinear models. *Biochimica et Biophysica Acta* (1991) 1056 (3): 317-320.
- Ruoff K., Karoui R., Dufour E., Luginbuhl W., Bosset J.O., Bogdanov S. & Amado R. Authentication of the botanical origin of honey by front-face fluorescence spectroscopy, a preliminary study. *J. Agric. Food Chem.* (2005) 53 (5): 1343-1347.
- Ruoff K., Luginbuhl W., Kunzli R., Bogdanov S., Bosset J., Von der Ohe K., Von der Ohe W. & Amado R. Authentication of the Botanical and Geographical Origin of Honey by Front-Face Fluorescence Spectroscopy. *J. Agric. Food Chem.* (2006) 54(18): 6858-6866.
- Russell M.D. & Gouterman M. Excitation-Emission-Lifetime analysis of multicomponent systems .1. principal component factor-analysis. *Spectrochimica Acta Part A-Molecular and Biomolecular Spectroscopy* (1988) 44 (9): 857-861.
- Sadecka J. & Tothova J. Fluorescence spectroscopy and chemometrics in the food classification—a review. *Czech J Food Sci* (2007) 546 (25) 159–173.
- Sanchez E. & Kowalski B.R. Tensorial resolution: A direct trilinear decomposition. *J. Chemom.* (1990) 4: 29-45.
- Sanchez E. & Kowalski B.R. Tensorial calibration: II. Second-order calibration. *J. Chemometrics* (1988) 2: 265–280.
- Sanchez E. & Kowalski B.R. Generalized Rank Annihilation Factor-Analysis. *Analytical Chemistry* (1986) 58 (2): 496-499.

- Schomacker K.T., Frisoli J.K., Compton C.C., Flotte T.J., Richter J.M., Nishioka N.S. & Deutsch T.F. Ultraviolet laser-induced fluorescence of colonic tissue: basic biology and diagnostic potential. *Lasers Surg Med* (1992) 12: 63–78.
- Sidiropoulos N. D. & Bro R. On the uniqueness of multilinear decomposition of N-way arrays. *J. Chemometrics* (2000) 14: 229–239.
- Sikorska E., Gorecki T., Khmelinskii I.V., Sikorski M. & Koziół J. Classification of edible oils using synchronous scanning fluorescence spectroscopy. *Food Chem* (2005) 89: 217–225.
- Silverman D.C. Corrosion prediction in complex environments using electrochemical impedance spectroscopy. *Electrochimica Acta* (1993) 38 (14): 2075-2078.
- Smilde A. K. & Doornbos D.A. 3-Way methods for the calibration of chromatographic systems - comparing parafac and 3-way pls. *Journal of Chemometrics* (1991) 5 (4): 345-360.
- Sterenborg H.J.C.M., Motamedi M., Wagner R.F., Duvic M., Thomsen S. & Jacques S.L. In vivo fluorescence spectroscopy and imaging of human skin tumours. *Lasers Med Sci* (1994) 9: 191–201.
- Strasburg G.M. & Ludescher R.D. Theory and applications of fluorescence spectroscopy in food research. *Trends in Food Science & Technology* (1995) 6 (3): 69-75.
- Ten Berge J.M.F. & Sidiropoulos N.D. On uniqueness in Candecom/Parafac. *Psychometrika* (2002) 67: 399–409.
- Thurstone L. L. The vectors of mind: Multiple-factor analysis for the isolation of primary traits. Chicago, IL, US: University of Chicago Press (1935).
- Uppal A. & Gupta P.K. Measurement of NADH concentration in normal and malignant human tissues from breast and oral cavity. *Biotechnol. Appl. Biochem.* (2003) 37: 45-50.
- Wold H. Path models with latent variables: The NIPALS approach. In: Blalock H. M. et al. (editors). Quantitative Sociology: International perspectives on mathematical and statistical model building. Academic Press, N.Y., 307-357, (1975).

Wold J.P., Mielnik M., Pettersen M.K., Aaby K. & Baardreth P. Rapid assessment of rancidity in complex meat products by front face fluorescence spectroscopy. *J Food Sci* (2002) 67: 2397–2404.

Wold H. Nonlinear estimation by iterative least squares procedures. *Research Papers in Statistics: Festschrift for J. Neyman* (1966) 411-444.

Zeković I., Lenhardt L., Dramićanin T., Bandić J. & Dramićanin M.D. Discrimination between melanoma, nevi and normal skin by using synchronous luminescence spectroscopy. *Appl. Spectrosc.* (2014) 68 (8): 823-830.

Zeković I., Lenhardt L., Dramićanin T. & Dramićanin M.D., Classification of intact cereal flours by front-face synchronous fluorescence spectroscopy. *Food Anal. Methods* (2012) 5: 1205-1213.

8. Prilog

Prilog A

A.1 Kronecker, Hadamard i Khatri-Rao proizvod

Za opis multidimensionalnih modela nisu dovoljni standardni matrični proizvodi. Stoga će ovde biti objašnjeno još tri tipa matričnog proizvoda: Kronecker (\otimes), Hadamard ($*$) i Khatri-Rao (\odot).

Kronecker-ov proizvod dve matrice \mathbf{A} ($I \times J$) i \mathbf{B} ($K \times M$) je definisan na sledeći način

$$\mathbf{A} \otimes \mathbf{B} = \begin{bmatrix} a_{11}\mathbf{B} & \dots & a_{1J}\mathbf{B} \\ \vdots & \ddots & \vdots \\ a_{I1}\mathbf{B} & \dots & a_{IJ}\mathbf{B} \end{bmatrix} \quad (\text{A.82})$$

iz čega sledi da je veličina rezultujuće matrice $\mathbf{A} \otimes \mathbf{B}$ ($IK \times JM$) i da Kronecker-ov proizvod je definisan čak i za slučajeve gde standardni matrični proizvod nije moguće primeniti (ako je $J \neq K$).

Primer A.1

Kao jednostavan primer Kronecker-ovog proizvoda dat je proizvode sledeće dve matrice

$$\mathbf{A} = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}; \quad \mathbf{B} = \begin{bmatrix} 0 & 7 \\ -1 & 8 \end{bmatrix}$$

$$\mathbf{A} \otimes \mathbf{B} = \begin{bmatrix} 1\mathbf{B} & 2\mathbf{B} \\ 3\mathbf{B} & 4\mathbf{B} \end{bmatrix} = \begin{bmatrix} 0 & 7 & 0 & 14 \\ -1 & 8 & -2 & 16 \\ 0 & 21 & 0 & 28 \\ -3 & 24 & -4 & 32 \end{bmatrix}$$

Za Kronecker-ov proizvod važe neki osnovni zakoni proizvoda:

$$\begin{aligned} \mathbf{A} \otimes \mathbf{B} \otimes \mathbf{C} &= (\mathbf{A} \otimes \mathbf{B}) \otimes \mathbf{C} = \mathbf{A} \otimes (\mathbf{B} \otimes \mathbf{C}) \\ (\mathbf{A} + \mathbf{B}) \otimes (\mathbf{C} + \mathbf{D}) &= \mathbf{A} \otimes \mathbf{C} + \mathbf{A} \otimes \mathbf{D} + \mathbf{B} \otimes \mathbf{C} + \mathbf{B} \otimes \mathbf{D} \end{aligned} \quad (\text{A.83})$$

ukoliko $\mathbf{A} + \mathbf{C}$ i $\mathbf{C} + \mathbf{D}$ postoje i

$$(\mathbf{A} \otimes \mathbf{B})(\mathbf{C} \otimes \mathbf{D}) = \mathbf{AC} \otimes \mathbf{BD} \quad (\text{A.84})$$

Ukoliko \mathbf{AC} i \mathbf{BD} postoje. Još neke korisne osobine Kronecker-ovog proizvoda su:

- a) $a \otimes \mathbf{A} = a\mathbf{A} = \mathbf{A}a = \mathbf{A} \otimes a$; gde a je skalar
- b) $(\mathbf{A} \otimes \mathbf{B})' = \mathbf{A}' \otimes \mathbf{B}'$
- c) $\mathbf{a}' \otimes \mathbf{b} = \mathbf{ba}' = \mathbf{b} \otimes \mathbf{a}'$
- d) $\text{tr}(\mathbf{A} \otimes \mathbf{B}) = \text{tr}(\mathbf{A})\text{tr}(\mathbf{B})$; za kvadratne matrice \mathbf{A} i \mathbf{B} (A.85)
- e) $(\mathbf{A} \otimes \mathbf{B})^{-1} = \mathbf{A}^{-1} \otimes \mathbf{B}^{-1}$; ukoliko \mathbf{A} i \mathbf{B} nisu singularne
- f) $r(\mathbf{A} \otimes \mathbf{B}) = r(\mathbf{A})r(\mathbf{B})$
- g) $(\mathbf{A} \otimes \mathbf{B})^+ = \mathbf{A}^+ \otimes \mathbf{B}^+$

Drugi proizvod koji se dosta koristi u višedimenzionalnoj analizi jeste Hadamard-ov proizvod koji je za dve matrice $\mathbf{A} (I \times J)$ i $\mathbf{B} (K \times M)$

$$\mathbf{A} * \mathbf{B} = \begin{bmatrix} a_{11}b_{11} & \dots & a_{1J}b_{1J} \\ \vdots & \ddots & \vdots \\ a_{I1}b_{11} & \dots & a_{IJ}b_{IJ} \end{bmatrix} \quad (\text{A.86})$$

Gde su a_{ij} i b_{ij} elementi matrica \mathbf{A} i \mathbf{B} , respektivno. Stoga se može zaključiti da je Hadamard-ov proizvod proizvod element po element.

Primer A.2

Za primer koji je prethodno korišćen za Kronecker-ov proizvod, Hadamard proizvod matrica \mathbf{A} i \mathbf{B} je

$$\mathbf{A} * \mathbf{B} = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} * \begin{bmatrix} 0 & 7 \\ -1 & 8 \end{bmatrix} = \begin{bmatrix} 0 & 14 \\ -3 & 32 \end{bmatrix}$$

Neke osobine Hadamard-ovog proizvoda su:

- a) $\mathbf{A} * \mathbf{B} = \mathbf{B} * \mathbf{A}$
- b) $(\mathbf{A} * \mathbf{B})' = \mathbf{A}' * \mathbf{B}'$
- c) $(\mathbf{A} * \mathbf{B}) * \mathbf{C} = \mathbf{A} * (\mathbf{B} * \mathbf{C})$ (A.87)
- d) $(\mathbf{A} + \mathbf{B}) * (\mathbf{C} + \mathbf{D}) = \mathbf{A} * \mathbf{C} + \mathbf{A} * \mathbf{D} + \mathbf{B} * \mathbf{C} + \mathbf{B} * \mathbf{D}$
- e) $\mathbf{A} * \mathbf{I} = \text{diag}(a_{11}, \dots, a_{nn})$

Treći proizvod koji je koristan za trodimenzionalnu analizu je Khatri-Rao proizvod koji je definisan na sledeći način:

$$\begin{aligned} \mathbf{A} &= [\mathbf{A}_1 \dots \mathbf{A}_K] \text{ i } \mathbf{B} = [\mathbf{B}_1 \dots \mathbf{B}_K] \\ \mathbf{A} \odot \mathbf{B} &= [\mathbf{A}_1 \otimes \mathbf{B}_1 \dots \mathbf{A}_K \otimes \mathbf{B}_K] \end{aligned} \quad (\text{A.88})$$

gde su \mathbf{A} i \mathbf{B} matrice podeljene na jednak broj delova odnosno matrica. Neke korisne osobine Khatri-Rao proizvoda su:

$$\begin{aligned} (\mathbf{A} \odot \mathbf{B}) \odot \mathbf{C} &= \mathbf{A} \odot (\mathbf{B} \odot \mathbf{C}) \\ (\mathbf{T}_1 \otimes \mathbf{T}_2)(\mathbf{A} \odot \mathbf{B}) &= \mathbf{T}_1 \mathbf{A} \odot \mathbf{T}_2 \mathbf{B} \end{aligned} \quad (\text{A.89})$$

Ukoliko su \mathbf{A} i \mathbf{B} podeljeni na njihovih K kolona, onda je to specijalan slučaj za koji dodatno važi

$$(\mathbf{A} \odot \mathbf{B})'(\mathbf{A} \odot \mathbf{B}) = (\mathbf{A}'\mathbf{A}) * (\mathbf{B}'\mathbf{B}) \quad (\text{A.90})$$

Primer A.3

Isti primer korišćen za Kronecker i Hadamard proizvod, U slučaju Khatri-Rao proizvoda je prikazan na sledeći način

$$\mathbf{A} \odot \mathbf{B} = \left[\begin{pmatrix} 1 \\ 3 \end{pmatrix} \otimes \begin{pmatrix} 0 \\ -1 \end{pmatrix} \otimes \begin{pmatrix} 7 \\ 8 \end{pmatrix} \right] = \begin{bmatrix} 0 & 14 \\ -1 & 16 \\ 0 & 28 \\ -3 & 32 \end{bmatrix}$$

A.2 Koncept linearnosti modela

Pretpostavimo da nam treba model koji će povezati parametre x_i (prediktor) i y_i (rezultat), gde i indeksira objekte i ima vrednosti od 1 do I . Takav model je linearan po parametrima i po varijablama ukoliko se može napisati na sledeći način

$$y_i = b_0 + b_1 x_i + e_i ; i = 1, \dots, I \quad (\text{A.91})$$

što predstavlja klasičan regresioni model sa presekom (b_0) i greškom (e_i). Primer modela koji je linearan po parametrima, ali je nelinearan po x je

$$y_i = b_0 + b_1 x_i + b_2 x_i^2 + e_i ; i = 1, \dots, I \quad (\text{A.92})$$

Koncept linearnosti se može proširiti na bilinearnost i trilinearnost. Ako nam treba model sa dva parametra za opis matrice \mathbf{X} ($I \times J$) sa elementima x_{ij} , onda se jedan takav dvokomponentni model može napisati na sledeći način

$$x_{ij} = t_{i1} p_{j1} + t_{i2} p_{j2} + e_{ij} ; i = 1, \dots, I ; j = 1, \dots, J \quad (\text{A.93})$$

Uočava se da modeliran deo matrice \mathbf{X} je linearan kako po t_{i1}, t_{i2} (za fiksirane p_{j1}, p_{j2}) tako i po p_{j1}, p_{j2} (za fiksirane t_{i1}, t_{i2}). Stoga se izraz (A.12) naziva bilinearan model matrice \mathbf{X} .

Slično prethodnom primeru može se prikazati i koncept trilinearnosti na modelu trodimenzionalne matrice $\underline{\mathbf{X}}$ ($I \times J \times K$) sa elementima matrice x_{ijk} :

$$x_{ijk} = a_{i1} b_{j1} c_{k1} + a_{i2} b_{j2} c_{k2} + e_{ijk} ; i = 1, \dots, I ; j = 1, \dots, J ; \\ k = 1, \dots, K \quad (\text{A.94})$$

gde je matrica $\underline{\mathbf{X}}$ ($I \times J \times K$) modelirana sa dve komponente. Modelirani deo x_{ijk} je trilinearan po parametrima a, b i c . Samim tim se izraz (A.13) može nazvati trilinearnim modelom matrice $\underline{\mathbf{X}}$, što je u stvari PARAFAC model matrice $\underline{\mathbf{X}}$.

A.3 Rang i k-Rang matrice

Rang matrice je broj r jednak najvećem redu kvadratne regularne podmatrice \mathbf{A} pri čemu važi da je $r \leq \min(m, n)$ gde su m i n broj redova i kolona, respektivno. Postoji više načina određivanja ranga matrice, u Primeru A.4 je prikazan jedan od načina.

Primer A.4

Određivanje ranga matrice i ranga kolona matrice se može jednostavno ilustrovati primerom. Ako imamo matricu

$$\mathbf{A} = \begin{bmatrix} 1 & 4 \\ 2 & 5 \\ 3 & 6 \end{bmatrix}$$

onda \mathbf{A} ima dve nezavisne kolone. Stoga je rang kolona matrice jednak 2. Uz pomoć matrice \mathbf{A} definiše se linearna transformacija za dobijanje opsega $\mathfrak{R}(\mathbf{A}) = \mathbf{Ax} = \mathbf{y}$ i primenom

$$\mathbf{Ax} = \begin{bmatrix} 1 & 4 \\ 2 & 5 \\ 3 & 6 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = x_1 \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} + x_2 \begin{bmatrix} 4 \\ 5 \\ 6 \end{bmatrix} = \mathbf{y}$$

kolone matrice \mathbf{A} formiraju bazu za $\mathfrak{R}(\mathbf{A})$. Stoga $\mathfrak{R}(\mathbf{A})$ je dvodimenzionalna i rang matrice je samim tim 2. Primer matrice koja ima rang jedan je matrica

$$\mathbf{A} = \begin{bmatrix} 1 & 3 \\ 2 & 6 \end{bmatrix}$$

Linearna transformacija ove matrice može biti napisana kao

$$\mathbf{Ax} = \begin{bmatrix} 1 & 3 \\ 2 & 6 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = x_1 \begin{bmatrix} 1 \\ 2 \end{bmatrix} + x_2 \begin{bmatrix} 3 \\ 6 \end{bmatrix} = (x_1 + 3x_2) \begin{bmatrix} 1 \\ 2 \end{bmatrix} = \mathbf{y}$$

samim tim je $\mathfrak{R}(\mathbf{A})$ jednodimenzionalan i rang matrice \mathbf{A} je jedan.

k -Rang matrice je koristan koncept pri odlučivanju da li je PARAFAC dekompozicija određene matrice jedinstvena. Ako uzmemo u obzir matricu \mathbf{X} ($I \times J$) koja ima rang R . Samim tim \mathbf{X} ima određeni set od R nezavisnih kolona. Međutim neki drugi set R kolona matrice \mathbf{X} može da bude nezavistan. Pretpostavimo da je svaki set R kolona nezavistan i nazovimo ovu osobinu nezavisnost univerzalnih R -kolona. Očigledno je da ako je matrica \mathbf{X} nezavisna univerzalnih R -kolona onda je $r(\mathbf{X}) \geq R$. Najveća celobrojna vrednost k za koju je matrica \mathbf{X} nezavisna univerzalnih k -kolona se naziva k -Rang matrice \mathbf{X} i označava se sa $k_{\mathbf{X}}$.

Primer A.5

Primeri koncepta k -Ranga su:

$$\mathbf{A} = \begin{bmatrix} 1 & 1 & 2 & 5 \\ 1 & 1 & 3 & 3 \\ 1 & 1 & 4 & 2 \end{bmatrix}; r(\mathbf{A}) = 3, k_{\mathbf{A}} = 1$$

$$\mathbf{B} = \begin{bmatrix} 1 & 2 & 3 \\ 2 & 3 & 4 \\ 3 & 4 & 5 \end{bmatrix}; r(\mathbf{B}) = 3, k_{\mathbf{B}} = 3$$

$$\mathbf{C} = \begin{bmatrix} 1 & 2 & 3 \\ 2 & 3 & 5 \\ 3 & 4 & 7 \end{bmatrix}; r(\mathbf{C}) = 2, k_{\mathbf{C}} = 2$$

Matrica \mathbf{A} ima rang tri. Svi mogući podsetovi od po dve kolone nemaju rang dva, stoga samo svi podsetovi od po jedne kolone imaju puni rang i samim tim se dobija da je $k_{\mathbf{A}} = 1$. Matrica \mathbf{B} ima rang tri. Svi podsetovi od po dve kolone imaju rang dva i postoji samo jedan podset od tri kolone odnosno cela matrica i zato je k -Rang matrice \mathbf{B} jednak 3. Matrica \mathbf{C} ima rang 2 jer je treća kolona suma prve i druge kolone. Svi podsetovi od po dve kolone imaju rang 2 odnosno pun rang. Podset od tri kolone

nema pun rang stoga je k -Rang matrice C jednak 2.

A.4 Singularna dekompozicija matrice

Neku su m i n ($m \geq n$) prirodni brojevi, a X proizvoljna realna matrica veličine $m \times n$ i neka je r rang matrice. Tada postoji dekompozicija

$$X = U\Sigma V^T \quad (A.95)$$

Gde je U ortonormalna matrica veličine $m \times m$ i V ortogonalna $n \times n$ matrica, dok je $\Sigma = \text{diag}(\sigma_1, \sigma_2, \dots, \sigma_r)$; $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_r \geq 0$ dijagonalna kvadratna matrica dimenzije $r \times r$ sa pozitivnim singularnim vrednostima. Kolone matrice U nazivamo levi singularni vektori, a kolone matrice V desni singularni vektori. Osnovna struktura singularne dekompozicije matrice se može predstaviti i kao suma

$$X = \sigma_1 u_1 v'_1 + \sigma_2 u_2 v'_2 + \dots + \sigma_r u_r v'_r = \sum_{i=1}^r \sigma_i u_i v'_i \quad (A.96)$$

Iz čega sledi da matrica X , ranga r , može da se predstavi kao linearna kombinacija r matrica $(u_i v'_i)$, ranga 1.

Neka bitna svojstva singularne dekompozicije matrice su:

- 1) Neka je $X \in \mathbf{C}^{m \times n}$, njena inverzna matrica je jedinstvena matrica X^+ , takva da važi

$$\begin{aligned} XX^+X &= X, & X^+XX^+ &= X^+ \\ (XX^+)^* &= XX^+, & (X^+X)^* &= X^+X. \end{aligned}$$

- 2) Neka je $m \geq n$ i $X = U\Sigma V^T$ singularna dekompozicija matrice X

- a) Ako je X simetrična matrica sa svojstvenim vektorima q_i , tj. $X = Q\Lambda Q^T$, pri čemu je $Q = [q_1, \dots, q_n]$, $\Lambda = \text{diag}(\lambda_1, \dots, \lambda_n)$. Tada je $X = U\Sigma V^T$ za $u_i = q_i$, $\sigma_i = |\lambda_i|$ i $v_i = \text{sign}(\lambda_i)q_i$.
- b) Svojstvene vrednosti $X^T X$ su σ_i^2 . Desni singularni vektori v_i su odgovarajući ortonormirani svojstveni vektori.

- c) Svojevredne vrednosti $\mathbf{X}\mathbf{X}^T$ su σ_i^2 i $m-n$ nula. Levi singularni vektori u_i su odgovarajući ortonormirani svojstveni vektori za svojevredne vrednosti σ_i^2 .
- d) Ukoliko \mathbf{X} ima pun rang, a b je proizvoljan vektor, onda je $x = \mathbf{X}^+b = \mathbf{V}\mathbf{\Sigma}^{-1}\mathbf{U}^Tb$ rešenje problema $\min_x \|\mathbf{X}x - b\|_2$.
- e) Važi da je $\|\mathbf{X}\|_2 = \sigma_1$. Ukoliko postoji \mathbf{X}^{-1} onda je $\|\mathbf{X}^{-1}\|_2 = 1/\sigma_n$.
- f) Neka je $\mathbf{X}_k = \mathbf{U}\mathbf{\Sigma}_k\mathbf{V}^T = \sum_{i=1}^k \sigma_i \mathbf{u}_i \mathbf{v}'_i$, gde je $\mathbf{\Sigma} = \text{diag}(\sigma_1, \sigma_2, \dots, \sigma_k, 0, \dots, 0)$. Tada matrica \mathbf{X}_k ima rang k te je ona najbliža matrici \mathbf{X} među svim matricama ranga k :

$$\|\mathbf{X} - \mathbf{X}_k\|_2 = \min_{\text{rang}(B)=k} \|\mathbf{X} - B\|_2$$

Takođe važi da je $\|\mathbf{X} - \mathbf{X}_k\|_2 = \sigma_{k+1}$.

- g) Pretpostavimo da za singularne vrednosti matrice \mathbf{X} važi $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_r \geq \sigma_{r+1} = \dots = \sigma_n = 0$, tada je $\text{rang}(\mathbf{X}) = r$.

Primer A.6

Singularna dekompozicija matrice veličine 3×2 .

$$\mathbf{X} = \begin{bmatrix} 1 & 2 \\ 2 & 3 \\ 3 & 4 \end{bmatrix} = \begin{bmatrix} -0.3381 & 0.8480 & 0.4082 \\ -0.5506 & 0.1735 & -0.8165 \\ -0.7632 & -0.5009 & 0.4082 \end{bmatrix} \begin{bmatrix} 6.5468 & 0 \\ 0 & 0.3742 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} -0.5696 & -0.8219 \\ -0.8219 & 0.5696 \end{bmatrix}$$

8. Biografija

Lea Lenhardt rođena je 2.5.1986. u Beogradu, gde je završila osnovnu školu i gimnaziju. Školske 2005/2006. godine upisuje osnovne studije na Mašinskom fakultetu Univerziteta u Beogradu. Osnovne akademske studije završava 2008. godine sa diplomskom radom pod nazivom "Senzori za detekciju raka". Iste godine upisuje master studije na Mašinskom fakultetu Univerziteta u Beogradu. 2010. godine završava master studije sa master radom pod nazivom "Primena fluorescentne spektroskopije u karakterizaciji i klasifikaciji biomolekula". 2010. upisuje prvu godinu doktorskih studija na Fizičkom fakultetu Univerziteta u Beogradu na studijskom programu – Primenjena i kompjuterska fizika. Od maja 2010. godine je zaposlena u Institutu za nuklearne nauke „Vinča“ u laboratoriji za radijacionu hemiju i fiziku "Gama". Od tada osnovni predmet istraživanja Lee Lenhardt je primena optičkih spektroskopskih metoda i multivarijantnih statističkih tehnika u biologiji i medicini.

Kandidat je koautor 6 naučnih radova u međunarodnim časopisima sa SCI liste. Objavila je značajan broj saopštenja iznetih na domaćim i međunarodnim konferencijama štampanim u celini u izvodu. Radovi korišćeni za tezu su pod rednim brojevima 1, 3 i 5.

Spisak naučnih radova:

1. Lenhardt L., Zeković I., Dramićanin T., Tešić Ž., Milojković-Opsenica D. & Dramićanin M.D. Authentication of the botanical origin of honey by infrared spectroscopy coupled with support vector machine algorithm. *Physica Scripta*, (2014) T162: 014042.
2. Zeković I., Lenhardt L., Dramićanin T., Bandić J. & Dramićanin M.D., Discrimination among melanoma, nevi and normal skin by using synchronous luminescence spectroscopy. *Applied Spectroscopy* (2014) 68(8): 823-830.

3. Lenhardt L., Zeković I., Dramićanin T., Dramićanin M.D. & Bro R. Determination of the botanical origin of honey by front-face synchronous fluorescence spectroscopy. *Applied Spectroscopy* (2014) 68(5): 557-563
4. Lenhardt L., Zeković I., Dramićanin T. & Dramićanin M.D. Artificial neural network for processing fluorescence spectroscopy data in skin cancer diagnostics. *Physica Scripta* (2013) T157: 014057.
5. Dramićanin T., Lenhardt L., Zeković I. & Dramićanin M.D. Support Vector Machine on fluorescence landscapes for breast cancer diagnostics. *Journal of Fluorescence* (2012) 22: 1281-1289.
6. Zeković I., Lenhardt L., Dramićanin T. & Dramićanin M.D. Classification of intact cereal flours by front-face synchronous fluorescence spectroscopy. *Food Analytical Methods* (2012) 5: 1205-1213.

Изјава о ауторству

Потписани-а _____ Леа Ленхардт _____

број индекса _____ D-5/2010 _____

Изјављујем

да је докторска дисертација под насловом

Паралелна факторска анализа флуоресцентних својстава вишекомпонентних

система

- резултат сопственог истраживачког рада,
- да предложена дисертација у целини ни у деловима није била предложена за добијање било које дипломе према студијским програмима других високошколских установа,
- да су резултати коректно наведени и
- да нисам кршио/ла ауторска права и користио интелектуалну својину других лица.

Потпис докторанта

У Београду, _____

Изјава о истоветности штампане и електронске верзије докторског рада

Име и презиме аутора _____ Леа Ленхардт _____

Број индекса _____ D-5/2010 _____

Студијски програм _____ Примењена физика _____

Наслов рада _____ Паралелна факторска анализа флуоресцентних својстава
вишекомпонентних система _____

Ментор _____ Проф др Мирослав Драмићанин _____

Потписани/а _____ Леа Ленхардт _____

Изјављујем да је штампана верзија мог докторског рада истоветна електронској верзији коју сам предао/ла за објављивање на порталу **Дигиталног репозиторијума Универзитета у Београду**.

Дозвољавам да се објаве моји лични подаци везани за добијање академског звања доктора наука, као што су име и презиме, година и место рођења и датум одбране рада.

Ови лични подаци могу се објавити на мрежним страницама дигиталне библиотеке, у електронском каталогу и у публикацијама Универзитета у Београду.

Потпис докторанта

У Београду, _____

Изјава о коришћењу

Овлашћујем Универзитетску библиотеку „Светозар Марковић“ да у Дигитални репозиторијум Универзитета у Београду унесе моју докторску дисертацију под насловом:

ПАРАЛЕЛНА ФАКТОРСКА АНАЛИЗА ФЛУОРЕСЦЕНТНИХ СВОЈСТАВА
ВИШЕКОМПОНЕНТНИХ СИСТЕМА

која је моје ауторско дело.

Дисертацију са свим прилозима предао/ла сам у електронском формату погодном за трајно архивирање.

Моју докторску дисертацију похрањену у Дигитални репозиторијум Универзитета у Београду могу да користе сви који поштују одредбе садржане у одабраном типу лиценце Креативне заједнице (Creative Commons) за коју сам се одлучио/ла.

1. Ауторство
2. Ауторство - некомерцијално
3. Ауторство – некомерцијално – без прераде
4. Ауторство – некомерцијално – делити под истим условима
5. Ауторство – без прераде
6. Ауторство – делити под истим условима

(Молимо да заокружите само једну од шест понуђених лиценци, кратак опис лиценци дат је на полеђини листа).

Потпис докторанта

У Београду, _____

1. Ауторство - Дозвољавате умножавање, дистрибуцију и јавно саопштавање дела, и прераде, ако се наведе име аутора на начин одређен од стране аутора или даваоца лиценце, чак и у комерцијалне сврхе. Ово је најслободнија од свих лиценци.
2. Ауторство – некомерцијално. Дозвољавате умножавање, дистрибуцију и јавно саопштавање дела, и прераде, ако се наведе име аутора на начин одређен од стране аутора или даваоца лиценце. Ова лиценца не дозвољава комерцијалну употребу дела.
3. Ауторство - некомерцијално – без прераде. Дозвољавате умножавање, дистрибуцију и јавно саопштавање дела, без промена, преобликовања или употребе дела у свом делу, ако се наведе име аутора на начин одређен од стране аутора или даваоца лиценце. Ова лиценца не дозвољава комерцијалну употребу дела. У односу на све остале лиценце, овом лиценцом се ограничава највећи обим права коришћења дела.
4. Ауторство - некомерцијално – делити под истим условима. Дозвољавате умножавање, дистрибуцију и јавно саопштавање дела, и прераде, ако се наведе име аутора на начин одређен од стране аутора или даваоца лиценце и ако се прерада дистрибуира под истом или сличном лиценцом. Ова лиценца не дозвољава комерцијалну употребу дела и прерада.
5. Ауторство – без прераде. Дозвољавате умножавање, дистрибуцију и јавно саопштавање дела, без промена, преобликовања или употребе дела у свом делу, ако се наведе име аутора на начин одређен од стране аутора или даваоца лиценце. Ова лиценца дозвољава комерцијалну употребу дела.
6. Ауторство - делити под истим условима. Дозвољавате умножавање, дистрибуцију и јавно саопштавање дела, и прераде, ако се наведе име аутора на начин одређен од стране аутора или даваоца лиценце и ако се прерада дистрибуира под истом или сличном лиценцом. Ова лиценца дозвољава комерцијалну употребу дела и прерада. Слична је софтверским лиценцама, односно лиценцама отвореног кода.