

Pregled metoda za predobradu spektralnih podataka

Hartmann, Hariett

Undergraduate thesis / Završni rad

2023

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

Permanent link / Trajna poveznica: <https://um.nsk.hr/um:nbn:hr:149:211278>

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

Download date / Datum preuzimanja: **2024-10-02**



Repository / Repozitorij:

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



SVEUČILIŠTE U ZAGREBU
FAKULTET KEMIJSKOG INŽENJERSTVA I TEHNOLOGIJE
SVEUČILIŠNI PRIJEDIPLOMSKI STUDIJ

Hariett Hartmann

ZAVRŠNI RAD

Zagreb, rujan 2023.

SVEUČILIŠTE U ZAGREBU
FAKULTET KEMIJSKOG INŽENJERSTVA I TEHNOLOGIJE
SVEUČILIŠNI PRIJEDIPLOMSKI STUDIJ

Hariett Hartmann

**PREGLED METODA ZA PREDOBRADU SPEKTRALNIH
PODATAKA**

ZAVRŠNI RAD

Mentor: Željka Ujević Andrijić, doc. dr. sc.

Članovi ispitnog povjerenstva:

doc. dr. sc. Željka Ujević Andrijić

prof. dr. sc. Nenad Bolf

doc. dr. sc. Matija Cvetnić

Zagreb, rujan 2023.

Prije svega, željela bih se zahvaliti mentorici, doc. dr. sc. Željki Ujević Andrijić, što mi je, svojim znanjem pomogla u izradi ovog završnog rada. Zahvaljujem joj na svoj strpljivosti koju je uložila u zadnjih par mjeseci. Hvala joj na ogromnom razumijevanju, susretljivosti i fleksibilnosti koje je pokazala. Hvala joj što je svojom stručnošću uvijek bila tu i uvelike mi olakšala stvaranje ovog završnog rada. Vjerujem da je i Vama bilo naporno, ali uspjeli smo!

Veliko hvala asistentici Matei Gavran na korisnim stručnim komentarima i uloženom trudu i vremenu prilikom pisanja cijelog rada, a ponajviše pri primjeni metoda predobrade podataka.

Hvala kolegicama s PBF-a koje su mi pružale ogromno razumijevanje i beskompromisnu podršku, ponekad se zbog vas stvarno pitam jesam li upisala pravi fakultet, ali odgovor je još uvijek DA.

Hvala svim članovima DPS-a što su mi pisanje ovog rada uljepšali predivnim izletima i šarolikom atmosferom i ogromnom vjerom u moj još uvijek upitni životni potencijal.

Hvala određenim zaposlenicima DVS-a na moralnoj podršci, neslanim šalama, finim kavama i vjeri koju ste mi pružali, a još je i sad preispitujem.

Zahvalila bih se i sama sebi iako sam bas razvukla pisanje ovog završnog rada, ali svejedno zaslužujem ogromnu zahvalu, jer sam bila uporna i nisam odustala, možda i jesam napravila jako mali korak, ali to je korak koji meni puno znači.

Hvala i mojim prijateljicama Lari, Dori i Anamariji na svim motivacijskim porukama i porukama utjehe, na svim pozivima i vremenu koje sam provela zabavljajući se s vama umjesto da sam učila! Vrijedilo je.

Ali najveće hvala zaslužuju moja braća koja su mi pružala podršku na neki svoj čudan način, hvala im što su uvijek bili tu i donekle imali razumijevanje za moje učenje. Hvala Ottu što je provodio vrijeme s našim psom na svježem zraku, kada ja za to nisam imala priliku. Hvala Mihovilu što mi je bio velika motivacija i potpora za vježbanje. Hvala Mati što mi je sate učenja uljepšao predivnom glazbom. Samo da vas podsjeti da nas čekaju još minimalno dvije ovakve godine! Znam da jedva čekate.

Hvala mojoj Lepojki što nikada nije preispitivala smjer u kojem idem, ili barem to nikada nije pokazala. Hvala joj što me održavala zdravom sa svojom nevjerojatnom hranom i hvala joj što mi je pružala slobodu da postanem osoba kakva jesam. Hvala joj što u rukama drži i svoje i moje konce! (I možeš javiti Brižnoj Tetki da sam jedan korak bliže njenoj definiciji uspjeha, možda rođendane ne slavim, ali ovo ćemo proslaviti.)

Hvala i tebi Tata za sve ono što je bilo i za sve ono što ću biti.

SADRŽAJ

1. UVOD	1
2. TEORIJSKI DIO	2
2.1. Razlozi predobrade podataka	2
2.2. Osnove o spektralnim podacima	3
2.2.1. Oblici vrhova	3
2.2.2. Upotreba informacija o obliku spektralnog vrha	6
2.3. Šum	7
2.3.1. Stacionarni šum	8
2.3.1.1. Homoskedastički šum	8
2.3.1.2. Heteroskedastički šum	8
2.3.2. Korelirani šum	9
2.3.3. Omjer signala i šuma	9
3. RAMANOVA SPEKTROSKOPIJA	11
3.1. O Ramanovoj i IR spektroskopiji	11
3.2. Predobrada podataka u Ramanovoj spektroskopiji	13
3.2.1. Usklađivanje spektralne osi	13
3.2.2. Uklanjanje kozmičkih zraka	13
3.2.3. Korekcija pozadine	14
3.2.3.1. Metoda temeljena na pokretnom medijan prozoru	16
3.2.3.2. Metoda temeljena na prilagodbi polinomom	17
3.2.3.3. Metoda temeljena na asimetričnoj metodi najmanjih kvadrata	17
3.2.3.4. Metoda temeljena na analizi frekvencijske domene	17
3.2.4. Izgladivanje podataka	18
3.2.4.1. Metoda pomičnog prosjeka	19
3.2.4.2. Hanning i Hamming filtri s pomičnim prozorom	20
3.2.4.3. Savitzky-Golay filtri	22
3.2.4.4. Zagladivanje podataka medijanom	24
3.2.5. Derivacije	25
3.2.5.1. Metoda temeljena na derivaciji	27
3.2.6. Konvolucija	31
3.2.7. Normalizacija Raman podataka	31
3.2.7.1. Vektorska normalizacija ili Euklidska norma	32

3.2.7.2.	Min-max normalizacija	32
3.2.7.3.	Normalizacija standardno normalnom varijacijom (SNV).....	33
3.2.7.4.	Normalizacija vrhova	34
4.	PREDOBRADA PODATAKA U FTIR SPEKTROSKOPIJI	35
4.1.	Korekcija pozadine	35
4.2.	Korekcija bazne linije	35
4.3.	Normalizacija FTIR podataka	37
4.4.	Multiplikativna korekcija signala (MSC)	37
4.5.	Metode Fourierove transformacije.....	41
4.6.	Utjecaj šuma	45
5.	REZULTATI	48
5.1.	Grafički prikazi primjene metoda predobrade podataka	48
5.1.1.	Primjene 1. i 2. derivacije spektra Savitzky-Golay filtrom	51
5.1.2.	Predobrada spektara svih mjerenih koncentracija djelatne tvari.....	53
5.1.3.	Primjena kombinacije različitih metoda predobrade spektara	54
6.	ZAKLJUČAK	57
7.	POPIS KRATICA	58
8.	LITERATURA	59

SAŽETAK

Metode vibracijske spektroskopije - spektroskopija bliskoga infracrvenoga zračenja (NIR) i Ramanova spektroskopija postale su vrijedan analitički alat u raznim područjima kako u znanosti tako i industriji. Razlog tomu je njihova sposobnost pružanja jedinstvenih informacija o molekularnoj strukturi, mogućnost beskontaktnog mjerenja, mogućnost analize u stvarnom vremenu, sposobnost neinvazivne analize složenih sustava, minimalna interferencija s vodom te komplementarnost primjene s drugim tehnikama. Međutim, Ramanovi spektri mogu sadržavati i negativne utjecaje izvan Ramanova spektra kao što su fluorescentna pozadina, šum i kozmičke zrake, koje se moraju ukloniti kako bi se osigurala pouzdanost mjerenja.

Da bi se prevladali izazovi interpretacije složenih spektara dobivenih Ramanovom i IR spektroskopijom i omogućio dublji uvid u podatke moraju se primijeniti različite metode predobrade podataka.

Ovaj završni rad daje teorijski pregled te analizu primjene metoda predobrade podataka u Ramanovoj i IR spektroskopiji poput korekcije bazne linije, normalizacije podataka, smanjenja šuma u podacima, detekcije odstupanja i ekstrakcije bitnih karakteristika iz spektra.

Uz pregled metoda i primjere iz literature u ovom završnom radu je na primjeru laboratorijskih podataka iz procesa kristalizacije djelatne tvari istražen utjecaj odabira različitih parametara i redoslijeda predobrade podataka na ukupnu kvalitetu i pouzdanost obrađenih spektralnih podataka.

Ključne riječi: vibracijska spektroskopija, Ramanova spektroskopija, IR spektroskopija, šumovi u spektru, metode predobrade spektralnih podataka, digitalni filtri, obrada signala

SUMMARY

Vibrational spectroscopy methods - near-infrared (NIR) spectroscopy and Raman spectroscopy - have become valuable analytical tools in various fields in both science and industry. This is due to their ability to provide unique information about molecular structure, the possibility of non-contact measurement, the possibility of real-time analysis, the ability of non-invasive analysis of complex systems, the minimal interference with water, and the complementarity of application with other techniques. However, Raman spectra contain additional negative influences beyond the Raman spectrum, such as fluorescent background, noise, and cosmic rays, which must be removed to ensure measurement reliability.

To overcome the challenges of interpreting complex spectra obtained by Raman and IR spectroscopy and to provide deeper insight into the data, various methods of data preprocessing must be applied.

This final thesis provides a theoretical overview and analysis of the application of data preprocessing methods in Raman and IR spectroscopy, such as baseline correction, data normalization, noise reduction, outlier detection and extraction of essential characteristics from the spectrum.

In this thesis, in addition to a method overview and examples from the literature, the influence of the selection of different parameter and data preprocessing sequences on the overall quality and reliability of the processed spectral data is investigated using laboratory data from the crystallization process of the active ingredients.

Key words: vibrational spectroscopy, Raman spectroscopy, IR spectroscopy, noise in the spectrum, spectral data preprocessing methods, digital filters, signal processing

1. UVOD

Matematičke metode predobrade podataka izuzetno su korisne tehnike za početnu obradu „sirovih“ spektralnih podataka. Cilj predobrade spektralnih podataka je ukloniti nedostatke uzrokovane raspršenjem svjetlosti, ukloniti fizikalne pojave u spektru, ukloniti slučajni šum, odnosno eliminirati ili svesti na najmanju moguću mjeru varijabilnost koja nije povezana sa svojstvom od interesa, a sve kako bi se ispitivani spektri mogli dalje učinkovitije modelirati. Pažljivim odabirom metode predobrade spektara može se znatno poboljšati robusnost i kvaliteta konačnog modela. [1]

Često su spektri iz raznih ispitivanih sustava dosta složeni, a ponekad i najmanje razlike u spektrima mogu sadržavati informacije od ključne važnosti. Interpretacija spektara dobivenih Ramanovom i IR spektroskopijom može biti dosta kompleksna te se da bi se postigla lakoća iščitavanja informacija iz spektra i omogućila dublja analiza podataka, primjenjuju različite kombinacije metoda predobrade podataka.

U predobradi spektara dosta je bitno povećanje razlučivosti (tj. rezolucije) spektara kako bi se izbjeglo zamućenje vrhova. U svrhu povećanja razlučivosti ponekad se spektri prikazuju u derivacijskom obliku kako bi se smanjila širina vrhova i povećala njihova rezolucija. Postoji niz brzih računalnih metoda koje korigiraju pozadinu, tj. uklanjaju baznu liniju. Neke od takvih metoda bit će opisane u ovom radu: metoda pomičnog prosjeka, zaglađivanje podataka medijanom (metoda koja je najučinkovitija pri uklanjanju šiljaka / kozmičkih zraka) i postupci temeljeni na metodi najmanjih kvadrata. Postoje i drugi pristupi izgladivanja podataka koji se temelje na prilagođavanju krivulje, kao npr. Savitzky-Golay algoritam i Fourierovi filtri. [2]

Važno je uspostaviti ravnotežu između različitih koraka predobrade te izbjeći potencijalni rizik od pretjeranog prilagođavanja ili izobličenja originalnih podataka kako se ne bi izgubile bitne informacije iz spektra. U ovom radu istražit će se utjecaj odabira različitih parametara predobrade spektara, redoslijeda koraka predobrade i tehnika normalizacije na ukupnu kvalitetu obrađenih spektralnih podataka. Pravilnom primjenom tehnika predobrade, može se iskoristiti puni potencijal spektralnih podataka, što dovodi do dubljih saznanja o proučavanim fenomenima.

2. TEORIJSKI DIO

2.1. Razlozi predobrade podataka

Cilj predobrade spektralnih podataka je poboljšati linearni odnos između spektralnih signala i koncentracije analita tako da se signali bolje pridržavaju Beerovom zakonu. Mnogi fizikalni i kemijski fenomeni mogu uzrokovati raspršenje odnosno odstupanje podataka od linearnog odnosa. Raspršenje u podacima može se pojaviti zbog prisutnosti čestica, interferencije, molekularnih interakcija, promjena u indeksu loma pri visokim koncentracijama, pomaka u kemijskoj ravnoteži kao funkciji koncentracije, prisutnosti stranog svjetla, promjena u veličini uzorka, itd. [3]

Predobradom spektra eliminiraju se neželjeni dijelovi signala kao što su fluorescencija, Miejevo raspršenje, šum detektora, pogreška kalibracije, kozmičke zrake, fluktuacije snage lasera, signali iz staničnih medija ili staklene podloge. Predobradom podataka intenziviraju se suptilne razlike između uzoraka što može biti od velike važnosti prilikom analize. [4]

Metode predobrade spektralnih podataka mogu se podijeliti u dvije velike skupine: tehnike koje koriste dostupne referentne vrijednosti namijenjene za postupak predobrade i one koje to ne čine. Referentno neovisne tehnike predobrade omogućuju primjenu općenitijih alata u istraživanjima kada nije dostupna referentna vrijednost. One se dalje mogu podijeliti u dvije podskupine: metode korekcije raspršenosti i metode spektralnih derivacija.

Raspršenost spektralnih podataka manifestira se kao pomak bazne linije spektra, koji može biti aditivan ili multiplikativan, ovisno o karakteristikama uzorka i fizičkoj interakciji uzorka sa svjetlošću. U metode korekcije raspršenosti ubrajaju se:

- multiplikativna korekcija signala (MSC), poznata i kao multiplikativna korekcija raspršenosti (uključujući prošireni MSC, inverzni MSC, i detrendiranje),
- skaliranje standardne normalne varijacije (SNV),
- normalizacija i korekcija bazne linije.

Derivacijske metode se prvenstveno koriste za preklapanje vrpce u spektru te uklanjanje velikih pomaka bazne linije. Također se koriste za ispravljanje aditivnih i multiplikativnih varijacija bazne linije u spektru. Međutim, pri tome dolazi do pojačavanja šuma, koji se može

ukloniti nekim od algoritama izgladivanja. Izgladivanje spektara uklanja slučajni šum iz spektralnih podataka i poboljšava vizualni aspekt spektra. [1] Derivacijske metode predobrade spektralnih podataka uključuju metode konačnih razlika, Savitzky–Golay i Norris–Williams tehnike. [3]

Budući da se Ramanova i infracrvena spektroskopija temelje na dva različita fenomena, izvori neželjenih doprinosa u neobrađenim spektrima također su različiti te su potrebni različiti koraci prethodne obrade podataka. [4]

Infracrvena spektroskopija proučava interakcije tvari i elektromagnetnog zračenja, a IR-spektar pokazuje ovisnost apsorbancije ili transmitacije o valnoj duljini ili valnom broju dok se Ramanova spektroskopija temelji na neelastičnom (Ramanovu) raspršenju elektromagnetskog zračenja zbog međudjelovanja s uzorkom. [5] Kod svih metoda analize podataka očekuje se da imaju isti Ramanov pomak osi za sve spektre. Dakle, vrlo je važno uskladiti sve spektre (engl. *Spectral axis alignment*) kako bi imali zajedničku spektralnu os. Osim spektralnog usklađivanja, iz Ramanovih spektara treba ukloniti kozmičke zrake (engl. *cosmic ray events*, CRE) prije daljnje predobrade spektralnih podataka. CRE se stvaraju zbog čestica visoke energije koje prolaze kroz naponom povezani uređaj (engl. *charge coupled device*, CCD) i generiraju mnogo elektrona, koje CCD tumači kao signal. Jedan od izazova u korištenju Ramanove spektroskopije kod analiza bioloških uzoraka je jaka intrinzična fluorescencija mnogih biomolekula. Fluorescentna pozadina često je intenzivnija od slabih Ramanovih signala te da bi se izdvojili informativni Ramanovi signali mora se ukloniti fluorescencija. Također, ambijentalno svjetlo i toplinski šum detektora mogu pridonijeti pozadinskom dijelu signala. U tu svrhu koriste se različite metode obrade kao što je prilagodba podataka polinomom (engl. *polynomial fitting*), derivacija prvog i drugog reda, filtriranje frekvencijske domene itd. Fluktuacije izvora/okoliša dodaju neke varijabilnosti spektrima koje nisu povezane sa kemijskim ili strukturnim razlikama u uzorcima. Za nadvladavanje ovih varijacija koriste se različite metode normalizacije. [4]

2.2. Osnove o spektralnim podacima

2.2.1. Oblici vrhova

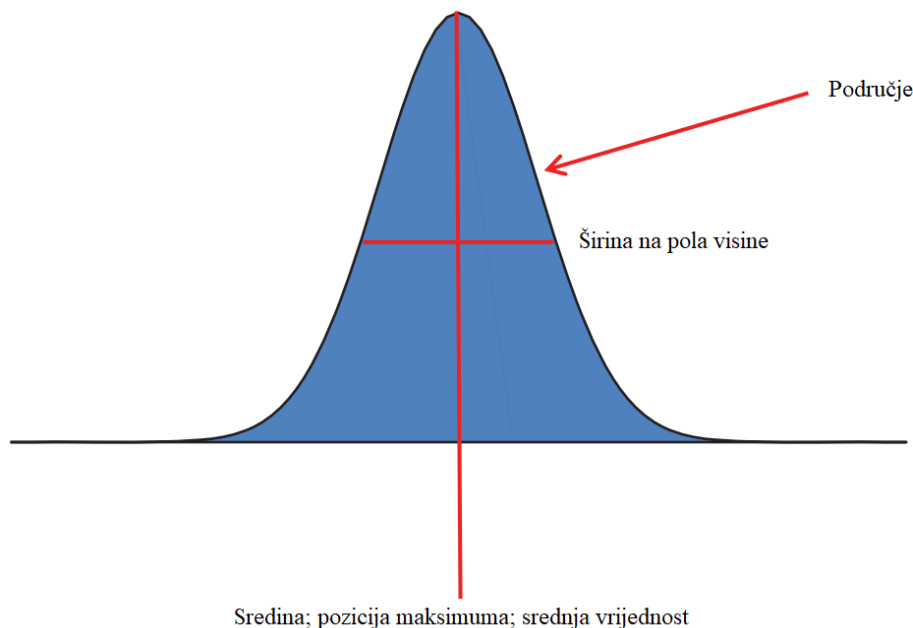
Spektralni podaci se uobičajeno sastoje od niza vrhova. Svaki vrh obično proizlazi iz karakteristične apsorpcije (kao što je kromofor) ili karakterističnog spoja. U većini slučajeva,

temeljni vrhovi su iskrivljeni zbog različitih razloga kao što su šum, zamućenje ili preklapanje sa susjednim vrhovima. Glavni cilj kemometrijskih metoda je dobiti temeljne, neiskrivljene informacije spektra.

Vrhovi se mogu karakterizirati na više načina, ali uobičajeni pristup za simetrične vrhove (**Slika 1**) je karakterizirati svaki vrh spektra pomoću:

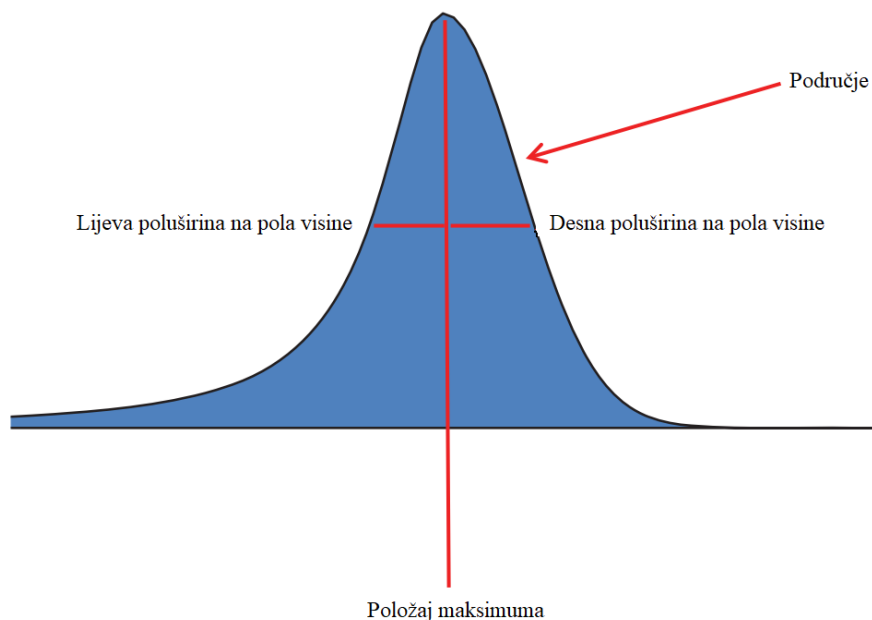
- 1) položaja u središtu (npr. vrijeme ispiranja (engl. *elution time*) ili spektralna frekvencija)
- 2) širine vrha, obično na pola visine
- 3) područja.

Za unimodalne simetrične vrhove položaj u središtu je također i položaj najveće visine i srednje vrijednosti. Odnos između površine i visine vrha ovisi o obliku vrha, iako je visine vrhova često lakše mjeriti. Ako niz vrhova ima isti oblik, tada su omjeri visina proporcionalni omjerima površina. Međutim, površina vrhova je obično bolja mjera kod kemijskih svojstava kao što je koncentracija te je prije svega važno dobiti precizne informacije o oblicima vrhova.



Slika 1. Parametri koji karakteriziraju simetrični vrh spektra

U mnogim spektroskopskim mjerenjima se oblici vrhova mogu vrlo precizno predvidjeti, kao na primjer kod NMR ili vidljive spektroskopije. U drugim situacijama, oblik vrha ovisi o složenim fizičkim procesima, kao u kromatografiji, i može se samo empirijski modelirati. [2]



Slika 2. Parametri koji karakteriziraju asimetrični vrh spektra

Za asimetrične oblike vrhova uobičajeno je koristiti maksimum položaja umjesto srednje vrijednosti i navesti lijevu i desnu poluširinu na polovici visine (**Slika 2**).

U praksi se u većini slučajeva susreće nekoliko uobičajenih oblika vrhova.

Gaussian i Lorentzian oblici spektralnih vrhova - Asimetrični oblici vrhova

Gaussianovi oblici vrhova su prisutni u većini vrsta kromatografije i spektroskopije. Pojednostavljena formula za Gaussian oblik vrha je prikazana jednadžbom (2.1):

$$x_i = A \exp\left(\frac{-(x_i - x_0)^2}{s^2}\right) \quad (2.1)$$

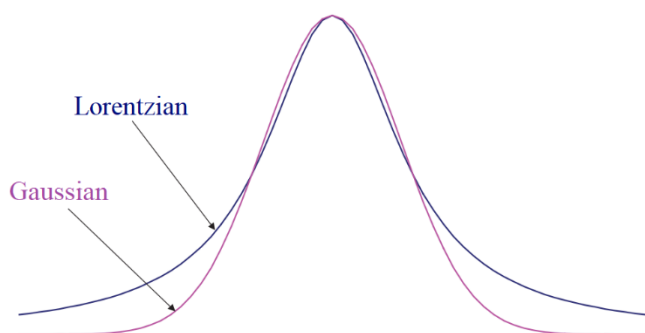
A je visina vrha u središtu, x_0 je položaj središta, s je širina vrha. Gaussianovi oblici se temelje na normalnoj distribuciji gdje x_0 odgovara srednjoj vrijednosti mjerenja, a s standardnoj devijaciji.

Lorentzianovi oblici vrhova odgovaraju statističkoj funkciji koja se naziva Cauchyjeva distribucija. Pojednostavljena formula za Lorentzianov oblik vrha:

$$x_i = A / \left(1 + \frac{(x_i - x_0)^2}{s^2}\right) \quad (2.2)$$

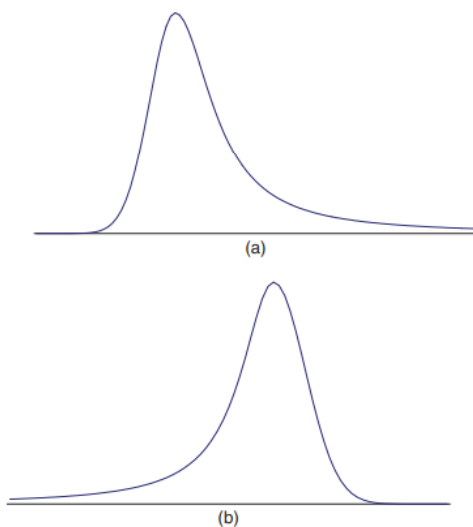
A je visina vrha u središtu, x_0 je položaj središta, s se odnosi na širinu vrha. Lorentzianovi oblici vrhova su rjeđi, ali se često pojavljuju u npr. NMR spektroskopiji. Glavna razlika između

Gaussianovog i Lorentzianovog oblika vrha je da Lorentzianov oblik ima veći rep, kao što prikazuje **Slika 3**, za dva vrha identičnih poluširina i visina.



Slika 3. Gaussian i Lorentzian oblici vrhova sa istim poluvisinama

Kod mnogih kromatografa je teško dobiti simetrične oblike vrhova. Iako je dostupan niz prilično sofisticiranih modela, vrlo jednostavna aproksimacija je Gaussian/Lorentzian oblik vrha. **Slika 4** (a) predstavlja repni (engl. *tailing*) oblik vrha, u kojem se lijeva strana može modelirati Gaussian, a desna strana pomoću Lorentzian oblika. **Slika 4**(b) prikazuje prednji (engl. *fronting*) vrh (takvi su vrhovi mnogo rjeđi). [2]



Slika 4. Asimetrični oblici vrhova. (a) Repni oblik: lijevi je Gaussian, a desni je Lorentzian (b) Prednji oblik (engl. *Fronting*): lijevi je Lorentzian, a desni je Gaussian

2.2.2. Upotreba informacija o obliku spektralnog vrha

Informacije o obliku spektralnog vrha mogu se koristiti na dva glavna načina:

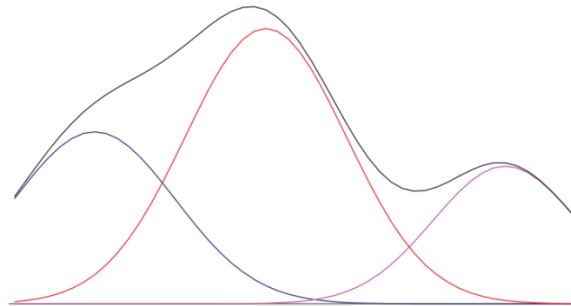
a) Prilagodba krivulje

Postoje različiti računalni algoritmi namijenjeni za prilagodbu podataka spektra krivuljom, a većina njih uključuje neku vrstu minimizacije metodom najmanjih kvadrata. Ako se sumnja (ili zna) da postoje tri vrha u klasteru Gaussova oblika, tada je potrebno pronaći devet parametara, tj. tri položaja vrhova, tri širine vrhova i tri visine vrhova. U bilo kojoj prilagodbi krivulje, važno je utvrditi postoji li određeno predznanje o oblicima vrhova, koliko vrhova postoji i imaju li određene značajke, na primjer, položaj svake komponente. Primjenom metoda prilagodbe krivulje može se dobiti bolji opis sustava; na primjer, poznavanjem područja vrhova, mogu se odrediti relativne koncentracije komponenti u smjesi. Ako ne postoji prethodno znanje o sustavu koji bi se potencijalno mogao sastojati od više vrhova koji se u nekom svom dijelu preklapaju, mnogi algoritmi za prilagođavanje krivulje neće dati zadovoljavajuće rezultate. [2]

b) Simulacije

Simulacije također imaju važnu ulogu u kemometriji. Ako rezultat kemometrijske metode rezultira rekonstrukcijom vrhova koji su bliski stvarnim podacima, tada temeljni oblici vrhova daju dobar opis. Simulacije se također koriste za istraživanje koliko dobro različite tehnike funkcioniraju i pod kojim okolnostima ne funkcioniraju.

U stvarnim analizama spektar se najčešće sastoji od nekoliko vrhova, na različitim položajima, različitih intenziteta i ponekad različitih oblika. **Slika 5** predstavlja skupinu od tri pika, zajedno s njihovim ukupnim intenzitetom. Iako je par vrhova s desne strane vrlo lako vizualno razlučiti, to ne vrijedi za par vrhova s lijeve strane, a bilo bi posebno teško identificirati položaj i intenzitet prvog vrha klastera bez korištenja nekog oblika analize podataka. [2]



Slika 5. *Tri vrha koji tvore klaster*

2.3. Šum

U statistici su priroda i podrijetlo šuma često nepoznati, ali se pretpostavlja da se vladaju po normalnoj distribuciji. Mnogi statistički testovi kao što su t -test i F -test to i pretpostavljaju i predstavljaju samo aproksimacije u nedostatku eksperimentalnih studija distribucije šuma. Međutim, središnji granični teorem sugerira da većina simetričnih distribucija teži normalnoj distribuciji u središtu, što je dobra aproksimacija u nedostatku drugih informacija.

Postoje dva temeljna izvora šuma u instrumentalnim mjerenjima. Prvi izvor šuma uključuje pripremu uzorka, na primjer, razrjeđivanje, vaganje i učinkovitost ekstrakcije i mogu se minimalizirati dobrim analitičkim postupcima. Drugi izvor šuma potječe od mjerne tehnike jer nijedan instrument nije savršen. Opaženi signal izražava se pomoću formule (2.3):

$$x = \tilde{x} + e \quad (2.3)$$

gdje je \tilde{x} „idealan“ ili pravi signal, a parametar e predstavlja šum. Cilj većine tehnika obrade signala je dobiti procjenu osnovnog signala bez šuma, to jest odvojiti signal od šuma. [2]

Postoje dvije glavne vrste mjernog šuma.

2.3.1. Stacionarni šum

Stacionarni šum je šum koji u svakoj sljedećoj točki (obično u vremenu) ne ovisi o šumu u prethodnoj točki. Postoje dvije glavne vrste stacionarnog šuma:

2.3.1.1. Homoskedastički šum

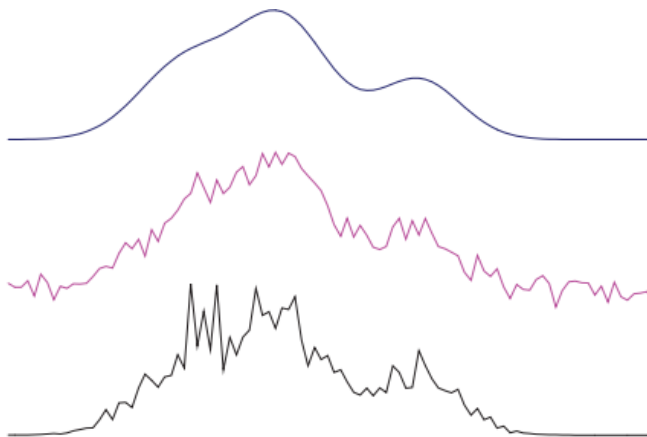
Značajke šuma (srednja vrijednost i standardna devijacija) ostaju konstantne tijekom cijelog niza podataka. Najčešća vrsta šuma dana je normalnom distribucijom (obično sa srednjom vrijednosti 0, osim ako ne postoji problem bazne linije) i standardnom devijacijom, ovisno o instrumentu. U većini situacija postoji nekoliko izvora šuma kod instrumenata, ali kombinacija različitih simetričnih distribucija šumova često teži normalnoj distribuciji.

2.3.1.2. Heteroskedastički šum

Ova vrsta šuma ovisi o intenzitetu signala i često mu je proporcionalna. Šum se i ovdje može predstaviti normalnom distribucijom, ali je standardna devijacija te distribucije proporcionalna intenzitetu osnovnog signala. Oblik heteroskedastičnog šuma se često pojavljuje ako se podaci transformiraju prije obrade, a uobičajena metoda je logaritamska transformacija koja se koristi u mnogim vrstama spektroskopije kao što je UV/Vis ili IR spektroskopija, od transmisije

do apsorbanije. U neobrađenim podacima prisutna je prava distribucija, ali ju transformacije podataka mogu iskriviti (engl. *distorts*).

Slika 6 ilustrira utjecaj obje vrste šuma na tipičan signal.



Slika 6. *Primjeri šuma. Od vrha prema dolje: osnovni signal, homoskedastičan i heteroskedastičan šum.*

U praksi se ne treba previše oslanjati na objavljene studije distribucije šuma od različitih instrumenata jer je svaki instrument drugačiji. Kako se poboljšava proizvodnja instrumenata, određene vrste šumova se smanjuju te je malo vjerojatno da će temeljita studija distribucije šuma provedena prije nekoliko desetljeća biti pouzdana i na modernijim instrumentima. U nedostatku određenog eksperimentalnog znanja, najbolje je držati se jednostavne normalne distribucije. [2]

2.3.2. Korelirani šum

Kod uzorkovanja podataka ponekad razina šuma u svakom uzorku može ovisiti o šumu iz prethodnog uzorka. Ova pojava je prilično uobičajena kod vođenja procesa. [2]

2.3.3. Omjer signala i šuma

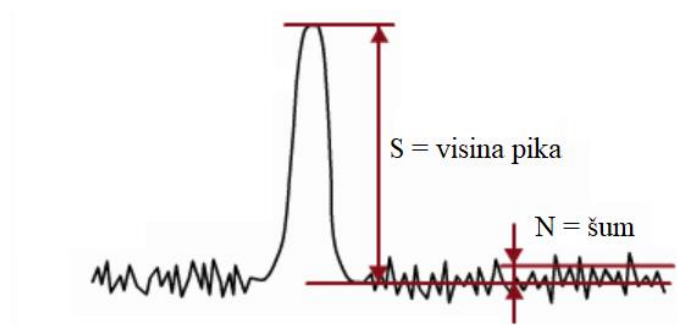
Omjer signala i šuma (engl. *signal-to-noise ratio*, SNR) koristan je parametar za mjerenje (**Slika 7**). Što je taj omjer veći, signal je intenzivniji u odnosu na pozadinu. Ovo je empirijski parametar, a najčešća definicija uključuje dijeljenje visine relevantnog signala (obično najintenzivnijeg ako ih ima nekoliko u skupu podataka) s korijenom srednje vrijednosti kvadrata šuma, izmjenjenog u području podataka gdje se zna da nema signala. Najčešće mjere omjera signala

i šuma pretpostavljaju da je šum homoskedastičan. Obično postoji područje signala u kojem nema vrhova i od njega mora se oduzeti bazna linija.

$$SNR = \frac{H_{MAX}}{\sqrt{\frac{\sum_{i=1}^n (\bar{x} - x_i)^2}{n-1}}} = \frac{S}{N} \quad (2.4)$$

U gornjoj jednadžbi brojnik predstavlja visinu promatranog signala, a nazivnik standardnu devijaciju šuma. [6] Sa **Slika 7** se vidi da će veći odnos S/N pridonijeti udjelu determinističkog dijela signala.

Ako je signal tvari manji od osnovnog šuma analitičke metode, tvar se ne prepoznaje. Upravo iz tog razloga omjer signala i šuma određuje granicu detekcije (engl. *limit of detection*, LOD). LOD se može definirati kao najmanja količina (ili koncentracija) analita koja se može detektirati s prihvatljivim omjerom signala i šuma [2]



Slika 7. Grafički prikaz SNR parametra [7]

3. RAMANOVA SPEKTROSKOPIJA

3.1. O Ramanovoj i IR spektroskopiji

Prilikom praćenja industrijskih procesa za dobivanje proizvoda odgovarajućih svojstava i kontrole kvalitete koriste se procesne analitičke tehnike koje se prema mjestu određivanja dijele na *off-line*, *at-line*, *on-line* i *in-line* tehnike. Farmaceutske industrije sve više primjenjuju *in-line* analitičke tehnike koje prate fizikalne i kemijske procese *in situ*. Takav pristup omogućuje manju izloženost opasnim materijalima, manju mogućnost kontaminacije ili uništavanja uzorka, omogućeno je snimanje u realnom vremenu, nema potrebe za uzimanjem alikvota i ne dovodi se u opasnost narušavanje ravnoteže u reakcijskom sustavu. Još neki od razloga zašto industrije biraju *in situ* tehniku za kontrolu i praćenje procesa su: obuhvaćanje širokog spektralnog područja, mogućnost snimanja u vodenim otopinama te analiza pomoću fleksibilne sonde prilagodljive eksperimentalnim uvjetima. Shodno tome razvija se sve više *in-line* tehnika, neke od njih se temelje na Ramanovoj spektroskopiji i njenim modifikacijama. Osim Ramanove spektroskopije, u najvažnije tehnike vibracijske spektroskopije ubraja se NIR-spektroskopija i ATR-MIR-spektroskopija koje se temelje na infracrvenoj spektroskopiji. [5]

Vibracijska spektroskopija se u suštini bavi identifikacijom molekula pomoću vibracija njihovih atoma, tj. periodičkih promjena kutova i udaljenosti atoma u molekulama. Ona se temelji na apsorpciji ili raspršenju elektromagnetskog zračenja i upravo je to glavna razlika Ramanove i IR spektroskopije. [5] Fenomen neelastičnog raspršenja svjetlosti, odnosno elektromagnetskog zračenja, poznat kao Ramanov efekt, objašnjava promjenu u valnoj duljini raspršenog zračenja. Ono ima različitu frekvenciju od one upadne zrake. Ova promjena valne duljine ovisi o kemijskoj strukturi molekula koje su odgovorne za raspršenje. [8]

Ozračivanje molekule monokromatskim svjetlom manifestira se kao dva tipa raspršenja svjetlosti, elastično i neelastično. Kod elastičnog raspršenja, ne dolazi do promjene u frekvenciji fotona ili promjene njegove valne duljine i energije. Suprotno tome, neelastično raspršenje je popraćeno pomakom frekvencije fotona zbog ekscitacije ili deaktivacije molekularnih vibracija te foton može dati ili primiti neku količinu energije. Dakle, mogu se pojaviti tri vrste fenomena. Prvi, kada svjetlost upadne na molekulu, može stupiti u interakciju s molekulom, ali neto razmjena energije (E) jednaka je nuli, pa je frekvencija raspršenog svjetla ista kao i upadnog svjetla ($E = E_0$). Ovaj proces je poznat kao Rayleighovo raspršenje. Drugi, svjetlost može stupiti u interakciju

s molekulom i neto izmijenjena energija je energija jedne molekularne vibracije. Ako se interakcijom poveća energija fotona tada će frekvencija raspršene svjetlosti biti veća od one upadne svjetlosti ($E = E_0 + E_v$). To je poznato kao anti-Stokes Ramanovo raspršenje. Treći, ako interakcijom molekula dobije energiju od fotona tada će frekvencija raspršene svjetlosti biti manja od upadne svjetlosti ($E = E_0 - E_v$), što je poznato kao Stokesovo Ramanovo raspršenje. [8]

Ramanov spektrometar sastoji se od izvora svjetlosti, monokromatora, držača uzorka i detektora. Čimbenici koji utječu na analizu Ramanovog spektra su visok omjer signala i šuma, stabilnost instrumenta i dovoljna rezolucija. Razvoj učinkovitih Raman spektrometara koji koriste NIR ili crveni ekscitacijski (pobudni) laser, riješio je problem izbjegavanja fluorescencije koja utječe na Ramanove signale. Glavne tehnologije koje se koriste za prikupljanje Ramanova spektra su disperzivna Ramanova spektroskopija i Fourierove transformacije, sa razlikama u njihovim laserskim izvorima i načinu na koji Raman detektira i analizira raspršenje. Obje ove tehnike imaju jedinstvene prednosti. U najosnovnijem smislu, molekula je ramanski aktivna kada postoji promjena polarizabilnosti tijekom vibracije. Osim toga, simetrija molekule također je jedan od uvjeta za dobivanje Ramanovog spektra, jer su u Ramanovim spektrima intenzivnija simetrična istežanja, s druge strane IR aktivna molekula je molekula kod koje postoji promjena dipolnog momenta, a za to je potrebno asimetrično razvlačenje u molekule. [8]

Vibracijska spektroskopija se kod farmaceutskih proizvoda najčešće izvodi pomoću Ramanove sonde. Pobudno zračenje prenosi se optičkim vlaknom za pobudu preko sonde do uzorka, a zatim se raspršeno zračenje skuplja te optičkim vlaknima prenosi do spektrometra i na kraju do detektora. Većina Ramanovih sondi posjeduje konfiguraciju uz povratno raspršenje. Kod procesnih Ramanovih sondi vlakna za pobudu i prikupljanje smještena su prostorno blizu, što omogućava praćenje procesa u realnom vremenu. [5]

Osim navedenih prednosti Ramanova spektroskopija ima i određene nedostatke. Jedan od njih je osjetljivost na nehomogenost uzorka zbog malog promjera laserske zrake. To se može umanjiti postupkom osvjetljivanja na široj površini, pritom se zračenje fokusira na šire područje sa većom žarišnom duljinom. U nedostatke ove tehnike ubraja se i nizak omjer signala i šuma. On se može povećati primjenjivanjem snažnijeg pobudnog zračenja kraće valne duljine, ali isto tako zbog takvog zračenja može nastati fluorescencija i degradacija uzorka. [5]

Ramanova spektroskopija je alat koji se primjenjuje u farmaceutici od karakterizacije formulacija lijekova do razjašnjenja kinetičkih procesa. Ramanova spektroskopija može pružiti brzo otkrivanje i prepoznavanje krivotvorenih lijekova. Standardne metode zajedno s Ramanovom spektroskopijom omogućuju proučavanje otpuštanja lijeka kod polukrutih i krutih formulacija. Ovaj pristup je također primjenjiv za *in vivo* studije i karakterizaciju strukture koloidnih sustava nosača lijeka. Ispitivanje kvalitete sastava i ujednačenosti tableta osjetljiva su područja u fazi proizvodnje farmaceutika. Ramanova spektroskopija smanjuje potrebno vrijeme za analizu sastava aktivne tvari u tabletama dok povećava broj testiranih tableta s povećanom razinom pouzdanosti, isto tako služi kao pouzdan alat u raznim industrijskim primjenama za praćenje i kontrolu procesa obrade krutih tvari. [8]

3.2. Predobrada podataka u Ramanovoj spektroskopiji

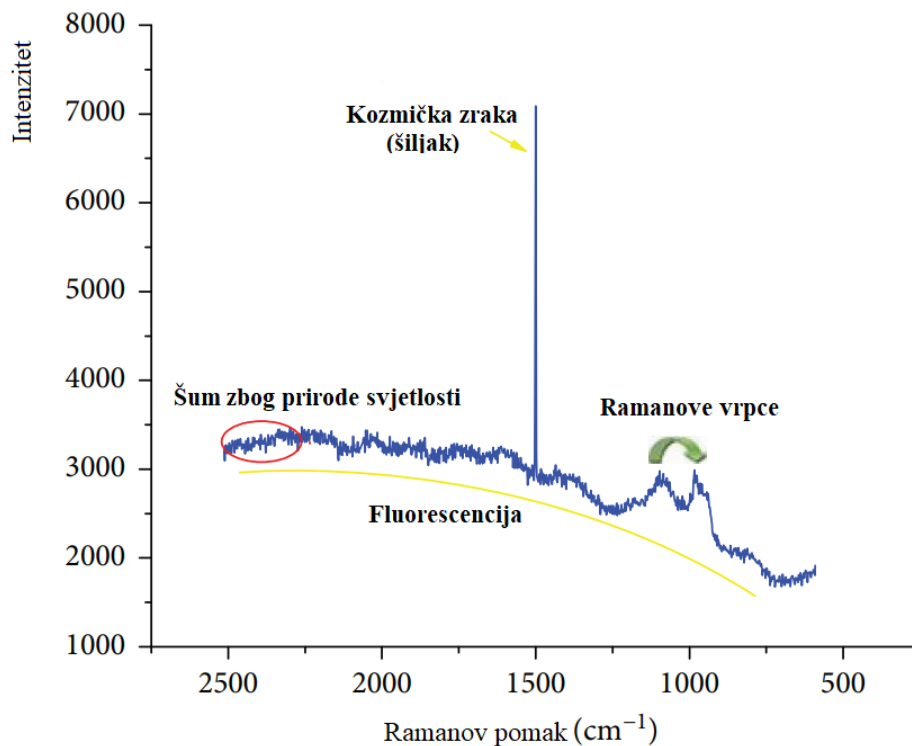
3.2.1. Usklađivanje spektralne osi

Kako se spektri od iste tvari mogu snimati tijekom nekoliko dana/mjeseci, vrlo je teško precizno kalibrirati Ramanov instrument kako bi imao istu Ramanovu os pomaka. Također, pojavljuju se i različite spektralne rezolucije stoga se spektri moraju poravnati na zajedničku os prije primjene bilo koje metode predobrade. [4]

3.2.2. Uklanjanje kozmičkih zraka

Vanjski neoptički izvori šuma stvaraju čestice visoke energije kao što su kozmičke zrake koje dopiru do detektora. Kozmičke zrake oslobađaju elektrone koji se ne mogu razlikovati od fotona. Elektroni generirani kozmičkim zrakama koncentrirani su u jednom ili dva elementa detektora i obično utječu samo na jedan piksel odjednom, što se manifestira kao vrlo uzak vrh visokog intenziteta. Ovi vrhovi su potpuno nasumični i teško ih je zamijeniti s Ramanovom vrpcom zbog karakteristične širine i oštine. **Slika 8** prikazuje primjer kozmičke zrake u spektru zajedno sa ostalim nepovoljnim utjecajima. Zraka se može otkriti skeniranjem spektra točku po točku tako da se provjerava da apsolutna vrijednost razlike između trenutne i sljedeće točke ne prelazi vrijednost od 500, u protivnom se u spektru pojavila zraka. Trenutačna točka, koja nije veća od točke gdje zraka počinje, oduzima se od sljedećih točaka sve dok je razlika veća od 500 kako bi se odredila spektralna širina zrake nazvana "*n*" (može biti između 1 i 4 točke). Nakon

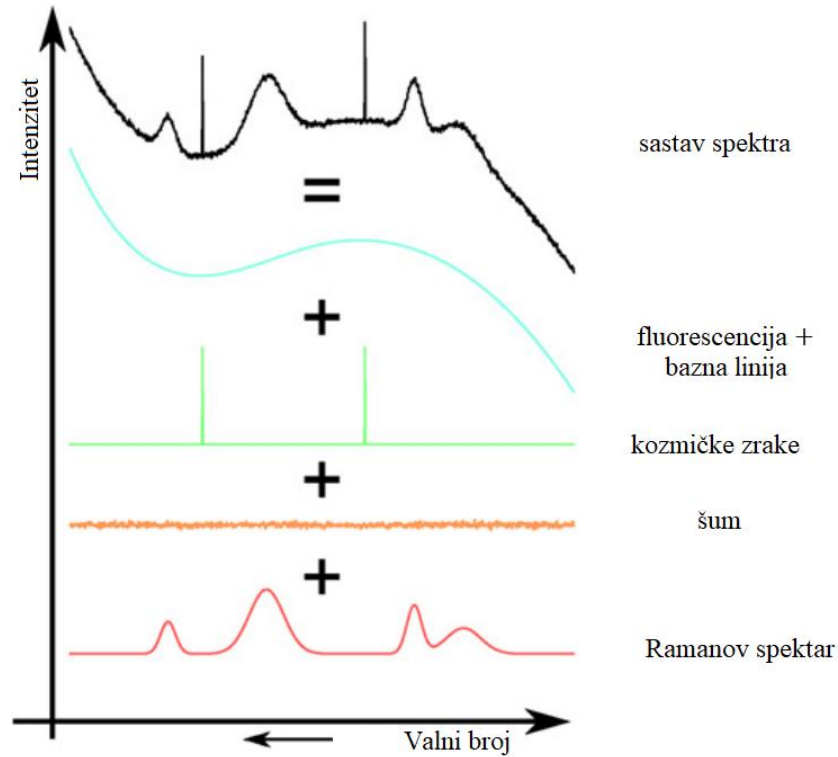
utvrđivanja širine zrake, trenutna točka postaje točka u kojoj počinje zraka, zatim se izračunava prosjek točke koja se nalazi n pozicija prije trenutne točke i n pozicija iza trenutne točke i vrijednost u trenutnoj točki zamjenjuje se vrijednošću izračunatog prosjeka. Ovaj postupak zamjene vrijednosti zrake ponavlja se n puta. [9]



Slika 8. *Nepovoljni utjecaji u spektru [9]*

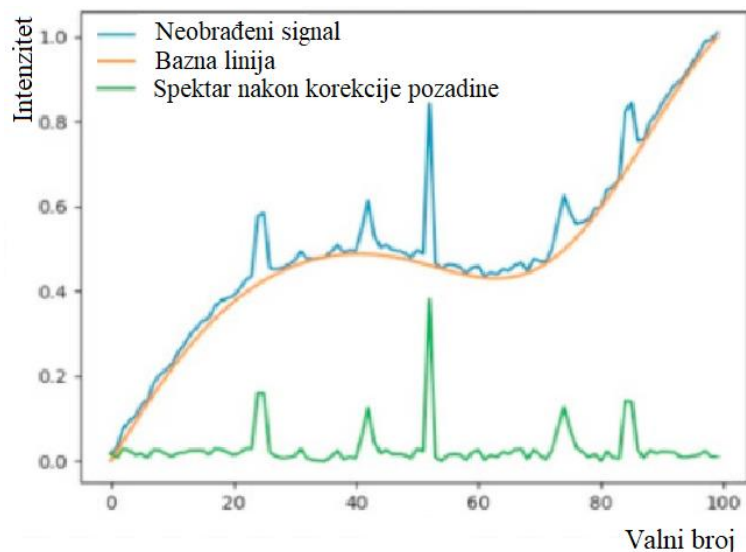
3.2.3. Korekcija pozadine

Korekcija pozadine / uklanjanje bazne linije (engl. *Background correction*) vrlo je važan dio predobrade spektralnih podataka. Razni fenomeni, prikazani na **Slika 9** i objašnjeni u poglavlju 2.1. poput fluorescencije, kozmičkih zraka i šumova izazivaju neravnomjerne pomake amplitude preko različitih valnih brojeva (Ramanovih pomaka). Ovi pomaci amplitude se moraju kompenzirati prije daljnje analize.



Slika 9. Sastav Ramanovog spektra. [10]

Pomicanje bazne linije je fluktuacija ukupne pozadine, koja zajedno s visokofrekventnim šumom čini glavnu smetnju signalu. Na primjer, u Ramanovom spektru (prikazanom na **Slika 10**), pik je karakterističan signal, a pozadinska fluktuacija označena narančastom bojom je bazna linija. Za daljnju analizu bitne su informacije o broju, položajima, visinama, površinama i drugim karakteristikama pikova signala. Osnovna linija preklapa se s vrhom signala, mijenja visinu svake podatkovne točke i interferira s oblikom. Da bi se iz signala očitale ispravni podaci, najprije se bazna linija mora identificirati u spektru, a zatim oduzeti od njega. Postupak uklanjanja naziva se korekcija pozadine/uklanjanje bazne linije, kao rezultat tog postupka očekuje se spektar s pozadinom blizu nule i da su samo vrhovi signala značajni. [11]

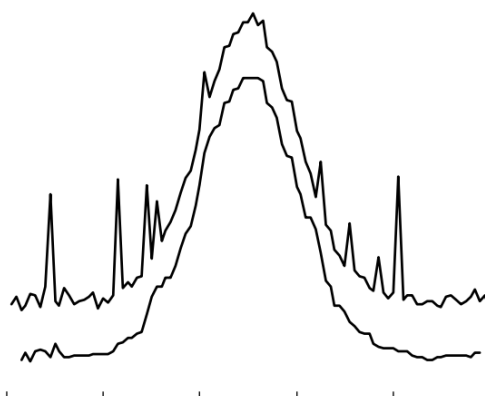


Slika 10. Korekcija pozadine [11]

U nastavku će se navesti neke od metoda korekcije bazne linije Ramanovih spektara.

3.2.3.1. Metoda temeljena na pokretnom medijanu prozoru

Ova se metoda temelji na pokretnom prozoru gdje se za svaku točku (Ramanov pomak) koristi samo nekoliko vrijednosti intenziteta (duljina prozora) oko točke za procjenu bazne linije. Prvo se izračunava medijan od vrijednosti intenziteta lokalnog prozora u svakoj točki. Ovaj niz vrijednosti medijana „uvija“ se (engl. *convolved*) sa Gaussovom funkcijom kako bi se osiguralo da u procijenjenoj baznoj liniji nema oštih diskontinuiteta. Iako je ova metoda bez modela (neparametarska), smatra se primitivnom za obrađivanje Ramanovih signala, a posebno je učinkovita u uklanjanju skokova, tj. kozmičkih zraka ili velikih odstupanja (**Slika 11**). [4]



Slika 11. Vrh sa šumovima (gornja krivulja) prije i (donja krivulja) nakon filtriranja medijan filtrom. [12]

3.2.3.2. Metoda temeljena na prilagodbi polinomom

Ovo je najčešće korištena metoda za uklanjanje bazne linije Ramanovih spektara. Kod metode temeljene na prilagodbi polinomom određene točke u spektru se odabiru kao bazne točke i polinom se prilagođava kroz te točke. Taj se polinom naknadno oduzima od Ramanovog spektra kako bi se eliminirali pozadinski učinci. Za jednostavan Ramanov signal kao što je Ramanov spektar nefluorescentnog krutog spoja, ravna linija koja prolazi kroz nekoliko točaka mogla bi biti dovoljna, dok bi za složeni Ramanov signal kao što je Ramanov spektar tkiva/stanica moglo biti potrebno mnogo baznih točaka zajedno s prilagodbom polinoma petog ili višeg reda. Odabir odgovarajućeg reda polinoma iznimno je važan jer pogrešno odabrani polinom višeg reda može procijeniti neke važne Ramanove vrpce kao pozadinu. Također, na prilagodbu polinoma višeg reda može utjecati visokofrekventni šum i stoga su pozadinske procjene nedosljedne. Za rješavanje ovih problema predlažu se neke modificirane tehnike pozadinske korekcije temeljene na prilagodbi s više polinoma. [4]

3.2.3.3. Metoda temeljena na asimetričnoj metodi najmanjih kvadrata

Ovdje se izgladeni signal procjenjuje iz neobrađenog Ramanovog signala kao bazne linije. Preostali signal (razlika) između neobrađenog Raman signala i procijenjene bazne linije je korigirani Ramanov signal. To se može postići uobičajenom metodom najmanjih kvadrata u kojoj je funkcija definirana kao zbroj kvadrata razlika između neobrađenog Raman signala i bazne linije koju treba procijeniti, a minimizira se iterativno. Kod metode najmanjih kvadrata jednaki prioritet se daje negativnim i pozitivnim pogreškama. Međutim, budući da bi bazna linija uvijek trebala dati pozitivne zaostale pogreške kako bi se osiguralo da su sve važne Ramanove vrpce neoštećene, metoda najmanjih kvadrata modificira se da bude pristranija pozitivnim zaostalim pogreškama. Ovdje se može dodati derivacija drugog reda kao još jedan izraz funkciji u svrhu izgladenosti procijenjene bazne linije. Postoji mnogo jednostavnih algoritama poput gradijentnog spusta koji procjenjuju takvu baznu liniju koja minimizira funkciju. Ova metoda je relativno brza čak i za složene signale, zahtijeva manje parametara i pokazala se učinkovitom za Ramanov spektar. [4]

3.2.3.4. Metoda temeljena na analizi frekvencijske domene

Kao što je ranije spomenuto, bazna linija je definirana kao glatka krivulja neobrađenog Ramanovog signala. Bazna linija varira s puno nižim vrijednostima u usporedbi s originalnim

Ramanovim vrpčama. Drugim riječima, bazna linija je niskofrekventna komponenta neobrađenog Ramanovog spektra, a Ramanove vrpce su puno uže i intenzivnije. Metode temeljene na frekvencijskoj domeni pokušavaju iskoristiti ova svojstva bazne linije i Ramanovih vrpce kako bi ih odvojile jedne od drugih. Ovdje se vremenski niz pretvara u frekvencijsku domenu, tj. u, nama tumačljiv, Ramanov spektar. Rezultat metoda zasnovanih na analizi frekvencijske domene sadrži informacije o promjenama frekvencija u vremenskim nizovima. Metode temeljene na frekvencijskoj domeni mogu se podijeliti u dvije kategorije, metode temeljene na Fourierovim transformacijama (FT) i metode temeljene na tzv. valićima (engl. *wavelet*). [4]

Neobrađeni spektralni podaci ne dobivaju se kao „razumljivi“ tj. interpretabilan spektar, već kao vremenski niz, gdje su sve spektroskopske informacije pomiješane te je potrebna matematička transformacija da bi se dobio „razumljivi“ spektar. Jedan od razloga za izvođenje FT spektroskopije je taj što se spektar prihvatljivog omjera signala i šuma može zabilježiti puno brže (i do stotinu puta brže) nego pomoću konvencionalnih spektrometara. Paralelno s Fourierovom transformacijom kreiran je veliki broj pristupa za poboljšanje kvalitete spektroskopskih podataka, često zvanih „Fourierove dekonvolucije“. Mnogi od ovih filtara se opisuju kao digitalni filtri. Fourierovi filtri mogu se povezati s linearnim metodama pomoću teorema o konvoluciji. [2]

3.2.4. Izgladivanje podataka

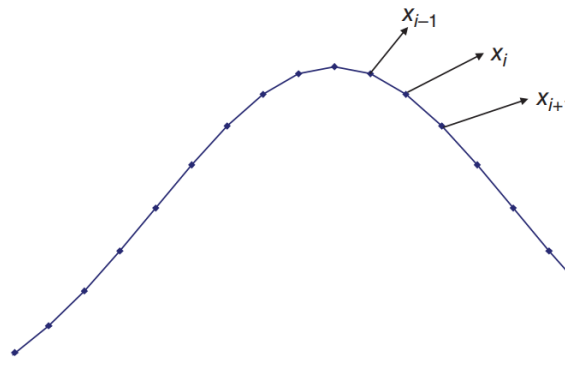
Nakon uklanjanja šuma iz podataka bitno je dobiti što je moguće informativniji signal. Kada se podaci dobivaju sekvencijalno u vremenu ili frekvenciji, fundamentalni signali često proizlaze iz zbroja glatkih, monotonih funkcija, dok je temeljni šum često nekorelirana funkcija. Važna metoda za otkrivanje signala uključuje izgladivanje (engl. *smoothing*) podataka; šum se izgladuje korištenjem „blagih“ metoda, dok će signal, budući da je širi od šumova, ostati. Ovaj pristup ovisi o tome imaju li vrhovi poluširinu od nekoliko točaka podataka; ako je digitalna razlučivost vrlo loša, signali se mogu pojaviti kao šiljci te ih se može zamijeniti sa šumom. Važno je odrediti optimalni filter za svaku pojedinu primjenu jer previše izgladivanja može smanjiti intenzitet i rezoluciju signala. Optimalna funkcija izgladivanja ovisi o širini vrha (u podatkovnim točkama) i o karakteristikama šuma.[2]

3.2.4.1. Metoda pomičnog prosjeka

Najjednostavnije metode izgladivanja su linearni filtri koji izgladuju izvorne podatke provlačeći ih kroz linearnu funkciju. Metoda pomičnog prosjeka (engl. *Moving average*, MA) uključuje korištenje okolnih točaka, npr. korištenje funkcije tri točke (**Slika 12**), kako bi se izračunala vrijednost za središnju točku i . Algebarski se takve funkcije izražavaju pomoću:

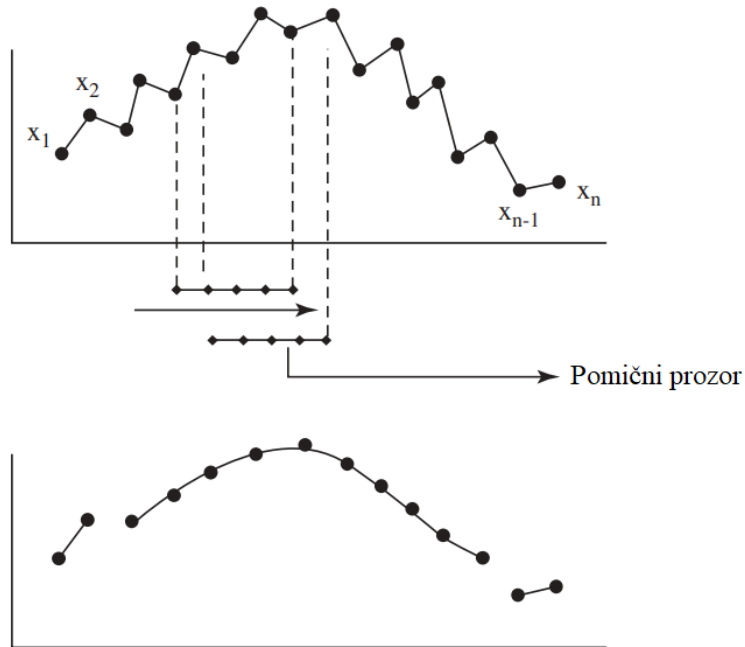
$$x_{i,novi} = \frac{1}{2M + 1} \sum_{j=-M}^M c_j x_{i+j}; M = 1; c_j = \frac{1}{3} \text{ za sve tri točke} \quad (3.1)$$

Jedna od najjednostavnijih funkcija izgladivanja je metoda pomičnih prosjeka sa tri točke. Svaka točka zamjenjuje se prosjekom točaka prije i poslije.



Slika 12. Odabir točaka koje će se koristiti u filtru pomičnog prosjeka od tri točke

Filtar se može proširiti na MA od 5 točaka ($M=2, c=1/5$), kao što je prikazano na **Slika 13**, MA od 7 točaka, itd. Što je više točaka u filtru, veće je smanjenje šuma, ali je i veća šansa zamučivanja signala. Broj točaka M u filtru često se naziva prozor ili širina filtra. Filtar se pomiče duž vremenske serije ili spektra, pri čemu se svaka podatkovna točka sukcesivno zamjenjuje odgovarajućom filtriranom podatkovnom točkom. Optimalni filter ovisi o distribuciji šuma i širini signala. Najbolje je eksperimentirati s nizom različitih širina filtara. Točke $(M-1)/2$ na početku i kraju serije se obično uklanjaju; npr. ako se koristi prozor širine 5 točaka (**Slika 13**), točke 1-2 se uklanjaju, a prva nova filtrirana točka je broj 3. [2] Sa **Slika 13** treba napomenuti da se za ekstremne točke ne mogu izgladiti podaci jer se koriste za izračunavanje prvog i posljednjeg prosjeka. Gornja slika prikazuje izvorni signal, a donja izgladeni signal.[13]



Slika 13. Filtar pomičnog prosjeka za prozor veličine $2M + 1 = 5$, odnosno $M = 2$.

3.2.4.2. Hanning i Hamming filtri s pomičnim prozorom

Ove metode filtriranja definiraju se postupkom ponderiranja, tim postupkom se izračunava vrijednost pojedine točke pomoću srednje vrijednosti, ali uzima se u obzir da unutar pomičnog prozora sa određenim brojem podatkovnih točaka neće svaka točka jednako pridonositi filtriranoj vrijednosti. Najveći utjecaj dodjeljuje se točki u centru prozora. Svakoj vrijednosti se određuje važnost sa ponderom (ponderacijskim koeficijentom). [13]

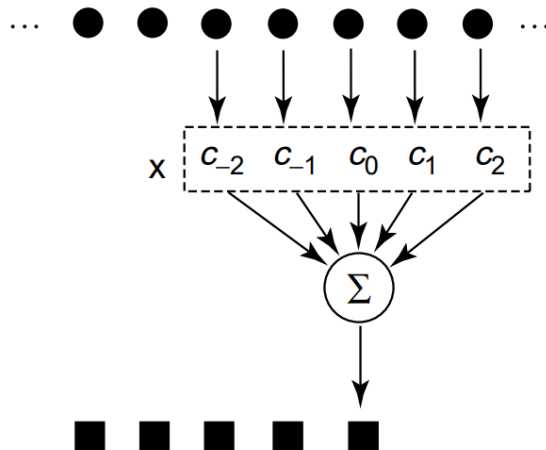
$$\hat{x} = \frac{1}{2M + 1} \sum_{j=-M}^M c_j x_{i+j} \quad (3.2)$$

U gornjoj jednadžbi c_j označava ponderacijski koeficijent, a zbroj svih pondera mora biti 1.

$$\sum_{j=-M}^M c_j = 1 \quad (3.3)$$

U metode ponderiranja spadaju Hanningovi prozori nazvani po Juliusu von Hannu i Hammingovi prozori nazvani po Richard W. Hammingu. Za veličinu Hanningova prozora od tri podatkovne točke, $M=1$, centralnoj točki pridodaje se iznos pondera od 0,5, a drugim dvjema

točkama 0,25. Za Hammingov prozor od pet podatkovnih točaka, $M=2$, ponderi su definirani iznosima od: 0,0357, 0,2411, 0,4464, 0,2411, 0,0357. Ova dva prozora imaju slične učinke i mogu se koristiti za drugačije veličine prozora. [2]



Slika 14. *Princip ponderirane metode pomičnog prozora*

Za Hammingov prozor sa gornje slike uzimaju se vrijednosti:

$$c_0 = 0,4464$$

$$c_{-1} = c_1 = 0,2411$$

$$c_{-2} = c_2 = 0,0357$$

Nedostatak MA filtara je što koriste linearnu aproksimaciju podataka što dovodi do odstupanja signala od izvornog oblika. Vrhovi spektara se često najbolje aproksimiraju polinomskim krivuljama, a posebno u središtu pika. Najčešće se koriste kvadratni, kubični ili čak modeli 4. reda. Međutim, kod aproksimacije polinomom regresijski proračun može biti dosta zahtjevan jer se spektar može sastojati od tisuću podatkovnih točaka, što bi značilo da bi trebali provesti približno tisuću linearnih regresija kako bi prilagodili polinom određenog stupnja kroz zadanu veličinu prozora, a to bi oduzimalo jako puno vremena. [2]

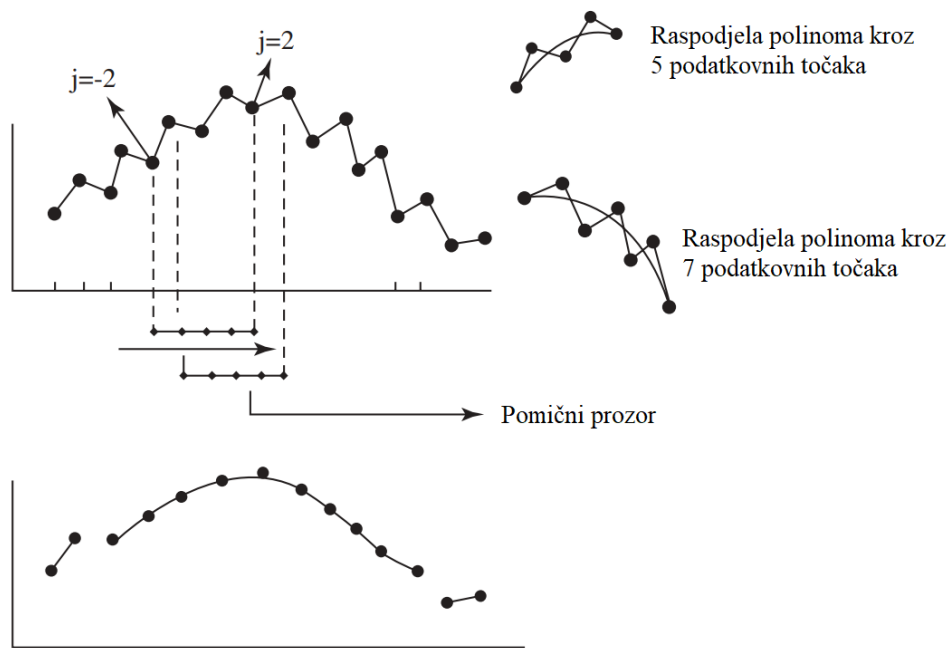
Ušteda vremena i jednostavnost mogu se postići metodom prilagodbe polinoma metodom najmanjih kvadrata pomoću koeficijenata koji definiraju filter. [14]

3.2.4.3. Savitzky-Golay filtri

Daleko najrašireniji digitalni filtar za izgladivanje podataka je prilagodba polinoma metodom najmanjih kvadrata, poznatiji kao Savitsky-Golay (SG) filtar. Savitsky-Golay filtar je filtar za izgladivanje koji se temelji na polinomskoj regresiji. Umjesto jednostavnog korištenja metode prosjeka, Savitsky-Golay filtar koristi regresijsko prilagođavanje za poboljšanje rezultata izgladivanja kao što je prikazano na **Slika 15**. Formulacija Savitsky-Golay filtra glasi:

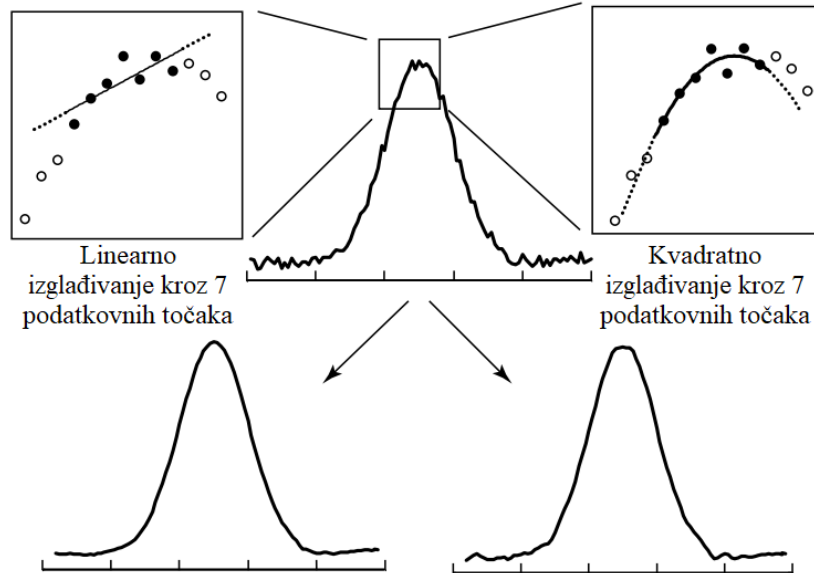
$$\hat{x} = \sum_{j=-M}^M c_j x_{i+j} \quad (3.4)$$

U gornjoj jednadžbi c_j označava Savitsky-Golayeve koeficijente.



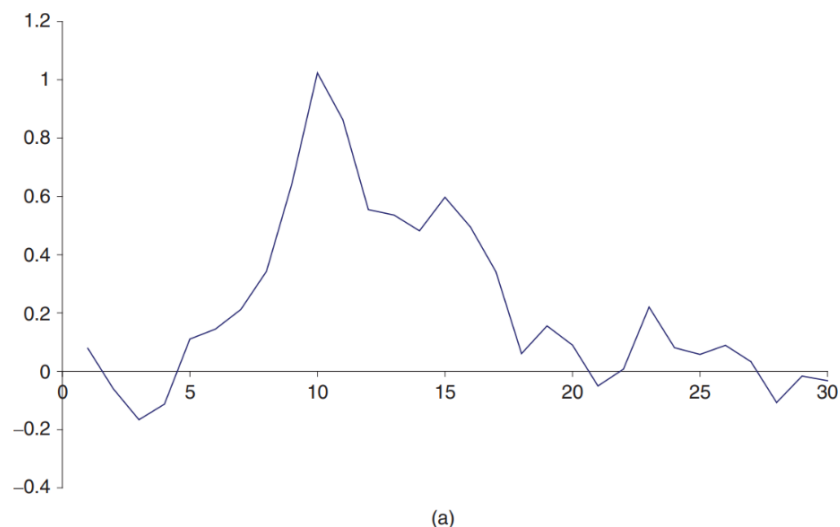
Slika 15. Savitsky-Golay filtar s veličinom prozora od $2M + 1 = 5$ točaka ($M=2$). Umjesto prosjeka podatkovnih točki u pokretnom prozoru, koristi se tehnika prilagođavanja polinoma. Gornja slika prikazuje izvorni signal, a donja izgladjeni, tj. obrađeni signal. [13]

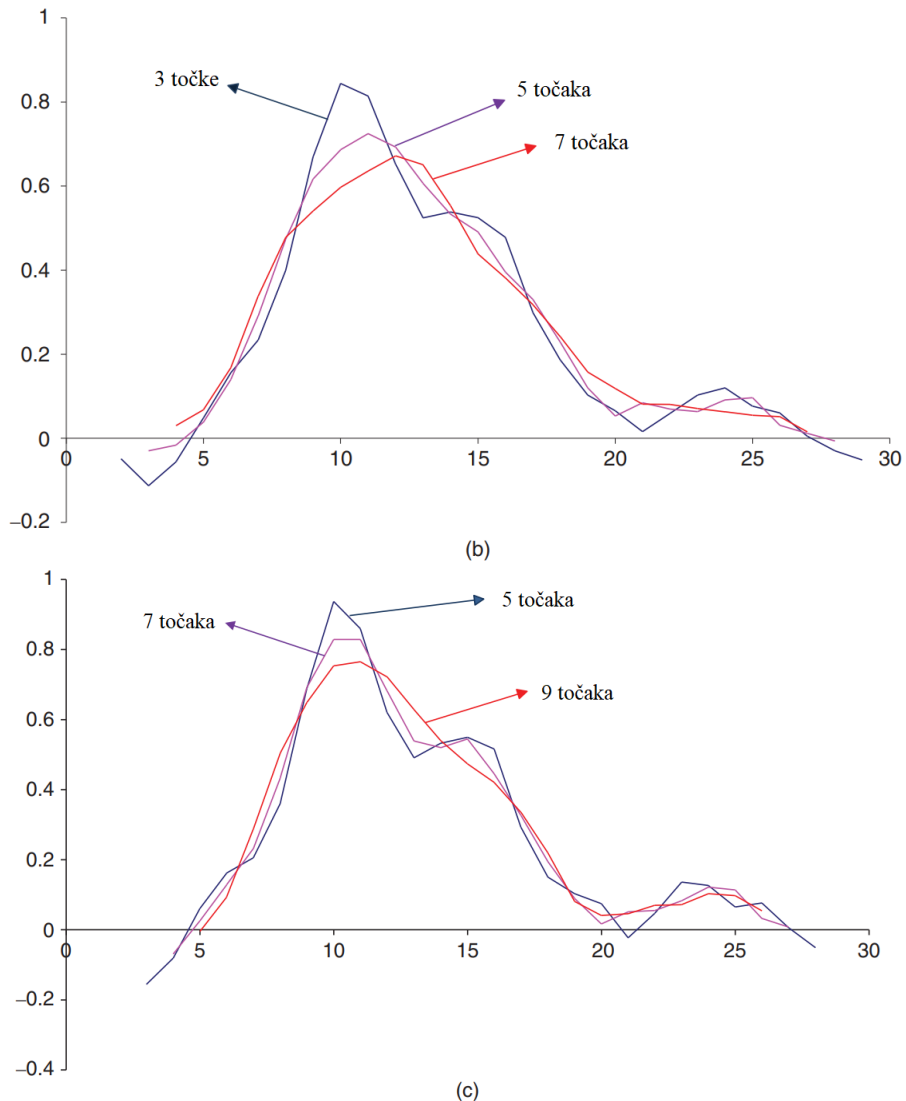
Primjena ovih filtara pretpostavlja da se područje skupa podataka na koje se fokusiramo, tj. prozor filtra, može prilagoditi polinom nižih redova, prema tome središnju točku unutar tog prozora procjenjuje izvođenje takvog uklapanja. Ovo je ilustrirano na **Slika 16** za polinom prvog i drugog reda. Uklapanjem podataka na ovaj način treba modelirati korelacije u čistom signalu dok se smanjuje utjecaj slučajnih fluktuacija šumova. [12]



Slika 16. *Ilustracija izgladivanja polinomom, metodom najmanjih kvadrata [12]*

Slika 17 prikazuje primjer primjene linearnih filtara. **Slika 17(a)** prikazuje neobrađene podatke. Na **Slika 17(b)** se može vidjeti da MA od tri točke zadržava rezoluciju, dok MA od pet točaka gubi rezoluciju i čini se da se klaster sastoji od samo jednog vrha. Nasuprot tome, kvadratni/kubični Savitzky–Golay filtri s pet i sedam točaka (**Slika 17(c)**) zadržavaju razlučivost dok ujedno smanjuju šum, a počinju gubiti razlučivost tek kada koriste funkciju s 9 točaka. [2]





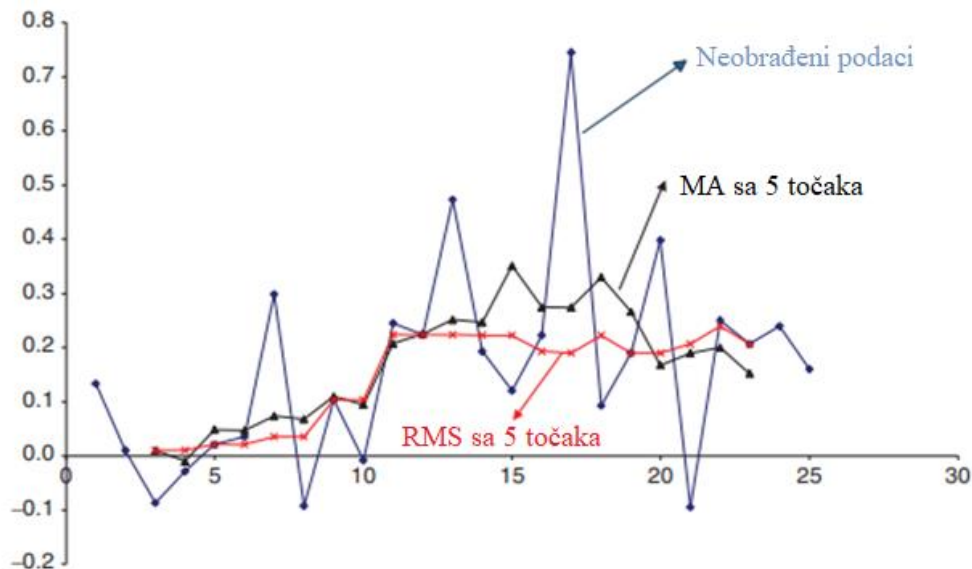
Slika 17. *Filtriranje podataka. (a) Neobrađeni podaci, (b) Filtri pomičnog prosjeka, (c) Kvadratni/kubični Savitzky–Golayevi filtri*

SG filtri koriste se i prilikom izgladivanja deriviranih signala na način da se najprije spektar uklopi u polinom, a zatim se polinom derivira, o čemu će biti riječi u nastavku.

3.2.4.4. Zaglađivanje podataka medijanom

Većina konvencionalnih filtara uključuje izračunavanje lokalnih multilinearne modela; međutim, u određenim područjima mogu postojati ekstremne vrijednosti u podacima, za koje je malo vjerojatno da su dio kontinuiranog procesa. Alternativna metoda zaglađivanja podataka uključuje korištenje RMS funkcija (engl. *Running median smoothing*) koje izračunavaju medijan umjesto srednje vrijednosti kroz prozor podataka. Primjer rezultata korištenja RMS funkcija

prikazan je na **Slika 18**. Pregledom neobrađenih podataka ne uočavaju se temeljni trendovi. Daljnja analiza mogla bi otkriti sustavni trend u podacima, no u većini situacija, prvi pregled je grafički. MA sa pet točaka sugerira sustavan proces, ali nije jasno povećava li se fundamentalni proces monotono s vremenom ili se povećava, a zatim smanjuje. RMS od pet točaka sugerira da se proces povećava s vremenom i puno je glađi od rezultata MA filtra.



Slika 18. Usporedba filtra pomičnog prosjeka (MA) i zaglađivanja podataka medijanom (RMS)

Svaka funkcija izgladivanja uklanja različite značajke u podacima, a često se preporučuje kombinacija nekoliko pristupa filtriranja, na primjer, srednje izgladivanje u pet točaka nakon čega slijedi Hanningov prozor u tri točke. [2]

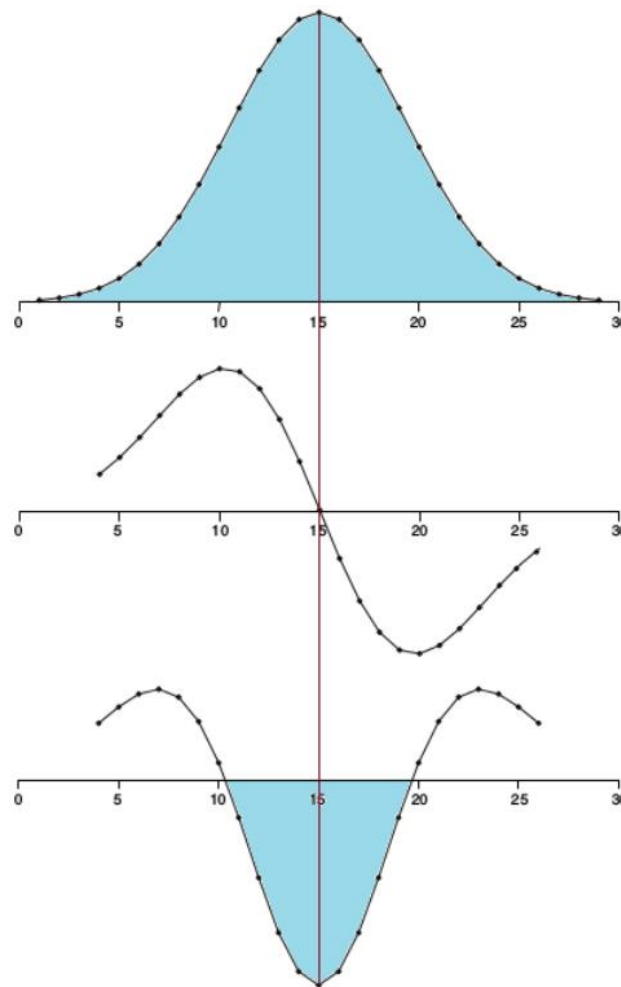
3.2.5. Derivacije

U odjeljku 3.2.4.3 govorilo se o Savitsky-Golayevom filtru za izgladivanje, koji se temelji na prilagodbi polinoma metodom najmanjih kvadrata kroz pomični prozor. Ovaj filter se također može koristiti za numeričko deriviranje. Budući da se SG filter temelji na prilagodbi polinoma, moguće je izravno izvesti derivaciju na polinomu i zatim dobiti težinski izraz za središnju točku u pomičnom prozoru kako bi se dobio derivacijski spektar. [13]

Prethodno opisane metode prvenstveno se odnose na uklanjanje šuma. Većina takvih metoda ostavlja nepromijenjene ili pak povećane širine vrhova, što je ekvivalentno zamućivanju. U analizi signala je važna potreba povećanje rezolucije. Jednostavan i često korišten pristup

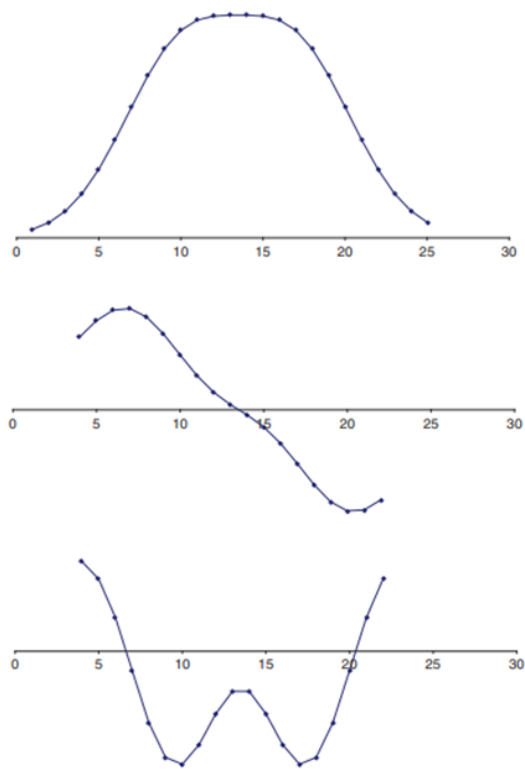
povećanju rezolucije je izračunavanje derivacija. Metode deriviranja signala korisne su pri razlučivanju detalja unutar samoga signala, tako da oni preklapljeni pikovi koji na prvi pogled tvore jedinstveni signal, postaju jasno vidljivi svaki za sebe. [2]

Kod prve derivacije, u deriviranom spektru, maksimum (i minimum) kod neobrađenog spektra postaju nulta vrijednost, a točke infleksije su pretvorene u maksimum i minimum. To bi zapravo značilo da je u središtu vrha prva derivacija jednaka nuli i dobar je način za točno određivanje položaja širokih vrhova, a prikazuje dvije točke infleksije. Druga derivacija postiže minimum u središtu vrha i prelazi nulu na točkama infleksije za prvu derivaciju. Valja napomenuti da se širina vrha smanjuje korištenjem derivacija. Prethodno opisano jasno je vidljivo sa **Slika 19** na primjeru Gaussovog pika. [2]



Slika 19. Prva i druga derivacija Gaussovog pika

Slika 20 ilustrira prvu i drugu derivaciju dva vrha koji se blisko preklapaju. Druga derivacija jasno ukazuje na dva vrha i omogućuje precizno određivanje njihovih položaja. Prva derivacija sugerira da vrh nije savršen. Moguće je izračunati treću i četvrtu derivaciju, ali izrazi mogu postati prilično komplicirani. No, postoje i dva nedostatka korištenja derivacija. Prvo, izračun derivacija može biti računalno zahtjevan jer je za svaku podatkovnu točku u spektru potreban novi izračun. Drugo, njihovim korištenjem se značajno pojačava šum i zahtijeva niske omjere signala i šuma. Ta se ograničenja mogu prevladati korištenjem Savitzky–Golay koeficijenata koji uključuju brzi izračun izglađenih viših derivacija. [2]



Slika 20. Dva pika koji se blisko preklapaju zajedno s njihovom prvom i drugom derivacijom

3.2.5.1. Metoda temeljena na derivaciji

Općenito, bazna linija je glatka krivulja sa niskofrekventnim komponentama u usporedbi s pravim Ramanovim vrpčama, tj. oštrijim vrhovima u spektru. Derivacija kod neobrađenog Ramanovog signala pojačava komponente više frekvencije (oštre), a komponente niže frekvencije kao što je pozadinska fluorescencija se potiskuju. Međutim, šum prisutan u neobrađenom Ramanovom signalu također sadrži komponente vrlo visoke frekvencije koje se također pojačavaju

zbog derivacije zajedno s pravim Ramanovim vrpcama. Ovi šumovi se mogu naknadno izgladiti (engl. *smoothing*). [4]

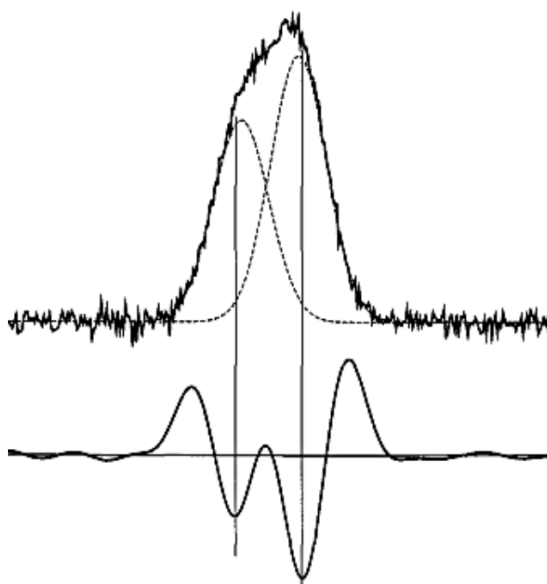
Najčešći postupak predobrade podataka derivacijama je primjena Savitzky-Golay (SG) tehnike. SG tehnika se osim za procjenu derivacije može koristiti i za izgladivanje podataka odnosno smanjenje šuma u podacima. Otklanjanje šumova proizašlih iz derivacija sa SG algoritmom kod računanja derivacija glasi:

$$\hat{x}_i^{(L)} = \sum_{j=-M}^M c_j^{(L)} x_{i+j} \quad (3.5)$$

gdje $\hat{x}_i^{(L)}$ označava filtriranu vrijednost L -te derivacije. [15]

Procjena derivacije izvodi se propuštanjem podataka kroz prozorski simetrični filtar (engl. *window-wise symmetric*). To se provodi kako se tijekom postupka derivacije ne bi pojačao visokofrekventni šum kao što je slučaj kod derivacije s konačnom razlikom. SG spektar je okružen (engl. *convoluted*) s prozorom koji sadrži $2g + 1$ točaka, gdje se svaki prozor koristi za procjenu središnje točke (s g točaka sa svake strane). Ove $2g + 1$ točke prilagođene su polinomom danog reda, a koeficijenti dobiveni ovom prilagodbom koriste se za procjenu nove vrijednosti na toj valnoj duljini (izgladivanjem ili izgladivanjem i derivacijom). Zbog navedenog postupka g točke na svakom kraju spektra će biti izgubljene. [3]

Slika 21 prikazuje simultano deriviranje i izgladivanje signala polinomnom aproksimacijom raspršenih točaka SG algoritmom. [15]



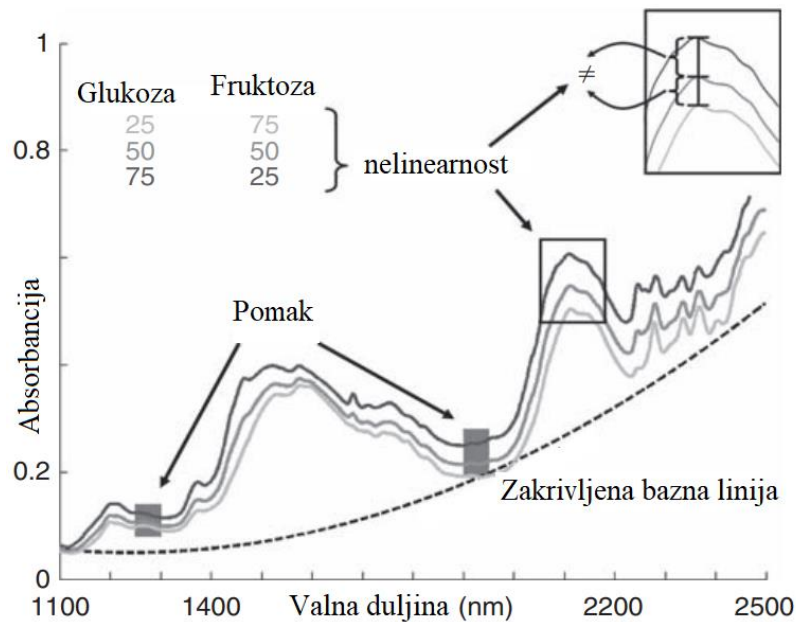
Slika 21. Detekcija pikova drugom derivacijom uz otklanjanje šuma pomoću SG algoritma

Procjena koeficijenata polinoma se provodi samo jednom, budući da se koeficijenti ne mijenjaju od jednog prozora do drugog kada su valne duljine postavljene ekvidistantno na osi, što znači da se isti vektor koeficijenata može koristiti u cijelom spektru (minus dva skupa g točaka na krajevima spektra). Koeficijenti polinoma lako se izračunavaju inverzijom matrice C :

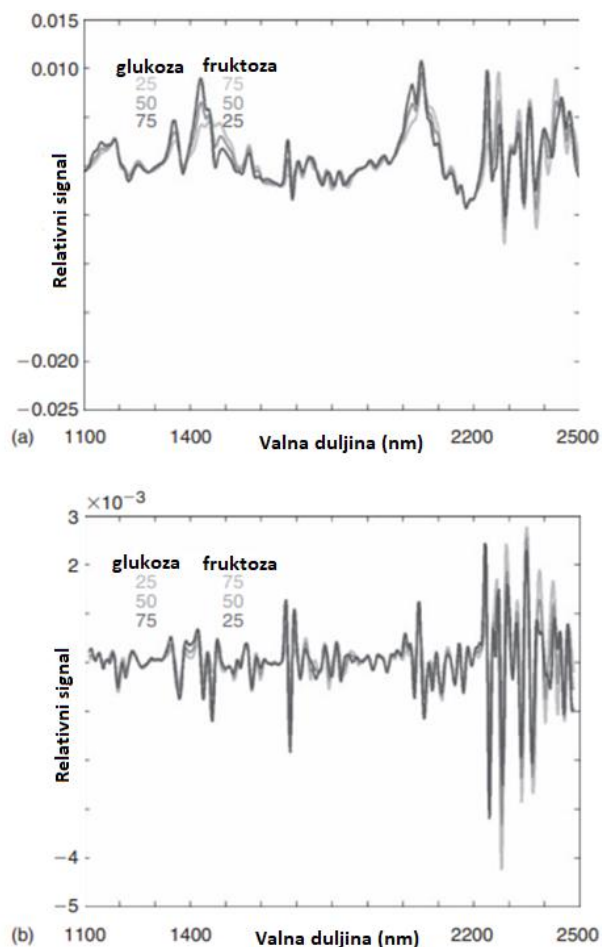
$$C = \begin{bmatrix} 1 & -g & \dots & (-g)^i \\ 1 & -(g-1) & \dots & [-(g-1)]^i \\ \dots & \dots & 0 & \dots \\ 1 & g-1 & \dots & (g-1)^i \\ 1 & g & \dots & g^i \end{bmatrix} \quad (3.6)$$

gdje je i red polinoma na koji su točke prilagođene (fitane). Veličina ove matrice je $(2g + 1) \times (i + 1)$. Kao što su Wentzell i Brown primijetili, parovi polinoma koji se podudaraju jedan s drugim daju istu procjenu za derivacije zbog redundantnosti u procjeni koeficijenata (tj. linearno i kvadratno prilagođavanje daje istu procjenu prve derivacije, dok kvadratno i kubno daju istu procjenu za drugu derivaciju, itd.). [3]

Podaci za glukozu i fruktozu prikazani na **Slika 22** koristit će se kao primjeri u teorijskom dijelu rada. **Slika 22** prikazuje neobrađeni spektar, te dva glavna efekta raspršenja, pomak i zakrivljenost bazne linije koji općenito uzrokuju nelinearnosti u spektru.



Slika 22. Primjer neobrađenog spektra



Slika 23. Savitzky–Golay procjena (a) prve derivacije i (b) druge derivacije spektra šećera. Oba su procijenjena izgladivanjem polinoma drugog reda od 9 točaka.

Kao što se može vidjeti na **Slika 23** primjenom 1. i 2. derivacijom porasla je količina šuma, a posebno na kraju spektra. Ta se pojava može izbjeći odabirom većeg prozora za izgladivanje, nauštrb gubitka važnih informacija na rubovima spektra. Obično je korištenje 7 do 11 točaka za izgladivanje te korištenje polinoma drugog ili četvrtog reda za postupak prilagodbe dovoljan za tipične spektralne podatke visoke razlučivosti. Općenito bi se za procjenu druge derivacije trebao koristiti veći broj točaka izgladivanja, budući da je ova procjena sklonija pojačanju šuma nego prva derivacija, smanjujući pritom ukupni omjer signala i šuma. Što je veći stupanj derivacije, potreban je veći broj točaka u prozoru za izgladivanje. Iako procjena druge derivacije daje priličan šum, postoje područja linearnosti u obje derivacije. Učinak bazne linije uklonjen je u drugoj derivaciji, dok je pomak (engl. *offset*) uklonjen objema tehnikama. [3]

3.2.6. Konvolucija

Konvolucija je matematički formalizam kojim se kombiniraju dvije funkcije stvarajući treću. Načela konvolucije su jednostavna. Dvije funkcije, f i g , su konvoluirane na način da daju h ako:

$$h_i = \sum_{j=-p}^{j=p} f_j g_{i+j} \quad (3.7)$$

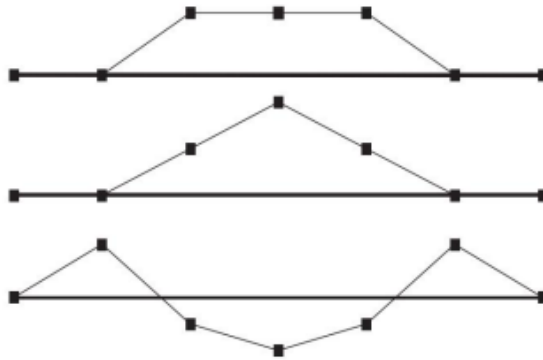
Ponekad se ova operacija piše pomoću operatora konvolucije označenog sa '*':

$$h(i) = f(i) * g(i) \quad (3.8)$$

Ovaj proces konvolucije potpuno je ekvivalentan digitalnom filtriranju:

$$x_{novi(i)} = x(i) * g(i) \quad (3.9)$$

gdje je $g(i)$ funkcija filtra. Naravno, moguće je konvoluirati bilo koje dvije funkcije jednu s drugom, pod uvjetom da je svaka iste veličine. **Slika 24** ilustrira funkciju (ili prozor) konvolucije za MA u tri točke, Hanningov prozor i Savitzky–Golay kvadratni/kubni filter druge derivacije u pet točaka. Rezultirajući spektar je konvolucija takvih funkcija s neobrađenim podacima. Konvolucija predstavlja opći matematički način rada s nizom metoda za poboljšanje signala. [2]



Slika 24. Od vrha prema dolje, pomični prosjek od tri točke, Hanningov prozor i Savitzky–Golay kvadratni/kubni filter druge derivacije u pet točaka

3.2.7. Normalizacija Raman podataka

Normalizacija je važan dio predobrade, budući da se različiti spektri iste tvari mogu snimati u različito vrijeme i pod različitim uvjetima instrumenta kao što su poravnanje i razina snage lasera. Dakle, spektri iste tvari mogu imati različite razine intenziteta. Normalizacija je proces koji vodi

računa o razlikama u razinama intenziteta osiguravajući da je intenzitet dane Ramanove vrpce iste tvari što je moguće sličniji u spektrima snimljenim pod istim eksperimentalnim parametrima, ali malo drugačijim uvjetima. U literaturi su dostupne brojne tehnike normalizacije. U nastavku je dan kratak opis uobičajenih tehnika normalizacije. [4]

Pretpostavimo da je spektar koji treba normalizirati definiran kao vektor „ S “, normalizirani spektar kao vektor „ SN “ gdje je

$$S = (s_1, s_2, \dots, s_N) \quad (3.10)$$

N je broj Ramanovih pomaka (spektralnih podatkovnih točaka)

$$SN = (sn_1, sn_2, \dots, sn_N) \quad (3.11)$$

gdje svaki element vektora predstavlja Ramanov intenzitet pri danom Ramanovom pomaku.

3.2.7.1. Vektorska normalizacija ili Euklidska norma

U vektorskoj normalizaciji, prije svega se izračunava '*norm*' spektra, koji je definiran kao kvadratni korijen zbroja kvadrata intenziteta spektra. Nadalje, svaki Ramanov intenzitet koji odgovara Ramanovu pomaku dijeli se s '*norm*' kako bi se dobio normalizirani spektar: [4]

$$norm = \sqrt{s_1^2 + s_2^2 + \dots + s_N^2} \quad (3.12)$$

$$sn_i = \frac{s_i}{norm}; i = 1, 2, \dots, N \quad (3.13)$$

3.2.7.2. Min-max normalizacija

Ovdje se prvo izračunavaju '*max*' i '*min*' vrijednosti svih intenziteta zadanog spektra. Zatim se svaki Ramanov intenzitet koji odgovara Ramanovom pomaku zamjenjuje novim intenzitetom dobivenim oduzimanjem '*min*' i dijeljenjem s '*raspon (maksimum-minimum)*': [4]

$$s_{max} = \max(s_1, s_2, \dots, s_N) \quad (3.14)$$

$$s_{min} = \min(s_1, s_2, \dots, s_N) \quad (3.15)$$

$$sn_i = \frac{s_i - s_{min}}{s_{max} - s_{min}}; i = 1, 2, \dots, N \quad (3.16)$$

3.2.7.3. Normalizacija standardno normalnom varijacijom (SNV)

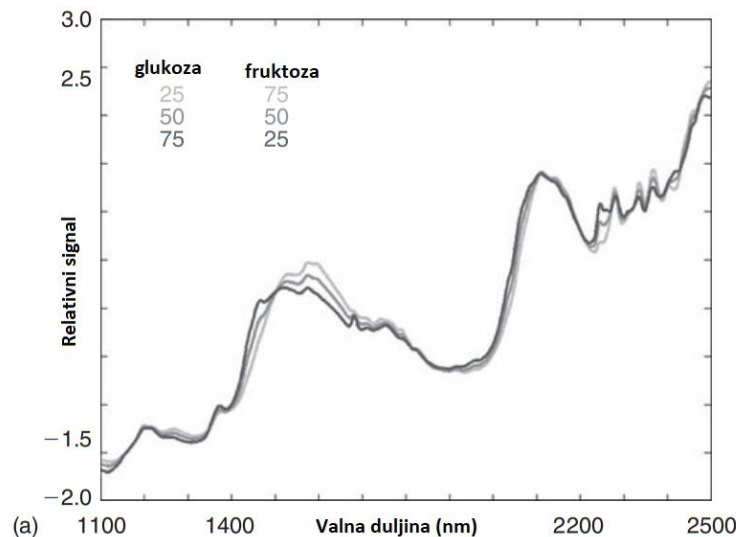
SNV (engl. *standard normal variate*, SNV) korigira raspršenja u spektru koja su nastala zbog veličine čestica i uklanja aditivna raspršenja bez da pri tome promijeni oblik izvornog spektra. [1] Ova tehnika je slična *Min-max* normalizaciji osim što se umjesto '*min*' koristi srednja vrijednost - '*mean*', a umjesto 'raspon' 'standardna devijacija (SD)': [4]

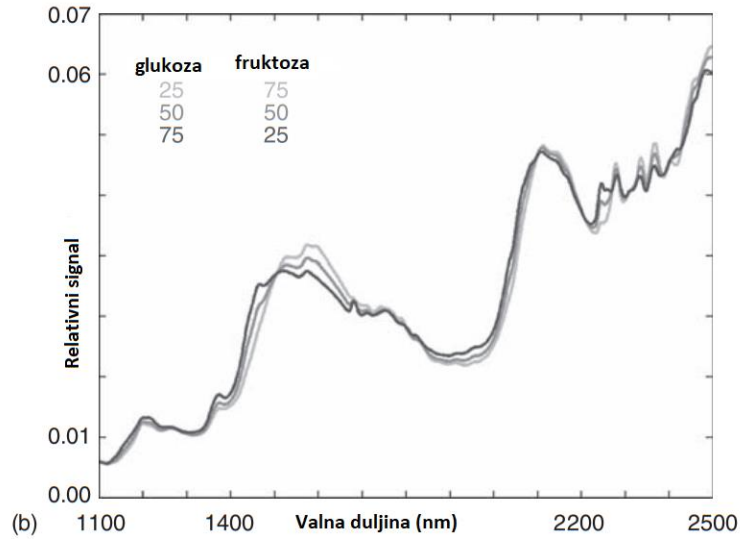
$$mean = \frac{s_1, s_2, \dots, s_N}{N} \quad (3.17)$$

$$SD = \sqrt{((s_1 - mean)^2 + (s_2 - mean)^2 + \dots + (s_N - mean)^2) / (N - 1)} \quad (3.18)$$

$$SN_i = \frac{s_i - mean}{SD}; i = 1, 2, \dots, N \quad (3.19)$$

Slika 25 ilustrira primjenu SNV tehnike na skupu podataka o šećeru. SNV je transformirao podatke na način da ispravljene podaci imaju linearniji odnos između signala i koncentracije. Budući da se normalizacija Euklidske baze temelji na kvadratnom korijenu kvadriranog zbroja kao u SNV, često daje rezultate slične SNV, kao što je jasno sa **Slika 25**. [3]





Slika 25. Podaci o šećeru prethodno obrađeni pomoću (a) standardne normalne varijance i (b) euklidske norme

3.2.7.4. Normalizacija vrhova

U normalizaciji vrhova, intenzitet koji odgovara središnjoj frekvenciji određene Ramanove vrpce koristi se kao referenca ' P '. Zatim se svaki Raman u intenzitetu spektra dijeli s ' P ' kako bi se dobio normalizirani spektar.

$$sn_i = \frac{S_i}{P}; i = 1, 2, \dots, N \quad (3.20)$$

Ova se metoda ne preporučuje kada postoji mogućnost pomaka položaja vrpce u spektru iz različitih ispitivanih uzoraka. [4]

4. PREDOBRADA PODATAKA U FTIR SPEKTROSKOPIJI

U ovom poglavlju se ne obrazlažu one metode koje su prikladne i za Ramanove i za IR podatke, a koje su opisane u prethodnom poglavlju, a to su metode pomičnog prosjeka, zaglađivanje podataka medijanom, Savitzky-Golay filtri, Hanning i Hamming filtri s pomičnim prozorom i metode temeljene na prilagodbi polinoma.

4.1. Korekcija pozadine

Zbog različitih instrumentalnih učinaka i pojave raspršenja, stvarni FTIR spektar se superponira na pozadinu. Slično Ramanovoj spektroskopiji, pozadina u FTIR signalima se sastoji od krivulje u području niske frekvencije. Metode korekcije pozadine mogu biti jednostavne poput oduzimanja pomaka (DC pomaka) ili uklanjanja bazne linije po dijelovima odabirom nekoliko točaka i spajanjem tih točaka ravnim linijama. Različite složene tehnike korekcije pozadine/uklanjanja bazne linije objašnjene ranije za Ramanovu spektroskopiju također se mogu koristiti za FTIR spektre, ali neke od tih tehnika su prikladnije od drugih. Takve tehnike uključuju prilagodbu polinoma (engl. *polynomial fitting*) i derivaciju na temelju SG filtera. Općenito, polinomi nižeg reda (drugog ili trećeg reda) su prikladni za FTIR spektroskopiju. Ova se tehnika može kombinirati s određenim tehnikama normalizacije kao što je korekcija multiplikativnog raspršenja (engl. *multiplicative scatter correction*, MSC) kako bi se istovremeno izvršila korekcija pozadine i normalizacija.

Budući da derivacije drugog reda mogu učinkovito ukloniti pozadinu prisutnu u FTIR spektrima, tehnike derivacije drugog reda temeljene na SG vrlo su popularne za korekciju pozadine FTIR spektara. Kako je SG filter nelinearna ponderirana funkcija izgladivanja, on osigurava da visokofrekventni šum pojačan tijekom derivacije drugog reda bude potisnut.[4]

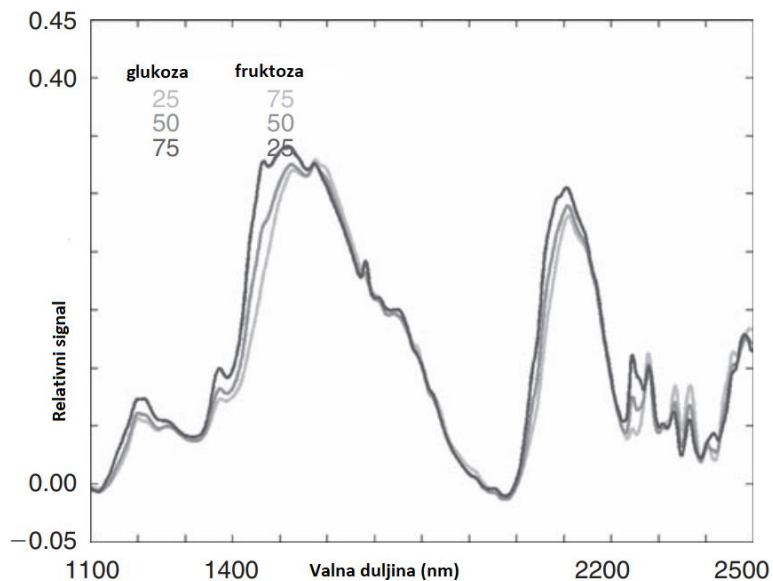
4.2. Korekcija bazne linije

Najčešća tehnika korekcije bazne linije koja se koristi u NIR (i IR) spektroskopiji je tehnika detrendiranja, njom se smanjuju nelinearni trendovi sa polinomskom jednadžbom za svaki spektar, koji se kasnije koristi za korekciju bazne linije. Uklanjanje trenda se izračuna kao razlika originalnog spektra i polinomske jednadžbe nove bazne linije. [1] Nešto sofisticiranije metode za

korekciju bazne linije su tzv. valići (engl. *wavelets*) i iterativno prilagođavanje bazne linije polinomom (IPBF). [3]

IPBF koristi isti princip kao i detrendiranje, a osnovna razlika je u tome što se kod primjene IPBF prilagođavanje bazne linije vrši iterativno. Najprije se bazna linija odabranog reda polinoma prilagođava spektru uzorka. Tada se mjerne točke koje leže iznad procijenjene bazne linije zamjenjuju predviđanjima iz prilagođene bazne linije. Ovaj novi (umjetni) osnovni spektar zatim se ponovno prilagođava s istim redom polinoma i ovaj se postupak ponavlja dok se ne zamijene nove točke uzorka. IPBF pretpostavlja da je bazna linija spektra dana najnižim točkama duž spektra, što nije nužno slučaj za NIR, no za IR spektre apsorbancije općenito vrijedi ova pretpostavka. [3]

Uspoređujući IPBF na **Slika 26** s detrendiranjem na **Slika 28** jasno je da ove dvije metode rade na sličnom principu. Što se tiče smanjenja trenda, IPBF je promijenio spektre, namještajući ih da imaju gotovo nultu baznu liniju. Uklonjeni su pomak i spektralni trend, ali je teško pronaći dijelove spektra koji se vode po Beerovom zakonu. Uspoređujući MSC i SNV korekcije spektara može se utvrditi da su MSC (**Slika 28**) i SNV (**Slika 25**) korigirani spektri vrlo slični, sa glavnom razlikom u mjerilu ordinatne osi. U usporedbi s MSC tehnikom, SNV tehnike ne trebaju referentni spektar i stoga ne trebaju odluku korisnika za izračun, ali kao nedostatak je moguća prisutnost bazne linije. Međutim, kako u SNV rješenju nema koraka najmanjih kvadrata, te su operacije sklonije pojavi šuma od robusnijeg MSC. Općenito, SNV i originalni MSC dovode do sličnih rezultata. [3]



Slika 26. Iterativna prilagodba bazne linije polinoma s polinomom drugog reda

4.3. Normalizacija FTIR podataka

Normalizacija je vrlo važan dio predobrade jer pokušava minimizirati učinke fluktuacija snage izvora (MIR izvora zračenja), raspršenja, varijacija u debljini uzorka, itd. Normalizacija pokušava istovremeno ispraviti različite efekte pomaka i skaliranja uzrokovane gore navedenim fenomenima. Neke od metoda normalizacije koje se koriste za Ramanove podatke vrlo su učinkovite i za FTIR spektroskopiju. Takve tehnike uključuju *Min-max* normalizaciju, SNV normalizaciju itd. Ove metode se primjenjuju pojedinačno na svaki spektar, pa se ove metode mogu klasificirati kao jednosmjerne metode (engl. *1-way methods*). Određene metode normalizacije koje uzimaju u obzir više od jednog spektra (recimo '*n*' broj spektara) u trenutku razvoja modela mogu se klasificirati kao *n*-smjerne metode (engl. *n-way methods*). Na primjer, tehnike poput multiplikativne korekcije raspršenja i proširenog MSC (EMSC) vrlo su popularne u FTIR spektroskopiji. Dok MSC pokušava normalizirati FTIR spektre, EMSC uklanja baznu liniju zajedno s normalizacijom.[4]

4.4. Multiplikativna korekcija signala (MSC)

Osnovno načelo MSC tehnike je ukloniti nelinearnost u podacima uzrokovanu raspršivanjem čestica u uzorcima. Postupak predobrade MSC tehnikom sastoji se od dva koraka:

a) procjena korekcijskih koeficijenata

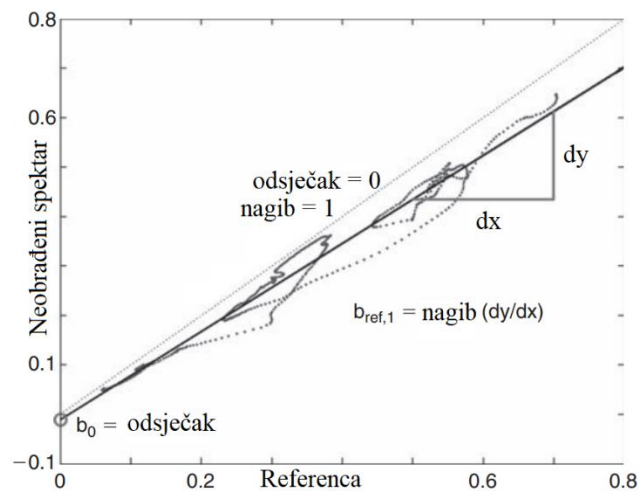
$$x_{org} = b_0 + b_{ref,1}x_{ref} + e \quad (4.1)$$

b) korekcija spektra

$$x_{corr} = (x_{org} - b_0)/b_{ref,1} \quad (4.2)$$

gdje su b koeficijenti korekcije, e je nemodelirani dio, a x_{org} , x_{ref} , i x_{corr} su izvorni, referentni i ispravljeni spektri, redom.

Postupak predobrade MSC tehnikom prikazan je na **Slika 27** u kojoj je nagib krivulje $b_{ref,1}$, dok je presjek s ordinatnom osi b_0 .



Slika 27. Procjena korekcijskih koeficijenata kod multiplikativne korekcije signala

Izvorno je predloženo da se korekcijski koeficijenti MSC tehnike mogu procijeniti iz manjeg spektralnog raspona koji ne sadrži kemijske informacije, ali nekada je dosta teško naći takav „prazan“ spektralni raspon, posebno u NIR spektroskopiji. U praksi se ' b ' koeficijenti često nalaze korištenjem cijelog spektra u jednadžbi (4.1). Korištenjem MSC tehnike na ovaj način nije potrebna subjektivna procjena spektra te će bilo koja dva korisnika dobiti iste rezultate MSC predobrade. Još jedan važan aspekt koji treba uzeti u obzir kod primjene MSC tehnike je definiranje referentnog spektra x_{ref} . Referentni spektar može biti jedan odgovarajući unaprijed definiran ili se mogu koristiti prosječni spektri od skupa uzoraka (npr. kalibracijski set).

MSC je kasnije proširen kako bi se uključile i korekcije valne duljine i korekcije za poznate spektre koji predstavljaju interferente ili željene dijelove (engl. *constituents*). Korekcije ovisnosti o valnoj duljini mogu se promatrati kao spajanje tehnika uklanjanja trenda (detrendiranja) s MSC tehnikom. Uključivanje ovisnosti valne duljine u MSC naziva se prošireni MSC (EMSC), dok se dodavanje poznatog spektra naziva oduzimanje spektralne interferencije. Ova dva proširenja mogu se lako uključiti u jednu jednadžbu, čime se istovremeno određuju svi korekcijski koeficijenti:

$$x_{org} = [1 \ x_{ref} \ x_{ref}^2 \ \lambda \ \lambda^2 \ x_{poznati,1} \ x_{poznati,2} \ \dots]b + e \quad (4.3)$$

$$b = x_{org}X_C^+ \quad (4.4)$$

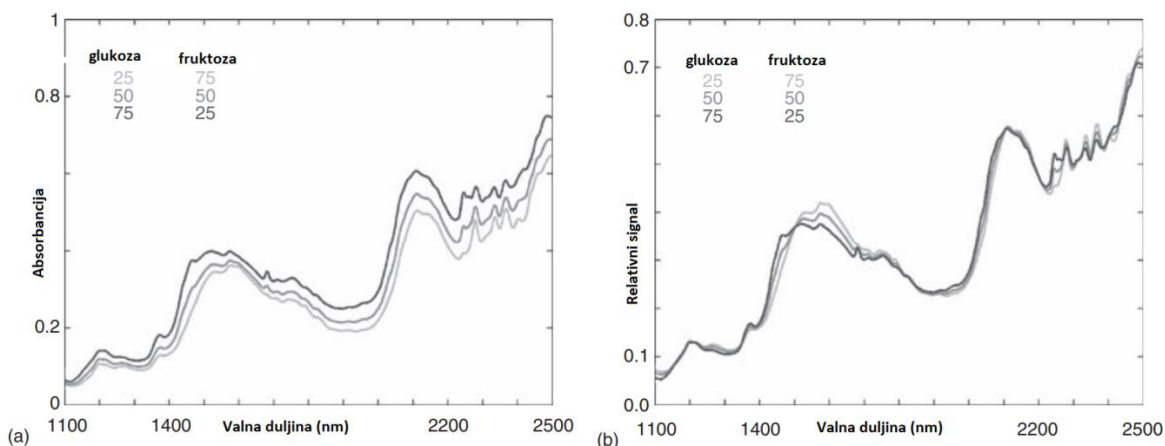
$$b = [b_0 \ b_{ref,1} \ b_{ref,2} \ b_{\lambda,1} \ b_{\lambda,2} \ b_{k,1} \ b_{k,2} \ \dots] \quad (4.5)$$

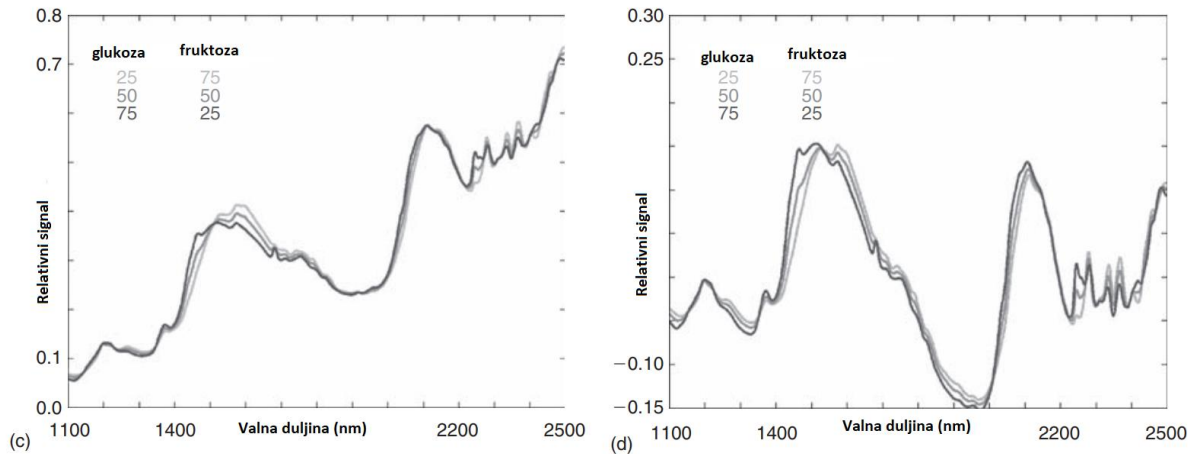
gdje vektor b sadrži EMSC korekcijske koeficijente, 1 je vektor koji se sastoji samo od jedinica, λ je valna duljina, a $x_{poznati}$ su poznati spektri. U jednadžbi (4.4) X_C označava cijelu matricu korekcije koja je u zagradama u jednadžbi (4.3), dok "+" označava da se za rješavanje sustava koristi tzv. Moore-Penroseov inverz. Jednadžba (4.5) pokazuje da je vektor korekcije sastavljen od jednog koeficijenta po članu korekcije u jednadžbi (4.3). Različiti $x_{poznati}$ vektori bi trebali biti ortogonalizirani kako bi izračun inverzne matrice u jednadžbi (4.4) bio numerički stabilan. Uz gore navedene primjere, moguće je dodati ovisnosti o valnim duljinama višeg reda ili referentne ovisnosti višeg reda. Međutim, viši redovi dovode do dva ili više mogućih rješenja za korekciju (koja proizlaze iz višestrukih korijena za proširenje polinoma) no samo jedno od rješenja ima fizikalno/spektroskopsko značenje. Primjerice, samo pozitivni korijen daje fizički/spektroskopski validan rezultat. Rješavanje jednadžbe za referentne ovisnosti drugog i trećeg reda može se riješiti analitički, dok se korekcija višeg reda može riješiti samo numerički.

Iz jednadžbe (4.3) može se vidjeti da se uklanjanjem svih članova osim prva dva, EMSC pretvara u izvorni MSC, dok se uklanjanjem referentnog x_{ref} i korekcije $x_{poznate}$ jednadžba pretvara u standardnu spektralno ovisnu jednadžbu detrendiranja. Stoga, jednadžba (4.3) sažima različite tehnike predobrade, a radi jednostavnosti označavat će se kao "MSC jednadžba".

Nekada nije baš jednostavno dobiti poznati spektar interferencija za korištenje u korekciji, budući da su uzorci složeni (posebno u analizi hrane NIR spektroskopijom) tako da posljednji dio jednadžbe (4.3) nije dostupan za mnoge praktične primjene MSC tehnike.

Učinak triju varijacija MSC tehnike ilustriran je na **Slika 28**, gdje su uspoređeni regularni MSC (linearna referentna ovisnost), EMSC (linearna referenca i ovisnost o valnim duljinama prvog i drugog reda) i detrendiranje (ovisnost o valnim duljinama prvog i drugog reda). Kao što je vidljivo iz **Slika 28** uključivanje referentnih spektara (x_{ref}) u korekciju dovodi do neznatnih promjena u usporedbi s neobrađenim spektrima. S druge strane, dodavanje korekcije valne duljine značajno mijenja izgled prethodno obrađenog spektra; spektri su postavljeni duž apscise za razliku od onih postavljenih duž dijagonale. Kao što se može vidjeti za MSC, **Slika 28 b**) i za EMSC **Slika 28 c**), predobrada podataka je dovela do smanjenja odstupanja između tri uzorka, a podaci su postali više linearno ovisni o koncentraciji šećera. Vršni dio spektra oko 1600 nm varira linearno s koncentracijom fruktoze, dok je manji vrh na 1500 nm linearno povezan s koncentracijom glukoze. U slučaju detrendiranja podataka (**Slika 28 d**) uklonjen je opći nagib u spektrima, dajući „ravan“ spektar, ali je nelinearnost između apsorbancije i koncentracije još uvijek očita. Pomak u podacima je ipak donekle smanjen.[3]

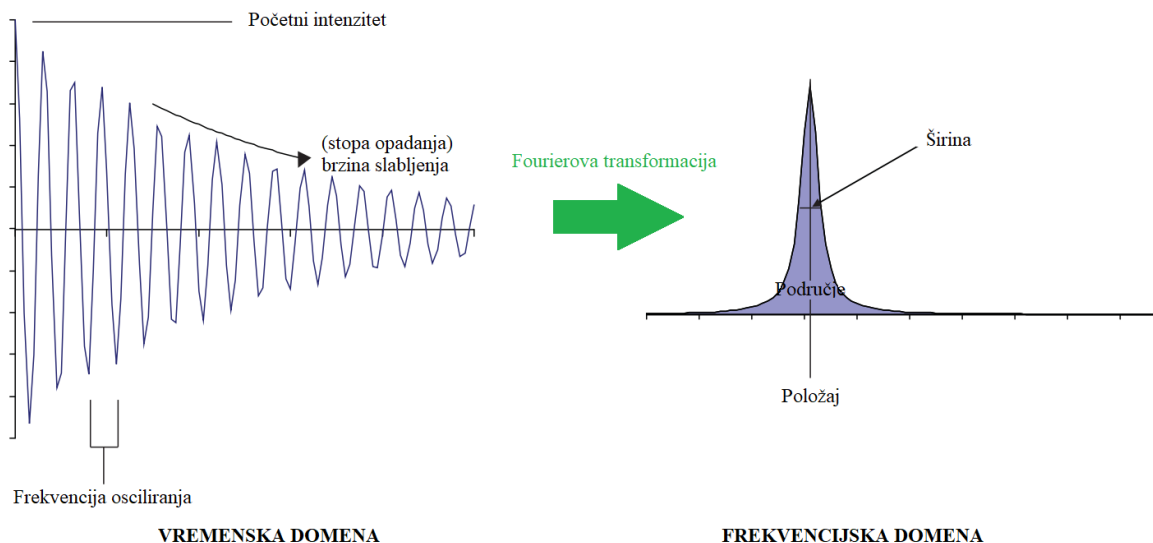




Slika 28. Učinak triju varijacija MSC na spektre smjese šećera. (a) neobrađeni podaci, (b) MSC podaci: $[1 x_{ref}]$, (c) EMSC podaci: $[1 x_{ref} \lambda \lambda^2]$, i (d) podaci s uklonjenim trendom: $[1 \lambda \lambda^2]$. (Matrične zagrade označavaju koji su dijelovi jednadžbe (4.3) uključeni.)

4.5. Metode Fourierove transformacije

Izvorna literatura o Fourierovim transformacijama podrazumijevala je primjenu na kontinuiranim skupovima podataka. Međutim, u kemijskoj instrumentaciji podaci se često ne uzorkuju kontinuirano nego u pravilnim vremenskim intervalima. Za obradu takvih vrsta podataka koristi se diskretna Fourierova transformacija (DFT). U FTNMR-u se podaci prikupljaju u određenim vremenskim intervalima, u tzv. vremenskoj domeni. Pomoću FTNMR puno je brže dobiti podatke nego korištenjem konvencionalnih metoda te se cijeli spektar može uzorkovati u svega nekoliko sekundi. Budući da vremensku domenu nije lako protumačiti javila se potreba za korištenjem DFT-ova. Svaki vrh u spektru može se opisati s tri parametra, visinom, širinom i položajem. Spektar se sastoji od zbroja vrhova i često se naziva frekvencijska domena. Neobrađeni podaci se bilježe u vremenskoj domeni i svaki vrh frekvencijske domene odgovara vremenskoj seriji koju karakterizira početni intenzitet, brzina oscilacije, tj. frekvencija osciliranja i brzina slabljenja (engl. *decay rate*). Vremenska domena se sastoji od zbroja vremenskih serija od kojih svaka odgovara vrhu u spektru. Fourierove transformacije pretvaraju vremensku seriju u prepoznatljivi spektar kao što je prikazano na **Slika 29**. Svaki parametar u vremenskoj domeni odgovara parametru u frekvencijskoj domeni (Tablica 1).



Slika 29. *Fourierova transformacija iz vremenske domene u frekventijsku domenu*

Tablica 1 Ekvivalencija između parametara u vremenskoj i frekventijskoj domeni

VREMENSKA DOMENA	FREKVENCIJSKA DOMENA
početni intenzitet	područje vrha
frekvencija osciliranja	položaj vrha
stopa opadanja	širina vrha

- Što je veća brzina oscilacije u vremenskoj seriji, to je vrh udaljeniji od ishodišta u spektru.
- Što je brža stopa opadanja u vremenskoj seriji, to je širi vrh u spektru.
- Što je veći početni intenzitet u vremenskoj seriji, veća je površina transformiranog vrha.

Oblik vrha u frekventijskoj domeni odnosi se na krivulju opadanja u vremenskoj domeni. Vremenska domena ekvivalentna Lorentzianovom vrhu je:

$$f(t) = A \cos(\omega t) e^{-t/s} \quad 4.6$$

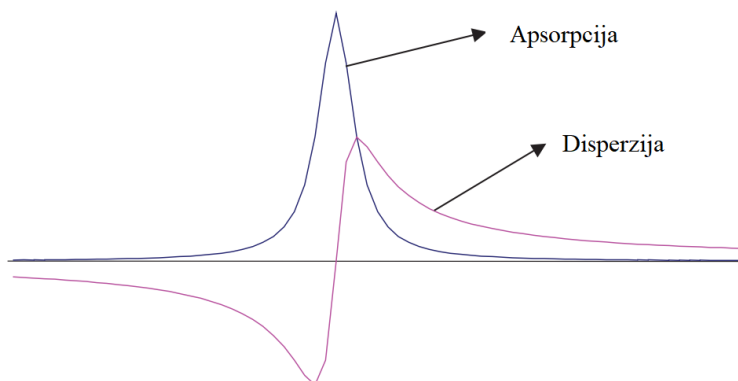
gdje je A početna visina (odgovara površini u transformaciji), ω frekvencija oscilacije (odgovara položaju u transformaciji) i s brzina opadanja (odgovara širini vrha u transformaciji).

Kod većine spektara prisutno je nekoliko vrhova, a vremenske serije izgledaju mnogo složenije nego na **Slika 29**, sastojе se od nekoliko superponiranih krivulja, kao što je prikazano na **Slika 30**.



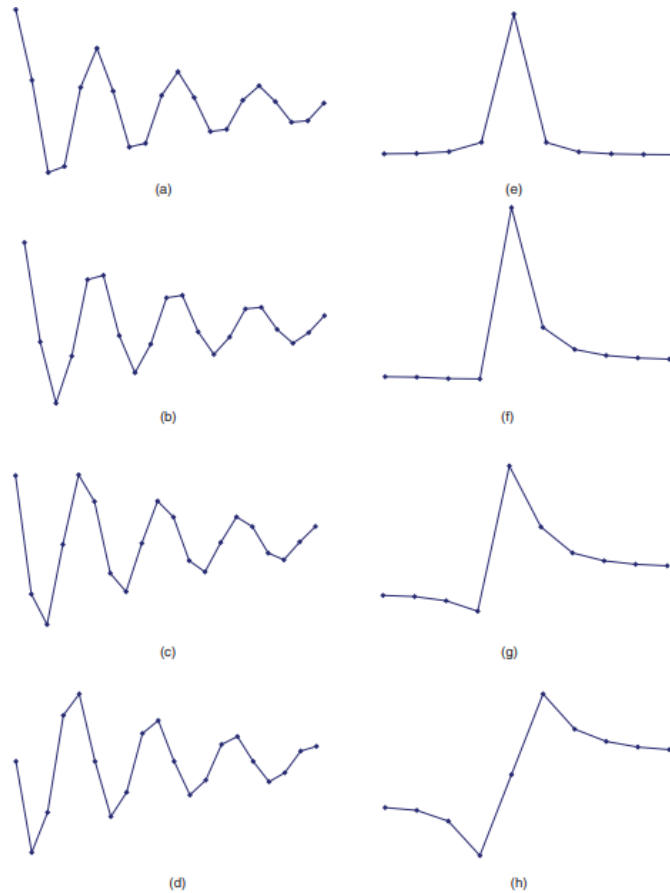
Slika 30. Tipični vremenski niz koji se sastoji od nekoliko komponenti

Fourierove transformacije pretvaraju neobrađene podatke (npr. vremenski niz) u dva spektra frekvencijske domene, od kojih se jedan naziva stvarnim spektrom, a drugi imaginarnim. Pravi spektar predstavljen je samo polovicom transformiranih podataka. U Fourierovoj transformaciji realnog vremenskog niza, razlikuju se oblici vrhova u realnoj i imaginarnoj polovici spektra. U idealnom slučaju, stvarni spektar odgovara obliku apsorpcijske linije, a imaginarni spektar disperzijske linije oblika, kao što to prikazuje **Slika 31**. Oblik apsorpcijske linije ekvivalentan je osnovnom obliku vrha kao što je Lorentzian ili Gaussian, dok oblik disperzijske linije pomalo nalikuje prvoj derivaciji. [2]



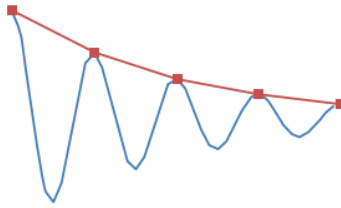
Slika 31. Oblici linija apsorpcije i disperzije

Međutim, često se ta dva oblika vrha miješaju zajedno u stvarnom spektru, zbog malih nesavršenosti u prikupljanju podataka, koje se nazivaju fazne pogreške. Razlog za to je taj što prikupljanje podataka ne počinje uvijek točno na vrhu kosinusnog vala, a u praksi bi izraz $\cos(t)$ trebao biti zamijenjen s $\cos(t+)$. Budući da fazni kut u vremenskom nizu od 90 pretvara kosinusni val u sinusni val, posljedica faznih pogrešaka je miješanje sinusne i kosinusne komponente stvarne i imaginarne transformacije za savršen oblik vrha. Kako se ovaj kut mijenja, oblik stvarnog spektra postupno se iskrivljuje (**Slika 32**).



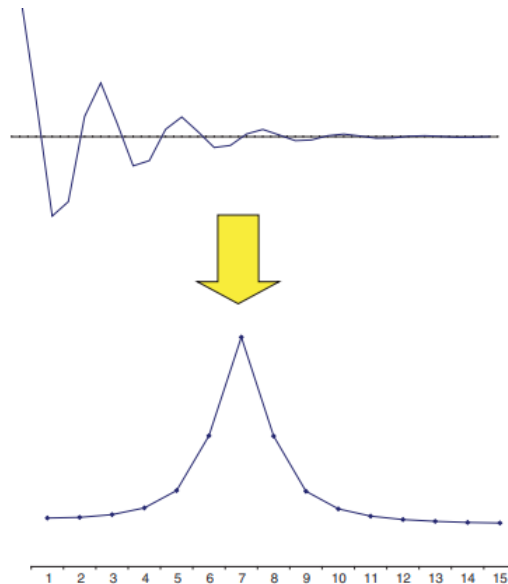
Slika 32. *Ilustracija faznih pogrešaka (vremenski niz (a–d) i realna transformacija (e–h))*

Važno svojstvo DFT-a odnosi se na brzinu uzorkovanja podataka. Na **Slika 33** prikazan je vremenski niz, pri čemu svaki kvadratić označava točku uzorkovanja. Ako se uzorkuje upola manjom brzinom, činit će se da nema oscilacija, jer će svaka alternativna podatkovna točka biti eliminirana. Stoga ne postoji način da se takav niz razlikuje od niza nulte frekvencije. Frekvencija osciliranja na **Slika 33** naziva se Nyquistova frekvencija. Sve što oscilira brže od ove frekvencije neće se razlikovati od sinusnog vala na nižoj frekvenciji. Brzina uzorkovanja utvrđuje raspon vidljivih frekvencija. Što je veća brzina uzorkovanja, veći je raspon vidljivih frekvencija. Kako bi se povećala spektralna širina, potrebna je veća brzina uzorkovanja.



Slika 33. Rijetko uzorkovana vremenska serija uzorkovana na Nyquistovoj frekvenciji. Plava: temeljna vremenska serija, crvena: rijetko uzorkovana vremenska serija

Filtiri se mogu primijeniti na neobrađenim podacima (vremenska domena) i prije Fourierove transformacije, što je uobičajena metoda u spektroskopiji za poboljšanje rezolucije ili omjera signala i šuma, a kao alternativa primjeni filtera izravno na spektralnu domenu. Širina vrha u spektru prvenstveno ovisi o stopi opadanja u vremenskoj domeni. Što je opadanje (engl. *decay*) brže, vrh je širi. **Slika 34** ilustrira široku vršnu vrijednost zajedno s odgovarajućom vremenskom domenom. Ako se želi povećati razlučivost, jednostavan pristup mogao bi biti promjena oblika funkcije vremenske domene tako da je opadanje sporije. [2]

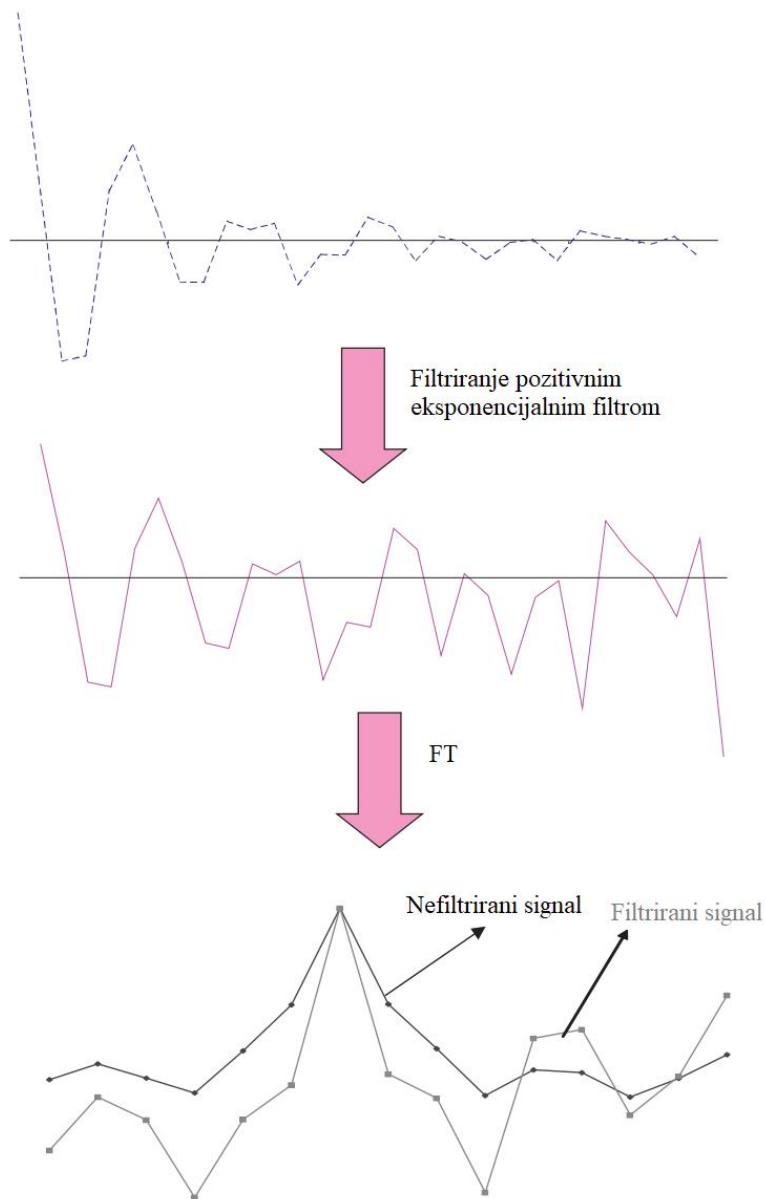


Slika 34. Fourierova transformacija brzo opadajuće vremenske serije

4.6. Utjecaj šuma

Realni spektri uvijek sadrže šum. Na **Slika 35** prikazan je jedan vremenski niz sa prisutnim šumom, proveden kroz eksponencijalni filter. Filtriranje vremenskog niza znatno pojačava šum

koji može interferirati sa signalima i otežati identifikaciju položaja vrhova. Iako se širina transformiranog vrha doista smanjila, šum se povećao.

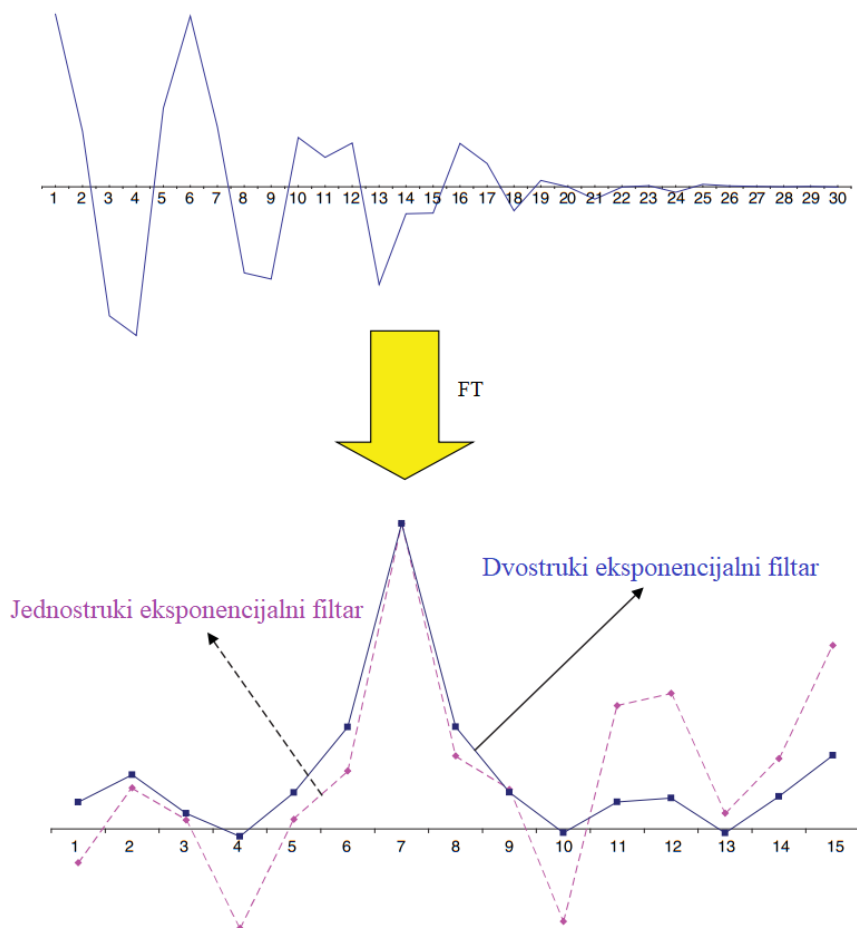


Slika 35. Rezultat provođenja pozitivnog eksponencijalnog filtra na vremenskom nizu (koji sadrži šum)

Jasno je da postoje ograničenja u smislu koliko se vrhovi smiju izoštriti, ali funkcija filtra može se poboljšati tako da se istovremeno primjenjuje smanjenje šuma i poboljšanje rezolucije. Jedna uobičajena metoda je množenje vremenske serije dvostrukim eksponencijalnim filtrom (**Slika 36**) u obliku:

$$g(t) = e^{+kt-vt^2} \quad (4.7)$$

gdje prvi (linearni) član eksponencijalno raste s vremenom i povećava razlučivost, a drugi (kvadratni) član smanjuje šum. Cilj je naglasiti sredinu vremenskog niza uz smanjenje kraja. Ova dva izraza mogu se optimizirati ako su širine vrhova i razine šuma unaprijed poznate. [2]



Slika 36. Dvostruko eksponencijalno filtriranje, istovremena primjena smanjenja šuma i poboljšanja rezolucije

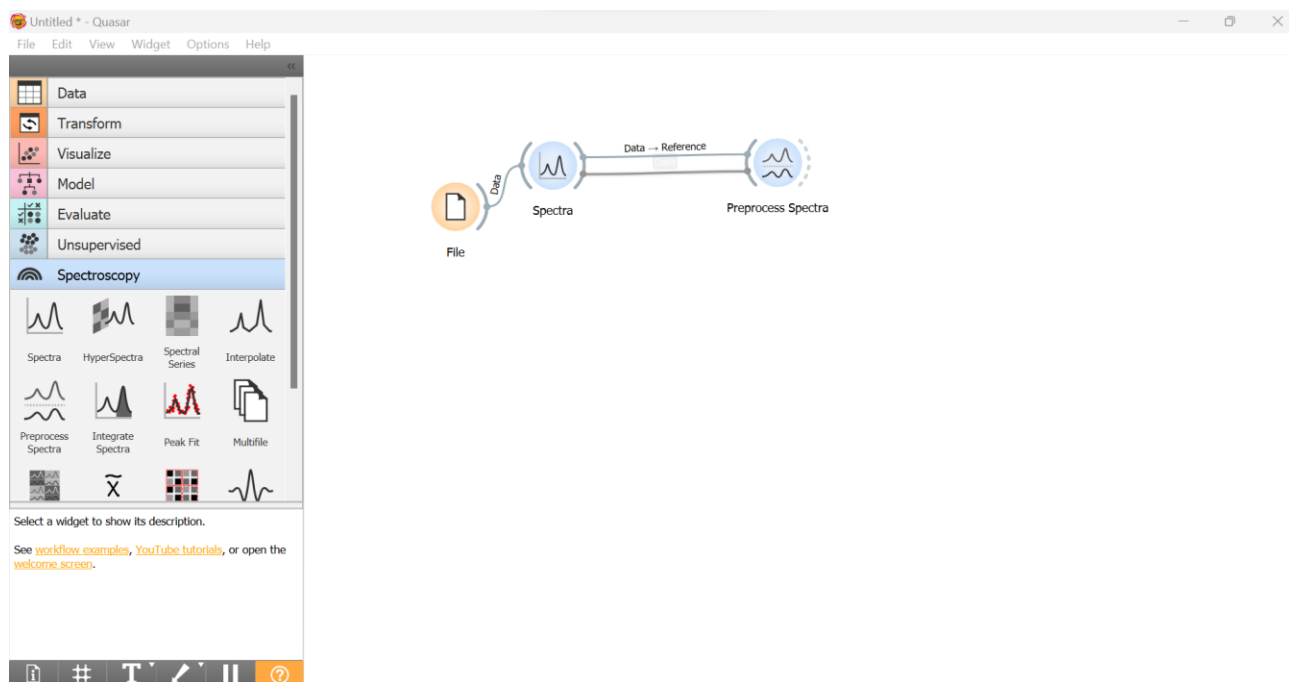
Slika 36 prikazuje dvostruki eksponencijalni filter, gdje isprekidana linija predstavlja rezultat jednostrukog eksponencijalnog filtra. Oblik vrha u transformaciji je gotovo jednako uzak kao onaj dobiven uporabom jednostrukog eksponencijalnog filtra, ali je šum izrazito smanjen. [2]

U mnogim primjenama spektroskopije, podaci se prikupljaju izravno kao vremenske serije i moraju se transformirati Fourierovim transformacijama da bi se dobio interpretabilan spektar. Prvi korak je transformacija podataka spektra u vremensku seriju koja fizički ne postoji, ali njome može upravljati računalo. Drugi korak je primjena Fourierovog filtra na vremensku seriju, treći korak je konačna transformacija natrag u spektar, što rezultira poboljšanom kvalitetom u podacima. Taj postupak se naziva **Fourierova samodekonvolucija**. [2]

5. REZULTATI

U ovom završnom radu je na primjeru laboratorijskih Ramanovih spektralnih podataka iz procesa kristalizacije djelatne tvari istražen utjecaj odabira različitih parametara i redoslijeda predobrade podataka na ukupnu kvalitetu i pouzdanost obrađenih spektralnih podataka.

Skupina Ramanovih spektralnih podataka na kojima je provedena predobrada podataka obrađena je u softverskom alatu *Quasar* koji je proširenje paketa *Orange*. *Orange* je popularan alat za vizualizaciju i analizu podataka otvorenog koda u Pythonu koji se koristi za interaktivnu analizu, vizualizaciju i strojno učenje. *Quasar* je zbirka alata namijenjena za analizu spektroskopskih podataka, a njegovo sučelje je prikazano na **Slika 37**. Procesi obrade podataka uključuju korekciju bazne linije, normalizaciju, FFT, EMSC, analizu vrhova diferenciranjem i izgladivanjem koje se ovom softverskom alatu vrlo jednostavno primjenjuju na učitane podatke.

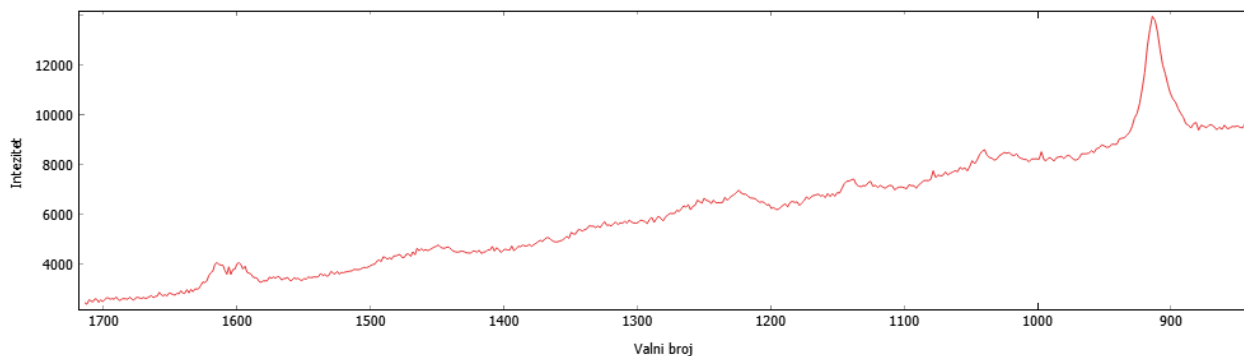


Slika 37. Prikaz sučelja programskog alata *Quasar* verzija: 1.7.0

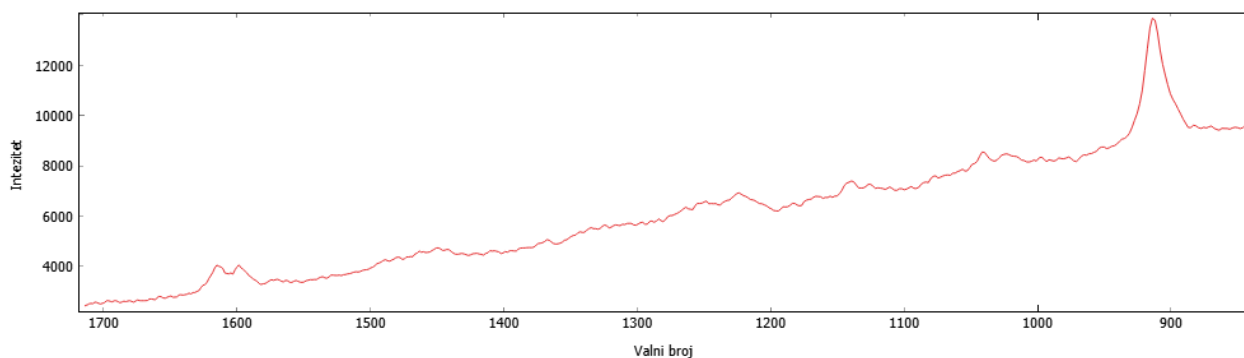
5.1. Grafički prikazi primjene metoda predobrade podataka

Na sljedećim slikama se mogu vidjeti grafički prikazi primjene metoda predobrade podataka. **Slika 38** prikazuje neobrađeni Ramanov spektar na kojemu se može vidjeti utjecaj bazne

linije, prisutnost šuma te nejasni pikovi. Primjenom metoda predobrade uklonit će se šumovi, korigirat će se bazna linija te povećati rezolucija podataka u svrhu boljeg uviđanja pikova.

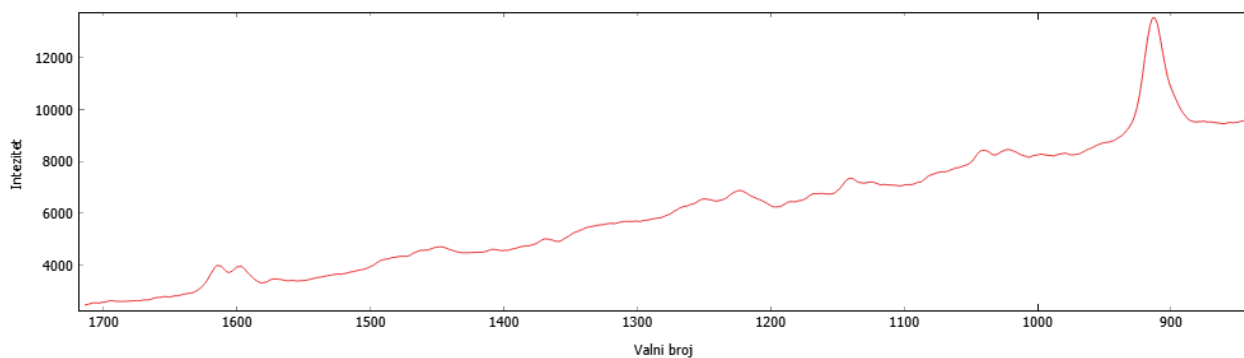


Slika 38. Neobrađeni spektar najviše koncentracije djelatne tvari



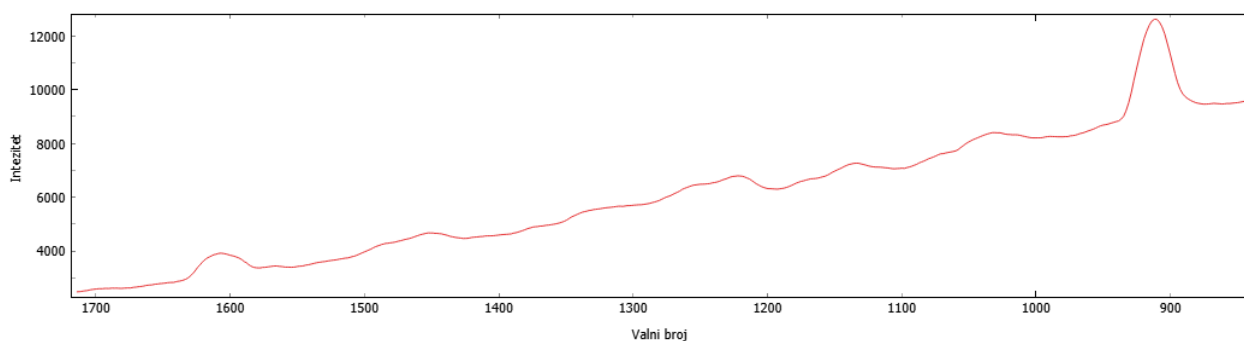
Slika 39. Primjena Savitzky-Golay filtra, prozor veličine 5 točaka

Na **Slika 39** prikazana je primjena SG filtra veličine prozora 5 točaka, polinoma 2. reda, derivacije 0. reda. Uočljivo je da je u spektru prisutno puno manje šuma, tj. da je obrađeni spektar puno izgladeniji.

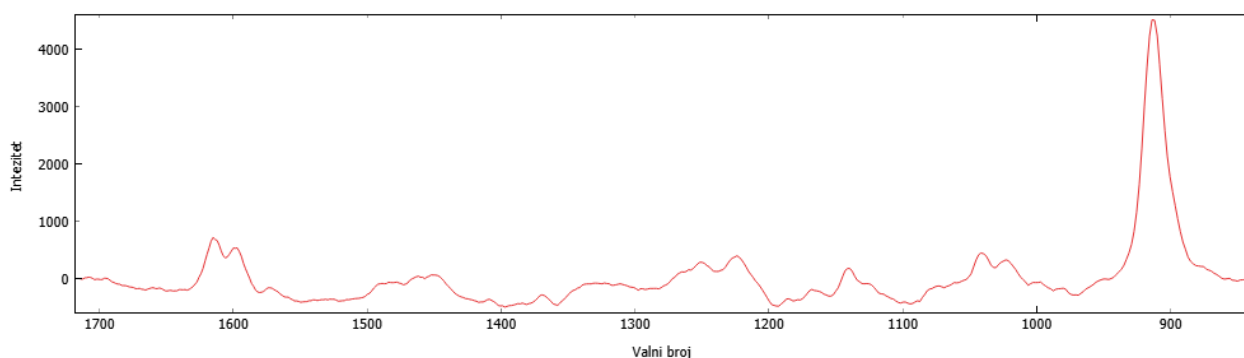


Slika 40. Primjena Savitzky-Golay filtra, prozor veličine 11 točaka

Na **Slika 40** prikazana je primjena SG filtra veličine prozora od 11 točaka. Odabirom većeg prozora za izgladivanje, očito je da se postiže veći stupanj izgladenosti spektra. SG filtar u ovom slučaju anulira šum u dijelovima spektra tako da se ne gubi informacija o karakterističnim pikovima, što je posebno vidljivo na **Slika 42** u području oko 1600 cm^{-1} . Na primjeru spomenutog područja vidljivo je anuliranje šumova i kako su dva karakteristična pika pročišćena i razvidnija. Uzimanjem većeg prozora filtra, što je prikazano na **Slika 41**, sa veličinom prozora 23 točke, karakteristična dva pika se gube, tj. stapaju u jedan pa je potrebno bolje prilagoditi veličinu prozora kako se ne bi izgubila analitička informacija.



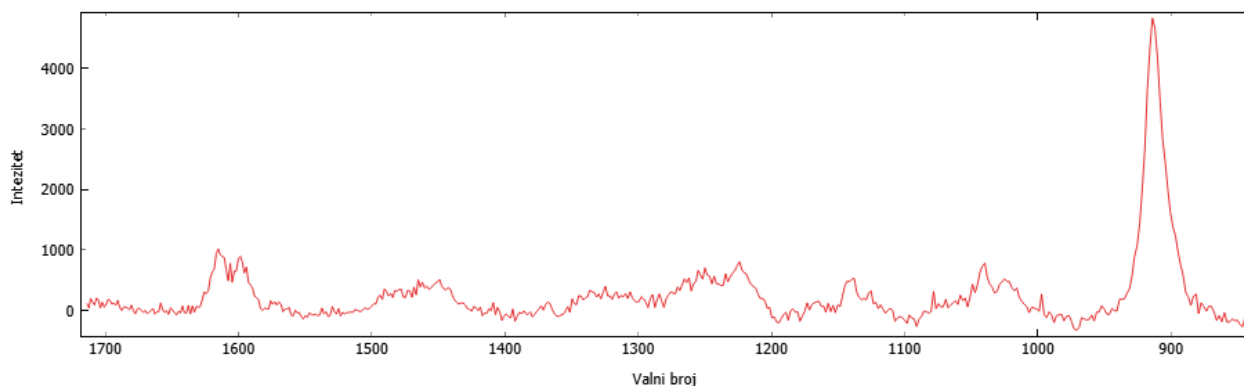
Slika 41. *Primjena Savitzky-Golay filtra, prozor veličine 23 točke*



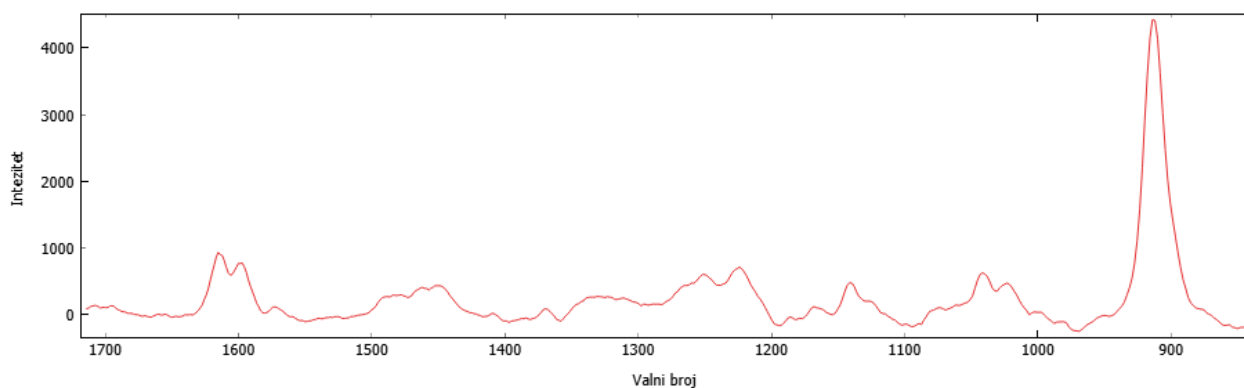
Slika 42. *Korekcija bazne linije*

Kako bi se mogle obraditi informacije iz spektra potrebno je korigirati baznu liniju radi uklanjanja neželjenih doprinosa kao što je fluorescencija, što je prikazano na **Slika 42**. Iako se neka razina spektralne analize može izvesti i bez korekcije bazne linije, uklanjanje bazne linije je u većini spektralnih predobrada, uvedena kao standardni korak. Korekcija bazne linije pomaže u preciznoj identifikaciji vrha, pouzdanoj kvantitativnoj analizi i smislenoj usporedbi između spektra.

Iako uklanjanje bazne linije nije nužan korak, važno je uzeti u obzir potencijalne posljedice i ograničenja izostave toga koraka. Korekcija bazne linije često se preporučuje u spektralnoj analizi, a posebno u tehnikama poput Ramanove spektroskopije, kako bi se poboljšala točnost i pouzdanost rezultata. Međutim, mogu postojati situacije u kojima korekcija bazne linije nije potrebna ili možda neće pružiti značajne prednosti. Na **Slika 43** prikazana je primjena metode uklanjanja bazne linije temeljene na asimetričnoj metodi najmanjih kvadrata na neobrađenom spektru. Vidi se da je bazna linija uklonjena, ali da spektar nije izgladen te da su prisutni šumovi, te se u svrhu izgladivanja spektra može primijeniti SG filtar (**Slika 44**).



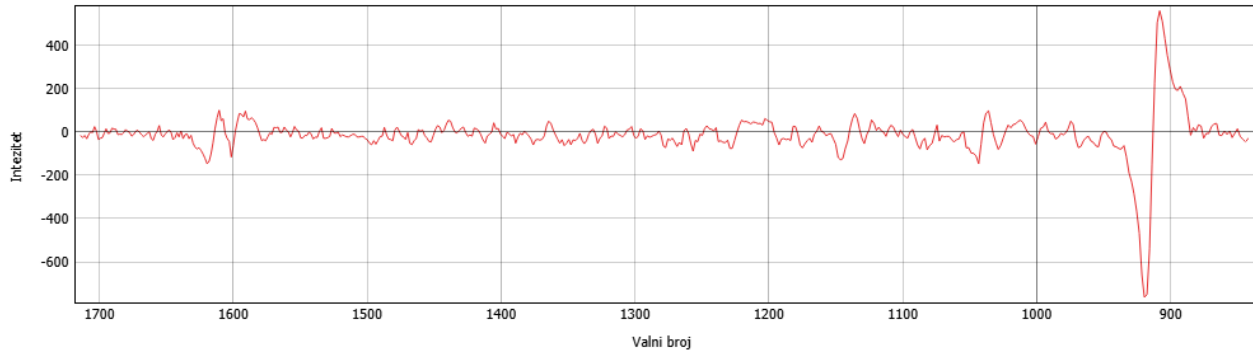
Slika 43. *Uklanjanje bazne linije temeljene na asimetričnoj metodi najmanjih kvadrata na neobrađenom spektru*



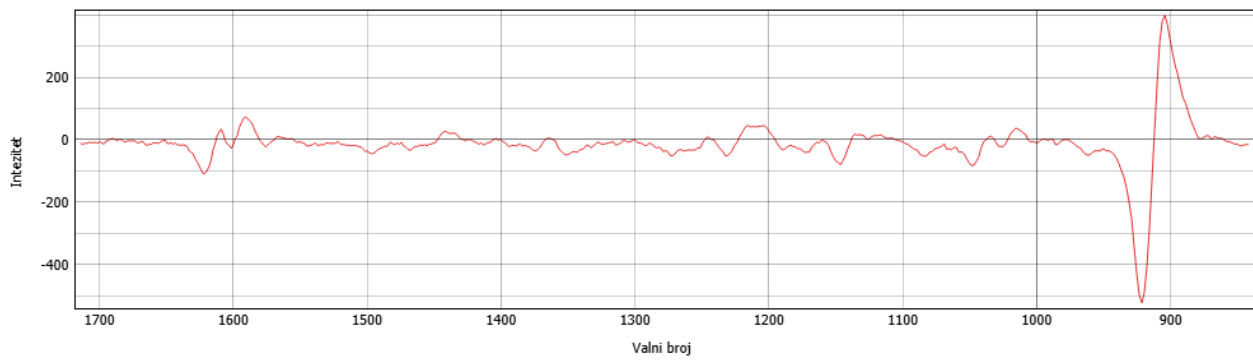
Slika 44. *Uklanjanje bazne linije temeljeno na asimetričnoj metodi najmanjih kvadrata na izgladenom spektru sa Savitzky-Golay filtrom, prozor veličine 11 točaka*

5.1.1. Primjene 1. i 2. derivacije spektra Savitzky-Golay filtrom

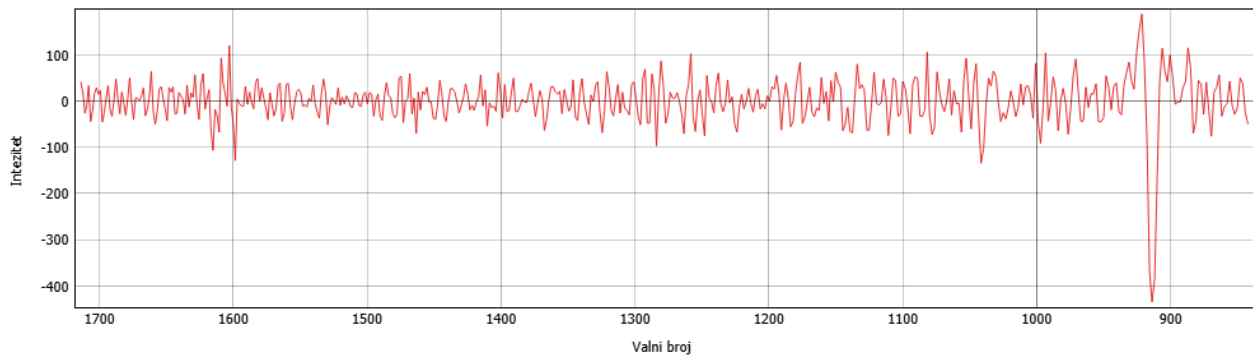
U nastavku će biti prikazane primjene 1. i 2. derivacije spektra Savitzky-Golay filtrom.



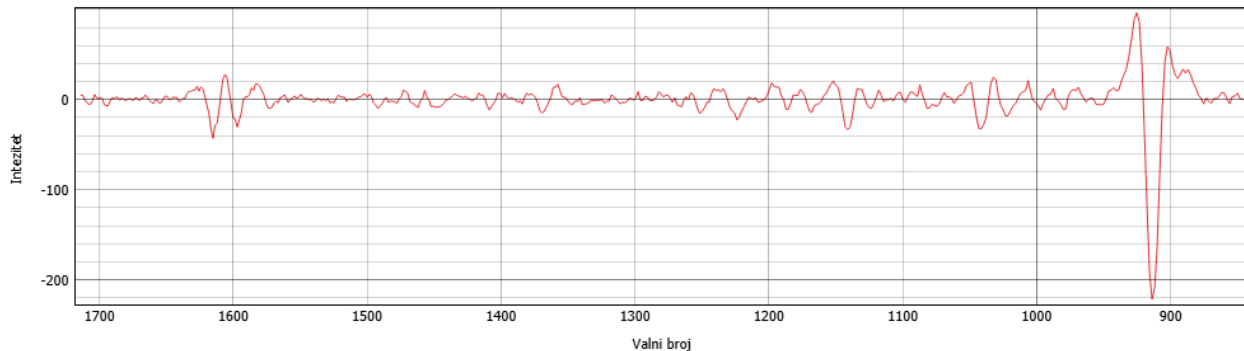
Slika 45. Prva derivacija spektra izgladenog Savitzky-Golay filtrom, prozor veličine 5 točaka



Slika 46. Prva derivacija spektra izgladenog Savitzky-Golay filtrom, prozor veličine 11 točaka



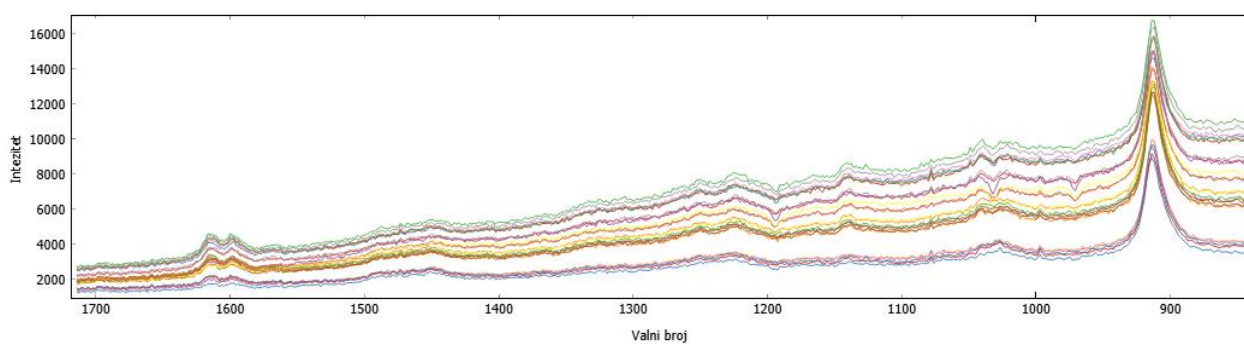
Slika 47. Druga derivacija spektra izgladenog Savitzky-Golay filtrom, prozor veličine 5 točaka



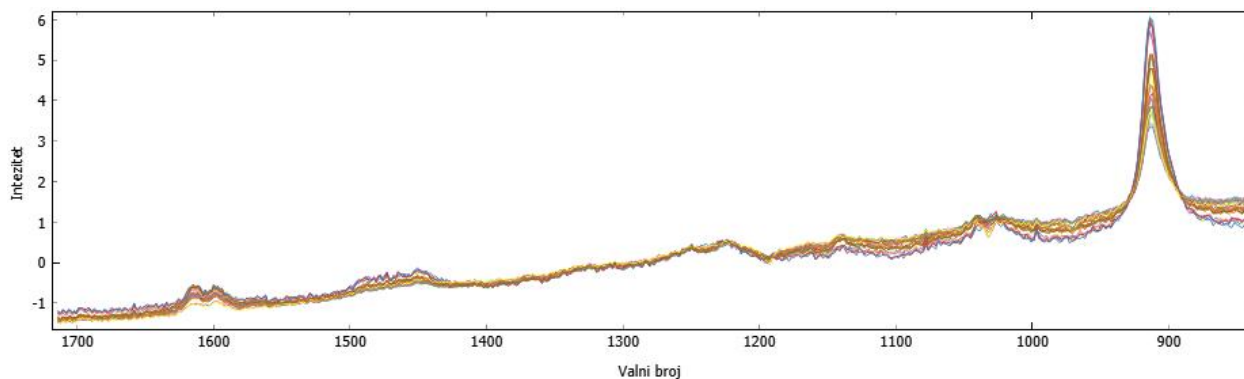
Slika 48. Druga derivacija spektra izgladenog Savitzky-Golay filtrom, prozor veličine 11 točaka

Na prikazima **Slika 45**, **Slika 46**, **Slika 47** i **Slika 48** spektar je proveden kroz prvu i drugu derivaciju. Iz slika se može vidjeti da je derivacija pojačala komponente više, tj. oštrije frekvencije, a zaobljenije komponente, tj. niže frekvencije je potisnula (fluorescencija i bazna linija). Uz pojačavanje pravih Ramanovih vrpca, može se vidjeti da je derivacija također pojačala i šumove koji su visoke frekvencije. Kada se uspoređuju prozori od 5 i 11 točaka, vrlo je uočljivo da spektar provučen kroz filter sa prozorom od 5 podatkovnih točaka sadrži puno više šuma nego onaj od 11 podatkovnih točaka, a kada se uspoređuje prva sa drugom derivacijom, primjećuje se da druga derivacija stvara puno više šuma nego prva. Također se može primijetiti da je primjenom derivacija uklonjen učinak bazne linije.

5.1.2. Predobrada spektara svih mjerenih koncentracija djelatne tvari

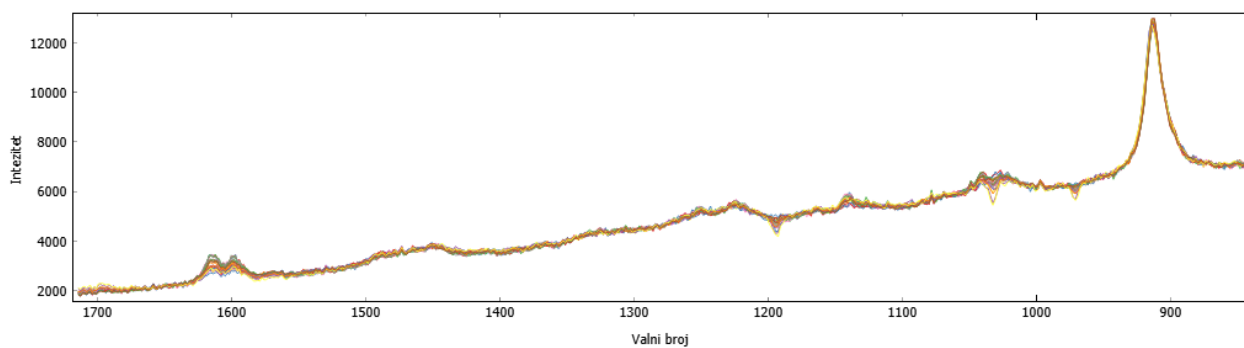


Slika 49. Neobrađeni spektri svih koncentracija djelatne tvari



Slika 50. *Primjena SNV metode na neobrađeni spektar*

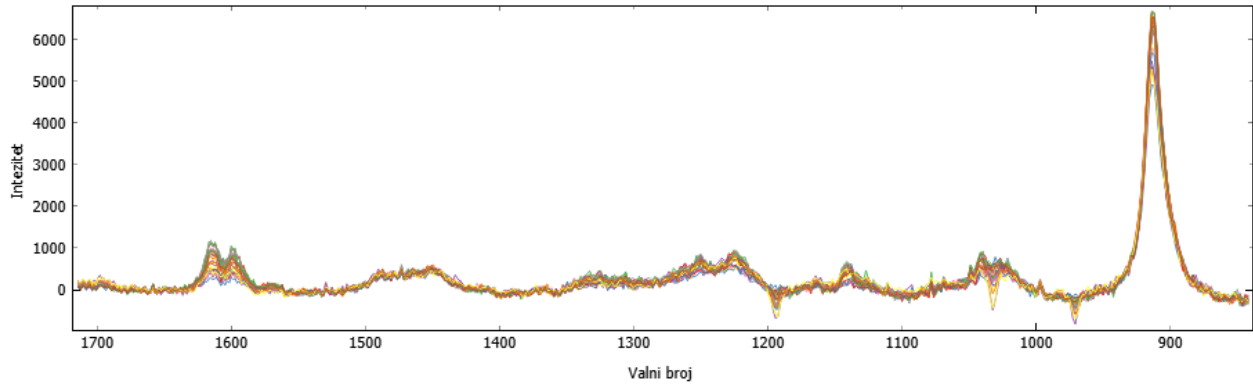
Na **Slika 49** prikazan je neobrađeni spektar, a na **Slika 50** je prikaz primjene SNV metode skaliranja na neobrađenom spektru. Sa **Slika 50** je vrlo jasno kako SNV metoda učinkovito uklanja raspršenja, tj. pomak (engl. *offset*) u spektru i dovodi spektre na istu skalu. Ono što se može primijetiti je da je pomak bazne linije još uvijek prisutan.



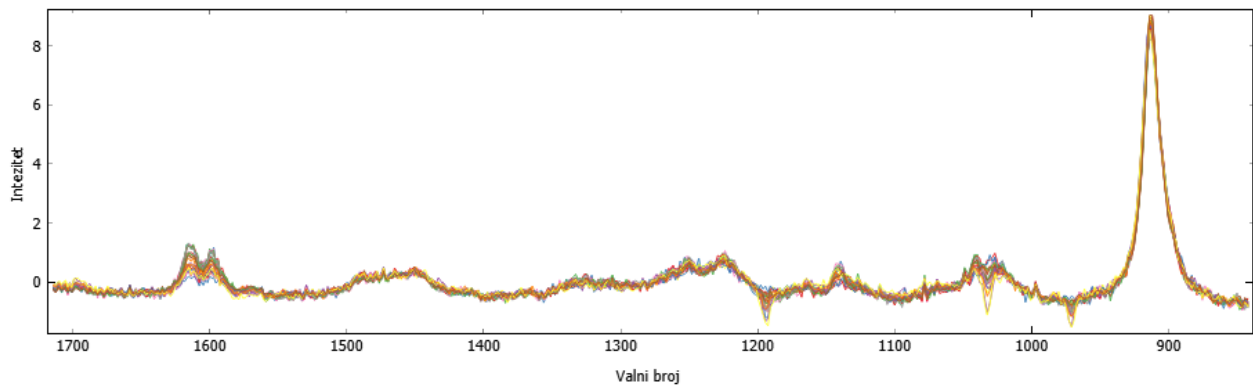
Slika 51. *Primjena EMSC metode na neobrađeni spektar*

Slika 51 prikazuje primjenu EMSC metode korekcije raspršenja. Na slici se vidi velika učinkovitost EMSC metode pri uklanjanju raspršenja i spektar je doveden na istu skalu. Kad bi se usporedile **Slika 50** i **Slika 51** vidljivo je da EMSC metoda bolje uklanja pomake u spektru.

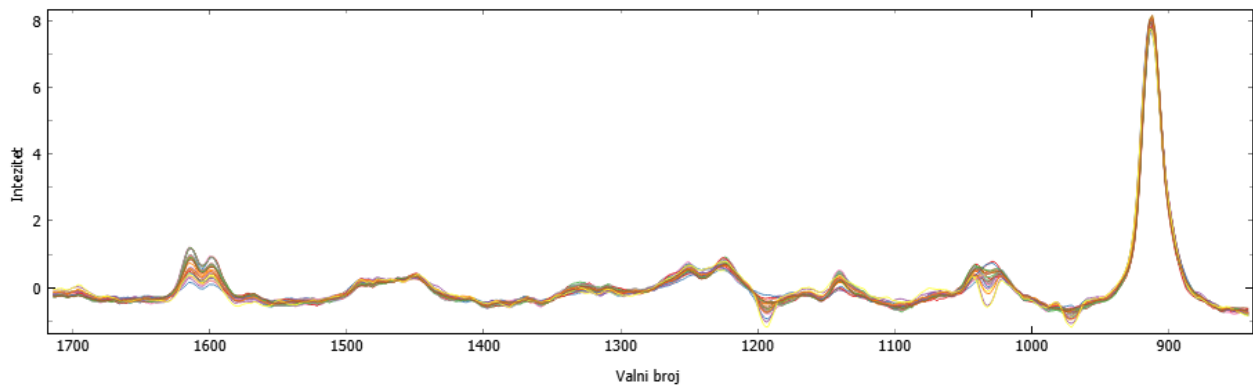
5.1.3. Primjena kombinacije različitih metoda predobrade spektara



Slika 52. *Uklanjanje bazne linije temeljene na asimetričnoj metodi najmanjih kvadrata na neobrađenom spektru*



Slika 53. *Uklanjanje bazne linije temeljene na asimetričnoj metodi najmanjih kvadrata na neobrađenom spektru i skaliranje SNV metodom*



Slika 54. *Uklanjanje bazne linije temeljene na asimetričnoj metodi najmanjih kvadrata na neobrađenom spektru, SNV skaliranje i SG sa veličinom prozora od 11 točaka*

Na prethodne tri slike (**Slika 52**, **Slika 53** i **Slika 54**) prikazana je primjena kombinacija raznih metoda predobrade na istom Ramanovom spektru. Najprije se uklonio pomak bazne linije pomoću metode temeljene na asimetričnoj metodi najmanjih kvadrata (**Slika 52**) zatim se na tom djelomično obrađenom spektru primijenila SNV metoda kako bi se uklonio pomak i raspršenje spektra (**Slika 53**), što je najviše vidljivo na piku kod valnog broja malo manjeg od 900 cm^{-1} . Nakon toga se pomoću SG filtra sa veličinom prozora od 11 točaka uklonio šum u spektru (**Slika 54**).

6. ZAKLJUČAK

U ovom završnom radu daje se pregled primjene najčešće korištenih metoda predobrade podataka u Ramanovoj i IR spektroskopiji. Predobrada podataka igra ključnu ulogu u izdvajanju pouzdanih informacija iz Ramanovih i IR podataka. Odabir adekvatnih metoda i pravilno odabranog redoslijeda predobrade spektralnih podataka omogućit će jednostavniju interpretaciju rezultata o promatranom uzorku. Prilikom provedbe predobrade potrebno je obratiti pozornost da se spektar ne preizgladi jer se time gube potencijalno vrlo bitni podaci. Poznavanje vrste šumova i njihovih izvora pomaže prilikom odabira metoda te njihovog temeljitog uklanjanja.

Iz primjene metoda predobrade podataka u ovom završnom radu može se zaključiti da je bitno odabrati optimalan stupanj i način primjene metoda predobrade.

Primjerice, kod primjene SG algoritma zaključeno je da je potrebno odrediti optimalnu veličinu prozora jer preveliki prozor može preizgladiti spektar i „progutati“ bitne pikove, tj. stopiti više pikova u jedan, dok primjena preuskog prozora neće dovoljno ukloniti šumove. Nadalje, što je veći stupanj derivacije potrebno je odabrati veću veličinu prozora s naglaskom da prozor ne smije biti prevelik jer će se izgubiti potencijalno bitni podatci iz spektra.

Postupak uklanjanja bazne linije, iako nije neophodan, može značajno pridonijeti lakoći iščitavanja bitnih informacija, a najviše pridonosi predobradi spektra uklanjanjem fluorescencije.

Ponekad je potrebno usporediti više metoda za obradu odabranih spektara kako bi se odabrala ona optimalna. Na primjeru predobrade spektralnih podataka djelatne tvari u ovom radu je EMSC metoda nešto bolje obradila spektralne podatke nego SNV metoda, što ne mora nužno vrijediti za sve spektre.

Kroz ovaj završni rad, istraženo je kako pravilna primjena predobrade spektralnih podataka može značajno poboljšati kvalitetu i pouzdanost analize, čineći ih neizostavnim koracima u raznovrsnim aplikacijama od znanosti materijala, kemije i biomedicine do farmaceutike i industrijskog inženjerstva.

7. POPIS KRATICA

IR	infracrveno zračenje
NIR	blisko infracrveno zračenje
MSC	multiplikativna korekcija signala
SNV	standardna normalna varijacija
CRE	kozmičke zrake
CCD	naponom povezani uređaj
NMR	nuklearna magnetska rezonancija (spektroskopija)
UV	ultraljubičasto zračenje
SNR	omjer signala i šuma
LOD	granica detekcije
ATR-MIR	srednja infracrvena prigušena totalna refleksija
<i>E</i>	Energija
FT	Fourierova transformacija
MA	metoda filtriranja pomičnim prosjekom
M	širina filtra
SG	Savitzky-Golay
RMS	metoda zaglađivanja podataka medijanom
FTIR	vezana tehnika infracrvene spektroskopije sa Fourierovim transformacijama
IPBF	iterativno prilagođavanje bazne linije polinomom
EMSC	proširena multiplikativna korekcija signala
DFT	diskretna Fourierova transformacija
FTNMR	vezana tehnika nuklearne magnetske rezonancije sa FT

8. LITERATURA

- [1] Mandac Zubak, A., Razvoj i validacija metoda vibracijske spektroskopije za identifikaciju pročišćenih meningokoknih polisaharida serogrupa A i C, Zagreb, 2021.
- [2] Brereton, R., Chemometrics: Data driven extraction for science, John Wiley & Sons, Hoboken, New Jersey, 2018., 121-181.
- [3] Rinnan, Å., Nørgaard, L., Berg, F.v.d., Thygesen, J., Bro, R., Engelsen, S. B., Data Pre-processing, u: Sun, D. W., Infrared Spectroscopy for Food Quality Analysis and Control, San Diego, Elsevier, 1 (2009) 29–50.
- [4] Gautam, R., Vanga, S., Ariese, F., Umopathy, S., Review of multidimensional data processing approaches for Raman and infrared spectroscopy. EPJ Techniques and Instrumentation, 2 (2015) 1-15.
- [5] Jednačak, T., Novak, P., Procesne analitičke tehnike temeljene na vibracijskoj spektroskopiji in-line i primjena u industriji, Kem. Ind., 62 (2013) 71-80.
- [6] Haswell, S., Practical Guide to Chemometrics, Marcel Dekker Inc., 1992., str. 241.
- [7] <https://www.shimadzu.com> (pristup 30. srpnja 2023.)
- [8] Das, R. S., Agrawal, Y. K., Raman spectroscopy: Recent advancements, techniques and applications, Vibrational Spectroscopy 57 (2011) 163–176.
- [9] Ferrer-Galindo, L., Sañu-Ginarte, A. D., Fleitas-Salazar, N., Ferrer-Moreno, L. A., Rosas, R. A., Pedroza-Montero, M., Riera, R., Denoising and principal component analysis of amplified Raman spectra from red blood cells with added silver nanoparticles, Journal of Nanomaterials, 2018.
- [10] Bocklitz, T., Walter, A., Hartmann, K., Rösch, P., Popp, J., How to pre-process Raman spectra for reliable and stable models?, Analytica chimica acta, 704 (2011) 47-56.
- [11] Liu, Y., Adversarial nets for baseline correction in spectra processing, Chemometrics and Intelligent Laboratory Systems, 213, 2021.

[12] Wentzell, P. D., Brown, C. D., Signal processing in analytical chemistry, Encyclopedia of Analytical Chemistry, 11 (2000) 12-15.

[13] Chau, Y., Liang, Z., Gao, J., Shao, X., Chemometrics: From Basics to Wavelet Transform, John Wiley & Sons, Inc., New Jersey, 2004., 10-69.

[14] Savitzky, A., Golay, M. J., Smoothing and differentiation of data by simplified least squares procedures, Analytical chemistry, 36, (1964) 1627-1639.

[15] Felinger, A., Data Analysis and Signal Processing in Chromatography, Elsevier Science B.V., Amsterdam, 1998., 143-183.