



## Sadržaj

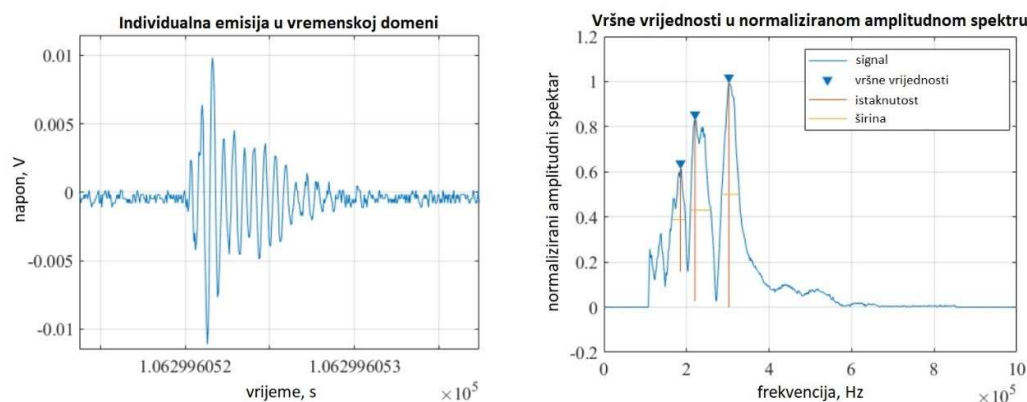
|   |    |
|---|----|
| Uvod .....  | 1  |
| 1. Opis korištenih algoritama.....  | 3  |
| 1.1. Proces odabira algoritma za grupiranje .....                                   | 3  |
| 1.2. OPTICS .....   | 5  |
| 1.3. Gradijentno grupiranje.....  | 9  |
| 1.4. Spajanje prozora .....   | 12 |
| 1.5. Sažimanje rezultata.....   | 14 |
| 2. Ispitivanje skupa algoritama nad sintetičkim podacima.....                       | 17 |
| 2.1. Metrika za klasifikaciju .....   | 18 |
| 2.2. Sintetički setovi podataka.....  | 19 |
| 2.3. Ovisnost točnosti o gustoći grupe, eksperiment 1 .....                         | 19 |
| 2.3.1. Eksperiment 1, prvi dio .....  | 20 |
| 2.3.2. Eksperiment 1, drugi dio .....   | 21 |
| 2.4. Ovisnost točnosti o odnosu veličine prozora i preklapanju, eksperiment 2 ..... | 23 |
| 2.5. Ovisnost točnosti o udaljenosti grupa, eksperiment 3 .....                     | 25 |
| 2.5.1. Eksperiment 3, prvi dio .....  | 25 |
| 2.5.2. Eksperiment 3, drugi dio .....   | 27 |
| 2.6. Ovisnost točnosti o broju grupa, eksperiment 4 .....                           | 28 |
| 2.7. Ovisnost točnosti o obliku i orijentaciji grupa, eksperiment 5 .....           | 29 |
| 3. Ispitivanje algoritma nad stvarnim podacima.....                                 | 31 |
| 3.1. Fiksni parametri.....  | 32 |
| 3.2. Adaptivno određivanje parametara iz satnog broja emisija.....                  | 33 |
| 3.3. Adaptivno određivanje parametara iz gustoće podataka .....                     | 33 |
| 4. Analiza implementacije na ugradbenom sustavu .....                               | 39 |
| 4.1. Vremenska analiza.....   | 39 |

|                      |    |
|----------------------|----|
| 4.2. Potrošnja ..... | 42 |
| 5. Zaključak .....   | 45 |
| Literatura .....     | 46 |
| Sažetak.....         | 47 |
| Summary.....         | 48 |

# Uvod

Tijekom kavitacije i drugih fizikalnih procesa u provodnom tkivu bilja, ksilemu, dolazi do embolizacije, odnosno pojave mjehurića plina koji blokiraju provod vode. Do kavitacije dolazi kada se tekućina unutar ksilema nalazi pod tolikim tlakom da se zrak unutar nje odvoji i stvara mjehuriće. Pucanje tako nastalih mjehurića proizvodi mehaničke valove ultrazvučnih frekvencija, ultrazvučne akustičke emisije (UAE). Premda je pojava ovakvih emisija uobičajena u životnom vijeku biljke, njihova učestalost se povećava pod povećanim vodnim stresom [1]. Vodni stres je pojam kojim se općenito označava nedostatak vlage u tkivu biljaka. Najčešći uzroci povećanog vodnog stresa su nedovoljno prikupljanje vode kroz sustav korijenja i povećana stopa transpiracije, odnosno isparavanja kroz listove [2].

Pojedina UAE najčešće je val vrlo kratkog trajanja s jednom ili više vršnih frekvencija u amplitudnom spektru (Slika 0.1).



a) Prikaz tipične UAE u vremenu

b) Prikaz amplitudnog spektra tipične UAE

Slika 0.1 Vremenski i amplitudni spektar UAE [3]

Primjer vremensko-frekvencijskog prikaza snimljenih UAE prikazan je na Slici 1.2. Točka na prikazu predstavlja emisiju određene frekvencije u nekom trenutku u vremenu. Analizom zabilježenih UAE može se pridonijeti razumijevanju fizikalnih procesa vezanih uz ponašanje biljnog tkiva tijekom povećanog vodnog stresa. Većina postojećih znanstvenih radova koji se bave obradom UAE, analizu temelje na praćenju određenih karakteristika u vremenskoj ili frekvencijskoj domeni. Primjer najčešćeg i najjednostavnijeg pristupa je praćenje kumulativnog broja UAE u vremenu [4] s pretpostavkom da proces sušenja prati

povećanje učestalosti emisija. U članku [5] autori porast apsolutne energije pojedinačnih UAE povezuju s gubitkom relativne vlage. Jedan od primjera frekvencijske analize opisan je u članku [6] gdje su autori automatiziranim algoritmom klasifikacije utvrdili postojanje dviju grupa emisija, visoko frekvencijske i nisko frekvencijske. Zaključeno je da embolizaciju ksilema u tkivu biljke prate nisko frekvencijske UAE (do 300 kHz), dok su visoko frekvencijske UAE najvjerojatnije rezultat fraktura tkiva drva.

Pristup analizi UAE opisan u ovome radu temelji se na prepoznavanju frekvencijskih linija (grupa) u vremensko-frekvencijskom prikazu UAE (Slika 1.3). Pretpostavka je da pojava grupiranja UAE započinje kao posljedica povećanog vodnog stresa uslijed sušenja biljke. Za obradu emisija je korišten skup algoritama opisan u poglavlju 1, od kojih je za rezultate najznačajniji OPTICS algoritam za grupiranje [7]. U poglavlju 2 navedeni su rezultati ispitivanja skupa algoritama nad sintetičkim podacima dok su u poglavlju 3 prikazani rezultati obrade za stvarne setove podataka. U poglavlju 4 opisana je vremenska analiza implementacije algoritma te su navedeni rezultati ispitivanja potrošnje ugradbenog sustava koji izvodi algoritam.

Konačni cilj ovoga diplomskog rada je implementacija navedenog skupa algoritama na ugradbenom sustavu s mogućnošću sekvencijalne obrade ulaznih podataka UAE u stvarnom vremenu. Ugradbeni sustav će u konačnici biti napajan baterijskim napajanjem te će povremeno, svakih nekoliko dana ili sati, skupom algoritama obrađivati podatke UAE. Rezultat obrade podataka bit će prepoznate grupe u vremensko-frekvencijskom prikazu, a za prijenos rezultata koristit će se neki oblik bežične komunikacije.

# 1. Opis korištenih algoritama

Za implementaciju skupa algoritama korišteni su programski paket Matlab i razvojni alat Keil uVision. Programski paket Matlab korišten je za razvoj testne verzije algoritma koja je služila za provjeru rada i pomoć pri izradi algoritma na ugradbenom sustavu. Također, Matlab je korišten pri analizi rezultata ispitivanja i prijenosu podataka preko UART veze s razvojnog sustava na računalo. Razvojni alat Keil uVision korišten je za implementaciju skupa algoritama na ugradbenom sustavu. Korišteni ugradbeni sustav je STM32 NUCLEO-144 sa STM32L4R5ZI mikrokontrolerom [8].

Skup algoritama za analizu UAE sastoji se od 4 cjeline (Slika 1.1):

- OPTICS algoritam,
- Gradijentno grupiranje,
- Povezivanje prozora,
- Sažimanje rezultata.

U nastavku poglavlja opisane su svaka od cjelina zasebno.

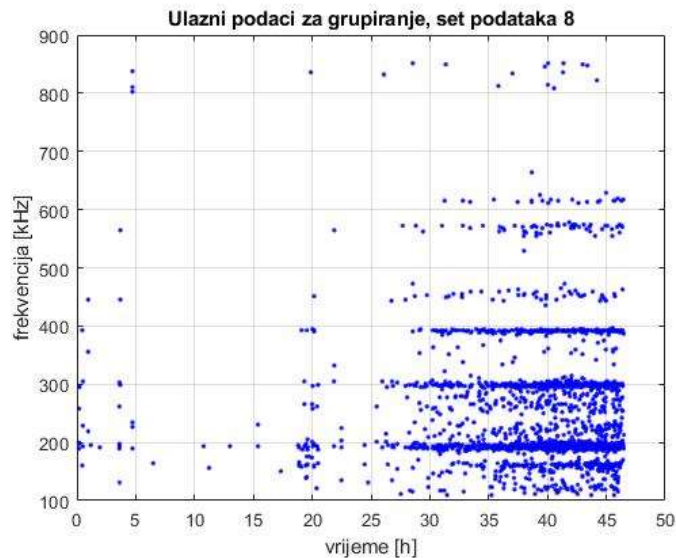


Slika 1.1 Blok dijagram skupa algoritama

## 1.1. Proces odabira algoritma za grupiranje

Ključni dio skupa algoritama za obradu ulaznih podataka snimljenih UAE je algoritam za grupiranje koji zadani skup podataka grupira prema određenim parametrima u grozdove (*clusteres*), odnosno grupe. Takvih algoritama danas postoji mnogo, no ključno je bilo odabrati algoritam koji optimalno odgovara zadanoj funkciji i tipičnim skupovima

podataka snimljenih UAE. Na Slici 1.2 prikazan je set podataka za grupiranje dobiven snimanjem UAE.



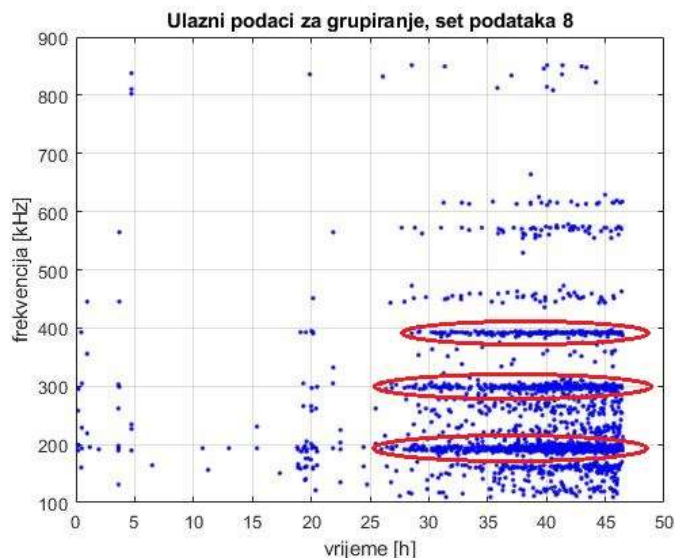
Slika 1.2 Vremensko-frekvencijski prikaz UAE

Glavni zadatak algoritma za grupiranje je prepoznati grupiranje točaka oko već spomenutih frekvencijskih linija (Slika 1.3). Iz glavnog zadatka proizlazi nekoliko zahtjeva koje algoritam treba biti u stanju ispuniti kako bi grupiranje bilo izvedeno kvalitetno i uspješno.

Zahtjevi na algoritam za grupiranje:

1. Neovisnost o broju grupa – algoritam mora grupirati bez prethodne informacije o broju grupa.
2. Neovisnost o obliku i orijentaciji grupa – algoritam mora biti u stanju prepoznati grupe svih oblika i orijentacija.
3. Mogućnost sekvencijalnog rada – algoritam mora biti u stanju prepoznati stvaranje ili prisutnost grupa na kraćim vremenskim intervalima.
4. Neovisnost o apsolutnom broju emisija – broj točaka u grupi ne smije utjecati na rezultate grupiranja. Algoritam mora biti u stanju prepoznati grupe različitog broja točaka (različitih gustoća).
5. Niska složenost – poželjno je da algoritam ima što manju vremensku složenost zbog rada u stvarnom vremenu (kraće trajanje i niža potrošnja).

6. Prepoznavanje šuma – Algoritam ne smije grupirati sve točke, mora biti u stanju prepoznati šum.



Slika 1.3 Zaokružene frekvencijske linije (grupe)

Možda najpoznatiji i najjednostavniji algoritam za grupiranje K-Means ne ispunjava čak nekoliko zadanih zahtjeva [9]. Kao parametar mu je potrebno zadati broj grupa, može prepoznati samo grupe kružnog oblika i grupira sve točke (ne prepoznaje šum). Algoritam EM-GMM pretpostavlja Gaussovu razdiobu točaka te ih grupira prema vjerojatnosti da pripadaju nekoj grupi. Međutim, EM-GMM se također može odbaciti s obzirom da mu je ulazni parametar broj grupa te nije u stanju prepoznati šum [10]. DBSCAN spada u kategoriju algoritama za grupiranje ovisnih o gustoći, što znači da u ulaznom setu podataka pronalazi područja veće gustoće točaka u ovisnosti o zadanim parametrima [11]. Kao takav DBSCAN ispunjava većinu zahtjeva, ali nije u stanju prepoznati grupe različitih gustoća.

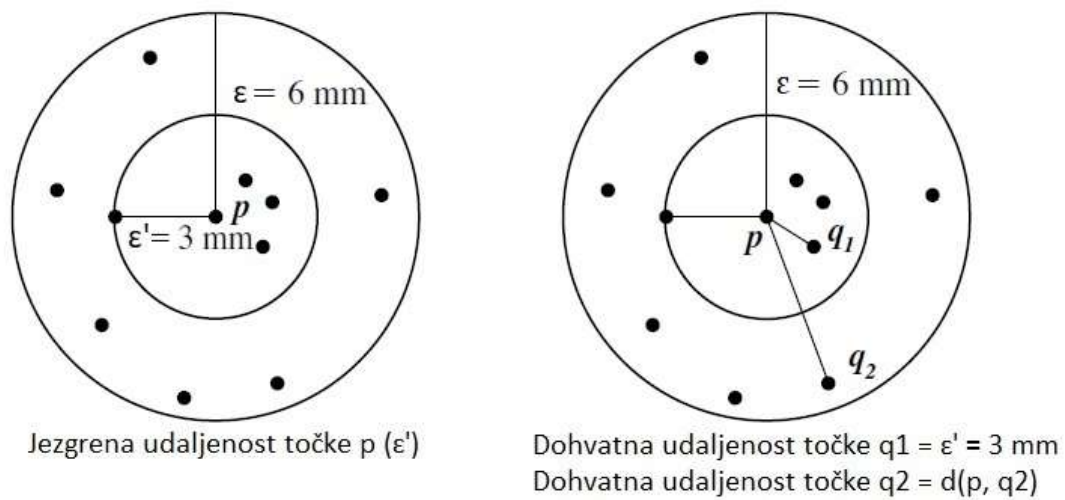
Odabrani algoritam za grupiranje korišten u ovome radu je OPTICS [7]. Algoritam za grupiranje ovisan o gustoći s mogućnosti prepoznavanja grupa različitih oblika, veličina, gustoća i orijentacija. OPTICS algoritmu nije potrebno znanje o broju grupa, može raditi sekvencijalno i u stanju je prepoznati šum. Jedini problem i neispunjeni zahtjev je niska složenost koja iznosi  $O(n^2)$ .

## 1.2. OPTICS

Prije opisa algoritma potrebno je definirati neke od korištenih pojmova. Na Slici 1.4 dan je prikaz dvaju pojmova koji su važni za razumijevanje rada OPTICS (*Ordering points*



to identify the clustering structure) algoritma, jezgrena udaljenost (*core-distance*) i dohvatna udaljenost (*reachability-distance*). Jezgrena udaljenost se definira kao minimalna euklidska udaljenost potrebna da bi točka „p“ imala minimalno  $N_{\min}-1$  susjeda unutar te udaljenosti. U slučaju da je ta udaljenost veća od parametra  $\varepsilon$ , jezgrena udaljenost ostaje nedefinirana. Ako točka „p“ zadovoljava uvjet  $N_{\min}$  na udaljenosti  $\varepsilon$ , za sve njene susjede mogu se izračunati dohvatne udaljenosti, a samu točku nazivamo jezgrena točka (*core point*). Dohvatna udaljenost se za točku „q“ definira kao veća od vrijednosti euklidske udaljenosti točaka „p“ i „q“ i jezgrene udaljenosti točke „p“. Dohvatna udaljenost ostaje nedefinirana ako iznosi više od parametra  $\varepsilon$ .

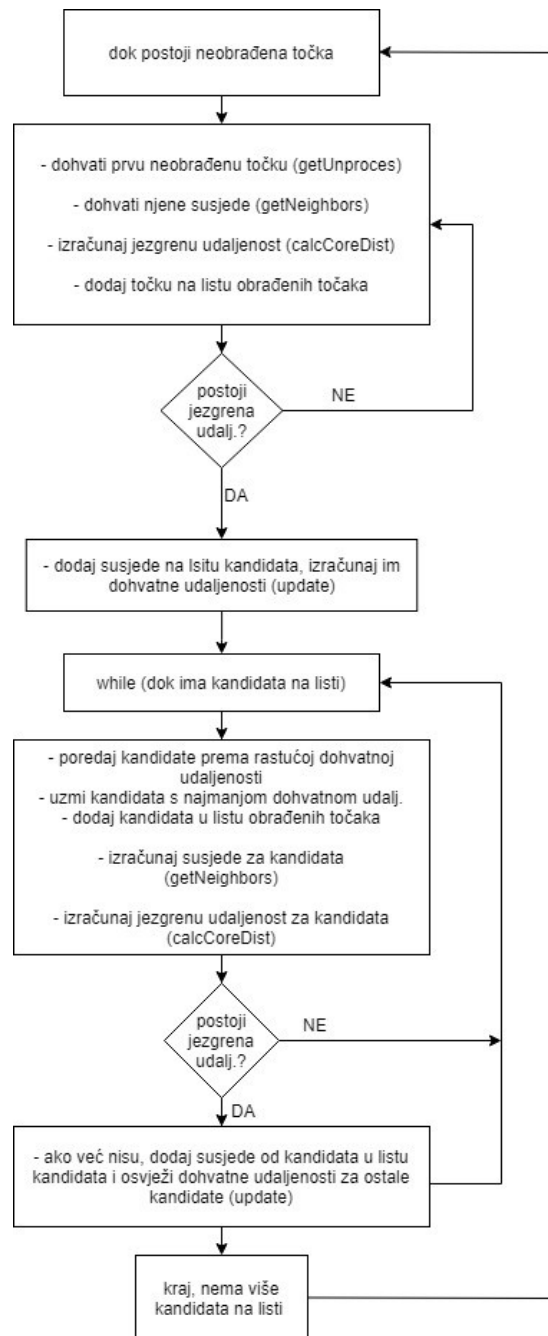


Slika 1.4 Prikaz pojmova dohvatne (*reachability*) udaljenosti i jezgrene (*core*) udaljenosti [12]

OPTICS algoritam ovisno o ulaznim parametrima  $N_{\min}$  (minimalni broj susjednih točaka koji tvore grupu) i  $\varepsilon$  (najveća udaljenost na kojoj se provjerava uvjet  $N_{\min}$ ) na svojem izlazu daje konačnu dohvatnu udaljenost za svaku točku u setu podataka te poredak kojim su točke obrađivane.

Dijagram toka OPTICS algoritma prikazan je na Slici 1.5, a implementacija prati pseudokod iz članka [7]. Algoritam započinje dohvatom prve neobrađene točke. Pronalaze se susjedi točke, računa se jezgrena udaljenost točke i indeks točke iz originalnog seta podataka prema se na listu obrađenih točaka. U slučaju da je jezgrena udaljenost definirana, algoritam nastavlja s radom tako da smješta sve susjedne točke na listu kandidata. U sljedećem koraku se za sve kandidate računa dohvatna udaljenost u ovisnosti o trenutno obrađivanoj točki. Kandidati se zatim poredaju prema rastućoj dohvatnoj udaljenosti, a za sljedeću točku za obrađivanje uzima se kandidat s najmanjom dohvatnom udaljenosti.

Postupak se zatim za odabranog kandidata ponavlja te se iterativno nastavlja sve dok u listi kandidata ima potencijalnih kandidata. U slučaju da je jezgrena udaljenost nedefinirana, algoritam dohvaća sljedeću neobrađenu točku. Važno je napomenuti da se tokom izvođenja algoritma lista kandidata cijelo vrijeme osvježava kako obrađujemo nove točke. Također, dohvatna udaljenost se za svaku točku osvježava ovisno o tome koji se njen susjed trenutno obrađuje.

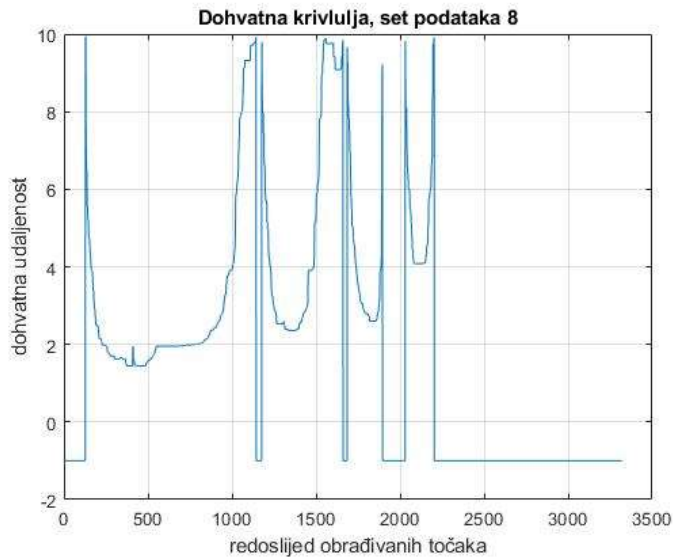


Slika 1.5 Dijagram toka OPTICS algoritma [7]

Algoritam se sastoji od nekoliko glavnih funkcija koje su na dijagramu toka (Slika 1.5) upisane unutar zagrada. Popis funkcija je sljedeći:

- `getNeighbors` – Funkcija za određenu točku traži sve susjede koji se nalaze unutar udaljenosti  $\epsilon$  od te točke. Traženje susjeda podrazumijeva računanje udaljenosti do svake druge točke u setu podataka. Funkcija kao izlaz daje indekse svih susjeda promatrane točke.
- `calcCoreDist` – Funkcija kao ulaz prima indekse susjeda trenutno obrađivane točke te pripadajuće udaljenosti. Ukoliko točka ima minimalno  $N_{\min}-1$  susjeda, na izlazu vraća jezgrenu udaljenost, a u suprotnome jezgrena udaljenost ostaje nedefinirana. Jezgrena udaljenost dobije se sortiranjem udaljenosti susjeda i uzimanjem  $(N_{\min}-1)$ -te udaljenosti kao rezultat.
- `update` – Funkcija služi za računanje i osvježavanje polja dohvatne udaljenosti za sve kandidate. Kao ulaz također prima indekse susjeda i pripadajuće udaljenosti. Računanje i osvježavanje dohvatne udaljenosti odrađuje se jedino u slučaju da je kandidat prethodno neobrađen (ne nalazi se na listi obrađenih točaka).
- `sortSeeds` – Funkcija koja se koristi za sortiranje kandidata kako bi se odredila sljedeća točka za obradu. Za sortiranje je implementiran Insertion sort algoritam s obzirom da su u većini slučajeva kandidati već uglavnom sortirani (više u poglavlju 4.1).
- `getUnproces` – Funkcija koja služi za dohvat sljedeće neobrađivane točke u slučaju da je lista kandidata prazna ili dohvata prve točke u slučaju početka rada algoritma.

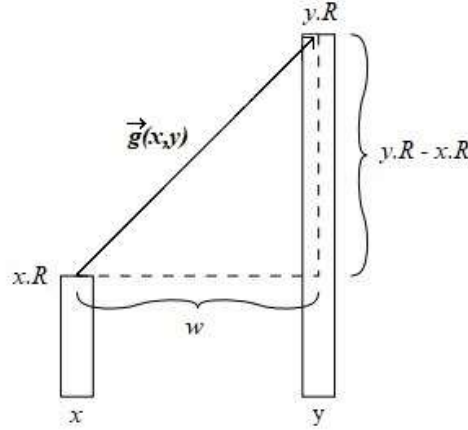
Konačni rezultat OPTICS algoritma naziva se dohvatna krivulja (*reachability curve*) (Slika 1.6), a same grupe možemo izdvojiti tako da prepoznamo udubine na krivulji. Točke kojima dohvatna udaljenost iznosi -1 (nedefinirana) ostale su negrupirane i smatraju se šumom. Kasnije se za potrebe ekstrakcije grupa iz dohvatne krivulje vrijednosti -1 zamjenjuju proizvoljnim, velikim brojem ( $2^{10}$ ). Ekstrakcija grupa iz dohvatne krivulje odrađuje se algoritmom gradijentnog grupiranja.



Slika 1.6 Grafički prikaz dohvatne krivulje

### 1.3. Gradijentno grupiranje

Postupak ekstrakcije grupa iz dohvatne krivulje računalu je nešto kompliciraniji proces nego golome oku. Gradijentno grupiranje (*Gradient Clustering*, GC) se temelji na određivanju početnih i krajnjih točaka grupa koje moraju zadovoljavati određene uvjete da bi ih se takvima prozvalo. Algoritam je implementiran prema pseudokodu iz članka [9]. Glavni uvjet je da točka početka ili kraja grupe mora biti infleksija (indeks infleksije veći od 0). Zatim se, ovisno o nagibu (gradijentu), može zaključiti o tome je li točka početak ili kraj grupe. Algoritam kao ulaz prima 2 parametra,  $t$  i  $w$ . Parametar  $t$  predstavlja kut nagiba između dva susjedna gradijentna vektora, a parametar  $w$  predstavlja fiktivnu udaljenost između susjednih uzoraka dohvatne krivulje. Na Slici 1.7 prikazana su dva uzorka dohvatne krivulje, označeni sa  $x$  i  $y$ . Iznosi njihovih dohvatnih udaljenosti označeni su sa  $y.R$  i  $x.R$ , a gradijenti vektor između dva uzorka sa  $\vec{g}(x,y)$ . Za svaku od točaka (osim prve i zadnje) dohvatne krivulje tako možemo izračunati indeks infleksije i gradijent pomoću izraza (1) i (2), gdje  $y$  predstavlja središnju točku,  $x$  prethodnu točku i  $z$  sljedeću točku dohvatne krivulje.



Slika 1.7 Prikaz gradijentnog vektora na dohvatnoj krivulji [9]

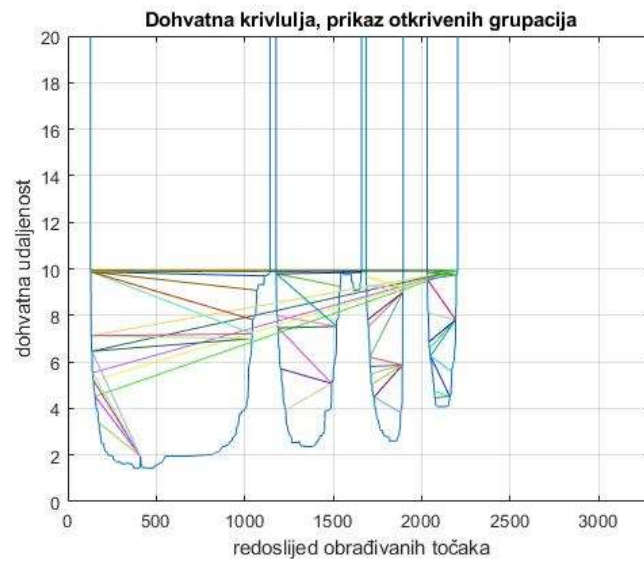
$$ind. infleksije = \frac{-w^2 + (x.R - y.R) * (z.R - y.R)}{||\vec{g}(x,y)|| * ||\vec{g}(y,z)||} \quad (1)$$

$$gradijent = w * (y.R - x.R) - w * (z.R - y.R) \quad (2)$$

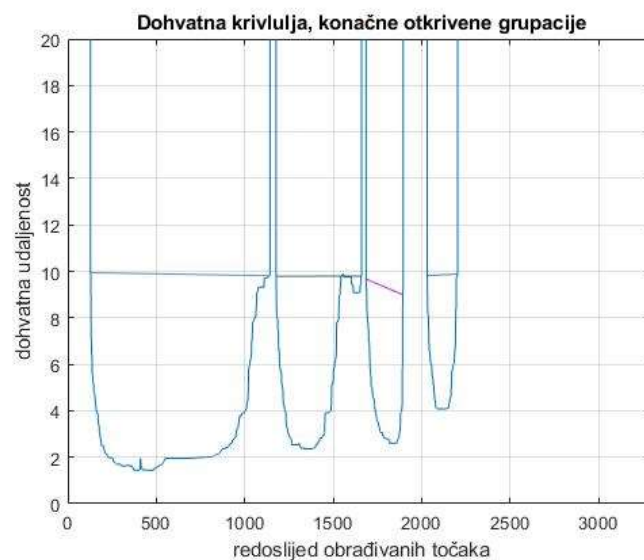
Prolaskom kroz sve točke dohvatne krivulje algoritam određuje početne i krajnje točke potencijalnih grupa, a primjer rezultata može se vidjeti na Slici 1.8. Svaka prepoznata grupa prikazana je crtom koja povezuje početak i kraj grupe. Vidljivo je da je broj grupa prevelik te su neke grupe detektirane više puta. U nastavku algoritma uklanja se višak grupa. Provođi se „spajanje“ grupa na način da se od više detektiranih sličnih grupa zadržavaju samo one najveće te se izbacuju grupe koje su detektirane, a sadrže točku definiranu kao šum. Konačni rezultat GC algoritma prikazan je na Slici 1.9.

Zbog male ovisnosti rezultata GC algoritma o parametrima  $t$  i  $w$ , oni su tijekom implementacije određeni na fiksne vrijednosti ( $w = 0.5$ ,  $t = 160$ ). Glavna razlika za drugačije kombinacije parametara  $t$  i  $w$  je u broju otkrivenih grupa na dohvatnoj krivulji, ali je konačni rezultat nakon uklanjanja viška grupa isti.

Indeksi originalnih podataka mogu se iščitati iz liste obrađivanih točaka. Drugim riječima svi indeksi spremljeni na listi obrađivanih točaka od pozicije početka grupe do pozicije kraja grupe tvore grupu. Tako se za svaku točku iz početnog skupa podataka može odrediti indeks grupe.

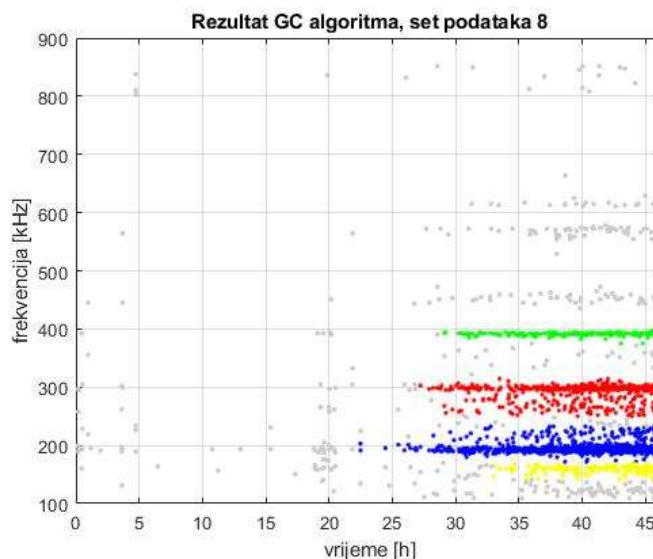


Slika 1.8 Prikaz svih otkrivenih grupa GC algoritmom na dohvatnoj krivulji



Slika 1.9 Prikaz konačnih grupa otkrivenih GC algoritmom na dohvatnoj krivulji

Konačni rezultat ekstrakcije grupa GC algoritmom prikazan je na Slici 1.10 gdje su različitim bojama prikazane otkrivene grupe, dok su sivom bojom prikazane nesvrstane točke, odnosno šum.



Slika 1.10 Prikaz rezultata GC algoritma

## 1.4. Spajanje prozora

Skup algoritama će u konačnoj implementaciji na ugradbenom sustavu raditi u stvarnom vremenu, obrađujući podatke blok po blok. Blok podataka od konačnog broja točaka zove se prozorom. OPTICS algoritam i GC su dovoljni za grupiranje jednog prozora podataka, ali za spajanje rezultata obrade dvaju susjednih prozora potreban je algoritam za spajanje prozora.

Kako bi algoritam za spajanje prozora pravilno funkcionirao, određen broj točaka na kraju prethodnog prozora mora se ponovno obraditi u trenutnom prozoru, odnosno prozori se u određenom broju točaka moraju preklapati.

Postupak provjere radi li se o istoj grupi u dva susjedna prozora započinje usporedbom rezultata dobivenih u oba prozora na točkama preklapanja. Algoritam prolazi kroz svaku točku preklapanja i uspoređuje indekse grupa dobivenih u prethodnom i trenutnom prozoru te rezultat sprema u matricu sličnosti. Na primjer, ako je točka u prethodnom prozoru pripadala grupi 1 (indeks grupe je 1), a u drugom prozoru pripada grupi 2, broj u matrici sličnosti na mjestu (1,2) povećat će se za 1. Dakle matrica sličnosti je  $N \times M$  matrica gdje  $N$  (broj redaka) predstavlja grupe prepoznate u prethodnom prozoru, a  $M$  (broj stupaca) grupe prepoznate u trenutnom prozoru.

Primjer matrice sličnosti za jednostavan slučaj gdje se 3 grupe iz prethodnog prozora nastavljaju u 3 grupe u sljedećem prozoru prikazan je na Slici 1.11.

|   |   |   |
|---|---|---|
| X | 0 | 0 |
| 0 | Y | 0 |
| 0 | 0 | Z |

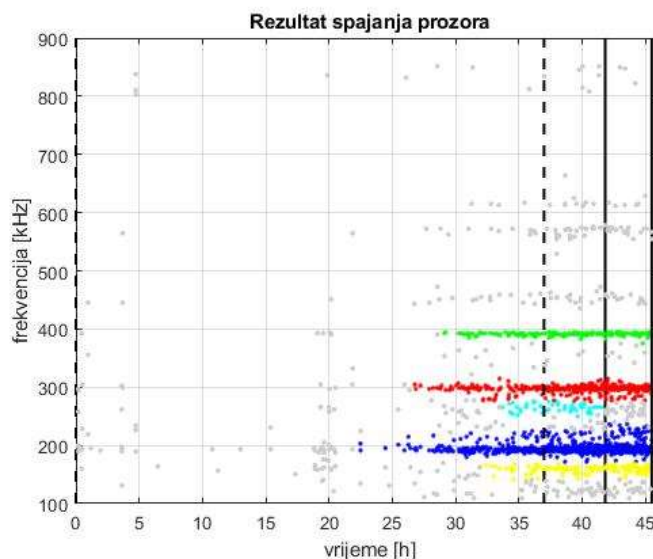
Slika 1.11 Primjer matrice sličnosti

X, Y i Z predstavljaju brojeve točaka koje grupe dijele dok su na ostalim poljima nule. Svaki stupac u matrici predstavlja grupu iz trenutnog prozora. Kako bi algoritam zaključio nastavlja li se grupa iz prethodnog prozora u trenutni, traži maksimum u stupcu promatrane grupe. Indeks retka u kojem se nalazi maksimum, indeks je grupe iz prethodnog prozora na koju se promatrana grupa nastavlja. Primjerice, za matricu sa Slike 1.11, maksimum X za prvi stupac nalazi se u prvom retku što znači da se grupa iz trenutnog prozora označena brojem 1 nastavlja na grupu iz prethodnog prozora također označenom brojem 1. Naravno, broj redaka i stupaca matrice sličnosti ne mora biti isti, kao u slučaju da se pojavi nova grupa. Tada ta grupa u odgovarajućem stupcu ima sve elemente 0. Također, ako je grupa iz prethodnog prozora u trenutnom prozoru neotkrivena, u odgovarajućem retku imat će sve elemente 0.

Broj ostalih elemenata u matrici (uz maksimume) ne mora nužno biti nula (u slučaju da su pojedine točke različito grupirane u dva prozora). Tada algoritam provjerava iznos maksimuma stupca u odnosu na ostale elemente stupca i retka kojem pripada.

Ova provjera nužna je u slučaju da se više grupa u sljedećem prozoru spaja u jednu ili u slučaju da se jedna grupa u sljedećem prozoru razdvaja na više njih. Ovakav slučaj događa se ako su rezultati OPTICS i GC algoritama „netočni“, odnosno ako su drugačiji od onog što bismo očekivali. Na primjer, očekivana je jedna veća grupa, no zbog drugačijeg rasporeda točaka u drugom prozoru, algoritam detektira dvije manje. Ako se dogodi ijedna od dviju navedenih situacija, algoritam novonastale grupe indeksira novim indeksom (smatra ih novim grupama). Primjer rezultata obrade u prozorima prikazan je na Slici 1.12 gdje pune vertikalne crte predstavljaju krajeve prozora, a isprekidane vertikalne crte početke prozora. Podaci su obrađeni u dva prozora veličine 2000 točaka s preklapanjem od 1000 točaka. Sivom bojom su označene točke prepoznate kao šum dok obojane točke predstavljaju različite grupe.





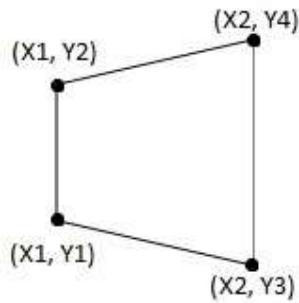
Slika 1.12 Prikaz rezultata algoritma spajanja prozora za dva prozora. Punom crtom prikazani krajevi prozora, a isprekidanom crtom počeci prozora

Korišteni pristup je jednostavan i brz, ali ima dva nedostatka. Prvi problem je što se dio točaka mora ponovno obraditi OPTICS i GC algoritmima što podrazumijeva povećanje potrošnje i dulje trajanje obrade. Zato je zanimljivo minimizirati preklapanje, međutim, premalim brojem točaka preklapanja riskiraju se pogreške pri povezivanju prozora. Drugi problem je što algoritam nije u stanju detektirati spajanje ili fragmentaciju grupa između susjednih prozora.

## 1.5. Sažimanje rezultata

Iz razloga navedenih u uvodu, vezanih za rad ugradbenog sustava, potrebno je minimizirati količinu informacija za prijenos. Zadatak algoritma za sažimanje rezultata je prikaz grupa dobivenih prethodno opisanim algoritmima pomoću trapezoidnih oblika. Svaki trapezoid određen je pomoću 4 točke (Slika 1.13). Koordinate ovih točaka određuju se ovisno o poziciji grupe unutar trenutno obrađivanog prozora, a postoji četiri moguća slučaja:

1. Grupa je nova, nastala u ovom prozoru,
2. Grupa se nastavlja iz prethodnog prozora,
3. Grupa završava u ovom prozoru,
4. Grupa se nastavlja u sljedeći prozor.

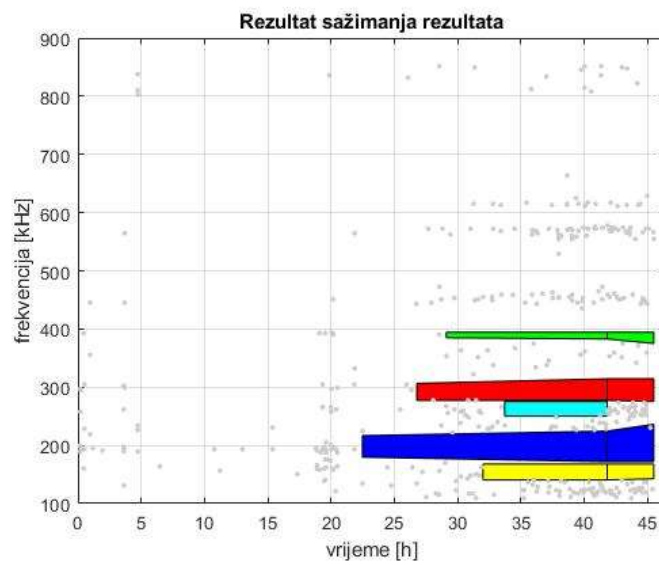


Slika 1.13 Primjer trapezoida određenog pomoću 4 točke

Za svaki spomenuti slučaj koordinate x-osi bočnih točaka trapezoida su iste. Drugim riječima bočne stranice trapezoida uvijek su vertikalne, paralelne s y-osi.

- U slučaju da je grupa novonastala, početak trapezoida u vremenu, koordinata  $X1$ , određena je pojavom prve točke te grupe. Koordinate  $Y1$  i  $Y2$  određene su minimalnim i maksimalnim frekvencijama u prvih  $N_{\min}$  točaka te grupe.
- U slučaju da se grupa nastavlja iz prethodnog prozora, početne točke  $(X1, Y1)$ ,  $(X1, Y2)$  određene su krajnjim točkama grupe na koju se nastavljaju. Drugim riječima, ako se grupe nastavljaju kroz više prozora, trapezoidi se također kroz prozore nastavljaju jedan na drugoga te tvore jedan oblik.
- Ako grupa završava u ovom prozoru, koordinata  $X2$  određena je vremenom pojave zadnje točke te grupe. Koordinate  $Y3$  i  $Y4$  određene su minimalnim i maksimalnim frekvencijama u zadnjih  $N_{\min}$  točaka te grupe.
- Ako se grupa nastavlja u sljedeći prozor, koordinata  $X2$  određena je vremenom pojave zadnje točke grupe u trenutno obrađivanom prozoru, a koordinate  $Y3$  i  $Y4$  određuju se kao u prethodnom slučaju.

Primjer konačnog izgleda rezultata skupa algoritama za set podataka 8 prikazan je na Slici 1.14. Trapezoidi različitih boja predstavljaju grupe, a sive točke su i dalje negrupirani šum.



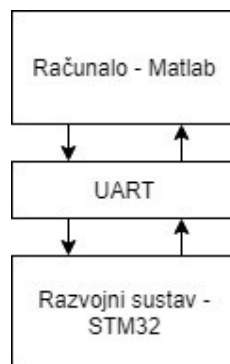
Slika 1.14 Primjer rezultata algoritma za sažimanje rezultata

## 2. Ispitivanje skupa algoritama nad sintetičkim podacima

Svrha ispitivanja skupa algoritama nad sintetičkim podacima je ispitati i potvrditi pretpostavke i pravilnosti vezane uz korišteni skup algoritama. U nastavku poglavlja opisano je pet eksperimenata provedenih nad sintetičkim podacima. Za sve sintetičke setove podataka iz ovih pet eksperimenata, ispituje se točnost prema metrici opisanoj u poglavlju 2.1. Vrijednosti dobivene ispitivanjem točnosti pokazatelji su kvalitete rada skupa algoritama te služe za potvrdu spomenutih pretpostavki.

Dva su glavna razloga korištenja sintetičkih podataka. Prvi razlog je taj što je broj dostupnih eksperimenata stvarnih mjerenja UAE relativno malen (sveukupno 11 setova podataka). Drugi razlog je taj što bi ispitivanje točnosti nad podacima za koje ne postoji točan rezultat grupiranja bilo sistematski pogrešno. Rezultati grupiranja algoritma bi se u tom slučaju uspoređivali s ručno određenim grupama. Kako bi se izbjegao ovaj problem i uklonili utjecaji pojedinih karakteristika setova podataka na rezultate, kreirano je nekoliko desetaka sintetičkih setova podataka. Svaki set podataka prilagođen je određenom eksperimentu, o čemu će biti riječi u nastavku.

Sustav korišten pri ispitivanju prikazan je na Slici 2.1. Skup algoritama za obradu podataka je implementiran na razvojnom sustavu dok je programski paket Matlab korišten za analizu rezultata. UART veza korištena je za prijenos podataka i parametara s računala na ugradbeni sustav te za prijenos rezultata s razvojnog sustava na računalo.



Slika 2.1 Prikaz sustava za ispitivanje skupa algoritama

## 2.1. Metrika za klasifikaciju

Problem analize točnosti grupiranja se u ovom radu promatrao kao problem višeklasne klasifikacije. Svaka se grupa u setu podataka smatra jednom klasom. Za svaku su klasu tijekom ispitivanja računati parametri ispravno pozitivnih (*True Positive*, TP), ispravno negativnih (*True Negative*, TN), pogrešno pozitivnih (*False Positive*, FP) i pogrešno negativnih (*False Negative*, FN) uzoraka [14], usporedbom rezultata grupiranja algoritma s anotacijom točaka dobivenom kreiranjem seta podataka. Definicije mjera su sljedeće:

- TP – Točka je točno grupirana (pridružena promatranoj klasi),
- TN – Točka nije grupirana, a nije ni trebala biti (pripada drugoj klasi ili šumu),
- FP – Točka je pogrešno grupirana (pridružena promatranoj klasi, a nije trebala biti),
- FN – Točka nije grupirana, a trebala bi biti (pripada promatranoj klasi, ali je svrstana u drugu klasu ili šum).

Iz iznosa TP, TN, FP i FN točaka, za svaku klasu računati su odziv (*recall*), preciznost (*precision*), točnost (*accuracy*) i F1 mjera (*F1-score*) prema izrazima (3, 4, 5, 6) [14].

$$recall = \frac{TP}{TP + FN} \quad (3)$$

$$precision = \frac{TP}{TP + FP} \quad (4)$$

$$accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (5)$$

$$F1 - score = \frac{precision * recall}{precision + recall} \quad (6)$$

Za cijeli set podataka računati su *Macro-recall*, *Macro-precision*, *Macro-accuracy* i *Macro-F1-score* kao usrednjene vrijednosti *recall*, *precision*, *accuracy* i *F1-score* rezultata svih klasa. U prikazu rezultata svih eksperimenata bit će korištene izračunate *Macro* vrijednosti.

## 2.2. Sintetički setovi podataka

Zajedničke vrijednosti za sve sintetičke setove podataka su veličina seta, raspon frekvencija i trajanje. Veličina sintetičkih setova podataka u svim eksperimentima iznosi 2000 točaka, od kojih je 200 točaka šum. Raspon frekvencija setova je od 0 do 400 kHz, a trajanje 100 h. U većini eksperimenata, osim drugog, rezultati su dobiveni obradom podataka u jednom prozoru od 2000 točaka i bez preklapanja, kako bi se uklonio njihov utjecaj na rezultate. Također, u svim eksperimentima, osim petog, grupe su orijentirane horizontalno, paralelno s vremenskom osi. Generalni pregled setova podataka za sve eksperimente prikazan je Tablicom 2.1, a detalji će biti navedeni u potpoglavlju svakog eksperimenta.

Tablica 2.1 Pregled karakteristika sintetičkih setova podataka korištenih u eksperimentima

| Eksperiment  | Broj setova | Broj grupa po setu       |
|--------------|-------------|--------------------------|
| 1, prvi dio  | 10          | 1                        |
| 1, drugi dio | 7           | 2                        |
| 2            | 16          | 4 seta *(3 do 6 grupa)   |
| 3, prvi dio  | 50          | 2                        |
| 3 drugi dio  | 50          | 4                        |
| 4            | 40          | 5 setova *(1 do 8 grupa) |
| 5            | 3           | 4                        |

## 2.3. Ovisnost točnosti o gustoći grupe, eksperiment 1

Iako sama po sebi nije parametar OPTICS algoritma, iz parametara  $N_{\min}$  i  $\varepsilon$  može se izračunati takozvana zadana gustoća prema izrazu (7). Nazivnik u izrazu (7) predstavlja površinu na kojoj OPTICS algoritam provjerava uvjet minimalnog broja susjeda,  $N_{\min}$  (poglavlje 1.2). Gustoća podataka, pa tako i grupe, definira se kao ukupni broj točaka na nekoj površini. Pretpostavka koja se ispituje ovim eksperimentom je da rezultat grupiranja ovisi o zadanoj gustoći (gustoća izračunata iz parametara). Ukoliko je grupa koju želimo otkriti algoritmom manje gustoće od zadane, ostat će neotkrivena (uvjet minimalne gustoće).

$$\text{zadana gustoća} = \frac{N_{\min}}{\varepsilon^2 * \pi} \quad [\text{točaka}/(\text{kHz} * \text{h})] \quad (7)$$

### 2.3.1. Eksperiment 1, prvi dio

U prvom dijelu eksperimenta algoritam je ispitan na 10 setova sintetičkih podataka. Svi setovi sadržavali su po jednu grupu pravokutnog oblika i točke šuma, a primjeri seta podataka prikazani su na Slici 2.2. Gustoće grupa za sve setove podataka navedene su u Tablici 2.2.

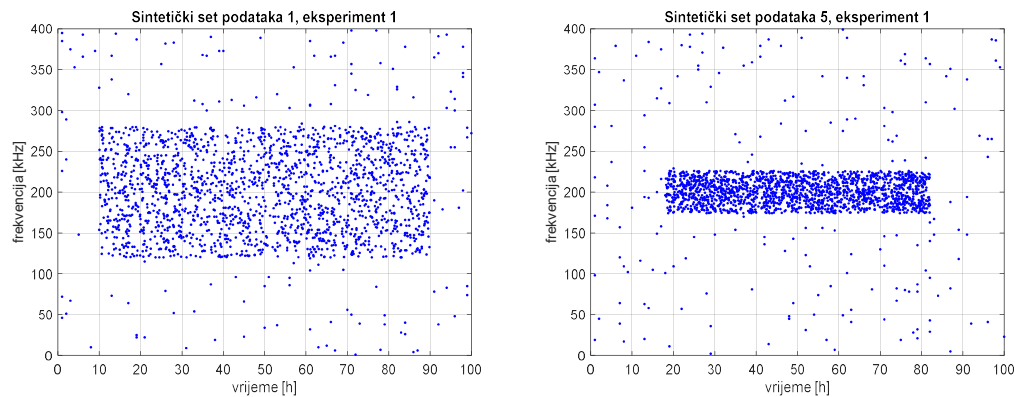
Tablica 2.2 Gustoće grupa za setove podataka, prvi dio eksperimenta 1

| Set     | 1     | 2     | 3     | 4     | 5     | 6     | 7     | 8     | 9     | 10    |
|---------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| Gustoća | 0,145 | 0,255 | 0,348 | 0,438 | 0,545 | 0,647 | 0,755 | 0,861 | 0,954 | 0,064 |

Variranje zadane gustoće izvedeno je variranjem parametara  $\varepsilon$  i  $N_{\min}$  u sljedećim rasponima:

- $N_{\min} = [15, 30, 45, 60, 75, 90, 105, 120, 135, 150]$ ,
- $\varepsilon = [5, 7, 9, 11, 13, 15, 17, 19, 21, 23]$ .

Svaka kombinacija parametara  $N_{\min}$  i  $\varepsilon$  predstavlja drugačiju zadanu gustoću. Cilj prvog dijela eksperimenta 1 je pokazati kako kvaliteta grupiranja ovisi o zadanoj gustoći.



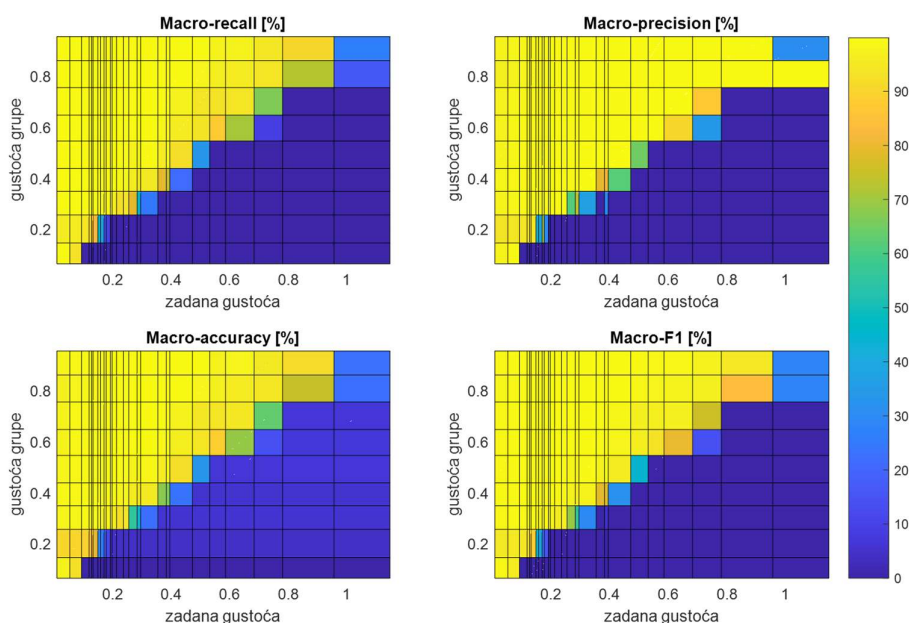
Slika 2.2 Primjeri sintetičkih setova podataka za eksperiment 1, prvi dio

Rezultat prvog dijela eksperimenta može se vidjeti na Slici 2.3. Može se uočiti, a vrijedit će i za rezultate svih ostalih eksperimenata, da *Macro-recall* i *Macro-precision* „prate“ jedan drugoga, odnosno nemaju recipročan odnos kakav bi imali kod tipične binarne klasifikacije. Premda naizgled nelogično, to je očekivani rezultat. Naime, za kvalitetno odrađeno grupiranje, za sve grupe se očekuje mali broj FP i FN točaka i velik broj TP točaka, što se preslikava u visoku vrijednost za *recall* i *precision*. U slučaju loše određenog grupiranja očekuje se mali broj TP točaka za sve grupe, što se preslikava u niske vrijednosti

za *recall* i *precision*. Najveća razlika u iznosima za *recall* i *precision*, kod rezultata svih eksperimenata, može se primijetiti u prijelaznim područjima, gdje su neke od grupa otkrivene djelomično ili gdje su pojedine grupe otkrivene, a pojedine ne.

U nastavku teksta će se kod referenciranja na kvalitetu grupiranja promatrati vrijednost *Macro-F1-score*, s obzirom da se *F1-score* računa direktno iz iznosa *recall* i *precision* i predstavlja sveobuhvatnu mjeru točnosti grupiranja.

Iz rezultata na Slici 2.3 vidljivo je kako prelaskom zadane gustoće preko iznosa gustoće grupe, kvaliteta grupiranja pada, odnosno grupa ostaje neprepoznata.



Slika 2.3 Prikaz ovisnosti točnosti o zadanoj gustoći

### 2.3.2. Eksperiment 1, drugi dio

U drugom dijelu eksperimenta skup algoritama je ispitan na 7 sintetičkih setova podataka od kojih je svaki set imao dvije pravokutne grupe različitih gustoća. Prva grupa, manje gustoće, u svim je setovima imala fiksnu gustoću, dok je gustoća druge grupe, u svakom setu, veća od gustoće manje za drugačiji omjer (Tablica 2.3).



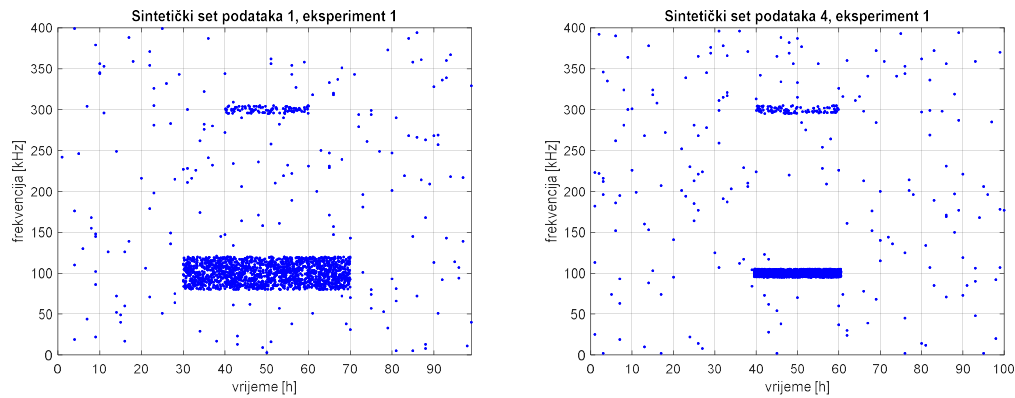
Tablica 2.3 Gustoće grupa za setove podataka, drugi dio eksperimenta 1

| Set                        | 1     | 2     | 3     | 4     | 5      | 6      | 7      |
|----------------------------|-------|-------|-------|-------|--------|--------|--------|
| Gustoća manje grupe        | 0,500 | 0,500 | 0,500 | 0,500 | 0,500  | 0,500  | 0,500  |
| Gustoće veće grupe         | 1,067 | 1,936 | 4,055 | 8,100 | 16,195 | 32,381 | 65,385 |
| Omjer gustoća (veći/manji) | 2,11  | 3,80  | 7,95  | 16,20 | 32,38  | 63,49  | 130,77 |

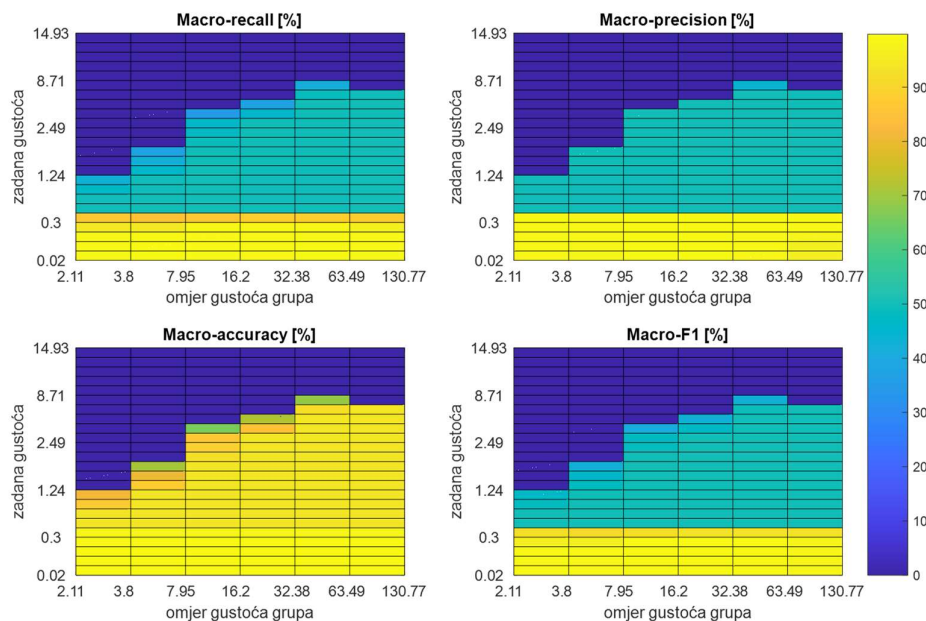
Parametar  $\varepsilon$  imao je fiksnu vrijednost ( $\varepsilon = 8$ ), a variranje zadane gustoće izvedeno je variranjem parametra  $N_{\min}$  u sljedećem rasponu:

- $N_{\min} = [5, 10, 20, 40, 60, 80, 100, 150, 200, 250, 300, 350, 400, 450, 500, 750, 1000, 1250, 1500, 1750, 2000, 2250, 2500, 2750, 3000]$ .

Primjeri setova podataka prikazani su na Slici 2.4. Cilj drugog dijela eksperimenta 1 je pokazati kako kvaliteta grupiranja ovisi o zadanoj gustoći i ima li razlika u gustoći grupa utjecaja na rezultate (poglavlje 1.1, zahtjev 4).



Slika 2.4 Primjeri sintetičkih setova podataka za eksperiment 1, drugi dio



Slika 2.5 Prikaz ovisnosti točnosti o zadanoj gustoći s grupama različitih gustoća

Rezultati sa Slike 2.5 su očekivani. Algoritam je u stanju prepoznati obje grupe kada je uvjet minimalne gustoće za obje grupe zadovoljen. U trenutku kada zadana gustoća preraste gustoću rjeđe grupe, algoritam otkriva samo jednu, gušću grupu, sve dok zadana gustoća ne preraste i njenu gustoću. Zanimljivo je primijetiti da algoritam za omjer gustoća 63,49 i 130,77 ne prepoznaje gušću grupu makar je uvjet minimalne gustoće zadovoljen (zadana gustoća 14,93). Razlog tome je taj što je pri proračunu zadane gustoće korišten  $\varepsilon = 8$  i  $N_{\min} = 3000$ . Drugim riječima, iako je zadana gustoća dovoljno mala i zadovoljava uvjet minimalne gustoće, parametar  $N_{\min}$  također mora biti manji od ukupnog broja točaka u grupi. S obzirom da gušća grupa ima manji broj točaka od  $N_{\min}$ , ostaje neprepoznata.

Oba dijela eksperimenta 1 potvrđuju pretpostavku s početka poglavlja da grupiranje neće biti uspješno odrađeno ukoliko nije zadovoljen uvjet minimalne gustoće.

## 2.4. Ovisnost točnosti o odnosu veličine prozora i preklapanju, eksperiment 2

Eksperimentom opisanom u ovom poglavlju ispituje se utjecaj veličine prozora i preklapanja na uspješnost grupiranja. U ispitivanju je korišteno 16 setova podataka koji se razlikuju po broju, obliku, gustoći, veličini i položaju grupa (Tablica 2.4). Primjeri setova prikazani su na Slici 2.6. Kod ispitivanja algoritma varirane su veličine prozora i preklapanja

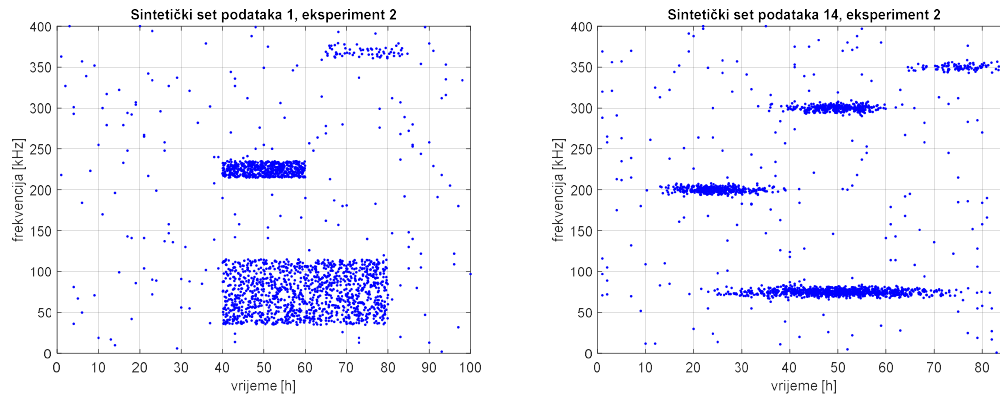
dok su parametri  $N_{\min}$  ( $N_{\min} = 25$ ) i  $\varepsilon$  ( $\varepsilon = 10$ ) bili fiksni, vodeći računa o tome da uvjet minimalne gustoće bude zadovoljen.

Tablica 2.4 Podaci o sintetičkim setovima podataka, eksperiment 2

| Setovi      | 1, 5        | 2, 6 | 3, 7 | 4, 8 | 9, 13  | 10, 14 | 11, 15 | 12, 16 |
|-------------|-------------|------|------|------|--------|--------|--------|--------|
| Broj grupa  | 3           | 4    | 5    | 6    | 3      | 4      | 5      | 6      |
| Oblik grupa | pravokutnik |      |      |      | elipsa |        |        |        |

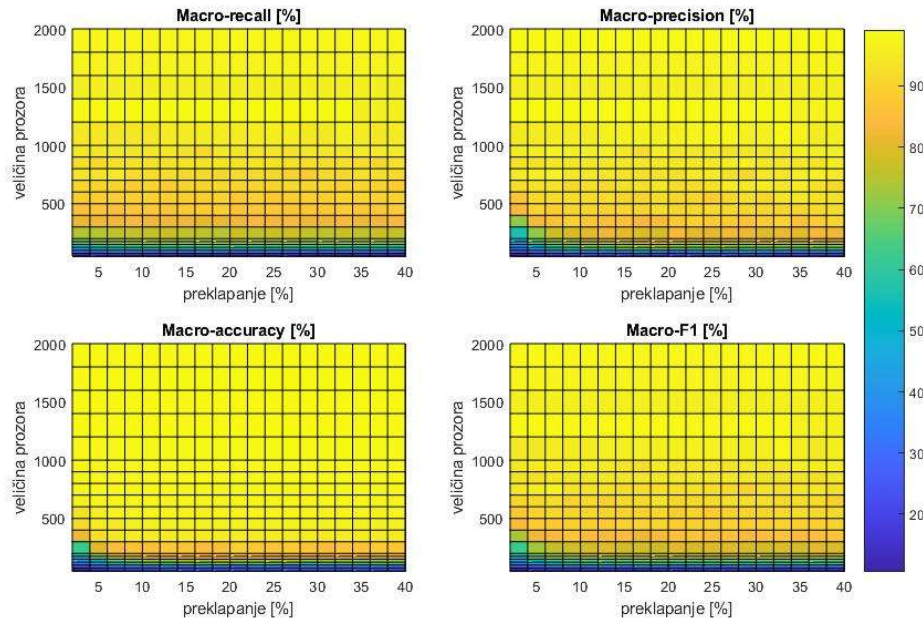
Parametri veličine prozora i preklapanja varirani su u sljedećem rasponu:

- vel. prozora = [50, 75, 100, 125, 150, 175, 200, 300, 400, 500, 600, 700, 800, 900, 1000, 1200, 1400, 1600, 1800, 2000],
- preklapanje = [2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40] [%].



Slika 2.6 Primjeri sintetičkih setova podataka za eksperiment 2

Rezultati eksperimenta prikazani su na Slici 2.7. Iz rezultata je vidljivo da je utjecaj veličine prozora zanemariv za veće prozore. Smanjenjem veličine prozora na oko 300 točaka, točnost počinje padati, a minimalna veličina prozora bila bi oko 100 točaka. Zanimljivo je i da preklapanje prozora ima mali utjecaj na točnost za veće veličine prozora. Zaključak je da se kod konačne implementacije algoritma veličina prozora može odabrati proizvoljno dok se preklapanje može minimizirati. Treba imati na umu da ovisno o rasporedu ulaznih točaka može doći do pogreške u spajanju prozora.



Slika 2.7 Prikaz ovisnosti točnosti o veličini prozora i preklapanju

## 2.5. Ovisnost točnosti o udaljenosti grupa, eksperiment 3

Pitanja na koja se traži odgovor ovim eksperimentom je sljedeće - koja je minimalna apsolutna udaljenost između dviju grupa za koju skup algoritama prepoznaje svaku zasebno i kakav utjecaj na točnost grupiranja ima parametar  $\epsilon$ ?

OPTICS algoritam sve točke koje se nalaze unutar udaljenosti  $\epsilon$  neke jezgrene točke grupira u istu grupu. Ako se dvije grupe nalaze na udaljenosti manjoj od  $\epsilon$ , OPTICS algoritam i GC će ih grupirati u jednu grupu. Stoga je pretpostavka da minimalna udaljenost između dviju grupa mora iznositi minimalno vrijednost  $\epsilon$ .

Eksperiment 3 proveden je u dva dijela gdje se ispitala ista pretpostavka, ali s razlikom u broju grupa u setovima podataka.

### 2.5.1. Eksperiment 3, prvi dio

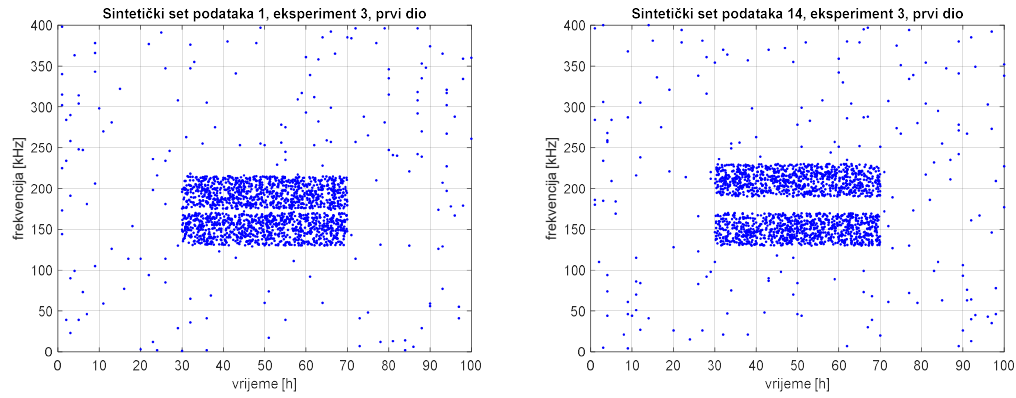
Prvi dio eksperimenta proveden je na 50 setova sintetičkih podataka koji su sadržavali dvije kvadratne grupe istog oblika i veličine (900 točaka), ali na 10 različitih međusobnih udaljenosti (po 5 setova podataka s grupama na istoj udaljenosti) (Tablica 2.5). Gustoća grupa u svim setovima je ista, 0,5625. Primjer setova podataka dan je na Slici 2.8.

$N_{\min}$  je odabran fiksno uzimajući u obzir da uvijek bude zadovoljen uvjet minimalne gustoće, a  $\varepsilon$  je variran u sljedećem rasponu:

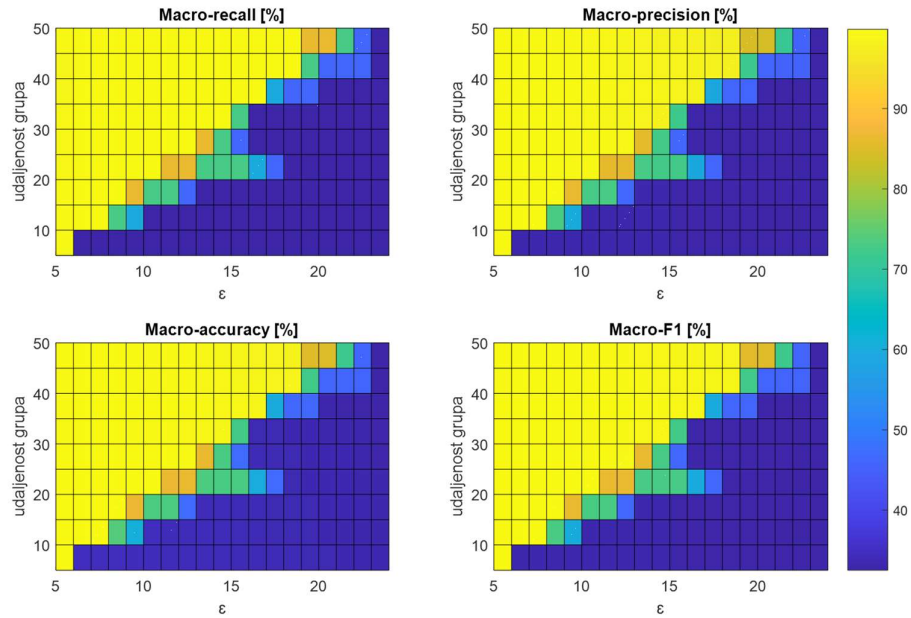
- $\varepsilon = [5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25]$ .

Tablica 2.5 Sintetički setovi podataka razvrstani prema udaljenosti grupa, eksperiment 3, prvi dio

|            |               |               |               |               |                |
|------------|---------------|---------------|---------------|---------------|----------------|
| Setovi     | 1,11,21,31,41 | 2,12,22,32,42 | 3,13,23,33,43 | 4,14,24,34,44 | 5,15,25,35,45  |
| Udaljenost | 5             | 10            | 15            | 20            | 25             |
| Setovi     | 6,16,26,36,46 | 7,17,27,37,47 | 8,18,28,38,48 | 9,19,29,39,49 | 10,20,30,40,50 |
| Udaljenost | 30            | 35            | 40            | 45            | 50             |



Slika 2.8 Primjeri sintetičkih setova podataka za eksperiment 3

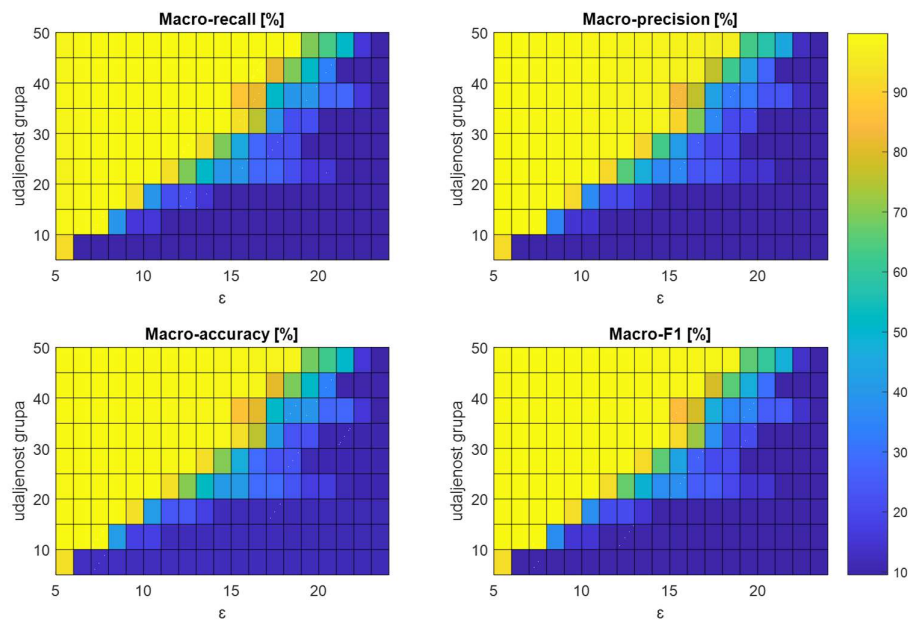


Slika 2.9 Prikaz ovisnosti točnosti o udaljenosti između dviju grupa

Iz rezultata na Slici 2.9 se može primijetiti kako je početna pretpostavka potvrđena - udaljenost između dviju grupa za kvalitetno grupiranje mora iznositi minimalno  $\varepsilon$  (uvjet minimalne udaljenosti grupa). Također je vidljivo da nerijetko ova udaljenost mora biti i veća od  $\varepsilon$ , negdje i do dva puta. Razlog je što se između grupa u svim setovima podataka nalazi šum, koji ovisno o gustoći i broju točaka može utjecati na rezultate. Ako se primjerice točka šuma nalazi između dviju grupa i zadovoljava uvjete za proglašenje jezgrenom točkom, može doći do spajanja dviju grupa.

### 2.5.2. Eksperiment 3, drugi dio

Drugi dio eksperimenta strukturiran je kao i prvi, s razlikom da su u setovima podataka po 4 grupe na istoj međusobnoj udaljenosti. Gustoća grupa je u svim setovima jednaka i iznosi 0,5625. Raspored setova po udaljenosti između grupa izgleda kao na Tablici 2.5. Rezultati su prikazani na Slici 2.10, a potvrđuju zaključke iz prošlog odlomka.



Slika 2.10 Prikaz ovisnosti točnosti o udaljenosti između četiriju grupa

## 2.6. Ovisnost točnosti o broju grupa, eksperiment 4

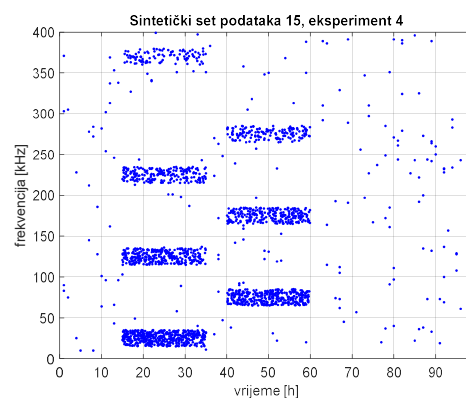
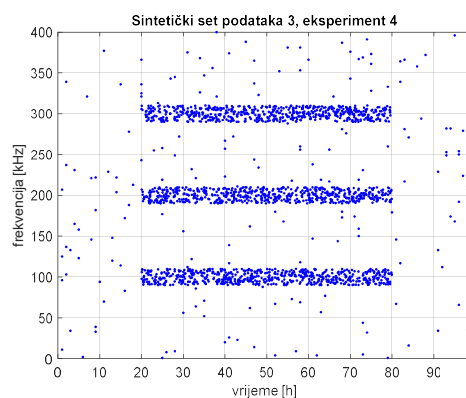
U poglavlju 2.1 objašnjeni su razlozi odabira OPTICS algoritma za grupiranje. Jedan od razloga je mogućnost grupiranja bez prethodnog znanja o broju grupa. Stoga je pretpostavka da točnost grupiranja ne bi trebala ovisiti o broju grupa.

Eksperimentom se točnost grupiranja ispituje na 40 setova podataka s različitim brojem grupa (Tablica 2.6). Primjer setova podataka za eksperiment 4 prikazan je na Slici 2.11. Kod ispitivanja  $N_{\min}$  i  $\epsilon$  odabrani su tako da zadovolje uvjet minimalne gustoće za sve grupe i uvjet minimalne udaljenosti između grupa ( $N_{\min} = 25$ ,  $\epsilon = 10$ ).

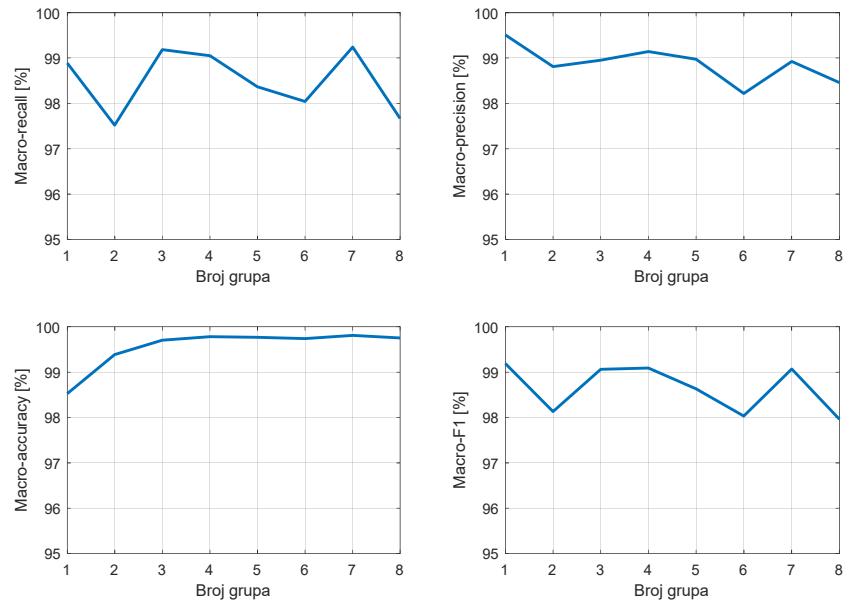
Rezultati su prikazani na Slici 2.12. Premda *Macro - F1-score* lagano varira mijenjanjem broja grupa, promjene su zanemarive i može se potvrditi da kvaliteta grupiranja ne ovisi o broju grupa unutar seta podataka.

Tablica 2.6 Podaci o sintetičkim setovima podataka, eksperiment 4

|           |             |         |         |         |         |         |         |         |       |    |
|-----------|-------------|---------|---------|---------|---------|---------|---------|---------|-------|----|
| Setovi    | 1,9,17      | 2,10,18 | 3,11,19 | 4,12,20 | 5,13,21 | 6,14,22 | 7,15,23 | 8,16,24 |       |    |
| Br. grupa | 1           | 2       | 3       | 4       | 5       | 6       | 7       | 8       |       |    |
| Oblik     | pravokutnik |         |         |         |         |         |         |         |       |    |
| Setovi    | 25,33       | 26,34   | 27,35   | 28,36   | 29,37   | 30,38   | 31      | 40      | 32    | 39 |
| Br. grupa | 1           | 2       | 3       | 4       | 5       | 6       | 7       | 8       | 8     | 7  |
| Oblik     | elipsa      |         |         |         |         |         |         |         | prav. |    |



Slika 2.11 Primjeri sintetičkih setova podataka za eksperiment 4



Slika 2.12 Prikaz ovisnosti točnosti o broju grupa

## 2.7. Ovisnost točnosti o obliku i orijentaciji grupa, eksperiment 5

Ovim eksperimentom želi se pokazati da uspješnost grupiranja OPTICS algoritma ne ovisi o obliku i orijentaciji grupe. Eksperiment se sastoji od tri seta podataka gdje svaki ima 4 grupe različitih oblika i orijentacija (Tablica 2.7).

Tablica 2.7 Sintetički setovi podataka, eksperiment 5

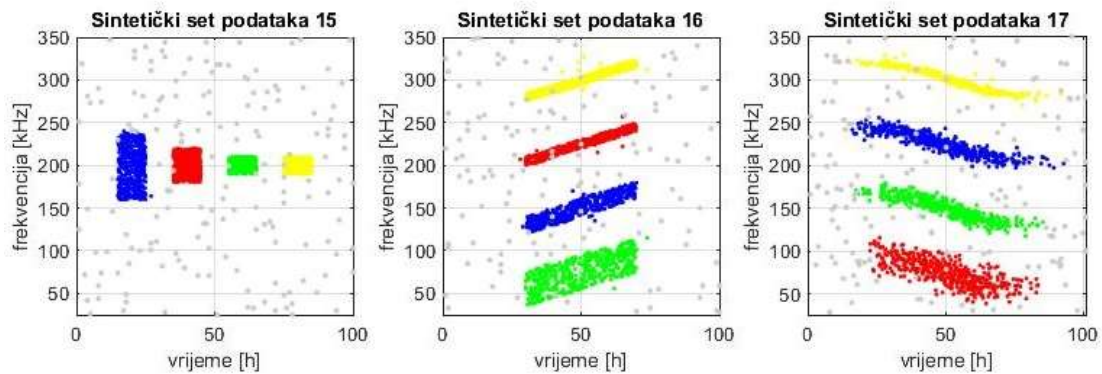
| Set          | 15          | 16           | 17                 |
|--------------|-------------|--------------|--------------------|
| Oblik        | pravokutnik | paralelogram | zakrivljena elipsa |
| Orijentacija | vertikalna  | horizontalna | horizontalna       |
| Nagib        | 0°          | 45°          | - 45°              |

Svaki set obrađen je jednom s optimalnim odabirom parametara, a oni su sljedeći:

- Set 15,  $N_{\min} = 25$ ,  $\varepsilon = 5$
- Set 16,  $N_{\min} = 25$ ,  $\varepsilon = 10$
- Set 17,  $N_{\min} = 25$ ,  $\varepsilon = 10$

Rezultati za sva tri primjera prikazani su na Slici 2.13.

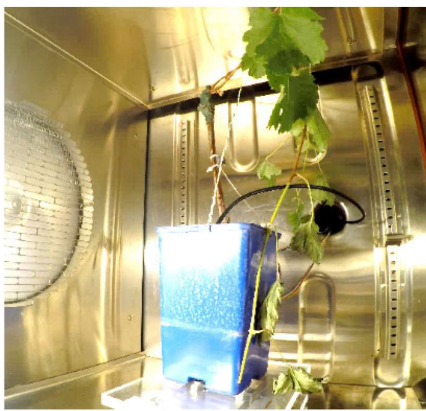




Slika 2.13 Prikaz grupiranja različitih setova podataka

### 3. Ispitivanje algoritma nad stvarnim podacima

Skupovi stvarnih podataka na kojima je testiran algoritam dobiveni su iz 8 različitih eksperimenata [3]. Neki od eksperimenata izvedeni su u nekontroliranim uvjetima gdje su biljke ostavljene da se postepeno suše prirodnim putem (eksperiment 6), dok su ostali eksperimenti izvedeni u klima komori (Slika 3.1). Klima komora je korištena zbog mogućnosti stroge kontrole uvjeta te smanjenja ukupnog trajanja eksperimenta. Tipični vremensko-frekvencijski prikaz zabilježenih UAE već je prikazan na Slici 1.2.



Slika 3.1 Postav eksperimenta u klima komori [2]

Setovi stvarnih podataka su različiti - različitog broja točaka, oblika grupa, vremena trajanja, i tako dalje. Obilježje zajedničko svim setovima podataka su, već spominjane, horizontalne frekvencijske linije (grupe) oko određenih frekvencijskih pojaseva. Kako će skup algoritama u konačnici biti implementiran na ugradbenom sustavu koji će obrađivati podatke bez ljudskog utjecaja, pitanje koje se samo po sebi nameće je kako odabrati parametre  $N_{\min}$ ,  $\epsilon$ , veličinu prozora i preklapanje za kvalitetno grupiranje neovisno o ulaznom setu podataka? Prema rezultatima ispitivanja algoritma u poglavlju 2.4, parametre veličine prozora i preklapanja možemo odabrati proizvoljno pa su tako za potrebe ispitivanja u ovome poglavlju korišteni veličina prozora od 1000 točaka i preklapanje od 20% (200 točaka). Premda je u poglavlju 2.3 pokazano da je za prepoznavanje grupe dovoljno parametrima zadovoljiti uvjet minimalne gustoće, ne može se parametrima zadati vrlo mala gustoća i očekivati dobre rezultate. Problem je što povećavanjem parametra  $\epsilon$  i smanjivanjem parametra  $N_{\min}$  (smanjenje zadane gustoće) grupiranje uključuje sve više i više točaka šuma, a može doći i do spajanja grupa.

Još jedna zajednička značajka svim setovima podataka (6, 7, 8, 9, 10, 12, 13, 14) je da u prvih nekoliko sati ili dana imaju određeni vremenski period prije početka samog grupiranja. Fizikalno značenje ove činjenice je da sušenje (grupiranje) obično ne započinje početkom mjerenja samih emisija. Drugim riječima, sustav za mjerenje trebao bi biti postavljen na biljku za vrijeme dok biljka nije pod povećanim vodnim stresom. Primjerice, u setu podataka 6, grupiranje započinje otprilike tek šesnaestog dana. Ova se činjenica može iskoristiti za izračun parametara iz samih ulaznih podataka.

U nastavku poglavlja opisana su tri različita pristupa odabiru parametara  $N_{\min}$  i  $\varepsilon$ :

1. Odabir fiksnih parametara,
2. Adaptivno određivanje parametara iz satnog broja emisija,
3. Adaptivno određivanje parametara iz kumulativnog broja emisija.

Važno je napomenuti da nijedna od adaptivnih metoda nije implementirana na ugradbenom sustavu, već su vrijednosti parametara dobivene u Matlabu, a ispitivanje nad podacima je odrađeno na razvojnom sustavu. U nastavku poglavlja rezultati će biti prikazivani bez sažimanja, kako bi se dobio bolji dojam o kvaliteti grupiranja. Također, zbog boljeg pregleda spomenutih prikaza, sve slike rezultata nalaze se na kraju poglavlja.

### 3.1. Fiksni parametri

Najjednostavniji pristup je odabrati  $N_{\min}$  i  $\varepsilon$  prema dosadašnjim saznanjima i koristiti ih kod grupiranja svih setova podataka. Na Slici 3.2 i Slici 3.3 prikazani su rezultati uz tri kombinacije parametara  $N_{\min}$  i  $\varepsilon$ :

1.  $\varepsilon = 10$ ,  $N_{\min} = 25$  (prosječan odabir),
2.  $\varepsilon = 10$ ,  $N_{\min} = 10$  (odabrani  $N_{\min}$  premali),
3.  $\varepsilon = 10$ ,  $N_{\min} = 100$  (odabrani  $N_{\min}$  prevelik).

Iz prikaza je vidljivo da prosječan odabir fiksnih parametara (kombinacija 1) više odgovara nekim setovima podataka od drugih, ali generalno govoreći rezultati su dobri. Kod nekoliko setova došlo je do spajanja grupa zbog nedovoljne udaljenosti, međutim, pitanje je koliko je na to utjecao odabir parametara, a koliko velika gustoća točaka između grupa. Kao relativno loš primjer odabira parametara dane su kombinacije 2 i 3. Vidljivo je da se premalim odabirom  $N_{\min}$  detektira sve više šuma i grupe se sve više spajaju, dok se prevelikim odabirom  $N_{\min}$  dio grupa ni ne prepozna.

### 3.2. Adaptivno određivanje parametara iz satnog broja emisija

Prva adaptivna metoda zasniva se na praćenju satnog broja emisija u vremenu bez pojave grupa. Za setove podataka 7, 8, 9 i 13 (koji predstavljaju kraća mjerenja) ovaj vremenski period je prva 24 h mjerenja, dok za setove 6, 12, 14 traje prvih 100 h. Period promatranja može se odabrati i drugačije, s obzirom na to koliko se brzo očekuje sušenje biljke. S obzirom da je eksperiment 6 jedini proveden izvan klima komore, vremenski period praćenja za buduće eksperimente može se prilagoditi prema njemu.

Parametar  $\varepsilon$  je pri proračunu odabran kao fiksna vrijednost 10, s obzirom da su sve grupe unutar promatranih setova udaljene barem za 10 kHz. Parametar  $N_{\min}$  se zatim izračuna preko izraza (8) gdje je  $N_{\text{avg}}$  prosječni broj emisija po satu u vremenskom periodu promatranja.

$$N_{\min} = 10 + N_{\text{avg}} * 3 \quad (8)$$

Pribrojnik 10 u izrazu služi za postavljanje apsolutnog minimuma, u slučaju da je srednji satni broj emisija malen. Faktor 3 odabran je proizvoljno, obzirom da za sve skupove podataka broj  $N_{\text{avg}}$  na promatranim vremenskim intervalima iznosi između 2 i 6. Navedenim izračunom će se tako za gušće setove podataka dobivati veći  $N_{\min}$ , a za rjeđe manji, ali će iznos i dalje ostajati u granicama očekivane normale. Teško će doći do izračuna pretjerano velikog ili pretjerano malog parametra  $N_{\min}$ . Rezultati za adaptivno određene parametre  $N_{\min}$  ovom metodom prikazani su na Slici 3.4 i Slici 3.5. Vertikalne isprekidane crte označavaju sat u kojem je parametar  $N_{\min}$  izračunat.

### 3.3. Adaptivno određivanje parametara iz gustoće podataka

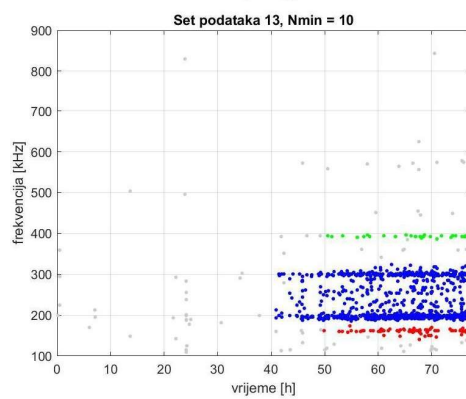
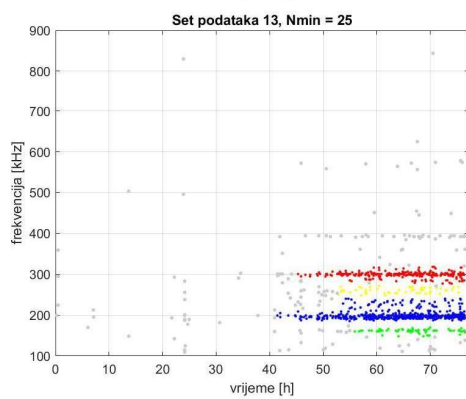
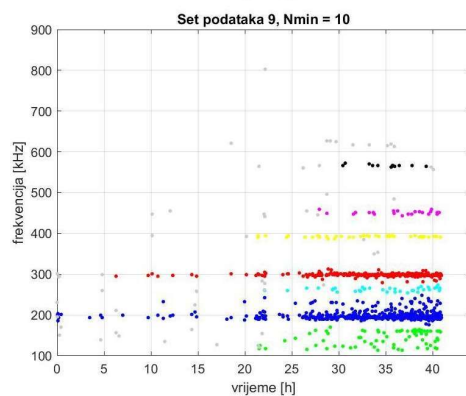
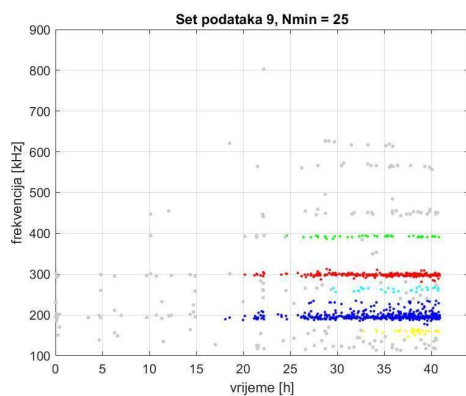
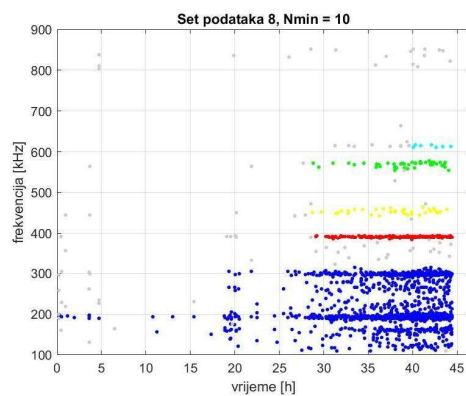
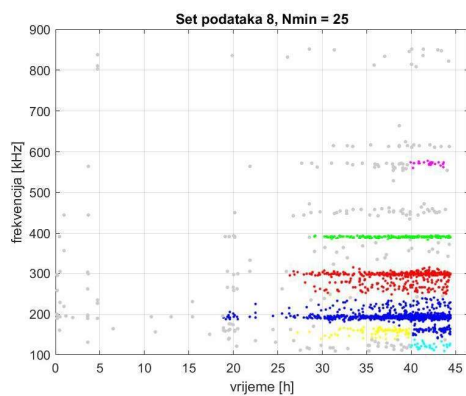
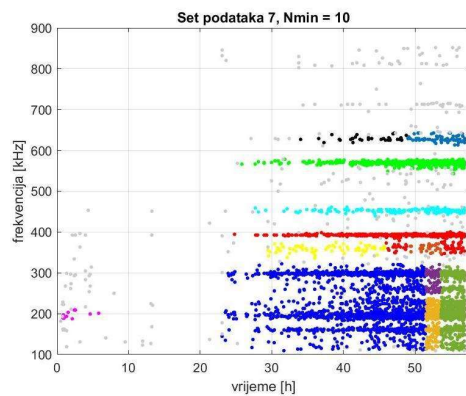
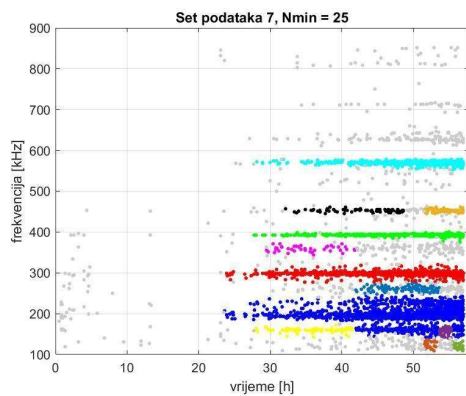
Drugi pristup adaptivnog određivanja parametra  $N_{\min}$  i  $\varepsilon$  nešto je kompliciraniji od prvog i zasniva se na određivanju parametara iz gustoće podataka u određenom satu. Slično kao i kod prethodne metode, postoji vremenski interval promatranja, ali se parametri ne određuju odmah nakon njegovog isteka. Umjesto prosječnog broja emisija po satu prati se njihov prosječni kumulativni broj po satu. Nakon isteka intervala promatranja, u satu u

kojem kumulativni broj emisija bude barem 4 puta veći od dotada izračunatog prosječnog kumulativnog broja emisija, parametar  $N_{\min}$  računa se prema izrazu (9).

$$N_{\min} = \text{maks. gustoća} * \varepsilon^2 * \pi \quad (9)$$

*Maksimalna gustoća* odnosi se na maksimalnu gustoću točaka unutar sata u kojem se zadovolji uvjet iz prethodnog odjeljka, a računa se na sljedeći način. Površina koju točke zauzimaju (1 h\*800 kHz) unutar promatranog sata podijeli se po frekvencijama na 8 dijelova od 100 kHz. 800 kHz je prosječan raspon frekvencija UAE (100-900 kHz). *Maksimalna gustoća* određuje se iz dijela koji sadrži najviše točaka tako da se broj točaka unutar njega podijeli s njegovom površinom  $P$  (1 h\*100 kHz). *Maksimalna gustoća* množi se s površinom kruga ( $\varepsilon^2 * \pi$ ) kako bi se dobio  $N_{\min}$ . Parametar  $\varepsilon$  opet je postavljen na fiksnu vrijednost 10. Rezultati za parametre određene ovom adaptivnom metodom prikazani su na Slici 3.4 i Slici 3.5.

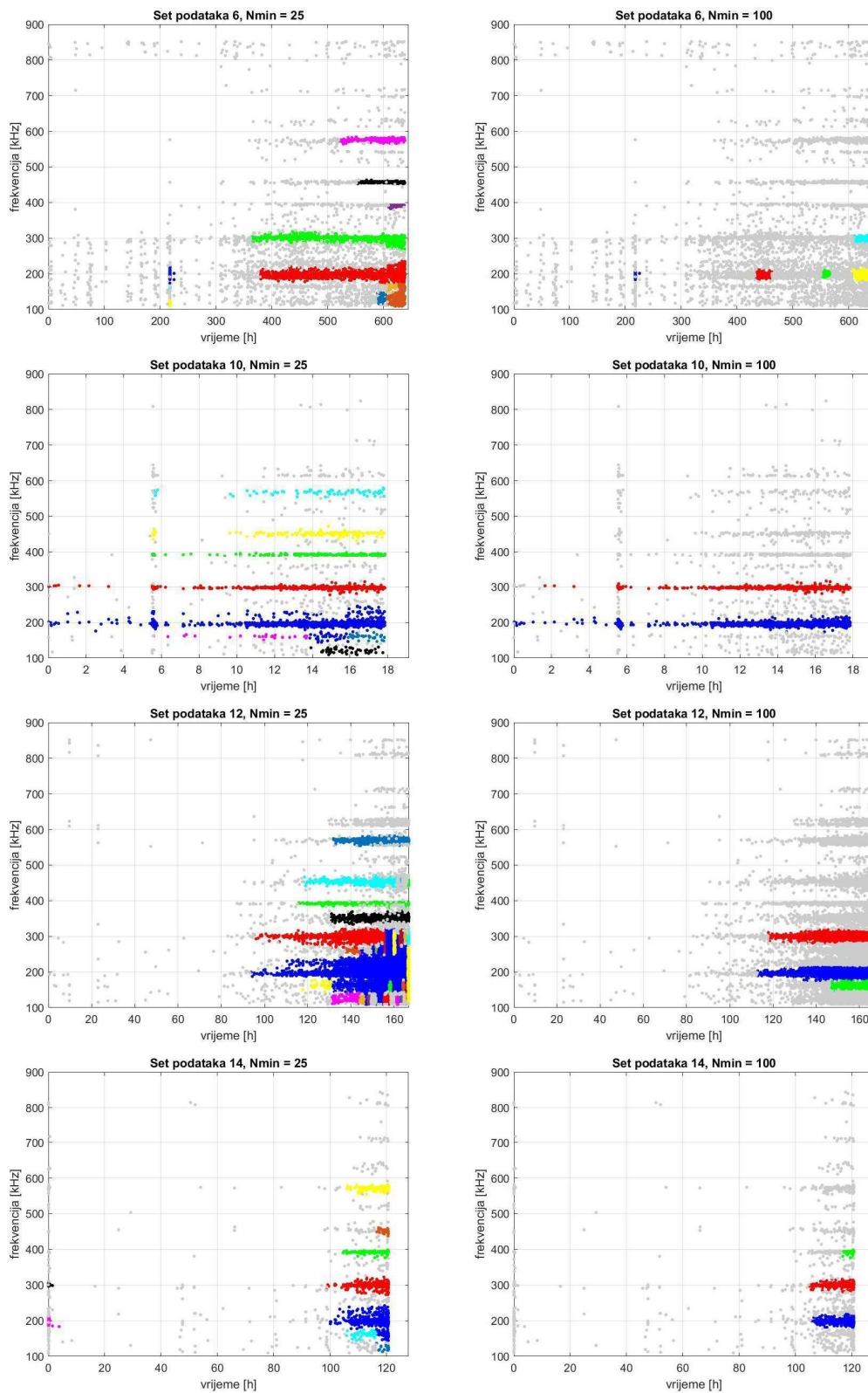
Nedostatak navedene metode je što parametar  $N_{\min}$  nije određen u zadanom trenutku, već se računa tek nakon dovoljnog porasta kumulativnog broja emisija pa može biti izračunat i kada je grupiranje već započelo kao što je na Slici 3.4 i Slici 3.5 prikazano isprekidanim vertikalnim crtama. U tom slučaju bi se na ugradbenom sustavu trebao implementirati i neki oblik pohrane podataka kako bi se naknadno mogli obraditi (kada se izračuna  $N_{\min}$ ).



a) Dobro odabran  $N_{\min}$

b) Loše odabran  $N_{\min}$

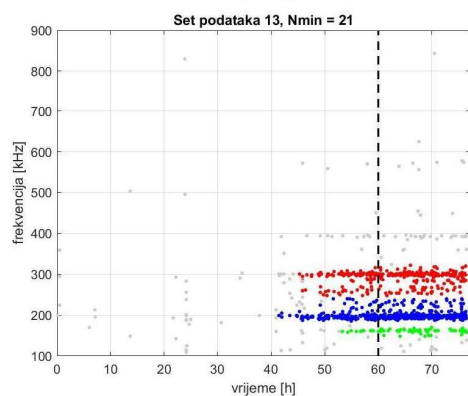
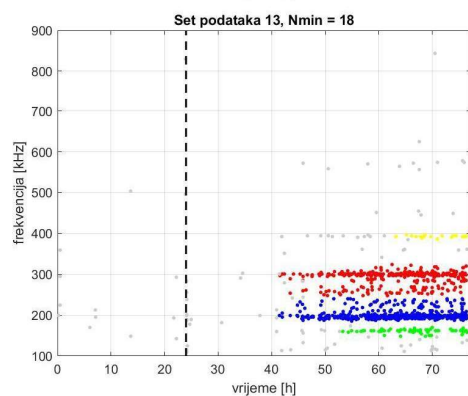
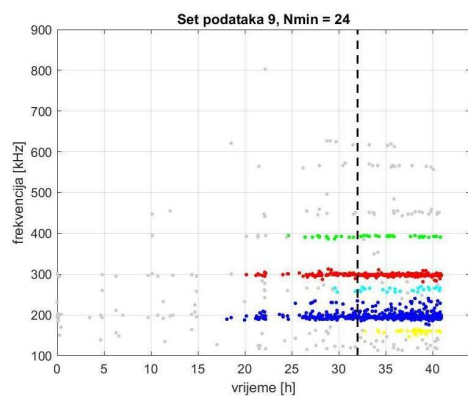
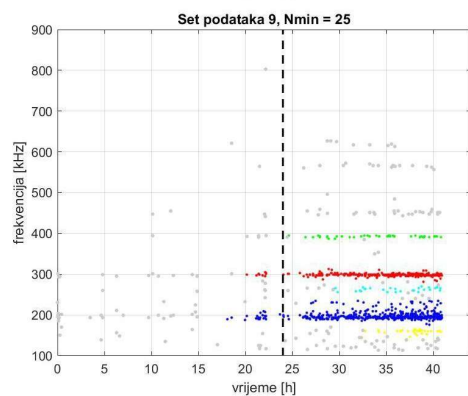
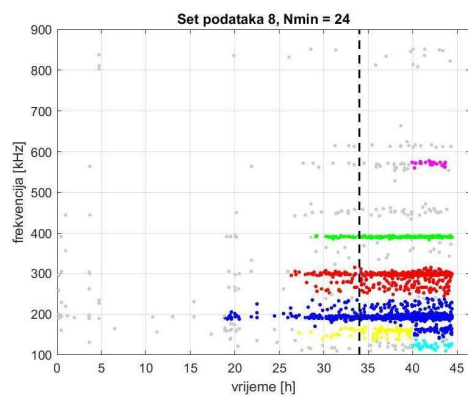
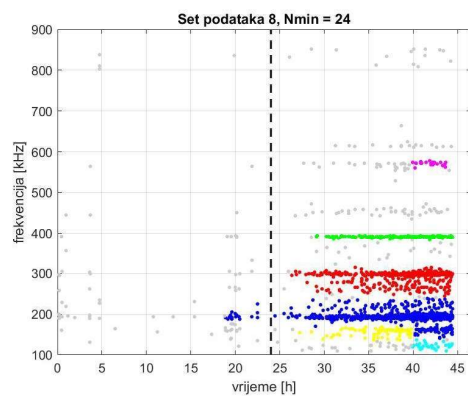
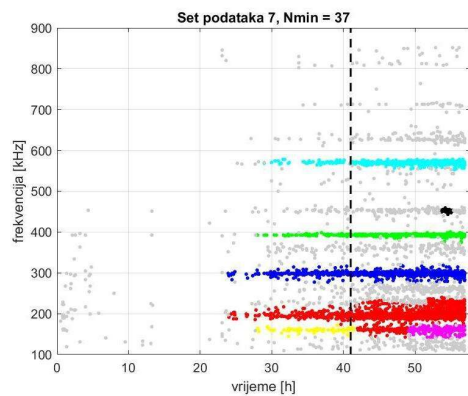
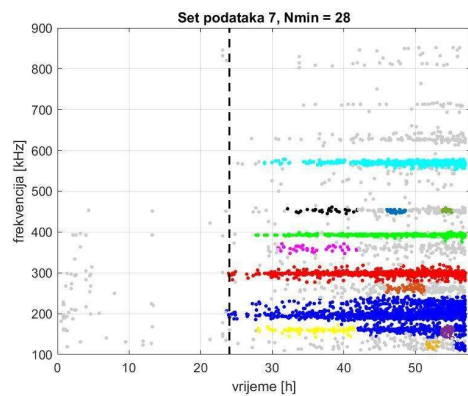
Slika 3.2 Rezultati algoritma za fiksno odabrane parametre. Setovi 7, 8, 9, 13



a) Dobro odabran  $N_{min}$

b) Loše odabran  $N_{min}$

Slika 3.3 Rezultati algoritma za fiksno odabrane parametre. Setovi 6, 10, 12, 14

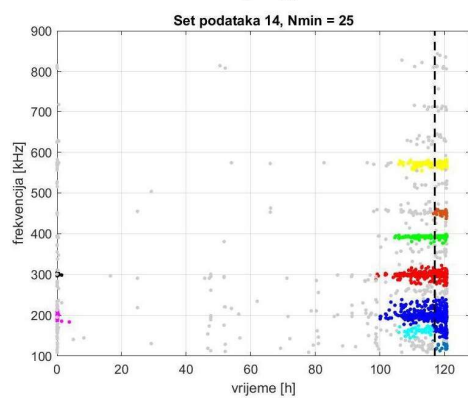
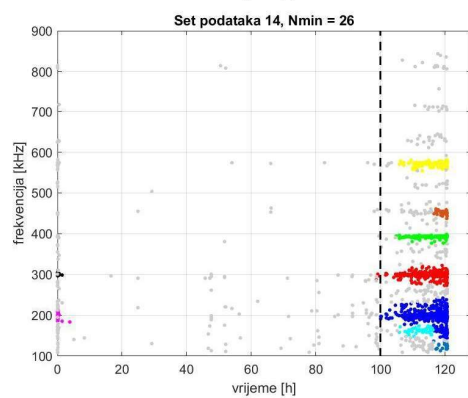
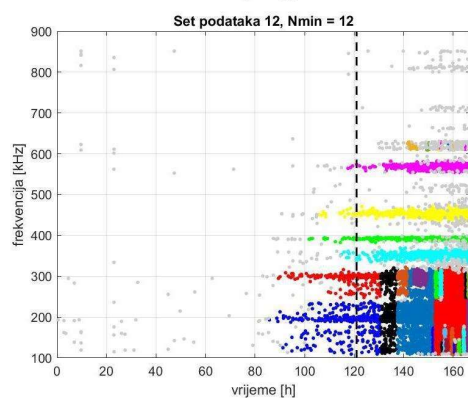
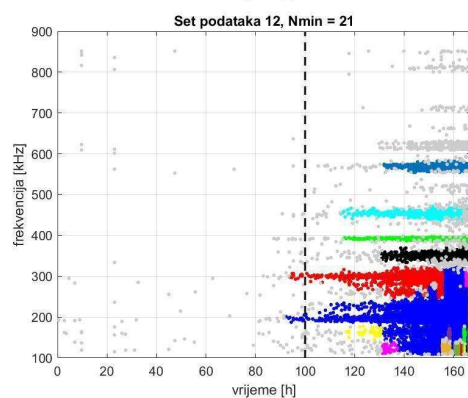
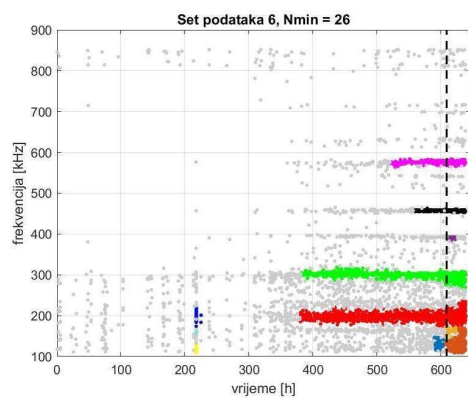
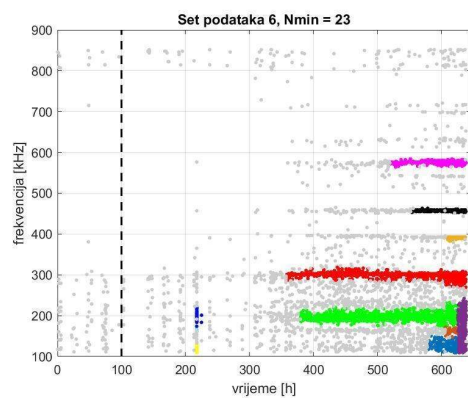


a) Metoda 1

b) Metoda 2

Slika 3.4 Rezultati algoritma za adaptivno odabrane parametre. Setovi 7, 8, 9, 13





a) Metoda 1

b) Metoda 2

Slika 3.5 Rezultati algoritma za adaptivno odabrane parametre. Setovi 6, 12, 14

## 4. Analiza implementacije na ugradbenom sustavu

U sljedeća dva potpoglavlja bit će riječi o vremenskoj analizi rada skupa algoritama te mjerenju potrošnje ugradbenog sustava na kojem je implementiran skup algoritama.

### 4.1. Vremenska analiza

Korištenjem Keil Event Recordera izmjerena su prosječna, minimalna i maksimalna vremena izvođenja za glavne dijelove skupa algoritama (Slika 1.1). Izmjerene vrijednosti su prikazane Tablicom 4.1. Vrijednosti iz tablice dobivene su mjerenjem trajanja obrade seta stvarnih podataka (set 7) po prozorima od 1000 točaka s preklapanjem od 200 točaka. Uz vremena trajanja prikazani su i brojevi instrukcija za sve dijelove.

Od 4 glavne cjeline algoritma daleko najdulje trajanje izvršavanja ima OPTICS algoritam. Složenost OPTICS algoritma je  $O(n^2)$ , a razlog tomu je repetitivno računanje udaljenosti između točaka. Naime, algoritam u prosječnom prolazu izračunava približno  $n$  udaljenosti za  $n$  točaka.

Tablica 4.1 Vremena izvođenja i broj instrukcija glavnih dijelova algoritma u jednom prozoru (1000 točaka), frekvencija procesora = 120 MHz

| Segment algoritma   | $T_{avg}$ [ms] | $T_{min}$ [ms] | $T_{max}$ [ms] | Br. instrukcija |
|---------------------|----------------|----------------|----------------|-----------------|
| OPTICS              | 784,83         | 694,83         | 810,03         | 82578648        |
| GC                  | 2,52           | 2,37           | 2,69           | 321872          |
| Spajanje prozora    | 0,47           | 0,37           | 0,51           | 43713           |
| Sažimanje rezultata | 0,50           | 0,47           | 0,62           | 74441           |
| Ukupno              | 788,32         |                |                | 83018674        |

Tablicom 4.2 prikazana su ukupna vremena izvođenja glavnih funkcija OPTICS algoritma. Vrijednosti u tablici su dobivene mjerenjem vremena izvođenja svih poziva navedenih funkcija unutar jednog prozora od 1000 točaka.

Najdulje ukupno vrijeme izvođenja ima funkcija `getNeighbors` zbog računanja udaljenosti između točaka. Izračunate vrijednosti se zatim privremeno spremaju i dalje koriste u funkcijama `calcCoreDist` i `update`. Kako je udaljenost definirana kao euklidska (Pitagorin poučak) ključna operacija je računanje korijena. Funkcija korištena za

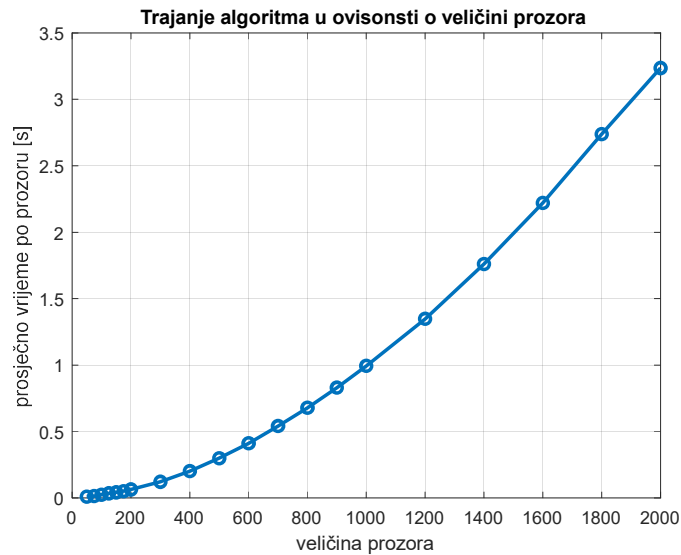
implementaciju na ugradbenom sustavu pripada CMSIS DSP biblioteci [16], a naziva se `arm_sqrt_f32`.

Tablica 4.2 Prosječno ukupno vrijeme izvođenja pojedinih funkcija OPTICS algoritma u jednom prozoru

| Funkcija u OPTICS algoritmu | $T_{avg}$ [ms] |
|-----------------------------|----------------|
| <code>getNeighbors</code>   | 361,28         |
| <code>calcCoreDist</code>   | 147,35         |
| <code>update</code>         | 20,76          |
| <code>sortSeeds</code>      | 16,17          |
| Ukupno                      | 542,56         |

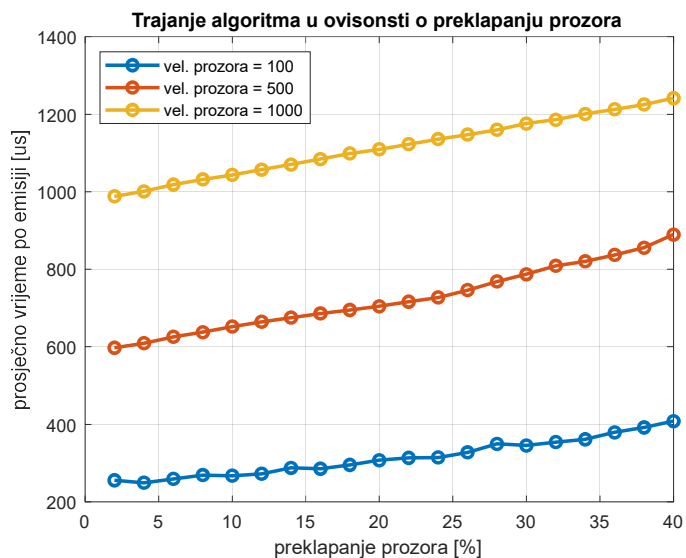
Sljedeća funkcija po duljini trajanja je `calcCoreDist`. Razlog tomu je sortiranje prethodno izračunatih udaljenosti kako bi se dobila jezgrena udaljenost. Korišteni algoritam za sortiranje je Quicksort zbog malog memorijskog zauzeća i brzine sortiranja. Kao što je već spomenuto u poglavlju 1.2, sortiranje se koristi i u funkciji `sortSeeds`. U tom slučaju koristi se Insertion sort s obzirom da tom tipu sortiranja pogoduje ako su podaci već pretežito sortirani.

Na vrijeme izvođenja skupa algoritama utječu njegovi parametri, najviše veličina prozora i preklapanje, i način implementacije. Parametri  $N_{min}$  i  $\epsilon$  također utječu na vrijeme izvođenja, ali je njihov utjecaj dosta teško procijeniti. Utjecaj veličine prozora na vrijeme izvođenja skupa algoritama prikazan je na Slici 4.1. Vremena su izmjerena kod izvođenja eksperimenta opisanog u poglavlju 2.4. Iz prikaza se može vidjeti kvadratna složenost OPTICS algoritma.



Slika 4.1 Prikaz ovisnosti trajanja algoritma o veličini prozora

Ovisnost o preklapanju prikazana je na Slici 4.2, ali je umjesto prosječnog vremena po prozoru korištena metrika prosječnog vremena po obrađenoj emisiji. Razlog tomu je što se većim preklapanjem produljuje ukupno vrijeme izvođenja za neki broj točaka, ali prosječno vrijeme po prozoru ostaje isto. Logično je očekivati da će izvođenje algoritma dulje trajati ukoliko je preklapanje veće zbog dvostrukog obrađivanja točaka preklapanja.



Slika 4.2 Prikaz ovisnosti trajanja algoritma o preklapanju

## 4.2. Potrošnja

Multimetrom Fluke 45 je mjerena struja napajanja ugradbenog sustava, a snaga je dobivena proračunom uz konstantan napon napajanja od 3,3 V. Potrošnja je izmjerena za četiri slučaja:

1. Ugradbeni sustav konstantno izvodi samo skup algoritama,
2. Ugradbeni sustav izvodi skup algoritama i prijenos podataka preko LPUART veze (*Low power* UART),
3. Ugradbeni sustav izvodi samo prijenos podataka jednog prozora preko LPUART veze,
4. Ugradbeni sustav je u stanju smanjene potrošnje (*Low power mode* STOP 0).

Dobivene vrijednosti prikazane su u Tablici 4.3. Kako se radi o *Low power* UART vezi, potrošnja za vrijeme prijenosa podataka manja je od potrošnje za vrijeme obrade podataka skupom algoritama. S obzirom da prijenos podataka jednog prozora traje dulje od njegove obrade, srednja struja je u slučaju 2 manja od struje u slučaju 1. Također, UART veza u konačnoj implementaciji na ugradbenom sustavu neće biti potrebna zbog direktne akvizicije podataka sa senzora pa je bitnije promatrati potrošnju za slučajeve 1 i 4.

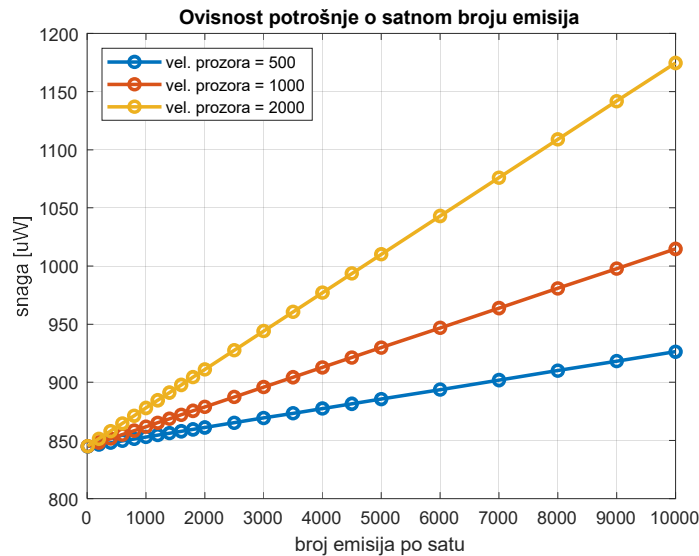
Izrada modela je nužna kako bi se mogla pokazati ovisnost potrošnje o satnom broju emisija za različite veličine prozora i preklapanje. Pri izradi modela su korištena vremena izvođenja algoritma dobivena eksperimentom opisanim u poglavlju 2.4. Također, kod proračuna vrijednosti se satni broj emisija smatrao konstantnim, odnosno za emisije se smatralo da se pojavljuju u jednakim vremenskim intervalima. Izvedeni model oponaša rad razvojnog sustava u stvarnom vremenu po sljedećem principu:

1. Sustav čeka dovoljan broj točaka da bi algoritmom obradio prozor (potrošnja za slučaj 4),
2. Sustav obrađuje podatke (potrošnja za slučaj 1),
3. Sustav ponovo čeka dovoljan broj točaka za obradu sljedećeg prozora.

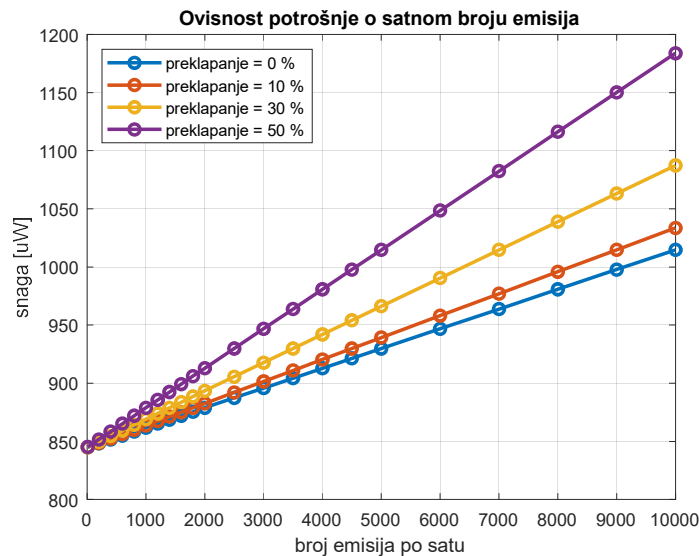
Tablica 4.3 Vrijednosti dobivene mjerenjem potrošnje

| Slučaj mjerenja: | Algoritam (1) | UART i alg. (2) | UART (3) | LPW STOP 0 (4) |
|------------------|---------------|-----------------|----------|----------------|
| Struja [mA]      | 22,600        | 20,000          | 20,000   | 0,256          |
| Snaga [mW]       | 74,500        | 66,000          | 66,000   | 0,845          |

Zanimljivo je promatrati ovisnost potrošnje o satnom broju emisija jer se većim brojem emisija smanjuje vrijeme koje ugradbeni sustav provodi u stanju niske potrošnje, a povećava vrijeme za obradu podataka. Pomoću izvedenog modela dobivena je ovisnost potrošnje o satnom broju emisija, Slika 4.3 i Slika 4.4.



Slika 4.3 Prikaz ovisnosti potrošnje o satnom broju emisija za različite veličine prozora, preklapanje = 0%



Slika 4.4 Prikaz ovisnosti potrošnje o satnom broju emisija za različito preklapanje prozora, veličina prozora = 1000 točaka

Vidljivo je da potrošnja s porastom broja emisija raste linearno, a također raste i s povećanjem veličine prozora i preklapanja zbog duljeg vremenskog izvođenja algoritma.

Iz ovako opisanog modela može se izračunati još jedan zanimljiv podatak - maksimalni broj emisija u satu ili minuti za koji skup algoritama može obraditi prozor prije pojave sljedećeg. Ako je, na primjer za obradu prozora od 1000 točaka potrebna 1 sekunda (Slika 4.1), minimalni broj za potpuni kvar iznosio bi  $3,6 \cdot 10^6$  emisija po satu. Naravno da je u stvarnosti ovaj broj puno manji s obzirom da se u proračun mora uzeti i broj točaka preklapanja te vrijeme potrebno za akviziciju samih točaka. Za usporedbu, najveći satni broj emisija u svim stvarnim setovima podataka iznosi 1298 (set podataka 12).

## 5. Zaključak

U ovom diplomskom radu je opisan skup algoritama za analizu ultrazvučnih akustičkih emisija u stvarnom vremenu na ugradbenom sustavu s ciljem praćenja vodnog stresa drvenaste biljke u pametnim sustavima za navodnjavanje. Ugradbeni sustav će svakih nekoliko dana ili sati, kada se prikupi dovoljan broj snimljenih emisija, podatke obrađivati skupom algoritama. Opisane su četiri glavne cjeline skupa algoritama: OPTICS algoritam, gradijentno grupiranje, spajanje prozora i sažimanje rezultata. Zadatak OPTICS algoritma i gradijentnog grupiranja je ekstrakcija frekvencijskih linija (grupa) iz vremensko-frekvencijskog prikaza ultrazvučnih akustičkih emisija. Spajanje prozora koristi se kod spajanja rezultata obrade međusobno susjednih blokova podataka, dok se sažimanje rezultata koristi za prikaz otkrivenih grupa jednostavnim geometrijskim likom.

Ispitivanjem skupa algoritama nad sintetičkim setovima podataka pokazano je da uz dobar odabir parametara, implementacija može postići visoke točnosti (preko 95%). Objašnjeno je kakav utjecaj na kvalitetu grupiranja imaju četiri glavna parametara: veličina prozora, preklapanje prozora, minimalni broj susjednih točaka koji tvore grupu ( $N_{\min}$ ) i udaljenost na kojoj se provjerava uvjet  $N_{\min}$  ( $\epsilon$ ). Dok se veličinu prozora i preklapanje prozora može odabrati skoro pa proizvoljno, kod odabira  $N_{\min}$  i  $\epsilon$  treba uzeti u obzir karakteristike seta podataka.

Skup algoritama ispitan je i nad setovima stvarnih podataka snimljenih ultrazvučnih akustičkih emisija. Ispitivanjem nad stvarnim podacima pokazano je kako kvalitetnim odabirom fiksnih parametara  $N_{\min}$  i  $\epsilon$  algoritam može uspješno odraditi grupiranje. Osim fiksnog odabira parametara, predložene su i dvije implementacije za adaptivan odabir  $N_{\min}$  i  $\epsilon$  iz karakteristika setova ulaznih podataka.

Analizom implementacije na ugradbenom sustavu prikazana su tipična vremena izvođenja skupa algoritama (994 ms) te tipična potrošnja ugradbenog sustava kod rada u stvarnom vremenu (869  $\mu$ W).

OPTICS algoritam može se dodatno optimizirati za kraće izvođenje, a gradijentno grupiranje može se unaprijediti za kvalitetnije prepoznavanje grupa iz rezultata OPTICS algoritma. Potencijalno se može razviti i drugi algoritam za spajanje prozora koji je u stanju prepoznati spajanje ili fragmentaciju grupa, a za rezultate mu nije potrebna ponovna obrada točaka preklapanja.



## Literatura

- [1] Holtta T., Vesala T., Nikinmaa E., Peramaki M., Siivola E., Mencuccini M. *Field measurements of ultrasonic acoustic emissions and stem diameter variations. New insight into the relationship between xylem tensions and embolism*, Tree Physiology, 25, 237–243 (2005), str. 1-3.
- [2] De Roo L, Vergelyn L, De Baerdemaeker N., Steppe K., *Acoustic Emissions to Measure Drought-Induced Cavitation in Plants*, (2016).
- [3] Oletić D. *Report on sensor response in laboratory conditions*, (2018).
- [4] Rosner S. *Acoustic Emission Related to Drought Stress Response of Four Deciduous Broad-Leaved Woody Species*, (2015).
- [5] Mayr S., Rosner S. *Cavitation in dehydrating xylem of Picea abies: energy properties of ultrasonic emissions reflect tracheid dimensions*, Tree Physiology, 31, 59–67 (2010), str. 63-66.
- [6] Vergelyn L., Sause M., De Baerdemaeker N., De Roo L., Steppe K., *Clustering reveals cavitation-related acoustic emission signals from dehydrating branches*, Tree Physiology, 36, 786–796 (2016), str. 789-792.
- [7] Ankerst M., Breunig M., Kriegel H., Sander J., *OPTICS: Ordering Points To Identify the Clustering Structure*, (1999).
- [8] Stranica proizvoda, Poveznica: <https://www.st.com/en/evaluation-tools/nucleo-l4r5zi.html>; pristupljeno 20. lipnja 2020.
- [9] MacQueen J. B. *Some Methods for classification and Analysis of Multivariate Observations*, (1967).
- [10] Sridharan R. *Gaussian mixture models and the EM algorithm*, (2014).
- [11] Ester M., Kriegel H. P., Sander J., Xu X. *A density-based algorithm for discovering clusters in large spatial databases with noise*, (1996).
- [12] Sinclair C., *Clustering using OPTICS*, Towards data science, (2019, siječanj). Poveznica: <https://towardsdatascience.com/clustering-using-optics-cac1d10ed7a7>; pristupljeno 21. lipnja 2020.
- [13] Brecheisen S., Kriegel H., Kroger P., Pfeifle M., *Visually Mining Through Cluster Hierarchies*, (2004), str. 406.
- [14] Knežević, K. *Binarna i višeklasna klasifikacija uporabom kartezijskog genetskog programiranja*. Poslijediplomski seminar. Sveučilište u Zagrebu Fakultet elektrotehnike i računarstva, 2017.
- [15] Shmueli B., *Multi-Class Metrics Made Simple*, Towards data science, (2019, srpanj). Poveznica: <https://towardsdatascience.com/multi-class-metrics-made-simple-part-i-precision-and-recall-9250280bddc2>; pristupljeno 21. lipnja 2020.
- [16] Stranica CMSIS-DSP biblioteke, Poveznica: <http://www.keil.com/pack/doc/CMSIS/DSP/html/index.html>; pristupljeno 21. lipnja 2020.

## Sažetak

### Prepoznavanje vremensko-frekvencijskog grupiranja u spektrima ultrazvučnih emisija

Pod povećanim vodnim stresom biljke, u njenom tkivu, ksilemu, dolazi do pojave mjehurića plina koji pucanjem proizvode mehaničke valove ultrazvučnih frekvencija, ultrazvučne akustičke emisije (UAE). Vremensko-frekvencijska analiza UAE u stvarnom vremenu na ugradbenom sustavu smještenom na biljci i procjena vodnog stresa, značajno mogu unaprijediti navodnjavanje u sustavima precizne poljoprivrede. Ovaj rad bavi se razvojem i implementacijom na ugradbenom sustavu skupa algoritama za stvarnovremensku analizu UAE i prijenos sažetih podataka, koji čine OPTICS algoritam, gradijentno grupiranje, spajanje prozora i sažimanje rezultata. OPTICS algoritam i gradijentno grupiranje koriste se za prepoznavanje frekvencijskih linija. Spajanje prozora koristi se za povezivanje rezultata obrade susjednih blokova podataka, dok se sažimanje rezultata koristi za prikaz otkrivenih frekvencijskih linija jednostavnim geometrijskim likom. Ispitana je točnost skupa algoritama nad sintetičkim podacima, gdje su dobivene točnosti preko 95 %. Opisana je mogućnost primjene rješenja nad stvarnim podacima te su predložene metode za adaptivno određivanje parametara skupa algoritama. Analizom implementacije na ugradbenom sustavu STM32 NUCLEO-144 određeno je prosječno vrijeme izvođenja (994 ms) i potrošnja (869  $\mu$ W).

Ključne riječi: OPTICS, stvarno vrijeme, ugradbeni sustav, vodni stres, precizna poljoprivreda

# Summary

## Recognizing Time-Frequency Clustering in Ultrasound Emission Spectra

Increased water potential in plants conducting tissue, xylem, produces gas bubbles, which when breaking produce mechanical waves of ultrasound frequencies, ultrasound acoustic emissions (UAE). Time-frequency analysis of UAE in real time on an embedded system mounted on a plant and water potential evaluation can greatly improve watering systems in high precision farming. This paper deals with the development and the implementation on an embedded system of an algorithm set for real time analysis of UAE and transferring compressed data. The algorithm set is made up of OPTICS algorithm, gradient clustering, window merging and result compression. The OPTICS algorithm and gradient clustering are used for extracting clusters. Window merging is used for connecting results of neighboring data blocks, while result compression is used for depicting the recognized clusters with a simple geometric shape. Accuracy of said algorithm set is tested on synthetic data with results over 95 %. The possibility of using the solution on real data is described and methods for adaptive parameter determining have been proposed. Analysis of the implementation on the embedded system SMT32 NUCLEO-144 shows average run time of 994 ms and power consumption of 869  $\mu$ W.

Keywords: OPTICS, real time, embedded system, water potential, precision farming