

Ventspils Augstskola
Informācijas tehnoloģiju fakultāte

Bakalaura darbs

**Vāju radioastronomisko objektu novērojumu
datu apstrāde - kalibrācija, filtrēšana un
rezultātu analīze**

Autors:

Ventspils Augstskolas
Informācijas tehnoloģiju fakultātes
bakalauru studiju programmas
datorzinātnēs
3. kursa students
Gints Jasmonts
Matrikulas Nr. 17020024

(paraksts)

Fakultātes dekāns:

Dr. sc. comp. Vairis Caune

(paraksts)

Zinātniskais vadītājs:

Mg. sc. comp. Karina Šķirmante

(paraksts)

Recenzents:

(Ieņemamais amats, zinātn. nosaukums, vārds, uzvārds)

(paraksts)

Ventspils, 2020

SATURS

ANOTĀCIJA	3
ABSTRACT	4
IZMANTOTIE SAĪSINĀJUMI UN TERMINI	5
IEVADS	6
1. RADIOASTRONOMISKO DATU NOVĒROJUMU APSTRĀDES METODIKA	8
1.1. Novērojuma plānošanas process	9
1.2. Datu nolasīšana	12
1.3. Kalibrācijas process	14
1.4. Doplera kompensācija un objekta pārvietošanās ātruma aprēķins	17
1.5. Veivletu izmantošana radioastronomisku signālu apstrādē	19
1.5.1. Veivletu definēšana	19
1.5.2. Veivletu pielietojums	20
1.6. MPI integrācija kalibrācijas procesā	22
1.7. Anomāliju detektācija	23
1.8. Citas trokšņa samazināšanas metodes	25
1.9. ACU žurnālēšanas faili	26
1.10. Nākotnes plāni un datu apstrādes tālākie soļi	29
2. DATU ANALĪZE UN IEGŪTIE REZULTĀTI	30
2.1. Novērojuma apstrāde	30
2.2. Iegūto rezultātu izvērtējums	36
2.3. Bakalaura darba izstrādes laikā novērotie kosmiskie objekti	40
3. ALGORITMA IZVĒRTĒJUMS	42
3.1. Datu apstrādes tehniskā informācija	42
3.2. Veiktspējas testi	42
3.3. Algoritma potenciālie uzlabojumi	46
SECINĀJUMI UN PRIEKŠLIKUMI	48
IZMANTOTĀS LITERATŪRAS UN AVOTU SARAKSTS	53

A	ATLAS C/2019 Y4 REDZAMĪBAS GRAFIKS	54
B	MPI ALGORITMA PIRMĀS VERSIJAS IZPILDES PROCESA ATTĒ- LOJUMS	55
C	MPI ALGORITMA FINĀLA VERSIJAS IZPILDES PROCESA ATTĒ- LOJUMS	57
D	BAKALAURA DARBA IETVAROS VEIDOTIE KODI	59
	GALVOJUMS	60

ANOTĀCIJA

Darba nosaukums:	Vāju radioastronomisko objektu novērojumu datu apstrāde - kalibrācija, filtrēšana un rezultātu analīze
Darba autors:	Gints Jasmonts
Darba vadītāja:	Mg. sc. comp. Karina Šķirmante
Darba apjoms:	54 lapas, 3 tabulas, 21 attēli, 17 formulas, 30 literatūras avoti, 4 pielikumi
Atslēgas vārdi:	OH MĀZERI, KOMĒTAS, VĀJI SIGNĀLI, DATU AP- STRĀDE, AUGSTAS VEIKTSPĒJAS SKAITĻOŠANA, ALGORITMI PARALĒLĀ REŽĪMĀ

Bakalaura darba mērķis ir veikt vāju radioastronomisko objektu datu apstrādi, to kalibrāciju, trokšņa filtrēšanu, analizēt un apkopot iegūtos rezultātus. Primāri tiek novēroti OH māzeri Saules sistēmas komētu iekšienē, kuri, tuvojoties Saulei, aktivizējas un uzsāk OH molekulu izstarošanu. Balstoties uz to, ka komētas daļēji sastāv no ledus maisījumiem, novērojumi var sniegt informāciju par ūdeni kosmosā, kā arī ūdens rašanos uz Zemes. Darbā tiek arī novēroti un apstrādāti daži ārpus Saules sistēmas objekti, kā mainzvaigznes un zvaigžņu veidošanās reģionu OH māzeri. Lai to realizētu, izmantots RT-32 teleskops, ar kuru tiek veikti novērojumi diapazonā, kurā atrodas OH molekulu spektrālās līnijas - 1665MHz un 1667MHz. Ņemot vērā, ka Ventspils Starptautiskais Radioastronomijas Centrā ar komētu novērojumiem darbs uzsākts tikai nesen, ir nepieciešams izveidot metodiku, kā plānot un apstrādāt iegūtos datus no minēto klasifikāciju radioastronomiskajiem avotiem.

Bakalaura darba praktiskajā daļā Python valodā tika izveidoti datu apstrādes metodikā iekļautie algoritmi (kalibrācijas, trokšņa mazināšanas un spektra iegūšanas, apvienojot vairāku novērojumu rezultātus), kas veic darbības paralēli, izmantojot Message Passing Interface protokolu, tādējādi efektīvāk pielietojot VSRC augstas veiktspējas skaitļošanas nodaļai pieejamos skaitļošanas resursus. Darbā algoritmi tiek aprakstīti un izvērtēti, kā arī tiek aprakstīti potenciālie uzlabojumi, ko iespējams veikt. Darbā tiek apskatīti arī potenciālie uzlabojumi un nākotnes plāni datu apstrādes metodikā.

Darba rezultātā ir iegūta metodika, uz kuras balstoties, ir iespējams apstrādāt vāju starojuma astronomisko objektu radio astronomiskos datus. Bakalaura darba apraksta ietvaros tiek apskatīts metodikas pielietojums uz ATLAS Y4/2019 komētas, kā arī tiek apskatīti rezultāti mainzvaigznei R Leonis Minoris un citiem novērotajiem objektiem.

ABSTRACT

Title:	Weak radioastronomical object observation data processing - calibration, filtration and result analysis
Author:	Gints Jasmonts
Academic Advisor:	Mg. sc. comp. Karina Šķirmante
Volume of the work:	54 pages, 3 tables, 21 images, 17 equations, 30 literature sources, 4 appendices
Keywords:	OH MASERS, COMETS, WEAK SIGNALS, DATA PROCESSING, HIGH PERFORMANCE COMPUTING, PARALLEL ALGORITHMS

The goal of this Bachelors degree thesis is to perform weak radioastronomic object data processing, which entails calibration, noise filtering and the analysis of results. Primarily, OH masers from within comets in Solar system are observed, which upon getting closer to Sun activate and start emitting OH molecules, but observations on objects outside Solar system, such as stars and OH masers from star forming regions, are also performed. Considering that comets are mostly made of ice, observations may give us more insight about water in space, which includes the appearance of water on Earth. To perform the task, RT-32 radio telescope is used to observe the radio wave region of which spectral lines of OH molecules may be observed. Since Ventspils International Radio Astronomy Centre has not performed comet observations until recently, there was no guidelines for data processing of such objects.

There are a number of algorithms for radio telescope data processing created during the writing period of this thesis. Algorithms are written using the Python programming language and are crucial for observation data transformation, calibration, signal noise reduction and data aggregation, some of which are parallelized using the Message Passing Interface to efficiently use the high performance computing resources of VIRAC. Algorithms are explained and the potential improvements are noted. Potential improvements on the guidelines of OH maser observation data processing are also looked at.

As a result of this thesis, a set of instructions for successful OH maser observation data processing is produced and examples from a couple of observed objects, such as comet ATLAS Y4/2019 and Mira variable star R Leonis Minoris are examined.

IZMANTOTIE SAĪSINĀJUMI UN TERMINI

MPI - Message Passing Interface protokols, Bakalaura darba ietvaros tiek izmantots, lai veiktu komunikāciju starp vairākiem *Python* procesiem gan vienas iekārtas, gan vairāku skaitļošanas mezglu ietvaros.

Novērojums - Teleskopa fiziski pozicionēšana pret objektu uz noteiktu laiku un datu ierakstīšana laika periodā (parasti 2-4 stundas).

Skans - Novērojuma fragments, kurā ietilpst visas fāzes aptuveni 2 minūšu novērojuma intervālā.

FFT - Fast Fourier Transform. Algoritms, kurš aprēķina diskrēto Furjē transformāciju, kura pārveido signālu frekvenču reprezentācijas modelī.

Ns - Furjē transformācijas rezultējošais garums

VSRC - Ventspils Starptautiskais Radioastronomijas Centrs

VIRAC - Ventspils International Radio Astronomy Centre

Māzeris - (Angliski *maser*). Microwave Amplification by Stimulated Emission of Radiation. Dažu mikroviļņu starojuma pastiprināšanās sakarā ar inducētā starojuma rašanos komētas molekulās.

CSV - Comma Separated Value. Faila formāts, kur vērtības atdalītas ar atdalīšanas simbolu, bieži - komatu.

Dat formāts - Formāts, kur kolonnās ierakstītas radioastronomiskā novērojuma rezultātā iegūtās vērtības.

Jēldati - Binārais datu formāts, kur spektrs ierakstīts reālu un imagināru skaitļu virknē.

ACU - Antenna Control Unit. Sistēma, kuru izmantojot tiek nodrošināta radioteleskopa pozicionēšana pret novērojuma objektu un pozīcijas datu žurnālēšanu.

SDR - Software Defined Radio, iekārta, kura pārveido radioteleskopa novēroto analogo signālu par digitālo signālu.

IEVADS

Komētas var uzskatīt par planetāro sistēmu veidošanās paliekām, līdz ar to, komētu struktūras izpēte sniedz ieskatu elementu kompozīcijā, no kuras sastāvēja pirmatnējie zvaigžņu miglāji, no kuriem ir veidojusies arī Zeme. Komētas pētījumus var veikt vairākos veidos, piemēram, veicot optiskus novērojumus, nosūtot zondes, kas pēta komētu struktūru, vai arī komētas novērot radio frekvenču diapazonā. Tuvojoties Saulei vai kādam citam objektam, kurš izstaro lielu daudzumu elektromagnētisko radiāciju, komētās esošais ledus uzsilst un tiek uzsāktas dažādas ķīmiskas reakcijas. Balstoties uz ķīmiskajām reakcijām, komētās aktivizējās OH māzeri un uzsāk OH molekulu izstarošanu, kuru iespējams novērot 1665MHz un 1667MHz frekvencēs, izmantojot radioteleskopu. Diemžēl komētas OH māzera starojums ir ļoti vājš radio frekvencēs, tipisku komētu OH māzeru enerģijas plūsma mērāma no 4 līdz 40 mJy.

Lai veiktu novērojumus, tiek izmantots VSRC īpašumā piederošais RT-32 radioteleskops, kura L(1 – 2 GHz) joslas frekvenču uztvērējs kopā ar Cassegrain antenu, veido trīs spoguļu sistēmu [1], kas atļauj veikt ļoti vāju starojumu novērojumus, ieskaitot komētas OH māzeru novērojumus. VSRC radioteleskops piemērots komētu novērošanai, jo novērojumus ir iespējams veikt 360 grādu leņķī, līdz ar to, ir nepieciešams tikai pārliecināties, vai objekts ir virs horizonta.

Radioastronomijā vāju radioastronomisku objektu datu apstrāde ir sarežģīts process, jo apkārtējie avoti ar savu starojumu ļoti ietekmē rezultējošo signālu, kas veido nepieciešamību izveidot pēc iespējas efektīvākas metodes signāla apstrādei un trokšņa filtrēšanai.

Lai realizētu Bakalaura darbā izvirzīto mērķi - veikt vāju radioastronomisko objektu datu apstrādi, to kalibrāciju, trokšņa filtrēšanu, analizēt un apkopot iegūtos rezultātus, tiek ieviesti sekojošie uzdevumi:

1. apgūt radioastronomisko novērojumu infrastruktūru un iegūt izpratni par datu tiem, kuri tiek izmantoti novērojumu datu saglabāšanā;
2. apgūt signālu apstrādes pamatus un implementēt novērojumu datu apstrādes algoritmus;
3. iepazīties ar pieejamiem datu avotiem komētu novērojumu plānošanai un izveidot algoritmus datu saglabāšanai un nolasīšanai no minētajiem avotiem;
4. apkopot visu novērojumu plānošanas un datu apstrādes procesu vienotā dokumentā, uz kuru balstoties, iespējams veikt vāju radioastronomisku datu apstrādi;

5. izstrādāt metodiku, ar kuru var veikt vāju starojumu datu apstrādi;
6. izstrādāt programnodrošinājumu, kas veic datu apstrādi, tai skaitā datu kalibrāciju, trokšņa filtrēšanu un spektru apkopošanu;
7. papildināt izstrādāto programnodrošinājumu, iekļaujot paralelizāciju.

Darbu uzsākot, tika apskatīta eksistējošo dokumentāciju novērojumu veikšanas infrastruktūrai ([2], [3], [4], [1], [5]), kā arī apskatīti algoritmus, kuri tiek izmantoti, lai realizētu datu ierakstīšanu.

Lai izpildītu otro uzdevumu, Bakalaura darba ietvaros tiek realizēta datu nolasīšana un apstrādāšana ar Furjē transformāciju un Blackman-Harris ([6], [7]) funkciju, datu kalibrācija, bojātu datu detektēšana, kā arī vairāku novērojumu datu apvienošana. Novērojumu plānošanai, kā arī rezultātu apvienošanai tiek izmantoti algoritmiski iegūtie dati no NASA HORIZONS JPL [8] datubāzes.

Nemot vērā, ka datu apstrāde ir resursietilpīgs process, bija nepieciešams apskatīt un implementēt paralelizācijas iespējas izmantojot Message Passing Interface protokolu [9]. MPI algoritms tika implementēts divās iterācijās un Bakalaura darba apraksta ietvaros tiek apskatīts to darbību pamatprincipi.

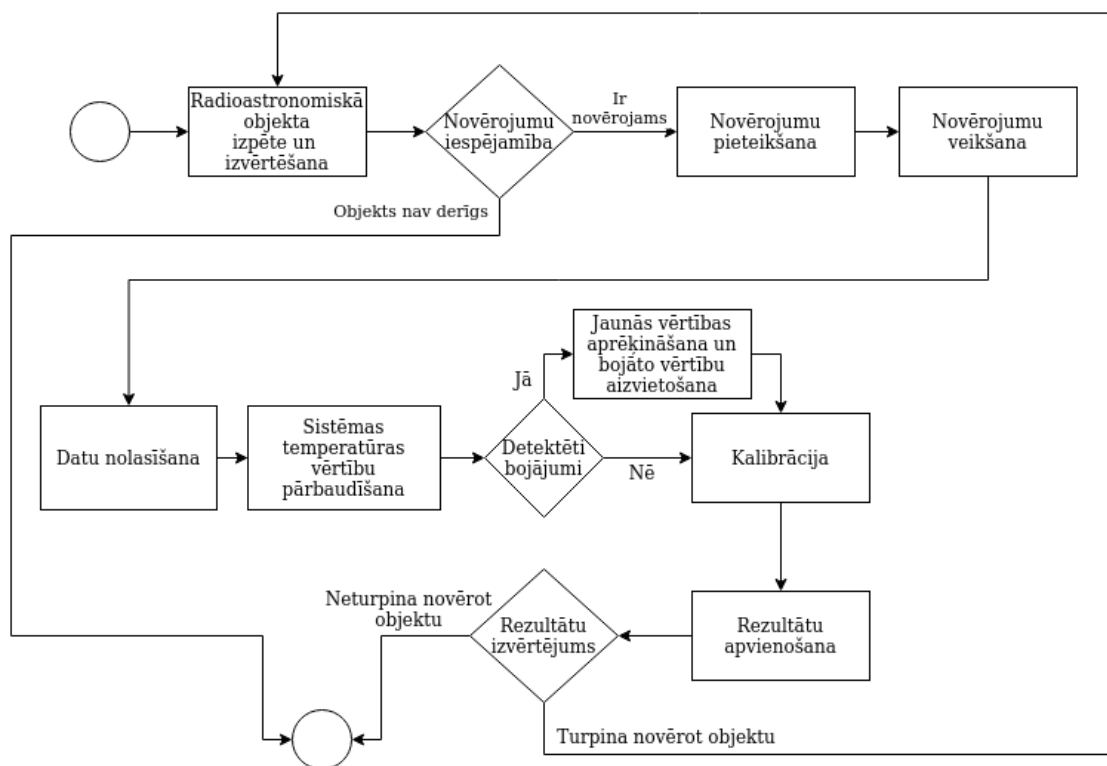
Novērojumu apstrādes process no sākuma līdz beigām tiek aprakstīts aprakstā, apskatot gan teorētisko pamatojumu darbībām ([10], [11]), gan iegūtos rezultātus, kā arī tiek apskatīti nākotnes plāni un metodikas potenciālie uzlabojumi.

Bakalaura darba apraksts strukturēta 3 nodaļās, kur pirmajā nodaļā tiek aprakstīts metodikas teorētiskais process, kā arī aprakstītas metodes procesu implementācijai. Otrajā nodaļā metodika tiek pielietota radioastronomisku novērojumu datiem un tiek attēlots piemērs metodikas izmantošanai. Lai izvērtētu augstas veiktspējas skaitļošanas algoritmu, kurš ir viens no primārajiem mērķiem darbā, tiek ieviesta trešā nodaļa, kur tiek veikti veiktspējas testi un aprakstītas algoritma potenciālie uzlabojumi un nepilnības.

Bakalaura darba ietvaros veidotie algoritmi ir pieejami repozitorijā pielikumā D, kur pieejams arī Bakalaura darba apraksta pirmkods, kā arī Bakalaura darba izstrādes laikā iegūto rezultātu augstākas izšķirtspējas attēli.

1. RADIOASTRONOMISKO DATU NOVĒROJUMU APSTRĀDES METODIKA

Ventspils Starptautiskais Radioastronomijas Centrs ar kosmosa objektu novērošanu radio frekvenču diapazonā nodarbojas jau vairāk kā 10 gadus, bet tikai pēdējā gaida laikā VSRC pētnieki veic komētu novērojumus, kas sevī ietver komētas kodolā izveidojušo OH māzeru novērojumus. Bakalaura darba ietvaros tika izveidota metodika, kuras process atspoguļots 1.1. attēlā, kuru izmantojot ir iespējams apstrādāt iegūtos OH māzeru novērotos datus. Tiek arī apskatīti komētu novērojumu veikšanas kritēriji, piemēram, redzamības noteikšanas virs horizonta, objekta spilgtums un citi kritēriji.



Att. 1.1.: Novērojumu apstrādes metodikas modelis

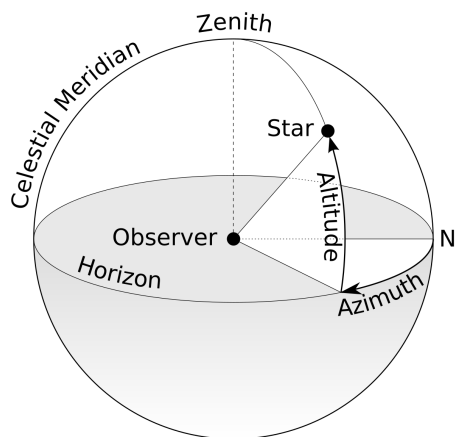
Bakalaura darba ietvaros tiek aprakstīti visi soļi, izņemot *Novērojumu pieteikšanas* un *Novērojumu veikšanas* soļi, jo minētie soļi ir aprakstīti VSRC un Ventspils Augstskolas nolikumos un iekšējos dokumentos. Visi pārējie soļi aprakstīti pēc sekojošās metodikas

- aprakstot teorētisko pamatojumu, realizāciju un iegūtos rezultātus. Radioastronomiskā objekta izpēte un izvērtēšana tiek apskatīta nodaļā 1.1., kur tiek apskatīti raksturlielumi objekta novērošanā, kā arī teorētiskais apraksts novērojumu plānošanai.

Datu nolasīšanas sadaļa aprakstīta nodaļā 1.2., sistēmas temperatūras pārbaudes aprakstītas anomāliju detektēšana nodaļā 1.7., bet pats kalibrācijas process aprakstīts 1.3. nodaļā. Minēto soļu praktisku piemēru var apskatīt 2.1. sadaļā. Novērojumu apvienošanas matemātiskais process aprakstīts 1.4. nodaļā, bet radioastronomisko datu apstrādes rezultātus var apskatīt nodaļā 2.2.

1.1. Novērojuma plānošanas process

Lai veiktu radioastronomiskā objekta novērošanu, ir jāņem vērā vairāki faktori, kas ietekmēs rezultātu precizitāti un iespējamību. Lai gan turpmāk aprakstītās nodaļas novērojumu veikšanas kritēriji ir svarīgi, jāņem vērā, ka komētu, kuras iespējams detektēt radio frekvenču diapazonā, ir salīdzinoši maz, balstoties uz to, ka komētu OH māzeri izstaro vājus signālus, salīdzinot ar citiem OH māzeriem, līdz ar to ir ieteicams novērot visas komētas, kuras potenciāli detektējamās. Balstoties uz to, ka komētas daļēji sastāv no ledus maisījumiem, novērojumi var sniegt informāciju par ūdeni kosmosā, kā arī ūdens rašanos uz Zemes. Komētu sastāva izpēte sniedz vairāk informāciju komētu zinātniskajai bāzei, tajā skaitā, informāciju par ūdens rašanos uz Zemes, jo komētu sastāvā ir sastopams ūdens ledus formā.

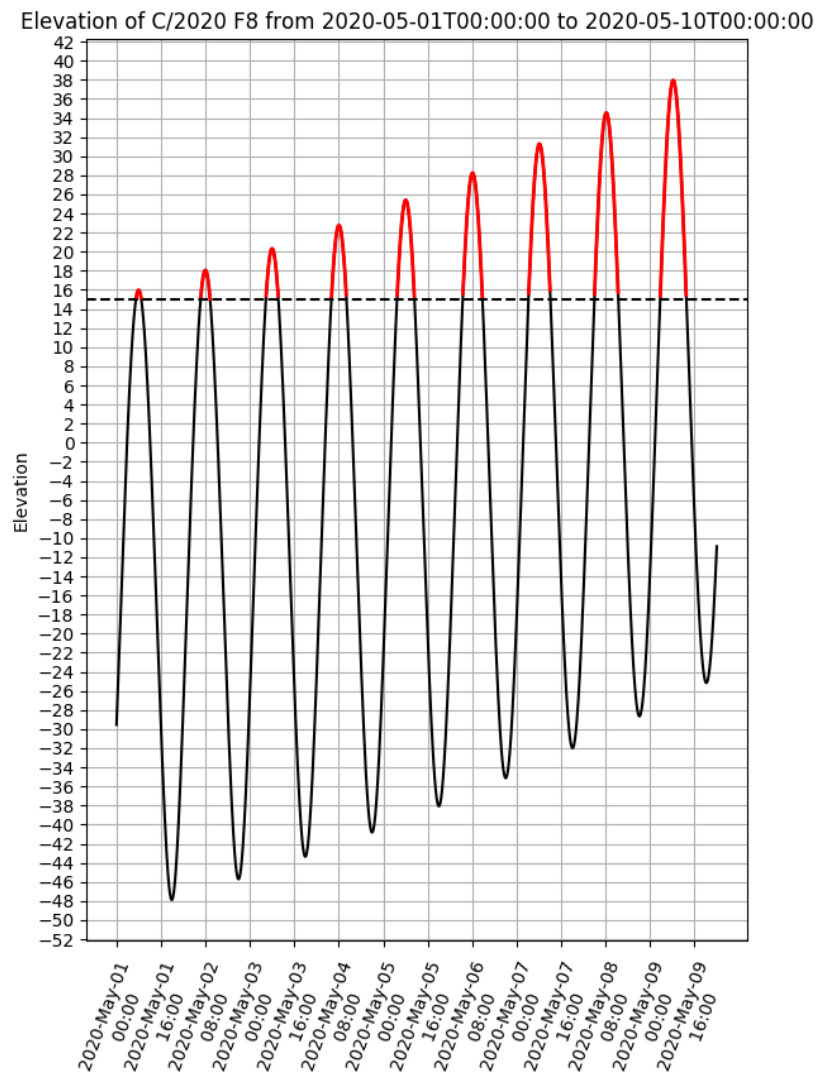


Att. 1.2.: Topocentriskās koordinātu sistēmas attēlojums [12]

Par svarīgāko faktoru komētas detektēšanai ar radioteleskopu var uzskatīt objekta pozīciju. Astronomijā tiek izmantota topocentriskā koordinātu sistēma, kurā ir divas neatkarīgas lenķiskās koordinātes - azimuts un elevācija. Sistēma attēlota attēlā 1.2., kur azimuts - Azimuth, bet elevācija - Altitude. Attēlā, objekts ir zvaigzne, bet zvaigznes vietā var būt jebkurš astronomisks objekts, kurš izstaro radio signālu. Bakalaura darba ietvaros izmantotais Irbenes RT-32 teleskops spējīgs novērot objektus rotējot 360 grādu lenķī, līdz ar to vienīgās vērtības, kuras jāņem vērā, lai novērojums būtu iespējams, ir elevācijas vērtība. Lai gan pēc radioteleskopa lietošanas specifikācijas minimālā elevācijas vērtība ieņem 2.7 grādu vērtību [2],

praksē pierādījies, ka elevācijas vērtībām jāpārsniedz 15 grādu robežām, lai iegūtais rezultāts būtu izmantojams, jo objektu novērošana tuvu horizontam nozīmē, ka tiek pastiprināti uztverti apkārtējie radio signālu avoti, kas atrodas uz Zemes virsmas, kā mobilie torņi, Ovīšu bāka un citi. Novērojumu sesijas ilgst vismaz 2 stundas, līdz ar to ir jāpārlicinās, ka elevācijas vērtības ir derīgas visa novērojuma laikā.

Objekta redzamības grafiku var apskatīt attēlā 1.3., kur apskatīts SWAN C/2020 F8 komētas elevācijas vērtības atkarībā no radioteleskopa un objekta pozīcijas no 1. maija līdz 10. maijam, ar sarkanu krāsu iezīmējot potenciāli iespējamos novērojumu laikus.¹ ATLAS komētas redzamības grafiku skatīt pielikumā A. Visi grafiki nodaļā tiek izveidoti, izmantojot *Matplotlib Python* bibliotēku [13].



Att. 1.3.: SWAN C/2020 F8 elevācijas atkarība no laika

1. Elevācijas vērtības iegūtas no NASA HORIZONS sistēmas

Veicot vāju radioastronomisku objektu novērojumus, ir nepieciešams novērojuma sesijas laikā novērot arī kādu spēcīgāku avotu jeb kalibratoru, kuru izmanto kontrolei. Tas nozīmē, ka arī kalibratoram ir jābūt redzamam sesijas laikā. Pēdējais nosacījums, objektu novērošanas iespējamībai fiziskā līmenī, ir radioteleskopa noslogojums. Ņemot vērā, ka uz teleskopa izmantošanu ir pieteikušies gan VSRC zinātnieki, gan citu valstu astronomisko institūtu zinātnieki, ir jāņem vērā citu zinātnieku vēlmes un jāplāno savstarpēji vislabākie laiki.

Komētu datu kvalitāti ļoti ietekmē komētas spilgtums. Balstoties uz to, ka komētas aktīvi tiek novērotas optiskajā diapazonā, komētu spilgtums bieži tiek mērīts magnitūdās, pēc kurām iespējams novērtēt komētas spektrālo plūsmas blīvumu novērojumos radio diapazonā. Komētu spilgtumu precīzi noteikt nav viegli, taču izmantojot vairākus avotus, var iegūt aptuveno objekta spilgtumu magnitūdu skaitu. Katra magnitūda atbilst 2.512 spilgtuma faktoram, kas nozīmē, ka 1 magnitūdas spilgtuma objekts ir aptuveni 100 reizes spilgtāks par 6 magnitūdas spilgtuma objektu. Bakalaura darba ietvaros tika novērotas komētas, kuru aprēķinātais spilgtums svārstījās robežās no 12 līdz 4 magnitūdām.²

Viens no noteicošajiem faktoriem komētu novērošanas procesā ir komētas pārvietošanās pa tās orbītas. Lai veiksmīgi iegūtu augstas kvalitātes rezultātus, izmantojot patreizējās tehnoloģijas, ir nepieciešams novērot objektu ilgstošā laika periodā. Novērošanas ilgums katrai komētai atšķiras, balstoties uz tās strukturālajiem parametriem un orbītas, pa kuru komēta pārvietojas, taču Bakalaura darba izstrādes laikā, galvenajām novērotajām komētām, minimālais novērojumu laiks bija 110 stundas. Pilnus novērojumu laikus, kā arī citu informatīvo informāciju var apskatīt tabulā 2.2..

Komētai tuvojoties Saulei, palielinās Saules ietekme uz komētu, kas izraisa ķīmiskas reakcijas komētas iekšienē. Komētai ir kodols, kas sastāv no H_2O un papildus ķīmisko elementu piemaisījumiem. Miera stāvoklī kodols ir neaktīvs un sasalis, bet komētai tuvojoties Saulei, Saules ultravioletais starojums ietekmē kodolu. Kodols tiek aktivizēts un tā iekšienē veidojās OH māzeris, kas izstaro OH molekulas. Minēto starojumu ir iespējams novērot radio diapazonā. Kodola aktivitāte ir proporcionāla starojuma spilgtuma jaudai.

Veicot novērojumu, nepieciešams ņemt vērā arī novērotā objekta pozīciju attiecībā pret Zemi. Ņemot vērā, ka OH māzera izstarotais signāls ir salīdzinoši vājš, situācijās, kad komēta ir aktīva, taču atrodas tālu prom no Zemes, nav iespējams objektu novērot citu objektu izstaroto radio starojuma dēļ. Pēc teorijas, radioteleskops spējīgs novērot avotus tūkstošiem astronomisko vienību attālumā, pieņemot, ka avots ir pietiekami spēcīgs. Praksē pierādījies, ka komētām jābūt maksimāli divu astronomisko vienību attālumā no novērošanas punkta, kur komēta, ideālā gadījumā, novietota starp Zemi un Sauli.

2. Spilgtuma vērtības iegūtas no NASA HORIZONS sistēmas

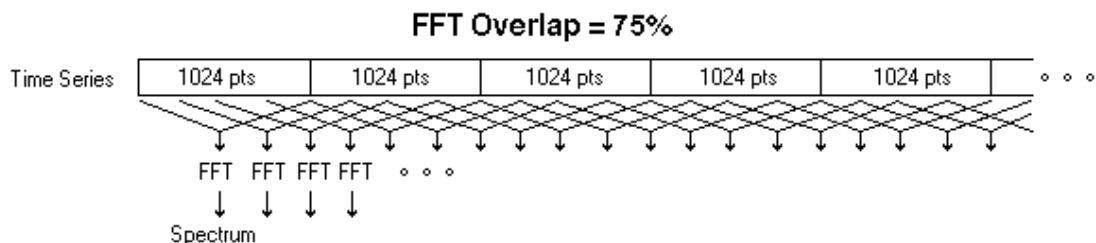
Neskatoties uz precīziem skaitļiem, ir jāņem vērā arī komētas popularitāte starp citiem novērotājiem. Jo vairāk astronomi novēro vienu objektu, jo vairāk informācijas pieejama, tādējādi radot iespēju salīdzināt iegūtos rezultātus, kā arī no vairākiem novērojumiem ir efektīvāk iespējams uzzināt objekta aktualitātes. Lai gan pārsvarā astronomi novēro komētas optiskajā diapazonā, rezultāti sniedz informāciju par komētas izmaiņām, piemēram, pateicoties minēto avotu informācijai, bija iespējams detektēt ATLAS Y4/2019 kodola fragmentāciju, kas ļoti ietekmē novērošanas plānus.

1.2. Datu nolasīšana

Lai uzsāktu datu apstrādi, ir nepieciešams vispirms izprast datu tipa uzbūvi. Bakalaura darbā tiek apskatīti divi datu tipi, kuri izmantoti komētu novērojumu datu ieguvē un izguvē - jēldati un dat tipa faili. Nodaļā tiks aprakstīta atšķirība starp tiem, kā arī apstrādes process datu izmantošanai.

Katrs novērojuma skans jēldatu gadījumā ir sadalīts astoņās atsevišķās datnēs, kur katrā ir ierakstīts spektrs $int16+int16$ reālu un imagināru skaitļu virknē. Lai nolasītu datus, pēc *SDR* spektrometra servera dokumentācijas [3], Bakalaura darba rakstīšanas laikā visai jēldatu apstrādei papildus Furjē transformācijai, jāizmanto *Blackman-Harris* logošanas funkcija [7], lai iegūtu rezultējošu spektru. Balstoties uz to, ka ieteiktā pārklāšanās robeža *Blackman-Harris* funkcijai ir 66.1% [6], tika izveidots algoritms, kas sadala nolasītos datus tvertnēs, kuru garums ir identisks žurnālēšanas failu ierakstītajai N_s vērtībai.

Katrai tvertnei tiek izsaukta *Fast Fourier Transform* (turpmāk FFT), kā arī *Blackman-Harris* funkcija. FFT realizēta, izmantojot *numpy FFT* pakotni [14], bet *Blackman-Harris* logošanas funkcija izmantojot *scipy signal* pakotni [15]. Pēc FFT veikšanas, tiek apvienotas spektra reālās un imaginārās daļas, un rezultāts ir pieskaitīts kopējai spektru summai. Kad visas datu tvertnes ir apstrādātas, nulles frekvences ir pārvietotas uz vidu un rezultējošais spektrs ir izdalīts ar tvertņu daudzumu, iegūstot vidējo spektru, kas ir skaitļu masīvs ar amplitūdu vērtībām, kuras iegūtas no konkrētā faila. Vizuāli process izklāstīts attēlā 1.4., kur attēlots pārklāšanās process izmantojot 75% pārklāšanās koeficientu. Lai gan attēlā attēlots 75% pārklāšanās rezultāts, datu apstrāde ar 66.1% pārklāšanās koeficientu ir ļoti līdzīga, jo mainās tikai kopējais spektru daudzums.



Att. 1.4.: Furjē transformācijas pielietojums uz datu tvertnēm ar 75% pārklšanās koeficientu. [16]

Dat formāta datu gadījumā, dati jau ir apstrādāti ar iepriekš minētajām funkcijām, līdz ar to tie ir ierakstīti *CSV* tipa failā trijās kolonnās. Pirmajā kolonnā ierakstītas frekvenču joslas vērtības konkrētajām amplitūdām, otrajā kolonna ierakstītas kreisās cirkulārās polarizācijas amplitūdu vērtības, bet trešajā ierakstītas labās cirkulārās polarizācijas amplitūdu vērtības. Tā kā abas polarizācijas ir saglabātas vienā failā, kopējais failu daudzums vienam skanam ir četri.

Kopējais failu skaits novērojumā atkarīgs no novērojuma ilguma, taču katram failam nav precīza ierakstīšanas ilguma. Četru stundu novērojumiem kopā ir 240 skani, bet divu stundu novērojumiem 120. Balstoties uz minēto informāciju, var detektēt situācijas, kur novērošanas sesijā bijušas tehniskas problēmas, jo patiesais failu daudzums neatbilst plānotajam, un izvērtēt, vai datus izmantot.

Lai gan dat formāta failu rezultāti ir neprecīzāki par jēldatu failu rezultātiem, tie tiek biežāk izmantoti sava ērtuma dēļ, jo bieži tiek apskatīti spēcīga signāla objekti un noklusējuma konfigurācija tiem ir apmierinoša. Priekšrocība jēldatiem ir to konfigurēšanas iespējas, taču jāņem vērā, ka to apstrādei nepieciešami vairāk skaitļošanas un laika resursi. Novērojot objektus un rezultātus ierakstot jēldatu formātā, nepieciešama salīdzinoši liela datu krātuve, jo divu stundu novērojums atbilst aptuveni 340 gigabaitiem atmiņas, kamēr dat tipa failu novērojumi aizņem tikai ap 70 megabaitiem.

Lai izvērtētu datu apstrādes veikspēju, tiek veikti veikspējas testi, kuri aprakstīti 3. nodaļā, kur apskatīts gan pirmās iterācijas, gan pēdējās iterācijas algoritmu izpildes laiks, kā arī procesu daudzuma ietekme uz pēdējās iterācijas algoritma izpildes laiku. Algoritma realizāciju var apskatīt pielikumā D.

1.3. Kalibrācijas process

Nodaļā aprakstītas kalibrācijas procesā izmantotās formulas, taču ņemot vērā, ka Bakalaura darba mērķis ir orientēts uz algoritmisko izpildi, nevis fizikālo pamatojumu, formulu izmantošanas pamatojums netiek sniegts. Aprakstītā kalibrācijas metode tiek izmantota spēcīgu māzeru novērojumos, līdz ar to, pamatalgoritms ir ticis izveidots jau iepriekš, ņemot par pamatu citās radio stacijās izmantotos kalibrēšanas algoritmus.

Kalibrācijas process ir izveidots, izmantojot rakstu *Unbiased flux calibration methods for spectral-line radio observations* [11], kurā aprakstītas metodes efektīvai datu kalibrācijai - pozīciju maiņa un frekvenču maiņa. Pozīciju maiņā teleskops tiek norādīts uz references pozīciju pirms tas tiek pozicionēts uz novērojumu objektu. Frekvenču maiņas metodē, tiek izmantoti divi lokālie oscilatori, kur viens no tiem tiek izmantots, lai sniegtu atskautes spektru. Veicot novērojumus teleskopa stars netiek fiziski bīdīts prom no avota, līdz ar to, nav iespējams veikt pozīciju maiņas pieeju, tāpēc tiek izmantota frekvenču maiņas metode.

Frekvenču maiņas metodes rezultātā veidojas 4 dažādas fāzes - P_{off}^{sig} , P_{on}^{sig} , P_{off}^{ref} , P_{on}^{ref} . Šie apzīmējumi tiek izmantoti formulās (1.2.1), (1.2.2), (1.2.3), (1.2.4), (1.2.5) un (1.2.6), kur:

P_{off}^{sig} = Lokālā oscilatora ierakstītais signāls ar izslēgtu trokšņa diodi

P_{on}^{sig} = Lokālā oscilatora ierakstītais signāls ar ieslēgtu trokšņa diodi

P_{off}^{ref} = References lokālā oscilatora ierakstītais signāls ar izslēgtu trokšņa diodi

P_{on}^{ref} = References lokālā oscilatora ierakstītais signāls ar ieslēgtu trokšņa diodi

Lai iegūtu datus par objekta plūsmas blīvumu, ir jāņem vērā kopējā sistēmas trokšņa temperatūra pēc formulas 1.2.1.

$$T_{sys}^{[cal]} = T_{bg} + T_{atm} + T_{atm} + T_{spill} + T_{loss} + T_{rx} + T_{cal} \quad (1.2.1)$$

kur:

$T_{sys}^{[cal]}$ = Kopējā sistēmas trokšņa temperatūra

T_{bg} = Mikrovilņu un galaktiskā fona troksnis

T_{atm} = Atmosfēras troksnis

T_{spill} = Zemes virsmas troksnis

T_{loss} = Novērojumā zaudētā informācija

T_{rx} = Uztvērēja trokšņa temperatūra

T_{cal} = Injicētais troksnis, izmantojot trokšņa diodi

Taču visas minētās vērtības iegūt nav vienkārši, līdz ar to tiek piedāvāts alternatīvs veids, kā aprēķināt sistēmas temperatūru no iegūtajiem datiem. Ja T_{sys}^{off} un T_{cal} tiek uzskatītas par konstantēm attiecībā pret laiku un frekvenci, iegūst vienādojumus:

$$T_{sys,ref}^{off} = T_{cal} \frac{(P_{on}^{ref} + P_{off}^{ref}) - \langle P_{on}^{ref} - P_{off}^{ref} \rangle_v}{2\langle P_{on}^{ref} - P_{off}^{ref} \rangle} \quad (1.2.2)$$

kur:

$\langle P_{on}^{ref} - P_{off}^{ref} \rangle_v$ = References oscilatora spektru starpības iekšējo 50 % vidējā vērtība

T_{cal} = Injicētais troksnis, izmantojot trokšņa diodi

$T_{sys,ref}^{off}$ = Sistēmas temperatūra izslēgtai trokšņa diodes references oscilatora signālam

$$T_{sys,sig}^{off} = T_{cal} \frac{(P_{on}^{sig} + P_{off}^{sig}) - \langle P_{on}^{sig} - P_{off}^{sig} \rangle}{2\langle P_{on}^{sig} - P_{off}^{sig} \rangle} \quad (1.2.3)$$

$\langle P_{on}^{sig} - P_{off}^{sig} \rangle_v$ = Lokālā oscilatora spektru starpības iekšējo 50 % vidējā vērtība

T_{cal} = Injicētais troksnis izmantojot trokšņa diodi

$T_{sys,sig}^{off}$ = Sistēmas temperatūra izslēgtai trokšņa diodes lokālā oscilatora signālam

Izmantojot skalāru T_{cal} vērtību, zināmu katram uztvērējam, var iegūt kalibrētu spektru izslēgtai trokšņa diodei:

$$T_{cal,ref}^{off} = T_{sys,ref}^{off} \cdot \frac{(P_{off}^{sig} - P_{off}^{ref})}{P_{off}^{ref}} \quad (1.2.4)$$

kur:

$T_{sys,ref}^{off}$ = Sistēmas temperatūra izslēgtai trokšņa diodes references oscilatora signālam

$T_{cal,ref}^{off}$ = Kalibrētas sistēmas temperatūra izslēgtai trokšņa diodes references oscilatora signālam

$$T_{cal,sig}^{off} = T_{sys,sig}^{off} \cdot \frac{(P_{off}^{ref} - P_{off}^{sig})}{P_{off}^{sig}} \quad (1.2.5)$$

kur:

$T_{sys,sig}^{off}$ = Sistēmas temperatūra izslēgtai trokšņa diodes lokālā oscilatora signālam
 $T_{cal,sig}^{off}$ = Kalibrētas sistēmas temperatūra izslēgtai trokšņa diodes lokālā oscilatora signālam

Līdz ar to, lai aprēķinātu sistēmas temperatūru ieslēgtai trokšņa diodei, papildus jāapskata trokšņa diodes temperatūra.

$$T_{cal,ref}^{on} = \left(T_{sys,ref}^{off} + T_{cal} \right) \cdot \frac{(P_{on}^{sig} - P_{on}^{ref})}{P_{on}^{ref}} \quad (1.2.6)$$

$$T_{cal,sig}^{on} = \left(T_{sys,sig}^{off} + T_{cal} \right) \cdot \frac{(P_{on}^{ref} - P_{on}^{sig})}{P_{on}^{sig}} \quad (1.2.7)$$

kur:

$T_{cal,ref}^{on}$ = Rezultējošā sistēmas temperatūra lokālā oscilatora references signālam ieslēgtai trokšņa diodei

$T_{cal,sig}^{on}$ = Rezultējošā sistēmas temperatūra lokālā oscilatora signālam ieslēgtai trokšņa diodei

$T_{sys,ref}^{off}$ = Kalibrētas sistēmas temperatūra izslēgtai trokšņa diodes lokālā oscilatora signālam izslēgtai trokšņa diodei

$T_{sys,sig}^{off}$ = Kalibrētas sistēmas temperatūra izslēgtai trokšņa diodes lokālā oscilatora references signālam

T_{cal} = Injicētais troksnis, izmantojot trokšņa diodi

Spektri ir apvienoti, lai gala rezultātā samazinātu troksni:

$$T_{cal}^{sig} = \frac{T_{cal,sig}^{on} + T_{cal,sig}^{off}}{2} \quad (1.2.8)$$

$$T_{cal}^{ref} = \frac{T_{cal,ref}^{on} + T_{cal,ref}^{off}}{2} \quad (1.2.9)$$

Lai iegūtu optimālu rezultātu T_{rez} vērtības, ir arī jāveic frekvenču pārvirzīšana, bet Bakalaura darba ietvaros tiek izmantota *numpy roll* funkcija, kurai tiek padots T_{rez} spektrs, kā arī pārbīdes koeficients. Rezultējošie spektri tiek apvienoti:

$$T_{rez} = \frac{T_{cal}^{sig} + T_{cal}^{ref}}{2} \quad (1.2.10)$$

Rezultāti tiek pārveidoti no kelviniem uz janskijiem, lai efektīvāk attēlotu plūsmas blīvumu. Jāņem vērā Zemes atrašanās vieta attiecībā pret objektu, līdz ar to tiek rēķinātas vērtības polinoma vērtības Zemes atrašanās vietā. Tas tiek darīts pēc formulas:

$$Sf = \frac{T_{rez}}{DPFU \cdot f(Gel, El)} \quad (1.2.11)$$

kur:

Sf	= Rezultējošais spektrālās plūsmas blīvums
T_{rez}	= Rezultējošā sistēmas temperatūra
$DPFU$	= (Degree per flux unit) Grādi uz plūsmas vienību
$f(Gel, El)$	= Aprēķinātā elevācijas polinoma vērtība

Nodaļā aprakstītās formulas tiek pielietotas divreiz, katrai polarizācijai. Kalibrācijas rezultātā tiek iegūts plūsmas blīvuma spektrs, kurš attēlo enerģiju, kas iegūta attālumā no novērotā objekta laukuma, kurš atrodas novērojuma brīdī veiktajā attālumā starp radioteleskopu un novēroto objektu. Algoritma implementāciju var apskatīt pielikumā D.

1.4. Doplera kompensācija un objekta pārvietošanās ātruma aprēķins

Nākamais solis vāju radioastronomisku objektu novērojumu procesā ir vairāku novērojumu apvienošana. Ņemot vērā, ka katrā novērojuma iterācijā stohastiskais troksnis dod signālam dažādas vērtības, apvienojot vairākus novērojumus, stohastiskais troksnis samazināsies ar katru iterāciju, bet novērojuma objekta vērtības nemainīsies. Lai to realizētu, ir jāņem vērā faktors, ka gan novērotais objekts, gan Zeme pārvietojas Saules sistēmā. Tas nozīmē, ka rezultējošo novērojuma frekvenci ietekmēs Doplera efekts un iegūtais frekvenču masīvs ir jāmodificē, lai pārbīde tiktu ņemta vērā.

Doplera efekta aprēķins ir atšķirīgs, atkarībā no objekta atrašanās vietas. Ņemot vērā to, ka Bakalaura darba ietvaros, tika apskatīti gan Saules sistēmas objekti, gan maiņzvaigņes un māzeri ārpus Saules sistēmas, tika izpētītas atšķirīgās pieejas nobīdes aprēķinam. Lai aprēķinātu Doplera efektu objektiem ārpus Saules sistēmas, ir nepieciešams zināt objektu pārvietošanās ātrumu. Tā kā novērotos radioastronomiskos objektus ārpus Saules sistēmas var uzskatīt par punktveida objektiem, visa ātruma vektoru ietekme tiek aprēķināta ar formulu:

$$V_{total} = (V_{sun} + V_{obs} + V_{Earth}) \quad (1.3.1)$$

kur:

V_{total} = Rezultējošais objekta ātrums

V_{sun} = Vidējais Saules pārvietošanās ātrums novērojuma laikā

V_{obs} = Vidējais objekta ātrums novērojuma laikā

V_{Earth} = Vidējais Zemes pārvietošanās ātrums novērojuma laikā

Lai iegūtu rezultējošās objekta pārvietošanās ātruma vērtības, tika izmantots jau eksistējošs programrisinājums, kas tika radīts radioastronomisku māzeru novērojumiem.[4] Metode izmanto māzera spektrālo līniju pīķus, lai iegūtu rezultējošo ātrumu pēc formulas:

$$V_{rez} = \left(\frac{f_{obs}}{f_{rest}} - 1 \right) \cdot c + V_{total} \quad (1.3.2)$$

kur:

V_{rez} = Rezultējošais ātruma masīvs, attiecīgajai frekvencei

f_{obs} = Novērojumā iegūtā frekvenču josla

f_{rest} = OH molekulu spektrālo līniju frekvence

c = Gaismas ātrums

V_{total} = Objekta rezultējošais pārvietošanās ātrums (skat. (1.3.1) formulu)

Ķermeņus Saules sistēmā nevar uzskatīt par punktveida objektiem, jo ātruma vektoru ietekme uz objektu ir lielāka. Lai iegūtu radioastronomiskā objekta ātrumu, ir iespējams izmantot Keplera orbitālos parametrus [10], taču lai atvieglotu darbu, kā arī mazinātu kļūdu iespējamību, tiek izmantota NASA HORIZONS sistēma [8], kura uzglabā lielu daudzumu informācijas par Saules sistēmas objektiem, to skaitā, objektu atrašanās vietu konkrētos laikos, kā arī objektu ātrumu konkrētajā laikā. Iegūtais ātrums novērojuma laika periodā tiek ievietots (1.3.2) formulas V_{total} mainīgajā.

Novērojumu veikšanas laikā, Doplera pārbīde jau tiek ņemta vērā, izmantojot aprakstīto metodiku, taču esošais algoritms paredzēts mazāka joslas platuma novērojumos, līdz ar to tiek izmantots vidējais ātrums visā novērojumā. Ņemot vērā minēto faktoru, spektrs tiek pārbīdīts atpakaļ par nobīdīto vērtību, sadalīts uz pusēm, kur pirmā puse reprezentē 1665MHz frekvences diapazona spektru, bet otra puse - 1667MHz frekvences diapazona

spektru, katrai daļai pārrēķinātas ātruma vērtības un attiecīgi pārbīdītas. Pareizā pārbīde tiek rēķināta pēc formulas:

$$df = f_0 - \frac{f_0}{1 - \frac{V_{object}}{c}} \quad (1.3.3)$$

kur:

- df = Aprēķinātā spektra nobīde
- f_0 = Novērojumā iegūtā frekvenču josla
- c = Gaismas ātrums
- V_{object} = Objekta radiālais ātrums relatīvs uztvērēja antenai

1.5. Veivletu izmantošana radioastronomisku signālu apstrādē

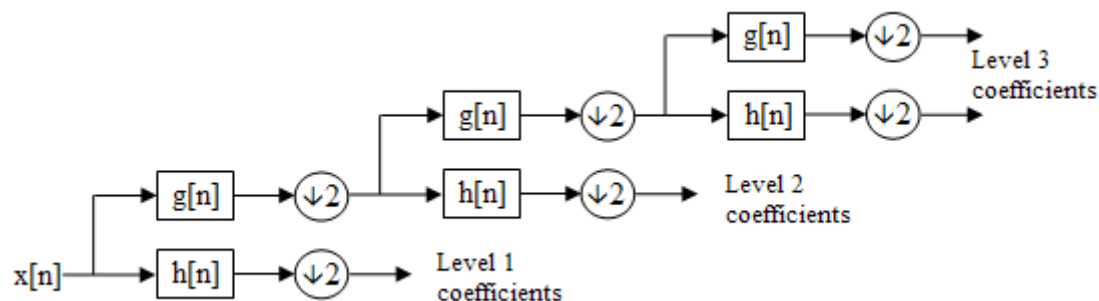
1.5.1. Veivletu definēšana

Lai precīzāk analizētu signālus ar pēkšņām pārmaiņām, ir nepieciešams izmantot ne-tradicionālākas metodes par Furjē transformāciju, kas reprezentē signālu kā sinusoīdu summu, kas nav lokalizēta laikā vai telpā. Lai veiktu signālu apstrādi, tika izmantota jauna funkciju klase – veivleti.

Veivlets ir viļņveida svārstība ar amplitūdu, kas sākas ar nulli, palielinās vai samazinās un gala rezultātā atgriežas nulles vērtībā. Veivletiem ir dažādi tipi un grupējumi, un tos arī iedala pēc īpašībām, taču tiem ir vienots nosacījums - veivleta amplitūdas vidējā vērtība ir nulle. Veivletu plašais formu klāsts ir viens no galvenajiem iemesliem veivletu izmantošanai signālu apstrādē. Lai vieglāk grupētu veivletus, tie ir iedalīti vairākās ģimenēs, taču konkrētie veivleti ir tikai visbiežāk izmantotie – veivletus var veidot jebkurš, ja vien izveidotais atbilst kritērijiem.

Viena no veivletu transformācijas formām ir diskrētā veivletu transformācija, kura visbiežāk tiek izmantota datu trokšņa samazināšanā un datu kompresijā. Process ir līdzīgs signāla salīdzināšanai ar filtru bankām. Diskrētā veivletu transformācija attēlota attēlā 1.5., kur modelēts apstrādes process.

Signāls tiek filtrēts, izmantojot *low-pass* un *high-pass* filtrus, lai iegūtu *low-pass* jeb aproksimācijas koeficientu, un *high-pass* jeb detaļu koeficientu apakšjoslas. Nākamajā



Att. 1.5.: Diskrētās veivletu transformācijas vizuālais attēlojums
[17]

dekompozīcijas līmenī, *low-pass* apakšjosla, tiek iteratīvi filtrēta, izmantojot to pašu metodi, kuru sākuma signāls. Koeficientu daudzums katrā apakšjoslā ir puse no koeficientiem no iepriekšējās apakšjoslas.

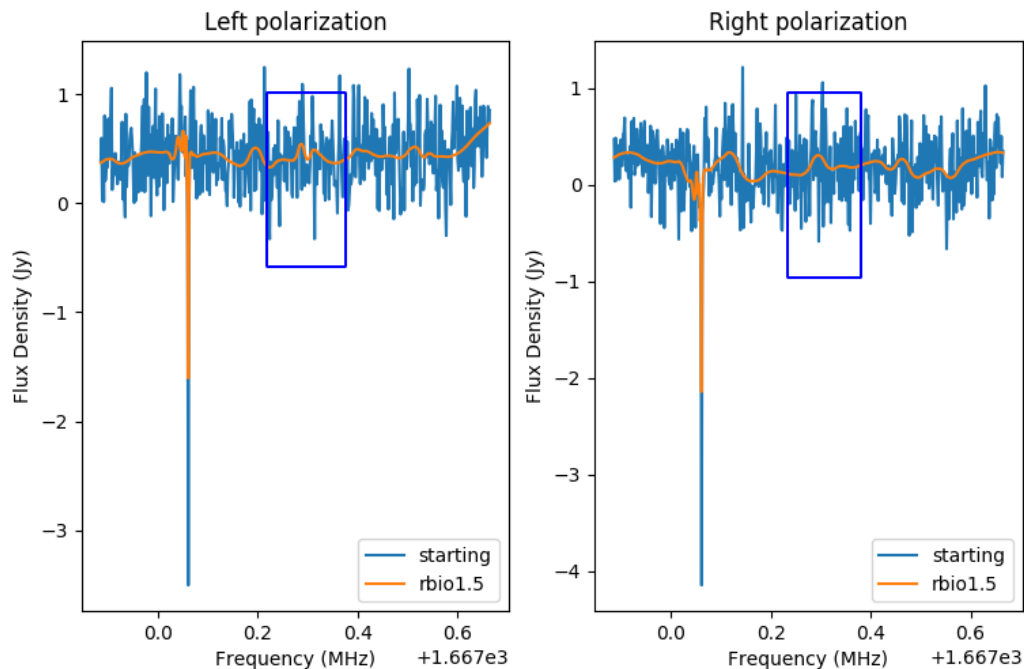
Lai gan veivletiem ir vairākas metodes, kā tos pielietot signāla trokšņa samazināšanai, Bakalaura darba ietvaros tika izmantota veivletu sliekšņošna. Ir apskatītas divas veida sliekšņošanas operācijas – mīkstā un cietā. Abos gadījumos koeficientiem ar vērtībām zem noteiktā sliekšņa tiek piešķirta vērtība nulle, bet atšķiras, kas notiek ar pārējiem koeficientiem. Mīkstajā metodē koeficienti, kas lielāki par norādīto sliekšni ir samazināti, atņemot sliekšņa vērtību no koeficienta vērtības, taču cietajā – koeficienti paliek nemainīgi. Bakalaura darba ietvaros pēc noklusējuma tiek izmantota mīkstā sliekšņošanas metode, jo metodes pielietojums bija piemērotākā uzdevumu veikšanai.

1.5.2. Veivletu pielietojums

Veivletu transformācija Bakalaura darba ietvaros tika realizēta izmantojot *PyWavelets* [18] Python bibliotēku. Bibliotēka tika izvēlēta, tās atvieglotās veivletu transformācijas realizācijas dēļ, kā arī plašo iebūvēto veivletu dēļ. Pēc noklusējuma tiek atbalstīti 106 veivleti, kuri atbalsta diskrēto veivletu transformāciju no sekojošām ģimenēm - bior, coif, db, dmey, haar, rbio, sym.

Plašo konfigurēšanas iespēju dēļ, veivletu pielietošana tika testēta gandrīz visos datu apstrādes procesos, ar atšķirīgu efektivitāti. Pielietojot šāda veida transformāciju, jāņem vērā, ka ar katru transformāciju, tiek zaudēti ar vien vairāk dati. Tas nozīmē, ka, ja transformācija tiks izmantota pārāk daudz, var tikt zaudētas spektra detaļas, kuras sākotnēji būtu redzamas. Arī jāņem vērā, ka iegūtais spektrs ir balstīts uz paša veivleta formu, līdz ar to, var rasties situācijas, kur veivlets izceļ pīķi konkrētā vietā, taču tas nav

meklētais signāls. Problēma tiek novērsta salīdzinot vairāku veivletu iegūtos rezultātus savā starpā, jo, ja vairākas formas veivletu spektri iegūst pīķi vienādās frekvences vērtībās, var uzskatīt, ka tā nav sagādīšanās.



Att. 1.6.: Pirmie iegūtie rezultāti uz R Leonis Minoris izmantojot rbio1.5 veivletu ar 0.5 sliekšni

Veivletu transformācijas pielietojuma iespējas ir ļoti plašas augstās konfigurēšanas iespējas dēļ, jo ir iespējams izmantot visus 106 iebūvētos veivletus un tiem norādīt savu precizitātes koeficientu, kā arī sliekšņošanas tipu. Tas nozīmē, ka katrs veivlets var tikt pielietots datiem, bet atkarībā no pielietotā veivleta rezultāts būs labāks vai sliktāks. Bakalaura darba pētījuma laikā tika atrasti vairāki veivleti, kurus iespējams efektīvi izmantot visos apskatītajos gadījumos.

Veivletu pielietošana Bakalaura darba ietvaros aizsākās ar mērķi novērst stohastisko troksni pirms vairāku novērojumu apvienošanas. Izmantojot veivletus, bija iespējams detektēt pirmās pazīmes maiņzvaigznes R Leonis Minoris OH māzera novērošanā, kuras attēlotas attēlā 1.6.. Grafikā attēlota plūsmas blīvuma attiecība pret frekvenci, kur, izmantojot sākotnējo signālu, ir izveidots jauns spektrs izmantojot rbio1.5 veivletu ar 0.5 sliekšni. Zilajā četrstūrī ir iezīmēts potenciāli novērotais OH māzera plūsmas blīvums, kurš, lai gan, nevar būt uzskatīts par detektētu signālu balstoties uz 3 sigmu likuma, sniedz informāciju, ka novērojot objektu ilgāku laiku būs iespējams iegūt detektējamu signālu.

Lai gan apstrādes metodika ir mainījies, darbs ar veivletiem ir atvēris vairākas iespējas datu apstrādes patreizējā metodikā gan 1.7. nodaļā aprakstītajā anomāliju detektēšanā, gan stohastiskā trokšņa mazināšanu komētu datu apstrādes rezultātiem.

1.6. MPI integrācija kalibrācijas procesā

Balstoties uz to, ka jēldatu apstrādes process ir laukietilpīgs, kā arī resursietilpīgs process, to ir nepieciešams efektīvi paralelizēt. Veids, kā tas tiek darīts Bakalaura darba datu apstrādes sistēmā, ir, izmantojot *Python* bibliotēkas *mpi4py* [9] iespējas, kura realizē *Message Passing Interface* jeb *MPI* protokolu python valodā, kas atļauj *Python* procesam izmantot vairākus kodolus vienlaicīgi.

Lai testētu *MPI* integrāciju kalibrācijas procesā, tika izveidots programrisinājums, kurš paredzēts izmantošanai ar deviņiem *MPI* procesiem, kur process nr. 1 atbild par failu apstrādes organizēšanu un masīvu apstrādi, bet procesi 2-9 veic jēldatu nolasīšanu, kura aprakstīta nodaļā 1.2. Izpildes procesa piemērs attēlots pielikumā B, kur pa soļiem attēlota datu apstrādes loģika un informācijas plūsma. Procesi attēloti, kā apli, kuru centrā norādīts fails, kurš patreiz tiek apstrādāts.

Nemot vērā to, ka sākotnējā *MPI* algoritma versija bija spējīga izmantot tikai 9 procesus datu apstrādei, bet pieejami bija daudz vairāk skaitļošanas resursu, algoritms tika pilnībā pārveidots. Algoritma versija Bakalaura nodošanas laikā spējīga datus apstrādāt neierobežotā kodolu skaitā, pilnvērtīgi izmantojot visus iedalītos kodolus.

Algoritma process nr. 1 uzlabotajā algoritmā atbild par kalibrēšanu un datu apvienošanu. Process nr. 2 - par pārējo procesu pārvaldību un failu nosaukumu nodošanu pārējiem procesiem. Procesam ar numuru 3 ir iedalīts speciāls uzdevums - nokalibrēt dat failus, jo dat failu apstrādes process ir ļoti ātrs un lietotāja ērtības dēļ, visas darbības tiek koncentrētas vienotā algoritmā. Kad process nr. 3 ir nokalibrējis dat failus, tas tiek pievienots pārējiem, jēldatu nolasīšanai paredzētajiem, procesiem. Tas nozīmē, ka algoritmu ir nepieciešams startēt, izmantojot vismaz 3 procesus, lai gan algoritms paredzēts daudz vairāku procesu izmantošanai. Izpildes procesa piemērs attēlots pielikumā C, kur attēlota procesu komunikācija un datu apstrādes process. Kā piemērs, tiek apskatīts potenciālā komunikācija starp astoņiem procesiem, taču līdzīgs rezultāts ir iegūstams jebkurā procesu daudzumā. Jāņem vērā, ka katrā algoritma izpildes reizē komunikācija var atšķirties, jo nolasīšanas ātrums nav konstants. Procesu pielikumā C attēloti kā apli, kuru centrā norādīts procesa pieejamības stāvoklis.

Process nr. 2 ir algoritma sarežģītākā daļa, jo ir nepieciešams nodrošināt, ka visiem iedalītajiem procesiem vienmēr ir iedalīts savs fails un iedalītais process strādā pie faila

nolasīšanas, kā arī jānodrošina, ka visi dati ir apstrādāti. Tas tiek realizēts ar masīvu, kur katrs elements atbilst procesa statusam. Sākotnēji procesu vērtības inicializētas ar vērtību 1, izņemot pirmās 3, jo minētie procesi jau ir aizņemti. Tad process, ciklā visu failu sarakstam, sūta failu ceļu pirmajam procesam, kurš ieņēmis stāvokli "1". Ja faila ceļš nosūtīts, atbilstošais procesa statuss attiecīgi ir mainīts uz 0, lai attēlotu, ka tas ir aizņemts. Ja visi procesi ir aizņemti, process 2 gaida uz kādu no pārējiem procesiem, lai tas nosūtītu apstiprinošu ziņu, ka dati nolasīti un iedala attiecīgajam procesam jaunu failu apstrādei. Kad visi faili nolasīti, pārējiem procesiem tiek aizsūtīta ziņa par datu apstrādes beigām un procesi apkopo savus datus vienā datu struktūrā, kuru nosūta procesam 1 tālākai apstrādei.

Algoritmam tika veikti veikspējas testi, kuri aprakstīti 3. nodaļā. Tiek apskatītas atšķirības no vecā algoritma versijas veikspējas, kā arī fināla versijas algoritma veikspējas atkarība no kodolu daudzuma.

1.7. Anomāliju detektācija

Veicot novērojumus, pastāv risks, ka konkrēts skans ir ticis bojāts kādu inženiertehniskū problēmu dēļ un nav izmantojams datu apstrādē. Lai realizētu veiksmīgu datu apstrādes procesu, bojātos kalibrācijas rezultātus ir nepieciešams identificēt. Gadījumos, kur novērojuma skans ir bojāts, rezultējošā spektra vērtības ieņem vērtības, kas ir daudz lielākas, par pareizu skanu rezultātiem. Augstās amplitūdas vērtības tiek iegūtas augstas sistēmas temperatūras vērtības rezultātā, tāpēc identificējot pēkšņas augstas sistēmas temperatūras vērtības, ir iespējams atrast bojātos skanus ar pietiekami augstu precizitāti, lai pārējo skanu kopas rezultāts būtu izmantojams.

Bakalaura darba ietvaros, datu anomālijas tiek detektētas, izmantojot standarta vērtību jeb *z-score*. *Z-score* ir statistikas rādītājs, kas norāda vērtības attiecību pret vidējo. Kā vidējā vērtība, šajā gadījumā tiek uzskatīta izslēgtas trokšņa diodes sistēmas temperatūras vidējā vērtība pēc formulas:

$$\overline{T_{sys}} = \frac{\overline{T_{sys,sig}^{off,left}} + \overline{T_{sys,sig}^{off,right}} + \overline{T_{sys,ref}^{off,left}} + \overline{T_{sys,ref}^{off,right}}}{4} \quad (1.7.1)$$

kur:

$\overline{T_{sys,sig}^{off,left}}$ = Lokālā oscilatora sistēmas temperatūra ar izslēgtu trokšņa diodi kreisajai cirkulārai polarizācijai

$\overline{T_{sys,sig}^{off,right}}$ = Lokālā oscilatora sistēmas temperatūra ar izslēgtu trokšņa diodi labajai

cirkulārai polarizācijai

$\overline{T}_{sys,ref}^{off,left}$ = References lokālā oscilatora sistēmas temperatūra ar izslēgtu trokšņa diodi
kreisajai cirkulārai polarizācijai

$\overline{T}_{sys,ref}^{off,right}$ = References lokālā oscilatora sistēmas temperatūra ar izslēgtu trokšņa diodi
labajai cirkulārai polarizācijai

Lai iegūtu precīzākus rezultātus, iespējams izmantot katras sistēmas temperatūras spektra vidējo vērtību, kopumā rēķinot *z-score* katram sistēmas temperatūras masīvam, taču, pēc noklusējuma tiek izmantots vidējā izslēgtas trokšņa diodes sistēmas temperatūra, jo otra metode nav pietiekami daudz analizēta, lai sniegtu pārliecinošus rezultātus.

Izmantojot vidējo vērtību un *scipy stats* piedāvāto funkciju *z-score* aprēķināšanai, tiek atlasītas visas vērtības, kur standarta vērtība ir lielāka par 2 jeb divas standartnovirzes lielāka par vidējo vērtību. Detektētās anomālijas tiek ierakstītas kalibrācijas log failā, kopā ar skana numuru un skana ierakstīšanas laiku turpmākai analīzei, lai palīdzētu identificēt potenciālos problēmu cēloņus nākotnē.

Sākotnējais anomāliju detektēšanas algoritms izmantoja individuālā spektra vidējo vērtību, kuru salīdzināja ar visu novērojumu vidējo vērtību. Lai gan šī metode detektēja spēcīgi bojātos skanus, bija problēmas ar iegūto rezultātu precizitāti. Dažkārt algoritms klasificēja nebojātos datus kā bojātos, kas samazināja rezultātu integritāti, līdz ar to bija nepieciešams atrast alternatīvu, šajā gadījumā, *z-score*, kurš, lai gan strādā uz līdzīga principa, bija spējīgs iegūt precīzākus rezultātus.

Tika izvirzīta hipotēze, kurā bojāto skanu ignorēšanas vietā, varētu prognozēt sistēmas temperatūru konkrētajā skanā un aizvietot bojāto rezultātu ar pareģoto. Pareģot temperatūru ir iespējams ļoti daudzus veidos, taču Bakalaura darba ietvaros tika apskatītas divas metodes - vismazāko kvadrātu polinoma pielāgošana jeb *polyfit* un veivletu pielietošana spektra formas iegūšanai.

Izmantojot *polyfit* funkciju, tika pareģota taisne, kura izmanto pirmo piecu un pēdējo piecu skanu vidējās vērtības. Šāda veida prognozēšana ir bīstama, jo ir iespējams, ka bojātais skans būs kādā no minētajām 10 vērtībām, līdz ar to, pareģotais rezultāts būs bojāts. Lai gan varētu vairāk modificēt minēto metodi, balstoties uz pieredzi citos novērojumos, tika izteikts priekšlikums izmantot veivletus.

Izmantojot zemas precizitātes sliekšni veivletu transformācijā, tiek iegūts spektrs, kas vāji atbilst sākotnējam spektra vērtībām, taču saglabā sākotnējo spektra formu. Līdz ar to, skanos, kur tiek detektēta augstāka sistēmas temperatūra, atbilstošā pareģotā vērtība būs augstāka, bet skanos ar zemāku temperatūru, vērtība būs zemāka. Skanos, kuros

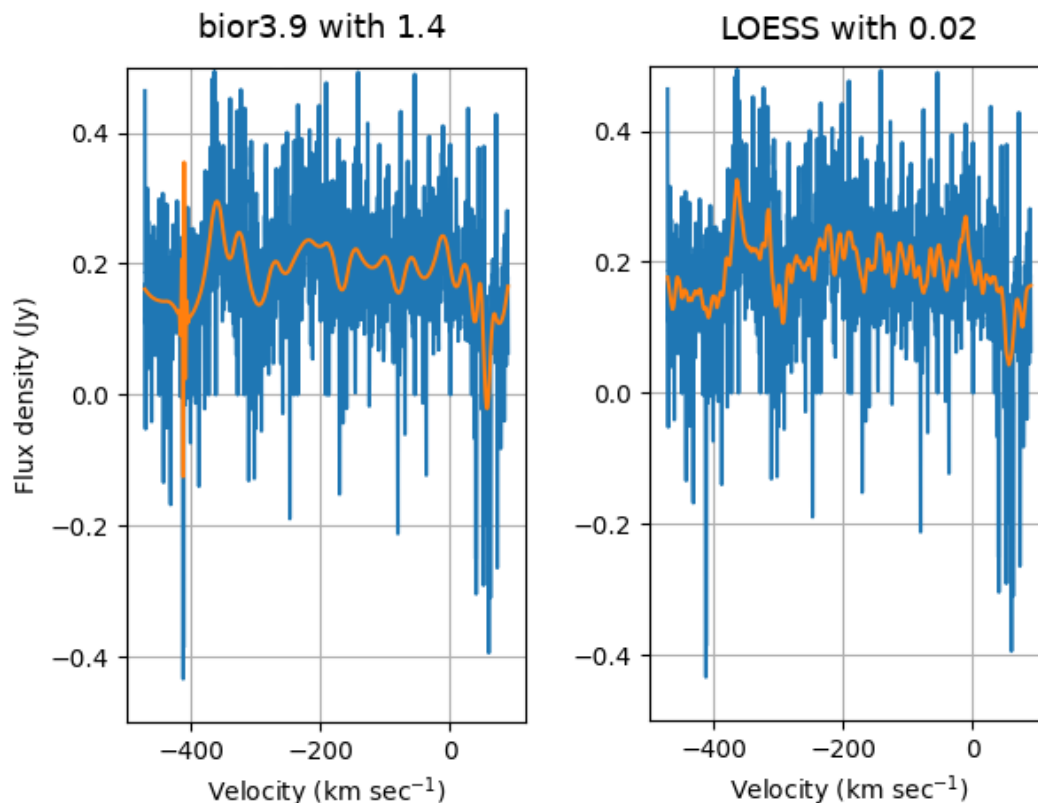
detektētas anomālijas, pateicoties veivleta zemajam precizitātes koeficientam, pareģotā vērtība sasniegs aptuveni maksimālo pieļaujamo vērtību. Nodaļā aprakstītā anomālijas detektācijas algoritma piemērs, kā arī iegūtais rezultāts, ir aprakstīts nodaļā 2.1.

1.8. Citas trokšņa samazināšanas metodes

Bakalaura darba ietvaros tika apskatītas vairākas trokšņa samazināšanas metodes, kuras Bakalaura darba sākuma posmā tika izmantotas, lai pārliecinātos par iegūto veivletu rezultātu integritāti. Metodes tika izmantotas uz apvienotajiem datiem, ar mērķi vieglāk atrast meklēto frekvenču pīķi un pārbaudīt veivletu iegūtos rezultātus pret citu trokšņu samazināšanas metožu rezultātiem.

Bakalaura darba ietvaros tika apskatītas sekojošās trokšņa samazināšanas metodes: Butterworth Low Pass [19], Savgol Filter [20] un Locally Weighted Scatterplot Smoothing (LOESS/LOWESS) [21]. Piemēra attēlā 1.7. apskatīts veivletu iegūtais rezultāts kreisajai cirkulārajai polarizācijai uz dažiem novērojumiem C/2017 T2 (PANSTARRS) komētai 1665 MHz joslā. Izmantojot abas metodes, var iegūt pīķi ap -5 km/s ātruma vērtības, kur arī pēc aprēķiniem varētu atrasties pīķis. Metodēs izmantots rbio3.9 veivlets ar 1.4 sliekšni izmantojot cietu sliekšņošanu. LOESS algoritmā, nepieciešams algoritmam sniegt daļu sākotnējā spektra datu, lai pareģotu katru y vērtību, līdz ar to algoritmam sniegts 2% datu.

Līdzīgi rezultāti iegūti pārējām aprakstītajām metodēm, taču Bakalaura darba apraksta ietvaros iegūtie rezultāti nav attēloti.



Att. 1.7.: Panstarrs C/2017 T2 datu kreisās cirkulārās polarizācijas metožu salīdzinājums

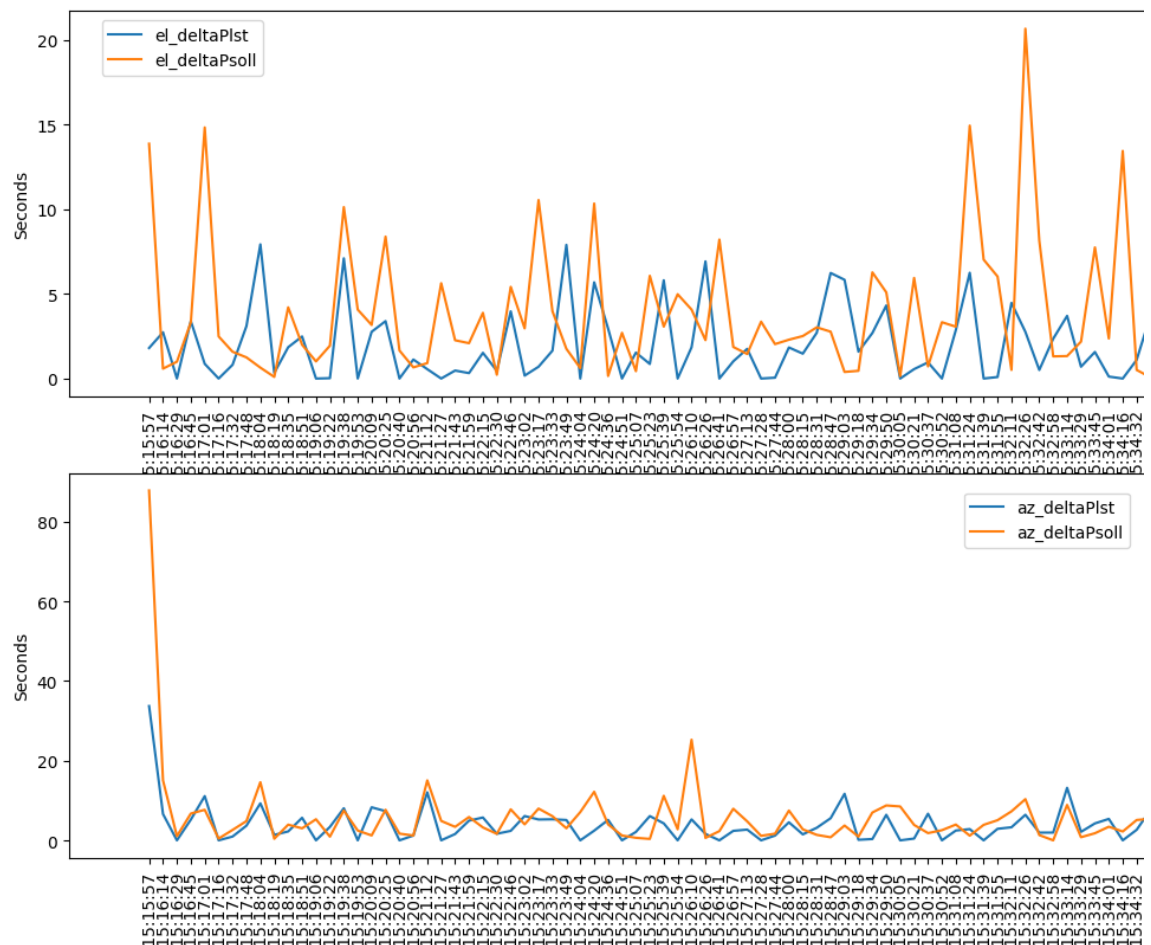
1.9. ACU žurnālēšanas faili

Bakalaura darba ietvaros, tika izveidotas vairākas kļūdu kontroles sistēmas, viena no tām - ACU žurnālēšanas failu pārbaudes. ACU jeb Antenna Control Unit ir sistēma, kura atrodas uz Fieldsystem aparatūras [22], kas nodrošina radioteleskopa pozicionēšanu pret novērojuma objektu un informāciju par vēlamu un faktisko teleskopa pozīciju (elevācijas un azimuta vērtības) apraksta žurnālēšanas failos jeb ACU log failos. Šāda veida dokumentēšana nodrošina iespēju zināt precīzu antenu stāvokli konkrētos laikos, līdz ar to, lai pārbaudītu datu integritāti, ACU log failu informācija tika salīdzināta ar novērojumu log failiem.

Plašais informācijas daudzums, kurš tiek ierakstīts ACU žurnālēšanas failos arī apgrūtinā vieglu informācijas nolasīšanu. Žurnālēšanas faila katra rinda sākas ar ierakstīšanas laiku, kuram seko informācija par ierakstu. Informācija ierakstīta *CSV* stilā, taču informācijai ir ap 200 kolonnām. Ne visas kolonnu vērtības ir vienā formātā, kas vēl vairāk

atrast kļūdas un dod iespēju informēt inženierus par laikiem, kuros anomālijas ir atrastas, lai novērstu problēmas nākotnē.

Grafika piemērs attēlots 1.9. attēlā, kur aprēķināts OH māzera g126p715 0p822 žurnālēšanas failā ierakstīto elevācijas un azimuta vērtību salīdzinājums ar ACU žurnālēšanas failā ierakstītajām. Rezultāts attēlots, kā sekunžu atšķirība starp abu žurnālēšanas failu elevācijas vērtībām, konkrētajos laikos. Kā *el_deltaPlst* un *az_deltaPlst* tiek attēlots elevācijas un azimuta tekošā pozīcija saskaņā ar enkoderi, bet ar *az_deltaPsoll* un *el_deltaPsoll* - uzdotā pozīcija.



Att. 1.9.: ACU žurnālēšanas faila salīdzinājuma rezultāts

1.10. Nākotnes plāni un datu apstrādes tālākie soļi

Vāju radioastronomisku signālu apstrāde ir sarežģīts process, līdz ar to ir dažādas pieejas katrai problēmai. To efektivitāti iespējams pārbaudīt tikai ar realizāciju, taču visas metodes nebija iespējams apskatīt Bakalaura darba ietvaros, līdz ar to, nodaļā aprakstīti potenciālie uzlabojumi un apstrādes tālākie soļi.

Potenciāls uzlabojums apstrādes rezultātos ir iespējams jau datu nolasīšanas brīdī, kur Furjē transformācijas procesu var aizvietot ar Karhunen—Loève Transform [23]. Izmantojot minēto apstrādes metodi, tehniski ir iespējams detektēt ļoti vājus signālus attiecībā pret troksni, jo metode izmanto matricu ar nejausi ģenerētiem skaitļiem un veic kompleksus lineārās algebras aprēķinus, lai iegūtu vislabākos koeficientus, kurus izmanto, lai iegūtu spektra frekvenču amplitūdu vērtības. Lai gan vēl nav izpētīts metodes pielietojuma iespējamība radioastronomiskajiem datiem, iespējams, ka, izmantojot minēto metodi, vāju objektu detektēšana būtu efektīvāka.

Lai nodrošinātu datu integritāti un iegūtu papildus informāciju par novērojamo objektu, tika apskatīti arī Stoka [24] parametri - vērtību kopa, kas apraksta polarizācijas radiācijas stāvokli. Balstoties uz rakstu par Nançay radioastronomijas observatorijas iegūtajiem rezultātiem 2007.-2009. gada intervālā [5], kurā tika aprakstīta informācija par laika periodā novērotajām maiņzvaigznēm, bija iespējams salīdzināt šos rezultātus ar Bakalaura darba laikā apskatītā objekta rezultātiem. Bakalaura darba ietvaros tika izpētīts *The Measurement of Polarization in Radio Astronomy* [25] raksts, lai iegūtu formulas cirkulārās polarizācijas Stokes parametru aprēķinam, taču iegūtos rezultātus nevarēja uzskatīt par pārliecinošiem, līdz ar to tie netiek aprakstīti Bakalaura darba ietvaros.

Stoka parametri sniedz informāciju arī par objekta magnētiskā lauka stāvokli, līdz ar to, papildus izmantojot iegūtos datus kopā ar optiskiem novērojumiem, iespējams veidot OH māzeru spilgtuma modeļus, kā arī veikt objekta monitoringu. Lai gan Bakalaura darbā sadarbība nav realizēta, nākotnē ir potenciāls veidot minēto sadarbību ar kādu optisko novērojumu observatoriju, piemēram, Baldones observatoriju.

Komētu novērojumiem arī ir potenciāls izmantot *Very-Long-Baseline Interferometry* metodes, kur signāls no viena astronomiska objekta tiek novērots ar vairākiem radioteleskopiem vienlaicīgi. Vairāku neatkarīgu staciju novērojumu veikšana vienlaicīgi samazina iespēju veikt inženiertehniskās kļūdas, kā arī sniedz daudz vairāk datu. Lai gan VSRC ir iesaistīts VLBI novērojumos, novērojumu veikšana ir process, kas prasa daudz staciju sadarbību, kā arī jaunu sistēmu ieviešanu, līdz ar to metodikas ieviešana prasītu ilgu laiku, taču varētu novest pie daudz labākiem rezultātiem.

2. DATU ANALĪZE UN IEGŪTIE REZULTĀTI

2.1. Novērojuma apstrāde

Novērojuma apstrādes procesu var sadalīt vairākos soļos, katrā solī iegūstot rezultātu. Nodaļā, kā piemērs, tiks apskatīts komētas ATLAS Y4/2019 novērojumu datu rezultāti no 15. marta novērojumu sesijas. Komētas novērojumi tika plānoti, balstoties uz 1.1. nodaļā aprakstītajiem plānošanas soļiem, bet iegūtie novērojuma dati tika apstrādāti balstoties uz 1. nodaļā aprakstītās metodikas.

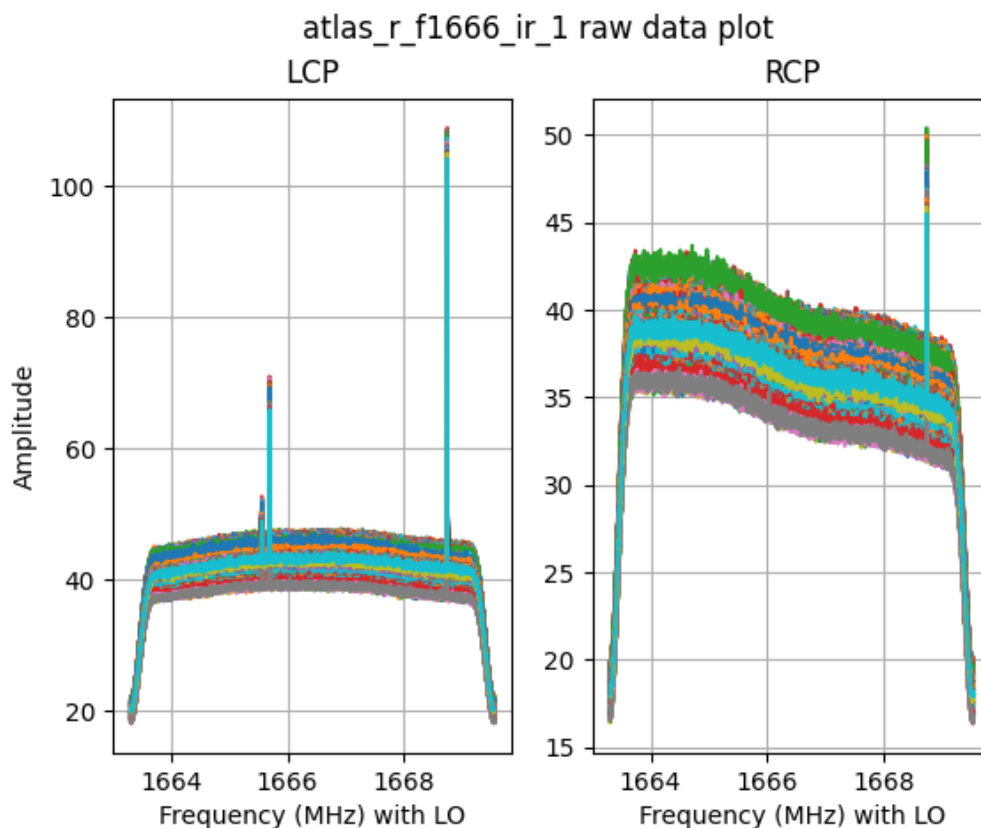
Tabula 2.1.: Nodaļā aprakstītā novērojuma informācija

Identifikators	atlas_r_f1666_ir_1
Sākuma laiks	2020-03-15T04:18:08
Beigu laiks	2020-03-15T06:34:16
Skanu daudzums	120
FFT garums	4096
Joslas platums	32 MHz
Kanāla platums	1.526 KHz

Konkrētais novērojums ir izvēlēts kā piemērs, jo ir iespējams attēlot visus 1. nodaļā aprakstītos datu apstrādes soļus. Novērojumam ir pieejami gan jēldati, gan dat formāta faili, kā arī dažos skanos ir detektējamās anomālijas. Novērojuma aktuālā informācija aprakstīta tabulā 2.1..

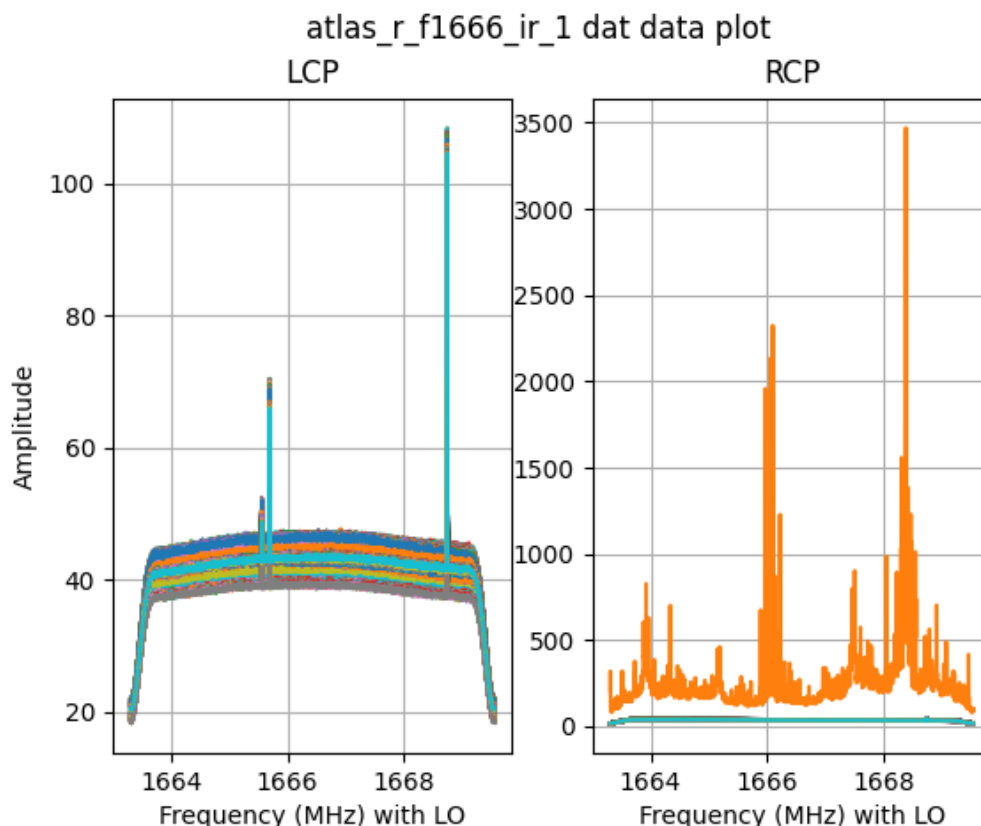
Pirmais iegūtais grafiku tips datu apstrādes procesā ir amplitūdu grafiks, kas ir ticis iegūts, balstoties uz 1.2. nodaļā aprakstītās datu nolasīšanas algoritma.

Nolasot visus datus, rezultāti tiek attēloti informatīviem mērķiem un minētās novērojumu sesijas gadījumā tiek iegūti divi grafiki, katrs savam datu tipam (skatīt attiecīgi attēlos 2.1., 2.2.). Papildus, rezultātus var izmantot, lai identificētu anomālijas jau pirmajā novērojumu apstrādes fāzē.



Att. 2.1.: Jēldatu nolasīšanā iegūtais rezultāts

Grafikos 2.1., 2.2. attēlotas abas cirkulārās polarizācijas, kā arī visu skanu rezultāti pēc datu nolasīšanas. Rezultātu attēlošanā uz y ass tiek norādīta amplitūda, bet uz x ass - frekvence. Rezultātu iegūšanas brīdī, aprēķini tiek veikti neizmantojot lokālo oscilatoru, taču šiem rezultātiem lokālā oscilatora vērtība tiek pieskaitīta, lai ērtāk orientētos frekvenču joslā. Lai gan novērojumu sesijā vienlaicīgi tiek iegūti abi datu formāti, teleskopu tehnisku grūtību dēļ, iegūtie rezultāti labajā cirkulārajā polarizācijā ir atšķirīgi. Individuālie grafiki apzīmē iegūto rezultātu katrā skanā, līdz ar to katrā polarizācijā ir iegūti 120 spektri.

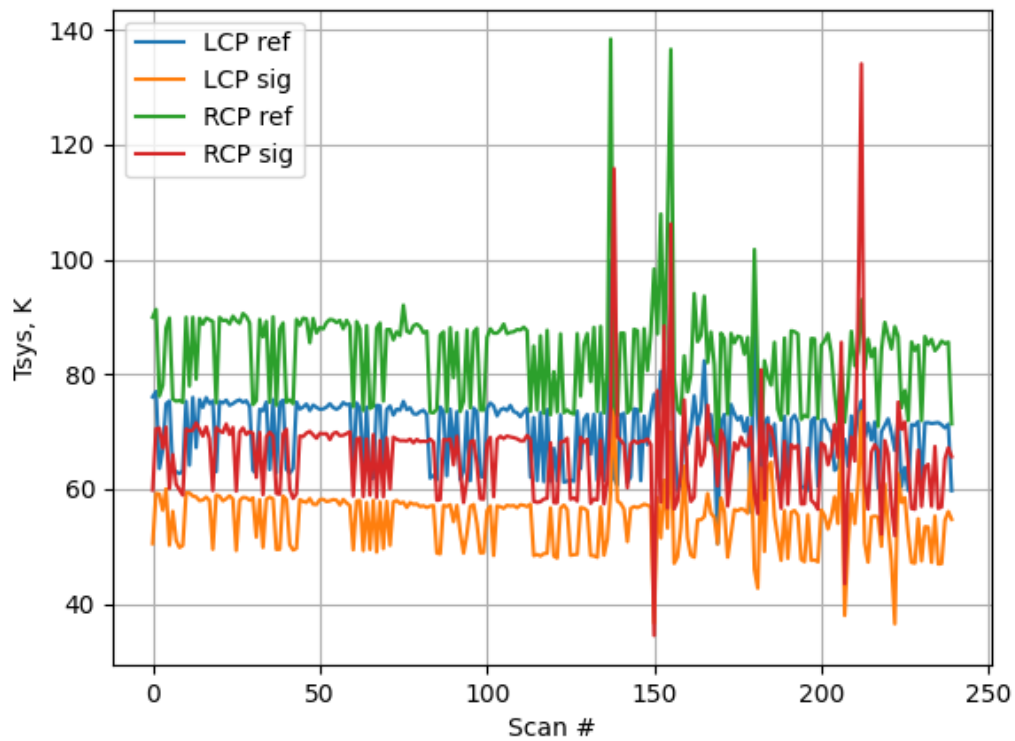


Att. 2.2.: Dat formāta datu nolasīšanā iegūtais rezultāts

Apstrādājot dat failus, iegūtas amplitūdas vērtības norāda, ka novērojuma laikā ir bijušas tehniskas problēmas, jo kreisajā cirkulārajā polarizācijā tiek iegūtas daudz augstākas vērtības noteiktos skanos, salīdzinot ar citiem skaniem. Minētās amplitūdas svārstības netika novērotas jēldatu apstrādes rezultātā, kur iegūtais spektrs nav bojāts, taču tas nenozīmē, ka šiem datiem nav nekādu problēmu. Izmantojot Furjē transformāciju, var iegūt amplitūdu pīkus frekvencēs, kurās avots staro. Pēc spektrālo līniju apraksta [26], OH spektrālajām līnijām ir konkrētas frekvences, kurās OH māzeri staro, līdz ar to, ideālā gadījumā, pīķi būtu novietoti 1665.402 MHz un 1667.359 MHz vērtībās. Aprakstītajā novērojumu sesijas gadījumā ir novērots arī kāds parazītisks signāls, kurš staro citās, bet tuvās frekvenču vērtībās, spēcīgāk par meklēto OH māzeri, kurās pīķis parādās un ietekmē rezultātus.

Lai tālāk analizētu iegūtos datus, ir nepieciešams apskatīt novērojuma sistēmas temperatūras vērtībās. Balstoties uz formulām 1.2.2 un 1.2.3, kuras aprakstītas 1.3. nodaļā, tiek iegūts 2.3. attēla grafiks. Bakalaura darba nodaļas ietvaros ir attēlots tikai jēldatu

iegūtais rezultāts, jo dat failu rezultāts ir ļoti līdzīgs. Grafikā attēlotas sistēmas temperatūra Kelvīnos katrā skana izpildē, kur *LCP ref* un *LCP sig* - kreisās cirkulārās polarizācijas references un objekta sistēmas temperatūras spektri, bet *RCP ref* un *RCP sig* - labās cirkulārās polarizācijas references un objekta sistēmas temperatūras spektri.

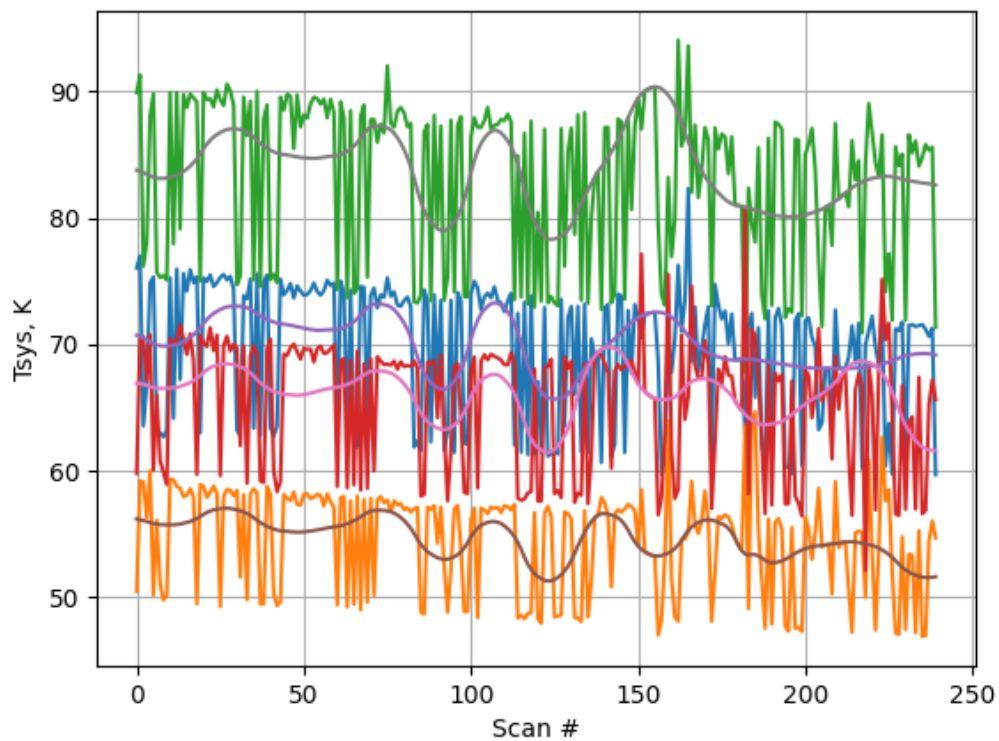


Att. 2.3.: Jēldatu formāta novērojumā aprēķinātā sistēmas temperatūra

Attēlojot sistēmas temperatūras vērtības grafiski, viegli ieraudzīt pēkšņas sistēmas temperatūras izmaiņas. Skani, kuros sistēmas temperatūra pārsniedz normu, spēcīgi negatīvi ietekmēs rezultējošo plūsmas blīvuma spektru, kurš tiek iegūts nākamajā datu apstrādes solī. Pēc 1.7. nodaļā aprakstītās metodikas, iespējams detektēt minētās izmaiņas algoritmiski. Sākotnēji bojātie skani netika ņemti vērā jēldatu apstrādē, tāpēc bija iespējams apstrādāt mazāku daudzumu datus. Līdz ar to, pareģojot bojāto sistēmas temperatūras vērtību, ir iespējams joprojām izmantot ierakstītos datus.

Izmantojot veivletu transformāciju, var iegūt spektra formu, kura vērtības iespējams izmantot bojātu sistēmas temperatūru gadījumos. Nodaļā aprakstītais rezultāts iegūts izmantojot *rbio1.5* veivletu ar sliekšni 3, kas arī ir noklusējuma konfigurācijā. Nodaļā

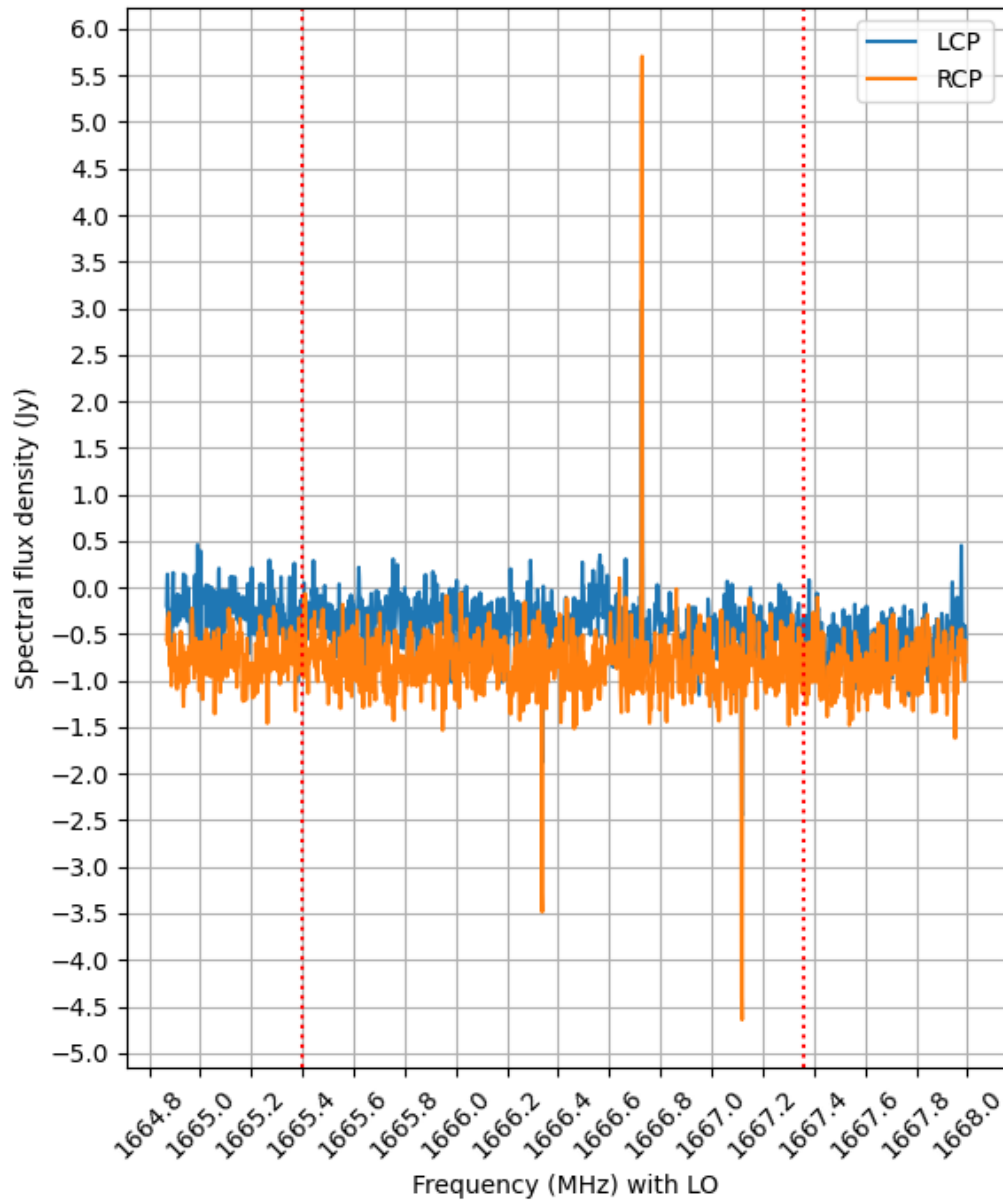
aprakstītā piemēra gadījumā bojātie dati ir detektēti skanos 137, 138, 150, 152, 153, 154, 155, 169, 179, 180, 181, 206, 207, 211, 212, 222. Neizmantojot aprakstīto metodi, bet saglabājot datu integritāti, būtu nepieciešams neapstrādāt 12 skanos jeb 10% no visiem novērojuma skaniem. Pēc anomāliju detektēšanas un vērtību aizvietošanas, tiek iegūts grafiks 2.4., kur papildus grafika 2.3. sistēmas temperatūras spektriem attēlots arī vievle-tu transformācijā iegūtie spektri katram sistēmas temperatūras grafikam. Visas iepriekš detektētās skanu sistēmas temperatūras vērtības ir aizvietotas ar vievletu transformācijas rezultējošo vērtību.



Att. 2.4.: Jēldatu aprēķinātā sistēmas temperatūra pēc anomāliju identificēšanas

Kad iegūtas derīgas sistēmas temperatūras vērtības, ir iespējams kalibrēt 2.1. un 2.2. grafikos iegūtos rezultātus pēc 1.3. nodaļā aprakstītās metodikas. Kalibrācijas procesā iegūtais rezultāts attēlots 2.5. grafikā, kur attēlots abu polarizāciju spektrālās plūsmas blīvums atkarībā no frekvences, izmantojot novērojuma jēldatus. Iegūtais rezultāts no dat tipa datiem ir ļoti līdzīgs, līdz ar to atskaitē nav apskatīts. Salīdzinot attēlos 2.1. un 2.2. iegūto frekvenču joslu, rezultējošā frekvenču josla ir mazāka, jo interesējošais

spektrs ir izgriezts no kopējā. Ar raustītām sarkanām līnijām ir iezīmētas OH māzeru spektrālās līnijas. Ideālā scenārijā, iezīmētajās frekvencēs iespējams iegūt pīķi, 3 reizes izteiktāku par pārējo troksni, bet viena novērojuma rezultātos diemžēl nav iespējams iegūt tik precīzas vērtības apskatītajam objektam.



Att. 2.5.: Spektrālās plūsmas blīvums atlas-r-f1666-ir-1 datiem.

Izmantojot iepriekš zināmo informāciju, iespējams aprēķināt trokšņa līmeni novērojumā pēc formulas

$$dS = \frac{SFED}{\sqrt{RBW \cdot t}} \quad (2.3.1)$$

kur:

$SFED$ = Teleskopa jūtība

t = Novērojuma integrācijas laiks

dS = Mērījuma troksnis

RBW = Kanāla platums

Pēc iekšējā pētījuma rezultātiem [1], pieņemot, ka teleskopa jūtība ir 700 Jy, kā arī pieņemot, pēc 3 sigmu likuma, ka, lai objektu detektētu, minimālajam objekta plūsmas jaudai jābūt vismaz 3 reizes spēcīgākam par mērījuma troksni, šajā novērojumu iterācijā iegūst aprēķinu:

$$Sf = 3 \cdot dS = 3 \cdot \frac{700(Jy)}{\sqrt{1526(Hz) \cdot 7200(s)}} \approx 3 \cdot 0.211(Jy) \approx 0.634(Jy) \quad (2.3.2)$$

kur:

Sf = Avota spilgtums

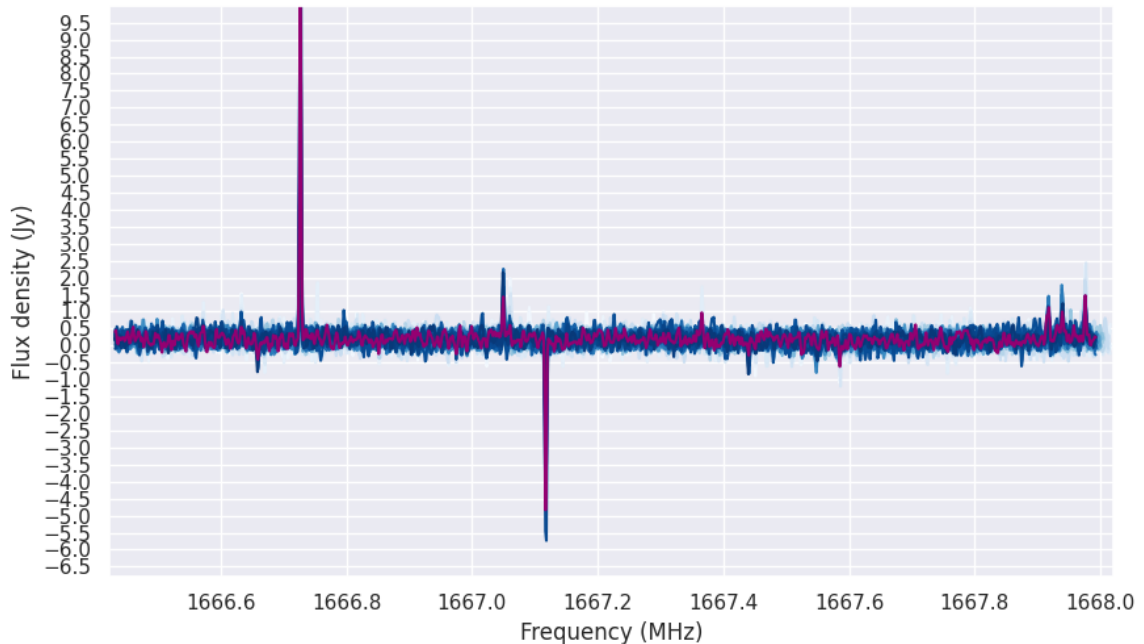
dS = Mērījuma troksnis pēc formulas 2.3.1

2.2. Iegūto rezultātu izvērtējums

Balstoties uz to, ka viena novērojuma rezultātā, ar patreizējām tehnoloģijām nevar detektēt signālu, pēc 1.4. nodaļā aprakstītās metodikas iespējams apvienot vairākus novērojumu rezultātus.

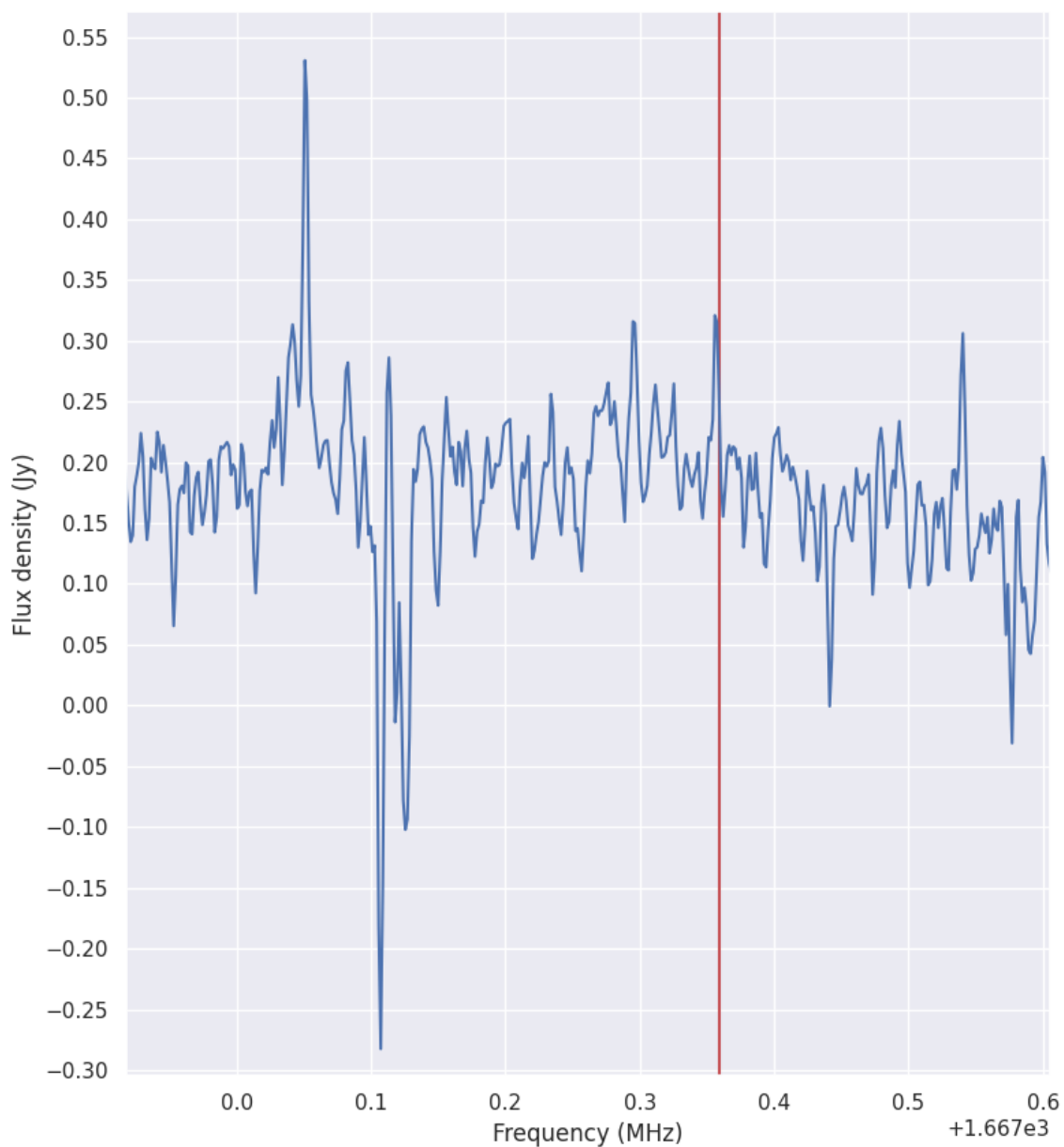
Apskatot iepriekšējā nodaļā novēroto objektu ATLAS Y4/2019 ilgākā laika periodā pēc teorijas iespējams mazināt kopējo trokšņa līmeni novērojumā. Ņemot vērā, ka objekta sadalīšanās veido rezultātu nestabilitāti, ar pielieto metodiku nav iespējams detektēt signālu. Lai iegūtu vairāk informācijas par objektu, ir nepieciešams apskatīt iegūtos datus noteiktos intervālos, kas Bakalaura darba ietvaros nav apskatīti. Lai iegūtu loģisko turpinājumu iepriekšējā nodaļā aprakstītajiem rezultātiem, tiek apskatīts 133 stundu laikā

iegūto datu apkopojums jeb 37 dažādu novērojumu rezultāts, atspoguļots 2.6. attēlā, kur ar gradienta palīdzību vecākie dati attēloti, izmantojot gaiši zilu krāsu, bet jaunākie - apzīmēti ar tumši zilu. Pats jaunākais novērojums atzīmēts atsevišķi ar violetu krāsu. Bakalaura darba atskaitē tiek apskatīta tikai kreisās polarizācijas, 1667 MHz daļas dati, taču iegūtais rezultāts ir pielīdzināms arī pārējām novēroto datu daļām.



Att. 2.6.: Visu individuālo novērojumu rezultāts ATLAS Y4/2019 komētas kreisajai cirkulārai polarizācijai.

Izmantojot attēlā 2.6. iegūtos rezultātus, nav iespējams detektēt signālu nevienā no individuālajiem novērojumiem, līdz ar to, apvienojot visus novērojumus, iegūst 2.7. attēlu, kur reprezentēts interesējošais apgabals kreisās cirkulārās polarizācijas 1667 MHz datu apgabalā, kā arī ar sarkanu līniju iezīmēta OH māzeru starošanas frekvence. Lai gan zemās pīķa maksimālā vērtības dēļ, datu apstrādes rezultāts nevar tikt uzskatīts kā neapstrīdams, ir iespējams, ka tiek detektēts signāls, taču Bakalaura darba ietvaros padziļināti apstrādes rezultāti netiek apskatīti.

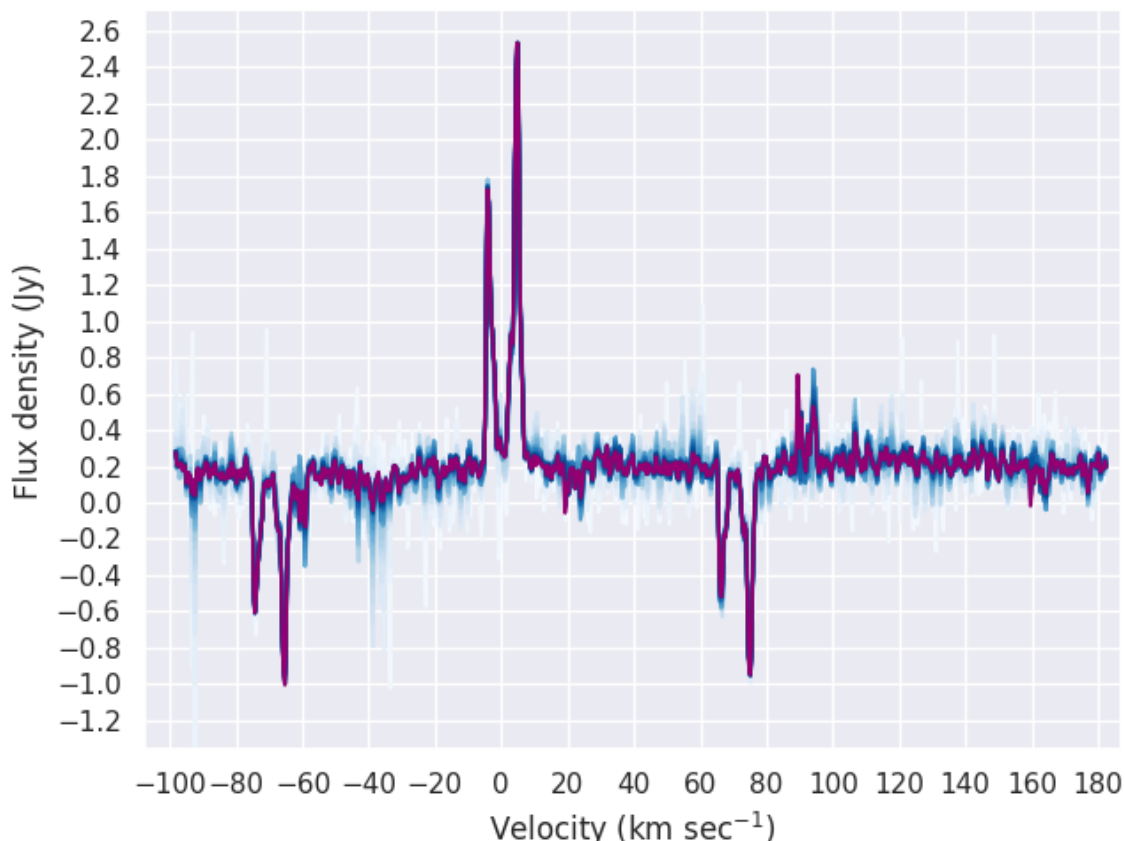


Att. 2.7.: Plūsmas blīvuma atkarība no objekta frekvences ATLAS Y4/2019 kreisās cirkulārās polarizācijas datiem.

Lai gan komētu novērošanā vēl ir daudz darba, Bakalaura darba ietvaros ir veikts liels progress citu vāju objektu novērošanā. Atsaucoties uz 1.5. nodaļā aprakstītajiem RLMI rezultātiem attēlā 1.6., izmantojot izveidoto metodiku ir iegūti daudz pārliecinošāki rezultāti. Rezultātus tieši nedrīkst salīdzināt, jo ir daudz ietekmējoši faktori, it īpaši novērojuma

ilgums un joslas platuma izmaiņas, taču iegūtie rezultāti liecina, ka objektu var detektēt novērojumu datos ar augstu ticības līmeni.

RLMi maiņzvaigznes patreizējā novērošanas iespējamība attēlota 2.8. attēlā, kur attēloti dati no 92 stundu novērojumiem. Lai attēlotu stohastiskā trokšņa samazināšanu, tiek izmantoti grafiki gradienta veidā, kur gaiši zilie spektri atbilst mazāk stundu apvienotajiem novērojumiem, bet tumši zilās atbilst vairāk stundu apvienotajiem novērojumiem. Galējais novērojumu rezultāts iekrāsots violetā krāsā.



Att. 2.8.: Plūsmas blīvuma atkarība no objekta ātruma maiņzvaigznes RLMi kreisās cirkulārās polarizācijas datiem.

Izmantojot 1.4. nodaļā aprakstīto metodiku, spektra frekvenču josla tiek pārrēķināta uz objekta ātruma vērtībām. Pēc datu apstrādes, rezultējošajā spektrā, ir iegūti pīķi -4 km/s un 5 km/s vērtībās, kas arī pēc aprēķiniem ir novērotā OH māzera pārvietošanās ātruma vērtības ar augstu trokšņa pret signāla attiecību.

2.3. Bakalaura darba izstrādes laikā novērotie kosmiskie objekti

Bakalaura darba ietvaros tika novēroti vairāki radioastronomiski objekti, kurus var iedalīt kategorijās - mainīgzvaigznes, komētas un zvaigžņu veidošanās apgabali.

Novērojot vājus objektus, lai nodrošinātu datu integritāti, nepieciešams kāds spēcīgāks avots, kuru vieglāk detektēt. Bakalaura darba novērojumu ietvaros tika izmantota mainīgzvaigzne R Leonis Minoris (turpmāk RLMi). Balsoties uz rakstu no Nancay radioastronomijas observatorijas Francijā, [5] tika salīdzinātas iegūtās vērtības ar Irbenē iegūtajām un tika iegūts līdzīgs rezultāts. Kopš tā brīža, RLMi tika novērots visās novērojumu sesijās, kur tiek novērota komēta Saules sistēmā.

Iegūstot rezultātus uz salīdzinoši spēcīgā avota RLMi, bija iespējams novērot arī komētas Saules sistēmā, kuras izstaro vāju OH māzera starojumu. Pirmā novērotā komēta, kura tika apskatīta Bakalaura darba ietvaros ir Panstarrs C/2017 T2 (turpmāk Panstarrs). Ņemot vērā, ka objekta orbīta nodrošina ilgstošu redzamību, komēta ir novērota visilgāko laiku. Lai gan komēta ir salīdzinoši vāja spilgtuma - vidēji ap 12 magnitūdām pēc NASA HORIZONS datiem novērojumu veikšanas laikā, ilga novērošanas laiks dod iespēju efektīvāk pārbaudīt ar vairāku novērojumu apvienošanu saistītu metodiku.

Asteroid Terrestrial-impact Last Alert System jeb ATLAS 2019. gada 28. decembrī atklāja jaunu komētu - ATLAS C/2019 Y4 (turpmāk Atlas). Gada beigās atklātajai komētai tika aprēķinātas augstas spilgtuma vērtības, kas sniedza iespēju daudziem zinātniekiem to novērot. Komētu klasificē kā Kreutz sungrazer tipa komētu, kas nozīmē, ka komētas perifērijas ir ļoti tuvu Saulei. Lai gan tas pozitīvi ietekmē komētas spilgtumu, tas palielina iespējamību komētām sadalīties, kas arī notika ar Atlas komētu ap 2020. gada 1. aprīli [27]. Komētas orbīta liecina, ka Atlas komēta ir bijusi daļa no Lielās komētas C/1844 Y1 [28], bet atdalījusies aptuveni piecu tūkstošu gadu atpakaļ. Komētai sadaloties, stipri palielinās komētas spilgtums, taču sadalīšanās brīdī komētai veidojas vairākas daļas, kur nereti katrai daļai ir savs kodols. Kodolā esošo OH molekulu daudzums vairs nav tik daudz, līdz ar to starojuma jauda samazinās, kas noved pie samazinātām komētas novērošanas iespējām.

Lai gan Atlas komētas sadalīšanās nozīmēja komētas novērošanas beigas, Saules sistēmā parādījās jauna spoža komēta - SWAN C/2020. Pirmo reizi atklāta, izmantojot Solar and Heliospheric Observatory jeb *SOHO*. Swan komēta tika atklāta, jo tā izvadīja kosmosā ļoti lielu daudzumu H_2O [29]. Komēta Bakalaura darba nodevuma laikā joprojām tiek novērota un darbs pie datu apstrādes turpinās.

Tabula 2.2.: Bakalaura darba izstrādes laikā ilgstoši novērotās komētas

Objekta nosaukums	Atlas C/2019 Y4	Panstarrs C/2017	Swan C/2020 F8
Tuvākā distance līdz Zemei	0.781 AU	1.5202 AU	0.557 AU
Tuvākā distance līdz Saulei	0.253 AU	1.615 AU	0.432 AU
Spilgtums perihēlija punktā	Sadalījies	8.6	7.6
Objekta novērojuma ilgums	133h 16m 47s	149h 21m 54s	110h 09m 57s
Novērojumu sākuma datums	2020-Mar-14	2020-Jan-26	2020-May-05
Pēdējais novērojumu datums	2020-Apr-11	2020-May-04	2020-May-20

Neskaitot primārās komētas, kuras aprakstītas 2.2. tabulā¹, tika arī novērotas sekojošās komētas: C/2018 N2 (ASASSN), C/2019 B1 (Africano), 284P/McNaught, taču minētie objekti bija pārāk vāji novērošanas laikā, vai bija tehniskas problēmas, kuras liedza veikt ilgstošus novērojumus.

Bakalaura darba izstrādes laikā, tika novērotas ne tikai komētas Saules sistēmā, bet arī radioastronomiski māzeri, kas staro spēcīgāk par komētām. Viens no tiem bija metanola māzeris Cepheus A zvaigžņu veidošanās apgabalā. Balstoties uz to, ka objekts tika novērots jau iepriekš, bija zināms, ka tā pīķis sasniedz 180 Jy vērtību, līdz ar to, objekts tika izmantots, lai testētu jēldatu apstrādes algoritmu uz cita veida datiem. Spēcīgu māzeru novērojumos parasti nav nepieciešams izmantot jēldatu formātu, jo dat faili sniedz pietiekami labu precizitāti, līdz ar to metodika, kā apstrādāt māzeru jēldatus nebija izveidota, taču šajā datu līmenī, starpība starp objektiem ir ļoti maza, līdz ar to, var izmantot to pašu apstrādes procesu. Māzeru apstrāde sniedza iespēju efektīvāk identificēt kļūdas, jo spektram bija iepriekš zināma forma, ar ko varēja salīdzināt iegūtos datus, kas nav iespējams komētu novērojumiem.

Papildus minētajiem novērojumiem, Bakalaura darba izstrādes laikā tika novērots vēl viens vājš OH māzeris - g126p715 0p822, kurš tika izmantots līdzīgiem nolūkiem, bet rezultātiem būtu jābūt līdzīgākiem komētu OH māzeru novērojumu rezultātiem. Māzera novērošanas laiks bija ļoti īss, kā arī novērojumu laikā bija tehniskas problēmas, kuras neļāva iegūt vēlamos rezultātus. Izmantojot datus no minētā māzera, kā arī dienu iepriekš novērotās C/2019 B1 (Africano) komētas, varēja identificēt, ka teleskopā ir problēmas, jo bija izteikts troksnis konkrētās frekvenču vērtībās.

1. Distances un spilgtuma dati no astro.vanbuitenen.nl

3. ALGORITMA IZVĒRTĒJUMS

3.1. Datu apstrādes tehniskā informācija

Nemot vērā, ka VSRC pārvaldībā ir nonācis jauns skaitļošanas klasteris, bija iespēja pielietot algoritmu uz jaunas paaudzes skaitļošanas klastera. Līdz ar to, bija iespējams veikt veikspējas testus uz lielāka resursu daudzuma, salīdzinot ar parasti pieejamo, un apskatīt algoritma veikspēju atkarībā no dažādu parametru konfigurācijas. Lai vairāk kontrolētu datu apstrādes vidi, tika izmantotas virtuālās mašīnas uz VSRC skaitļošanas klastera.

Tabula 3.1.: Veikspējas testos izmantoto individuālo virtuālo mašīnu apraksts

Procesora ligzdas	2
Procesora kodoli ligzdā	48
BIOS	SeaBIOS
Procesora tips	kvm64
Atmiņa	325.96 GB

Balstoties uz to, ka klastera mezgli izmanto *Proxmox* operētājsistēmu, kas piedāvā iebūvētas virtuālo mašīnu veidošanas tehnoloģijas, tika izveidotas trīs virtuālās mašīnas ar tabulā 3.1. aprakstītajiem parametriem. Virtuālajām mašīnām tika ielādēta *Ubuntu 19.10 Live Server* versija, kā arī papildus noklusējuma bibliotēkām, tika instalētas *openmpi* pakotnes, kā arī nepieciešamās bibliotēkas datu apstrādes skripta realizācijai. Skaitļošanas mezgli savā starpā saslēgti izmantojot 10GB/s optiskos kabeļus. Dati atrodas ārējā datu serverī, kas pieslēgts katrai virtuālajai mašīnai. Lai veiktu veikspējas testus, tiek izmantoti dati no 2.2. nodaļā aprakstītā novērojuma.

3.2. Veikspējas testi

Veicot veikspējas testus, ir jāņem vērā daudz faktori, līdz ar to ideālu veikspējas testu izpilde nav viegla. Faktori, kā ielasīšanas ātrums ir noteicoši konkrētā veida datu apstrādes algoritmos, līdz ar to ir limitācijas, cik ļoti ir iespējams uzlabot algoritma veikspēju. Izveidojot pilnībā simulētu vidi, zūd algoritma veikspējas uzticamība, salīdzinot ar veikspēju reālajā pasaulē.

Veikspējas testi ir veikti, izmantojot *Linux* pakotni *hyperfine*[30]. Pakotne veidota, lai ērti nodrošinātu veikspējas testu veikšanu, izmantojot nepieciešamos parametrus. Lai

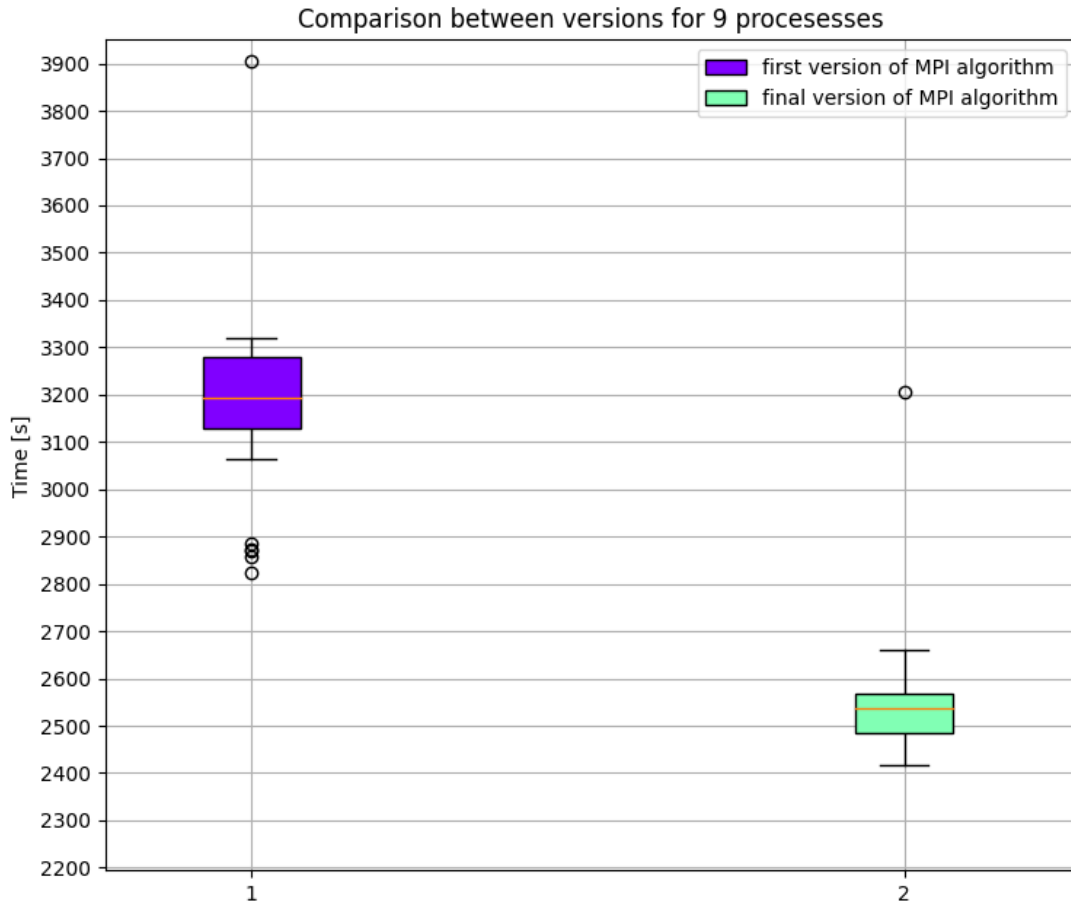
iegūtu salīdzinoši kontrolētus rezultātus, visiem nodaļā aprakstītajiem veikspējas testiem tiek izmantotas 5 *warmup* iterācijas, kā arī 30 rezultātu mērītās iterācijas.

Lai vieglāk atsauktos uz algoritma versijām, kuras tiek apskatītas, nodaļas ietvaros *pirmā* versija MPI algoritmam apzīmē 1.5 nodaļā aprakstīto pirmo algoritmu, kurš attēlots pielikumā B, bet *pēdējā* versija apzīmē 1.5 nodaļā aprakstīto otro algoritmu, kurš attēlots pielikumā C.

Lai attēlotu veikspējas uzlabojumus no *pirmās* MPI algoritma algoritma iterācijas un *pēdējās*, ņemot vērā vienādu kodolu skaitu (9), rezultāti tiek atspoguļoti kastveida diagrammā 3.1. attēlā, kur ar violeto krāsu tiek attēloti rezultāti no *pirmās* algoritma iterācijas, bet ar zaļo tiek attēlots *pēdējās* algoritma iterācijas veikspēja. Tehniski jēldatus *pēdējā* versijā apstrādā par vienu procesu mazāk nekā *pirmajā* versijā, taču tiek ņemts vērā, ka abi algoritmi katrs aizņem deviņus procesus kopā.

Ņemot vērā, ka datos ir izlecošās vērtības, biežākais izpildes laiks tiek mērīts, izmantojot mediānas vērtību. Pēc veikspējas testu rezultātiem, *pirmās* versijas algoritms visbiežāk izpildīsies aptuveni 3193.2148 sekundēs, kamēr *pēdējā* algoritma versija izpildīsies aptuveni 2535.8932 sekundēs.

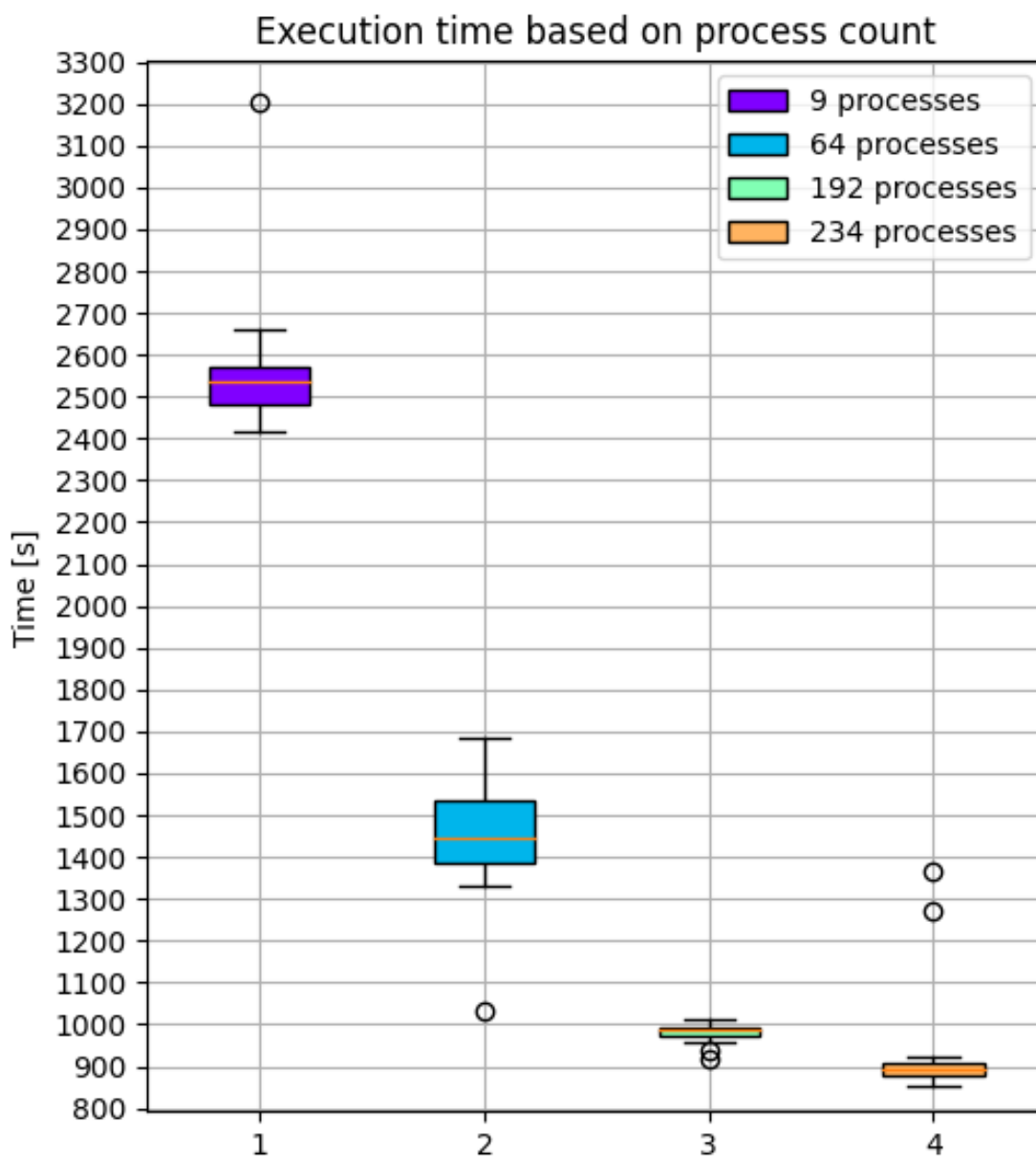
Jāņem arī vērā, ka *pirmajā* algoritmā nav integrētas anomāliju pārbaudes, kuras aprakstītas nodaļā 1.7., kā arī papildus datu izveide, līdz ar to, ja *pēdējā* algoritma darba skaits tiktu pielīdzināts *pirmajam*, tā veikspēja būtu vēl sliktāka. Lai gan *pirmā* algoritma darbībai, pēc teorijas, vajadzētu būt ātrākai, *pēdējā* algoritma ietvaros tiek izmantoti dažādi augstas veikspējas risinājumi problēmām, it īpaši ciklu vektorizācija, kas izteikti minimizē darbību ilgumu.



Att. 3.1.: Pirmā datu apstrādes algoritma veikspējas salīdzinājums ar pēdējās versijas veikspējas salīdzinājumu.

Pēdējais algoritms ir tieši atkarīgs no procesu skaita un failu daudzuma. Ideālā gadījumā, datu apstrādē algoritmam būtu iespējams apstrādāt katru failu savā procesā, taču, lai to realizētu 4 stundu gariem novērojumiem, būtu nepieciešams paredzēt 1920 kodolus datu apstrādei, kas patreizējā situācijā ir reti iespējams. Pēc teorijas, palielinot kodolu daudzumu, izpildes laiks samazināsies.

Lai attēlotu *pēdējās* versijas algoritmu, tiek apskatīts veikspējas rezultāts atkarībā no pieejamiem resursiem. Attēlā 3.2. tiek atspoguļoti rezultāti izmantojot 9, 64, 192 un 234 procesus. Iegūtie rezultāti no 9 un 64 procesu datiem ir apstrādāti vienas virtuālās mašīnas ietvarā, bet 192 un 234 procesu rezultāti, izmantojot visu virtuālo mašīnu kopējos skaitļošanas resursus.



Att. 3.2.: *Pēdējā* algoritma veikspēja atkarībā no procesu skaita

Lai gan Bakalaura darba ietvaros nav apskatīts pats sākotnējais algoritms, kurš realizēts bez MPI iekļaušanas, tipiska novērojuma apstrādāšanai bija nepieciešamas aptuveni 5 stundas.

3.3. Algoritma potenciālie uzlabojumi

Algoritma veikspējas uzlabošanas lielākais šķērslis ir algoritma procesu pārvaldības procesa optimizācija, jo apskatot kodolu noslodzi, var noteikt, ka algoritma izpildes sākuma posmā visi izsauktie kodoli izpildās uz 100% noslodzi, taču, kad pirmie faili nolasīti, uz pārvaldības procesu vienlaicīgi tiek sūtītas vairākas ziņas no apstrādes procesiem par nolases pabeigšanu. Tas nozīmē, ka ir nepieciešams apstrādāt visus ienākošos procesus, piešķirot atbrīvojušajiem procesiem jaunu failu apstrādei, kā arī klausīties jeb nodrošināt pārējo procesu statusu monitoringu. Visas aprakstītās darbības prasa skaitļošanas laiku, līdz ar to nevar efektīvi pārvaldīt visus pieejamos procesus. Problēmas risinājums būtu ieviest vairāk-pārvaldības procesus, lai gan minētais uzlabojums varētu ļoti sarežģīt algoritma loģiku. Lai gan patreizējā *numpy* vektorizācija ir daudz efektīvāka, salīdzinot ar ciklu pieeju, ir ieteicams izmantot iebūvētās *MPI* funkcijas, kuras nodrošinātu efektīvāku procesu komunikāciju.

Veicot datu apstrādes algoritmu optimizāciju, ir iespējams izmantot divas pieejas, kur katrai ir savi izmantošanas plusi un mīnusi. Pirmā pieeja, kuru savā ziņā arī izmanto *pirmais MPI* algoritms, ir statiski norādīt instrukcijas katram procesam un katru reizi algoritma izpildē izmantot statistisku procesu daudzumu. Minētais risinājums ļoti uzlabo kopējo veikspēju, jo tiek samazināts komunikāciju daudzums starp procesiem, kā arī realizācija ir daudz vienkāršāka, izmantojot nemainīgas instrukcijas, tādējādi veidot pilnībā kontrolētu vidi. Bet kopumā aprakstītais risinājums padara algoritmu nemērogojamu. Izmantojot *pirmo MPI* algoritmu, nav iegūti augstākas veikspējas rezultāti par jaunākās versijas rezultātiem, jo starp iterācijām tika apgūtas dažādas augstas veikspējas labās prakses, kas krasi uzlaboja veikspēju.

Nemot vērā, ka pieejamais skaitļošanas resursu daudzums var mainīties, nav garantēts konkrēts kodolu skaits algoritma izpildei, bija nepieciešams izmantot otro pieeju - zaudēt statistiskās instrukcijas katram procesam un dinamiski atļaut procesu pārvaldību. Minētā pieeja vienmēr samazinās kopējo veikspēju, jo procesiem savā starpā jāsazinās un jāinformē par procesa stāvokli, bet visbiežāk, izveidojot algoritmu pēc minētās metodes, ir iespējams iegūt labāku veikspēju kopumā, vienkārša iemesla dēļ - programrisinājumam ir pieejami vairāk skaitļošanas resursi.

Nemot vērā, ka Bakalaura darba ietvaros koda optimizācija un kopējā veikspēja ir svarīgs faktors, jāņem vērā, ka patreizējā sistēma pilnībā ir realizēta Python programmēšanas valodā. Tā tika izvēlēta atvieglotai algoritma realizācijai, plašo bibliotēku atbalsta un vienkāršotās sintakses dēļ, bet valodai dēļ tās izpildes laika interpretatora ir izteiktākas tehniskās limitācijas. Izmantojot *numpy* bibliotēku var efektīvi vektorizēt darbības, kā

ciklu izpildi un matricu manipulāciju, bet pilnībā optimizēts kods visticamāk būs lēnāks par pilnībā optimizētu kodu zemāka līmeņa valodā, kā C, pateicoties faktam, ka kods tiek interpretēts izpildes laikā. Diemžēl nav iespējams izvērtēt algoritma pārrakstīšanas guvumu neieguldot daudzas stundas pārrakstīšanā, līdz ar to konkrētus rezultātus konkrētajā jautājumā nevar iegūt.

SECINĀJUMI UN PRIEKŠLIKUMI

Darba gaitā ir sasniegts izvirzītais mērķis - veikt vāju radioastronomisko objektu datu apstrādi, to kalibrāciju, trokšņa filtrēšanu, analizēt un apkopot iegūtos rezultātus. Veicot darbu, autors secināja:

1. Lai realizētu datu apstrādes algoritmu, bija nepieciešams apgūt radioastronomisko novērojumu teorētiskos pamatus, kā arī Irbenes novērojuma stacijā lietoto infras-truktūru, ieskaitot datu formātus un ierakstīto informāciju.
2. Algoritmu izveide, balstoties uz zinātniskajiem rakstiem, nereti ir izaicinošs process, jo informācija rakstā var tikt interpretēta dažādos veidos. Minētais faktors, kopā ar nepilnīgām zināšanām radioteleskopa darbībā, noveda pie grūtībām pilnībā izprast kalibrācijas procesa darbību, kas sarežģīja uzlabojumu ieviešanu.
3. Veicot datu apstrādi, nepieciešams pārliecināties par iepriekšējiem datu apstrādes procesiem, jo nepietiekama informācija noved pie kļūdainas tālākas apstrādes. Ņemot vērā, ka nav pieejamas specifiski aprakstītas dokumentācijas par iepriekšējiem datu apstrādes procesiem, tas noveda pie apjukuma situācijās, kur, piemēram, tika apskatīta Doplera kompensācija, kas daļēji jau bija realizēta datu ierakstīšanas brīdī.
4. Apskatot vairākas apstrādes metodes, dažkārt ir vērts ieguldīt ilgāku laiku metodes izpratnei, jo metode var sniegt pielietojumu dažādās jomās. Tas pierādījās apskatot veivletu transformāciju, kuru bija iespējams pielietot ne tikai stohastiskā trokšņa samazināšanā, bet arī sistēmas temperatūru bojāto vērtību pareģošanā.
5. Apstrādājot datus, ir nepieciešams izvērtēt potenciālās paralelizācijas iespējas, kā arī izpētīt vairākus iespējamus variantus paralelizācijas realizācijai. Izmantojot MPI paralelizāciju, vairāku stundu ilgā datu apstrāde tika aizstāta tikai ar maksimāli pus stundu garu datu apstrādes procesu.
6. Veicot algoritmu paralelizāciju, ir nepieciešams ņemt vērā algoritmisko nestabilitāti. Vienmēr jāpārliecinās, ka izsauktais process pilda instrukcijas, kas procesam ierādītas. Pat ja algoritms ir paralelizēts, tam ir iespējams veikt dažādus vienkāršus uzlabojumus, kā piemēram ciklu vektorizēšanu, kas noved pie daudz labākas veiktspējas. Bakalaura darbā tas izpaudās algoritma iterāciju salīdzināšanā, kur, lai gan pirmās iterācijas algoritms veica mazāk darbu, izpildījās lēnāk par pēdējās iterācijas algoritmu.

7. Vāju radioastronomisko objektu datu apstrādes metodes dažkārt sniedz nepārlicinošus rezultātus, līdz ar to ir nepieciešams salīdzināt iegūtos rezultātus ar citu metožu rezultātiem, piemēram, ar rezultātiem no optiskajiem novērojumiem.
8. Veicot zinātnisku algoritmu izveidi, nepieciešams konsultēties ar jomas ekspertiem rezultātu analizēšanai, kā arī potenciālu uzlabojumu ieteikumiem. Radioastronomisko datu apstrādes rezultāti, bieži ir sarežģīti uzskatāmi, jo ne vienmēr ir skaidrs, kā izskatīsies rezultāts, taču izmantojot metodes no iepriekšējās pieredzes, ir iespējams efektīvāk apstrādāt datus, kā arī gūt pārliecību par iegūtajiem rezultātiem.
9. Iegūstot datus no ārējiem avotiem, ir jāpārliecinās par avotu uzticamību, kā arī nepieciešams būt pārliecinātam, ko iegūtie dati reprezentē. Resursi, kā piemēram, NASA HORIZONS satur ļoti daudz informācijas, taču dažkārt datu daudzums izraisa apjukumu. Tiek ierakstītas dažādas ātruma vērtības, atkarībā no citiem objektiem, rezultātā, iegūstot līdzīgas, bet tomēr, nedaudz atšķirīgas vērtības. Apskatot arī datus, kā pareģotā objekta spilgtuma vērtība, bieži dažādos resursos ir atšķirīga, lai gan objekts ir viens. Dati iegūti no dažādām datubāzēm un bieži tiek izmantoti atšķirīgi algoritmi, lai iegūtu pareģotas vērtības, kas izraisa atšķirīgu gala rezultātu.
10. Populārās programmēšanas valodās, kā piemēram Python vai C, populārām problēmām bieži ir augstas veikspējas risinājumi, kurus izmantojot, kopējā algoritma veikspēja var vairākas reizes uzlaboties.
11. Vāju radioastronomisku objektu novērošana ir sarežģīts darbs, līdz ar to, ir nepieciešams pārbaudīt vairākas un atšķirīgas signālu apstrādes metodes uz kādiem konkrētiem datiem, lai pārliecinātos par iegūto rezultātu pareizību.
12. Veicot datu apstrādi, ir nepieciešams aprakstīt iegūtos rezultātus dokumentācijā, kas autoram ļauj labāk izprast kopējos rezultātus, kā arī atskaites atļauj citiem pētniekiem labāk izprast algoritmiski realizētos problēmu risinājumus.

Līdz ar to, no iepriekš izdarītajiem secinājumiem, autors sniedz šādus priekšlikumus:

1. Izstrādātā algoritma procesu komunikācijas realizācija efektīvākā veidā nodrošinātu daudz labāku veikspēju datu apstrādes algoritma izpildē. To iespējams realizēt, izmantojot statiskas instrukcijas noteiktam procesu daudzumam vai ieviešot papildus pārvaldes procesus.
2. Algoritma veikspēju iespējams optimizēt, izmantojot zemāka līmeņa valodu, taču izstrādes process prasītu ilgu laiku, jo liela daļa funkciju realizētas izmantojot *Python* pakotnes.

3. Bakalaura darba ietvaros realizēto Doplera nobīdes aprēķināšanu Saules sistēmas objektiem būtu nepieciešams izvērtēt, jo iegūtie rezultāti ir nepārliciecināmi.
4. Autors iesaka apskatīt papildus datu apstrādes metodes, piemēram, Karhunen—Loève Transform metodi Furjē transformācijas vietā, kas, iespējams, sniegtu precīzākus rezultātus, jo metode paredzēta vāju signālu apstrādei, izmantojot nejaušu skaitļu un statistikas elementu aprēķinus. Transformācijas rezultātā, varētu būt iespējams samazināt stohastiskā trokšņa ietekmi, apstrādātajos rezultātos.
5. Papildus apskatītajiem novērojuma datiem, ir nepieciešams apskatīt objekta Stoka parametrus, kas atļautu iegūt vairāk informāciju par objekta polarizācijas radiācijas stāvokli un objekta magnētiskā lauka stāvokli.
6. Ņemot vērā faktoru, ka radioteleskops ir ļoti kompleksa sistēmu kopa, novērojumu laikā var rasties inženiertehniskas problēmas. Lai gan biežāko problēmu risinājumi ir ieviesti, ir nepieciešams padziļināti apskatīt problēmu sekas uz iegūtajiem datiem, kā arī veikt attiecīgās pārbaudes.
7. Komētas ir ieteicams novērot, izmantojot vairākas stacijas, gan optiskajā diapazonā, gan izmantojot citu radioteleskopu staciju datus, veidojot VLBI tipa novērojumus, jo tas sniegtu daudz vairāk datu, kā arī novērstu potenciālās inženiertehniskās problēmas.

IZMANTOTĀS LITERATŪRAS UN AVOTU SARAKSTS

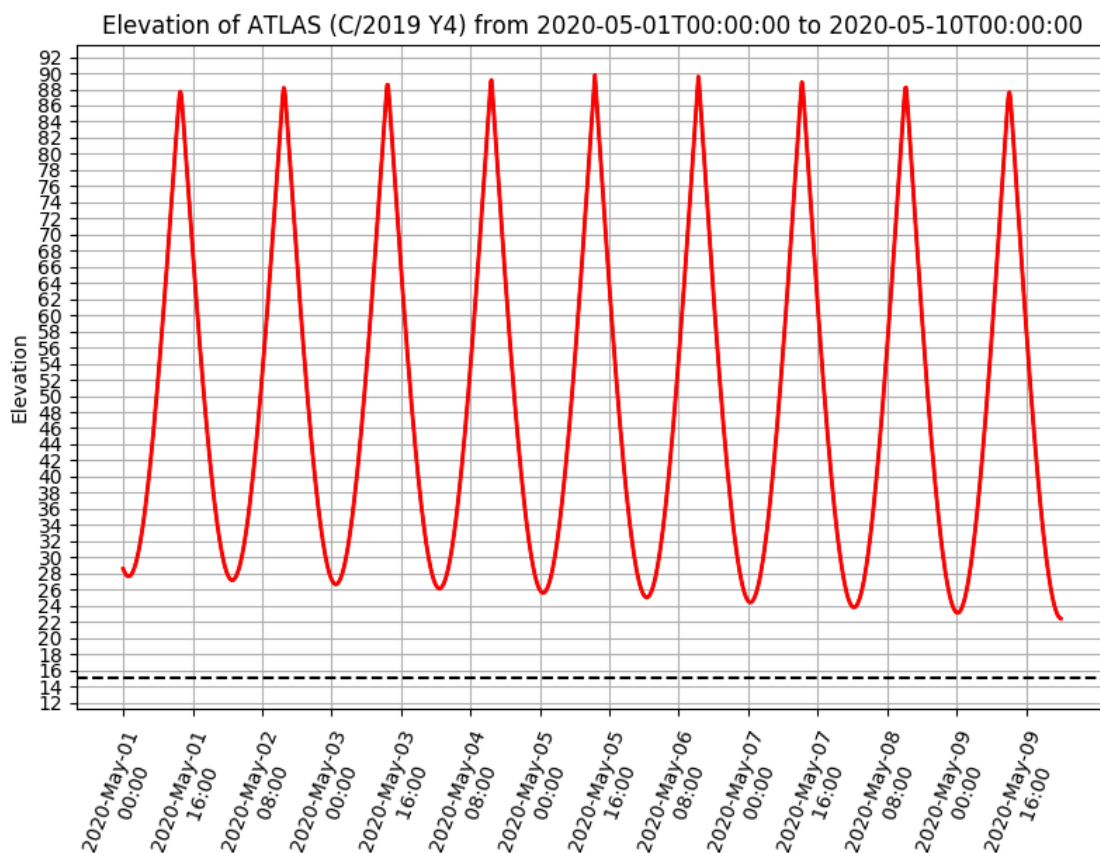
- [1] N. Jēkabsons K. Šķirmante M. Bleiders, A. Bērziņš and V. Bezrukovs. Low-Cost L-Band Receiving System Front-End for Irbene RT-32 Cassegrain Radio Telescope. *Latvian Journal of Physics and Technical Sciences Volume 56: Issue 3*, 2008.
- [2] MT Mechatronics GmbH. USER MANUAL - RADIO TELESCOPES RT-16/RT-32, VENTSPILS. *Tehniskā atskaite*, 2016.
- [3] M. Bleiders. USRP X300/310+TwinRX spectrometer progress v0.2. *Projekta nodevums/tehniskā atskaite*, 2018.
- [4] Calculation of Doppler effect for astronomical maser observations outside solar system. <https://github.com/sklandrausis/DataProcessingForMaserObservation/blob/master/src/vlsr.py>, skatīts 12.02.2020.
- [5] M. I. Pashchenko G. M. Rudnitskiĭ and P. Colom. Polarization Observations of Circumstellar OH Masers. *ISSN 1063-7729, Astronomy Reports, 2010, Vol. 54, No. 5*, 2010.
- [6] Albrecht Rüdiger Gerhard Heinzel and Roland Schilling. Spectrum and spectral density estimation by the Discrete Fourier transform (DFT), including a comprehensive list of window functions and some new at-top windows. *395068 [Max Planck Society]*, 2018.
- [7] Spectral Audio Signal Processing free books, Blackman-Harris window family. https://www.dsprelated.com/freebooks/sasp/Blackman_Harris_Window_Family.html, skatīts 11.03.2020.
- [8] NASA HORIZONS system. https://ssd.jpl.nasa.gov/horizons_batch.cgi, skatīts 19.11.2019.
- [9] Mpi4py library documentation. <https://mpi4py.readthedocs.io/en/stable/>, skatīts 08.02.2020.
- [10] Alexey G. Butkevich. Proper Motion and Secular Variations of Keplerian Orbital Elements. *Frontiers in Astronomy and Space Sciences Vol. 5, p.18*, 2018.

- [11] A. Kraus B. Winkel and U. Bach. Unbiased flux calibration methods for spectral-line radio observations. *arXiv:1203.0741 [astro-ph.IM]*, 2012.
- [12] Horizontal coordinate system representation. https://commons.wikimedia.org/wiki/File:Azimut_altitude.svg, skatīts 18.05.2019. File:az-el.png.
- [13] Documentation of matplotlib library. <https://matplotlib.org/>, skatīts 15.07.2019.
- [14] Numpy library documentation. <https://numpy.org/doc/stable/reference/generated/numpy.fft.fft.html#numpy.fft.fft>, skatīts 15.07.2019.
- [15] Scipy implementation of Blackman-Harris function. <https://docs.scipy.org/doc/scipy-0.14.0/reference/generated/scipy.signal.blackmanharris.html>, skatīts 11.03.2020.
- [16] Representation of FFT overlap percentage. https://www.spectraplus.com/DT_help/overlap_percentage.htm, skatīts 02.12.2019. File:OVERLAP2.png.
- [17] Discrete wavelet transform representation. https://upload.wikimedia.org/wikipedia/commons/2/22/Wavelets_-_Filter_Bank.png, skatīts 20.08.2019. File:DWT-fig.png.
- [18] PyWavelets library documentation. <https://pywavelets.readthedocs.io/en/latest/>, skatīts 20.08.2019.
- [19] Scipy implementation of Butterworth filter. <https://docs.scipy.org/doc/scipy-0.14.0/reference/generated/scipy.signal.butter.html>, skatīts 04.02.2020.
- [20] Scipy implementation of Savgol filter. https://docs.scipy.org/doc/scipy-0.16.1/reference/generated/scipy.signal.savgol_filter.html, skatīts 04.02.2020.
- [21] Statsmodels implementation of loess function. https://www.statsmodels.org/stable/generated/statsmodels.nonparametric.smoothers_lowess.lowess.html, skatīts 04.02.2020.
- [22] Fieldsystem documentation by NASA. https://web.archive.org/web/20161203064328/https://lupus.gsfc.nasa.gov/software_fs_main.htm, skatīts 01.11.2019. Pieejama tikai arhivēta lapas versija.

- [23] Claudio Maccone. The KLT (Karhunen–Loève Transform) to extend SETI searches to broad-band and extremely feeble signals. *Acta Astronautica Volume 67, Issues 11–12, December 2010, Pages 1427-1439*, 2010.
- [24] George Gabriel Stokes. On the Composition and Resolution of Streams of Polarized Light from different Sources. *Transactions of the Cambridge Philosophical Society, Vol. 9, p.399*, 1851.
- [25] Timothy Robishaw and Carl Heiles. The Measurement of Polarization in Radio Astronomy. *arXiv:1806.07391 [astro-ph.IM]*, 2018.
- [26] Internation Astronomical Union defined list of important spectral lines. <http://www.astrosurf.com/luxorion/radioastronomy-lines.htm>, skatits 10.07.2019.
- [27] Observation information for ATLAS C/2019 Y4. <http://astro.vanbuitenen.nl/comet/2019Y4>, skatits 20.04.2020.
- [28] Quan-Zhi Ye Man-To Hui. Observations of Disintegrating Long-Period Comet C/2019 Y4 (ATLAS) – A Sibling of C/1844 Y1 (Great Comet). *arXiv:2004.10990 [astro-ph.EP]*, 2020.
- [29] Minor Planet Center discovery of SWAN C/2020 F8. <https://minorplanetcenter.net/mpec/K20/K20G94.html>, skatits 15.04.2020.
- [30] Source code of hyperfine package for Linux. <https://github.com/sharkdp/hyperfine>, skatits 18.02.2020.

A ATLAS C/2019 Y4 REDZAMĪBAS GRAFIKS

Attēlā apskatīts Atlas C/2019 Y4 komētas elevācijas vērtības, atkarībā no laika. Ar sarkanu iekrāsots potenciālo novērojumu laiku intervāls, kas komētai iespējams visu laiku, jo visas vērtības ir virs 15 grādu sliekšņa. Attēls tiek apskatīts, jo attēlo atšķirīgu komētu potenciālo novērojumu iespēju. Komētu novērojumi iespējami daudz vieglāk realizējami nekā 1.3. attēlā apskatītās SWAN C/2020 F8 novērojumi. Dati iegūti no NASA HORIZONS sistēmas.

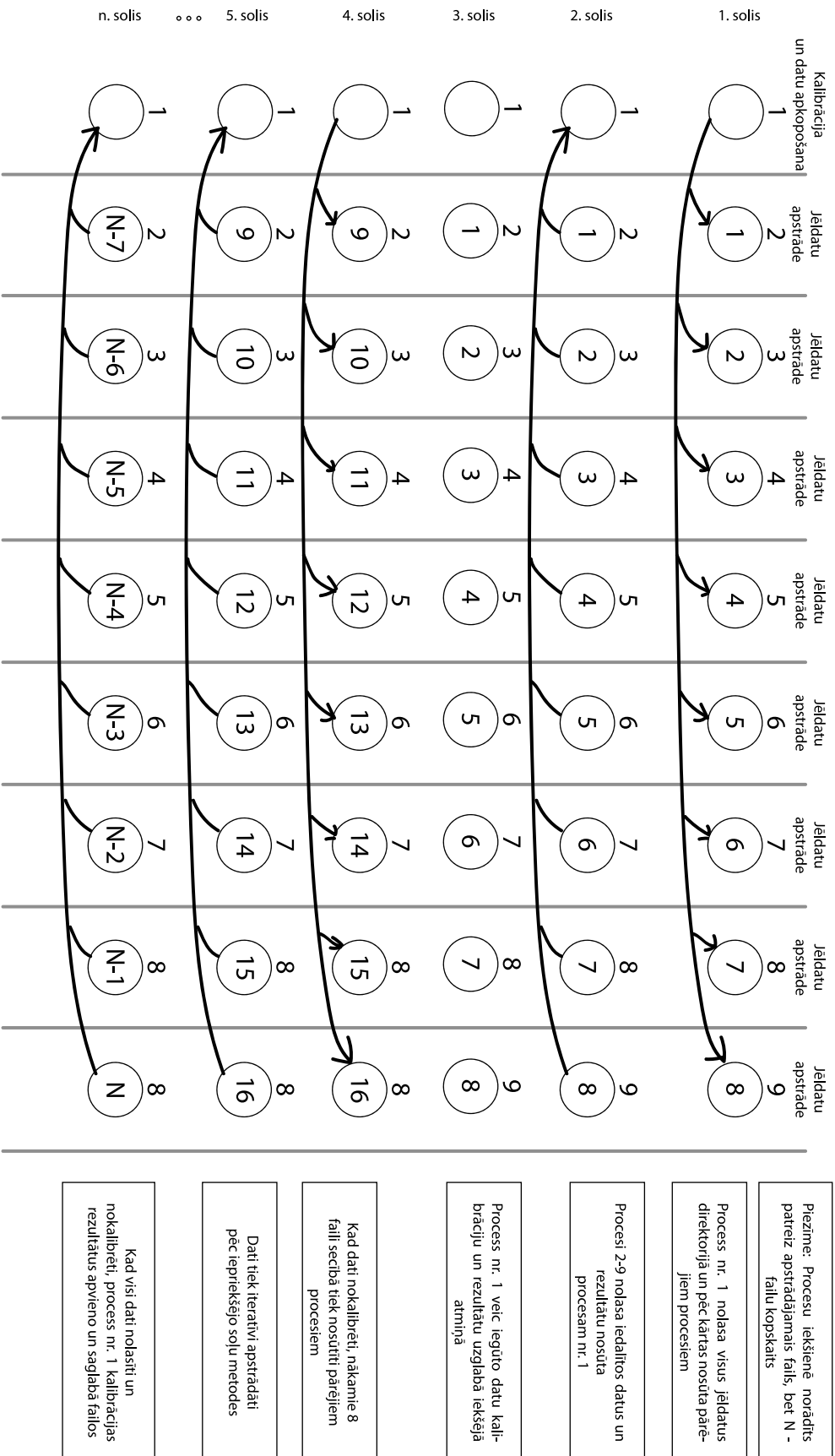


B MPI ALGORITMA PIRMĀS VERSIJAS IZPILDES PROCESA ATTĒLOJUMS

Attēlā tiek apskatīts, kā notiek procesu komunikācija un datu apstrādes plūsma pirmajā MPI algoritma versijā. Procesi sadalīti celiņos, kur katrs celiņš atbilst katram procesam. Pirmais process veic datu kalibrāciju un apkopošanu, kamēr pārējie - jēldatu apstrādi. Process katrā solī tiek attēlots kā aplis, kura iekšienē tiek norādīts, kurš fails no kopējās novērojumu datu kopas tiek apstrādāts.

Katrā datu apstrādes ciklā, tiek apstrādāti 8 faili, kuru failu nosaukumus process nr. 1 nosūta pārējiem procesiem. Kad dati nolasīti, katrs process nosūta iegūtos rezultātus pirmajam procesam, kurš rezultātus nokalibrē un uzglabā atmiņā. Pēc kalibrācijas, tiek izsūtīta nākamā 8 failu kopa procesiem un process tiek atkārtots, līdz visi faili apstrādāti. Kā N zīmējumā tiek apzīmēts kopējais failu skaits.

Procesa kodu var apskatīt D pielikuma repozitorijā, kur mapē *scripts* atrodas fails *MPI_calibr_v1.py*, kurā ir veikts pilns datu apstrādes process izmantojot pielikumā aprakstīto metodi.



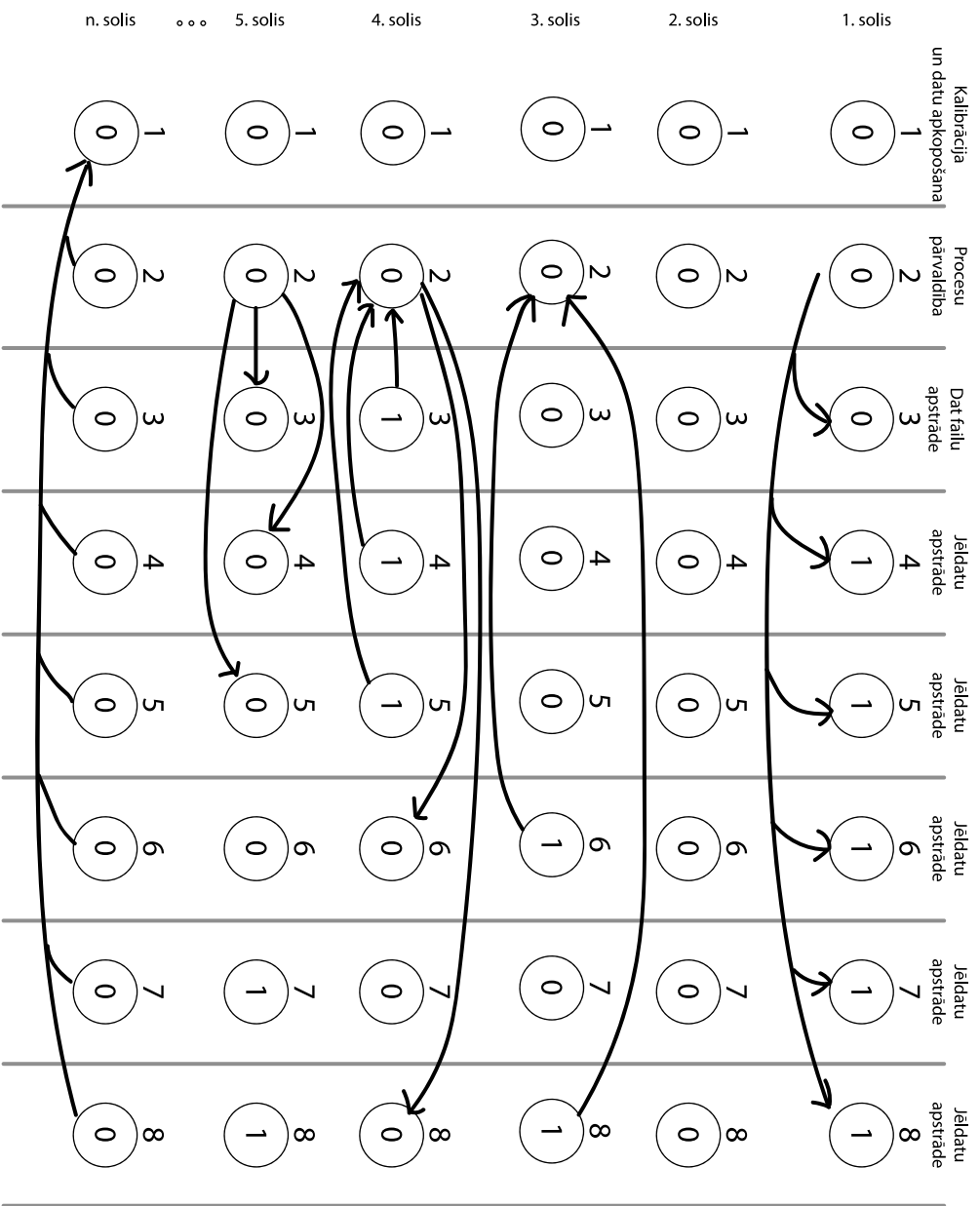
C MPI ALGORITMA FINĀLA VERSIJAS IZPILDES PROCESA ATTĒLOJUMS

Attēlā tiek apskatīts, kā notiek procesa komunikācija un datu apstrādes plūsma pēdējā MPI algoritma versijā. Procesi sadalīti celiņos, kur katrs celiņš atbilst katram procesam. Pirmais process veic datu kalibrāciju un apkopošanu, otrais process realizē procesa pārvaldību, trešais process - dat formāta failu apstrādi, kamēr pārējie - jēldatu apstrādi. Trešais process arī veic jēldatu apstrādi pēc dat formāta datu apstrādes beigšanas.

Process katrā solī tiek attēlots kā aplis, kura iekšienē tiek norādīts procesa statuss. Ja process ieņem vērtību "1", process ir brīvs un to var izmantot datu apstrādei.

Balstoties uz faktoru, ka katra faila nolasīšanas laiks var nedaudz atšķirties, arī komunikācija starp procesiem var atšķirties. Piemērā pirmie procesi, kuri beidz datu nolasīšanu ir process nr. 8 un nr. 6, bet realitātē nav iespējams pateikt kurš process beigs datu apstrādi pirmais. Nosacījums izpildās arī pārējos soļos. Pārvaldības process izsūta visus failu ceļus datu apstrādes procesiem, bet kad visi faili izsūtīti, pārējiem procesiem aizsūta ziņu par datu apstrādes beigām. Procesi saņemot ziņu, nosūta datu apstrādē iegūtos rezultātos procesam 1, kurš veic tālāku datu apstrādi un beidz darbību.

Procesa kodu var apskatīt D pielikuma repozitorijā, kur mapē *scripts* atrodas fails *MPI_calibr_v2.py*, kurā ir veikts pilns datu apstrādes process izmantojot pielikumā aprakstīto metodi.



Piezīme: Procesu iekšienē norādīts procesa statuss

Procesi 4-8 inicializēti kā brīvie procesi

Process nr. 2 detektē, ka ir brīvi procesi un tiem iedala failus

Process nr. 2 gaida uz pārējo procesu signāliem, par datu apstrādes beigām

Procesi nr. 8 un nr. 6 beidz jēldatu nolaišanu, uzglabā rezultātu un nosūta procesam nr. 2 ziņu par darba beigām, kas atzīmē procesus par pieejamiem

Process nr. 2 iedala procesiem nr. 8 un nr. 6 jaunus failus. Procesi 3, 4 un 5 ir beiguši savu datu apstrādi un ziņo par to procesam nr. 2

Arī process nr. 3 ir beidzis dat failu apstrādi, līdz ar to, tas savus rezultātus ieraksta failos un saņem instrukcijas jēldatu nolaišanai

Process nr. 2, sekojot līdzīgam principam, izsūta visus failus, komunicējot ar pārējiem procesiem

Kad visi faili izsūtīti, process nr. 2 izsūta ziņojumu pārējiem procesiem par darba beigām un katrs process apstrāda beidzamos failus un nosūta visus iegūtos rezultātus kopā ar faila nosaukumu procesam nr. 1

D BAKALAURA DARBA IETVAROS

VEIDOTIE KODI

Ņemot vērā kodu apjomu, lai iegūtu labāku algoritmu uzskatāmību, kods nav apskatīts Bakalaura darba apraksta ietvaros, bet pieejams publiskā, Bakalaura autora veidotā *github* repozitorijā. Repozitorijā ir divas direktorijas - *scripts* un *docs*, kur kodi apskatāmi *scripts* direktorijā, bet Bakalaura gala apraksta Latex pirmkods, kopā ar darbā izmantotajiem vizuālajiem materiāliem, apskatāms *docs* direktorijā.

Piekluve repozitorijam iespējama, izmantojot:
<https://github.com/dot361/Weak-radio-signal-data-processing>, vai pēc pieprasījuma.

GALVOJUMS

Ar šo es, Gints Jasmonts, galvoju, bakalaura darbs ir izpildīts patstāvīgi un bez citu palīdzības. No svešiem pirmavotiem ņemtie dati un definējumi ir uzrādīti darbā. Šis darbs tādā vai citādā veidā nav nekad iesniegts nevienai citai pārbaudījumu komisijai un nav nekur publicēts.

2020.gada _____._____

Es, Gints Jasmonts, atļauju Ventpils Augstskolai savu bakalaura darbu bez atlīdzības ievietot un uzglabāt Latvijas Nacionālās bibliotēkas pārvaldītā datortīklā Academia (www.academia.lndb.lv), kurā tie ir pieejami gan bibliotēkas lietotājiem, gan globālajā tīmeklī tādā veidā, ka ikviens tiem var piekļūt individuāli izraudzītā laikā, individuāli izraudzītā vietā.

Piekrītu _____

Nepiekrītu _____

2020.gada _____._____