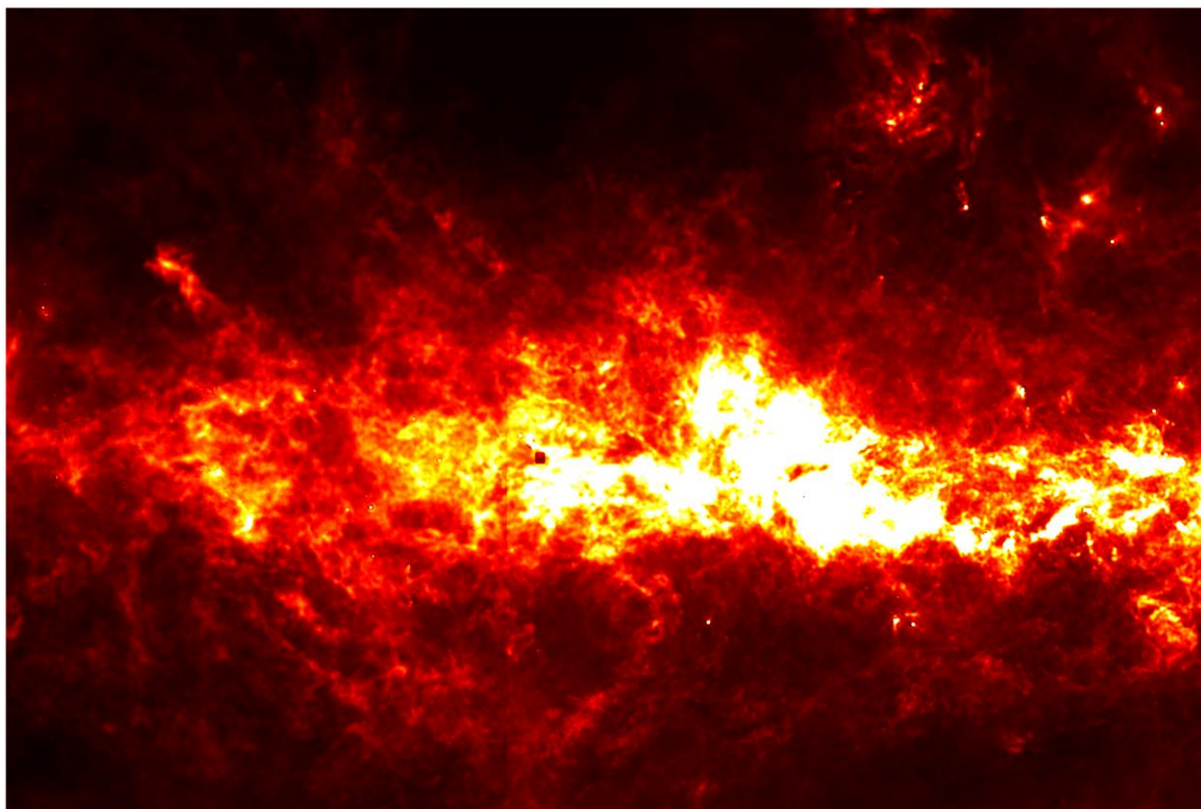




CHALMERS



Identifiering av föregångare till stjärnor i det interstellära mediet

En utmaning inom digital bildanalys

Kandidatarbete inom Rymd-, geo-, och miljövetenskap

Cecilia Erneman, Marcus Johansson, Joakim Kiselman,
Marcus Pettersson, Tage Tykesson, Ben Wooldridge

INSTITUTIONEN FÖR RYMD-, GEO-, OCH MILJÖVETENSKAP

KANDIDATARBETE 2023

Identifiering av föregångare till stjärnor i det interstellära mediet

En utmaning inom digital bildanalys

Cecilia Erneman, Marcus Johansson, Joakim Kiselman,
Marcus Pettersson, Tage Tykesson, Ben Wooldridge



CHALMERS

Institutionen för Rymd-, geo-, och miljövetenskap
CHALMERS TEKNISKA HÖGSKOLA
Göteborg, Sverige 2023

Identifiering av föregångare till stjärnor i det interstellära mediet
En utmaning inom digital bildanalys
Cecilia Erneman, Marcus Johansson, Joakim Kiselman, Marcus Pettersson,
Tage Tykesson, Ben Wooldridge

© Cecilia Erneman, Marcus Johansson, Joakim Kiselman, Marcus Pettersson,
Tage Tykesson, Ben Wooldridge, 2023.

Handledare: Jouni Kainulainen, Institutionen för Rymd-, geo- och miljövetenskap
Examinator: Magnus Thomasson, Institutionen för Rymd-, geo- och miljövetenskap

Kandidatarbete 2023
Institutionen för Rymd-, geo- och miljövetenskap
Chalmers tekniska högskola
412 96 Göteborg, Sverige

Omslag: Utdrag ur PROMISE-data som visar molekyln i det interstellära mediet

Typsatt i L^AT_EX
Göteborg, Sverige 2023

Identifiering av föregångare till stjärnor i det interstellära mediet

En utmaning inom digital bildanalys

Cecilia Erneman, Marcus Johansson, Joakim Kiselman, Marcus Pettersson, Tage Tykesson och Ben Wooldridge

Institutionen för Rymd-, geo- och miljövetenskap

Chalmers tekniska högskola

Sammandrag

Täta kärnor är små områden i så kallade molekylnmoln i det interstellära mediet som har en betydligt högre densitet än sin omgivning. Det första steget i bildandet av en stjärna är när en tät kärna börjar kollapsa under sin egen gravitation.

I följande projekt har ett opublicerat dataset kallat PROMISE använts som är en kombination av data från Spitzer-teleskopet och Herschel-teleskopet. Syftet med följande projekt är att skapa ett bildanalytiskt verktyg som kan identifiera täta kärnor utifrån PROMISE-datan och skapa en katalog av dessa. I PROMISE-datan finns det även artefakter som ska identifieras och särskiljas från de täta kärnorna. Projektet har avgränsats till att använda sig av traditionella digitala bildanalysmetoder. Programmet som skrivits är anpassat efter PROMISE-datan och kan därmed inte nödvändigtvis appliceras oförändrat på andra dataset.

Det slutgiltiga programmet använder de bildanalytiska teknikerna oskarp maskering och wavelettransform med *Haar* som basfunktioner. Oskarp maskering utnyttjar antingen en applicering av gaussisk utjämning eller ett lågpasfilter i frekvensdomänen där fouriertransform används. Som testmetod för att uppskatta de olika bildanalytiska metodernas säkerhet konstruerades artificiella kärnor och artefakter. Det undersöktes sedan hur många av dessa de olika metoderna kunde identifiera.

Resultatet som tagits fram är en katalog bestående av 23 345 stycken identifierade täta kärnor samt deras position, massa och radie. Genom metoden med gaussisk utjämning identifieras 18 994 stycken täta kärnor, med wavelettransform identifieras 4273 stycken täta kärnor och med fouriertransform identifieras 503 stycken täta kärnor. Endast 120 stycken täta kärnor identifierades av samtliga metoder. Metoden med gaussisk utjämning identifierade den största andelen artificiella kärnor på 82,8%. Metoden med fouriertransform identifierar den största andelen artificiella artefakter på 94,6%. Förändringar av programmets parametrar hade stor inverkan på resultatet.

Utifrån resultatet kunde slutsatsen dras att de valda bildanalytiska verktygen lämpar sig för att identifiera täta kärnor från PROMISE-datan. Samtliga bildanalysmetoder gav i stor utsträckning ett tillförlitligt resultat. Oskarp maskering genom gaussisk utjämning gav det bästa resultatet i testmetoden och identifierade flest täta kärnor i datan.

Nyckelord: Molekylnmoln, täta kärnor, PROMISE-data, artefakter, gaussisk utjämning, fouriertransform, oskarp maskering, wavelettransform, digital bildanalys.

Identification of the precursors to stars in the interstellar medium

A challenge in digital image processing

Cecilia Erneman, Marcus Johansson, Joakim Kiselman, Marcus Pettersson, Tage Tykesson och Ben Wooldridge

The Department of Space, Earth and Environment

Chalmers University of Technology

Abstract

Dense cores are small regions in so-called molecular clouds in the interstellar medium with a significantly higher density than their surroundings. The first step in the formation process of a new star is when a dense core collapses under the weight of its own gravity.

This project makes use of the unpublished PROMISE dataset which is a large map of molecular clouds derived using a combination of data from the Spitzer telescope and the Herschel telescope. The project aims to create a tool to identify and catalogue dense cores that appear in the PROMISE dataset, utilizing image analysis. Another aim is to identify artefacts in the data and differentiate them from the dense cores. The project has been limited to using only traditional digital image analysis techniques. The program that identifies the cores has been adapted to the PROMISE data and will therefore not work without adjustments on different datasets.

The developed program utilizes multiple digital image processing methods, namely unsharp masking and wavelet transform with the *Haar* basis functions. Unsharp masking is achieved through both the application of a linear gaussian kernel and usage of fourier transform. To determine how correct the produced catalogue was, artificial dense cores and artefacts were inserted into the dataset. The artificial cores were then identified using the program and compared with the input, thus enabling an estimated measure of how reliably the objects are identified.

The result is a catalogue consisting of 23 345 identified dense cores and their position, mass and radius. With the use of unsharp masking with gaussian smoothing 18 994 dense cores were identified, with wavelet transform 4273 dense cores were identified and with unsharp masking with fourier transform 503 dense cores were identified. Only 120 dense cores were identified by every method. Unsharp masking with gaussian smoothing identified the most artificial cores, namely 82,8%. Unsharp masking with the fourier transform identified the most artificial artefacts, namely 94,6%. Changing the parameters of the program had a big impact on the result.

The conclusion drawn is that all chosen image analysis methods work well for identifying dense cores in the PROMISE data. All methods seemed to yield reasonable results in regards to the objects identified. Unsharp masking using gaussian smoothing yielded the best results and identified the most dense cores in the dataset.

Keywords: Molecular clouds, dense cores, PROMISE dataset, artifact, smoothing, unsharp masking, Fourier transform, wavelet transform, digital image analysis.

Innehåll

1	Inledning	1
2	Astronomisk bakgrund	3
2.1	Stjärnbildning	3
2.2	PROMISE-datan	5
2.3	Artefakter i PROMISE	8
3	Teori om bildanalysmetoder	9
3.1	Utgjämmande filter	9
3.2	Fouriertransform	10
3.3	Oskarp maskering	12
3.4	Intensitetströskling	13
3.5	Wavelettransform	14
4	Metod	17
4.1	Hantering av data	18
4.2	Initial bildbehandling	18
4.3	Wavelettransform	19
4.4	Filtrering från standardavvikelser	20
4.4.1	Villkor för klassificering av en tät kärna	20
4.4.2	Beräkning av baslinjen	20
4.5	Hantering av artefakter	22
4.5.1	Identifiering av artefakter	22
4.5.2	Säkerhetsskala	22
4.6	Artificiella kärnor och artefakter	25
4.6.1	Massa-radieförhållande för täta kärnor	25
4.6.2	Konstruktion och insättning av artificiella kärnor	26
4.6.3	Konstruktion och insättning av artificiella artefakter	27
4.7	Sammanställning av program	32
5	Resultat	33
6	Diskussion	43
6.1	Resultatanalys	43
6.2	Begränsningar och förbättringsområden	44
6.2.1	Allmänna begränsningar	44
6.2.2	Intensitetströskling	45
6.2.3	Fouriertransform	45
6.2.4	Wavelettransform	46
6.2.5	Filtrering från standardavvikelser	46
6.2.6	Identifiering av artefakter	47

6.2.7	Radie- och massafunktion	47
6.2.8	Testmetoder	48
6.3	Förslag på framtida studier	50
6.4	Slutsats	50
Literaturförteckning		55
A Appendix		I
A.1	Pseudokod	I
A.2	Val av parametrar för identifiering av täta kärnor	III
A.3	Val av parametrar för artificiell data	IV
A.4	Programkod och fullständig katalog	V

1. Inledning

Studien av stjärnor utgör en central roll inom astronomin. Stjärnors ålder och fördelning i galaxer ger kunskap om galaxers historia, dynamik och evolution, vilket är områden som innefattar många olösta mysterier. Dessutom så står stjärnor för tillverkningen och fördelningen av tyngre material som till exempel kol, kväve och syre [1]. Själva stjärnformationen har en avgörande roll för hur en stjärnas livscykel kommer att se ut, därmed är stjärnformation av stort intresse att studera.

De astronomiska fenomen som är av intresse för detta projekt är samlingar av gas som med tiden kan komma att bilda nya stjärnor. Dessa gasansamlingar hittas i det interstellära mediet och befinner sig i stadiet av stjärnbildning innan de blir så kallade protostjärnor, alltså innan deras egen gravitation leder till att materialet i deras kärnor hettas upp [1]. Dessa samlingar av gas kan sedan observeras av teleskop genom att undersöka hur mycket infraröd strålning som blockeras och kan då urskiljas från övrig mindre kompakt gas [2], och kommer framöver att benämnas täta kärnor.

Ett dataset som tagits fram med ändamålet att kunna observera dessa gasansamlingar är PROMISE-projektet [3] (**P**RObing the **O**rigins of **M**assIve molecular cloud **S**tructur**E**s). PROMISE-datan är en kombination av två dataset, ett är samlat med Spitzer teleskopet [4] med en våglängd på $8\mu\text{m}$ och ett är samlat med Herschel-teleskopet [5] med våglängder mellan $70\mu\text{m}$ och $500\mu\text{m}$. Vid kombination av dessa dataset utnyttjas Spitzer-datans höga upplösning samt Herschel-datans goda kalibrering och intensitetskänslighet. Resultatet är ett dataset som kallas för PROMISE och utgör en unik galaktisk karta som beskriver kolumndensiteten av gas i tusentals kalla moln i det interstellära mediet. Kolumndensitet är mängden materia som finns i vägen mellan observatör och föremål som observeras.

I tidigare forskningsprojekt har föregångare till stjärnor kunnat identifieras med hjälp av Herschel-teleskopet som arbetar i det infraröda spektrumet [5]. Ett exempel på ett sådant projekt presenteras i artikeln *A far-infrared survey of molecular cloud cores* [6] där det tagits fram en katalog av täta kärnor. I artikeln har ett antal bilder av molekylnmoln studerats. Från bilderna togs allt brus bort och det studerades dels var det fanns intensitetstoppar som inte översteg en viss maxbredd och bilderna undersöktes även visuellt för att ta fram katalogen. Idag finns det däremot inga publicerade forskningsprojekt där PROMISE-datan har använts för att identifiera täta kärnor. Den galaktiska kartan från PROMISE-datan har hög upplösning, god kalibrering och känslighet, vilket innebär att enklare bildanalytiska metoder kan användas för att identifiera täta kärnor.

Syftet med projektet är att skapa ett bildanalytiskt verktyg som kan hitta täta kärnor utifrån PROMISE-datan. Projektet syftar att identifiera täta kärnor och

att särskilja dem från artefakter i datan. Olika bildanalytiska tekniker har använts för att uppnå det här syftet. Projektet är av en utforskande natur på grund av avsaknaden av en redan existerande katalog för PROMISE-datan. Därför testas flera bildanalytiska metoder för att undersöka vad som bäst framhäver täta kärnor i PROMISE-datan. Metoderna testas sedan mot konstgjord data som liknar den verkliga, för att kunna komma fram till med vilken säkerhet metoderna hittar de täta kärnorna.

Projektet avgränsar sig till att inte använda någon form av maskininlärning utan att istället använda mer traditionella tekniker för digital bildanalys. I detta ändamål används programspråket python med diverse bibliotek för att kunna implementera valda bildanalytiska tekniker på ett smidigt sätt. Projektet avgränsar sig också till den tillgängliga och tidigare nämnda PROMISE-datan och programmet kan inte nödvändigtvis appliceras oförändrat på andra dataset utan förändringar i parametrar och eventuella algoritmer.

2. Astronomisk bakgrund

I kommande kapitel kommer den astronomiska bakgrund som krävs för projektet att beskrivas. Kapitlet innefattar den teoretiska bakgrunden kring stjärnbildning, datans konstruktion samt relevanta egenskaper av PROMISE-datan.

2.1 Stjärnbildning

För att digital bildanalys av rymdbilder vid identifiering av täta kärnor ska få mening krävs en viss förståelse för stjärnors livscyklar, samt vad det finns för information att hämta från dataseten. Då Vintergatans materia studeras påträffas dels alla stjärnor och dels det interstellära mediet, ISM, som är den materia som finns mellan stjärnorna.

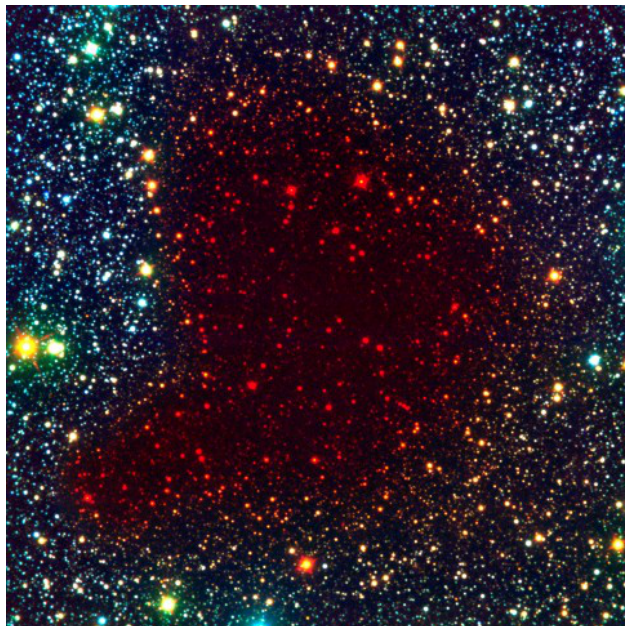
Det interstellära mediet består av gas, stoft och kosmisk strålning. Dessutom består ISM av elektromagnetisk strålning, gravitationsfält och magnetfält. Den största delen av materia i ISM är i gasform och kan observeras i fyra olika stadium. Ett av dessa stadium är kall och tät till största del molekylär gas som bildar större moln [7]. Molekylmolnen är relativt massiva med en diameter som kan variera mellan 0.1 pc till 100 pc. I en vardaglig kontext skulle dessa molekylmoln inte anses särskilt täta med en densitet på bara 10^{12} molekyler/m², men jämfört med övrigt ISM vars densitet är runt 10^6 atomer/m² har molekylmolnen en betydligt högre densitet än sin omgivning [7]. För synligt ljus kan dessa moln vara helt ogenomträngliga och bilda mörka fläckar på stjärnhimlen, se Figur 2.1 [7]. Molekylmolnen framstår som mörka eftersom stoft i molnen absorberar ljus som utstrålas från de bakomliggande stjärnorna. Att molnen syns som mörka fläckar kan underlätta att hitta dem, men det underlättar inte observerandet av objekt inuti molnen.

I stora delar av ISM så existerar väte som ensamma atomer, men i de tätare molekylmolnen är vätet istället i den molekylära formen H₂. I dessa moln är temperaturen ofta så kall att det i stort sett är omöjligt för vätemolekylerna att exciteras och själva skicka ut strålning. Den kalla temperaturen innebär att molnen endast med stor ansträngning kan observeras direkt, så istället utnyttjas infrarött ljus för att undersöka hur mycket av ljuset från bakomliggande stjärnor som blockeras av molekylmolnen. Infrarött ljus kan utnyttjas då det inte blockeras ut helt av molnet utan beror på kolumndensiteten på molnet [2], se Figur 2.2.

Det är i dessa molekylmoln som stjärnor bildas, och i dem finns områden där densiteten är högre än omgivningen [7]. I dessa områden är många stoft- och gaskärnor lokaliserade och de kallas för täta kärnor. En tät kärna breder ut sig på ungefär 0.1 pc och har en massa runt ett tiotal solmassor [10]. De täta kärnorna kan börja kollapsa under sin egen gravitation om området är tillräckligt tätt, det vill säga har en tillräckligt stor massa för att gravitationen ska vara större än det utåtriktade

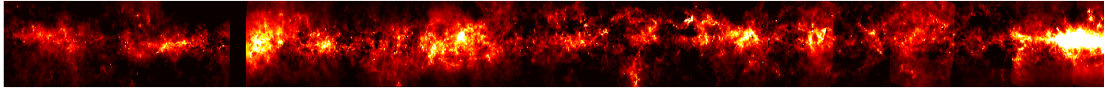


Figur 2.1: Molekylmolnet "Barnard 68" i synligt ljus. Bilden visar hur molekylmolnet blockerar ut allt synligt ljus. Molekylmolnet består av damm som är partiklar av fast materia. Dampartiklarna är mindre än en mikrometer i storlek men dess höga densitet i molekylmolnet gör molnet ogenomträngligt för synligt ljus. Från [8], CC-BY.



Figur 2.2: Molekylmolnet "Barnard 68" i infrarött ljus. Bilden visar ljuset från stjärnorna bakom molekylmolnet som blir synliga vid infraröd våglängd. Från [9], CC-BY.

trycket [7]. En tät kärna som kollapsar under sin egen gravitation är det första steget i bildandet av en stjärna.



Figur 2.3: Kompletta PROMISE-datan som har använts i projektet. PROMISE-datan har en upplösning på (120,000 x 7,000) pixlar

2.2 PROMISE-datan

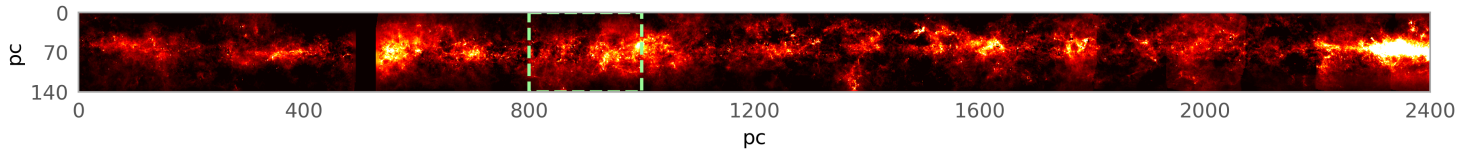
För att hitta täta kärnor har opublicerad data kallad PROMISE använts, se Figur 2.3. PROMISE beskriver kolumndensiteten av tusentals gasmoln som befinner sig på galaktisk longitud $1,5^\circ < l < 40^\circ$ och galaktisk latitud $-1^\circ < b < 1^\circ$. Datan är en sammanslagning av data insamlad från Spitzerteleskopet och Herschelteleskopet. Spitzerteleskopet samlar in data med våglängden $8\mu\text{m}$ medan Herschelteleskopet samlar in data i våglängderna $70\text{--}500\mu\text{m}$. Genom att kombinera data från de båda teleskopen kan deras bästa egenskaper utnyttjas för att få ett så bra dataset som möjligt. I PROMISE-datan utnyttjas Spitzers höga upplösning och Herschels noggranna kalibrering och intensitetskänslighet för att få ett unikt dataset med en upplösning på två bågsekunder.

De olika molnen befinner sig på ett avstånd mellan 2000 pc och 5000 pc [11] från jorden. Ifall det utifrån detta antas ett typiskt avstånd på 3500 pc för alla objekt så kan PROMISE-datans värden mätas i fysiska enheter på runt 0.02 pc per pixel och en pixels intensitetsvärde kan konverteras till solmassa. Ett pixelvärde på 1 kan antas innehålla 0.0081 solmassor av gas [11]. Detta innebär att molekylnmoln samt täta kärnor, vars storleksordning beskrevs i avsnitt 2.1 kan ses med en god upplösning i PROMISE-datan. PROMISE-datan förväntas innehålla tusentals täta kärnor [11]. På grund av det stora antalet kärnor som förväntas finnas i datan krävs en automatisk identifikation av de täta kärnorna.

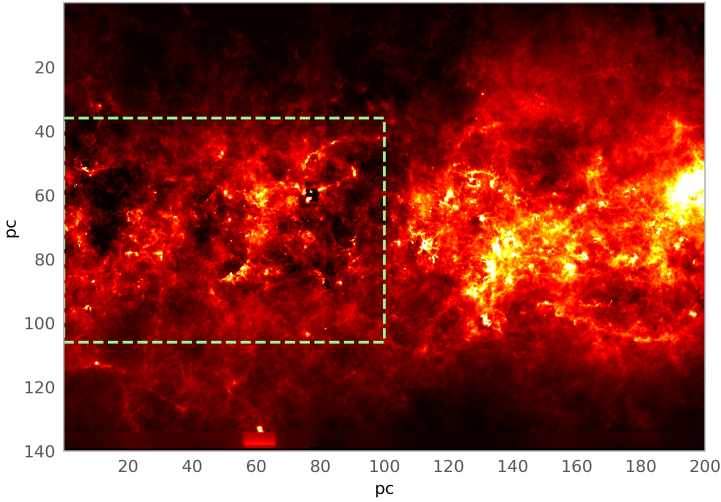
Ett exempel på skalan av PROMISE-datan kan ses i Figur 2.4 som innehåller flera molekylnmoln som varierar med storleksordning från 100 pc i bredd till 5 pc i bredd. I dessa moln kan täta kärnor observeras.

Pixelintensiteten i PROMISE-datan har stor variation och ligger mellan värden på $-5,70 M_\odot/\text{px}$ och $55,62 M_\odot/\text{px}$. De negativa värdena beror på att en perfekt kalibrering av nollvärdet inte har uppnåtts vid konstruktion av datan [11]. En konturkarta över ett område i PROMISE-datan går att se i Figur 2.5 vilket illustrerar variationen i värdena.

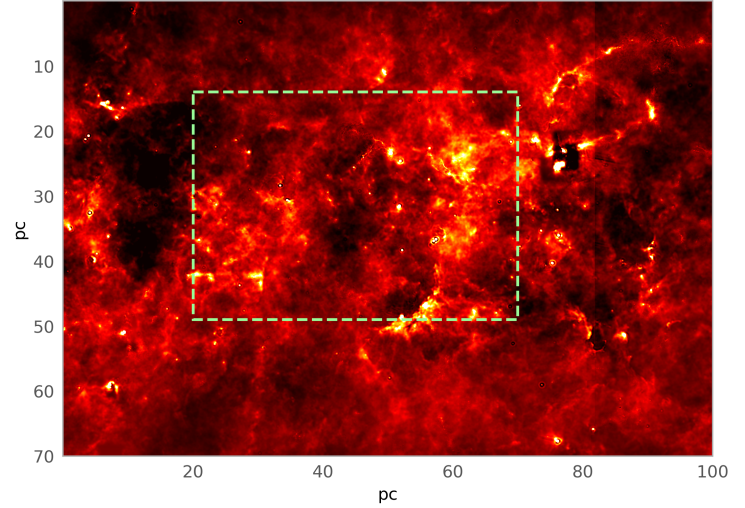
2. Astronomisk bakgrund



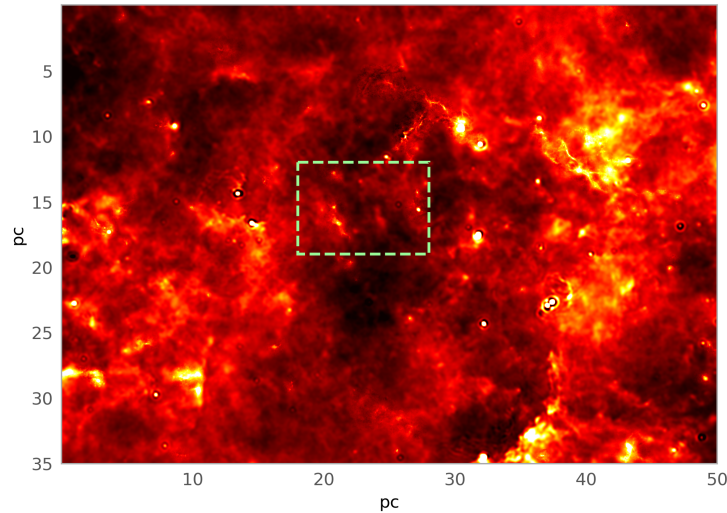
(a) Den kompletta PROMISE-datan modifierad med en grön ruta som illustrerar vart i bilden det har zoomats in för att få figur (b)



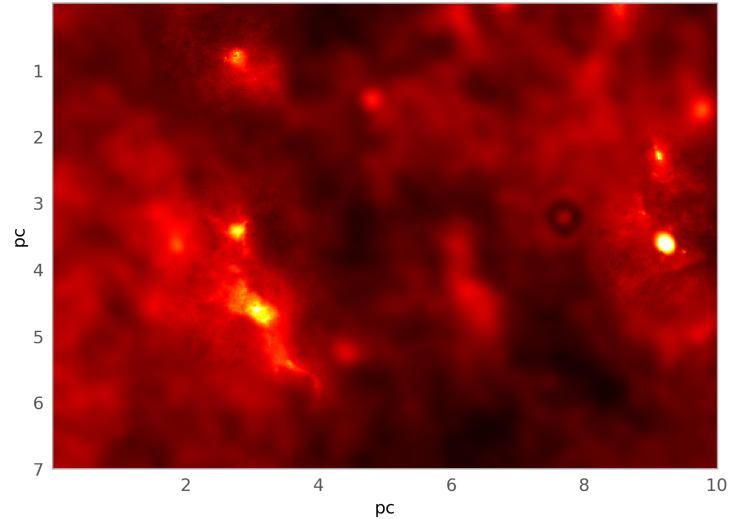
(b) PROMISE-datan, uppskalning med faktor av 12



(c) PROMISE-datan, uppskalning med faktor av 48

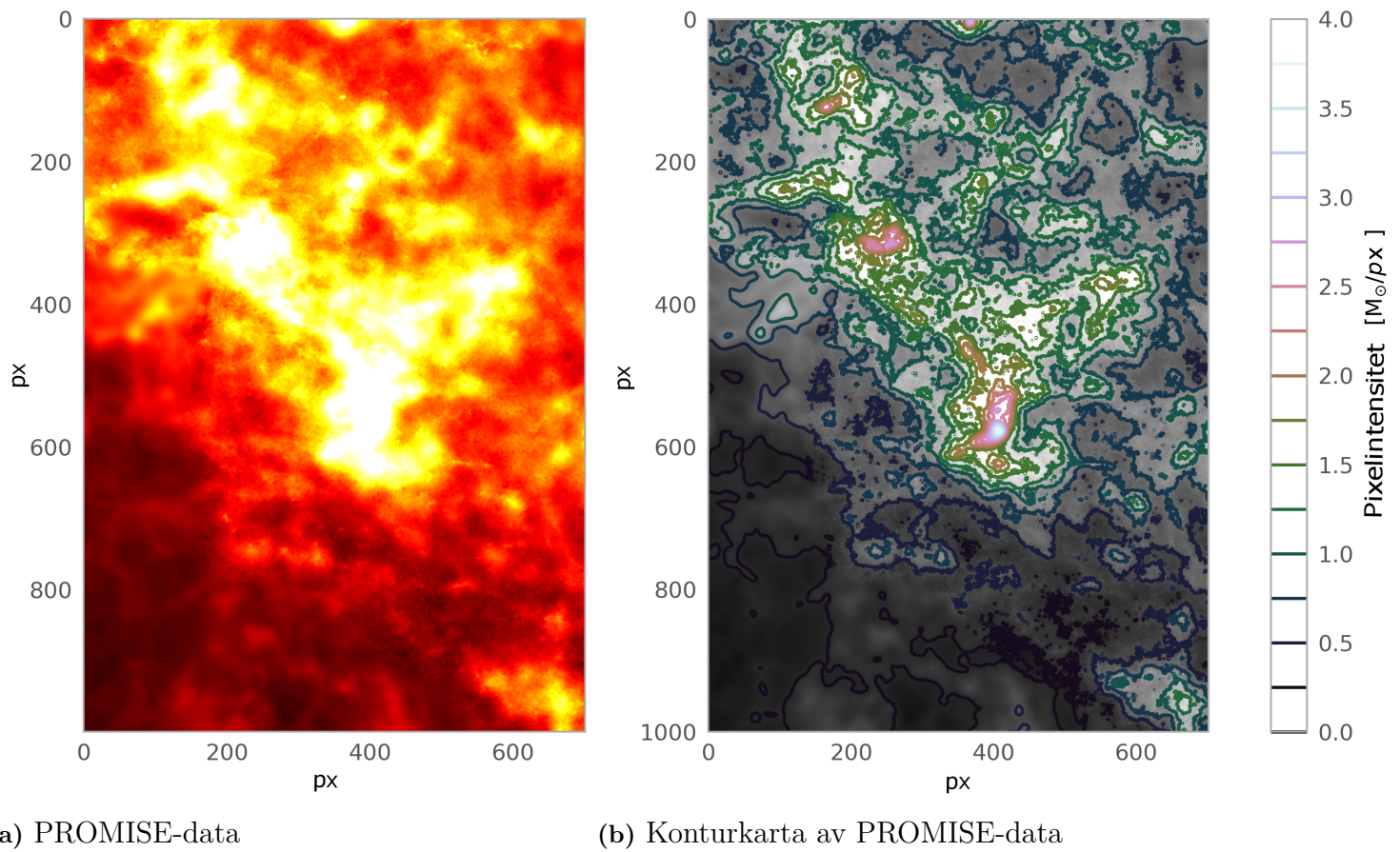


(d) PROMISE-datan, uppskalning med en faktor av 192



(e) PROMISE-datan, uppskalning med en faktor av 4800

Figur 2.4: Utdrag ur PROMISE-datan uppskalat på olika nivåer där de vertikala och horisontella axlarna mäts i enheten parsek. Den gröna rutan i varje figur illustrerar det område som har zoomats in på för att få nästa delfigur. Från figur (b), (c) och (d) går det att se ett flertal större och mindre molekylnmoln. I figur (e) går det att se ett antal täta kärnor i utkanten av molekylnmolnet, i mitten av figur (d). Det går även att se en tydlig artefakt i figur (e).



Figur 2.5: En konturkarta över ett urval av PROMISE-datan där figur (a) visar originaldatan och figur (b) visar samma område efter att en konturkarta blivit applicerad. Till höger om figuren illustreras hur pixelintensiteten i figuren varierar med färgskalan mätt i M_{\odot}/px

2.3 Artefakter i PROMISE

PROMISE bygger delvis på kolumndensitetsdata som härleds ifrån Herschel-data över strålning från rymdstoft [12]. Den härledda kolumndensitetsdatan från Herschel innehåller artefakter som propagerar till PROMISE-datan. I PROMISE syns dessa artefakter som mättade pixlar vilket gör att de kan förväxlas med täta kärnor. De måste därför hanteras. För att kunna hantera artefakterna underlättar det att förstå hur de uppkommer, vilket förklaras nedan.

Herschel-datan består av flera observationer av vid olika våglängder (70, 160, 250, 350 och 500 μm). Teleskopets upplösning påverkas av den observerade våglängden vilket ger de olika observationerna upplösningar på 8,5; 13,5; 18,2; 24,9 och 36,3 bågsekunder för de respektive våglängderna.

Det vanliga sättet att härleda kolumndensitet från dessa observationer är att kurvanpassa varje pixels spektralfördelning till en svartkropp. Herschel-datan som används i PROMISE bygger istället på en metod som använder flera temperaturer för att härleda kolumndensiteten [12]. I PROMISE används en modell där den totala spektralfördelningen beräknas utifrån tolv temperaturer mellan 8 och 50 K. Detta leder till en högre noggrannhet för kolumndensiteterna eftersom metoden tar hänsyn till temperaturförändringar längs teleskopets siktlinje.

Spektralfördelningen för stoft som har temperaturer kring metodens maximum (50 K) har sitt maximum vid de kortaste våglängderna som Herschel observerar. För kallare temperaturer hamnar istället spektralfördelningarnas maximum vid längre våglängder. Detta innebär i praktiken att varmt stoft kan ses i bättre upplösning än kallt stoft.

Metodens maxtemperatur på 50 K ger också upphov till artefakter i datan. Om stofttemperaturen i ett litet område överskrider 50 K fallerar metoden vilket leder till att pixeln och de närliggande pixlarna blir mättade. Om metoden fallerar misslyckas den med beräkningen av kolumndensiteten vid samtliga temperaturer vilket, på grund av skillnaden i upplösning för olika temperaturer, påverkar områden med olika storlekar. Det betyder att för 50 K påverkar metodens misslyckande enbart en pixel men för lägre temperaturer påverkar det ett större område. Detta leder till att artefakterna har ett vågliknande mönster som propagerar ut från dess centrum.

Eftersom dessa artefakter syns i datan som små områden med högre intensitet än sin omgivning förväxlas de lätt med täta kärnor. Det som tydligt skiljer dem från täta kärnor är deras vågliknande mönster. Vid närmre inspektion liknar det vågliknande mönstret en sinc-funktion vilket gör att artefakterna i två dimensioner kan approximeras till:

$$A = C \frac{\sin(\omega x)}{\omega x}. \quad (2.1)$$

3. Teori om bildanalysmetoder

I följande kapitel kommer relevant teori för de bildanalysmetoder som används i projektet att beskrivas. Nedan beskrivs teori kring utjämnande filter, oskarp maskering, tröskling, fourier- och wavelettransformer samt dess syfte i projektet.

Då den stora majoriteten av PROMISE-datan är områden av ISM och andra typer av signaler som för projektets syfte kan klassas som ointressant brus är det av intresse att kunna framhäva vissa delar av datan och bortse från andra delar. Ett typiskt sätt att uppnå detta är genom bildanalysmetoder och att skapa masker som markerar de delar av datan som uppfyller något krav.

3.1 Utjämnande filter

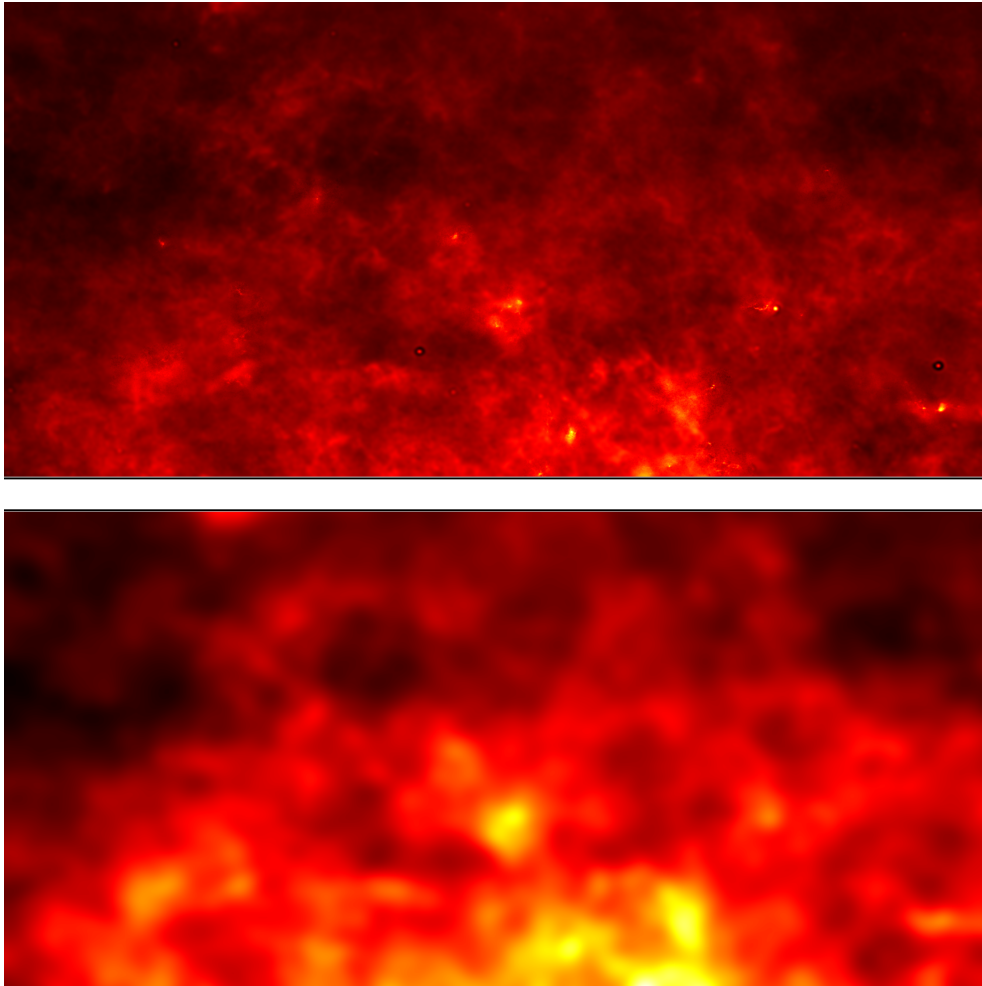
Utgjämnande filter används, som namnet föreslår, för att få fram en utjämnad bild av datan. Mer exakt fås ett resultat där alla ingående pixlars intensitetsvärde blir påverkat av pixelvärden som finns i det närliggande området. Det utjämnande filtret fungerar genom att falta den tvådimensionella datan med en kärna som släpper igenom låga frekvensinnehåll, även kallat lågpasfilter. Storleken och formen på kärnan avgör hur utjämnad bilden blir och hur mycket av bakgrunden som framhävs. Det är ofta fördelaktigt att använda utjämnande filter vid bildanalys då det framhäver bildens större strukturer och agerar som en form av brusreduktion. Som exempel är en utjämnad och mindre brusig bild mer eller mindre nödvändig om man ämnar använda oskarp maskering på den, se Figur 3.3.

En välfungerande och effektiv metod för att jämna ut data är gaussisk utjämning, som uppnås genom att falta bilden med en gaussisk kärna. En gaussisk kärna fungerar genom att sampla en gaussisk funktion som Ekvation 3.1 kring dess centrum, där σ är standardavvikelsen [13]. Urvalet avgör kärnans storlek.

$$G(x, y) = \frac{1}{2\pi\sigma^2} e^{-\frac{x^2+y^2}{2\sigma^2}} \quad (3.1)$$

Faktorn $\frac{1}{2\pi\sigma^2}$ i ekvation (3.1) är en normaliseringsfaktor som ser till att den genomsnittliga färgskalan som originalbilden har förblir densamma efter den gaussiska utjämningen har genomförts [14].

Vilket värde som väljs för standardavvikelsen påverkar i vilken utsträckning en pixel jämnas ut med sin omgivning [13]. Då värdet på standardavvikelsen ökas så minskar information med höga frekvenser runt pixeln. I Figur 3.1 illustreras en gaussisk utjämning på PROMISE-datan.



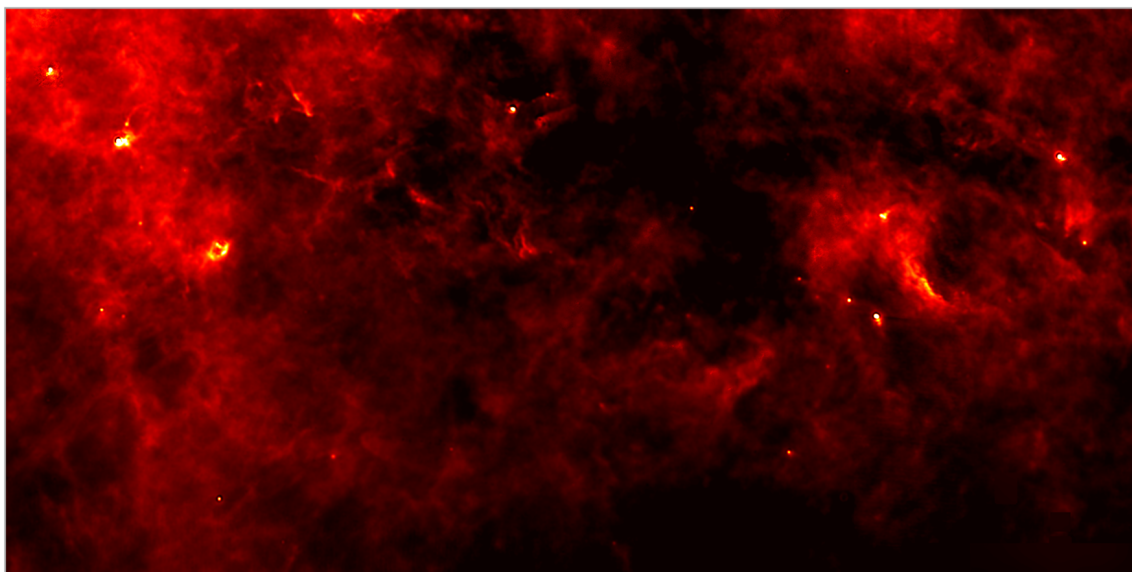
Figur 3.1: Figur som visar gaussisk utjämning på PROMISE-datan med $\sigma = 25$. Övre bild visar originaldata utan förändring, undre bild visar resultatet efter att en gaussisk utjämning har utförts på datan.

3.2 Fouriertransform

Bland alla olika metoder att behandla bilder på så är fouriertransform en av de mest kraftfulla och mest vedertagna. Anledningen är att när en fouriertransform görs på en signal så förloras ingen information, utan signalen uttrycks i frekvensdomänen helt komplett. Då kan vissa frekvenser selektivt filtreras ut vilket förändrar originalsignalen på ett specifikt sätt beroende på vilka frekvenser som filtrerats ut.

Förutsatt att typen av fouriertransform är lämplig för det går det med detta att behandla bilder genom samma princip. De olika frekvenserna motsvarar då olika egenskaper hos bilden. En enkel tolkning är att beskriva låga frekvenser som de större strukturerna utan starka kontraster medan de högre frekvenserna beskrivs som mindre strukturer med hög kontrast. Ett lågpasfilter har därmed effekten att den jämnar ut bilden och ger den mindre kontrast medan ett högpassfilter framhäver bildens mindre detaljer där det finns större kontrast [15]. Det finns i stort sett ingen

gräns på antalet olika sätt att filtrera bilder i frekvensdomänen, men dessa faller utanför projektets omfattning.

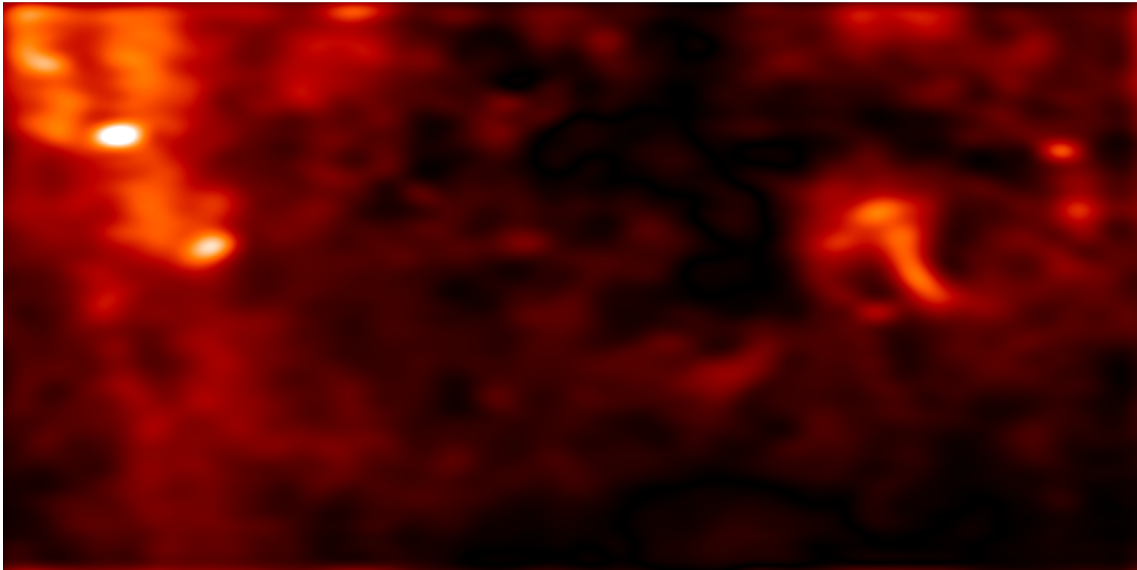


Figur 3.2: Exempel på del av datan innan den transformerats till frekvensdomänen och filtrerats genom ett lågpasfilter.

Den typ av transform som används för bildbehandling är tvådimensionell diskret fouriertransform enligt ekvation 3.2, där (k, l) representerar en koordinat och N och M kan tolkas som bildens dimensioner i höjd och bredd [15]. På samma sätt är $f(k, l)$ bilden i rumsdomänen, och $-2i\pi(\frac{ka}{N} + \frac{lb}{M})$ basfunktionen relaterat till varje punkt $F(k, l)$ i frekvensdomänen. Detta kan tolkas som en multiplikation av bilden i rumsdomänen och dess korresponderande basfunktion, som sedan summeras och ger resultatet av värdet i punkten $F(k, l)$. Basfunktionerna är sinus- och cosinuskurvor med olika frekvens. Frekvenserna stiger så att $F(0, 0)$ motsvarar bildens genomsnittliga ljusstyrka och $F(N - 1, M - 1)$ motsvarar den högsta frekvensen.

$$F(k, l) = \sum_{a=1}^{N-1} \sum_{b=1}^{M-1} f(a, b) e^{-2i\pi(\frac{ka}{N} + \frac{lb}{M})} \quad (3.2)$$

Ekvation härledd från [15] sida 257.

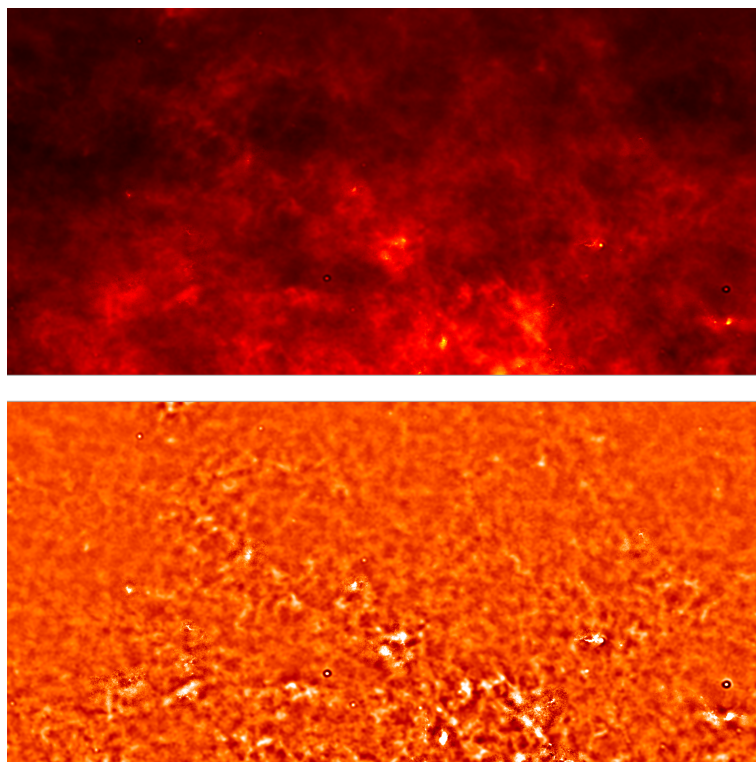


Figur 3.3: Samma del av datan som i Figur 3.2 efter den körts genom ett lågpasfilter i frekvensdomänen som sedan transformerats tillbaka till rumsdomänen.

3.3 Oskarp maskering

Oskarp maskering är en väletablerad metod som används för att framhäva detaljer i bilder och göra dem skarpare [15]. Metoden fungerar genom att först ta fram en utjämnad variant av datan, vilket till exempel kan göras genom metoder som gaussisk utjämning eller fouriertransform, för att sedan subtrahera den från originaldatan för att ge en oskarp mask. Denna mask läggs sedan tillbaka på originaldatan för att förstärka de detaljer som masken innehåller och på så sätt göra originaldatan skarpare.

Den oskarpa masken är i projektets applikationer relevant i sig utan återapplikering på originaldatan. Den oskarpa masken framhäver genom att subtrahera de låga frekvenserna istället de höga frekvenserna och kan på så sätt användas för att hitta detaljer som är karakteristiska för täta kärnor. Om tekniken appliceras korrekt kan den oskarpa masken sedan användas för att hitta de områden där en tät kärna troligtvis existerar.



Figur 3.4: Figur som visar oskarp maskering på samma område som Figur 3.1. Övre bild visar originaldatan, undre bild visar resultatet som fås genom att subtrahera den utjämnande masken från originaldatan.

3.4 Intensitetströskling

Målet med intensitetströskling är att dela upp data i objektpunkter och bakgrundspunkter baserat på om att de överskrider ett visst intensitetsvärde [15]. Utefter detta villkor skapas en mask med de punkter som överstiger intensitetsvärdet. Om värdet i en pixel överskrider intensitetsvärdet som det jämförs med skrivs korresponderande värde i masken till ett, annars skrivs värdet över till noll för att på så sätt kunna få en mask som endast innehåller pixlar som uppfyller villkoret. På så sätt fås en binär mask som kartlägger de områden i originalbilden som är av intresse.

Hur intensitetsvärdet beräknas varierar beroende på vilken metod som används. Med en global tröskelvärdesmetod så beräknas ett medelvärde för hela datan eller regionen medan för en lokal eller adaptiv tröskelmetod så beräknas det för varje punkt [15]. Detta gör att en lokal tröskling kan fungera bättre för data med stor variation i intensitetsnivåer i olika regioner som drar upp medelvärdet. Den adaptiva metoden kan då få med även lokala toppar i intensitetsnivåer som den globala metoden skulle kunna missa.

3.5 Wavelettransform

Wavelettransformen liknar fouriertransformen på det sätt att den låter en viss funktion över ett visst intervall representeras i termer av en ortonormal bas [15]. En fördel med wavelettransform jämfört med fouriertransform är att den har en så kallad temporal upplösning, med det menas att den fångar upp både information om frekvens och lokalisering [16].

Serieexpansionen för en endimensionell funktion $f(x)$ med waveletfunktion $\psi(x)$ och skalfunktion $\varphi(x)$ definieras som

$$f(x) = \sum_k c_{j_0}(k) \varphi_{j_0,k}(x) + \sum_{j=j_0}^{\infty} \sum_k d_j(k) \psi_{j,k}(x) \quad (3.3)$$

där j_0 är en godtycklig startskala, $c_{j_0}(k)$ kallas för approximationskoefficienter och $d_j(k)$ kallas för detaljkoefficienter [15].

Wavelettransformen består av ett ortonormalt par av basfunktioner samt tillhörande skalfunktioner. Vid wavelettransform i två dimensioner används en tvådimensionell skalfaktor $\varphi(x, y)$ och tre stycken tvådimensionella waveletfunktioner $\psi^H(x, y)$, $\psi^V(x, y)$ och $\psi^D(x, y)$ [15]. De tre stycken waveletfunktionerna är riktningskänsliga, där $\psi^H(x, y)$ mäter intensitetsvariationer horisontellt, $\psi^V(x, y)$ vertikalt och $\psi^D(x, y)$ diagonalt.

De skalade och translaterade basfunktionerna definieras i två dimensioner som

$$\varphi_{j,m,n}(x, y) \varphi(2^j x - m, 2^j y - n) \quad (3.4)$$

$$\psi_{j,m,n}^i(x, y) = 2^{j/2} \psi^i(2^j x - m, 2^j y - n) \quad (3.5)$$

där index i definierar vilken riktningswavelet som åsyftas, $i = H, V, D$ [15].

Den diskreta wavelettransformen för en bild $f(x, y)$ av storlek $M \times N$ är

$$W_{\varphi}(j_0, m, n) = \frac{1}{\sqrt{MN}} \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} f(x, y) \varphi_{j_0,m,n}(x, y) \quad (3.6)$$

$$W_{\psi}^i(j, m, n) = \frac{1}{\sqrt{MN}} \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} f(x, y) \psi_{j,m,n}^i(x, y) \quad (3.7)$$

där $i = \{H, V, D\}$ och j_0 är en godtycklig startskala. Koefficienterna $W_\varphi(j_0, m, n)$ definierar en approximation för $f(x, y)$ vid skala j_0 [15]. Koefficienterna $W_\psi^i(j, m, n)$ definierar horisontella, vertikala och diagonala detaljer för skalor $j \geq j_0$.

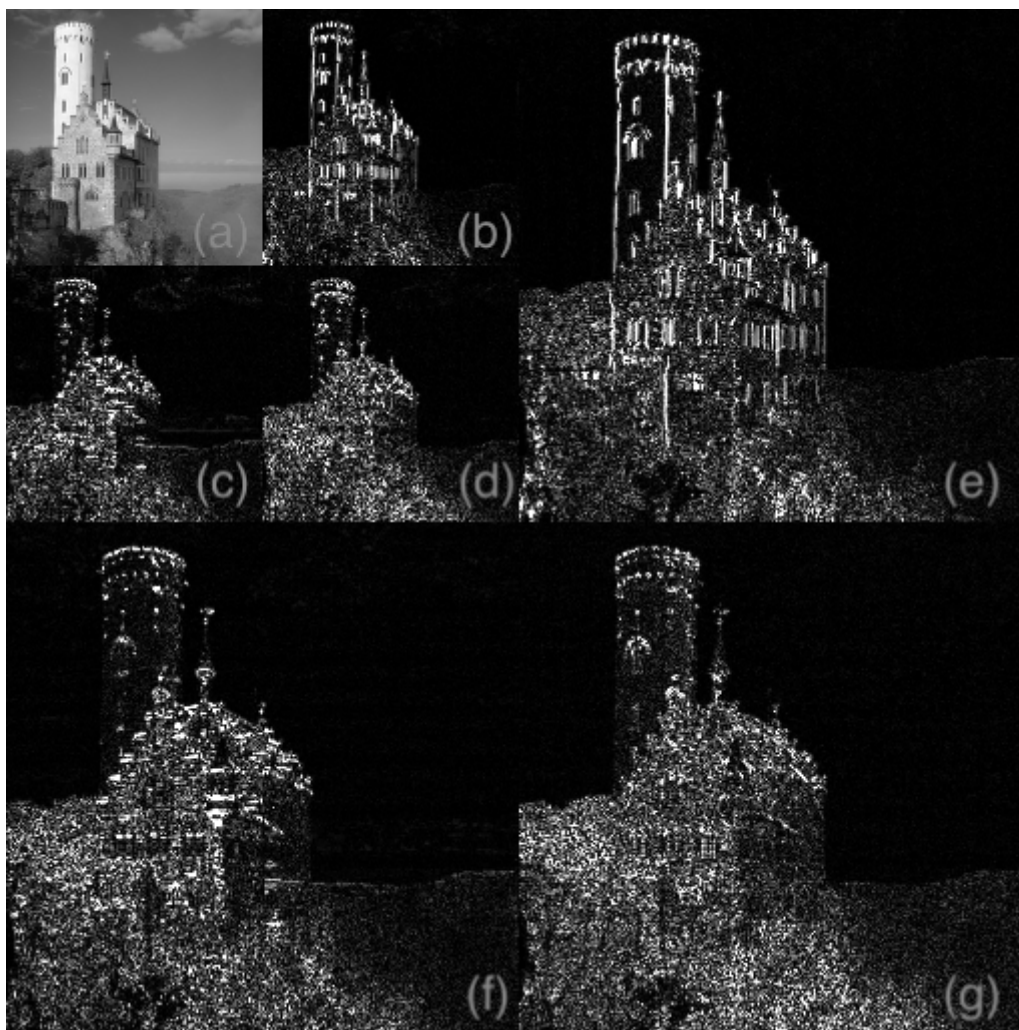
Wavelettransform kan inom bildanalys användas för att få ut önskade detaljer ur en bild. Dessa detaljer kommer från att falta den ursprungliga bilden med en waveletfunktion, och de delar av bilden som liknar waveletfunktionens mönster får ett högt värde i faltningsresultatet. Faltningsresultatet visar då vart detta mönster förekommer i bilden. Beroende på waveletfunktionens mönster eller från vilket håll faltningen sker, kan olika objekt framhävas, så som täta kärnor.

Transformen fungerar som så att skalfunktionerna och waveletfunktionerna sönderdelar datan till så kallade approximationer och detaljer. Approximationer avser en slät version av den ursprungliga datan som har samplats ner med en faktor av 2^n , där n avser hur många ordningar av wavelettransformen som har genomförts. Detaljer avser den skillnad som fås mellan originalet och approximationen. Även detaljerna skalas ner med en faktor av 2^n . Det som fås är första ordningens wavelettransform. Processen repeteras vid den nya skalan och ger nästa ordnings wavelettransform. Det slutgiltiga resultatet som fås är då en bild där brus har reducerats och detaljer har framhävts [17]. Ett exempel på andra ordningens wavelettransform av Figur 3.5 illustreras i Figur 3.6.



Figur 3.5: Originalbild som har använts för att illustrerar andra ordningens wavelettransform i Figur 3.6. Från [18], CC-BY-SA 3.0

Vilken bas som används vid wavelettransform påverkar hur väl olika typer av detaljer framhävs. Den ortonormala basen *Haar* utformas av kvadratformade funktioner av olika skalor och utgör den enklaste formen av wavelet [15].



Figur 3.6: Illustration över andra ordningens wavlettransform där orginalbilden visas i Figur 3.5. Figur (a) visar approximationen vid andra ordningens waveletttransform. Figur (b) illustrerar vertikala detaljer, Figur (c) horisontella detaljer och Figur (d) diagonala detaljer vid första ordningens waveletttransform. Figur (e) illustrerar vertikala detaljer, Figur (f) horisontella detaljer och Figur (g) diagonala detaljer vid andra ordningens waveletttransform. Från [19], CC-BY-SA 3.0.

4.1 Hantering av data

Ett viktigt steg för utförandet av projektet var hanteringen av datan, där datans storlek ibland var utmanande. PROMISE-datan var given som en *Flexible Image Transport System* (FITS) fil, ett filformat som ofta används inom astronomi, och har en storlek på 3.12 gigabyte [20]. Storleken på filen gjorde den svår att hantera i sin helhet med de datorer som fanns tillgängliga under projektet.

För att underlätta hanteringen av datan delades datan upp i mindre segment med hjälp av python. När metoder testades användes endast ett av dessa segment till en början, för att sedan appliceras på fler segment och till slut hela datan. Även när kataloger på hela datan togs fram analyserades regioner av datan var för sig för att sedan sättas ihop till en komplett katalog, istället för att försöka ladda in hela filen i datorns minne.

Ett verktyg som användes för att visuellt analysera datan är datorprogrammet SAO Image DS9 som har utvecklats för astronomisk data [21]. Till skillnad från python med biblioteket matplotlib [22] gjorde DS9 det enkelt att zooma in på datan. DS9 användes därför för att manuellt uppskatta rimligheten av uträknade metoder på ett snabbt och smidigt sätt. Det gick till exempel att efter att en potentiell tät kärna hittats, enkelt undersöka den i DS9 för att bedöma ifall programmet fungerar korrekt, istället för att skriva en ny pythonfunktion med det syftet.

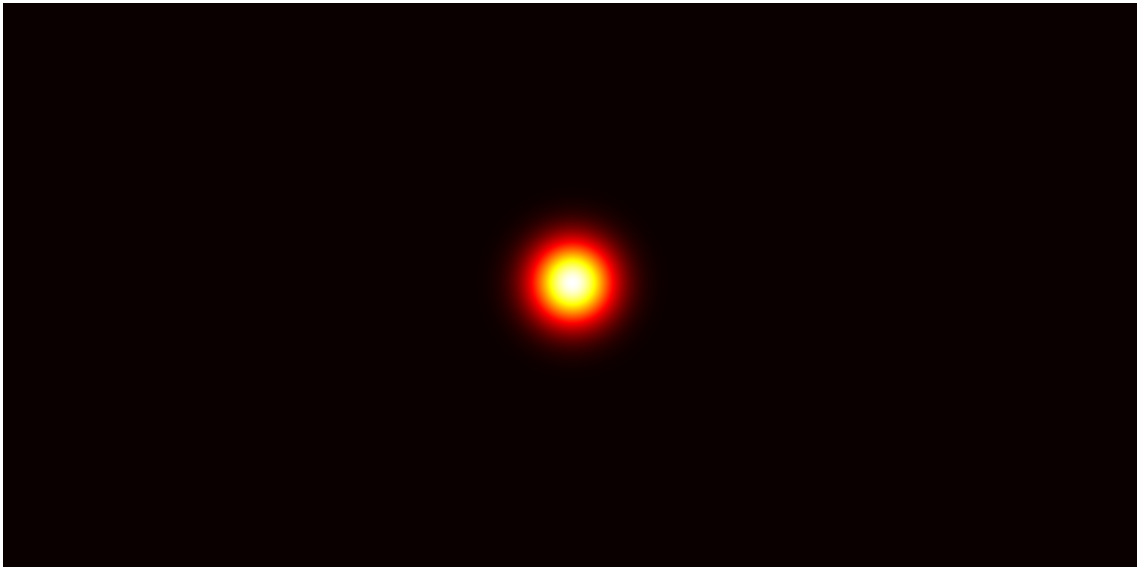
4.2 Initial bildbehandling

För att reducera mängden brus och framhäva datans större strukturer användes två olika typer av utjämning, gaussisk utjämning och utjämning genom ett lågpasfilter i frekvensdomänen, som testades var och en för sig.

För gaussisk utjämning användes den inbyggda funktionen *gaussian* från pythonbiblioteket *scikit image* [23], vilket är en gaussisk utjämning som fungerar enligt teorin beskrivet i avsnitt 3.1. Funktionen tar emot data och ett standardavvikelsevärde för att ge en utjämnad bild.

Den andra typen av utjämning är appliceringen av ett lågpasfilter i frekvensdomänen, där datan genomgått en tvådimensionell diskret fouriertransform så att ett filter kan appliceras. Mer bestämt användes en så kallad *fast fourier transform* som är en form av fouriertransform som optimerats för att minska mängden beräkningar och därmed exekveringstiden [24]. Lågpasfiltret skapas med en cirkulär gaussisk distribution av intensitet (Se Figur 4.2). För att lågpasfiltret ska kunna appliceras bör frekvenserna av bilden centreras, med de lägsta frekvenserna i mitten på bilden. Efter att lågpasfiltret har applicerats så inverteras centreringsen, och datan transformerades med den inversa tvådimensionella fouriertransformen för att återställa datan till rumsdomänen. En gaussisk fördelning av intensitet behövdes i lågpasfiltret för att undvika oönskade effekter som uppstod vid användning av

ideala filter.



Figur 4.2: Lågpasfilter med gaussisk distribution av intensitet i frekvensdomänen.

Efter att de utjämnade filtrena hade skapats så subtraherades den utjämnade datan från originaldatan för att skapa en skarp bild där mindre detaljer har framhävts. Detta trösklades sedan efter ett valt parametervärde för att få tillbaka en binär mask med relevanta punkter.

4.3 Wavelettransform

Utöver att datan behandlades med utjämnande filter och oskarp maskering för att få fram de högfrekventa pixlarna, behandlades den även med wavelettransform. Datan wavelettransformerades med Haar-waveletfamiljen till en nivå, vilket framhävde de horisontella, vertikala och diagonala detaljerna separat. Resultatet var tre olika masker, en för varje typ av detalj, som hade hälften så stor bredd och höjd som datan. Därefter interpolerades maskerna för att återuppnå datans originalstorlek. Alla tre typer av detaljer förekommer i täta kärnor då de har en rund form, så alla tre masker summerades. Därefter trösklades den summerade masken och de pixlar som hade höga förändringar i horisontella, vertikala och diagonala led lämnades kvar, alltså pixlarna på de platser där täta kärnor potentiellt var.

Den mask som skapades efter att wavelettransformen hade trösklats, visade inte exakt var de identifierade objekten var i den originala datan. Bilden hade skalats ner med en faktor två, och sedan tillbaka till original storlek, så koordinaterna i den riktiga bilden var inom en osäkerhetsmarginal. Osäkerhetsmarginalen ansågs vara inom en 2×2 matris. När koordinaterna från Waveletmetoden sedan användes för att beräkna radie och massa, användes istället koordinaten från den maximala värdet i matrisen, då den ansågs vara toppen.

4.4 Filtrering från standardavvikelser

Oskarp maskering genom gaussisk utjämning och fouriertransform, samt bildbehandling genom wavelettransform framhävde alla på olika sätt de detaljer i bilden som var förekommande bland täta kärnor, men inte om dessa detaljer hade en väsentligt högre intensitet än sin omgivning. Till exempel kunde områden med mycket brus där nivåerna i bilden förändrades mycket från punkt till punkt framhävas i bildmetoderna och därefter ta sig igenom trösklingen. Det blev alltså viktigt att filtrera bort de punkter som förmodligen var brus.

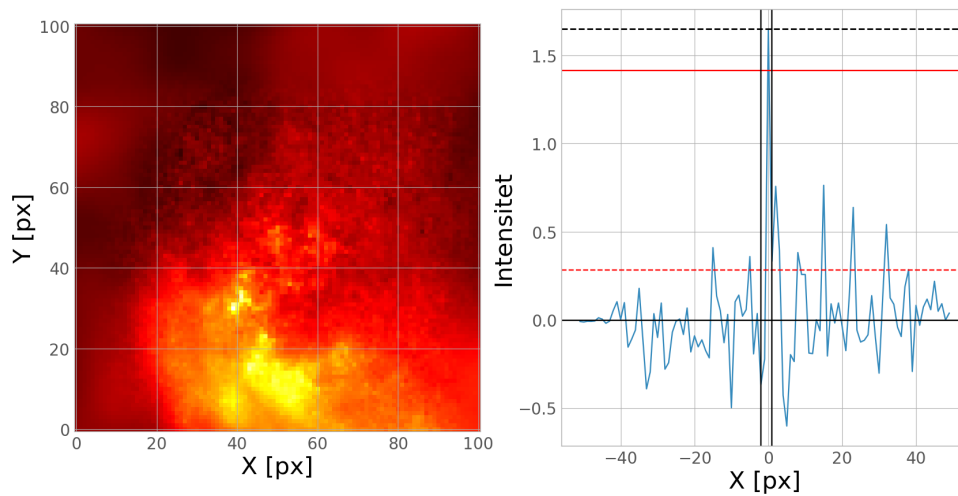
Maskerna som resulterade efter tröskling behövde även filtreras ytterligare. Detta eftersom maskerna i vissa fall innehöll flera punkter per tät kärna. För att endast ha en punkt per tät kärna inom masken, behölls enbart de punkter som är lokala maxima i originaldatan.

4.4.1 Villkor för klassificering av en tät kärna

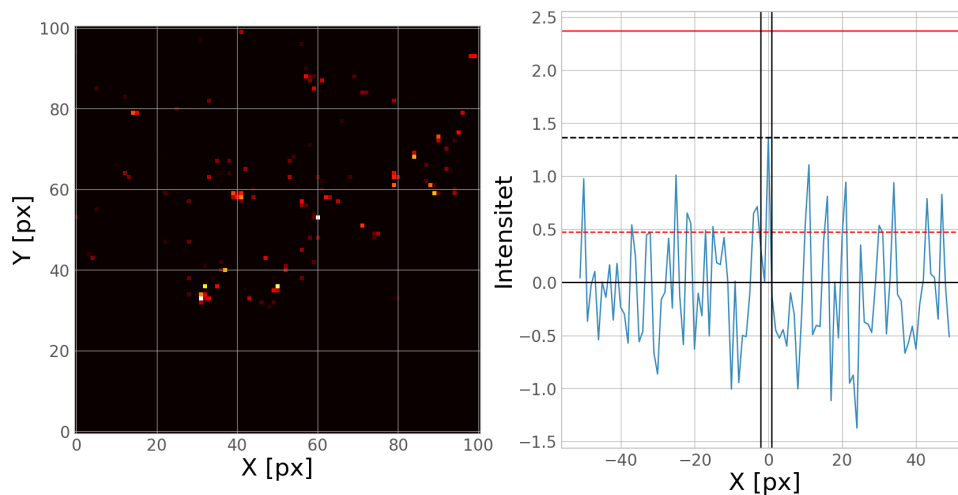
För att kunna avgöra om en given punkt är toppen av en tät kärna har programmet en algoritm med förinställda villkor för vad som ska klassas som en tät kärna. Först beräknades en baslinje kring en given punkt vilket syftade till medelvärdet av området runt punkten. Baslinjernas standardavvikelse beräknades sedan för att kunna avgöra om en förändring av intensiteten ansågs vara signifikant i förhållande till sin omgivning. Därefter villkorstestades topparna: om toppens högsta värde var fem gånger större än standardavvikelsen adderat med baslinjen, ansågs den vara signifikant högre än sin omgivning och behölls i masken, annars togs den bort. Eftersom detta är sista steget i databehandlingskedjan, blir detta i princip den slutgiltiga definitionen av en tät kärna i datasetet.

4.4.2 Beräkning av baslinjen

För att beräkna baslinjerna användes följande algoritm vid varje topp. En topp var mittpunkt till en kvadrat med sidlängd 61, låt oss kalla denna kvadrat för kvadrat 1. De värden som ingick i den hämtades och standardavvikelsen beräknades. Därefter skapades en ny kvadrat, låt oss kalla denna för kvadrat 2. Kvadrat 2 hade också toppen som mittpunkt, med en sidlängd på l , vilken till en början var tre. l ökades successivt, och för varje iteration beräknades standardavvikelsen hos snittet mellan kvadrat 1 och kvadrat 2. Den nya standardavvikelsen jämfördes mot den som beräknades i förgående iteration, och ifall de skiljde sig väldigt lite ansågs standardavvikelsen för baslinjen ha nåtts. Standardavvikelsen och medelvärdet i snittet mellan kvadrat 1 och kvadrat 2 tillhörde då baslinjen och kunde jämföras mot toppens värde. Exempel på uträknad baslinje och standardavvikelse kan ses i Figur 4.3 och 4.4.



Figur 4.3: Till vänster: en bild på en del av datan med ett identifierat objekt i mitten. Till höger: En graf som visar intensitetsförändringen i ett tvärsnitt i x -led från data behandlad med Gaussmetoden, med det identifierade objektet i mitten. Den horisontella svarta linjen visar baslinjen. Den streckade röda linjen visar standardavvikelsen för baslinjen + dess medelvärde. Den röda horisontella linjen visar 5 gånger standardavvikelsen för baslinjen + dess medelvärde. De två vertikala svarta linjerna visar området mellan dem som plockats bort för att beräkna standardavvikelsen. Den streckade svarta linjen visar vilken intensitet som analyseras.



Figur 4.4: Till vänster: en bild på en del av datan med ett ej identifierat objekt i mitten. Till höger: En graf som visar intensitetsförändringen i ett tvärsnitt i x -led från data behandlad med Gaussmetoden, med det ej identifierade objektet i mitten. Den horisontella svarta linjen visar baslinjen. Den streckade röda linjen visar standardavvikelsen för baslinjen + dess medelvärde. Den röda horisontella linjen visar 5 gånger standardavvikelsen för baslinjen + dess medelvärde. De två vertikala svarta linjerna visar området mellan dem som plockats bort för att beräkna standardavvikelsen. Den streckade svarta linjen visar vilken intensitet som analyseras.

4.5 Hantering av artefakter

Datan innehåller även artefakter som beskrivs i Avsnitt 2.3 som behövde särskiljas från täta kärnor. Nedan beskrivs metoden för hur dessa hittas och hur de delas upp i en skala mellan säkra och osäkra artefakter.

4.5.1 Identifiering av artefakter

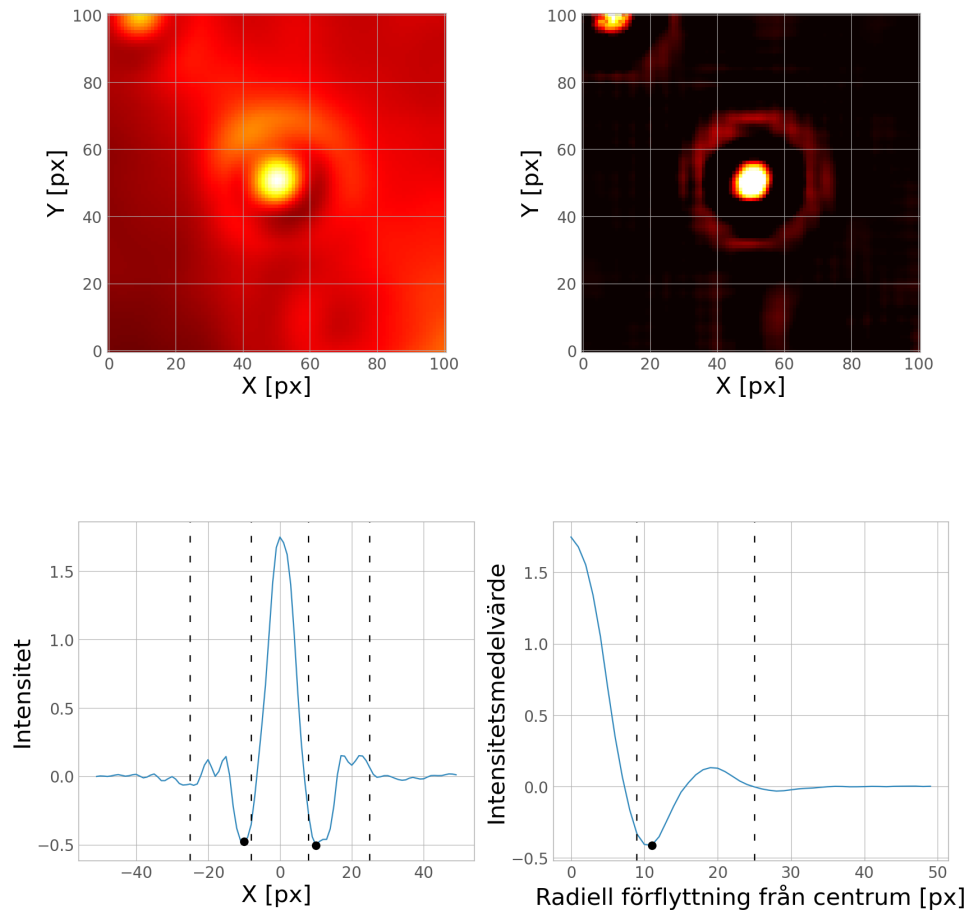
Artefakter hade ofta ett karakteristiskt drag när de plottades som ett tvärsnitt av intensitetsvärden av datan, den liknade en sinc-funktion, se Figur 4.5. Följdes toppen från höger och vänster nerför kurvan kom först en minimipunkt inom ett visst intervall, därefter en maximipunkt inom ett annat intervall, för att slutligen stabiliseras. Om ett objekt hade minimipunkter inom 8 och 25 pixlar från toppen, klassades den som en artefakt.

En annan metod som användes utnyttjade den cirkulära symmetrin hos artefakterna. För att göra detta valdes cirkelns mittpunkt till toppens koordinater och därefter beräknades medelvärdet på cirkelkonturen vid radien r . Radien på cirkelkonturen r ökades successivt från 0 till 50 pixlar och medelvärdet beräknades vid varje iteration. På så vis kunde sedan intensiteten plottas mot radien och även här ansågs objektet vara en artefakt om det första lokala minimat låg mellan 8 och 25 pixlar från cirkelns mitt. Klassade denna metod eller den förra att en punkt var en artefakt, bestämdes punkten som en artefakt.

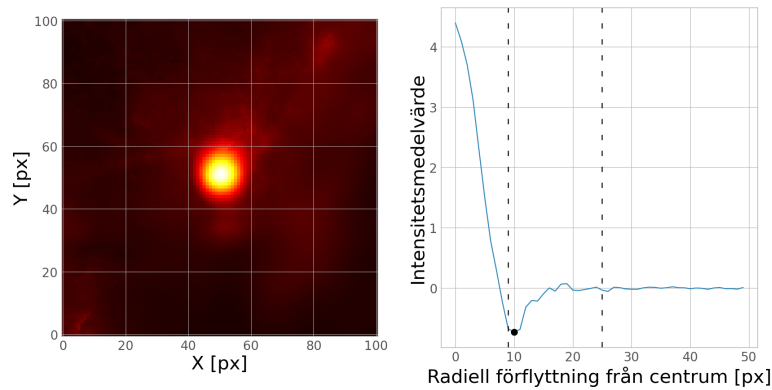
4.5.2 Säkerhetsskala

För att förmedla hur säker programmet var att en klassificerad artefakt faktiskt var en artefakt, introducerades en säkerhetsskala. Skalan var från 1-3, där 3 motsvarade en väldigt säker artefakt, se Figur 4.5, 2 en säker, se Figur 4.6, och 1, se Figur 4.7 en osäker artefakt. Klassades en artefakt på det sätt som nämns i 4.5.1, ansågs den åtminstone vara nivå 1 eller 2 beroende på om den hämtats med Fourier- eller Waveletmetoden respektive Gaussmetoden. Anledningen till denna indelning var att den senare metoden ansågs mer träffsäker än de andra. För att klassas som nivå 3 behövde dess radiograf likna Figur 4.5 och gå igenom följande krav:

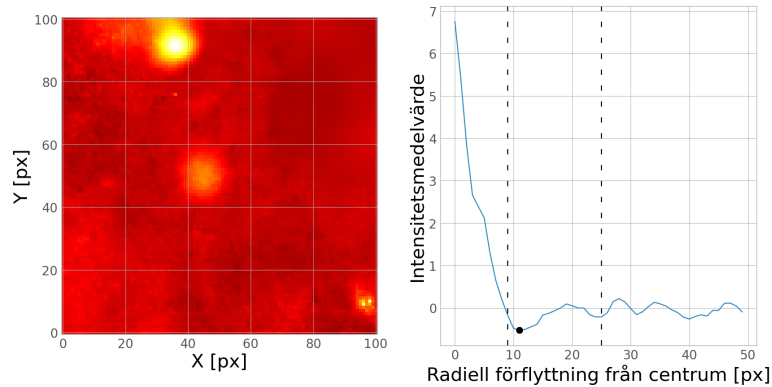
1. Det första lokala minimat behövde ha en lägre amplitud än baslinjen och minst på ett avstånd ekvivalent med artefaktens amplitud dividerat med nio. Avståndet behövde även vara högre än fem gånger baslinjens standardavvikelse.
2. Det första lokala maximat behövde vara inom 30 pixlar från den centrala pixeln och även högre än fem gånger baslinjens standardavvikelse adderat med baslinjen.



Figur 4.5: Identifierad artefakt med säkerhetsnivå 3. Bilderna i det övre vänstra och högra hörnet visar objektet i originaldatan respektive oskarp maskerad data med gaussisk utjämning. Graferna i de nedre vänstra och högra hörnet visar intensitetsförändringen i ett tvärsnitt i x -led respektive intensitetsmedelvärdet kring centrumpunkten, där båda är i den oskarp maskerade datan med gaussisk utjämning. De streckade linjerna visar gränserna som första minimivärdena, markerade med svarta punkter, behöver befinna sig i för att klassas som artefakt. Detta objekt blev klassat som artefakt eftersom minimipunkterna är inom dess respektive gränser, och nivå 3 eftersom det är en tydlig nedgång och därefter uppgång, nedgång igen och slutligen stabilisering i den nedre högra grafen.



Figur 4.6: Identifierad artefakt med säkerhetsnivå 2. Bilden till vänster visar objektet i originaldatan. Grafen till höger visar det radiella intensitetsmedelvärdet kring centrumpunkten, i den oskarp maskerade datan med gaussisk utjämning. De streckade linjerna visar gränserna det första minimivärdet, markerat med en svart punkt, behöver befinna sig i för att klassas som artefakt. Detta objekt blev klassat som artefakt eftersom minimipunkten är inom gränsen, men uppfyllde inte alla krav för nivå 3.



Figur 4.7: Identifierad artefakt med säkerhetsnivå 1. Bilden till vänster visar objektet i originaldatan. Grafen till höger visar det radiella intensitetsmedelvärdet kring centrumpunkten, i wavelettransformerad data. De streckade linjerna visar gränsen det första minimivärdet, markerat med en svart punkt, behöver befinna sig i för att klassas som artefakt. Detta objekt blev klassat som artefakt eftersom minimipunkten är inom gränsen, men uppfyllde inte alla krav för nivå 3. Den blev heller ej nivå 2 eftersom den inte identifierades genom Gaussmetoden. Visuellt ser objektet inte ut som en typisk artefakt.

4.6 Artificiella kärnor och artefakter

För att bättre förstå hur väl vald en metod och dess parametrar var konstruerades en testmetod med artificiella täta kärnor och artificiella artefakter. Testmetoden fungerade genom att sätta in konstruerad data i PROMISE-datan och spara x - och y -positionerna där de placerades i datan. Ett program kunde sedan testas mot den manipulerade datan för att undersöka hur väl metoden hittar de konstruerade kärnorna och artefakterna. Andelen av den konstruerade datan som hittades användes som ett mått på hur väl metoden hittar täta kärnor och artefakter.

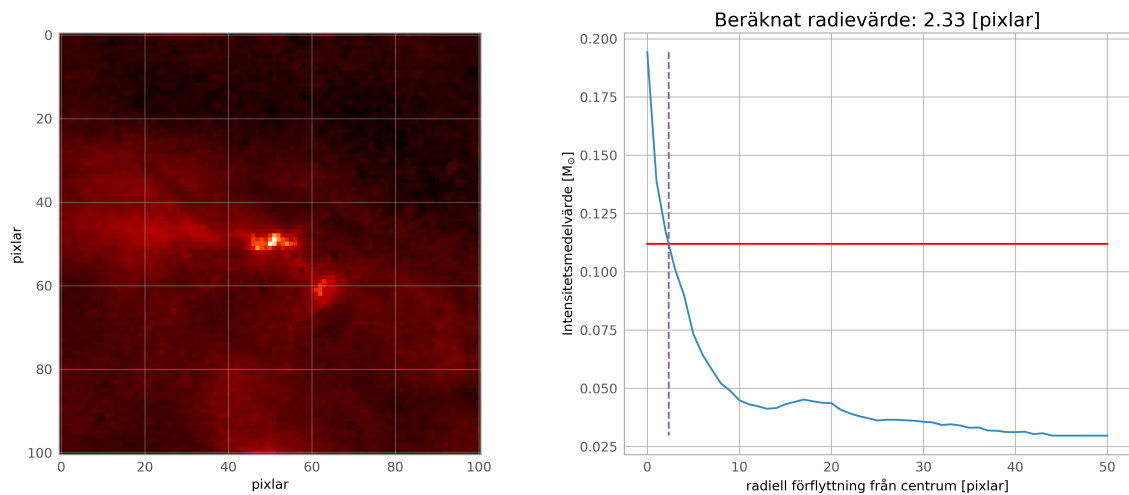
Desto bättre de konstruerade kärnorna och artefakterna representerar de riktiga ju säkrare blir denna metod. För att få rimliga approximationer av täta kärnor och artefakter undersöktes relevanta egenskaper som sedan efterliknades. Den mest relevanta egenskaperna är storlek, form och intensitet. Artefakterna har till skillnad från de täta kärnorna en relativt bestämd storlek, form och intensitet och analyserades därför inte i samma utsträckning.

4.6.1 Massa-radieförhållande för täta kärnor

De täta kärnorna som existerar i PROMISE-datan har olika storlek och olika intensitet, vilket leder till att olika kärnor också har olika massa. Att hitta ett ungefärligt mått på hur många täta kärnor av olika storlek och intensitet som finns i datan ger ett mått på vilken fördelning av storlekar och massa som bör genereras.

Radievärdet bestämdes genom att hitta avståndet från det maximala värdet till punkten med hälften av det maximala värdet, och sätta det som radie. Detta är en metod som ofta kallas "Half width half maximum"-metoden (HWHM-metoden) och kommer refereras till som HWHM-metoden i resten av rapporten. Ett radiellt intensitetsmedelvärde räknades ut för alla punkter kring centrumpunkten fram tills ett valt stopp, på samma sätt som i Avsnitt 4.5, sedan uppskattades en baslinje till det minsta värdet på området. Baslinjen subtraherades från alla värden på intervallet för att kunna hitta en radie med HWHM-metoden. Där intensitetsmedelvärdet var lika med halva toppen identifierades det som radien. Figur 4.8 visar det radiella intensitetsmedelvärdet kring en identifierad kärna samt vart halva maxvärdet hittades.

Efter manuell granskning observerades det att metoden överskattade radien för vissa kärnor, ofta var det små intensiva regioner i närheten av större objekt som gjorde att radien överskattades, se Figur 4.9. Av denna anledning sattes en övre gräns för radie på 20 pixlar (approximativt 0.4 pc) då ingen identifierad kärna efter manuell inspektion uppvisade högre radievärde. Ifall metoden hittade en radie på mer än 20 pixlar så sattes den täta kärnans radievärde till 0 och räknades som ej identifierbar med HWHM-metoden, detsamma gjordes för kärnor som identifierades precis i utkanten på PROMISE-datan med högre beräknad radie än avstånd från kantlinjen i antingen x - eller y -led.



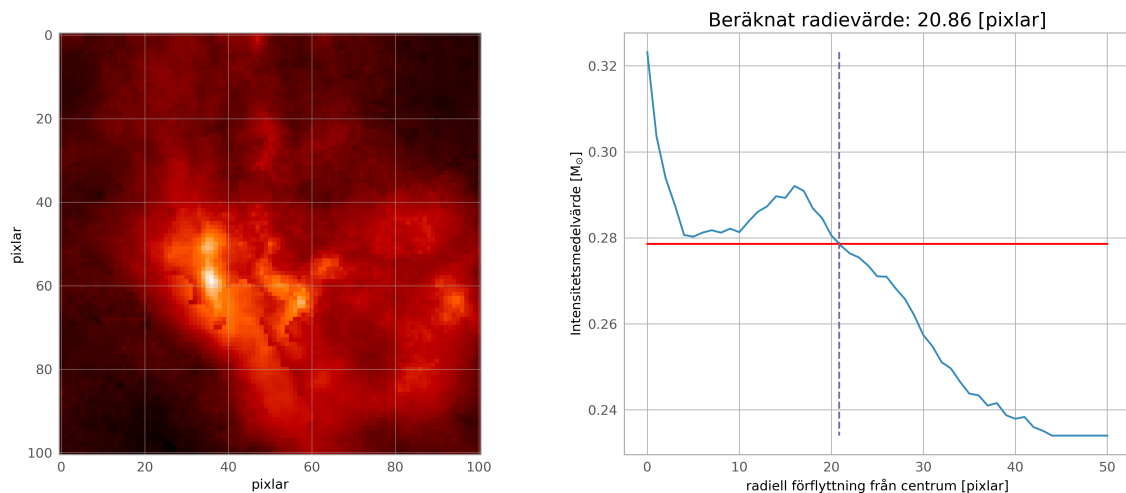
Figur 4.8: Bild över en identifierad tät kärna och det radiella intensitetsmedelvärdet kring centrumpunkten i bilden. Den röda tillsammans med den streckade linjen visar var halva maxvärdet hittas. Radien uppmäts till 2 pixlar.

För att beräkna massa skapades en algoritm för att summera alla pixlars intensitetsvärden innanför det beräknade radievärdet. Det summerade intensitetsvärdet multiplicerades sedan med 0.0081 för att få resultatet i enheten solmassa, se avsnitt 2.2. Resultatet på radie-massaförhållandet för täta kärnor kan ses i Figur 4.10.

4.6.2 Konstruktion och insättning av artificiella kärnor

Formen på de artificiella kärnorna approximerades till en tvådimensionell symmetrisk gaussisk funktion med standardavvikelse 0.25 sigma för att få en liknande intensitetsspic som de täta kärnorna i datan. De artificiella täta kärnorna konstruerades utifrån en gaussisk funktion med storlek på 1001 pixlar. Funktionen radie och massa beräknades med samma metoder som beskrivet i Avsnitt 4.6.1. För att se till att värdena längs kanterna när en artificiell kärna sätts in inte påverkas för mycket vid höga intensitetsskalningar applicerades även en cirkulär binär mask på funktionen för att sätta värden utanför en cirkel med radien 500 pixlar till noll.

Utifrån den framtagna radie-massafördelningen framtagna i Avsnitt 4.6.1 valdes det slumpmässigt uniformt ut radie-massapar och den gaussiska funktionen skalades om för varje valt par. Skalningen skedde i både storlek och intensitetsvärde. Den omskalade gaussiska funktionen sparades i en lista. Detta upprepades i ett förbestämt antal iterationer. Listan med storlekssskalade gaussiska objekt sattes sedan in i datan med slumpmässiga x- och y- koordinater. Koordinaterna sparades i en separat lista för att sedan kunna testa ifall programmet korrekt kunde hitta de insatta objekten. Fördelningen på insatta täta kärnor kan ses i Figur 4.11 och ett exempel på insatta täta kärnor i Figur 4.12



Figur 4.9: Bild som visar en identifierad kärna som får ett överestimerat radievärde med HWHM-metoden, notera att den funna kärnan ligger precis i mitten av den vänstra bilden. Den röda och den streckade linjen visar vart halva maxvärdet hittas. Närliggande områden påverkar HWHM-metoden och radievärdet mäts upp till 21 trots att radien ser ut att vara betydligt mindre.

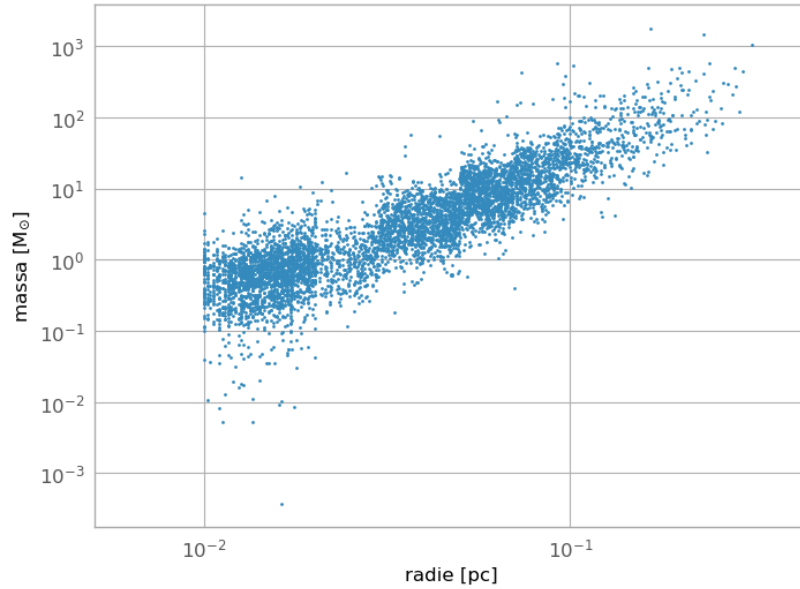
4.6.3 Konstruktion och insättning av artificiella artefakter

Konstruktionen av artificiella artefakter utnyttjar artefakternas symmetri och storlek. Som beskrivet i avsnitt 2.3 efterliknar artefakten en tvådimensionell sinc-funktion. Därför valdes det att approximera artefakten enligt Ekvation 4.1 där k är en skal-faktor.

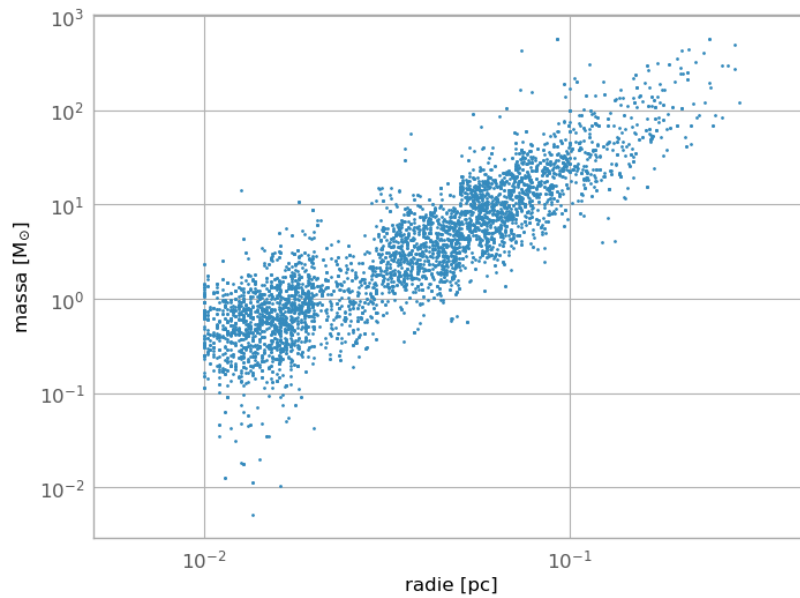
$$\text{sinc}(x, y) = \frac{\sin(\pi kx) \sin(\pi ky)}{\pi^2 k^2 xy} \quad (4.1)$$

Till skillnad från de täta kärnorna har artefakter en relativt bestämd storlek och varierar inte mycket från detta. Som referensvärde mättes en artefakt till att approximativt vara 50 pixlar bred med en mitt på 15 pixlar. Sinc-funktionen anpassades sedan efter dessa observationer och $k = 1/9$ för en bra approximation. Det lades sedan på en cirkulär mask för att få bort värden längs ytterkanten. Resultatet går att se i Figur 4.13.

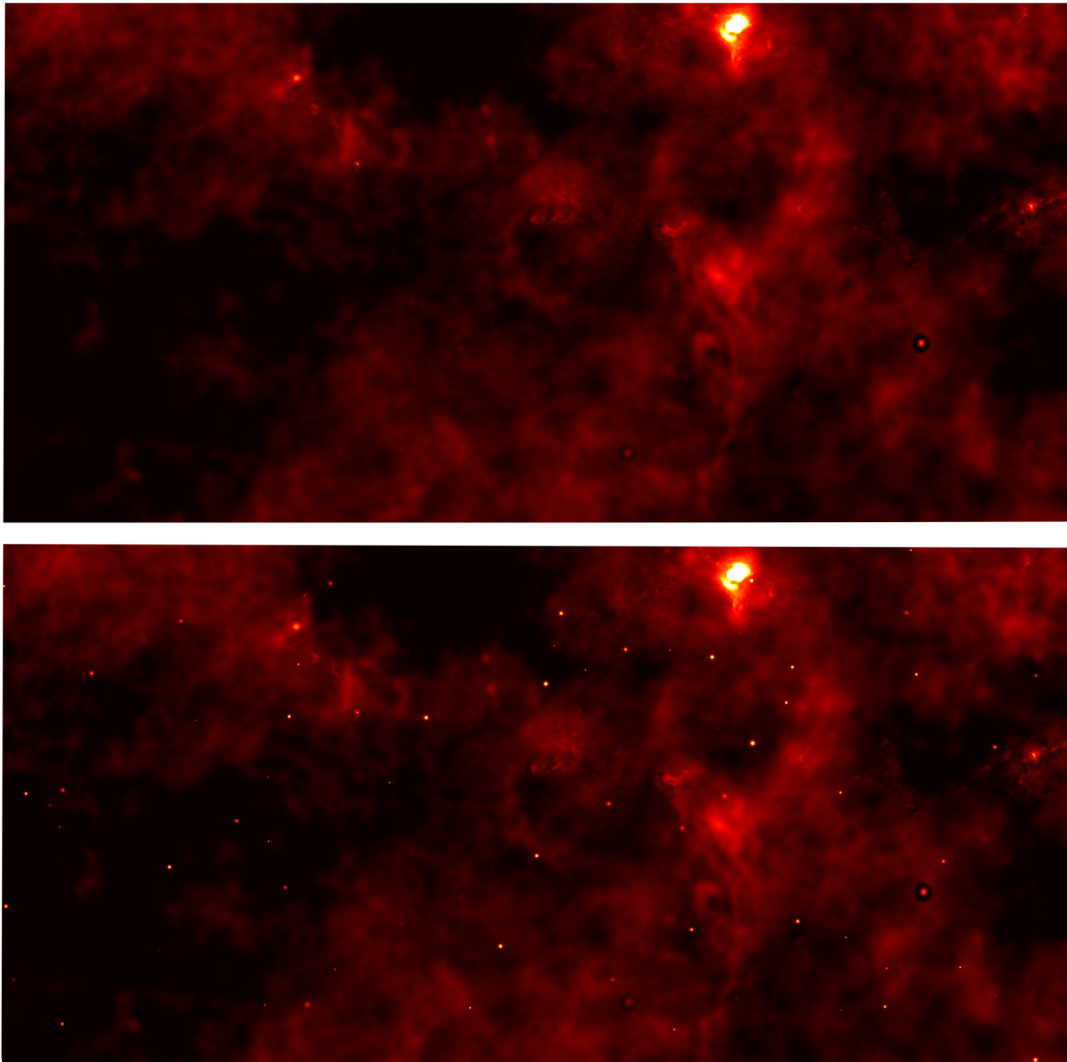
Varje konstruerad artefakt multiplicerades sedan med ett uniformt slumpmässigt värde mellan 25 och 75 innan insättning för att få rimliga intensitetsvärden. Artefakterna läggs sedan in i datan med samma metod som de artificiella kärnorna. Figur 4.14 visar ett antal artificiella artefakter insatta i PROMISE-datan. I Appendix A.3 presenteras de valda parametrarna för den artificiella datan i det slutgiltiga programmet.



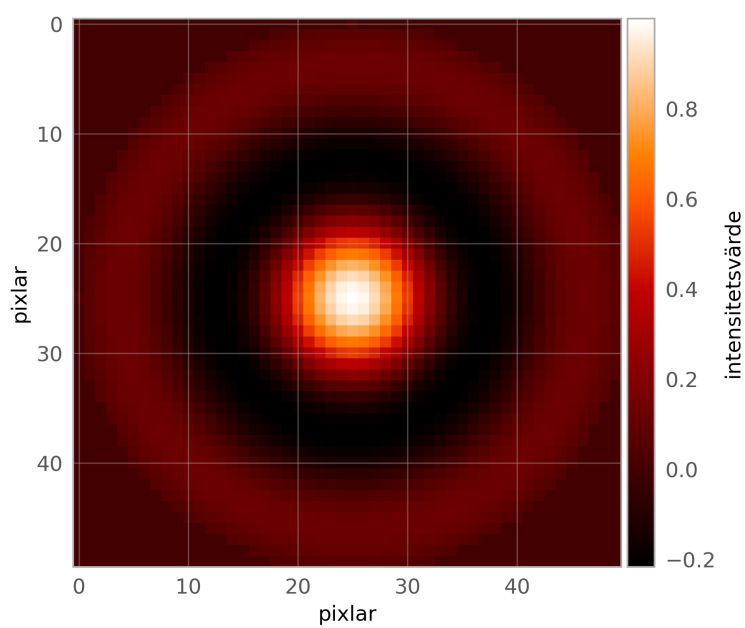
Figur 4.10: Massa-radie förhållande för täta kärnor identifierade genom Gaussmetoden, från horisontellt pixelvärde 0 till 40 000. Radie uträknad med HWHM-metod. En övre gräns har satts på 25 pixlar (0.4 pc).



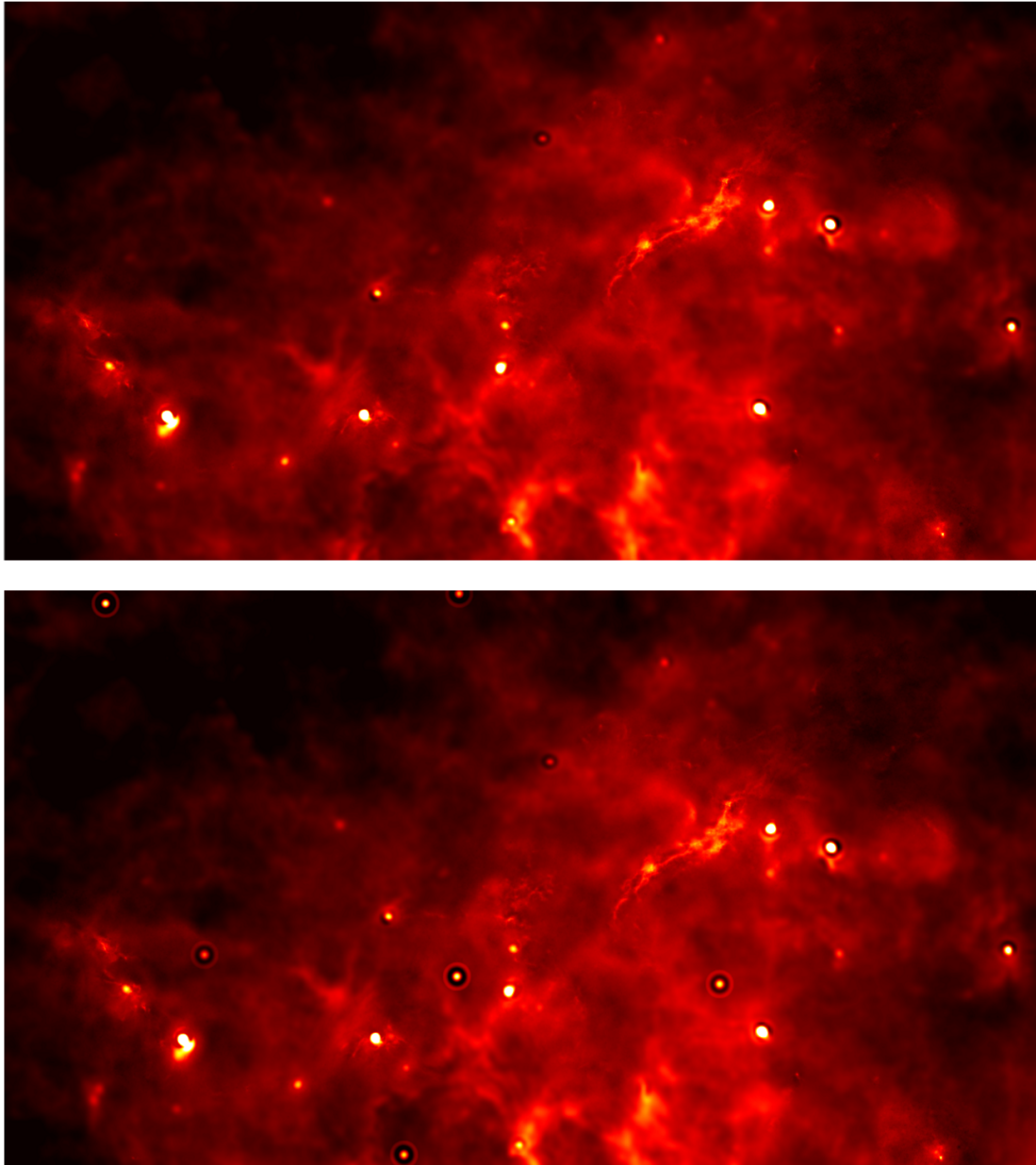
Figur 4.11: Massa-radie förhållande för artificiella kärnor. Radie uträknad med HWHM-metod.



Figur 4.12: Övre bild visar ett område i PROMISE-datan utan manipulation. Nedre bilden visar samma område med tillagda artificiella artefakter.



Figur 4.13: Bilden visar en artificiellt konstruerad artefakt anpassad efter en sinc-funktion med en cirkulär mask applicerad, därav är lamterna nollsatta. Axlarna visar storleken på artefakten mätt i pixlar. Färgskalan visar intensitetsvärden för pixlarna.



Figur 4.14: Övre bild: Område i originala PROMISE-datan som innehåller flertal artefakter med olika intensitetsnivåer. Nedre bild: Artificiella artefakter insatta i samma område för att illustrera likheter och skillnader.

4.7 Sammanställning av program

Nedan beskrivs hur det fullständiga programmet är sammanställt med de olika steg som beskrivits ovan. Den fullständiga koden går att hitta genom Appendix A.4.

I första steget definierades de parametrar som kom att användas i programmet, sedan valdes det ifall programmet skulle köras med eller utan att generera artificiell data.

I nästkommande steg påbörjades datamanipuleringen genom att utföra antingen Gaussmetoden, Fouriermetoden, eller Waveletmetoden. Efter detta fortskred programmet likadant oavsett vald metod, genom att tröskla den genererade masken och sedan filtrera datan utifrån masken i kombination med villkorstestning av topparna enligt 4.4.1. Sedan körs artefaktidentifiering, samt uträkning av radie och massa för de identifierade täta kärnorna, i den ordningen, därefter sparades katalogen. Pseudokod för programmet kan ses i Appendix A.1.

De bästa framtagna katalogerna med varje metod sammanfogades sedan till en större gemensam katalog som inkluderar både artefakter och täta kärnor, för att kunna enkelt hantera datan och observera likheter och skillnader mellan metoderna. Utöver kolumner för position, radie och massa, finns även kolumner som visar vilka metoder som identifierade punkten, samt en kolumn som visar ifall punkten är en artefakt enligt 4.5.2. Formatet går att se i tabell 5.1.

Ifall två metoder identifierade kärnor inom 7 pixlar av varandra i x- och y-led, räknades de till samma artikel i katalogen.

5. Resultat

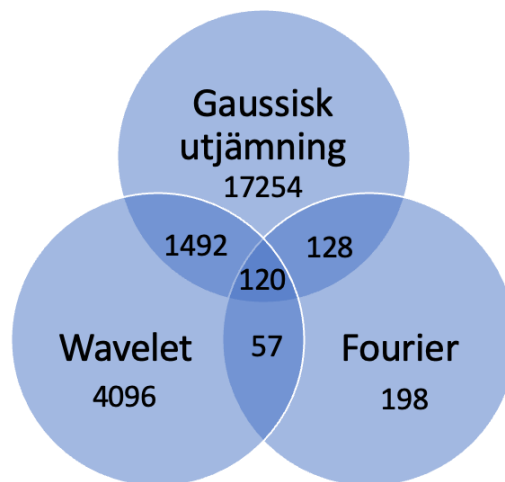
I följande kapitel presenteras projektets resultat för de olika metoderna. Det presenteras ett utdrag ur den sammanställda katalogen, de täta kärnornas radie- och massafördelning samt med vilken säkerhet programmet anses kunna identifiera täta kärnor och artefakter utifrån skapade testmetoder.

Den framtagna katalogen, se Tabell 5.1 för ett utdrag, består av en lista på 32 737 objekt, varav 23 347 klassats som täta kärnor och 9 390 klassats som artefakter. Av de klassade artefakterna var 2100 väldigt säkra (nivå 3), 1953 säkra (nivå 2) och 5337 osäkra (nivå 1). Den metod som hittade flest täta kärnor var Gaussmetoden som hittade 18 994 täta kärnor, näst mest hittade Waveletmetoden med 4 273 täta kärnor och minst hittade Fouriermetoden som hittade 503 täta kärnor. Överlappet mellan de olika metoderna kan ses i Figur 5.1.

Tabell 5.1: Urval ur katalog för Gaussmetoden, Fouriermetoden och Waveletmetoden. Koordinaterna i PROMISE-datan presenteras i x - och y -led med enheten pixel (px). Artefaktkolumnen visar en säkerhetsnivå på upphittade artefakten där 3 är väldigt säkert, 2 är säkert och 1 är osäkert. Inom parentes i artefaktkolumnen presenteras vilken eller vilka metoder som har identifierat artefakten. Sista kolumnen visar vilka metoder som har hittat objektet; g för Gaussmetoden, f för Fouriermetoden och w för Waveletmetoden.

Index	x [px]	y [px]	Radie [pc]	Massa [M_{\odot}]	Artefakt	Metod
2870	111699	1426	0,02	0,95	-	$g w$
3222	63231	1524	0,17	45,68	-	f
135	88344	339	0,11	4,65	3 (f)	f
266	75382	490	0,11	15,9	-	w
5900	73008	2006	0,15	754,71	3 (w)	w
187	5268	414	0,01	-0,02	-	w
13	16646	80	0,01	0,42	-	g
447	109890	612	0,01	0,06	-	w
2870	111699	1426	0,02	0,95	-	$g w$
26329	48598	4112	0,07	25,69	-	$w f$
3148	77327	1503	0,12	32,48	-	$g w$
135	88344	339	0,11	4,65	3 (f)	f
187	5268	414	0,01	-0,02	-	w
447	109890	612	0,01	0,06	-	w
287	59443	516	0,06	5,75	1 (f)	f

Några exempel från katalogen på tydliga täta kärnor i olika storlekar går att se i Figur 5.2, Figur 5.3 och Figur 5.4, fullständiga katalogen går att hitta i Appendix A.4. Ett exempel på en tydlig artefakt som har identifierats med Fouriermetoden och



Figur 5.1: Ett venndiagram över antalet identifierade täta kärnor för Gaussmetoden, Waveletmetoden och Fouriermetoden.

Gaussmetoden går att se i Figur 5.5. Det finns också exempel på när programmet felklassificerar objekt. Ett exempel visas i 5.6 där Waveletmetoden har identifierat potentiellt brus i datan som en tät kärna. Ytterligare ett exempel kan ses i Figur 5.7 där ett potentiellt fel i datan klassas som en tät kärna och i 5.8 där en tät kärna felaktigt klassificerats som en artefakt.

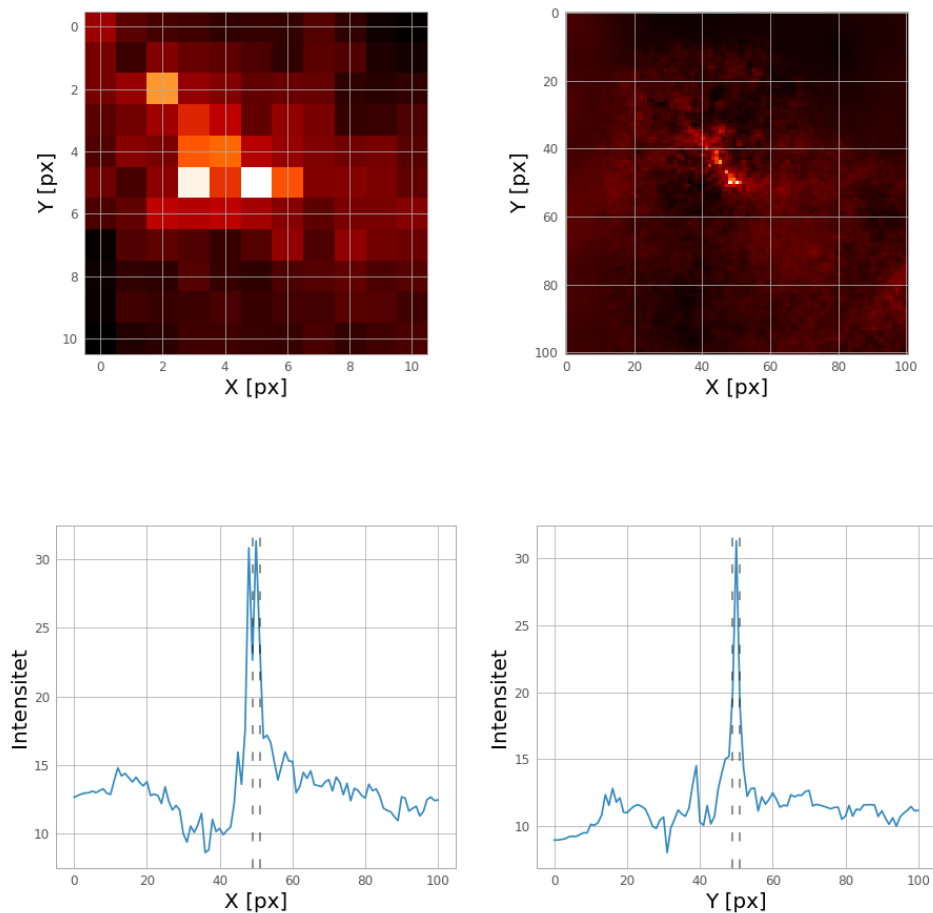
Täta kärnor hittades i storlekar mellan 0,01 pc och 0,21 pc där färre täta kärnor hittades ju större radien blev. Radiefördelningen på de identifierade täta kärnorna går att se i Figur 5.9. De täta kärnorna hade massa mellan $-0.04 M_{\odot}$ och $550,47 M_{\odot}$, där de flesta har värden mellan $0.1 M_{\odot}$ och $5 M_{\odot}$, se Figur 5.10.

En säkerhetsgrad har tagits fram för de olika bildanalytiska metoderna genom de konstruerade testmetoderna i Avsnitt 4.6. I Tabell 5.2 presenteras hur stor andel av de artificiella kärnorna och artefakterna som identifierades med respektive bildanalytisk metod. Resultatet av testmetoden visar att Gaussmetoden hittar flest artificiella kärnor men har svårt att identifiera vilka av dessa som är artefakter. Fouriermetoden missar många kärnor men är bäst på att identifiera artefakterna medan Waveletmetoden ligger däremellan.

Tabell 5.2: Andel identifierade artificiella kärnor och artefakter med de olika bildanalytiska metoderna i det slutgiltiga programmet.

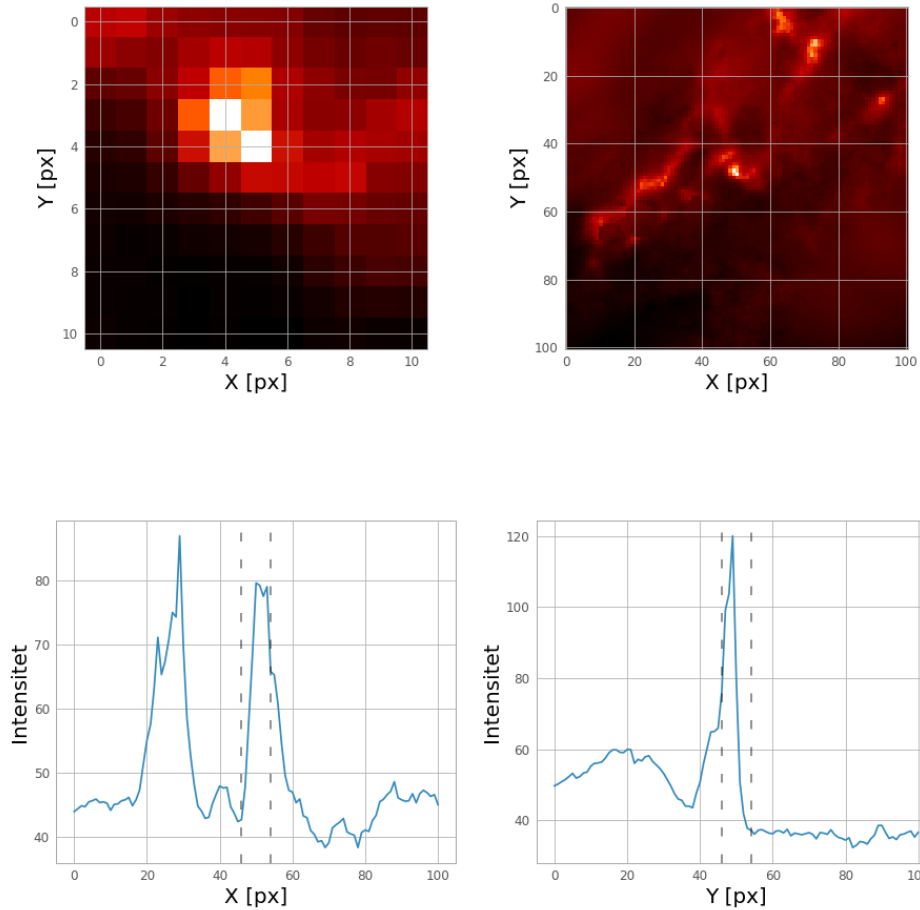
Bildanalytisk metod	Andel identifierade artificiella kärnor	Andel identifierade artificiella artefakter
Gaussmetoden och oskarp maskering	82,8%	70,2%
Fouriermetoden och oskarp maskering	33,7%	94,6%
Waveletmetoden	71,2%	92,4%

Gaussisk utjämning, Wavelet, x: 111699px, y: 1426px, radie: 0.02pc, massa: $0.95M_{\odot}$



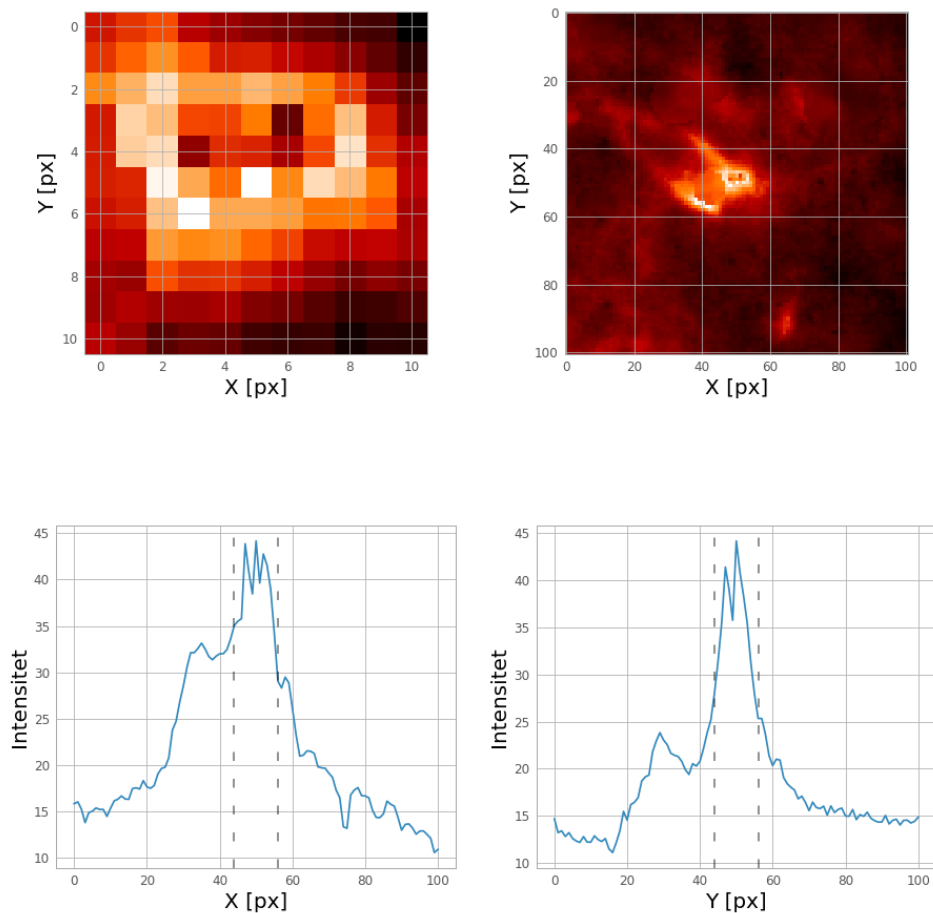
Figur 5.2: Ett exempel på en tät kärna som identifierats av Gauss- och Waveletmetoden. De övre bilderna visar den identifierade kärnan i mitten, med 10 respektive 100 pixlar i bredd och höjd. Båda dessa bilder är centrerade runt det som programmet uppfattat som en tät kärnas centrum eller högsta värde. De nedre graferna visar intensitetsförändringen i ett tvärsnitt genom kärnan i x - respektive y -led. Den streckade svarta linjen visar vad som innefattar diametern och därmed inom vilken radie som massan beräknas.

Wavelet, Fourier, x: 48598px, y: 4112px, radie: 0.07pc, massa: 25.69M_⊙



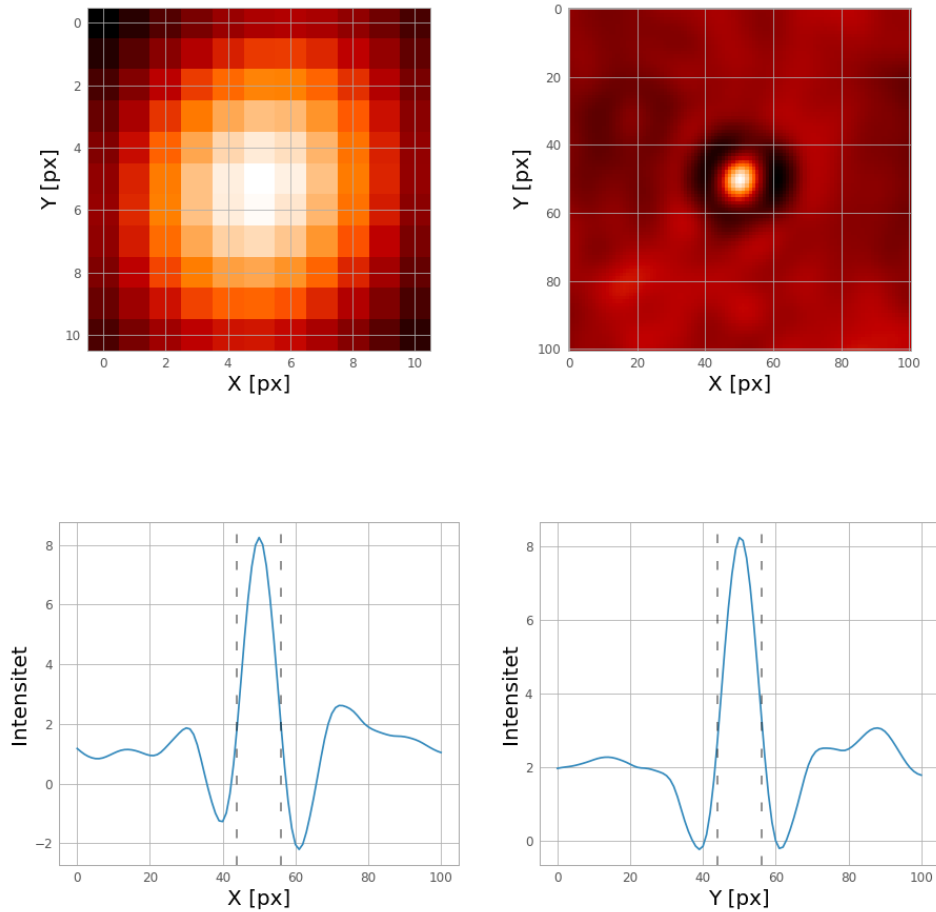
Figur 5.3: Ett exempel på en medelstor tät kärna som identifierats av Wavelet- och Fouriermetoden. De övre bilderna visar den identifierade kärnan i mitten, med 10 respektive 100 pixlar i bredd och höjd. Båda dessa bilder är centrerade runt det som programmet uppfattat som en tät kärnas centrum eller högsta värde. De nedre graferna visar intensitetsförändringen i ett tvärsnitt genom kärnan i x - respektive y -led. Den streckade svarta linjen visar vad som innefattar diametern och därmed inom vilken radie som massan beräknas.

Gaussisk utjämning, Wavelet, x: 77327px, y: 1503px, radie: 0.12pc, massa: 32.48M_☉



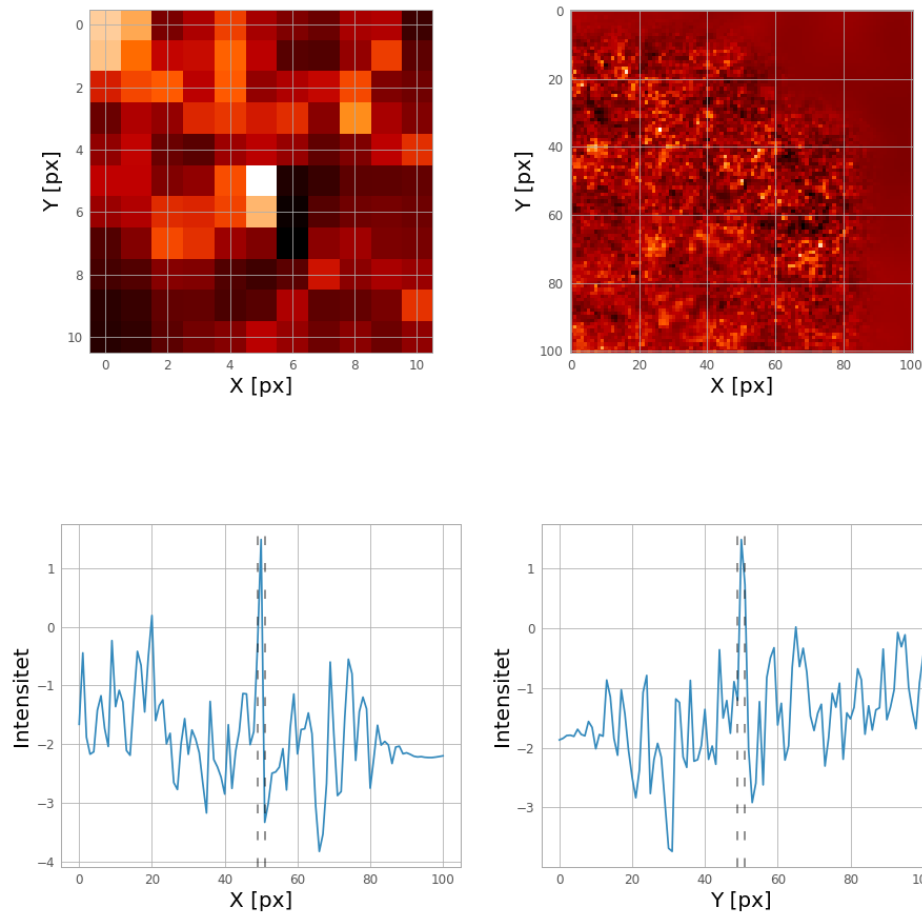
Figur 5.4: Ett exempel på en större tät kärna som identifierats av Gauss- och Waveletmetoden. De övre vänstra och högra bilderna visar den identifierade kärnan i mitten, med 10 respektive 100 pixlar i bredd och höjd. Båda dessa bilder är centrerade runt det som programmet uppfattat som en tät kärnas centrum eller högsta värde. De nedre graferna visar intensitetsförändringen i ett tvärsnitt genom kärnan i x - respektive y -led. Den streckade svarta linjen visar vad som innefattar diametern och därmed inom vilken radie som massan beräknas.

Fourier, x: 88344px, y: 339px, radie: 0.11pc, massa: $4.65M_{\odot}$, artefakt säkerhet: 3



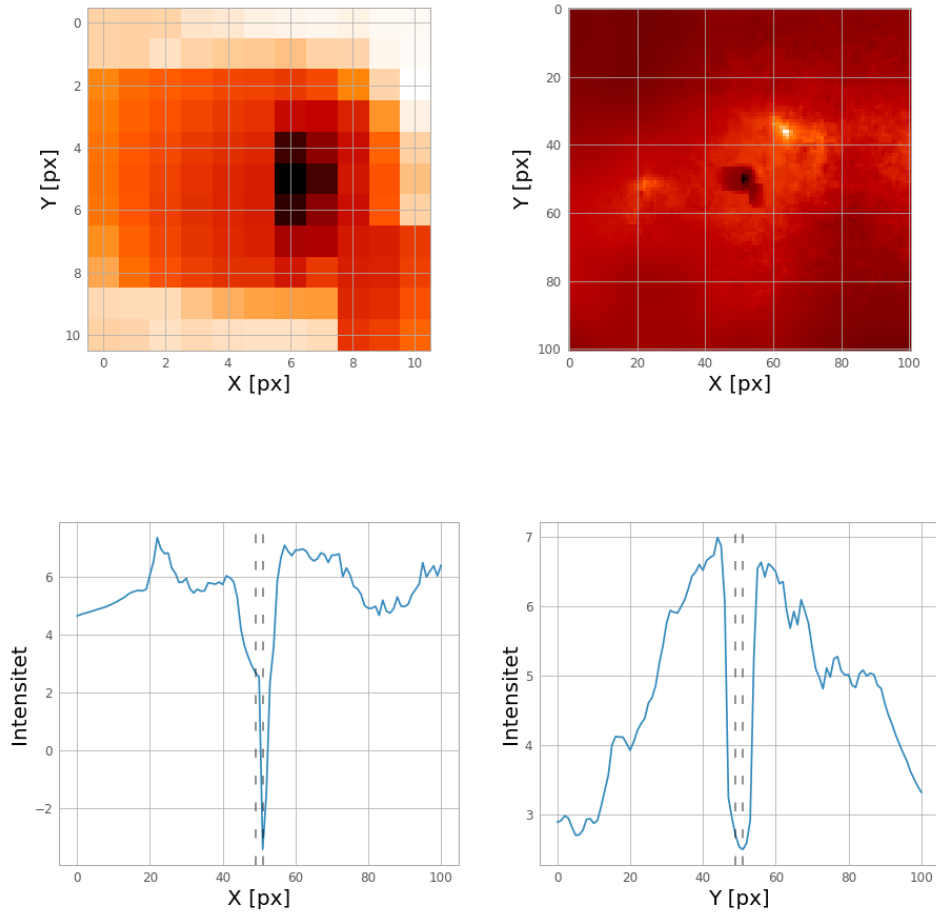
Figur 5.5: Ett exempel på en artefakt som identifierats av Fouriermetoden och Gaussmetoden. De övre bilderna visar den identifierade artefakten i mitten, med 10 respektive 100 pixlar i bredd och höjd. De nedre graferna visar intensitetsförändringen i ett tvärsnitt genom artefakten i x - respektive y -led. Den streckade svarta linjen visar vad som innefattar diametern och därmed inom vilken radie som massan beräknas.

Wavelet, x: 5268px, y: 414px, radie: 0.01pc, massa: $-0.02M_{\odot}$



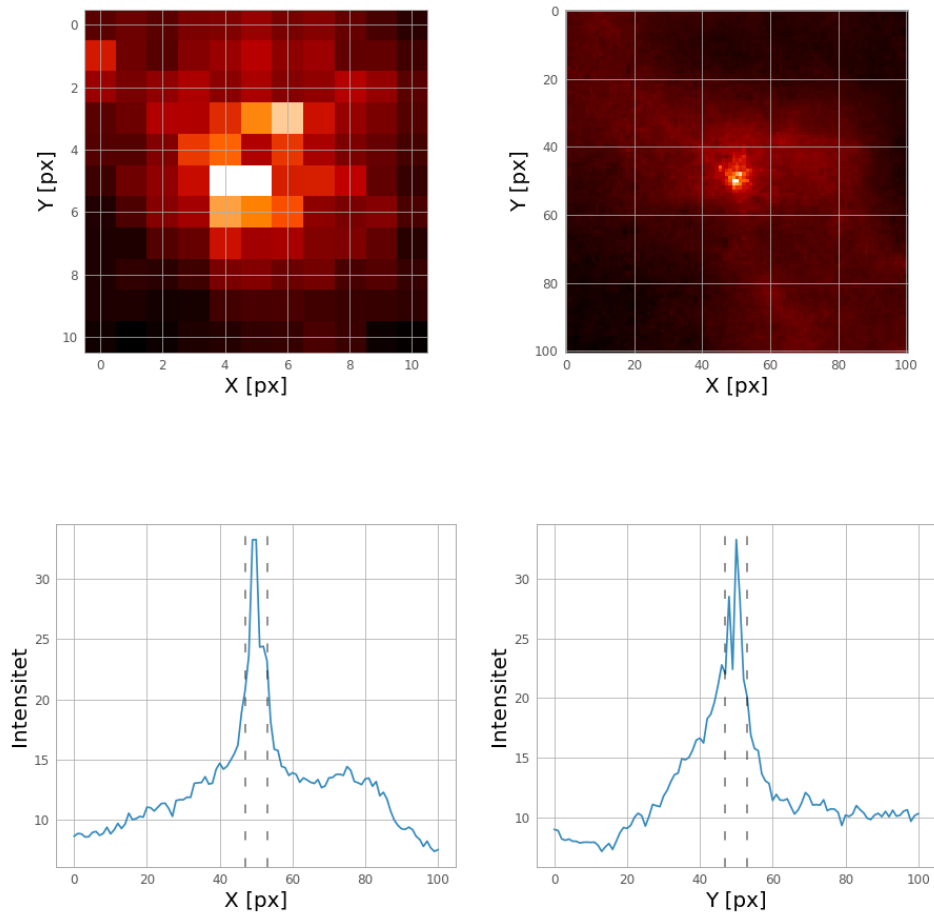
Figur 5.6: Ett exempel på en identifierad tät kärna med Waveletmetoden, men som på grund av dess omgivning potentiellt är brus. De övre bilderna visar den identifierade kärnan i mitten, med 10 respektive 100 pixlar i bredd och höjd. De nedre graferna visar intensitetsförändringen i ett tvärsnitt genom kärnan i x - respektive y -led. Den streckade svarta linjen visar vad som innefattar diametern och därmed inom vilken radie som massan beräknas.

Wavelet, x: 109890px, y: 612px, radie: 0.01pc, massa: 0.06M_☉

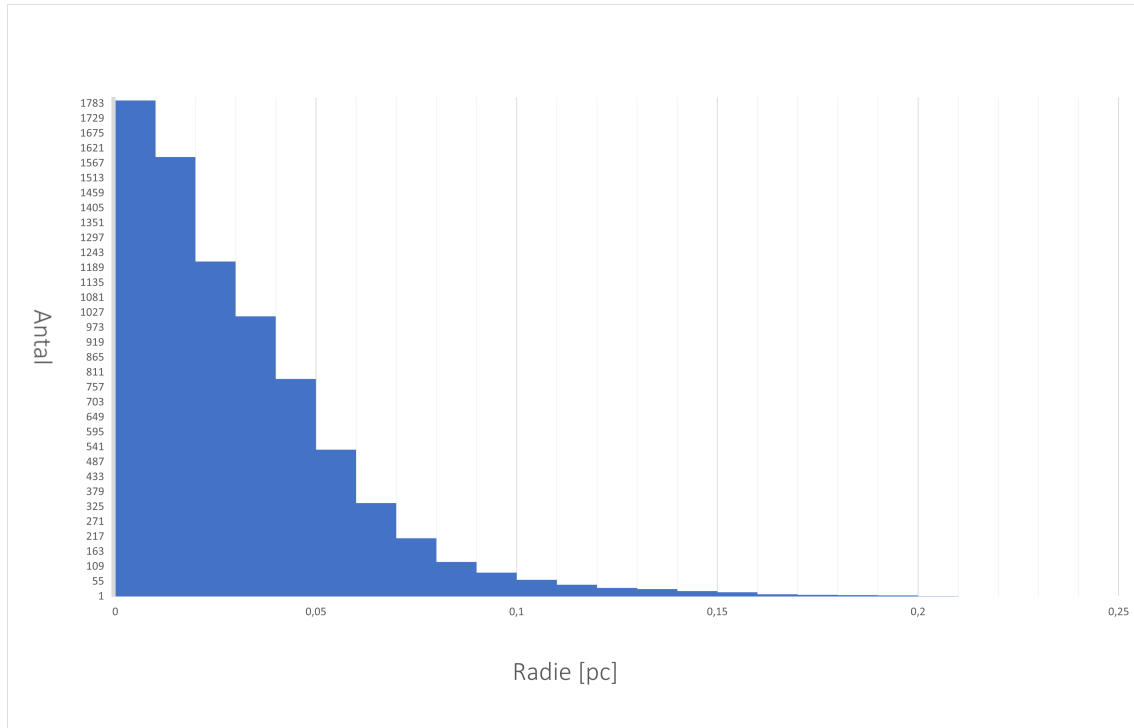


Figur 5.7: Ett exempel på en identifierad tät kärna med Waveletmetoden, men som är ett möjligt fel i datan. De övre bilderna visar den felidentifierade kärnan i mitten, med 10 respektive 100 pixlar i bredd och höjd. De nedre graferna visar intensitetsförändringen i ett tvärsnitt genom kärnan i x - respektive y -led. Den streckade svarta linjen visar vad som innefattar diametern och därmed inom vilken radie som massan beräknas.

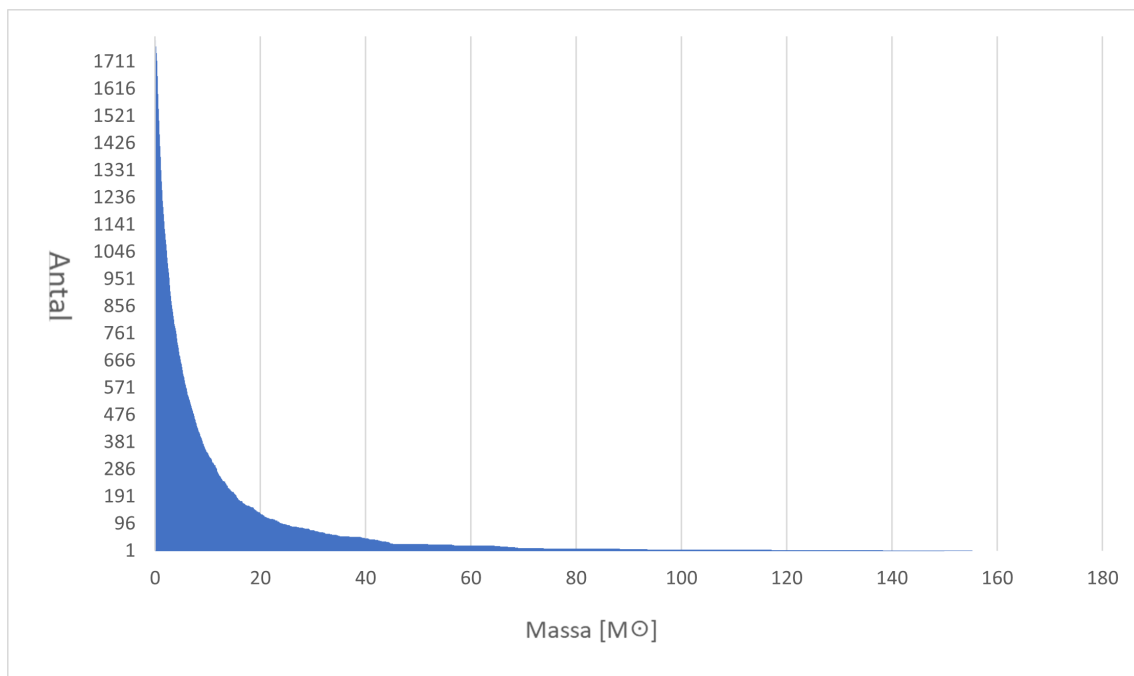
Fourier, x: 59443px, y: 516px, radie: 0.06pc, massa: $5.75M_{\odot}$, artefakt säkerhet: 1



Figur 5.8: Ett exempel på en identifierad artefakt med Fouriermetoden men som utseendemässigt efterliknar en tät kärna. De övre bilderna visar den felidentifierade artefakten i mitten, med 10 respektive 100 pixlar i bredd och höjd. De nedre graferna visar intensitetsförändringen i ett tvärsnitt genom kärnan i x - respektive y -led. Den streckade svarta linjen visar vad som innefattar diametern och därmed inom vilken radie som massan beräknas.



Figur 5.9: Ett stapeldiagram över radievärden för alla täta kärnor i katalogen.



Figur 5.10: Ett stapeldiagram över massavärden för de täta kärnor i katalogen som har en massa mellan 0 M_{\odot} och 180 M_{\odot} .

6. Diskussion

I kommande kapitel förs diskussion kring resultat och metod. Detta innefattar vilka möjliga förbättringar som finns och olika begränsningar med metoden.

6.1 Resultatanalys

Programmet fungerar överlag som förväntat. De tre bildanalytiska metoderna identifierar till stor del täta kärnor med en betydligt högre densitet än sin omgivning, vilket gör det teoretiskt troligt att metoderna hittar riktiga täta kärnor. Katalogen i sin helhet innehåller också den förväntade storleksordningen på antalet täta kärnor i PROMISE-datan, vilket stärker resultatets trovärdighet. Även artefaktidentifieringen verkar generellt sett pålitlig. Endast ett fåtal av de identifierade artefakterna med säkerhetsnivå 3 och 2 skulle kunna vara felklassade täta kärnor. Inom säkerhetsnivå 1 är det något större andel som skulle kunna vara felklassade. Med detta verkar säkerhetsmetoderna också ge ett till stor del önskvärt resultat även om det finns möjligheter till förbättringar. Att fler täta kärnor klassas som artefakter än tvärtom är önskvärt. Detta eftersom ett mer tillförlitligt resultat som underskattar antalet täta kärnor anses bättre än ett osäkert resultat som överskattar antalet.

En av de intressantare sakerna att notera från resultatet är att de olika metoderna skiljer sig i vilka täta kärnor de detekterar. Det finns många exempel där en metod tycks ha hittat en uppenbar tät kärna, men resterande metoder misslyckats med att detektera den. På så sätt är det tydligt att ingen av de prövade metoderna är väldigt nära att perfekt kunna identifiera samtliga täta kärnor. Det tyder också på att för projektets syfte kan det vara rent av nödvändigt att utnyttja en mängd olika bildanalytiska tekniker för att öka säkerheten bakom resultatet.

Radie- och massavärden kan generellt uppfattas som bra efter manuell analys av de täta kärnor som finns i katalogen. Däremot finns det vissa kärnor med potentiellt felvärderade radie och massa i katalogen. Ett fåtal täta kärnor uppskattades ha negativ massa vilket är ofysikaliskt och beror på felkalibrering av PROMISE-datan, se Avsnitt 2.2. Däremot är detta ingen stor begränsning eftersom endast fyra täta kärnor i den totala katalogen uppmättes ha negativ massa. De täta kärnor vars radie uppmättes till 0,14 pc eller högre låg i många fall i närheten av en större tät kärna. Detta ledde till att radievärdet överskattades genom HWHM-metoden dock gäller det endast ett litet antal täta kärnor av hela katalogen.

Ett förvånande resultat är att användningen av Fouriermetoden tar fram betydligt färre täta kärnor än de andra två metoderna. Detta är förvånande eftersom att både metoden som utnyttjar fouriertransform och den som utnyttjar gaussisk utjämning använder sig av oskarp maskering. Det enda som då skiljer dessa metoder åt är vil-

ken teknik som används för att jämna ut bilden och reducera bruset. Det finns ingen uppenbar anledning till varför det skulle skilja så mycket mellan gaussisk utjämning och lågpasfilter genom fouriertransform. I teorin borde båda dessa tekniker ha väldigt snarlika effekter på datan, men av resultatet så verkar det uppstå oförutsedda skillnader. Varför detta faktiskt skedde vid körning av programmet kunde inte fastställas, men likväl var det fortfarande vissa täta kärnor som endast upptäcktes av Fouriermetoden. Det som troligtvis har skett är att datan omvandlats på ett sådant sätt att karaktärsdragen hos täta kärnor i många fall har försvunnit.

Ett antal felidentifierade täta kärnor identifieras med Gaussmetoden och Waveletmetoden, där de toppar som passerat testet för att vara signifikant högre än sin baslinje (se Avsnitt 4.4.2) endast kan vara en form av brus, se Figur 5.6. I figuren har de närliggande punkterna negativt värde och den punkt som klassats som en tät kärna endast har ett värde av ungefär 1. Absolutvärde är så lågt att det råder tvivel om det i verkligheten handlar om en tät kärna. Omgivningen runt den identifierade täta kärnan är också rätt brusig, och det är svårt att övertyga sig att densiteten är signifikant högre än omgivningen. Detta kan tyda på en brist i beräkning av baslinjen eller dess standardavvikelse, som kanske hämtar omgivningen över ett för stort område, som istället hade behövt vara mer lokalt.

Även abnormaliteter i PROMISE-datan verkar påverka resultatet i katalogen något. Figur 5.7 är ett exempel på en sådan abnormalitet. Dessa tycks vara felaktigheter i datan utöver de artefakter som beskrivs i Avsnitt 2.3 som inte sorterats bort av programmet då det troligtvis rör sig om en helt annan form av artefakt i datan. Dessa övriga abnormaliteter hittas främst av Waveletmetoden och Gaussmetoden. Antagligen upptäcks de inte av Fouriermetoden då den är extremt konservativ med vad den identifierar som täta kärnor överlag. För att även ta bort dessa skulle troligtvis mer avancerade bildanalysmetoder behöva användas.

6.2 Begränsningar och förbättringsområden

I följande avsnitt analyseras programmets begränsningar och möjliga förbättringsområden med de olika metoderna presenteras.

6.2.1 Allmänna begränsningar

När datan segmenteras, görs detta genom att ta vertikala skivor av datan, när den slutgiltiga katalogen togs fram delades datan in i tio segment. Dessa tio segment medförde nio vertikala områden i datan där täta kärnor potentiellt skulle kunna missas eller alternativt dubbelt detekteras, ifall de är placerade mitt på denna indelningssträcka. Då källdatan är 120 000 pixlar bred har denna risk bedömts som försumbar.

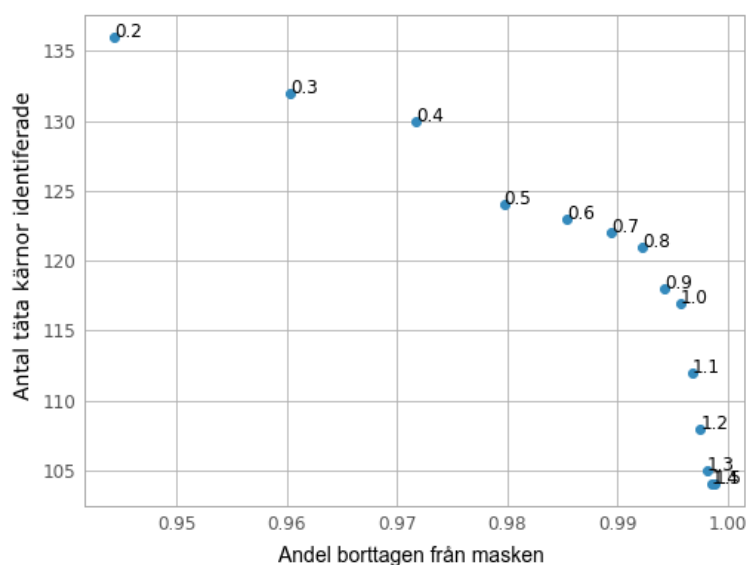
Det slutgiltiga programmet som tagits fram för syftet är inte en perfekt lösning. Det finns i vissa fall i princip ingen gräns på hur precist parametrar och tröskelvärden kan väljas. Detta kan ses som en begränsning i sig och ger många möjligheter att

förbättra metoden för ett säkrare resultat. Det finns även vissa metoder och tekniker som hade kunnat ge ett bättre resultat som inte har utforskats i detta projekt.

6.2.2 Intensitetströskling

Som en del av datahanteringen sker tröskling av intensitetsvärdena i varje mask, vilket exkluderar områden från masken med ett mycket lågt värde utifrån tröskelvärdesparametrar. Detta steg i datahanteringen medför förlust av information, mer specifikt så riskeras det att vissa små täta kärnor med mycket låg kontrast emot kringliggande område inte längre täcks av bildmasken. Detta görs primärt för att spara in på exekveringstid. Eftersom dessa täta kärnor med låg kontrast per definition inte har en intensitet mycket högre än bakgrundsbruset, är informationsförlusten minimal då de inte uppnår algoritmens definition för en tät kärna.

I Figur 6.1 visas hur olika tröskelvärden påverkar mängden data som behöver filtreras vid varje tröskelvärde, samt hur många kärnor som identifieras, med Gaussmetoden som identifieringsmetod. Figuren demonstrerar hur en mycket kraftig minskning i mängd behandlade pixlar inte nödvändigtvis korrelerar med mycket färre identifierade kärnor. Tröskelvärdet har därför valts till ett värde som både möjliggör identifiering av många täta kärnor och ger en hanterbar mängd pixlar att behandla.



Figur 6.1: Exempel på hur tröskelvärde påverkar antal detekterade kärnor via Gaussmetoden. Värdena ovanför datapunkterna visar dess tröskelvärde.

6.2.3 Fouriertransform

Teorin bakom appliceringen av fouriertransform i programmet kan inte ses som nödvändigtvis bristfällig däremot finns det troligtvis brister i praktiken. Genom att förändra filtret och att välja andra storlekar och standardavvikelsevärden för den

gaussiska fördelningen borde det teoretiskt sett gå att få stora förändringar i utjämning och resultat. Däremot fås inte dessa resultat i praktiken. Få täta kärnor hittas med både stora och små värden på den gaussiska fördelningens storlek och standardavvikelse, även med låga tröskelvärden. Utjämningar och oskarpa masker med Fouriermetoden kan med manuell analys inte uppfattas som felaktiga. Möjligtvis är Fouriermetoden därför något felimplementerad då väldigt få av täta kärnor alltid hittas, detta har dock inte kunnat verifieras inom projektets tidsramar och är något som skulle kräva ytterligare analys.

Det skulle också finnas möjligheter att använda andra typer av filter än ett lågpassfilter, men sådana filter undersöktes inte under projektets gång. Överlag skulle också en möjlighet finnas att utnyttja fouriertransform i andra delar av programmet. Möjligvis går det att hitta mönster i de täta kärnornas utseende i frekvensdomänen och på det sättet kunna utnyttja fouriertransform ytterligare i programmet.

6.2.4 Wavelettransform

I Avsnitt 3.5 konstaterades att valet av basfunktioner har en avgörande roll för hur wavelettransformen kommer att se ut och därmed hur väl täta kärnor identifieras. I den framtagna algoritmen används den ortonormala basen *Haar*, vilken inte är optimalt utformad för att framhäva cirkulära detaljer. Det skulle därmed kunna argumenteras för att en basfunktion av annat utseende, mer anpassad för att identifiera cirkulära detaljer, hade gett ett förbättrat resultat. Anledningen till att basen *Haar* ändå användes var för att detta utgör den enklaste formen av waveletfunktion vilket innebar att problem som hade uppstått med andra, mer avancerade, basfunktioner kunde undvikas.

Ett fel som tycks uppstå då wavelettransform används är att positionen för de identifierade täta kärnorna inte verkar bli helt korrekt. En trolig anledning till uppkomsten av detta fel är nedskalningen som sker av bilden vid användning av wavelettransform. Då bilden skalas ner komprimeras varje pixel, vilket troligtvis orsakar en viss osäkerhet exakt var en identifierad tät kärna är lokaliserad. Om den täta kärnans position blir inkorrekt orsakar det ytterligare fel i den beräknade radien och massan. När bilden sedan skalades upp och koordinaterna för de identifierade objekten hämtades, antogs de vara inom en osäkerhetsmarginal på en 2×2 matris, och koordinaten för det högsta värdet fick avgöra var den egentliga koordinaten var. Denna metod verkar inte lösa problemet helt, eftersom det fortfarande finns flera objekt som inte placeras på den verkliga toppen.

6.2.5 Filtrering från standardavvikelser

Ett av de mest avgörande momenten i analysprocessen är hur den slutliga filtreringen från standardavvikelsemetoden av vad som klassificeras som en tät kärna är utformad. Detta blir i praktiken kulminationen av transformmetoderna och den slutliga jämförelsen av intensitetsvärde emot baslinjen av det kringliggande området som beskrivs i Avsnitt 4.4. I projektets slutgiltiga program räknas denna baslinje ut

utifrån en kvadratrisk ruta runt den intressanta punkten vilket beskrivs i Avsnitt 4.4.2. Då de flesta täta kärnor har en till mestadels cirkulär form finns det en tydlig förbättringsmöjlighet genom att istället göra dessa uträkningar utifrån ett cirkulärt kringliggande område. Detta gäller för alla masker som i den slutgiltiga versionen av programmet beräknas som fyrkanter i koden.

Hur standardavvikelsen räknats ut (se Avsnitt 4.4.2) är inte heller en perfekt metod, till exempel så ger metoden opålitliga resultat ifall pixeln som mäts är omgiven av en större, lika intensiv, plata av värden. I det fallet kommer metoden att ta platans värde som baslinje istället för det kringliggande. Detta kan bland annat förekomma vid större artefakter och kring vissa stora täta kärnor. En möjlig lösning på detta kan vara att när baslinjen beräknas, kräva en oförändrad standardavvikelsen med flera iterationer på rad, istället för endast en som den nuvarande algoritmen gör. På så sätt måste standardavvikelsen vara oförändrad inom ett större område, och bortse från plataer eller andra jämna, mindre områden.

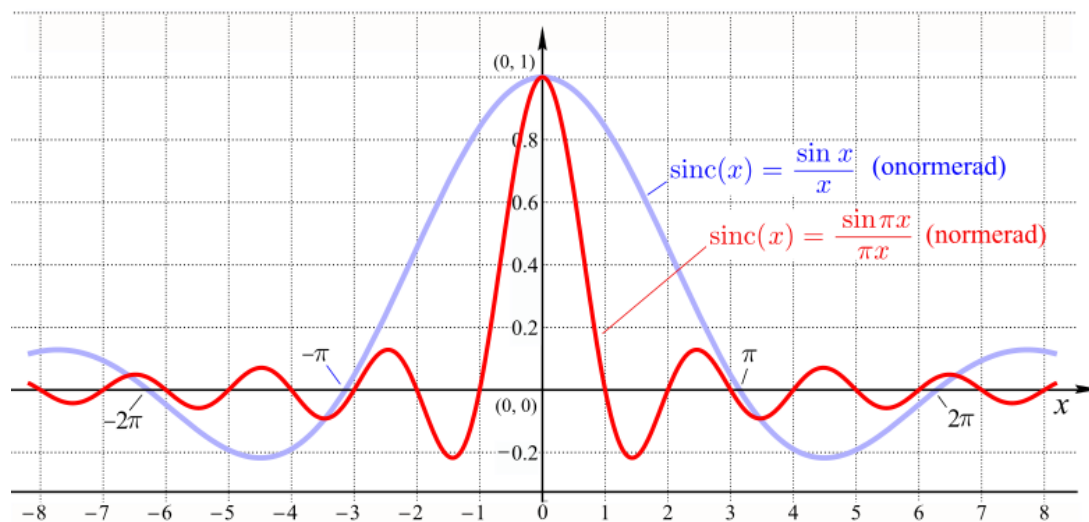
6.2.6 Identifiering av artefakter

Med den metod som använts i detta projekt är sannolikheten ganska liten att en perfekt algoritm för att sortera ut alla artefakter kan uppnås. Detta för att den förlitar sig på vissa förinställda parametrar och beroende på dessa kommer att vara olika träffsäker. Detta blir ett problem eftersom artefakterna i datan inte är uniforma, varken i storlek eller exakt form vilket gör det i stort sett omöjligt att exakt kunna hitta parametrar som omfattar endast alla artefakter utan täta kärnor. Svårtolkade artefakter fångades upp med de få och enkla parametrarna som användes för den lägsta säkerhetsnivån, men med konsekvensen att flera täta kärnor blev klassade som de. De mer komplexa parametrarna i säkerhetsnivå 3 var till för säkerställa artefakter, men tog inte med flera självklara artefakter. Ändrades en parameter för att lösa en felklassificering, kunde det felklassificera en annan. Parametrarna kan optimeras mer genom att analysera vad som utmärker de självklara artefakter som placerades som osäkra och tvärtom, men helt optimala parametrar är svårt att uppnå.

En möjlig förbättring av metoden är att utnyttja att artefakternas karaktärsdrag liknar de som finns hos sinc-funktioner (Se Figur 6.2). Det som potentiellt skulle kunna göras är att efter de potentiella täta kärnorna har identifierats först transformera dessa med fouriertransform till frekvensdomänen för att sedan utnyttja att de då borde likna en sinc-funktion i frekvensdomänen. Detta skulle potentiellt kunna göras antingen med någon form av regressionsanalys eller med ett sinc-filter som antingen filtrerar bort eller släpper igenom frekvenserna för sinc-funktioner. På så sätt skulle det vara möjligt att avgöra hur nära en sinc-funktion en potentiell tät kärna är och sedan kunna klassa denna som artefakt eller inte.

6.2.7 Radie- och massafunktion

Då projektet inte tar hänsyn till djup i PROMISE-datan och antar att objekt ligger i på ett avstånd av 3500 pc så kommer täta kärnors radie och massa ibland att



Figur 6.2: Figur över normerad och icke-normerad sinc-funktion i två dimensioner. Observera att sinc-funktionen skulle behöva vara i tre dimensioner för att kunna direkt jämföras med datan. Från [25], CC BY-SA 4.0.

felvärderas då det verkliga avståndet är moln på avstånd mellan 2000 pc och 5000 pc som beskrivet i Avsnitt 2.2. I beräkning av radie antas det att kärnor även är cirkulära och symmetriska, det tas inte hänsyn till asymmetriska objekt eller andra astronomiska fenomen påverkar deras utseende såsom bakomliggande gas. De uträknade radie och massavärdena kan därför inte förväntas vara helt träffsäkra, utan kan istället fungera som en första uppskattning.

En annan begränsning för uttänkandet av radie och massa är upplösningen av PROMISE-datan, då en stor andel täta kärnor mäts upp till att ha radie på mindre än en pixel genom HWHM-metoden. Det leder till att massan beräknas genom maximalt fyra pixlar för en övre avrundning och genom en pixel för en nedre avrundning. Ett annat tillvägagångssätt hade kunnat vara att approximera massan genom pixelandelar och att använda den interpolerade radiedatan som beskrivs i Avsnitt 4.6.1. Massan skulle på så sätt kunna beräknas med andelar av pixelvärden istället för hela pixelvärden. En summering av alla punkterna under den interpolerade radiefunktionen skulle leda till en approximativ massa som använder sig av pixelandelar.

6.2.8 Testmetoder

Ett av projektets överliggande svårigheter har varit att säkerställa att programmet faktiskt hittar exempel på riktiga täta kärnor eller artefakter och inte ointressanta delar av datan. Eftersom det saknas en komplett existerande katalog har de metoder som funnits tillgängliga för att testa om programmet fungerar varit att avgöra med ögat eller att skriva en egen testmetod i kod. Båda dessa har använts och har båda sina begränsningar.

De täta kärnorna är tillräckligt karaktäristiska för att det ofta är fullt möjligt för en

människa att avgöra om det är en tät kärna eller inte på bild. Den mänskliga faktorn måste tas hänsyn till, men denna testmetod är ofta tillförlitlig bortom större tvivel. Där testmetoden fallerar är dock i mängden potentiella täta kärnor som går att testa på detta sätt. Det hade varit alltför tidskrävande att manuellt undersöka vardera exempel programmet tar fram. Därmed saknas det med den mänskliga metoden tillräckligt med underlag för att säkert kunna uttrycka sig om hur bra en given metod är på att hitta potentiella kärnor. Däremot kan helhetsbilden uppskattas från ett urval, och när programmet tar fram något som uppenbarligen inte är en tät kärna så är det enkelt att se genom manuell analys.

De testmetoder som har implementerats i programmet med artificiellt konstruerade kärnor och artefakter har också vissa begränsningar. Testets framgång är helt beroende på huruvida de antaganden som görs om täta kärnor stämmer. Om de avviker från täta kärnors och artefaktors egenskaper skulle detta ge ett felaktigt mått på hur väl programmet kan identifiera de objekten som faktiskt existerar i datan.

En av de största begränsningarna i testmetoden är hur storlek och intensitet hos de täta kärnorna approximeras för att kunna göra artificiella kärnor, som sedan kan användas enligt Avsnitt 4.6. Då det saknas en komplett lista över täta kärnor i datan är det enda sättet att få fram någon approximation att använda det resultat som projektets dåvarande program får. Det krävs alltså ett antagande att det dåvarande programmet ger ett tillräckligt tillförlitligt svar för att kunna skapa ett meningsfullt test med någorlunda överensstämde egenskaper hos de täta kärnorna. Det ska dock sägas, som tidigare nämnts, att en manuell inspektion först skett för att kontrollera att en tillräckligt betydande mängd av resultatet faktiskt innefattade täta kärnor. Problemet kvarstår dock att det är svårt att hitta de täta kärnor som aldrig detekterats av programmet sedan tidigare, eftersom att dessa då aldrig kommer med i beräkningen för storlek och intensitet hos de artificiella kärnorna.

En lösning för att få en säkrare testmetod skulle kunna vara att göra det till en iterativ process, där för varje förbättrad identifieringsmetod så skapas det också ett nytt test utifrån den nya fördelningen av funna kärnor. Det nya testet kan sedan användas för att göra identifieringsmetoden bättre och processen upprepas fram tills ett resultat där inga större förändringar sker med ytterligare uppdatering av testmetodens fördelning. Det finns en risk i både den nuvarande och den iterativa processen med att programmet kan optimeras för att bli väldigt bra på att hitta specifikt testdatan, men om testdatan inte skulle representera de riktiga kärnorna kan det leda till en försämring i detekterandet av riktiga täta kärnor.

Något säkrare är testmetoden för artefakterna, då deras storlek, form och intensitet inte varierar speciellt mycket. Därmed har testmetoden för artefakter större säkerhet från början, för att om programmet korrekt identifierar en artefakt så är sannolikheten också stor att den korrekt identifierar stora delar av de resterande.

6.3 Förslag på framtida studier

I inledningen 1 beskrevs ett tidigare projekt, *A far-infrared survey of molecular cloud cores* [6], där författarna utifrån data från Herschel-teleskopet skapat en katalog av täta kärnor. Däremot finns det inga publicerade projekt som använt sig av PROMISE-datan eftersom detta dataset är opublicerat. Därmed hade framtida studier där PROMISE-datan används för att skapa en katalog av täta kärnor varit av intresse för att jämföra med resultatet för detta projekt. Speciellt så hade det varit intressant om andra metoder användes, som till exempel maskininlärning, för att kunna jämföra och analysera hur väl de bildanalytiska metoderna i detta projekt faktiskt fungerar.

Ytterligare något som hade varit intressant att undersöka i framtida studier är hur väl valda parametrarna i det slutgiltiga programmet är. Troligtvis hade flera parametrar i programmet kunnat finjusteras för att få bästa möjliga resultat, men på grund av projektets tidsbegränsningen kunde fokuset inte ligga på att testa olika parametervärden i alltför stor utsträckning.

Ännu ett exempel på något som hade kunnat arbetats vidare med i framtiden är exekveringstiden för programmet. I detta projektet var syftet inte att skapa ett så snabbt program som möjligt utan snarare att lära sig om bildanalytiska metoder och utifrån dessa få ut en katalog av täta kärnor. För framtida studier hade det varit intressant att arbeta fram en snabbare exekveringstid för att programmet ska bli lätthanterligt, så att det till exempel blir mer tidseffektivt när parametrar i programmet ska finjusteras.

6.4 Slutsats

Projektet kan anses ha lyckats i målet att skapa ett bildanalysitkst vekttyg som kan identifiera täta kärnor med viss säkerhet och skilja dem från artefakter. Även om individuella metoders prestationer för att hitta täta kärnor eller artefakter kan vara bristfälliga fungerar det kombinerade programmet väl för att både identifiera täta kärnor och skilja dem från artefakter.

Sammanfattningsvis kan det konstateras att en bildanalytisk metod är lämplig för ändamålet att detektera täta kärnor i PROMISE-datan. De olika tekniker som använts i projektet verkar alla till viss grad vara lämpliga, med Gaussmetoden som den teknik som tycks ge bäst resultat. Dock visar resultatet hur de olika bildanalysteknikerna inte hittar samma täta kärnor, vilket kan tyda på att det är lämpligt att utnyttja flera olika tekniker och jämföra resultaten för att uppnå ett så tillförlitligt resultat som möjligt. Vid manuell analys av resultatet framgår det även att en majoritet av vad det slutgiltiga programmet klassar som artefakter också tycks vara det, vilket påvisar att projektet varit framgångsrikt i den aspekten.

Det har också framkommit genom projektet hur svårbesvarad frågan om att ka-

talogisera täta kärnor kan vara, då det finns brist på en objektiv skala att mäta resultatet mot. Det går endast att bedöma hur korrekt den framtagna katalogen är genom att jämföra med den genererade artificiella datan, som inte har någon garanti på att matcha verkligheten. Däremot går det att uppnå relativt stor säkerhet genom manuell analys av både den artificiella datan och de framtagna kärnorna, vilket leder till större säkerhet i att de framtagna kärnorna är verkliga. Det är dock svårbedömt hur många täta kärnor som helt har missats av analysmetoderna, då de kan ha mycket låg intensitetsnivå och mycket låg kontrast till sin omgivning.

Litteraturförteckning

- [1] NASA science, “Stars.” [Online]. Tillgänglig: <https://science.nasa.gov/astrophysics/focus-areas/how-do-stars-form-and-evolve>. Hämtad: 2023-03-07.
- [2] M. R. Krumholdz, *Star Formation*. World Scientific, 2017.
- [3] Kainulainen J., Rezaei Kh. S., Spilker A., Orkisz J., Olivier E., Rener M., Trevino-Morales S., Zhang M., “PROMISE.” [Online]. Tillgänglig: <https://web.archive.org/web/20230124004706/http://promise.jounikainulainen.com/>. Hämtad: 2023-02-07.
- [4] NASA, “Spitzer space telescope.” [Online]. Tillgänglig: <https://nssdc.gsfc.nasa.gov/nmc/spacecraft/display.action?id=2003-038A>. Hämtad: 2023-04-25.
- [5] European Space Agency (ESA), “Herschel science and legacy.” 2019. [Online]. Tillgänglig: https://sci.esa.int/documents/33622/35995/1567260390853-ESA_Herschel_Science_and_Legacy_brochure_2019.pdf. Hämtad: 2023-04-25.
- [6] N.E. Jessop, D. Ward-Thompson, “A far-infrared survey of molecular cloud cores.” *Monthly Notices of the Royal Astronomical Society*, vol. 311, nr. 1, ss. 63–74, jan. 2000. [Online]. Tillgänglig: <https://doi.org/10.1046/j.1365-8711.2000.03011.x>.
- [7] D. Ward-Thompson and A. P. Whitworth, *An Introduction to Star Formation*. Cambridge, UK: Cambridge University Press, 2011.
- [8] European Southern Observatory, “How to become a star.” [Online]. Tillgänglig: <https://www.eso.org/public/images/eso0102a/>. Hämtad: 2023-02-07.
- [9] European Southern Observatory., “Bok globule b68.” [Online]. Tillgänglig: <https://www.eso.org/public/images/eso0102b/>. Hämtad: 2023-02-07.
- [10] S. W. Stahler, *Dense Core*, pp. 1–2. Berlin, Heidelberg: Springer Berlin Heidelberg, 2014.
- [11] J. Kainulainen. Privat kommunikation. Maj. 2023.
- [12] K.A. Marsh et al, “Multitemperature mapping of dust structures throughout the Galactic Plane using the PPMAP tool with Herschel Hi-GAL data,” Nov.

2017.

- [13] S. Misra, H. Li, J. He, *Machine Learning for Subsurface Characterization*. 50 Hampshire Street, 5th Floor, Cambridge, MA 02139, United States: Gulf Professional Publishing, 2020.
- [14] Wisconsin University of Wisconsin-Madison, “The gaussian kernel.” [Online]. Tillgänglig: <https://pages.stat.wisc.edu/~mchung/teaching/MIA/reading/diffusion.gaussian.kernel.pdf.pdf>. Hämtad: 2023-04-28.
- [15] R. W. R. Gonzalez, *Digital Image Processing, tredje upplagan*. Boston, USA: Pearson Education International, 2008.
- [16] “Wavelets: Applications,” in *Encyclopedia of Mathematical Physics* (J.-P. Francoise, G. L. Naber, and T. S. Tsun, eds.), pp. 420–426, Oxford: Academic Press, 2006. [Online]. Tillgänglig: <https://www.sciencedirect.com/science/article/abs/pii/B012512666200242X?via%3Dihub>, Hämtad: 2023-04-25.
- [17] Y. Chen, “An introduction to wavelet analysis with applications to image and jpeg 2000,” in *2022 4th International Conference on Intelligent Medicine and Image Processing, IMIP 2022*, (New York, NY, USA), p. 49–57, Association for Computing Machinery, 2022. Hämtad: 2023-04-18.
- [18] Wikipedia Commons, “File:lichtenstein img processing test.png.” [Online]. Tillgänglig: https://commons.wikimedia.org/wiki/File:Lichtenstein_img_processing_test.png. Hämtad: 2023-04-18.
- [19] A. Damato, “An example of the 2D discrete wavelet transform that is used in JPEG2000.” [Online]. Tillgänglig: https://en.wikipedia.org/wiki/Wavelet_transform#/media/File:Jpeg2000_2-level_wavelet_transform-lichtenstein.png. Hämtad: 2023-04-18.
- [20] FITS Working Group, “Definition of the Flexible Image Transport System (FITS).” [Online]. Tillgänglig: https://fits.gsfc.nasa.gov/standard40/fits_standard40aa-1e.pdf. Hämtad: 2023-04-18.
- [21] W. A. Joye and E. Mandel, “New Features of SAOImage DS9,” Jan. 2003. Provided by the SAO/NASA Astrophysics Data System.
- [22] Matplotlib Development Team, “Matplotlib: Visualization with python.” [Online] <https://matplotlib.org>. Hämtad: 2023-05-10.
- [23] sci-kit image, “Module: filters.” <https://scikit-image.org/docs/stable/api/skimage.filters.html#skimage.filters.gaussian>. Hämtad: 2023-05-02.

- [24] E. O. Brigham, *The Fast Fourier Transform and Its Applications*. USA: Prentice-Hall, Inc., 1988. ISBN 0133075052.
- [25] svjo, “File:sinc-function.png.” [Online]. Tillgänglig: <https://commons.wikimedia.org/wiki/File:Sinc-function.png>. Hämtad: 2023-04-18.

A. Appendix

A.1 Pseudokod

I nedan visas pseudokod för Program 1.

Program 1 Pseudokod för identifieringsprogrammet

▷ Ställ in parametrar

$def_parametrar \leftarrow \{\dots\}$

▷ Lägg potentiellt in artificiell data

if *Artificiell_data* **then**

$data, art_kärnor_koordinater \leftarrow artificiella_kärnor(data)$

$data, art_artefakter_koordinater \leftarrow artificiella_artefakter(data)$

end if

▷ Val av metod

if *Gaussisk_utjämning* **then**

▷ Oskarp mask genom gaussisk utjämning samt tröskling

$oskarp_mask \leftarrow data - gaussisk_utjämning(data, def_parametrar)$

$trösklad_mask \leftarrow oskarp_mask > OSKARP_MASK_TRÖSKELVÄRDE$

end if

if *Fourier* **then**

▷ Fourier högpasfilter och tröskling

$fourier_mask \leftarrow data - fourier_lågpassfilter(data, def_parametrar)$

$trösklad_mask \leftarrow fourier_mask > FOURIER_TRÖSKELVÄRDE$

end if

if *Wavelet* **then**

▷ Summera horisontella, vertikala, och diagonala wavletvärden

$wavelet_sum \leftarrow diskret_2d_wavelet(data)$

$trösklad_mask \leftarrow wavelet_sum > WAVELET_TRÖSKELVÄRDE$

end if

▷ Filtrera värden inom masken utifrån standardavvikelsen från baslinjen

$täta_kärnor_mask \leftarrow filtrering_från_standardavvikelser(data, trösklad_mask, def_parametrar)$

▷ Identifierar artefakter och täta kärnor

$artefakter_mask \leftarrow identifiera_artefakter(täta_kärnor_mask)$

$täta_kärnor_mask \leftarrow täta_kärnor_mask - artefakter_mask$

```
täta_kärnor_koordinater ← täta_kärnor_mask == 1  
artefakter_koordinater ← artefakter_mask == 1
```

▷ Räkna ut radie och massa

```
radie_lista ← beräkna_radie(data,täta_kärnor_koordinater)  
massa_lista ← beräkna_massa(data,täta_kärnor_koordinater),radie_täta_kärnor)
```

▷ Spara katalogen

▷ Testa katalogen ifall artificiell data har satts in

```
if Artificiell_data then  
    artificiellt_testvärde ← test_mot_artificiell_katalog(katalog,artefakt_katalog,  
    ,artificiella_koordinater)  
end if
```

A.2 Val av parametrar för identifiering av täta kärnor

För det slutgiltiga programmet valdes parametrar till följande:

Maximalt antal pixlar att expandera kring varje punkt vid uträkning av standardavvikelse: **61**

Faktor större en standardavvikelsen en punkt måste uppnå: **5**

Antal nivåer att köra wavelettransform: **1**

Minimalt avstånd mellan två täta kärnor (pixlar): **5**

Absoluta tröskelvärde för summerade waveletmasken: **2**

Absoluta tröskelvärde för den linjära oskarp maskerings-masken: **0.5**

Sigma (styrkan) för linjär oskarp maskering: **1**

Absoluta tröskelvärde för masken från oskarp maskering genom fourier: **2**

Radie för lågpassfiltret vid oskarp maskering genom fourier: **500**

A.3 Val av parametrar för artificiell data

- Antal artificiella täta kärnor att generera (per segment): **5000**
- Storlek på begränsande låda för artificiella kärnor: **1001**
- Radie-Massa distribution av artificiella kärnor: Se Figur 4.10
- Antal artificiella artefakter att generera (per segment): **500**
- Minimum intensitet av artificiella artefakter: **25**
- Maximum intensitet av artificiella artefakter: **75**

A.4 Programkod och fullständig katalog

Koden samt fullständiga katalogen finns online vid:

<https://github.com/Powana/AoP-80-Birthplace-of-Stars>

INSTITUTIONEN FÖR RYMD-, GEO-, OCH MILJÖVETENSKAP
CHALMERS TEKNISKA HÖGSKOLA

Göteborg, Sverige 2023
www.chalmers.se



CHALMERS