MASTERARBEIT / MASTER'S THESIS

Titel der Masterarbeit / Title of the Master's Thesis

„Millimeter variability of protoplanetary disks: ALMA observations of the SODA sample“

verfasst von / submitted by

David Krebs BSc BSc

angestrebter akademischer Grad / in partial fulfilment of the requirements for the degree of

Master of Science (MSc)

Wien, 2024 / Vienna, 2024

Studienkennzahl lt. Studienblatt / degree programme code as it appears on the student record sheet:	UA 066 861
Studienrichtung lt. Studienblatt / degree programme as it appears on the student record sheet:	Masterstudium Astronomie
Betreut von / Supervisor:	Ass. Prof. Alvaro Hacar, PhD

"In all times and epochs the greatest happiness for man has been to take part in new discoveries."

- Yuri Alekseyevich Gagarin

Abstract

Introduction: Since its completion, the Atacama Large Millimeter Array (ALMA) has enabled astronomers to conduct surveys with increasing amounts of sources at high detail compared to the surveys before. Young stellar objects (YSOs) due to their dusty nature are prime targets for such surveys.
The Survey of Orion disks with ALMA (SODA) is the largest survey to date looking at 873 unresolved YSO Class II in the L1641 and L1647 clouds. van Terwisga et al. (2022) have conducted this survey with the goal of finding even weak trends as well as the dust mass distribution in the disks while maintaining a low statistical error.
Goals: The primary goal is the reproduction of the dust mass distribution using the Vienna Scientific Cluster (VSC). Further, we aim to measure spectral indices for each source within one band utilising two different channels and detect possible variability within the disks, as the sources have been observed multiple times. Given the large sample size available with SODA, this is one of the first systematic detections of variability in protoplanetary disks.
Methods: The ALMA data was reduced with the VSC5 utilising a self written code. For the analysis and the visualisations we used a server of the Institute for Astrophysics (IfA).
Results: We achieved the disk mass results from van Terwisga et al. (2022) within reasonable error.
Due to high statistical scatter, we discarded the results of the spectral indices as they are not reliable for almost all sources.
Amongst the ≈400 detections, we find 21 sources with a variability above 5σ at timescales of ≈ one hour to a few days with an amplitude of up to a factor if two.

Kurzfassung

Einleitung: Die Fertigstellung des Atacama Large Millimeter Array (ALMA) ermöglichte Astronomen erstmals, Surveys (dt. Durchmusterungen) mit immer mehr Quellen in großer Genauigkeit durchzuführen. Young stellar objects (YSOs) sind durch ihren großen Staubgehalt eines der geeignetsten Ziele solcher Surveys. Die Survey of Orion disks with ALMA (SODA) ist die bisher größte Durchmusterung von 873 unaufgelösten YSOs Class II in den Wolken L1641 und L1647. van Terwisga et al. (2022) führten diese Durchmusterung mit dem Ziel durch, auch schwache Trends und die Staubmassenverteilung in den Scheiben zu erkennen, und dabei einen geringen Fehler zu behalten. **Ziele:** Das grundlegende Ziel besteht darin, die Ergebnisse von van Terwisga et al. (2022) bezüglich der Staubmasse der über 850 YSOs zu reproduzieren. Dafür verarbeiten wir die Daten am Vienna Scientific Cluster (VSC). Außerdem wollen wir die Spektralindizes für jede Quelle innerhalb eines Bandes mit zwei verschiedenen Kanälen messen und die mögliche Variabilität der Scheiben messen, da die Quellen mehrfach beobachtet wurden. Angesichts der großen Stichprobengröße, die SODA mit sich bringt, ist dies eine der ersten systematischen Entdeckungen von Variabilität in protoplanetaren Scheiben. **Methoden**: Der Großteil der Berechnungen wird mit dem Vienna Scientific Cluster (VSC) durchgeführt, genauer gesagt mit dem VSC5. Die Analysen wurden mit einem Server des Instituts für Astrophysik (IfA) durchgeführt.

Ergebnisse: Die Scheibenmassen von van Terwisga et al. (2022) konnten, innerhalb sinnvoller Abweichungen, reproduziert werden. Aufgrund der hohen Streuung wurden die Ergebnisse der Spektralindizes verworfen, da diese für nahezu alle Quellen nicht verlässlich waren. Unter den ≈400 Quellen fanden wir 21, die eine Variabilität des Flusses über einer 5σ Grenze zeigten. Die Variabilität zeigte Zeitskalen von etwa einer Stunde bis einige Tage auf und besaß eine Amplitude von bis zu zwei.

Contents

1 Introduction

The interstellar medium (ISM) is arguably the most important constituent of space to understand when looking into astrophysics below a galactic scale, for example the vital process of star formation. The cycle of a stars life is involved in most phenomena that are studied in astronomy and are responsible for the creation of planets, further stars, nebulae, small bodies and much more.
When looking at how a star is born, one of the evolutionary stages is the protoplanetary disk. It forms when a cloud in the ISM has partially collapsed into a star and the surrounding material has formed or is forming an accretion disk. Objects that are going through this process are called young stellar objects, or YSOs for short, that have been discovered in the 1930s and 1940s by i.e. B.J. Bok or V.A. Ambartsumian in Ambartsumian (1947) with the modern classification and understanding first laid out in Lada (1987), with later additions by Andre et al. (1993) and Greene et al. (1994).
The understanding of these YSOs and their protoplanetary disks is paramount for getting an idea about later planet formation and are important for models of later population synthesis as explained in Benz et al. (2014).
The advent of the Atacama Large Millimeter Array (ALMA) and an improvement of physical and chemical disk simulations made it possible to get a more detailed idea of protoplanetary disks, but the way this was done usually only focused on few sources, lacking large statistically relevant samples, particularly in the mm range as explained in Miotello et al. (2022).
ALMA, as can be read in the technical handbooks Cortes et al. (2023) or Remijan et al. (2021), is located in the Atacama desert in Chile and is operated by the European Southern Observatory, or ESO for short and stands at roughly 5000m above mean sea level which grants it partial immunity to the absorption effect the atmosphere has on radio wavelengths in lower elevations. It has the big advantage of using interferometry to make observations, greatly increasing its capabilities compared to standard observations with a single telescope or antenna.

Usual observations with ALMA have few sources with great detail, which typically satisfies the goals of other studies but leaves a lot on the table in terms of statistics and comparability. Not only that, but big mm surveys themselves are not something that has frequently been done historically but plays an ever increasing role today. See for example Fischer et al. (2023) and Miotello et al. (2022) who talk about various surveys and how specifically ALMA has furthered our understanding of YSOs, or Mairs et al. (2019) for a sub-mm survey that monitored star forming regions in Orion finding variable sources above the expected magnitude in the JCMT Transient survey.
In 2019, the largest sample of YSO Class II in the mm-band was observed using ALMA. This was done to study the relatively cold dust within the earlier mentioned protoplanetary disks in the close by Orion nebula with one large survey.

This survey is the 'Survey of Orion Disks with ALMA', or the SODA survey (also sometimes called the SODA sample). It is the reason for this thesis to exist and is a sample of 873 mostly YSO Class II sources in the L1641 and L1647 clouds in Orion. The sample size of 873 protoplanetary disks makes it especially interesting for statistical analyses since it is the biggest of its kind and stands in contrast to the usually constrained observations in mm of Orion, allowing for very small error even in sub-samples, see van Terwisga et al. (2022).
In van Terwisga et al. (2022), one of the final goals is to determine the disk mass distribution, an indication for how massive the resulting stars will be and valuable parameter in physical and chemical simulations about planet formation, amongst other fields. This is interesting not only because Orion is a very near star forming region and is thus a prime target for observation, but also because the remarkable capabilities of ALMA are enabling a more focused study on the different regions within the two clouds.
The original basis of this sample stems from a catalogue introduced in Megeath et al. (2012), who used Spitzer to identify dusty YSOs within Orion. That is also the reason why all the sources are referred to as *[MGM2012]number* (for example [MGM2012]512) in van Terwisga et al. (2022) and as *M12_number* (M12_512) in this work.
Every source from Megeath et al. (2012) was checked with ALMA in van Terwisga et al. (2022) and its mass was calculated from the received fluxes.

The aims of this work are

1. to reproduce the distribution of the disk masses from van Terwisga et al. (2022) with the catalogue from Megeath et al. (2012) albeit with slightly different methods and settings
2. to inspect the variability of every object, draw statistical conclusions and compare it to already known properties where available
3. to get information about the spectral slope achieved through the parameter β introduced later

After we reach similar results to van Terwisga et al. (2022), the variability and the spectral indices will be checked. To investigate points 2 and 3, each objects' individual observations will be analysed and a flux will be extracted. Note that there are not the same number of observations for every object. Also note that the observations were not made to look at the respective variabilities, so it is possible that only a small amount of variable sources are found amongst the ~900 objects, or even none at all.
When we talk about variability, we mean a change in flux over time. The variability of YSOs is strongly dependent on the timescale it occurs, as for example laid out in Fischer et al. (2023) in general and more specifically for radio and X-ray in Mairs et al. (2019), where it is more closely tied to magnetic activity. Generally speaking, the longer the time of one full cycle of the change in flux, the higher in amplitude that change is.
Something noteworthy about variability is that the available data usually spans from a few days to decades, but observations on shorter scales are somewhat rare but sometimes show very strong increases in flux densities already at small timescales of minutes, see Vargas-González et al. (2023), Vargas-González et al. (2021) and Forbrich et al. (2017), who find various high amplitude

variability of for example over a factor of 20 in Vargas-González et al. (2023) in only 15 minutes. We would like to gain a better understanding of variability because it can give us hints for physical conditions and processes occurring in the disk and shine a light on the interactions between it and the protostar. As for example laid out in Fischer et al. (2023), some variability patterns link to accretion, some to a possibly long lasting change in chemistry in the disk and others to the rotation of the young star. Benz & Güdel (2010) explain how a change in the magnetic field of the star can lead to variability, usually over shorter times of minutes to days.

We will also look at spectral indices. All observations were done in four channels, two higher channels (which will from now on be referred to as 'windows' as it is called that in the software) at roughly 241GHz and two lower channels at 225GHz. From this, the flux is extracted for both of these windows, which allows for the creation of 'spectral indices'[1]. From these indices, we hope to draw some conclusions about the surrounding dust and its size distribution, which was not yet done by van Terwisga et al. (2022) or follow up papers with the SODA sample.
The spectral indices can give insight into the shape of the emission continuum, which in turn can be linked to the temperature as well as the dust size and its distribution, should the quality of the data allow for it.

For the reduction of the data, it is advantageous to use parallel computing or, as van Terwisga et al. (2022), a supercomputer. This work will thus implement the reduction on the Vienna Scientific Cluster, or VSC for short, as the amount of data and the heavy computational load almost require supercomputing power. To be specific, the newest cluster VSC5 was used.
Note that the reproduction of van Terwisga et al. (2022) is of course to be seen relative. The further aims require a somewhat different algorithm for the reduction of the data that we will cover in more detail later. It is reasonable to assume that the end results will differ due to the changes in reduction. Still, for example the flux values of van Terwisga et al. (2022) acted as a guide stone for the data processing of this work.

[1]This is the most basic kind of spectral index since it only uses two points. The SODA survey is not meant for these kinds of studies, but we will try anyhow.

2 Underlying Theory

This chapter will cover all the required astronomy, physics and observation techniques to create a bigger picture and motivate the later data reduction as well as explain the measurements that were made.
The calculations presented in 3.5 and executed in 4 are based on the physics explained in this chapter.

2.1 Physics

This section will cover the fundamental physics for the understanding of this work. We will go into some of the physics behind electromagnetic radiation and dust, as well as necessary additions.

2.1.1 Radiation

Black body radiation

The black body spectrum, one of the big problems of physics at the end of the 19th century, and first correctly described by Max Planck at the very beginning of the 20th century, is the conventional model when asking about the frequency distribution of emitted radiation at a given temperature.

$$B_\nu(T) = \frac{2h\nu^3}{c^2} \frac{1}{\exp\left(\frac{h\nu}{k_B T}\right) - 1} \tag{2.1}$$

In formula 2.1 taken from Williams (2021), $B_\nu(T)$ is the emitted radiation at a given frequency ν and temperature T. h is Planck's constant, c is the speed of light and k_B is Boltzmann's constant. Note that for longer wavelengths of >250μm (when $h\nu << k_b T$), formula 2.1 can be reasonably approximated by $B_\nu(T) = 2k_B T \nu^2/c^2$ which is known as the Rayleigh-Jeans regime. As the data for this thesis has been observed at ~1.3mm, this approximation will be used from here on.
Note however that the black body is not the best model to approximate the actual radiation from dusty sources. This is due to ongoing absorption of higher frequencies and then re-emission at lower frequencies. The better model for this is called grey body radiation.

Grey body radiation

The dust clouds that surround early type stars drastically change the emitted spectrum from an ordinary black body. The reason for that is that the dust has a high absorption efficiency for

optical and UV photons, that then get emitted as FIR and radio photons. This process depends on the temperature distribution of the cloud as well as the amount of dust present.
To analyse the emission and absorption of a body, one usually turns to the radiative transfer equation[1]:

$$I_\nu = I_\nu(0)e^{-\tau_\nu} + B_\nu(T)(1 - e^{-\tau_\nu}) \tag{2.2}$$

Where in equation 2.2 I_ν is the specific intensity and $I_\nu(0)$ is the initial intensity before extinction, τ is the optical depth which significantly depends on the dust grain sizes and the actual amount of dust and $B_\nu(T)$ is the black body radiation that is emitted.
Note that

$$I_\nu = B_\nu(T)(1 - e^{-\tau_\nu}) \approx B_\nu(T) \cdot \tau_\nu \text{ for } \tau << 1 \tag{2.3}$$

is what is called grey body radiation, which is the part we can measure from earth, assuming a constant dust temperature we can change T to T_d from here. This temperature is usually set to 20K in literature, as it is the effective temperature of dust. See for example Draine (2006), van Terwisga et al. (2022) or Fischer et al. (2023) to name just a few that employ this assumption[2].

Flux and β

From equation 2.3 we know that $I_\nu = B_\nu(T) \cdot \tau_\nu$, where $d\tau_\nu = \kappa_\nu \cdot ds$ where ds is the depth of the dust. Please note that κ_ν can be expressed as $\kappa_\nu^{\text{dust}} \cdot \rho_d$ due to normalisation. ρ_d is the density of the dust. If we now think of the radiated flux as $F_\nu = \int I_\nu \cdot d\Omega$, so the integral of the intensity over all angles $d\Omega$. $d\Omega$ may also be expressed as dA/d^2, where d is the distance at which we are observing the flux and A is the area over which the radiation is equally distributed.
If we combine these two expressions, we can get a formula for the observed flux at a given frequency.

$$F_\nu = \int_\Omega I_\nu \cdot d\Omega = \int_A B_\nu(T)\rho_d\kappa_\nu^{\text{dust}} \cdot dA/d^2 = \kappa_\nu^{\text{dust}} B_\nu(T)\frac{M_{\text{dust}}}{d^2} \tag{2.4}$$

The flux emitted from the observed protoplanetary disk in the frequency ranges observed for this work thus stems almost exclusively from thermal radiation. The flux emitted by such a dusty disk would be:

$$F_\nu = \frac{\kappa_\nu^{\text{dust}} M_{\text{dust}} B_\nu(T)}{d^2} \tag{2.5}$$

Where equation 2.5 can also be found in Draine (2006). d is the distance to the object, M_{dust} is the mass of the dust and κ_ν^{dust} is the dependence of the opacity on the frequency, usually expressed as $\kappa_\nu^{\text{dust}} \propto \nu^\beta$ and thus with the aforementioned Rayleigh-Jeans regime from before

$$F_\nu \propto \nu^\alpha \tag{2.6}$$

where $\alpha = 2 + \beta$. β is the slope of the grea body spectrum and is also the parameter that will be given in the analysis of the spectral indices later. M_d is the dust mass, T_d is the dust temperature and d is the distance.

[1] also known as Kirchhoff's radiation law

[2] In the case of Fischer et al. (2023) they say that the temperature is in the ranges of 10K to 20K towards the outside of the disk.

From equation 2.6 and remembering that there are two spectral windows giving us two fluxes, the parameter α and thus with it β can be determined.

$$\frac{F_{high}}{F_{low}} = \left(\frac{\nu_{high}}{\nu_{low}}\right)^{2+\beta} \longrightarrow \beta = -2 + \frac{\log\left(\frac{F_{high}}{F_{low}}\right)}{\log\left(\frac{\nu_{high}}{\nu_{low}}\right)} \tag{2.7}$$

Where in equation 2.7 the F_{high} is the flux from the higher frequency window of the two, and ν_{high} the frequency of the higher window. The β that is given in chapter 4 is the β as from equation 2.7. Typically, β is close to 0.9 ± 0.7 in most cases at this frequency range, according to Draine (2006) for protoplanetary disks. From that we can already tell that the dust grains are bigger than in the surrounding ISM, where β is ≈ 1.7. Please note that we can find other values for β in recent literature. van Terwisga et al. (2022) themselves not that it is usually at 1, with some literature giving a lower values of ≈ 0.4(Pascucci et al. (2016) who measured 66 disk masses in Cha I).
The mass of said disks can also be calculated looking at equation 2.5.

$$M_{\text{dust}} \approx \frac{F_\nu d^2}{\kappa_\nu^{\text{dust}} B_\nu(T)} \tag{2.8}$$

All of the factors in equation 2.8 are known or follow from the observations, apart from T_d and κ_ν.

As laid out previously in section 2.1.1 and practised in van Terwisga et al. (2022), T_d may be reasonably estimated to be 20K for all following calculations, leaving only κ_ν^{dust} as an unknown. From van Terwisga et al. (2022), κ_ν at 230GHz is given as $2.3\text{cm}^2\text{g}^{-1}$ with a dependence on $\kappa_\nu^{\text{dust}} = \kappa_0 \cdot \nu^\beta$, where $\beta = 1$. It should be noted that in recent literature β is sometimes stated to be around 0.4, see van Terwisga et al. (2022). This range includes the value for β stated in Draine (2006), but is notably smaller.
This seems like it would make a large difference, however the frequency range will be small, so the consequences remain small too.

2.1.2 Dust

This subsection will present the necessary knowledge about dust needed to understand both later statements and some of the bigger picture, and comes in this form after Williams (2021) and Draine (2006).
About 1% of the mass of the interstellar medium (ISM) is composed of dust. Dust in this context are agglomerates of available molecules, predominantly made from Si, C, O, Fe, Al and similar elements.
While dust only makes up a small fraction of the ISM mass, it is emits thermal radiation that makes up 30% of the radiation of the galaxy, where the gas primarily emits in narrow lines. This comes from the fact that dust emits efficiently at lower frequencies, particularly in the far infrared and in radio.
It does this because it absorbs light from surrounding stars and re-emits it in mostly IR and radio. This simple mechanism, absorbing light from starts and re-emitting in lower frequencies, makes

it highly useful to study the stars within and around it. The brighter the star, the more flux we should receive in these wavelengths.
The dust later plays a fundamental role in the creation of planets and other large as well as small bodies.

2.2 Molecular clouds & Young stellar objects

2.2.1 Historical context

In the 1930s and 1940s it became apparent that there is ongoing star formation within our galaxy. Most notably first suggested by Ambartsumian (1947) based on earlier observations. Around 40 years later this evolution from material such as gas and dust clouds to main sequence stars was first classified with a scheme based on the spectrum of these objects in Lada (1987), calling them young stellar objects or YSOs for short.
In a paper Lada (1987) introduced three so called Classes of YSO, where Class I described the earliest stage and Class III the latest. The Classes are based on the slope of the spectrum using the α that we introduced in equation 2.6 but only utilising the portion between roughly 2.2μm and 20μm. All the Classes and some additions to them will be explained in detail at a later point. This was a significant step up in terms of comparability and was followed a few years later by Andre et al. (1993) who in their paper show that there is also an evolutionary stage before the Class I that they called Class 0, where a cold molecular cloud contracts but importantly has not formed a star or a disk yet.
Again a year later, Greene et al. (1994) have published a paper adding a middle ground between Class I and Class II that is commonly referred to as a 'flat spectrum', since the α is nearly flat. We exclude sub-Classes on purpose as they play no role in this thesis.

Class 0	λ undetectable at $< 20\mu m$
Class I	$\alpha > 0.3$
Flat spectrum	$0.3 > \alpha > -0.3$
Class II	$-0.3 > \alpha > -1.6$
Class III	$\alpha < -1.6$

Table 2.1: A table of the classification, commonly found in literature and presented in Greene et al. (1994), Andre et al. (1993) and Lada (1987). More modern sources list $20\mu m$ instead of $10\mu m$ for Class 0.

2.2.2 YSO Class 0

Class 0 sources are, again according to Andre et al. (1993), faint sources that have a strong sub-millimetre feature and "remain undetected at $\lambda<10\mu m$"[3]. The way we currently see these YSO Class 0 sources, after Werner & Eisenhardt (2019), is a molecular cloud that has a collapsing

[3] Other (more modern) sources give $\lambda<20\mu m$ here, for example `https://www2.mps.mpg.de/solar-system-school/lectures/origin_solar_system_2009/Dasi.pdf` accessed 17.01.2023

core in the centre with the surrounding dust falling inwards. We once again emphasise that neither a star nor a disk exist yet.
Prior to this section we talked about the black and grey body radiation from equation 2.1.1 and equation 2.3, respectively. Although for the understanding of the following, only the black body spectrum is necessary. During the collapsing of a cloud, the material in said cloud will contract increasing the density and thus opacity. The thermal radiation from the heated centre can no longer escape the cloud and the interior will start to heat up. This means that now there exist dust that is warmer than before, from which we would expect black body (or grey body) radiation.

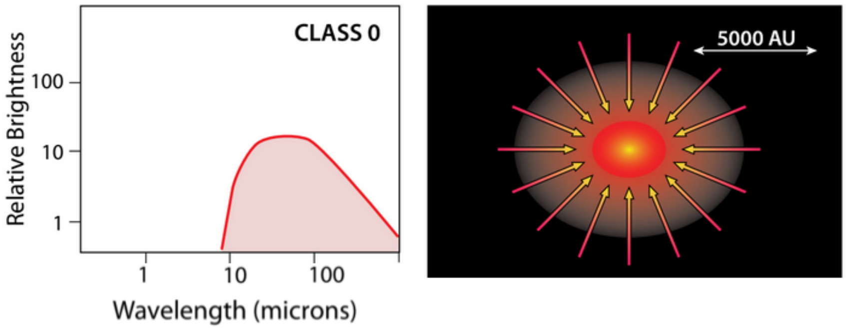

Figure 2.1: Sketch of a YSO class 0, from Werner & Eisenhardt (2019). In the left panel we can see the spectrum of the dust at around 20K in red, while in the right panel we can see a sketch of its morphology. The arrows show the movement of the cloud material.

In the left panel of figure 2.1 we see the spectrum of the dust, with the x-axis as microns (=μm) and the y-axis as relative brightness the peak of the black body spectrum is at $\approx 100\mu m$, indicating the temperature of about 20K. It is important to note here that the plot in the left panel of figure 2.1 only goes to $1000\mu m$ or 1mm, where the observations that we will use in this work are at roughly 230GHz or 1.3mm. This is because in Werner & Eisenhardt (2019) the focus is on infrared radiation, but the sketches are still useful to gain a deeper understanding about the morphology and overall trend of the spectrum.
The time for the YSO to go through this class is at the order of $\approx 10^4$ years according to Dauphas & Chaussidon (2011), who give an overview of YSOs in their paper.

2.2.3 YSO Class I

After the collapsing cloud of the YSO Class 0, a central protostar has already formed. The surrounding dust is in the process of forming an accretion disk and dust further out keeps falling inwards.
The star forms as a consequence of the cloud collapsing further under its own gravity. The disk around the protostar forms, according to the book about ISM and its physics by Williams (2021), mainly due to the conservation of angular momentum. The scale of the YSO changes by several orders of magnitude until it becomes an 'ordinary' stellar system.
The angular momentum is $J \propto MR^2\Omega$ where M is the mass, R is the distance of the mass from the central force and Ω is the angular velocity. We see that for any change in R the angular velocity Ω changes together with it. In case of a gravitational collapse this continues until eventually the centrifugal force equals the gravitational attraction, forming a disk around the protostar since this balance is strongest along the rotational axis. The material on the outskirts of this disk is falling onto it and the material even further out replenishes the infalling dust.

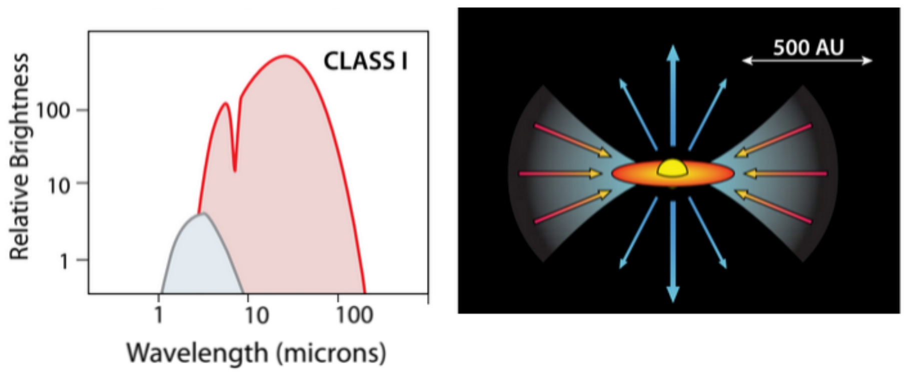

Figure 2.2: The spectrum of a YSO Class I in the left panel and the schematic of its morphology in the right panel, again from Werner & Eisenhardt (2019). The red spectrum is from the dust while the grey spectrum is from the central star.

In figure 2.2, the left panel again shows the spectrum. We can now see two spectra however, the red one coming from the dust and the grey one from the protostar. The peaks of the respective spectra are different due to the star having a higher temperature.
Compared to figure 2.1, the spectrum of the dust has become significantly brighter, both from heating of the material from the protostar and accretion. The red arrows in figure 2.2 again hint at the movement of the dust, in this onto the disk. The new blue arrows signify the outflow, sometimes called "jets", perpendicular to the rotation axis of the disk mostly caused by magnetic interactions. This phase lasts for some $\approx 10^5$ years.
When we look at the dust spectrum, at around $10\mu m$ we see a steep decrease in the relative brightness. This is because near the star, the dust disk is not continuous. a The part of the disk with the major contribution at that wavelength of the spectrum is "missing".

Note that the dust spectrum in the left panel in figure 2.2 does end before it reaches anything that would be considered radio-wavelengths. This is, as mentioned previously, because Werner & Eisenhardt (2019) focus solely on what they call the IR spectrum. There is of course radiation at radio-wavelengths, albeit significantly weaker than the IR radiation.
In order to give some context, we decided on creating a simple black body spectrum similar to the ones in figure 2.2 to show that there is of course a detectable spectrum at radio-wavelengths[4].

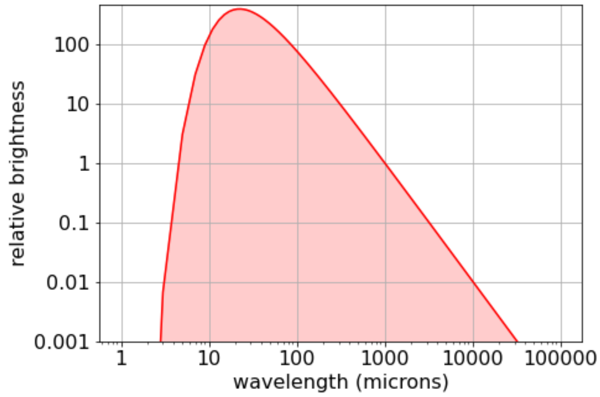

Figure 2.3: This plot shows a black body spectrum of one source with a temperature of ≈ 125K. This is just to visualise that there is radiation reaching radio-wavelengths, beginning at 1000 microns.

In figure 2.3, which was modelled after equation 2.1.1, we can see that the spectrum in the figures from Werner & Eisenhardt (2019) are just cut off at a certain brightness. If we allow the spectrum to extend further, we can see that radiation reaches the radio-wavelengths.

2.2.4 YSO Class II

Please note that we will not talk about the "flat spectrum" Class introduced in Greene et al. (1994). YSO Class II have evolved further, with the disk now smaller and more compacted than before. As for example described in Werner & Eisenhardt (2019), the envelope that was replenishing the accretion disk material has all but disappeared by now. The star is moving towards the main sequence in terms of mass, temperature and luminosity.
In the protoplanetary disk clumps of dust are slowly transitioning into planetesimals, what absorb more dust from surrounding areas in the disk due to gravity. This phase lasts longer at around 2Myr, as pointed out in Evans et al. (2009), who studied YSOs with Spitzer to determine star formation rates and efficiencies.

[4]For our purposes, "radio" wavelengths start at 1mm.

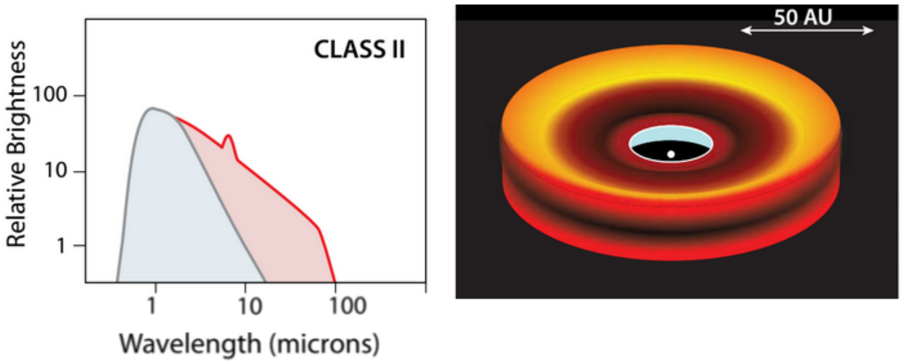

Figure 2.4: In the left panel of this figure from Werner & Eisenhardt (2019) we can again see the spectrum of the dust in red and that of the star in grey. In the right panel we see the morphology as a sketch.

Figure 2.4 shows us that the dust disk has shrunk by a factor of 10 compared to Class I. This is because the dust is falling into the star or forming into planets.
It should be mentioned that the disk in figure 2.4 is not flat as the sketch may suggest, but we will talk about the disk structure in general after going through the classes.
Since there is less material surrounding the star its brightness is now for the first time higher than that of the disk, as we can see in the left panel of figure 2.4. The black body spectrum of the disk has decreased due to its change in size, making the spectrum look almost like a side-lobe to the expected spectrum of the star.

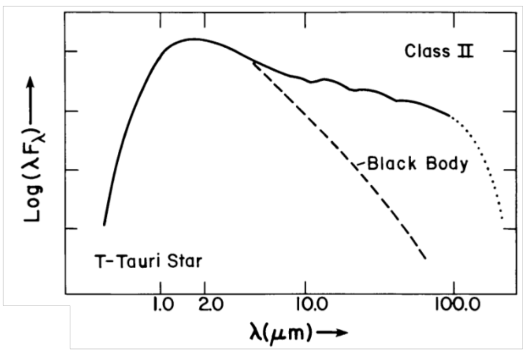

Figure 2.5: The original sketch of a black body spectrum of a YSO Class II from Lada (1987).

In figure 2.5, the mentioned additional radiation and its deviation from a typical black body can be seen better. Also in figure 2.5 we see that this is the usual spectrum of T-Tauri stars, which as described in Lada (1987) are relatively young stars with an age of less than 3 Myr and between spectral class F to M.
As the vast majority of the sources from the SODA sample are YSO Class II, we would like to point out a few things specifically relevant to this thesis.

1. As mentioned earlier the data later used is at a wavelength of 1.3mm. When looking at figures 2.5 or 2.3, it becomes apparent that the bulk of the radiation received at those wavelengths will come from the dust relatively inside the disk.

2. The sources in the SODA sample are for the most part unresolved. This means that any possible structure in the disk cannot be seen in this work. We will elaborate more on the design of the SODA survey at a later point.

3. As laid out in van Terwisga et al. (2022) itself, the cold ≈ millimetre sized dust from the inside of the disk was observed, not reflected light from the star as would be the case in infrared surveys, for example in Megeath et al. (2012) who used Spitzer, whose catalogue were later reused in van Terwisga et al. (2022).

2.2.5 YSO Class III

In the final stage of the evolutionary track of star and planet formation the former cloud looks almost like an ordinary stellar system. The star stars with hydrogen fusion, as for example mentioned in Williams (2021) and thus reaches the main sequence within a few million years, planet formation is in its final stages, the accretion disk has disappeared, being swept up by forming planets or being pushed back by stronger radiation pressure from the evolving star.
This remaining debris can in some cases still be seen in the infrared, as for example in current catalogues using Spitzer in Chen et al. (2014) or in Hughes et al. (2018) who show a review of recent advancements and list typical parameters of the debris disks.

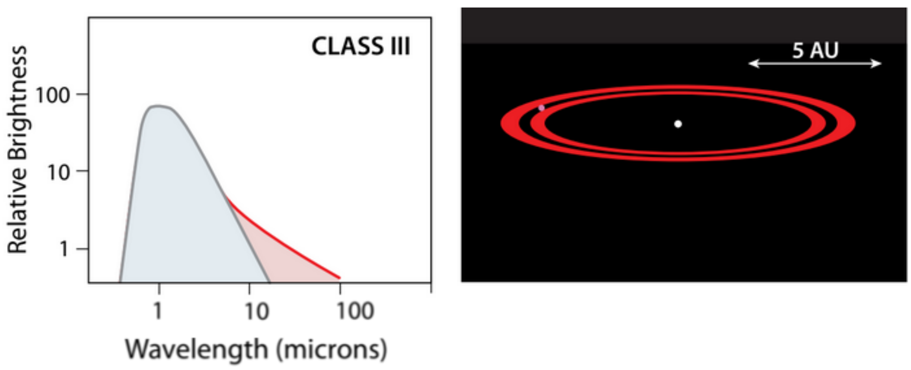

Figure 2.6: The final stage of YSO. In the left panel we once again see the spectrum of the star in grey and that of the disk in red. The right panel shows the sketch of the morphology, almost looking like an ordinary stellar system.

In figure 2.6 we can see in the left panel that the spectrum of the system is very close to a black body spectrum, only showing clear signs of the debris disk in the mid-, to far infrared. The system has further shrunk considerably, by about another order of magnitude. YSO Class III last somewhere in between 3 to 5 Myr.

2.2.6 Importance of YSO research

Armed with the knowledge from the chapter so far, we would like to elaborate further why these kinds of observations are carried out in the first place.
As mentioned in the description of the YSO Class III, the final product of the process is a stellar system. That means that we are left with one or multiple stars, planets and some smaller bodies both within the system and possibly an envelope around it.
Thus by getting a handle on for example the disk masses, we can make direct predictions of the later properties of planetary systems.
One point we would like to emphasise in particular is how the disk mass distribution changes over time, as we will later be determining disk masses.

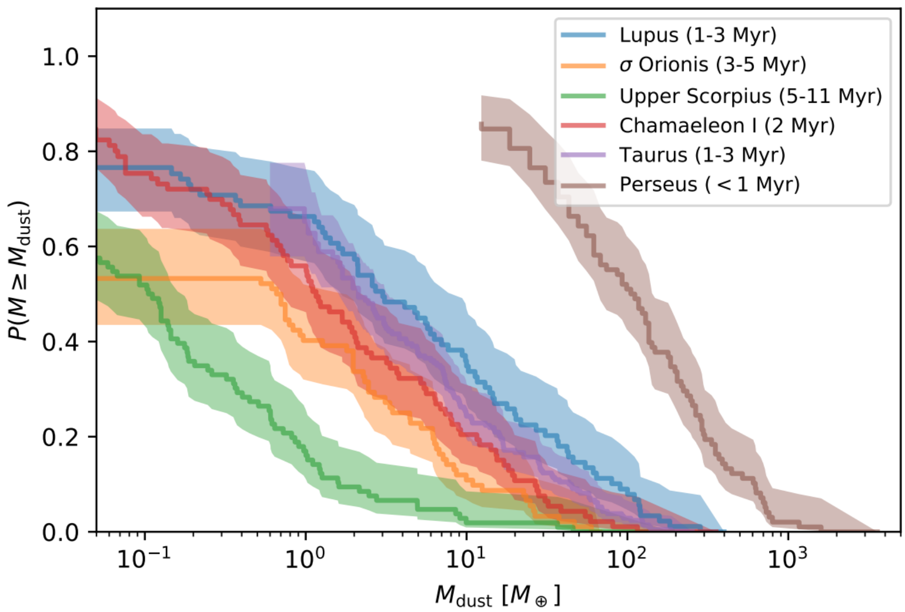

Figure 2.7: This figure shows multiple cumulative mass distributions of YSO disks. It is important to note that the younger disks are more massive than the older ones. The evolution of the shape of the distribution shows that over time it moves to the less massive side. Taken from Miotello et al. (2022).

In figure 2.7 taken from Miotello et al. (2022) we can see that the various disk mass distributions have the same general shape; relatively few high-mass disks with a lot of low-mass disks.

Lombardi, M. et al. (2006) state that the most common core[5] mass is at around $1M_\odot$ being about 30 times more common than a core mass of $20M_\odot$. The distribution for the younger populations being more massive overall. This is evidence for the fact that over time, these sources evolve and that they evolve similarly to each other. We can would expect this from the established theory thus far, as the disk masses decrease over time since it forms planets or falls into the star. This means that in figure 2.7, we would expect older populations to be further left than younger ones. Miotello et al. (2022) note that envelope contributions from Class0 and maybe Class I sources are however difficult to disentangle from the disk emissions.
van Terwisga et al. (2022) have a similar figure to Miotello et al. (2022).

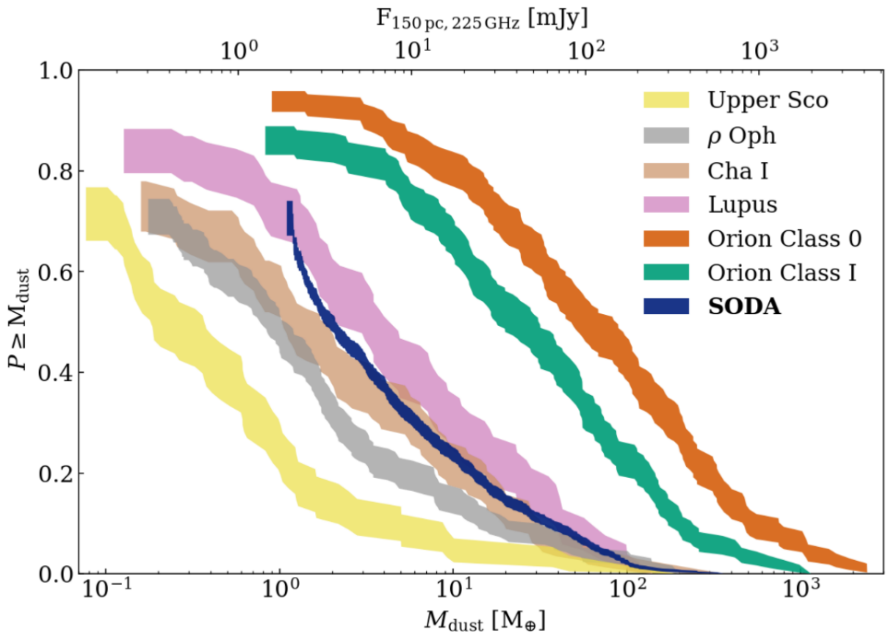

Figure 2.8: This figure shows multiple cumulative mass distributions of YSO disks, this time from van Terwisga et al. (2022). It should be noted that the colours are not the same as in Miotello et al. (2022).

In figure 2.8 we can see clearly that Class 0 and Class I sources are on the right, massive side while the SODA data is in the middle range. This is expected, as the SODA sample consists of Class II sources and fits established theory.

2.2.7 Structure of protoplanetary disks

As we have already gone through the various stages of the YSO classification scheme, it may be advantageous to also spend time elaborating further on the structure of the disks that will be

[5]Core here refers to a core inside molecular clouds that has started contracting. We would refer to it as a YSO Class 0.

worked with later.

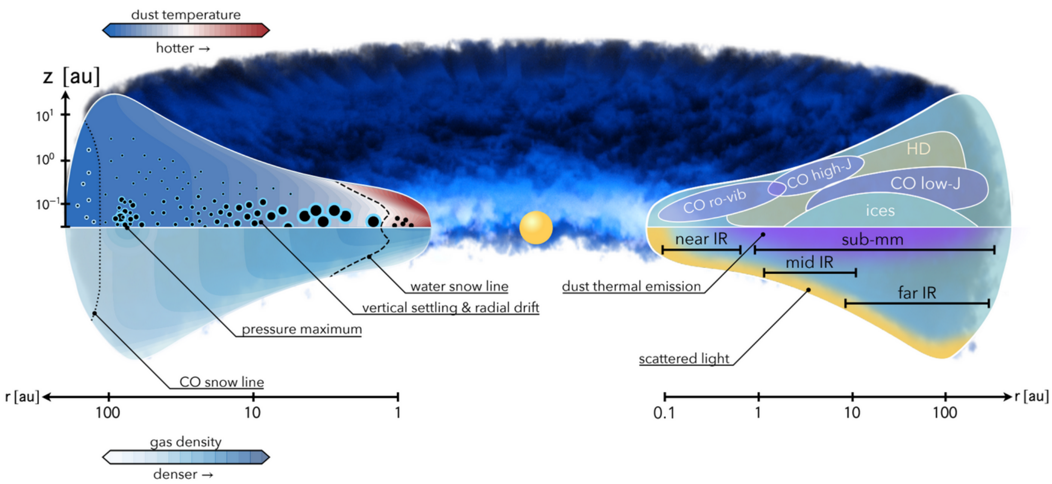

Figure 2.9: A side-view of a typical YSO disk. Please note that solely from morphology this is likely between Class I or Class II. The left side shows information about dust grain size and dust temperature, as well as gas density. The right hand side shows information about typical emission regions of molecular lines as well as continuum emission from either thermal radiation in the sub-mm and from scattered light in the IR.

Figure 2.9 consists of two illustrations. On the left side we can see the information about the dust grain size here depicted as black circles, the gas density in the lower half as well as the dust temperature in the upper half. We can also see a radius
The more interesting side for our purpose is the right side however, as there we find information on the emission regions. In the upper half we can see the typical line emissions from molecules, which play no role in the data that will be used later. The lower half is very important however, and reminds of a point we made earlier already. From the lower half we can see where it says "sub-mm", which is at the inside of the disk, not on the surface like the various IR regions are here. This once more emphasises that with ALMA we see the dust on the inside of the disk.
If we remember that the SODA survey analysed non-resolved YSOs of Class II in Orion with a distance of 454pc (see van Terwisga et al. (2022)) and the radius is here given at roughly 100au, we can check if that is indeed a valid size using the resolution of ALMA. van Terwisga et al. (2022) used the most compact 12m array of cycle 7 (C43-1), with a resolution of 1.47” according to Remijan et al. (2021). At a distance of 454pc, a disk with the radius of 100au would have an angular size of ≈0.45”, so it makes sense that almost all sources are unresolved.

Talking about ALMA, we also want to mention direct observations of YSO Class I and Class II, primarily in ALMA Partnership et al. (2015), who showed the first results of resolved YSOs with ALMA.

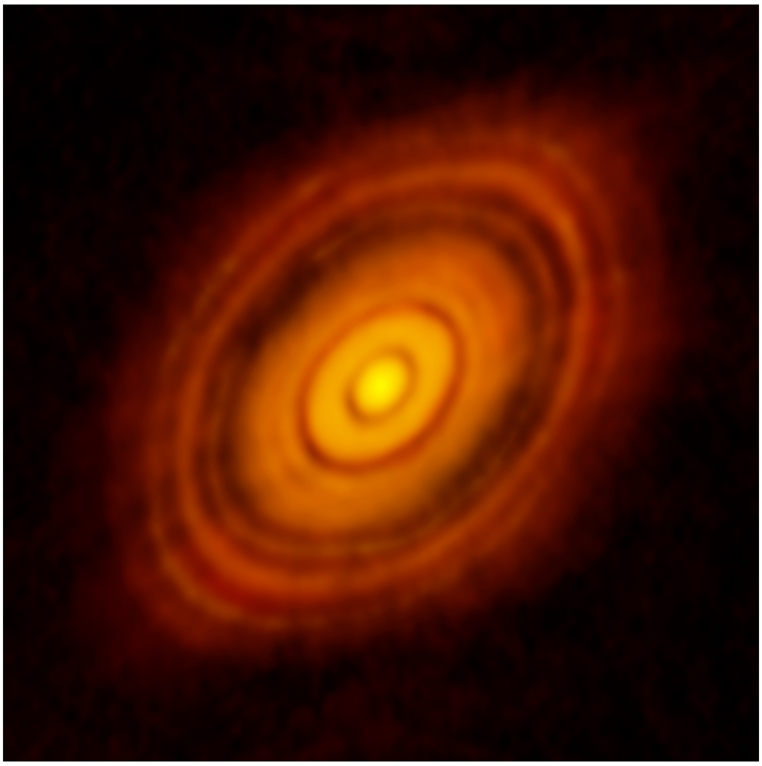

Figure 2.10: The ALMA image of HL Tauri, a Class I or Class II YSO at resolutions of a few au. We can see that the dust envelope has for the most part disappeared and importantly there are gaps visible in the accretion disk. This is likely due to a forming planet "clearing" the material in an ongoing planet formation process.

In figure 2.10 we can see an image made with ALMA from ALMA Partnership et al. (2015)[6] that shows clear signs of planet formation on the gaps of the accretion disk. The most likely scenario is that planetesimals have cleaned this region of material during the planet formation process. This goes to show the remarkable capabilities of ALMA, that we would like to talk about in greater detail.

2.2.8 Variability of YSOs

Now that we have covered the structure and evolution of the YSOs and have already mentioned some of the physical processes in them, we can talk about variability and its causes. We do this because one of our goals is looking for variable sources later in this thesis.
Variability can have many physical causes at very different amplitudes and timescales, that we would like to divide into four groups after Fischer et al. (2023), who give a summary of variability in their paper. We also reference Benz & Güdel (2010) who talk in more detail about magnetic

[6]taken from `https://public.nrao.edu/gallery/hl-tau-birth-of-planets-revealed-in-astonishing-detail-2/accessed18.01.2023`

causes of variability.

Change in accretion

As pointed out in Fischer et al. (2023), the main changes in luminosity come from a change in the accretion rate from the disk onto the star. Biazzo et al. (2014) give an overview of accretion rates comparing their measurements of accretion with some done previously, most notably Gandolfi et al. (2008), and find variability due to a change in accretion rates of a factor of three over a few years. Manara et al. (2023) explain the usual way of measuring the accretion rate, direct observations of shocked gas interacting with the photosphere of the pre-main sequence star (PMS) in the UV or visual. As both UV and visual wavelengths are prone to extinction, red visual or even infrared proxy observations are sometimes used for more embedded sources (as for example in Gully-Santiago et al. (2017)who use such observations to measure a heavily spotted T-Tauri star or White & Hillenbrand (2004) who use red optical observations to measure accretion in Class I sources.). For even more embedded sources, we must turn to far infrared or millimetre observations that observe either re-emitted light from the dust on the surface in the infrared or the thermal emission from the dust itself in millimetre.
We would like to briefly talk about the consequences of a variable accretion rate and how it may be detected.
Magnetic accretion: Stellar magnetic fields direct gas towards the star, it interacts with the photosphere. As mentioned earlier, this can be seen in the UV and visual, with Fischer et al. (2023) noting that X-ray heating occurs in the photosphere so that it may also be seen by X-ray telescopes.
Envelope emission: The envelope around a YSO is the material furthest away from the developing star. It can replenish dust that is accreted onto the star and otherwise remains around the forming system.
This envelope, particularly the outer part, is not heated strongly by the star, but primarily by the interstellar radiation field (Fischer et al. (2023)). Through careful measurements in long wavelengths of the envelope, changes in the its emission peak can be used to track the accretion.
Heated dust in the disk: The dust in the accretion disk is heated mostly by the star. The closer to the star, the higher the temperature will be with the closest region measuring > 1000K (Hartmann et al. (2016)) or $\approx$1400K (Fischer et al. (2023)).
Further outwards, the temperature declines to the order of 15K. Those 15K (in Fischer et al. (2023) its 10K-20K, papers utilising radio-observations have agreed on a convention of 20K, as can be seen in van Terwisga et al. (2022) or Draine (2006) for the sake of comparability) can be seen at long wavelengths in the millimetre range. As those temperatures change with the radiation output of the star, over a long time a change in accretion can translate to a change in flux from long wavelength observations.
Viscous gas heating: The gas at the inside of the disk can be heated to extreme temperatures that may even exceed that of the star through high accretion rates. This phenomenon is visible at UV and visible wavelengths and may even dominate the spectrum in some cases. Once extinction is taken into account, any change in brightness can be linked to accretion at those wavelengths. See Bai (2011) for the mechanism of viscous heating.

Stellar phenomena

Flares: Benz & Güdel (2010) explain that flares in YSOs come from a build-up of magnetic energy and later annihilation of it, resulting possibly in extremely bright radiation at all wavelengths usually over short times of less than a day to a few days in giants and subgiants, while lasting as long as 3 weeks in radio observations of less massive stars as for example in Slee et al. (1987).
Starspots: Starspots on the surface of a star can cause cyclic variation in the measured brightness due to them being colder than the surrounding material of the star according to Gully-Santiago et al. (2017) who try to measure the temperature of a spotty T-Tauri star. Starspots have a lifetime of weeks to months according to Namekata et al. (2019), with rare cases as long as a few years. It must be noted here that the perceived change in flux from a starspot is dependent on the rotation of the star, but is generally held to be in the order of one day to a few weeks.
Pulsations: In later stages of the evolution of an YSO, the expansion and contraction of the stars outer layers can also be responsible for a change in brightness, as pointed out in a review by Bhardwaj, A. et al. (2019) and in a review of pulsation in all types of stars in Eyer & Mowlavi (2008), who also state that the brightness change due to pulsation is usually low, at the order of a few 0.1mag at best and rarely last longer than a few hours.

Dust related causes

By "dust related causes" we mean any kind of change in brightness due to either a change in extinction in the line of sight or a process in the disk itself emitting radiation.
Dips: The so called narrow dips originate in a particularly dusty region in the disk or the envelope that briefly obscure the star. This can notably also be caused by a magnetically warped disk, as described in Bouvier, J. et al. (2007) who used spectra and photometry to find magnetic accretion rates and effects on the disk. This is of course only an observation artefact and not inherent to the source.
These dips can be at a timescale of a few hours to weeks.
Synchrotron radiation: Benz & Güdel (2010) and Vargas-González et al. (2023) as well as Mairs et al. (2019) note that high YSO birghtness changes may be seen in X-ray and radio-wavelengths at the same time, most likely due to (gyro)synchrotron radiation when electrons gyrate around a particularly magnetic region either next to the star on in the disk itself. This cause is not directly related to dust, but does manifest in a change of measured flux.
This can happen at extremely short timescales (< hour), as in Vargas-González et al. (2023) who record a millimetre flare in Orion with a duration of only $\approx$10 minutes.

Binaries

Whilst there are a number of interactions between binaries and the disk that can influence the brightness, only one really fits into the timescales that we will be working with later.
Binary Eclipse: Not exclusive to the context of YSOs, two binaries have a chance of eclipsing each other in such a way that one blocks the light from the other relative to us. This can of course also be seen in YSOs, as there is a chance of multiple stars forming in a system.
None of this of course changes the stellar variability, but it still manifests as a change in flux.

Binary Interactions: When two stars form as a binary system, their different evolution times can influence the behaviour of the respective binary partner. From Park et al. (2021) we know of instabilities linked to binary system that greatly change the flux coming from them. Those can come from a change in radiation pressure, magnetic field, gravitational attraction and planetary formation in later stages. Those processes can last for centuries but are typically on the scale of the orbital period.

Long term variability

We will only mention long term (many months to decades) variability causes briefly here, as the maximum observation time of the SODA sample is in the order of a week.

Outbursts: Outbursts in general, independent of their various types, are caused by long term accretion changes from the disk onto the star. According to Kenyon & Hartmann (1995) this may last for as long as a century or even longer. The various types of outbursts come from different properties of the stars in mass, evolutionary stage and environment. FU Ori outbursts for example are related to binarity and last for decades to centuries, while EX Lup outbursts only last for a year to a few years.
Fades: According to Shulman & Grinin (2022), fades are traditionally interpreted either as a cold relatively dense cloud eclipsing the YSO as for example in Natta & Whitney (2000), who model such clouds for later simulations or as a puffy disk close to the star, where material near the centre sublimates, as in Dullemond & Monnier (2010). Those fades last typically for months to decades.

Timescales and amplitudes

We would now like to a figure in Fischer et al. (2023), that nicely shows the amplitudes and timescales at which YSO variability can usually be observed.

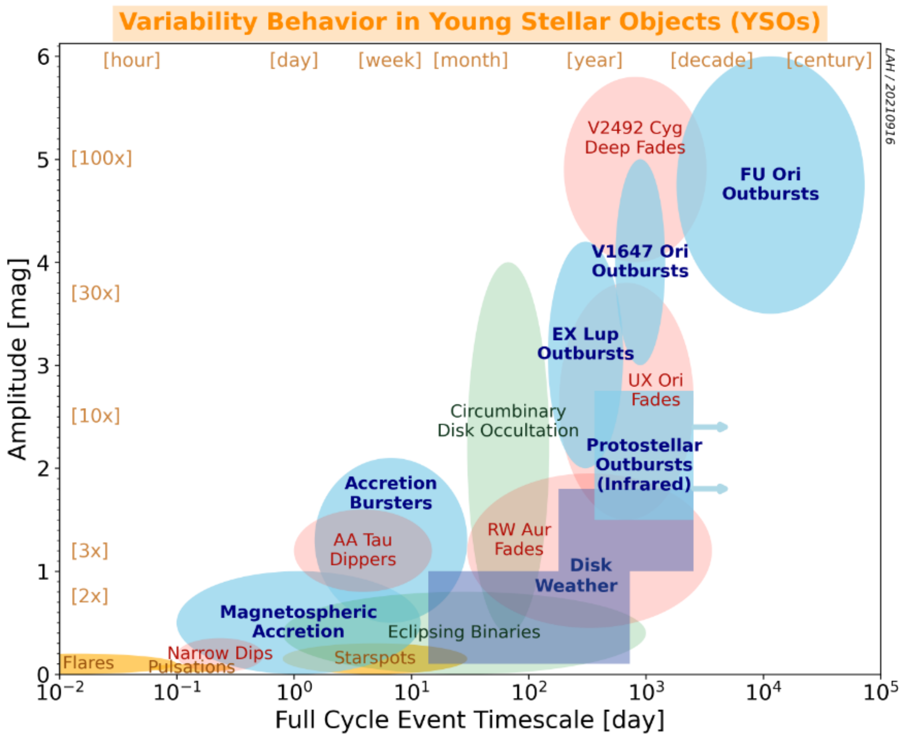

Figure 2.11: In this figure from Fischer et al. (2023) the x-axis is the full cycle event timescale, so the time for a periodic event to go through one full circle, and on the y-axis are two things, one the one hand a change in amplitude in mag as well as in orange an indication of how much the brightness changes in relative terms. Please note that the amplitudes given here are are the usual ranges and not the highest recorded ones. The yellow areas are variability due to stellar phenomena, red means changes in extinction, blue is related to accretion and green is for binary interactions. "Disk Weather" shows routine variability at longer wavelengths.

In figure 2.11 we can see the typical amplitudes of YSO variabilites along their respective timescales. Given that the longest timescale in the data from the SODA sample is around 5 days, the causes for potential variability are limited to everything at shorter timescales than the disk weather.

Please note that in figure 2.11 the gyrosynchrotron radiation from Benz & Güdel (2010) is missing. The changes in amplitude from this can, as in Vargas-González et al. (2023), Mairs et al. (2019) and Forbrich et al. (2017), be at an order of magnitude within less than an hour. Should any significant short term variability arise, it is likely due to gyrosynchrotron radiation.

2.3 The Atacama Large Millimeter Array (ALMA)

Note that this section is not meant to give a technical introduction, but rather give an overview of ALMA as well as underline its role in current research.

Prior to the start of operations of ALMA, other millimetre arrays achieved resolutions of somewhere in the range of some 0.1” (for example IRAM with 0.4”), which corresponds to tens of au in nearby star forming regions (Miotello et al. (2022)). This was enough to partially resolve big disks, maybe in the Class II stage, certainly enough for Class I, but only in the near SFR regions. With ALMA, the resolution has dropped significantly to 20mas (=0.02”) at the shortest wavelength of ≈0.3μm, corresponding to as little as an au. This is enough to fully resolve a disk and allow for images such as the ones from the NRAO from ALMA Partnership et al. (2015) in figure 2.10.

ALMA, as the name suggests, sits in the Atacama desert in northern Chile. The location was chosen on the one hand for very low humidity which would interfere at some wavelengths and on the other hand its height above sea level. At ≈5000m above mean sea level, the atmospheric window for radio-wavelengths allows otherwise impossible observations.
The earlier mentioned remarkable resolution of ALMA comes from its long baselines (a baseline is the distance between two radio antennas) of up to 16km according to the ALMA technical handbook Cortes et al. (2023).
ALMA can deploy up to 66 antennas of either 12m or 7m. The 12m antennas give a high resolution and the compactly arranged 7m antennas enhance the capabilities when observing extended targets.
This makes ALMA unique, both in number of capable antennas as well as their baselines. Combined with a somewhat big frequency range, from band 1 at around 40GHz (≈7mm) to band 10 at roughly 850GHz (≈0.3mm), ALMA can study multiple phenomena from planetary scales up to galactic scales, in either great detail or with large sample sizes, making it one of the most capable instruments currently in use.

2.4 Surveys of YSOs

Since we have covered YSOs, their importance and parts of the current understanding of them, we must now talk about how observations play into our understanding, particularly in view of this work.
Comparability and statistical significance are the two most important qualities for data to have when it comes to understanding bulk properties of one kind of source. Individual observations are certainly important for detailed observations of rare objects or phenomena, see for example the first resolved YSO images with ALMA that we mentioned earlier in ALMA Partnership et al. (2015). However for some purposes, it is paramount to carry out systematic surveys with the same instrument in order to gain a handle on the bulk properties.

2.4.1 Types of surveys

When conducting a survey, we typically weigh three factors against each other: time, details and sample size.
With a given observation time, one can either choose to observe many sources on a more superficial basis or look at relatively few sources in greater details or the middle ground of the two. Note that due to the nature of a survey its never exactly *few* sources, but few in comparison to what less focused surveys offer.
Of course if a project is given more time, it can be more extensive. Additionally, with new telescopes that have better sensors, are larger or employ other wavelengths these factors can be influenced without changing anything in terms of survey design.

Focused surveys

Focused surveys for our purposes are surveys that observe comparatively few sources for long times, such as the survey outlined in Vargas-González et al. (2021), where they observed six regions for in total four hours over multiple years each with the VLA with the goal of finding compact radio populations and detect long term variability in a densely populated field. This survey is considered focused, as the observation times are rather long and the number of detections, $\approx$520 are relatively low for the observation time.

Broad Surveys

A very broad (covering many different sources in a larger region) survey of protostars is the JCMT transient survey (see Herczeg et al. (2017) for details on the project), where sub-mm emissions from the ONC and other star forming regions are monitored for many years, as partially described in Johnstone et al. (2022). The findings in this survey range from millimetre flares in Mairs et al. (2019) to finding multiple variable sources that are thought to be linked to accretion in Mairs et al. (2024). It should be noted here that the JCMT transient survey is a single dish-telescope, and thus is not an interferometer.

Another broad survey conducted in the infrared by Megeath et al. (2012) finds as much as $\approx$3500 likely protostars with Spitzer in 9deg^2 of area in the sky. This survey is also where van Terwisga et al. (2022) got the positions of their sources from.

2.4.2 The SODA survey

The SODA survey, as described in van Terwisga et al. (2022), is a rather broad survey of YSO Class II sources in the L1641 and L1647 clouds. This survey is special for a few reasons:

1. The sample size of almost 900 sources of just one Class of YSO stands out in literature, as other surveys with comparable amounts take many years, as for example the 1200 infrared selected YSOs from Johnstone et al. (2022) with the JCMT.

2. The limited region of the YSOs, being located only in two clouds with a high number of sources, yields a very high statistical certainty compared to surveys with similar targets. Tobin et al. (2020) for example list a study of YSO Class 0 and Class I sources with ALMA and VLA, finding around 470 sources in the Orion molecular cloud. Whilst of course respectable, it is half the sources of van Terwisga et al. (2022), divided into two Classes.

3. As pointed out by Miotello et al. (2022) or Manara et al. (2023) millimetre observations of YSOs are rather important and have recently become available with ALMA, not a lot of large scale surveys of such sources with it have been done.

The SODA survey is the largest survey of its kind with a sample size of unresolved 873 YSO Class II carried out in the millimetre wavelengths. It allows, as done in van Terwisga et al. (2022), for the creation of statistically relevant subsamples with a still small error. Furthermore it is possible to asses variability as well as spectral indices due to the structure of the observations.

3 Methods & Tools

This chapter will elaborates on how we reduce the raw data and which tools and programs we used in the process. we also talk about how we handle the variability, spectral indices and where the errors of the later chapters come from.

3.1 Interferometry

In this section we would like to introduce interferometry and explain how it works in the context of ALMA. The contents of this section are taken from the ALMA technical handbook for cycle 10 Cortes et al. (2023) and from Thompson et al. (2017).
In simple terms, interferometry uses multiple images from separate antennas or telescopes. The signals of them are interfered, allowing for a sampling rate that is smaller than that of the corresponding single antenna or telescope. By sampling rate we mean the mapping of brightness of the sky. An interferometer really measures emission that is of the scale of the angular extent of a sinusoidal wavelength, which is what is called a spatial frequency. That wavelength is inversely proportional to the distance between the individual antennas, which is called a baseline. When observing an object with an interferometer, we really sample a selection of different angular sizes by using more than just two antennas. The rotation of the earth even helps with that, as the projected distance between antennas change with it, leading to more different angular sized to be sampled. This means that it is advantageous to space out multiple observations of one source over the day to get as many different angular sizes as possible.
After the observations are done, the data is combined using a Fourier transform to get the sky brightness distribution; a map of the sky that shows brightness for different positions.

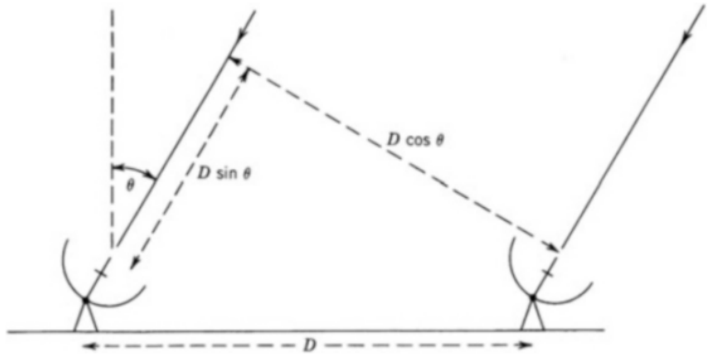

Figure 3.1: A sketch of the simples possible interferometer. D is the baseline and θ is the direction of the wavefront that is measured. Taken from Thompson et al. (2017).

The two antennas in figure 3.1 are separated by a baseline of D, and are observing the source at an angle of θ away from zenith. As already explained, the interferometer just observes a sinusoidal that is sometimes referred to as a 'fringe', the spacing of which is given by

$$\text{Fringe spacing} = 1/(D \cdot \cos(\theta)) = \lambda/(L \cos\theta) \tag{3.1}$$

In equation 3.1 D is the physical distance between the two antennas and θ is the angle away from zenith. λ is the observed wavelength and L is the distance between the antennas in units of λ. From this we can see directly that the fringe spacing is inversely proportional to antenna distance, thus compact arrays would observe extended structures a lot better.

3.1.1 Image & Deconvolution

The limit to what we can see in astronomy, but particularly in radio-astronomy is dictated by diffraction. The FWHM of the minimally resolvable disk, called an Airy disk, is give n by

$$FWHM \approx 1.22\lambda/d \tag{3.2}$$

Where in equation 3.2, d is the diameter of the telescope aperture. From this we can see that one single antenna is always limited by how big we can build them. After a certain point, our ability to build ever larger antennas will be limited by the construction itself, not to mention increasing problems with upkeep and repairs that come with larger antennas.
If we want to obtain higher resolutions, we need to combine the signals from separate antennas. This process is commonly referred to as 'aperture synthesis'. With it, we can increase the d in equation 3.2 as high as the distance of the antennas!
For any observation with ALMA, what we really get in terms of data is a convolution $I(a)$ between the point spread function (PSF)$PSF(a)$ and the actual signal $O(a)$ of the image. $I(a) = PSF(a) * O(a)$ If the system is only influenced by diffraction, the PSF would be the Airy disk. In general the PSF is the response of the system to a singe unresolved point. In simpler terms, it describes how light from a single point in the object space is spread out or distributed in the

image space. The PSF is affected by factors like diffraction, aberrations, and the optical design of the system, as explained in great detail in Thompson et al. (2017).
The issue is that we would like to see only the actual object $O(a)$, and not the convolution with the PSF that becomes $I(a)$. Luckily, ALMA observations also come with information of the PSF so that we can take the convolved image $I(a)$ and deconvolve it to get the best possible image of the actual source $O(a)$. We will not go into the details of the deconvolution, as the exact process is somewhat complicated and requires more consideration, but the basic algorithm, called 'CLEAN', is as follows, after Högbom (1974).

1. Take the brightest pixel in the image.
2. Create a 'clean map' through addition of the idealised PSF at that point.
3. Create a residual map through substraction of the actual PSF ('dirty beam') from the image.
4. Repeat until a certain residual threshold is reached or a certain number of iterations were finished.

What we are left with afterwards is a sum of the clean map and the residual map. Ideally, the residual map is very faint compared to the clean map. Please note that this method is not suitable for extended targets, but that is out of the scope of this work.

3.2 Observations with ALMA

In this section we will explain how ALMA observes targets, how the data is structured and how the observations we got fit into it.
The structure of ALMA data starts with the observing project, which is the whole suite of observations for one project. Below that are science goals, groups and members respectively. A science goal can have multiple groups, who could in turn have multiple members. The members are broken up into scheduling blocks (SB). The blocks are fed to a control software that creates so called execution blocks (EB), that contain information on the parameters and measurements of the condition for that SB. A SB may have multiple EBs, if the conditions require several observations until a certain integrated signal strength is achieved. The EBs are broken up further into scans and subscans, which in turn consist of individual integrations. The calibration for each of the EBs are carried out together with each scan or subscan. All EBs of one SB can be merged into one data cube with the calibrations taken into account of the individual scans.

3.2.1 Data structure

ALMA data are stored in folders according to the structure, so project → science goal → group → member, and so on. SBs come in folders that contain one file for each EB, that are so called .ms files. The ms stands for 'multiple sets', as multiple observations are packed into one of those files. Every SB comes with its own calibration, the information for the calibration source and the raw data of the observation itself. Additionally, there is information on the weather and some information about the technical equipment itself, for example the temperature of the receivers

during each measurement.
The .ms files themselves really are tables of the various sources, their respective fluxes, observation times (beginning, end and integration), spectral windows (spw) and information about the observer, the array, names of the sources and some others that are of no concern here.

3.2.2 Our data

The data we got (project ID: 2019.1.01813.S, PI: S. van Terwisga) consist of observations carried out between the 21^{st} of November 2019 and the 2^{nd} of December 2019. As previously mentioned, 873 YSO Class II sources were observed by van Terwisga et al. (2022) as part of ALMA Cycle 7. All observations have been carried out with the most compact formation of the 12m-array, C43-1 to be specific. All observations had a minimum of 41 antennas at any given time, with average water vapour of precipitated water vapor = 1.5 - 3.0mm. All sources have been observed in Band 6 with 4 spectral windows of 224, 226, 240 and 242 GHz which correspond to spectral window 17, 19, 21 and 23, respectively. The bandwidth of each window is at 1.8GHz and the primary beam for this ALMA configuration is around 1.1”. As stated in van Terwisga et al. (2022), the reason for this to maximise the detection of line-free continuum emissions from point sources with the shortest possible integration time of only 30s each. Every 20 minutes, phase calibrations and atmospheric calibrations are carried out using J0541-0541. The observations are carried out in 6 SBs, that are later referred to as ’batches’. In total, there are 13 EBs, 3 for SB one, two and four, and 2 for SB three, five and six. Thus, for example, M12_152 was observed three times, one for each EB as it is in SB one.

3.3 VSC

We used the Vienna Scientific Cluster (VSC) for the reduction of the data, more specifically we used the VSC5 for all the reductions. The VSC5 consist of 1540 central processing units (CPUs), each of the of the AMD EPYC Milan series. The VSC5 is the fastest supercomputer in Austria available to the university of Vienna.
The parallelisation itself was done in the simplest way possible. That is, the CASA script was run simultaneously in six batches of files. So while CASA itself did not run in parallel, the process was still technically parallelised. This arguably simple approach already yielded a significant reduction in the time needed. On the server of the IfA, also called ’licancabur’, one reduction takes, depending on the number of observations and the object itself, somewhere between 15 to 25 minutes.
On the VSC5, one single reduction takes 4 to 6 minutes, without any means of parallelisation implemented.

3.3.1 Settings

In order to use the VSC, one typically uses the cluster management system called ’slurm’. Slurm can execute files submitted to a queue depending on how many CPU cores are available and how

many cores the script will need.
To find out the most optimal settings for running the VSC, we tried different configurations. The aim was to find out two things:

- The optimal number of CPU cores: While it might seem obvious that more cores mean shorter times, the issue is to find the optimal point, where the reduction is fast but slurm does not schedule the process too far back due to the number of cores requested. We tested one, two, four, eight and sixteen cores for this purpose.

- The location of the .ms files: Should the .ms files be copied over to the temporary directory with every reduction, increasing run time due to copying but decreasing run time due to utilising the faster solid state drive (SSD) speeds, or should they be read from the hard disk drive (HDD) data directory?

We tried each combination with four identical sources to find the ideal settings.

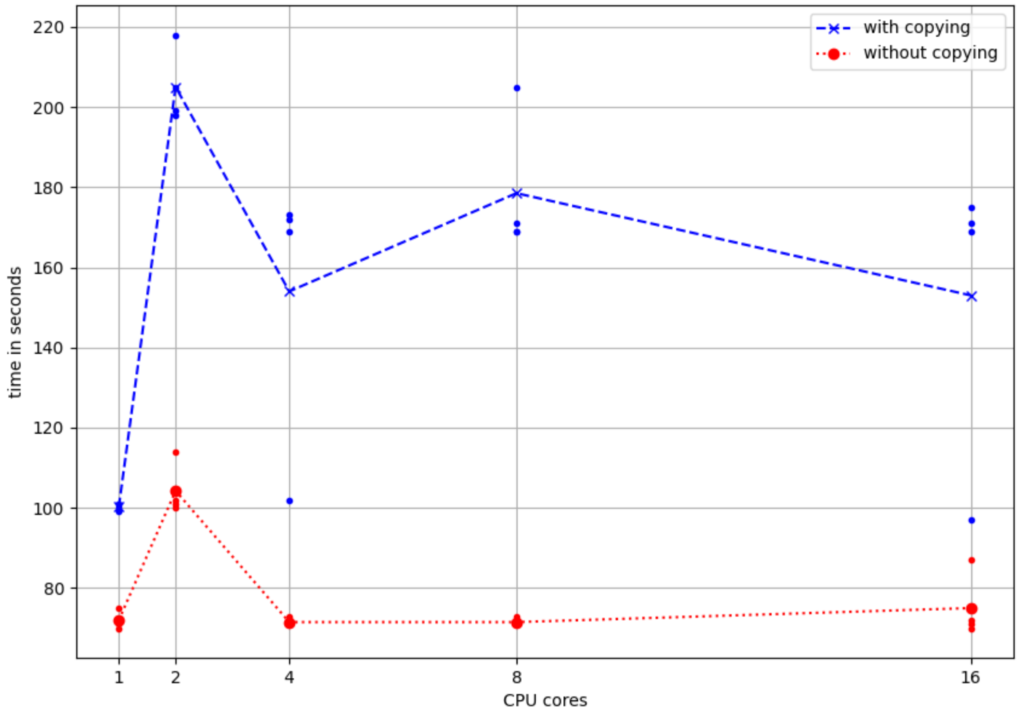

Figure 3.2: Measured time (sys) on the VSC5 for each settings. The red points are without copying the .ms files meaning that they were directly read from the slower HDD but saved time by not copying them to the directory. The blue crosses are the times for when the files were coppied to the SSD. The smaller points of the respective colour are the individual times, the larger markers are the average of four different tests.

Figure 3.2 shows a comparison of core-counts as well as two different approaches to reading the data. The y-axis is the 'sys' time in seconds, the exact nature of which we will explain a bit later. On the x-axis we can see the number of processor cores. The blue dashed line denotes the time it took for the reductions with copying the files from the HDD to the SSD, and the red dotted line shows the same with reading the files off of the SSD directly. The small dots in both colours are the individual times, since 4 reductions were carried out for each try. The blue crosses and red larger dots mark the average time for each core count, respectively.
From figure 3.2 we can clearly conclude that directly reading the files from the hard disk is the faster of the two in all cases. It should also be noted that, somewhat unexpectedly, more cores are not always faster. This is most likely due to CASA and not due to the VSC as it is designed to be faster when utilising multiple cores. We would also like to explain the 'sys' time: 'sys' represents the CPU time utilised within the kernel by a process. This encompasses the execution time spent

in system calls within the kernel, distinct from library code execution in user-space. Note that it is not the actual time the whole process takes, which was usually about 4-6 minutes, but still offers comparability.
For the times[1] we got for the non-copy method, the number of cores other than two were all close, so we decide on one CPU core per task as the scheduling time in slurm is lower for low core counts.

3.4 The Common Astronomy Software Applications Package

Common Astronomy Software Applications, or CASA for short, is the main data reduction and analysis tool for ALMA and VLA data as explained in The CASA Team (2022). It is the primary software with which we will extract the fluxes and analyse the variability and spectral indices. To be specific, the deconvolution as well as the handling of the file creation is done solely by CASA. We also use CASA on the VSC5 as a stable python environment that has access to the inbuilt packages such as astropy and numpy. We used CASA version 6.5 and 6.3, with CASA 6.3 being the version installed on the VSC5 and 6.5 on the server of the IfA.

3.5 Reduction

The reduction process happens in three steps. Note that it was started with the already calibrated data from van Terwisga et al. (2022) and provided by A. Hacar. Also note that the reduction process differs from van Terwisga et al. (2022) which will yield different results. The exact difference in reduction will be explained later in section 3.5.1. We believe that those differences should be small however and will check this during the process as well as afterwards.

3.5.1 Creating .fits files

First, the files need to be transformed from the calibrated raw data, so the automatically calibrated data that is given by ALMA[2], given as the .ms files to .fits files of the individual sources. There are multiple .ms files for each batch (SB) of sources; either two or three (EBs) depending on when a certain threshold is reached. This also is the basis of the inspections on variability, since a given object might change its flux in the time between observations.
The goals as described in chapter 1 require the creation of several files for each object, namely:

- One file for each time step in both windows combined. We need these files to check for variability, the exact spectral windows for that are not relevant.
- One file with all times in each window. This is what we used to check for spectral indices. Note that all observations from all times are combined into one for each spectral window.

[1] Note that linux outputs 3 times; 'real', 'user' and 'sys'. The earlier stated 4 to 6 minutes are describing the real time, and the y-axis in figure 3.2 describes the sys time.
[2] That was supplied by A.Hacar in our case

- One file with all time steps combined in both windows. This is what we use for the analysis of the total flux and what we use to calculate the disk masses.

Note that in this work the combined files from both windows may be referred to as *full* files.
For each of the files described above, there also exists a copy of the file with a narrower cut around the centre for purposes of noise analysis.

To create these files from the .ms files, we first use the functions `split` and `concat` from CASA. These two functions serve a similar purpose, just in opposite ways. `split` splits up a .ms file based on some parameter, i.e. a certain source or field, and `concat` 'glues' it back together into one file. This was used to only extract the observations at one spectral window from the whole .ms file for example. In this manner, we extract all the different parameters that will later be tunred into .fits files. The next step is using `tclean` to convert the data from the interferometric images back to the deconvolved image of the source. Deconvolving our data without using `tclean`[3] is not possible. It deconvolves the image from the original interferometric data received by ALMA, as explained in the interferometry part of this section. This function takes a lot of parameters laid out later and makes the bulk of the computing time together with splitting and concatenating. Every observation has to be entered into the `tclean` function of CASA with a few important parameters, namely:

- `threshold`, a threshold in Jansky (= $10^{-26}W \cdot Hz^{-1}m^{-2} = 10^{-23}erg \cdot s^{-1}Hz^{-1}cm^{-2}$) for CASA to know when to stop. It stops, when the residual reaches this value. For this work, `threshold` was chosen to be 0.001mJy after some testing. For most sources, this threshold is reached rather quickly and the error is hardly different when compared to an even lower threshold.

- `niter`, a number for how many iterations CASA is supposed to do in case the threshold isn't reached. In we decided on `niter` = 10000, as most sources reach the threshold earlier and thus don't require that many iterations.

- `imsize`, the number of pixels used for a given observation. For all observations an image size of 500 by 500 pixels was chosen after some experimentation with different image and pixel sizes. 500 by 500 was the best middle ground in terms of computation time and low errors.

- `cell`, the size of one individual pixel in arcseconds. The pixels are chosen to be 0.08" on each side. It is strictly necessary that the pixel size is at most $\frac{\text{beamsize}}{5}$ due to the Nyquist criterium [4], which is easily fullfilled as the beamsize is at least $\sim 1.45''$.

- `major` and `minor`, needed for the `imsmooth` function. It indicates which dimensions the smoothed source is supposed to have if it was a 2D gaussian. In this case, both axes are the same size at 2.1". .We chose 2.1" after analysing the whole sample once. It was the overall

[3] Note the similarity to the earlier noted CLEAN algorithm of Högbom (1974), see section 3.1.1.

[4] in `https://science.nrao.edu/facilities/alma/naasc-workshops/nrao-cd-austin/CASA_Imaging_Austin.pdf`(accessed 1.1.2024) state three to seven pixels per beam.

biggest extent of one source, and thus the lowest possible one for all of them. This is also the reason why the angle of the object does not need to be specified. Note that this value stems from the objects itelf, as CASA cannot deconvolve over a source that is larger than the given size, so naturally the largest object size from the SODA sample was found and chosen as the `imsmooth` size for all objects.

The settings were determined by trying multiple variants of each and weighing the accuracy to the calculation times. The `cell` and `imsize` parameters had little influence on the resulting value with moderate influence on the error, but large influence on the computation time. 500 by 500 with 0.08 arcseconds is a reasonable middle-ground, as the errors were small and the computation time was moderate. Also, the `imsize` of 500 by 500 pixels is valuable for later direct inspection, particularly when thinking about variability, binarity and irregularities in general.
After the `tclean` is done, we correct the deconvolved image with the `impbcor` function. This command is used for what is called 'primary beam correction', which is necessary as the deconvolution adds noise towards the edge of the images. The primary beam[5] is the central gaussian-like feature, and arises naturally during the reduction process in CASA. Due to the primary beam response, the sensitivity of the observation further away form the centre of the primary beam becomes less. With the `impbcor`, we can correct for this step, but the noise away from the centre of the primary beam becomes amplified, which is also the reason why we cut it off, the details of which we will explain later. with a cutoff of 0.8 inside CASA for the files with which we analyse the errors. This means that if the amplification by the primary beam correction reaches $1/0.8 \approx 1.25$, CASA will automatically set a cut at this radius away from the centre. The cutoff value for the normal files is set to 0.4 after some testing.
We then smooth the images using the `imsmooth` function of CASA. `imsmooth` allows for convolution of the deconvoluted image with some additional possibilities, like choosing the new size of the source. It also reduces some of the noise in the image, but of course loses detail, should the target be (partially) resolved.. Note that this was not done by van Terwisga et al. (2022). The smoothing was carried out in order to avoid fitting asymmetric 2D gaussians, as the shape of the gaussian would have differed in the different windows, making comparisons and spectral indices difficult if not impossible, particularly in limited time. After this, the image is turned into a .fits file using the `exportfits` function to be analysed later.
We have now converted the .ms files that were supplied by ALMA into usable .fits files to be inspected further.

3.5.2 Parsing of the .fits files

For every object, multiple files are created. In case of three observations 12 and in case of two observations 10 files are created. The 12 files consist of one file for the separate times in all windows (3), one at all times in separate windows (2) and one for all times in all windows (1). We do the same for pure noise analysis, making it 12.
The basic outline of the analysis is as follows:

1. We make use of the python package 'astropy' (see Astropy Collaboration et al. (2022)) to import and work with the .fits files through the whole process.

[5] sometimes called antenna beam size

2. Each pixel that is within 20 pixels of the centre is now compared to four times the standard deviation of the noise image with the 0.8 primary beam cutoff. If the flux of a certain pixel is higher than this 4σ, the pixel position will be used for the further calculations in all other files. This comparison is done with the full file of all time steps in all windows. The results of this are in flux per beam.
   We chose 4σ through some tests early on and found that it yielded good results while still clipping reliably. Please note that, from pure noise, the number of pixels above the cut based on our image size of 250000 pixels would be around 16 with a 4σ cutoff. This may seem like a problem, but since we only check for contributions to flux within 20 pixels of the centre, the number of 'random' pixels above the 4σ cutoff lowers to 0.09, which is not an issue.

3. If the flux of the pixel is not higher than four sigma, it is used for the calculation of the noise. Note that several `nan` values are simply discarded from the border of the image due to the circular shape of the aperture.

4. With the knowledge of the pixel size and the beam size the flux per beam is being converted into the total integrated flux using

$$F = \sum_i \frac{x_i}{\frac{A_{beam}}{A_{pixel}}} \tag{3.3}$$

   where the area of the beam $A_{beam} = \frac{\pi}{4} d_{beam}^2$ and the pixel area $A_{pixel} = a_{pixel}^2$ since the pixels are squares. d_{beam} is the diameter of the major beam and is extracted from the .fits header, and a_{pixel} is the side of one pixel, fixed at 0.08" in the reduction process. x_i is the flux of the pixel above the sigma cutoff.
   Note that d_{beam} varies throughout all sources, increasing with later batches. It does that because d_{beam} depends on the weather[6] which seems to have gotten worse for later observations.

5. The noise is calculated using

$$\sigma = \frac{rms}{\sqrt{N_{clip}}} \tag{3.4}$$

   where N_{clip} is the number of all the pixels above the clipping and rms is the root mean square of every pixel that is not above the 4σ clip. The rms is calculated as $rms = \sqrt{\sum_i (x_i - \mu)^2 / n}$, where x_i are the respective flux values from every pixel, μ is the median of the the flux distribution over the whole picture and n is the number of pixels under the threshold. For practical purposes, μ is always close to 0.

6. With the total integrated flux and its respective noise, the spectral index can be analysed as well. The spectral windows[7] as we explained earlier are at 224, 226, 240 and 242 GHz each. From these windows the lower and upper two are combined, and the average is taken for the

[6] And on some other factors like the projected angle of the array, but the weather seems to be the biggest contributor.
[7] A spectral window could also be called a channel, but since CASA calls them window, we will also stick to it.

frequency, so in the end the windows with which the spectral index will be calculated are the lower window at 225GHz and the higher window at 241GHz. β is calculated as stated in equation 2.7. Please note that the frequency is given in GHz in all further calculations. The errors of β were calculated with the gaussian error propagation law.

7. All of the relevant values are now written to text files for statistical analysis and visualisation.

3.5.3 Data Processing

For the process we utilise the VSC5 as well as the server of the IfA. For some of the visualisation after the data has been reduced and is written to a file, no server or similar is required. For the reduction on the VSC5, we make use of 'slurm'. Slurm stands for 'Simple Linux Utility for Resource Management' and is the software with which the VSC5 schedules the tasks and creates log files. The basic sequence of the reduction process is as follows:

1. We create the necessary files with a python script
2. The files are started and scheduled inside the VSC5 by slurm
   a) A temporary directory is created inside the VSC5
   b) The necessary files are copied into the temporary directory [8]
   c) The reduction is carried out in the directory as usual
   d) The .fits and .txt files that are required for analysis are moved out of the temporary directory
   e) The rest of the files are deleted automatically by the VSC5
3. We copy the final files for the visualisation manually to the server of the IfA.

3.5.4 Testing for differences in CASA versions

As outlined in the previously, there is a possibility that the different CASA versions cause the a difference in the flux. This would be a grave issue within CASA that we could not leave unchecked. For this purpose, three of the higher flux observations will be reduced with the settings explained in 3.5 and subtracted from the .fits files of van Terwisga et al. (2022). They used CASA 5.6 while we used CASA 6.5 and 6.3.

[8] Note that the .ms files of the observation itself are not copied there but are just read from the VSC5 data directory, more on that later.

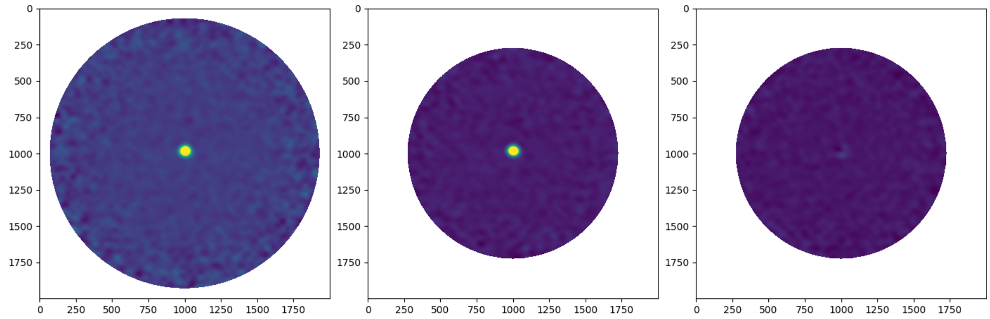

Figure 3.3: M12_49; the left panel is SODA data, the middle panel is the data as reduced in 3.5, the right panel is the residual. Note that the x and y axis are in pixels, which is sufficient for the purpose of comparison.

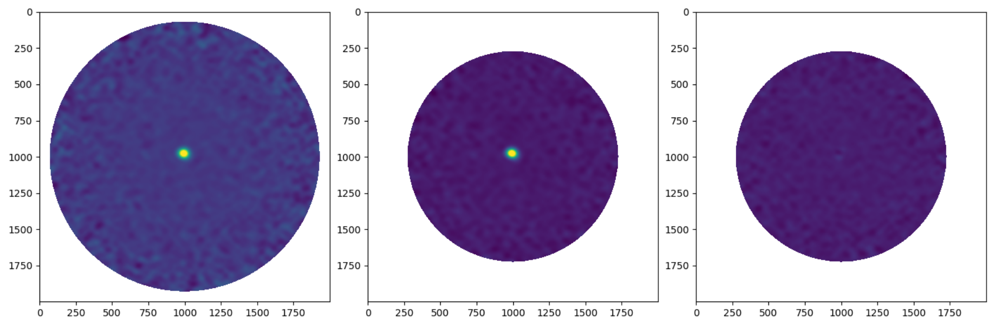

Figure 3.4: M12_52; the left panel is SODA data, the middle panel is the data as reduced in 3.5, the right panel is the residual.

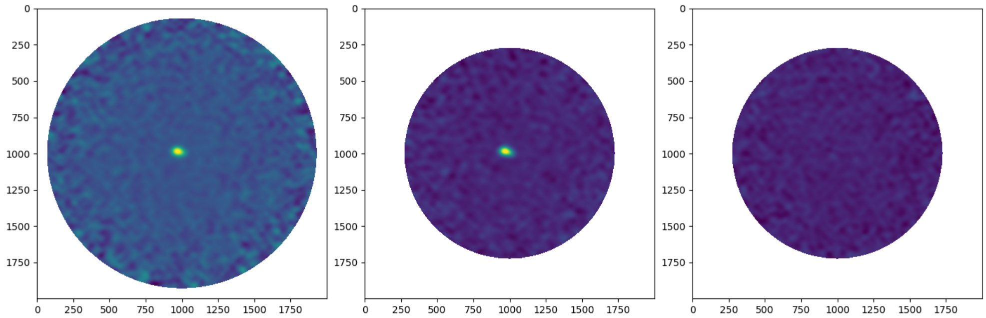

Figure 3.5: M12_105; the left panel is SODA data, the middle panel is the data as reduced in 3.5, the right panel is the residual.

In the three figures above, the left panel is the SODA data from van Terwisga et al. (2022), the middle panel is the reduction from this work and the right panel is the residual. The colour scale is the same for each source.
As apparent in figures 3.3, 3.4 and 3.5, the residuals are small enough around the central area to rule out a significant contribution from the different CASA versions. Note that the aperture circles are of a different size, due to our cutoff explained in section 3.5. This is not important since the outer regions do not contribute to the flux measurements.
Please note that we did the same subtraction with some other objects, not just these three.

3.6 Alternative attempts

Briefly, we would like to mention some other ideas we tried to implement, particularly in regards to the analysis.
In the beginning, for a still unknown reason, there were issues with the calculation of the fluxes. Batches 1, 2 and 3 were similar to what they are in the final reduction, but for whatever reason the other three batches had extremely large errors, almost 10 times the ones with the current reduction, that made any analysis of those sources impossible for practical purposes.
This is why the exact process of the reduction has changed multiple times throughout the creation of this thesis. Originally, we did not do the the circularisation (setting the minor axis of the beam to be the same size as the major axis) that was carried out with the `imsmooth` function. We then decided to do the circularisation in order to get more consistent results and to simplify the calculation of the amount of pixels contributing to the flux. After this step, the problems started arising. Fixing the errors and their inconsistent behaviour took multiple weeks. The issues got even worse when all batches behaved like the last three, making the whole sample almost unusable. The only thing that helped was partially rewriting some of the code for the error extraction.

Later in the analysis, the few binaries in the sample caused some issues. They behaved in-

consistent and yielded high variablilities that were not visible in manual inspection. To counteract this, we experimented with fitting a 2D gaussian to some sources. While not working perfectly, this would be doable in principle, however it was clear early on that for a lot of the sources the fit would not work, at least not properly. The consistent implementation of this fitting would have required more time than available at that time in September.
The issue with the binaries was resolved later with the ’forcing of the centre’ technique, that forces the program to only look for the brightest pixel within a 20 pixel radius from the centre.

4 Results

4.1 Comparison to van Terwisga et al. (2022)

In this section we will compare the results in terms of flux and thus the calculated disk mass against the original paper by van Terwisga et al. (2022). We do this to check for drastic differences in methodology and how our results differ. Should the results differ by a lot, this could hint at mistakes made on either side.

4.1.1 Flux

The analysis of the fluxes is part of the first aim of this work and we will thus talk about it in detail. The flux is also the first check of whether or not our methods for extraction of the fluxes were successful.
When we compare the received fluxes from this work to the fluxes from van Terwisga et al. (2022), there is a moderate difference. Throughout all fluxes, we see that the method we chose yielded lower fluxes, but consistently so. This is not unexpected since, as mentioned before in section 3.5, the reduction processes are different by quite a bit.

M12_X	van Terwisga et al. (2022) [mJy]	our values [mJy]
3	0.38±0.10	-
7	0.49±0.09	0.36±0.02
8	0.85±0.10	0.75±0.03
11	4.09±0.10	4.06±0.04
17	10.11±0.09	9.85±0.05
20	17.49±0.11	17.39±0.03
22	0.96±0.08	0.25±0.04
26	0.96±0.08	-
28	0.92±0.10	0.64±0.05
29	0.38±0.09	-
33	0.40±0.09	-
36	1.19±0.10	1.1±0.06
38	1.30±0.10	1.21±0.03
43	14.53±0.12	14.27±0.03
45	0.67±0.09	-
47	2.39±0.11	2.30±0.03
49	22.41±0.11	21.85±0.04
51	0.66±0.09	0.30±0.04
52	13.79±0.11	14.23±0.04

Table 4.1: A comparison of the received fluxes in the lower window at 225GHz. The data from van Terwisga et al. (2022) is from their table 1.

From table 4.1 a couple of trends that are constant throughout the data can already be seen, for example that we get less flux on average and that our errors are smaller. We can also see that the faintest sources remain undetected with our flux extraction method. We only list the data from table 1 of van Terwisga et al. (2022) here, of course we compared all data points during the analysis process.
For the sake of comparison and better visualisation, we plot the extracted flux with our method against that of van Terwisga et al. (2022) as well as a 1:1 line throughout the figure. We also fit a linear slope to the resulting data, which turned out to be 1:1.08 over the whole range.

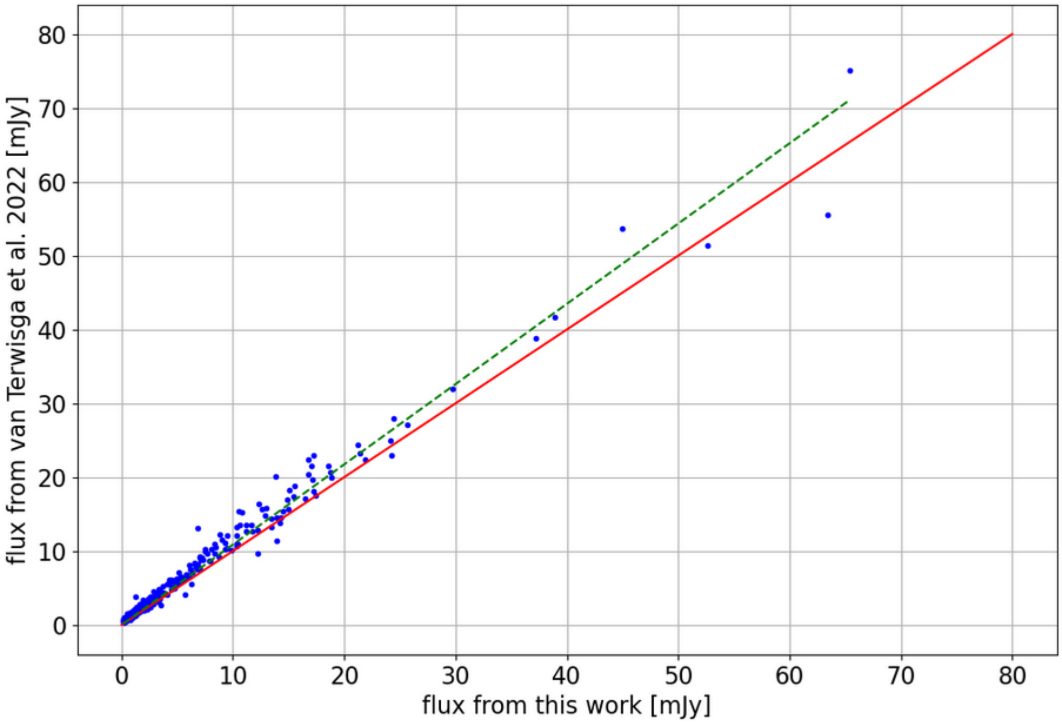

Figure 4.1: The flux comparison between van Terwisga et al. (2022) and this work. The blue dots are the individual fluxes and the red line is a 1:1 relation. The green dashed line is the actual best linear slope through the data at ≈1:1.08.

In figure 4.1 the y-axis are the received fluxes from van Terwisga et al. (2022) and the x-axis are the fluxes from this work. We can see that we get less flux on average, but consistently so. Over the whole sample, we get roughly 8% less flux for each source.
Nonetheless figure 4.1 shows clearly that our fluxes are consistent and allow for further analysis. Over the whole sample the lowest extracted flux is just 0.15mJy, with the highest flux of 65.43mJy. The median over the whole sample is 1.91mJy, which reflects what we can make out in figure 4.1, as there are by far more sources in the lower flux area between ≈ 0mJy and 7mJy.

4.1.2 Dust Mass

One of the big goals of van Terwisga et al. (2022) was to obtain the mass distribution of the whole sample. This can be directly gained from the fluxes, as is described in equation 2.8. For the sake of comparison and checking for viability, we did the same procedure. If we plot the two distributions over each other, we can check for differences also acting as a check of the methods from van Terwisga et al. (2022).
The median for the calculated masses is $12.3 M_{earth}$, the lowest mass is at $0.9 M_{earth}$ and the highest is at $407.5 M_{earth}$.

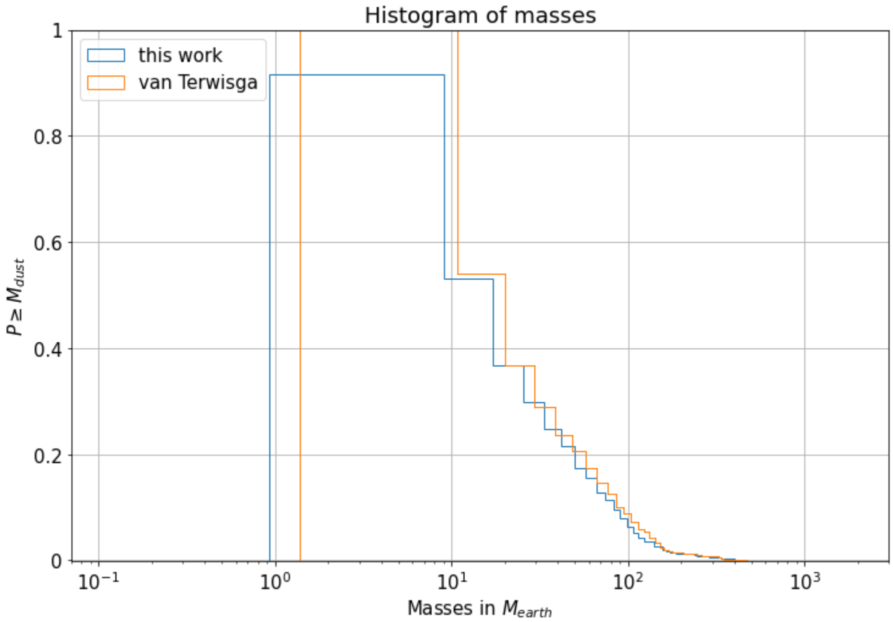

Figure 4.2: The distribution of disk masses. The orange line is van Terwisga et al. (2022) and the blue one is this thesis. The normalisation was done with our data, so the orange one can go above one as they have more objects in total.

Before talking about what we want to say with figure 4.2, the axis should be explained. On the x axis there is just the mass in units of the mass of the earth. The y axis is the probability of a mass being this massive or more massive. For example: the mass of 10 earths is at a P of ≈ 0.55, so around 55% of our masses are this massive or more massive.
Now let us explain what we see in this plot. The 'steps' visible in figure 4.2 are bins of various mass, and every new bin is added onto the last one to get a cumulative mass distribution.
This begs the question of why our graph does not reach 1, which is due to the bin size. The nature if the bins is such that the probability is correct for the central values of each bin. Thus the highest bin is not exactly at one but a bit below. Please note that van Terwisga et al. (2022) have more detections at lower masses, and are thus above 1 which is normalised for our detections.

4.2 Spectral indices

As laid out earlier near equation 2.7, the prediction from theory about β is about 1, where van Terwisga et al. (2022) states that some recent literature has set the value as low as 0.4 for disks, with Draine (2006) putting it at 0.9 but with an uncertainty of ± 0.7.
From this we can already see the large range of possible β values from literature only.
When analysing our βs, it became quickly apparent from the high scatter that they vary too much, as they have a range of ≈20 for the lowest fluxes, and still a scatter of ≈2 for moderate fluxes.

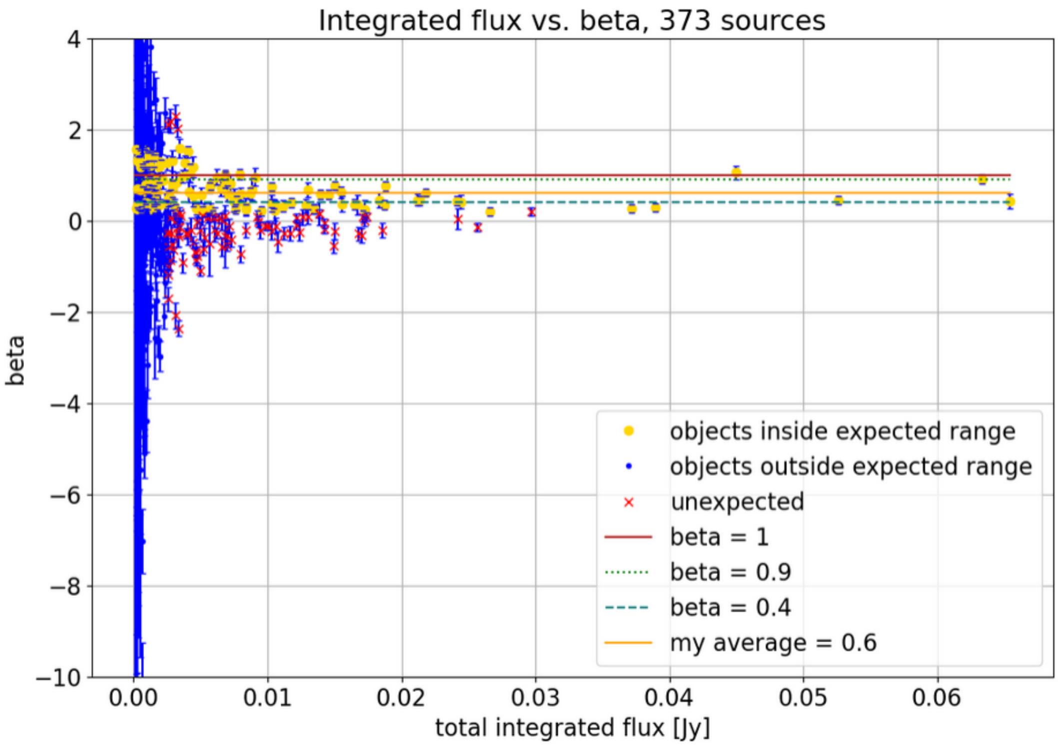

Figure 4.3: The β of each source plotted against the total flux in Jy. The multiple lines are β predictions from diverse sources in literature, as listed in the text above. The yellow dots on the plot are sources that lay within the somewhat expected range of Draine (2006) and the blue ones lie outside. The red crosses are outside of the expected range but with more flux than the blue dots.

Figure 4.3 shows the β value for each source against the respective flux. The various points are the different data points for each source. The blue dots are sources that have little flux and the scatter is too high to take them into account in any of the analyses. The yellow dots are sources that are inside the expected range of Draine (2006) of 0.9 ± 0.7. There are also sources that have at least moderate flux and are outside of the expected range, here signified as red x's. It should be noted that the distribution is roughly symmetrical around 0, hinting at this just being the natural scatter of such data.

The horizontal lines are the various beta values from literature, the dark red line is the conventional value for β of 1 as found in van Terwisga et al. (2022), the green dotted line is the one from Draine (2006) and the blue dashed line is the one from Pascucci et al. (2016). It becomes clear that none fit our distribution of β, apart from maybe the 0.4 mentioned in van Terwisga et al. (2022) from Pascucci et al. (2016). The average over all sources inside the expected range is here shown as a yellow horizontal line and stands at 0.6.

The analysis of the spectral indices presented in figure 4.3 results in no clear trend. It is inconclusive if the β parameter is accurate due to the scatter in the lower flux region on the left and due to the lack of a substantial number of bright source on the right.

For the sake of comparison, we calculate the average of β at a flux of > 5mJy. This amounts to a β of 0.11±0.49, putting this work below the established values, but within error of Draine (2006) and the 0.4 mentioned in van Terwisga et al. (2022). Only at the brightest end is the β from this thesis compatible with established literature, but given the lack of sources at this brightness, no statement can be made.
At the end we conclude that **we cannot make any definitive statements in regards to the spectral indices** of the sample.

4.3 Variability

Next section, we will touch on variability. The variability is here defined as the change in received flux with time. This is possible only since every object was observed at least twice, the majority even three times. So, one can look at the fluxes of one object at all times and see how much it changes.
In order to quantify variability, we take each individual flux from each time and look at this as a distribution. The standard deviation σ is then what we call the variability. We then take the σ from each source and plot the signal to noise (S/N) against that σ and fit the resulting distribution in figure 4.4.
From statistics, it would be reasonable to expect a $\frac{1}{a \cdot x}$ relationship, with a being some constant and x being the S/N. This is because we would expect the noise, here expressed as standard deviation, to stay roughly the same throughout all non-variable sources, and the signal to increase with flux. Thus the signal to noise ration S/N should behave like $\frac{1}{a \cdot x}$, with a factor a added for the effects of the normalisation.

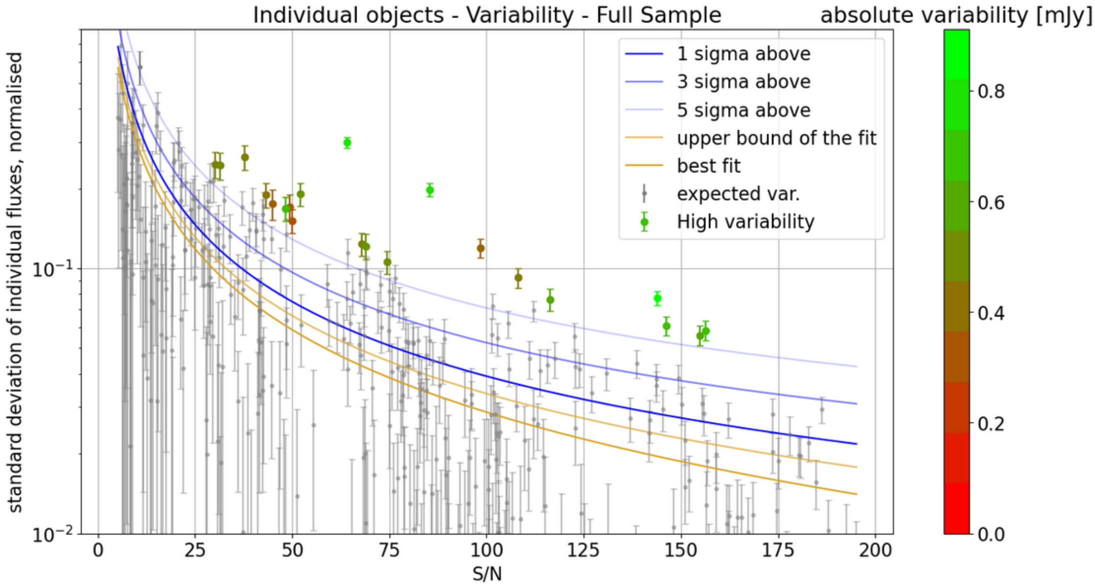

Figure 4.4: The variability as the standard deviation on the y-axis against the S/N of each source on the x-axis. The blue lines signify various σ away from the best fit. Sources that are above 5σ are considered variability candidates and are thus in colour. The colour in this plot is determined by the absolute variability, according to the colourbar on the right. The absolute variability is the difference between the highest and the lowest flux of each source.

We decided on this type of visualisation here as the relation between the standard deviation and the S/N is, as explained above, known. If some sources deviate significantly from this general trend, they may be variable sources. We are conservative on purpose with a 5σ threshold including the error to make sure that no non-variable sources would be above the threshold. This is in principle possible, as the data was not meant for such analyses.
In figure 4.4, the signal to noise ratio is on the x-axis and the normalised standard deviation is on the y-axis. The lowest of the two yellow lines is the best fit with the formula described earlier. The yellow line above is the upper bound of the fit in terms of uncertainty, including the error. This difference between the two fits is what we call 'sigma' within this plot. One sigma away from the upper bound is the first blue line. We do this for one, three and five sigma, where five sigma is the furthest we go away from the upper bound.
As we can see in figure 4.4, the amount of sources that can conservatively be called 'variable' is higher than expected from literature outlined for example in figure 2.11. To be specific, 21 sources are above this 5σ line. It should be noted that a source does not only need to be above the line, but it must be above including its lowest error. As you can see for example near the 100 S/N ratio, there is a source which has enough variability to be above 5σ, but not within its error. Over the whole sample, we get a maximum S/N ratio of 195.2, a minimum of 5.2 and a median of 75.0, which means that the S/N is spread rather equally.
It should also be noted that one of the sources that was initially above the line was taken out due

to it being a binary with the stars close to each other, confusing our algorithm. It had a jump in flux from one time to the next but with no visible increase on the images when we inspected it. From the results we think that the 21 sources are indeed variable, some by a factor of 2 over less than an hour (M12_661). We believe this to be likely linked to synchrotron radiation, as the other mechanisms do not offer a reasonable timescale for such an increase, but more on that in chapter 5.

Timescales

After creating figure 4.4, it would be interesting to know not only how the absolute variabilities behave, but also how the objects vary with time. It *may* be interesting to know if there are 'groups' of variabilities with time or if they all behave similarly.
Although this is of course true, the limiting factor for this kind of parameter is the number of observations. Even with three observations, it is not easy to make any definitive statement about the behaviour of a source, let alone with just two observations.
With this said, we calculated the variability over time with the highest change in mind. For this, our code takes each change of the individual timesteps and takes the one with the highest change in flux over time. In figure 4.5 the colourbar is exactly this highest change over time.

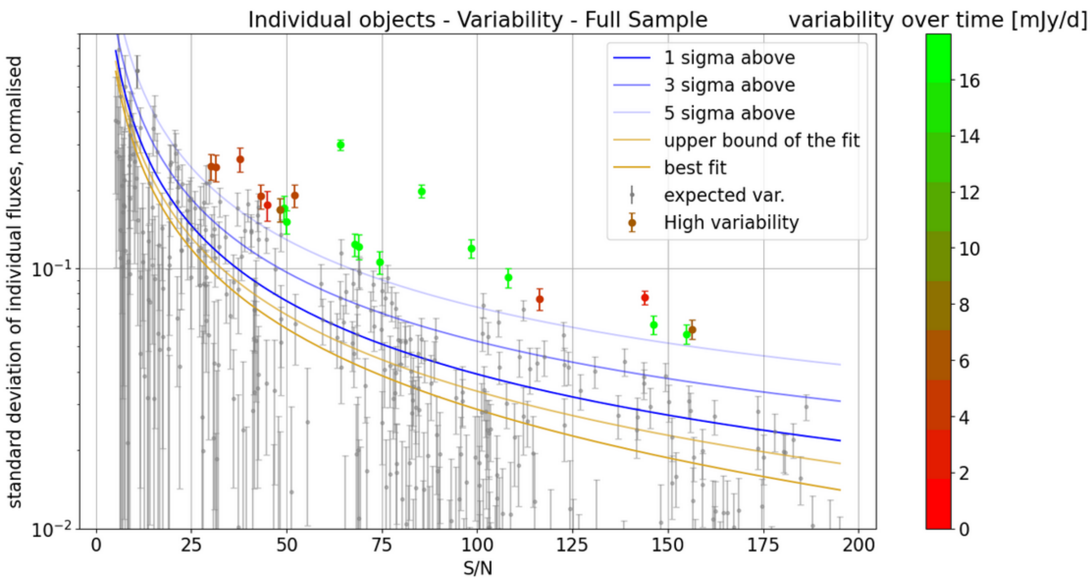

Figure 4.5: The variability against the S/N ratio again, this time with a different colourbar. The colours are referring to the change over time in mJy/day.

In figure 4.5 we can see the same plot as in figure 4.4 just with a different colourbar. This time, the colour indicates the variability rate, or the variability over time. It should be noted that there appears to be two groups, one with low variability over time in brown and red and one with a higher change rate in green. Importantly, this is likely to be a consequence of the structure of the data, where some sources were only observed within one day and some over multiple days. This fact alone may be sufficient to produce groups like this.

4.3.1 Variability Candidates

With the results from figures 4.4 and 4.5, there are 21 candidates up for inspection. Let us now take a look at some of these variable objects, both visually and with data.

M12_X	total flux [mJy]	σ variability [mJy]	variability over time [mJy/d]
648	0.65	0.465	4.8
62	0.81	0.495	4.6
208	0.66	0.460	4.2
584	0.89	0.420	4.3
357	0.71	0.347	2.4
463	0.85	0.287	6.3
448	0.99	0.339	7.4
294	0.80	0.302	6.6
36	1.10	0.528	4.9
661	1.11	0.819	17.6
164	1.43	0.473	8.7
120	1.50	0.531	8.1
656	1.98	0.516	11.2
330	1.48	0.784	9.6
395	1.54	0.370	8.1
187	1.78	0.455	8.9
98	2.81	0.601	3.6
194	4.42	0.911	1.7
404	5.68	0.690	15.0
340	3.46	0.603	7.4
11	4.06	0.689	6.3

Table 4.2: The data for the variables in figures 4.4 and 4.5 sorted after the ascending S/N ratio. The first column is the source number, the second column is the total flux, the third column is the absolute variability and the fourth column is the highest slope between two measurements for one source.

All sources above the 5σ line in figures 4.4 and 4.5 are given in data in table 4.2, including their absolute variability and the steepest slope for one source. From this it is clear that some of the sources show high variability over short times, particularly M12_661, M12_404 and M12_656, which show a slope of more than 10mJy/d. The slope is just the difference of the flux divided by the time in between the highest change.

We will show two of the more variable sources to give an impression of the sample. Of course we inspected all of them closely during the candidate analysis process.

M12_36

M12_36 for example shows an increase in integrated flux over about one day. The first increase occurs over just an hour by about 25%, with a later increase over the course of a day with another 35%. We chose M12_36 here as a pretty representative source for the variables that show a continuous change.
A total increase of around 60% over one day is in the range of what we would expect in those timescales from Fischer et al. (2023).

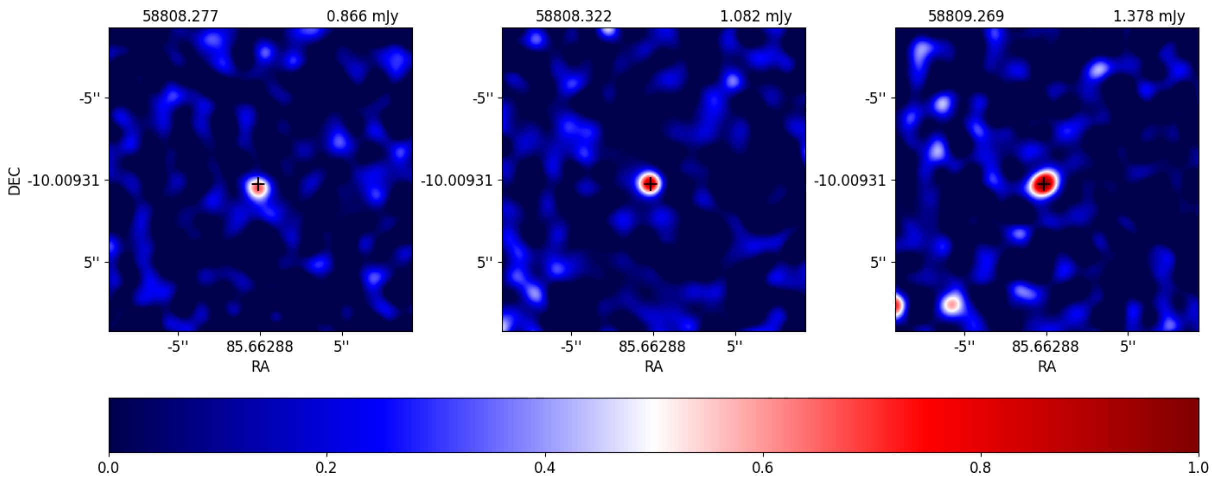

Figure 4.6: M12_36 as seen at all three times. The colourbar is normalised to the brightest pixel in all three observations. One can clearly see an increase in brightness from left to right with the background staying roughly the same.
At the top left of each panel is the observation time given as the mjd and at the top right there is the flux of each individual timestep.
The colour for the observations is normalised to the highest flux.

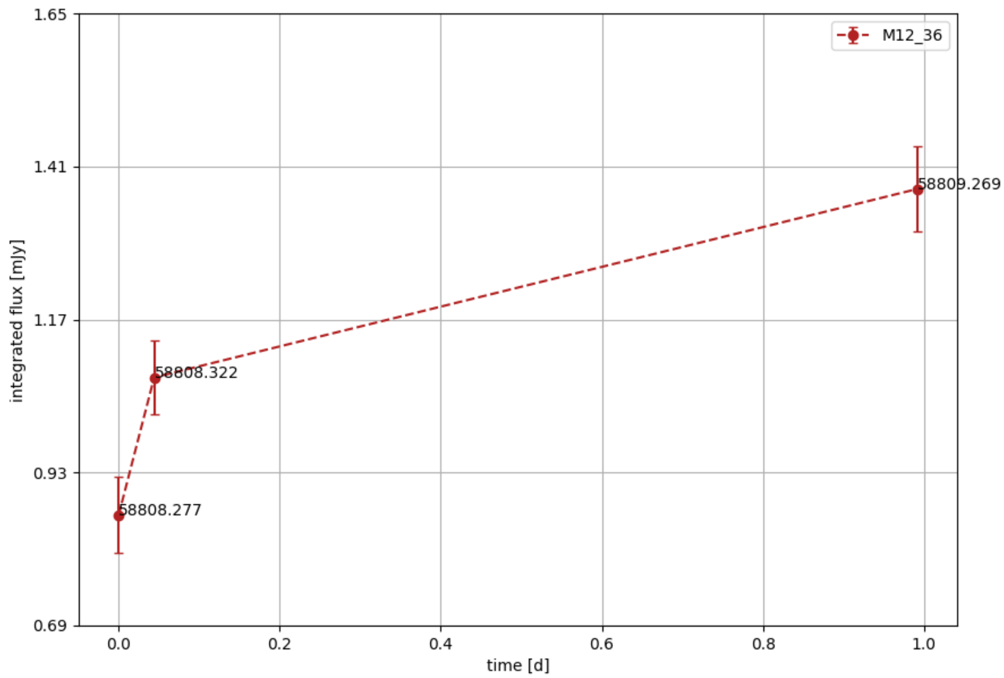

Figure 4.7: M12_36 in data. The sharp increase clearly reflects in actual integrated flux data. The error bars are 3σ.

We can clearly see from figure 4.6 that the intervals between observations are not equally spaced as already mentioned before. This is true for all sources and even the total time in between observations is not always about one day as it is here.

M12_661

M12_661 is one of the most variable objects in general, we would like to particularly point out the rapid increase in flux between the first and second measurement within only an hour, making M12_661 the source with the most rapid change.

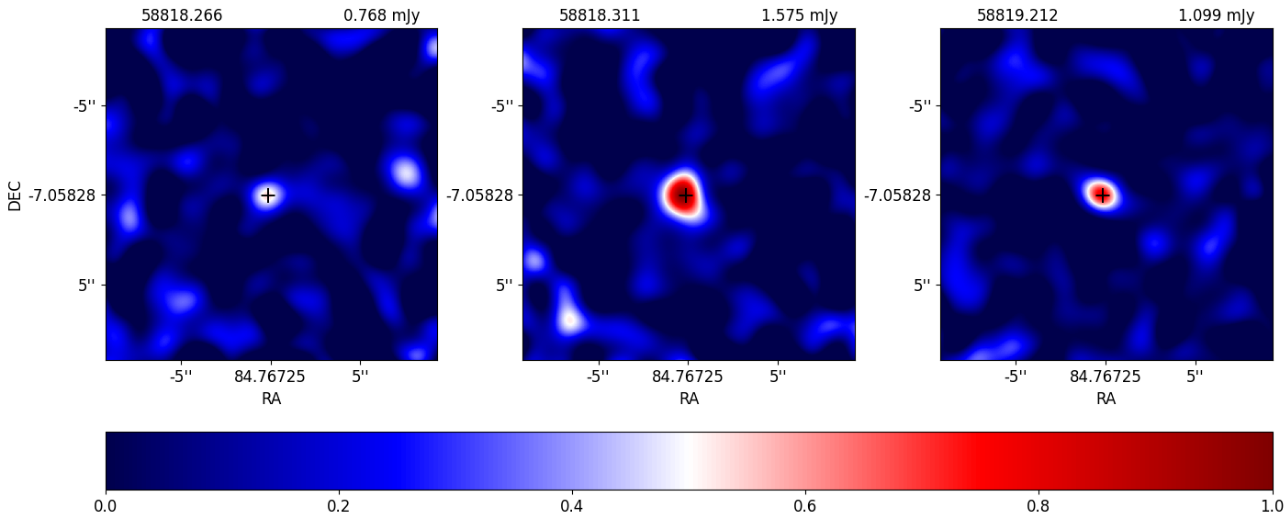

Figure 4.8: M12_661 as seen at all three times. The quick increase in flux from the first to the second picture is very clear.

As we can seen clearly in figure 4.8, the flux from the object increases sharply from the first to the second observation. Note that this change occurs at only an hour of time difference. M12_661 is thus also the source with the steepest increase with respect to time, 16 mJy/day to be exact.

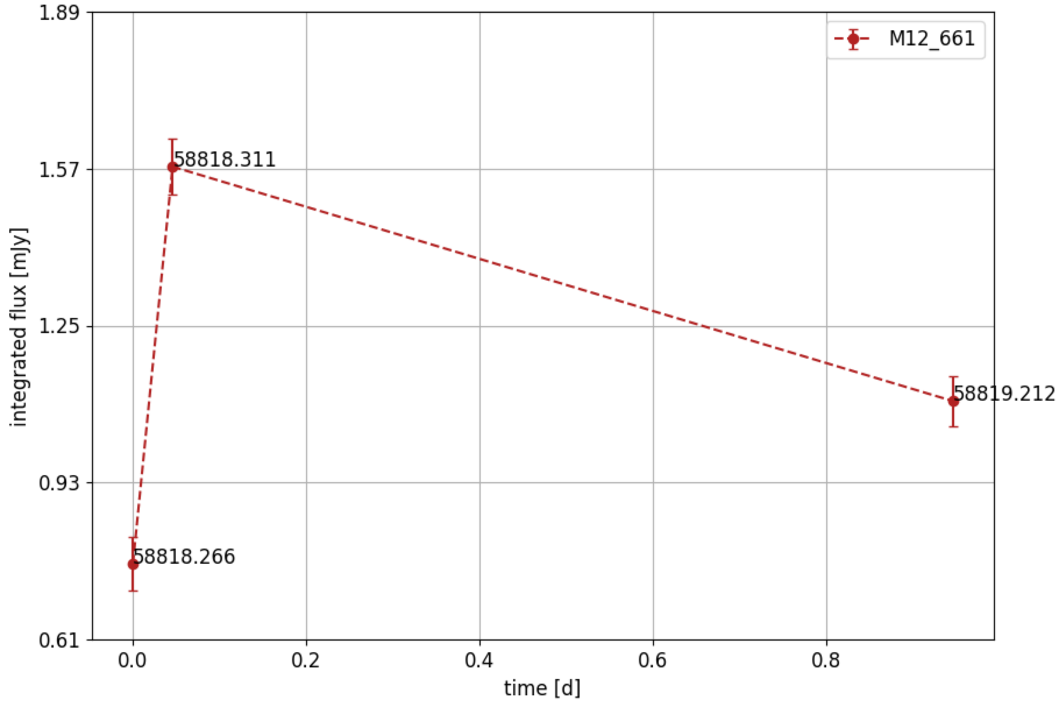

Figure 4.9: M12_661 in data. It is clear that the flux increases, in this case by a factor of ≈two in the first two measurements.

4.3.2 Comparison to other wavelengths

Here, we will compare our findings in regards to variability with other wavelengths. We do this to possibly identify variability of our candidate sources in other wavelengths and at other timescales. Note that this was the result of spontaneous talks with members of the IfA and its visitors and thus should not be seen as a big part of this work.

Gaia G-Band

Thanks to Ágnes Kóspál's and Péter Ábrahám's visit to the IfA in October of 2023 and a few talks with them, we could compare our variability to that of the G-band of Gaia. The G-band is between the UV at ~330nm and the Infrared at ~1050nm.
The comparison builds upon the magnitude measurements and its errors of the Gaia Data release 3. Gaia has observed so many objects that one can plot the error of the magnitudes against the magnitudes and get a very well defined curve, particularly for lower flux objects starting at roughly 13mag.
This defined curve expresses the inherent error of the measurement. This means that if a source is above this curve in the diagram, the source is likely variable of some sort. This is done with the 21 sources above the 5σ line from figures 4.4 and 4.5 and all the non variables on the sample as well. Of the 21 variable sources, 18 were found in the Gaia DR 3 and its magnitude, the error

thereof and the number of individual observations was extracted from the Gaia archive.
The error here is a measure for variability as the error rises with different measured fluxes.

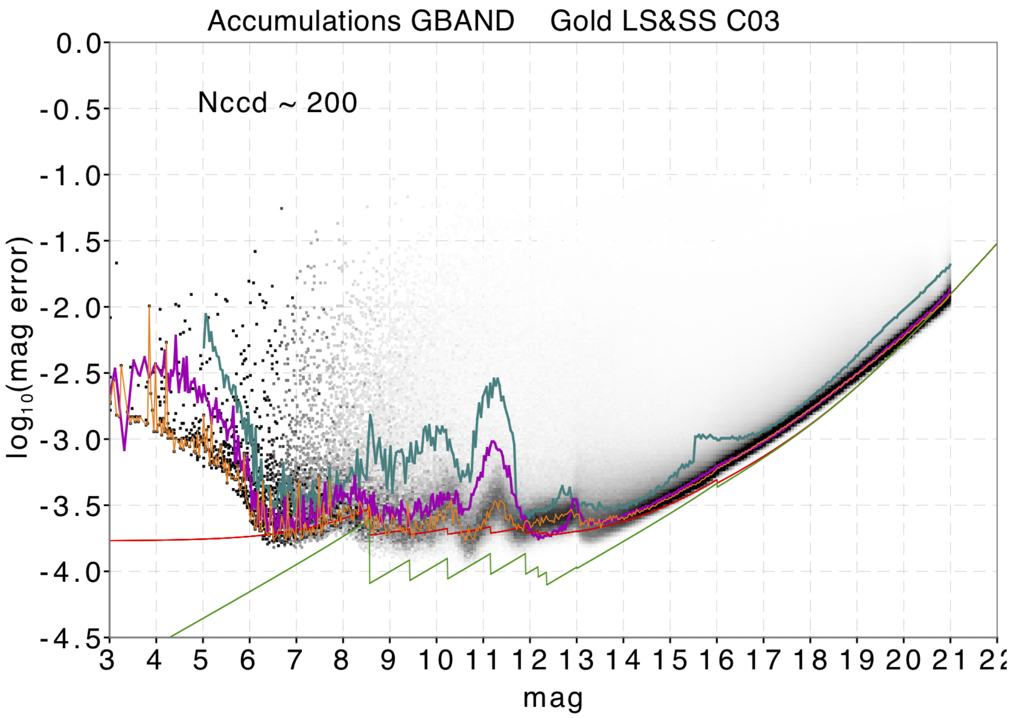

Figure 4.10: The data of all observations from all three Gaia releases. The x-axis is the magnitude, in this case of the G-band while the y-axis is the magnitude error in $\log_{10}$. The orange line is the one from DR3, being the one that is relevant in this context. It indicates the limit of what GAIA can reliably detect. Please note that there are no given uncertainties for the lines. From `https://gea.esac.esa.int/archive/documentation/GEDR3/Data_processing/chap_cu5pho/cu5pho_sec_photProc/cu5pho_ssec_photVal.html` (accessed 21.10.2023)

In figure 4.10 the x-axis is in magnitudes and the y-axis is the magnitude error. The magnitude error is the indicator for variability in our case since if a source is variable, its flux recorded by Gaia will have a higher error.
The curves in figure 4.10 are the lowest achievable errors for each magnitude and each data release (DR), where the orange line is for the current DR3. We can see data from all Gaia data releases, but for our purposes we are only interested in the orange curve from DR3. We will try to find our sources within the Gaia DR3 catalogue and plot them in a similar way in figure 4.11.

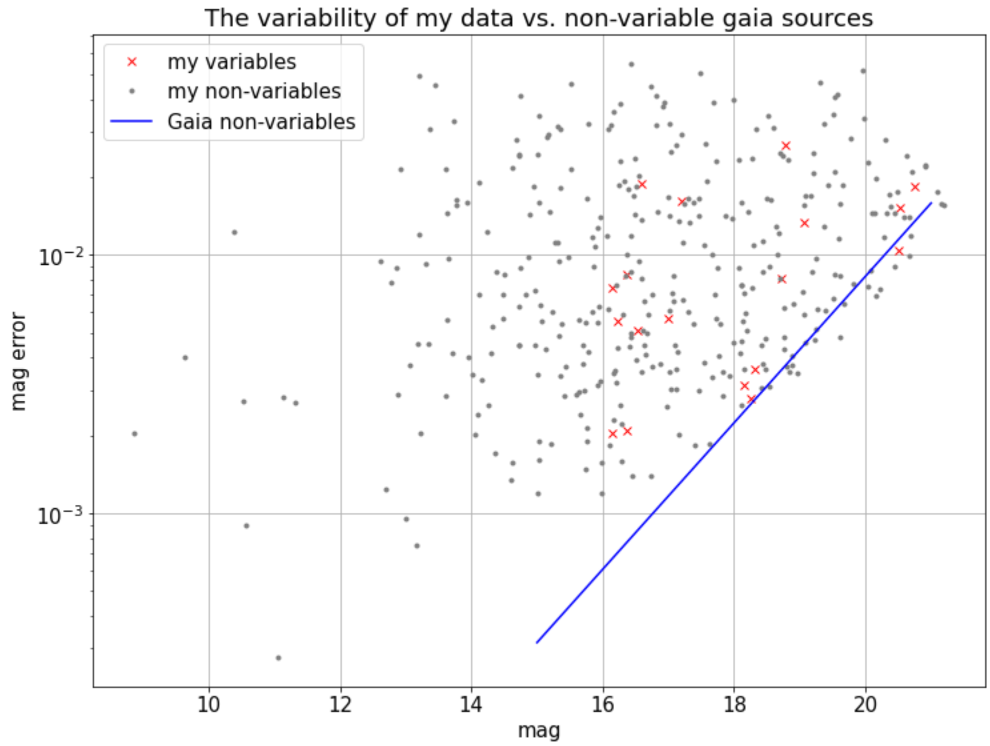

Figure 4.11: The comparison in the optical between our data and the rough Gaia curve for the whole DR3. The blue line was reconstructed after figure 4.10 thanks to a suggestion by Á. Kóspál.

Figure 4.11 shows the sources of the SODA sample from this work as observed by Gaia. The x-axis is again in magnitudes and the y-axis is the magnitude error in $\log_{10}$. This figure is supposed to be like figure 4.10. The red crosses are the variable sources from this work while the grey dots are the non-variables. The blue line is the lowest bound of what can reasonably be detected and in figure 4.10 is the orange line.
From figure 4.11 we can see that at least a majority of the sources deemed variable earlier is also showing variability in the G-Band of Gaia. Please note that the 'Gaia non-variables' line in blue in figure 4.11 was only reconstructed by us by reading off the values, but is enough for this purpose. Especially the sources near mag 16 and close to 10^{-2} in mag error are well above the curve, indicating variability.
But as we can also see, pretty much all sources are above the non-variable line, making the statements about variability meaningless.
As can be seen in figure 4.12, no clear variability can be made out for any object. It is safe to say that, at least in the infrared over those longer times, the variability from the mm does not correlate with the variability in infrared. When looking once again at figure 2.11, this is not totally unexpected, as the timescale of the observed data is a lot higher for both WISE and Gaia, meaning that the mechanism from which the variability originates is different, explaining the non-correlation of the variability.

WISE

Thanks to the explanation of Stefan Meingast, another reasonable comparison can be done with the data from the wide-field infrared survey explorer (WISE) and NEOWISE. Its observations were carried out in over a decade of time in W1 at 3.4μm and W2 at 4.6μm as described in more detail in Meisner et al. (2023). We compare our observations of the ones from WISE to check for possible correlations between the variability in millimetre and at the IR.
The observations from WISE/NEOWISE were carried out over multiple years about twice a year. Thus the data is per design of the survey investigating a different kind of variability. In general terms, the short kind of variability in the SODA survey is due to flares, while longer term variability can have many dozens of causes.
With this said, thanks to the unTimely catalog explorer[1] used in Meisner et al. (2023), the data for all objects that were detected with our algorithm was analysed. Since every object was observed about 20 times, the variability can be extracted from the data as the standard deviation, as was done with the ALMA data.
Plotting this σ against the average flux, any eventual outlier and thus variable object should be easily detectable.

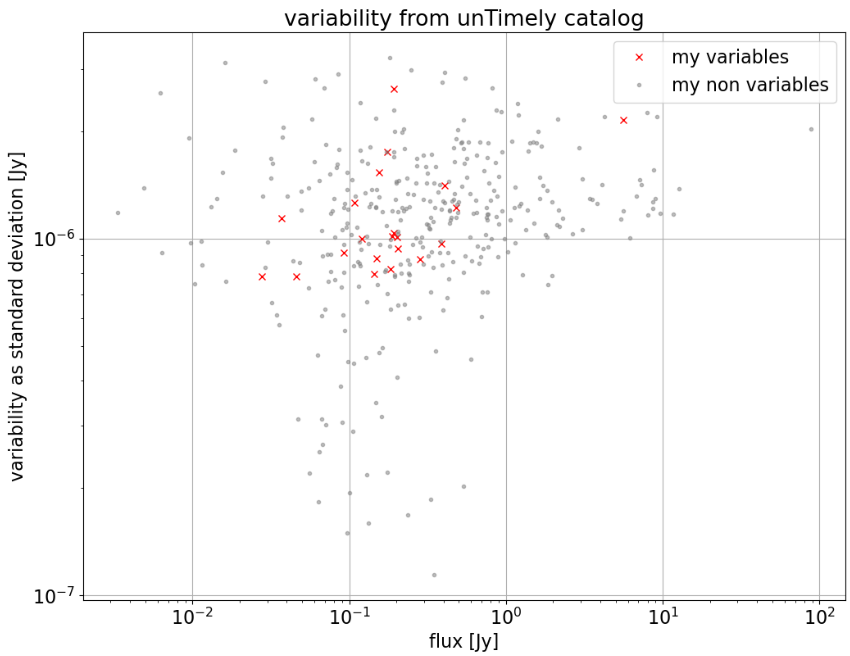

Figure 4.12: The variability of our sample against their respective flux from the unWISE sample.

In figure 4.12, the variability of the WISE data is plotted against the flux from each source. The

[1]https://github.com/fkiwy/unTimely_Catalog_explorer (accessed 24.10.2023)

red crosses are again the variable sources from the SODA sample, whereas the grey points are the non variables.
It should be noted that the axes were originally in nanomaggies (nMgy), but were transformed into Jansky with help of the SDSS III[2] documentation, listing a conversion factor of 1nMgy = $3.631 \cdot 10^{-6}$Jy.
As with the data from Gaia, our variables seem to be no different than our non-variables when observed with WISE.

4.3.3 Flux change as temperature change

The flux change amongst the separate sources *may* stem from a change in temperature. This is because, as elaborated earlier in chapter 2, the simplest explanation for an increase in flux is an increase in temperature. We also do this to see how a change in flux translates to a deviation from the established 20K dust temperature. If we attribute the variability solely to a rise in temperature, we get the following new temperatures:

M12_X	σ variability [mJy]	corresponding temperature [K]
648	0.465	34.3
62	0.495	32.2
208	0.460	33.9
584	0.420	29.5
357	0.347	29.8
463	0.287	26.7
448	0.339	26.8
294	0.302	27.6
36	0.528	29.6
661	0.819	34.8
164	0.473	26.6
120	0.531	27.1
656	0.516	25.2
330	0.784	30.6
395	0.370	24.8
187	0.455	25.1
98	0.601	24.3
194	0.911	24.1
404	0.690	22.4
340	0.603	23.5
11	0.689	23.4

Table 4.3: The variable objects from figures as previously given in table 4.2, but with the relative variability expressed as a change in temperature. This table is again sorted after the S/N.

[2] https://www.sdss3.org/dr8/algorithms/magnitudes.php(accessed 25.10.2023)

We calculated the temperature values from table 4.3 using equation 2.5. To be specific, if we take two masses with different flux, set them equal and get rid of the constants, we get

$$\frac{F_1}{T_1} = \frac{F_2}{T_2} \rightarrow \frac{F_1}{20} = \frac{F_1 + \triangle F}{T_2}$$

where we treat $\triangle F$ as the relative variability and T_2 is the 'new' temperature under the assumption that the flux change is only due to change in temperature. Rearranging for T_2 and simplifying further, we get the formula with which the temperatures from table 4.3 were calculated.

$$T_2 = \left(1 + \frac{\triangle F}{F_1}\right) \cdot 20 \tag{4.1}$$

Equation 4.1 is the rearranged formula for the 'new' temperatures.

4.4 Miscellaneous

4.4.1 M12_512 is an infalling streamer

The source M12_512 is one of the brighter sources in the sample with an average flux of ≈14.9mJy through both observed times. Very recently[3] there was a dedicated study carried out with M12_512 as a target, revealing that it actually is a YSO class I with an infalling streamer[4], detailed in Cacciapuoti et al. (2023).

They used the same SODA catalogue for part of there observations and also had a look at the spectral index, listing a $\alpha = \beta + 2 = 3.2$, meaning that $\beta = 1.2$ for M12_512. With our method, we get a β of -0.56. This only underlines our conclusion that the quality of the data within just one band is not enough to reliably draw conclusions about spectral indices. Cacciapuoti et al. (2023) also used Band 3 (B3) combined with more observations in B6 that are not from van Terwisga et al. (2022) to get their spectral index.

With this said out of pure curiosity, we decided to have a look at M12_512 for ourselves.

[3] in late November 2023

[4] A piece of nearby cloud that is accreted onto the disk but was not originally a part of it or a near flyby of a similar system. This may be interesting for planet formation as it introduces significant perturbations to the system that could increase planetesimal formation.

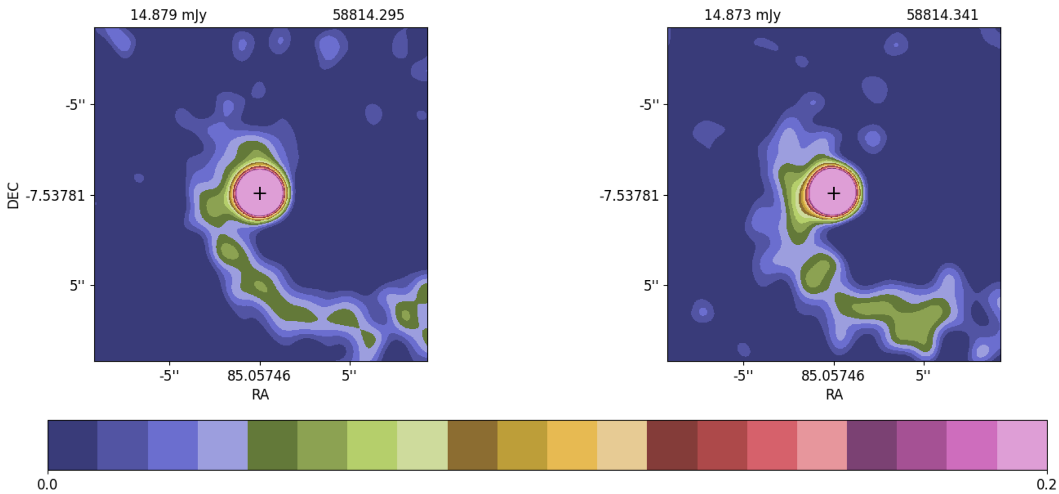

Figure 4.13: M12_512 in the SODA data. It should be noted that the extended structure was almost not visible without tweaking the overall colourscale to have a maximum of just 0.2 of the maximum flux. At the top left of the panel there is the flux and at the top right there is the time of observation in mjd.

As we can see in figure 4.13, there is indeed an extended structure visible when tweaking the scaling in our dataset.
The findings surrounding M12_512 and the sizeable disagreement between Cacciapuoti et al. (2023) and our values (β = 1.2 and -0.56, respectively) once again show the importance of having a large frequency range before making statements about spectral indices.

5 Discussion

5.1 Flux and Masses

The flux and thus the calculated masses were, as mentioned previously, one of the big goals in van Terwisga et al. (2022). This is because the mass of the dust in an accretion disk can be used in simulations, can give us insight into later planet formation and can teach us lessons about the mass distribution, as we can see in figure 4.2 for our work and in figures 2.8 and 2.7 of various other clouds. It became clear early on that this would also be one of the main goals in this work, not only because van Terwisga et al. (2022) analysed it as well and we could thus check our methods, but also because it was one of the more easy things to achieve in this work and still gives us quite relevant information on the sources.
When comparing the fluxes between us and van Terwisga et al. (2022) in figure 4.1, it becomes clear that there is a systematic difference. We mentioned above that the overall difference on average is 8%, but since the trend seems to be continuous and well expressed by the linear correlation, we believe that our results are fit for any later analyses.
We would still like to list some of the key differences and maybe shortcomings of our approach:

- Less objects made the cut above the detection threshold in our work then in van Terwisga et al. (2022). We already talked about this in section 3.5 and is thus no surprise. Our flux extraction methods were different from van Terwisga et al. (2022), so a difference is expected.

- Our measured fluxes are on average lower, as laid out in figure 4.1. This is also due to a different reduction mechanism as a whole.
  In their paper, van Terwisga et al. (2022) use the peak value and assume that it is the whole flux as this is the case for a non resolved gaussian (which is a reasonable thing to do, as the vast majority of the sources are unresolved). However in this work, we decided against it and used the method described in section 3.5 instead. We did this mostly because the method in van Terwisga et al. (2022) is more prone to noise as it only depends on one pixel and may miss some partially resolved sources or it may detect sources that are as a whole too faint, especially for low flux sources.

- We have smaller errors in regards to the flux. The data in table 4.1 shows smaller errors than the ones from van Terwisga et al. (2022) in their respective table. First of all, they list 2σ, while we give our data with 3σ. In their paper, some sort of noise map is used according to Sierk van Terwisga when asked about the error, this also follows from one of the code snippets we received from him, where a noise map is referenced. We do not know how van Terwisga et al. (2022) came up with the map, but it is safe to say that it results in higher errors.

- Importantly, the fluxes of multiple sources are, within reasonable error, the same; enabling comparisons and further analysis as stated in chapter 1.

Note that the general shape of our mass distribution is very similar to van Terwisga et al. (2022), clearly mimicking the slope and the extent at roughly $200 M_{earth}$. The whole distribution fits well into the general picture, compared with figure 2.7 from Miotello et al. (2022). In a similar figure from van Terwisga et al. (2022) we can see how this fits in with various population ages, where lower age populations have a higher disk mass, just as would be expected.

5.2 Variability

The variability over the whole sample shows a number of partially surprising results.

1. There are 21 variable source over 373 detected sources, translating to roughly 5.6% of sources that can be seen for example in table 4.2. Please note that this is with a conservative 5σ threshold. Compared to literature, particularly the previously mentioned Mairs et al. (2019) and Vargas-González et al. (2021) have vastly lower variability, with <0.1% and <1%, respectively. The variability here only lines up somewhat with the Chandra observations from Forbrich et al. (2017), who have 13 out of 222 sources (≈5.8%) variable. The data is very unlikely to have a systematic undetected error in it, so we assume that the variability is indeed how we show it.

2. No clear bias can be made out amongst the variable candidates. The distribution of variable candidates is equal[1] amongst the whole S/N range, the lack of variables in the lowest S/N area is explainable by 5σ being unrealistically high at low fluxes as well as the high relative error for low flux sources.

3. The magnitude of the variability is higher than the $< 1\%$ we expected. From literature, i.e. Fischer et al. (2023), we know that over a few hours, the expected variability is definitely below a factor of two and more of the scale of 20-40%, contradicting some of the variability that was extracted in this work, especially for M12_661 and M12_330 that have a change in flux of a factor two. Note that this is not a strong contradiction, as over a day the flux can vary by that much.
   With that said, some flares documented in i.e. Vargas-González et al. (2023) or Forbrich et al. (2017) also are a factor 10 higher than the typical ranges stated in Fischer et al. (2023), as are the more theoretical considerations by Benz & Güdel (2010) who mention possible variability of more than a factor 2 in those short times. We thus conclude that the variability is unlikely to stem from accretion variability and is due to gyrosynchrotron radiation, as it is in Vargas-González et al. (2023). In figure 2.11, we can see that our data at 230GHz

[1] or at least equal enough given the number of variable objects

with a variability timescale of usually a few hours to an hour does not really fit into the given areas of variability.

The 21 variable sources are very unlikely to have been random. From pure noise, the chance of one source having a variability of the 5σ explained in chapter 4 in figure 4.4 is just $\approx 5.7 \cdot 10^{-5}\%$. For our roughly 1000 sources, the chance of one source being that variable by accident is thus $\approx 5.7 \cdot 10^{-2}\%$, so very low.

In light of the information in figure 2.11, its is most likely that the variabilities stem from either pulsations or flares, or maybe magnetic accretion changes, while some of the shorter variabilities likely originate from synchrotron radiation.
Now with this said, there are studies of YSOs at lower frequencies ($\approx$ 10s of GHz with i.e. VLA, VLBA) and some with x-ray observations that do in fact note flares at that magnitude and even higher at timescales of minutes to a few hours, as briefly mentioned in the introduction. See Forbrich et al. (2017) who find radio flares in Orion and link them to corresponding x-ray variability, or Vargas-González et al. (2023) who report on a highly variable source in the ONC. So far, no such observations have yet been made in the mm-range at around 200 GHz to the best of our knowledge.
It should be noted that there appears to be a bimodality in figure 4.5 in regards to the variability over time, with around half of the candidates exhibiting high variability over time, and the rest low variability over time. This is likely not due to any inherent property of the sources, i.e. some ongoing mechanism for variability that is either slow or fast. Just from our data, we have sources that are observed over multiple days and some that have only been observed in one day. Naturally, the sources that were observed on only a day have higher variability rate than the ones that were observed over multiple days by pure chance.
As stated before, the observations were conducted at different times. This was not done in a systematic way however, so the difference of time in between each individual observation varies greatly. Some objects were observed three times in slightly more than a day and some were observed three times in over five days.
This alone is enough to create the dichotomy from figure 4.5, but a more detailed observation would surely be interesting.

5.2.1 Variability in Literature

From Vargas-González et al. (2021) we know that some radio-sources in and near the ONC exhibit high variability by as much as a factor of 10 in flux within only minutes in some cases. It must be noted that their sensitivity is a lot better, as they measure as little as 3μJy, and the observations were a lot more focused in short times of as little as a minute and carried out with the very large array (VLA) with as many as 41 observations per source. To be specific, some of their peak flux densities are as low as 0.15 mJy, which happens to be the flux of the lowest detection in this work. Safe to say that the vast majority of sources from them would not be detected in our data.

Vargas-González et al. (2023) makes similar observations using ALMA, but having around 1.15 hours on each target, they achieve sensitivity in the 300μJy range and lower, which would only be amongst the faintest sources in our sample. There they report 133 variable sources with periods reaching from mere minutes to days, more or less in agreement with the findings in this work. They also find sources that change by an order of magnitude within as little as four minutes, which would also have gone unnoticed in this work since the shortest time in between observations was just shy of an hour.

Somewhat in contrast to this, Johnstone et al. (2022) have analysed over 5 years worth of YSO observations at 450μm and 850μm done with the James Clerk Maxwell Telescope (JCMT) in the JCMT Transient Survey. The total number of YSOs observed was about 1200 and they only observed one burst event, with some other more focused observations of 100 YSOs in Ophiuchus yielding not one such event over 70 days worth of observations, usually in intervals of days-weeks.
They also point out that the one flare "remains a unique and rare event, perhaps related to binarity". The typical effects of binary are, as explained earlier, not relevant for our observations as they usually occur on longer timescales.

5.2.2 Some speculation on future plans

After talking to Jan Forbrich[2] he suggested that there are some other observations also in other wavelengths that show similar variability rates, as outlined in some of the cited papers (particularly Vargas-González et al. (2023) and Forbrich et al. (2017)). This is again unexpected when compared to the usual range of $< 1\%$, as for example seen in figure 2.11.
It would be very advantageous to check the timescales and their respective correlations with amplitudes in a systematic way, which has not been done yet.
However, I plan on doing an analysis of exactly that after the completion of this thesis with the help of J. Forbrich.

[2] co-author of Vargas-González et al. (2021), Vargas-González et al. (2023) and Johnstone et al. (2022), author of Forbrich et al. (2017), as well as investigating variability in shorter timescales in other wavelengths, talked to in late November 2023

6 Conclusion

The main goals of the thesis as laid out in the beginning have all been assessed:

1. The extraction of the fluxes for each source and the calculation of the masses from them, as well as analysing the mass distribution.

2. The analysis of spectral indices for each source.

3. The investigation of variabilities in the sample.

Working to achieve these goals, we present the results:

- We looked into 873, non-resolved YSO Class II disk sources from the SODA sample, the ALMA observations of the Megeath catalogue from Megeath et al. (2012) and first described in van Terwisga et al. (2022)

- We successfully reproduced the mass distribution achieved by van Terwisga et al. (2022), with different methods and therefore within reasonable differences, showing that our flux extraction algorithm works reliably.

- We carries out a new data reduction and utilising the Vienna Scientific Cluster (VSC) for the large computational load and the parallelisation capabilities during the extraction of the fluxes of the ALMA data. CASA (common astronomy software applications) was used to reduce the data from the SODA sample. To this end, custom code was developed with python/casa as well as bash.

- We tried to analyse the spectral indices by extracting the flux from two separate frequency windows at 225GHz and 240GHz respectively. This was not yet done by any other papers that use the SODA sample. However, the results suggest a very unreliable spectral index throughout the sample with large statistical scattering. We thus came to the conclusion that checking spectral indices with just one single band is not reliable.

- We analysed the yet unchecked variabilities of every source by looking at the various fluxes of the multiple observations per source, and found an excess of 21 sources (roughly 5.6% of all detections) compared to the expected amount from literature of <0.1% in Johnstone et al. (2022) and <1% in Vargas-González et al. (2021)[1]. Some of the sources show variability of a factor > 2 at short timescales of only an hour.

[1]It does however fit more or less into the finding of ≈5.8% in Forbrich et al. (2017) in the X-ray

- The measured variability, expressed as the standard deviation of the distribution of the individual fluxes is higher than expected from typical literature and does not seem to appear at other time scales in other wavelengths from our comparisons with optical and infrared data. This result is significant in that no such variability was seen or mentioned in large surveys with ALMA previously.
- The timescales of the variability is sometimes too short to be a change in temperature or accretion. We believe that the change is due to (gyro-)synchrotron emission instead, agreeing with previous studies of flares in YSOs like in Vargas-González et al. (2023) and elaborated in Benz & Güdel (2010).
- **Whitin the SODA sample, we find that disk fluxes vary at a rate of 5.6% most likely due to synchrotron radiation in timescales of 60 minutes to 5 days, in some cases by as much as a factor of two.**

The most remarkable of the points listed above are the high occurrence of variable sources as well as some of their respective variabilities, which warrants further studies on the subject, possibly also in other surveys and maybe even follow-up observations to make sure that this does indeed originate from the sources. Further, a deeper look at the topic of the recorded variabilities over time is necessary and will be carried out outside of the scope of this thesis.

7 Acknowledgements

The computational results presented have been achieved in part using the Vienna Scientific Cluster (VSC).
I thank Sierk van Terwisga for the explanation of several topics as well as his readiness to provide parts of his processing and code snippets.
I thank Ágnes Kóspál and Péter Ábrahám for giving us the idea and providing explanations for the comparison with Gaia data.
I thank Stefan Meingast for elaborations on the NEOWISE/unTimely tools and data.
I thank the whole Emerge team at the IfA for continued opinions and suggestions on the topic and on my visualisations.
I thank all the friends and family that have helped to proof-read this thesis.
This work has made use of data from the European Space Agency (ESA) mission Gaia (`https://www.cosmos.esa.int/gaia`), processed by the Gaia Data Processing and Analysis Consortium (DPAC, `https://www.cosmos.esa.int/web/gaia/dpac/consortium`). Funding for the DPAC has been provided by national institutions, in particular the institutions participating in the Gaia Multilateral Agreement.
I thank Prof. Hacar for offering to supervise my thesis and doing so in a supporting and engaging manner. Without his suggestions and help, this work would not exist.

Bibliography

ALMA Partnership, Brogan, C. L., Pérez, L. M., et al. 2015, , 808, L3

Ambartsumian, V. A. 1947, Communications of BAO, 69, translation from 2022 of the original paper in russian

Andre, P., Ward-Thompson, D., & Barsony, M. 1993, , 406, 122

Astropy Collaboration, Price-Whelan, A. M., Lim, P. L., et al. 2022, , 935, 167

Bai, X.-N. 2011, The Astrophysical Journal,, 739, 50

Benz, A. O. & Güdel, M. 2010, , 48, 241

Benz, W., Ida, S., Alibert, Y., Lin, D., & Mordasini, C. 2014, in Protostars and Planets VI, ed. H. Beuther, R. S. Klessen, C. P. Dullemond, & T. Henning, 691–713

Bhardwaj, A., Panwar, N., Herczeg, G. J., Chen, W. P., & Singh, H. P. 2019, AA, 627, A135

Biazzo, K., Alcalá, J. M., Frasca, A., et al. 2014, Astronomy amp; Astrophysics, 572, A84

Bouvier, J., Alencar, S. H. P., Boutelier, T., et al. 2007, AA, 463, 1017

Cacciapuoti, L., Macias, E., Gupta, A., et al. 2023 [[arXiv]2311.13723]

Chen, C. H., Mittal, T., Kuchner, M., et al. 2014, , 211, 25

Cortes, P., Vlahakis, C., Hales, A., et al. 2023, ALMA technical handbook 10 (ESO)

Dauphas, N. & Chaussidon, M. 2011, Annual Review of Earth and Planetary Sciences, 39, 351

Draine, B. T. 2006, The Astrophysical Journal, 636, 1114

Dullemond, C. & Monnier, J. 2010, Annual Review of Astronomy and Astrophysics, 48, 205

Evans, Neal J., I., Dunham, M. M., Jørgensen, J. K., et al. 2009, , 181, 321

Eyer, L. & Mowlavi, N. 2008, Journal of Physics: Conference Series, 118, 012010

Fischer, W. J., Hillenbrand, L. A., Herczeg, G. J., et al. 2023, in Astronomical Society of the Pacific Conference Series, Vol. 534, Protostars and Planets VII, ed. S. Inutsuka, Y. Aikawa, T. Muto, K. Tomida, & M. Tamura, 355

Forbrich, J., Reid, M. J., Menten, K. M., et al. 2017, The Astrophysical Journal, 844, 109

Gandolfi, D., Alcalá, J. M., Leccia, S., et al. 2008, The Astrophysical Journal, 687, 1303

Greene, T. P., Wilking, B. A., Andre, P., Young, E. T., & Lada, C. J. 1994, , 434, 614

Gully-Santiago, M. A., Herczeg, G. J., Czekala, I., et al. 2017, The Astrophysical Journal, 836, 200

Hartmann, L., Herczeg, G., & Calvet, N. 2016, , 54, 135

Herczeg, G. J., Johnstone, D., Mairs, S., et al. 2017, The Astrophysical Journal, 849, 43

Högbom, J. A. 1974, , 15, 417

Hughes, A. M., Duchêne, G., & Matthews, B. C. 2018, Annual Review of Astronomy and Astrophysics, 56, 541–591

Johnstone, D., Lalchand, B., Mairs, S., et al. 2022, The Astrophysical Journal, 937, 6

Kenyon, S. J. & Hartmann, L. 1995, , 101, 117

Lada, C. J. 1987, Star formation: from OB associations to protostars.

Lombardi, M., Alves, J., & Lada, C. J. 2006, A&A, 454, 781

Mairs, S., Lalchand, B., Bower, G. C., et al. 2019, The Astrophysical Journal, 871, 72

Mairs, S., Lee, S., Johnstone, D., et al. 2024, The JCMT Transient Survey: Six-Year Summary of 450/850 μm Protostellar Variability and Calibration Pipeline Version 2.0

Manara, C. F., Ansdell, M., Rosotti, G. P., et al. 2023 [[arXiv]2203.09930]

Megeath, S. T., Gutermuth, R., Muzerolle, J., et al. 2012, , 144, 192

Meisner, A. M., Caselden, D., Schlafly, E. F., & Kiwy, F. 2023, The Astronomical Journal, 165, 36

Miotello, A., Kamp, I., Birnstiel, T., Cleeves, L. I., & Kataoka, A. 2022 [[arXiv]2203.09818]

Namekata, K., Maehara, H., Notsu, Y., et al. 2019, The Astrophysical Journal, 871, 187

Natta, A. & Whitney, B. A. 2000, , 364, 633

Park, W., Lee, J.-E., Peña, C. C., et al. 2021, The Astrophysical Journal, 920, 132

Pascucci, I., Testi, L., Herczeg, G. J., et al. 2016, , 831, 125

Remijan, A., Biggs, A., Cortes, P., et al. 2021, ALMA Cycle 7 Technical Handbook

Shulman, S. & Grinin, V. 2022, Open Astronomy, 31, 67

Slee, O., Nelson, G., Stewart, R., et al. 1987, Monthly Notices of the Royal Astronomical Society, 227, 467

The CASA Team, e. a. 2022, CASA, the Common Astronomy Software Applications for Radio Astronomy

Thompson, A. R., Moran, J. M., & Swenson, George W., J. 2017, Interferometry and Synthesis in Radio Astronomy, 3rd Edition

Tobin, J. J., Sheehan, P. D., Megeath, S. T., et al. 2020, , 890, 130

van Terwisga, S. E., Hacar, A., van Dishoeck, E. F., Oonk, R., & Zwart, S. P. 2022, Astronomy & Astrophysics, 661, A53

Vargas-González, J., Forbrich, J., Dzib, S. A., & Bally, J. 2021, Monthly Notices of the Royal Astronomical Society, 506, 3169

Vargas-González, J., Forbrich, J., Rivilla, V. M., et al. 2023, Monthly Notices of the Royal Astronomical Society, 522, 56

Werner, M. & Eisenhardt, P. 2019, More Things in the Heavens: How Infrared Astronomy Is Expanding Our View of the Universe (Princeton University Press)

White, R. J. & Hillenbrand, L. A. 2004, The Astrophysical Journal, 616, 998

Williams, J. P. 2021, Introduction to the Interstellar Medium (Cambridge University Press)