



Using Structure Functions to Investigate the Long-Timescale Variability of Young Stellar Objects

Tom Joshi-Cale
tj294@ex.ac.uk
670000715

MPhys Dissertation

Supervisor: Prof. Tim Naylor
Project Group: Ryohei Matsumoto, Chris Reid, Mike Spencer

Abstract

Optical observations of 67 Class I and 396 Class II YSOs are presented, alongside infrared observations of 226 Class I and 272 Class II YSOs, and an investigation is performed into the long-timescale variability of these objects and how it relates to class and wavelength. Structure functions are used to quantify the variability of young stars observed in optical wavelengths by the Zwicky Transient Facility and in infrared by the Vista Variables in the Via Lactea survey. It is found that Class I objects are between 1.5-2.5 times more variable than Class II objects over all wavelengths, but also that, counter-intuitively, Class II objects are twice as variable in infrared than they are in optical. It is suggested that the increased level of variability of Class I objects implies that they have a more variable accretion rate than older Class II objects, and that the height of the accretion disc around Class II objects may vary with time.

CONTENTS

1	Introduction	4
2	Background	5
2.1	Young Stellar Objects	5
2.2	Variability in YSOs	6
2.3	Structure Functions	7
2.4	Prior Work	8
3	Data Collection	10
4	Method	13
4.1	Validating the Method Using Sergison et al.	13
4.2	Adapting the Method for Collected Data	15
4.3	Applying the Method	17
4.4	Cumulative Distribution Graphs	20
5	Final Results	22
6	Conclusions	23
7	Future Work	24
8	Acknowledgements	24
	Appendices	29
A	VVV data SQL query	29
B	Median SF Flow Chart	30

LIST OF FIGURES

1	YSO Class Diagram	5
2	Sergison Median Structure Functions	8
3	Non-Variable Star Criteria	11
4	Validation of FORTRAN SF code	13
5	Validation of Median SF Code	14
6	Marton Median SF Comparison	16
7	Broos Sample Class I Median SF	18
8	ZTF Class II Magnitude Bin Medians	19
9	ZTF Class I and II cumulative frequency graphs	20
10	ZTF and VVV Variability Amplitude Comparison	21
11	All Wavebands Median Structure Function Comparison	21

1 INTRODUCTION

Star formation and the very early stages of stellar evolution are dynamic processes and the newly-forming stars undergo many rapid structural changes early in life, leading to observations of magnitude variability in young stars. Many different types of variability have been observed, and are thought to have a wide array of causes. A large amount of astrophysical objects, for example Cepheid variables (Freedman et al. 1994), RR Lyrae variables (Lafler & Kinman 1965) and δ Scuti variables (Breger 1979), are observed to exhibit variability with a well-defined periodicity and while there have been many investigations into periodicity of the variability of young stars (Littlefair et al. 2010; Cody et al. 2014; Audard et al. 2014), only roughly one-third of Young Stellar Objects (YSOs) are observed to exhibit periodic behaviour (Herbst et al. 2002; Littlefair et al. 2010). The remaining objects appear to exhibit aperiodic variability, which has multiple proposed causes, including obscuration by circumstellar material (Herbst et al. 1994), instabilities within the circumstellar disc which surrounds the star (Alencar et al. 2010) and unsteady accretion rates of material falling onto the star (Scholz et al. 2009; Venuti et al. 2014).

The variability of YSOs occurs on a wide range of timescales; Sacco et al. (2008) suggests that some variability can occur on the order of a few minutes and there have been many studies into variability on timescales on the order of a few days to a few years (Herbst et al. 1994; Rice et al. 2012; Wolk et al. 2013). So called ‘FUor events’ (Reipurth 1990; Contreras Peña et al. 2019) can even result in magnitude increases on the timescales of 10^4 to 10^5 years. Since long-timescale variability is likely to be linked to the accretion rate (Vorobyov & Basu 2005; D’Angelo & Spruit 2012; Scholz et al. 2013; Contreras Peña et al. 2017) characterising it is useful to build up an understanding of the flow of material through, and the structure of, the circumstellar disc. Since planets are known to form before a star enters its main sequence (Morbideilli & Raymond 2016) this helps to create a more thorough understanding of the circumstances in which planets form.

By characterising the variability of Class I and Class II YSOs, this report aims to gather insight into how accretion rate changes as YSOs evolve through analysis of the variability of objects observed across the whole sky, and to build up a general picture of the variability of each class. It will attempt to build upon the findings presented in Sergison et al. (2020), as laid out in Section 2.4. YSO observations will be presented in two wavebands, to allow comparison between variability in the optical and in the infrared and further investigate the underlying causes of long-timescale variability.

2 BACKGROUND

2.1 YOUNG STELLAR OBJECTS

‘Young Stellar Object’ (YSO) is an umbrella term for any stars that are yet to reach the ‘Main Sequence’ stage of their life. YSOs evolve through multiple classes as they age, and these classes can be identified by the object’s ‘Spectral Energy Distribution’ (SED) which is used to describe the structure and composition of the object, as shown in Figure 1. Class I objects consist of a protostar embedded in an envelope of gas and dust, and surrounded by a circumstellar disc of gas, dust and rock. Once the envelope has been fully accreted onto the star it becomes a Class II object; it has no envelope, but still an accretion disc consisting of gas, dust and rock. A Class III object can be observed to have channels in this disc through which planets orbit the star. This suggests that planets have already formed when the object reaches Class III, which points towards Class I and II as being the site of planetary formation.

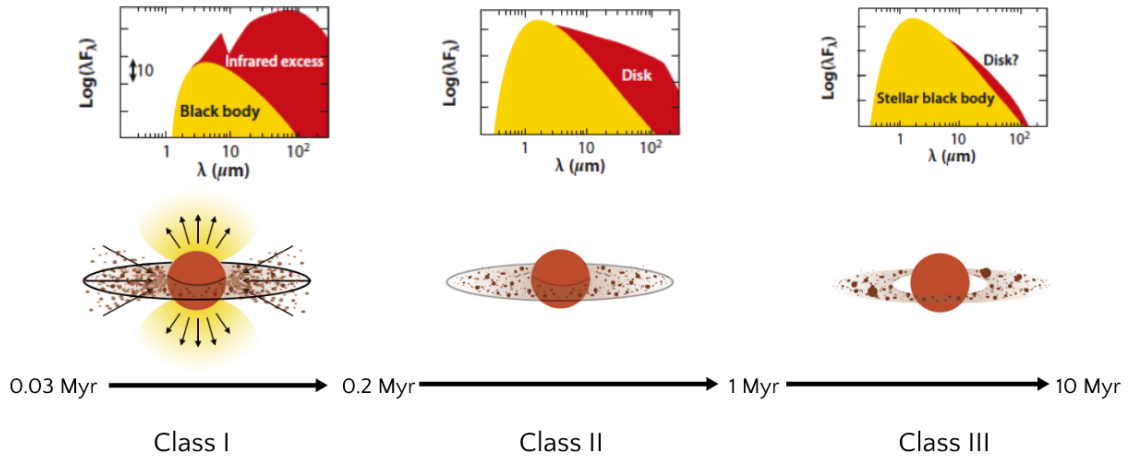


Figure 1: A diagram showing the SEDs (Maddison 2017) and structure (Guenther 2009) of YSOs as they evolve through their formation stages.

As mentioned above, a YSO is identified by the shape of its SED. The SED is found by plotting the flux of the observed light against wavelength of the observation. The SEDs of YSOs were first characterised by Lada & Wilking (1984), and are shown in Figure 1. The protostar itself radiates like a black-body, and so the deviation from a black-body curve seen in the SED of an object gives an insight into the additional features of the object, allowing us to place it into a YSO class. The Class I objects have a large excess due to the envelope of gas and dust which is heated by the star and radiates in the infrared, leading to the large increase in received flux in redder (longer) wavelengths. The Class II objects

no longer have that envelope, so the infrared excess is much less, however they still have a circumstellar disc of gas and dust that is heated by the star, leading to the excess shown in Figure 1. A large portion of what we see as infrared flux from the YSO is in fact light and heat from the protostar that has been absorbed and re-radiated by the surrounding disc or envelope. Finally, the Class III objects have much less of an infrared excess, the regions of the disc closest to the star have been accreted so the disc’s ‘edge’ is further from the star than for a Class II, and as a result it is heated less and gives off less infrared radiation. Class III objects have the closest SED to a black-body. The different ‘peaks’ in flux for different classes of YSO result in the younger objects being easier to spot in the infrared and harder in the optical, and vice versa for the older objects.

2.2 VARIABILITY IN YSOS

YSOs of all classes have been observed to vary in magnitude with time - i.e. they get brighter and dimmer. The variability is observed to be aperiodic, and can take the form of sudden, drastic increases or reductions of brightness taking place over a few days or can be gradual changes over hundreds or even thousands of years. The primary causes of magnitude variability are rotation and accretion ([Audard et al. 2014](#)).

Rotation driven variability arises due to geometric effects on the surface of the star. The points at which columns of material accrete onto the star get hotter than their surroundings, and as the star rotates these ‘hot spots’ move in and out of view, causing the bluer, optical flux received from the star to vary as the star rotates. Since most YSOs have rotation periods on the order of a few days ([Littlefair et al. 2010](#)), the variability associated with rotation is short-timescale, and as such is mostly outside the scope of this investigation.

The long-timescale variability is driven mostly by accretion. The mass of the star increases as it accretes material, and as such its gravitational potential increases. In order to balance against the newly increased mass, the radiation pressure of the star must also get stronger, leading to an increase in the brightness of the star. The faster the rate of accretion, the more mass is added to the star in a unit time, and so a greater change in radiation pressure is needed. This means that the brighter the star gets, and the shorter the timescale the change occurs over, the faster the rate of accretion. This increase in radiation pressure can cause material to be ejected from the star, causing it to lose mass. This means the radiation pressure needs to decrease to balance against the newer, lower gravitational potential, and so the star gets dimmer. The ejection of material associated with these ‘outbursts’ can also cause ripples of material to travel through the disc, creating dynamic instabilities in the movement of material through the disc, which are thought to be contributing factors to how, when and where in the disc planetary objects form. By characterising the long time-scale variability of YSOs, a picture can be gained of the rate

of mass accretion onto an object. One method of characterising this variability is through the use of structure functions.

2.3 STRUCTURE FUNCTIONS

A ‘structure function’ (SF) is a statistical method of quantifying the variability of a data set. Structure functions have been applied to astrophysical objects in extra-galactic astronomy (de Vries et al. 2003), but were only recently applied to YSO variability in Sergison et al. (2020) (henceforth Sergison), in which they were used to characterise the i-band variability of YSOs over 6 orders of magnitude. The main benefit of structure functions as a method of quantifying variability is that they take into account the relationship between all possible points of a lightcurve. The SF is calculated in discrete logarithmically spaced time-scale bins by taking all pairings of points τ_i and τ_j in a given bin, where

$$\tau_1 < \tau_i - \tau_j < \tau_2,$$

and τ_1 and τ_2 are the lower and upper limits of each time-scale bin respectively. The SF can then be calculated as

$$S(\tau_1, \tau_2) = \frac{1}{N(\tau_1, \tau_2)} \sum \frac{(F_i - F_j)^2}{F_{\text{med}}^2}, \quad (2.1)$$

where the summation is calculated over all $N(\tau_1, \tau_2)$ pairs of points (i, j) , which have fluxes F_i and F_j . The summation is then divided by the median flux of all the pairs of points within the bin. It can be common to divide by the mean flux of all pairs, but the median is used in this case as it is less susceptible to change due to outlying values. Each calculated SF is labelled $S(\tau)$ where τ is the geometric mean of τ_1 and τ_2 . When plotting structure functions, a logarithmic scale of $\sqrt{S(\tau)}$ against τ is used, as the square root allows the SF to be analogous to root mean square variability.

Structure functions have an advantage against other signal processing techniques such as Fourier Transforms as they are calculated within the time-domain, meaning their reliance on a consistent sampling interval is greatly reduced. Since YSO lightcurves are by necessity discrete and sparsely sampled - and often have a wide variety of cadences - structure functions become an extremely useful tool through which to analyse variability. Another advantage to using structure functions over something like a Fourier transform is that structure functions are sensitive to all variability, including aperiodic variability, whereas Fourier transforms are preferentially sensitive to periodic variability. Since the causes of variability in YSOs are being investigated, it is best to implement a method sensitive to all variability so that the results are not biased towards a certain type of variability through the method.

One problem when applying structure functions to stellar observation is that at short timescales, lightcurves from stars are subject to background variation introduced through

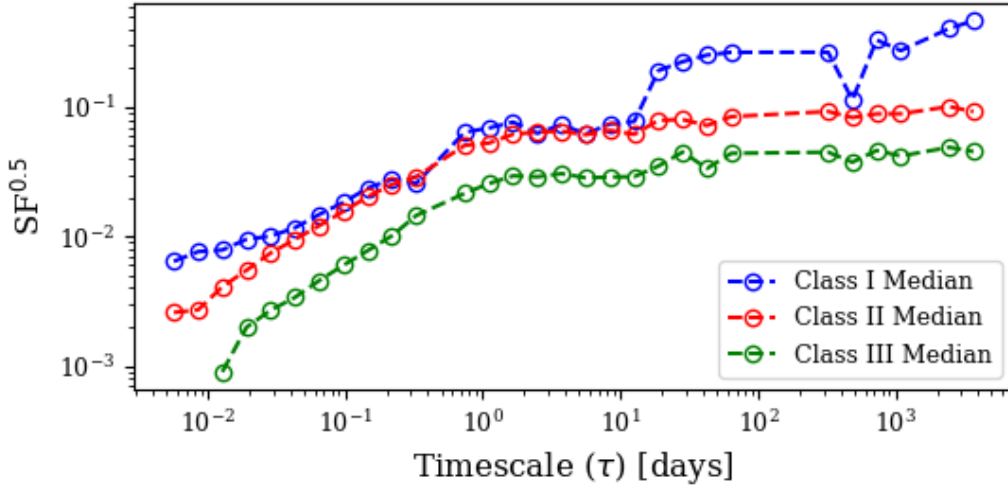


Figure 2: Adapted from [Sergison et al. \(2020\)](#), this plot shows the median SF for each YSO class.

instrumental and systematic noise. The structure function will interpret that noise as short-timescale variability, and as such any variability on short timescales shown by the structure function will be swamped by noise and unlikely to show anything physical. This effect can be reduced by observing a known non-variable object and calculating its structure function, where all variability will be due to instrumental or systematic noise, and then subtracting the effects demonstrated by the non-variable structure function from the structure function of the YSO. Such structure functions will be referred to as ‘Background Subtracted Structure Functions’ or SSFs.

2.4 PRIOR WORK

[Sergison et al. \(2020\)](#) investigated the OB association Cep OB3b using observations on timescales from 1 minute to 10 years. They found over 800 YSOs within the cluster, and took near-infrared i-band ($\lambda = 7700\text{\AA}$) observations of the objects using the Isaac Newton telescope in La Palma. After categorising the YSOs following the [Allen et al. \(2012\)](#) classifications based on *Spitzer* colours, they obtained a sample of 12 Class I, 500 Class II, and 274 Class III objects. Structure functions were then calculated and used to create Figure 2, which plots median SFs for the different classes of YSO observed to allow for class to class comparison. It was found that Class II and III objects displayed a power-law increase in variability up to a timescale on the order of a few days, which was found to correspond with the rotational period of the objects found by [Littlefair et al. \(2010\)](#), after which the structure function levels out. For Class I objects an approximately $\tau^{0.8}$ increase in variability was observed continuing over all timescales, with less evidence of the structure function leveling off above the rotational period - however it is highlighted

in Sergison that due to the small sample size of Class I objects this may not be the typical behaviour of these types of object. It was suggested that all YSOs have two primary components to their variability. The first is an underlying power-law variability spectrum driven by accretion rate changes, and the second is an approximately sinusoidal spectrum resulting from the geometric effects driven by the rotation of the star. It was suggested by [Sergison et al. \(2020\)](#) that for objects with dominant long-timescale variability, like the Class I objects, the rotational modulation is smaller as they are seen at different lower inclinations or have more complex magnetic field geometries.

This report intends to further investigate the behaviour of Class I objects relative to Class II objects, and attempts to verify whether the different behaviour for Class I objects at long-timescales is physically significant, or if it is a result of the small sample size used for Class I objects (12 objects) compared with Class II and III (500 and 274 respectively). To do this, YSOs will be selected from the whole sky and observed in both optical and infrared wavelengths to gain a larger sample of Class I objects which can be compared more confidently against Class II and III YSOs.

3 DATA COLLECTION

To investigate the variability of YSOs in different wavelengths, observations of star forming regions were needed in these wavelengths, so Zwicky Transient Facility (Masci et al. 2019) (ZTF) and Vista Variables in the Via Lactea (Saito et al. 2015) (VVV) observations were chosen. The ZTF data contains optical r-band ($\lambda = 6200\text{\AA}$) observations taken between March 2018 and December 2019, and the VVV data contains infrared K-band ($\lambda = 21500\text{\AA}$) observations taken between October 2013 and October 2015. To allow accurate classification of the observed objects, various surveys were used to determine which objects were likely YSO candidates, and what their respective classes were.

The first survey used to classify YSOs was the Marton et al. (2016) Support Vector Machine selection of YSO candidates, which combined Wide-Field Infrared Survey Explorer (WISE) observational data with Planck dust opacity values to obtain a likely classification of the objects in the survey, identifying 742,586 Class I, II and III YSOs. Photometric criteria from Koenig & Leisawitz (2014) were then applied to the Marton sample to refine the estimation of the likely class of each object. This method identified 10,233 likely Class I objects and 21,666 likely Class II objects. Unfortunately, due to reasons laid out in Section 4.3 and Section 4.4, the validity of the Marton sample was called into question, as data suggested that the sample was contaminated by faint objects with periodic variability, perhaps pulsating red giants. McBride et al. (2020) supported this conclusion, applying a trained neural network model to identify pre-main-sequence stars and then checking its results against other surveys. For the 2016 Marton sample, McBride found that of the 742,586 objects identified by Marton, only 3,722 sources - or 0.5% of the sample - were identified by the neural network within a confidence limit $> 70\%$, and only 17,472 sources - 2.3% - were identified without any quality checks on the data.

Due to the unreliability of the Marton sample, the Massive Young Star-forming Complex Study in Infrared and X-rays (MYStIX) catalogue (Feigelson et al. 2013) was used, which collects samples of YSOs based on their Infrared and X-ray profiles. More specifically, two components to the MYStIX catalogue were used. First, the stars identified by Broos et al. (2013) were used, a sample which identifies young stars through two pathways, the first consists of X-ray detections of young and massive stars with coronal activity/strong winds through a Bayesian classifier, and the second characterises infrared excess via their SEDs, as covered in Section 2.1. Later in the project, as laid out in Section 4.3, the Broos sample was further refined using Povich et al. (2014), which performs further analysis on infrared and near-infrared data. By combining these MYStIX surveys, a data set was obtained containing 67 Class I objects and 396 Class II objects in ZTF, and 226 Class I and 272 Class II objects in VVV.

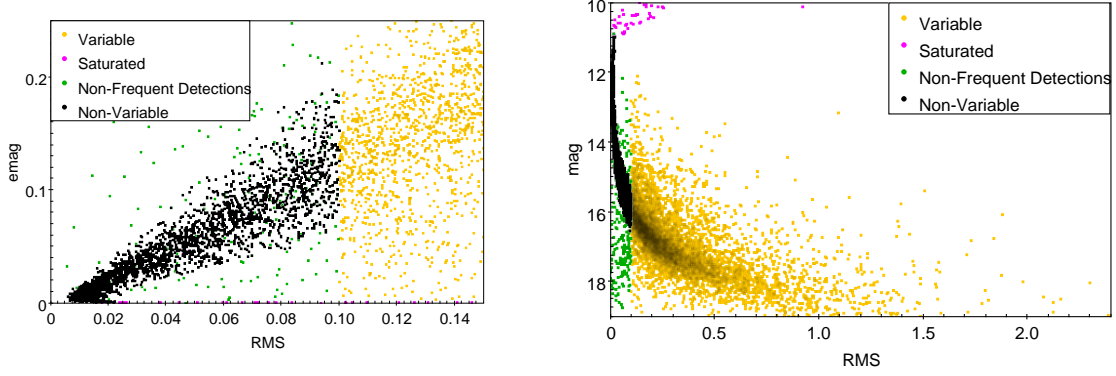


Figure 3: *Left:* the ‘emag’ vs RMS plot showing evidence for the RMS < 0.1 cut-off. *Right:* the full sample of stars taken from the ‘Golden Region’ and how it has been constrained. Image from Matsumoto, R., private communication.

To allow for the creation of Background Subtracted Structure Functions (Section 4.2), it was also necessary to obtain a sample of non-variable stars with a range of magnitudes, and to be confident that these stars displayed no variability. A region of the galactic plane with low amounts of star formation was obtained via private communication with Wilson, A. and nicknamed the ‘Golden Region’. The Golden Region lies at $160' < l < 162', 0' < b < 4'$, and a cone search of radius 600 arcseconds extracted 9621 objects. A minimum number of necessary observations was set at $\nu = 40$ to ensure there were sufficient observations to create an accurate structure function for each star, and a maximum magnitude of observation in VVV was set at the 11th magnitude, since observations brighter than this often saturated the telescope’s CCD leading to errors in photometry. A restriction on RMS variability was set to be < 0.1 ; this cutoff was chosen because if the stars were truly non-variable, the RMS variability should be zero, but in reality the effects of the atmosphere and calibration differences between observations introduce some non-physical variance, termed ‘emag’, where RMS should equal emag if all observed variability is due to non-physical variations. As can be seen in Figure 3 (left), below RMS = 0.1 the gradient of the sample is much more clearly defined and close to 1 than above that cut-off. The total sample of stars from the Golden Region is shown with all constraints applied in Figure 3 (right), and the final sample of selected non-variable stars contained 2,356 objects.

After a sample of stars had been identified, the lightcurves for each star needed to be obtained and sorted to ensure each object used was within a useful magnitude range and had enough good data points (observations that did not saturate the CCD or have large errors in magnitude) to be used in the creation of structure functions. After identifying the stars for which observations were required through the above sampling methods, the Right Ascension (RA) and Declination (DEC) of the stars were extracted from the surveys. Constraints were then imposed on the data to be downloaded through a series of criteria that were developed over the course of the study with the aim of getting the best possible

lightcurve data points while also keeping a good number of observations and minimising contamination. To obtain ZTF observations of the stars, the RA and DEC co-ordinates were converted into an IPAC file to be used on the IRSA (Infrared Science Archive) database, and then the criteria imposed were:

Criteria	Value	Effect
mag	< 20	only stars brighter than 20th magnitude
magerr	< 0.05	only stars with an error in magnitude less than 0.05 mags
ngoodobsrel	> 50	only stars with over 50 good observations released
catflags	== 0	stars that had been identified by ZTF as having “good” observations

For VVV, the process is slightly different. The RA and DEC co-ordinates are obtained in the same way, then uploaded to the VISTA Science Archive (VSA) Cross-ID form to crossmatch against VVV observations, and the `sourceID` of the available objects is saved, and applied to an SQL query which is submitted to the VSA Region Search form. The SQL query can be found in Appendix A, but the most important arguments are `aperMag5 > 11` and `ppErrBits < 256`, which set the brightest magnitude obtained at 11th magnitude (due to CCD saturation above this magnitude as shown in Figure 3), and constrains the photometric quality to the recommended value in the documentation.

4 METHOD

4.1 VALIDATING THE METHOD USING SERGISON ET AL.

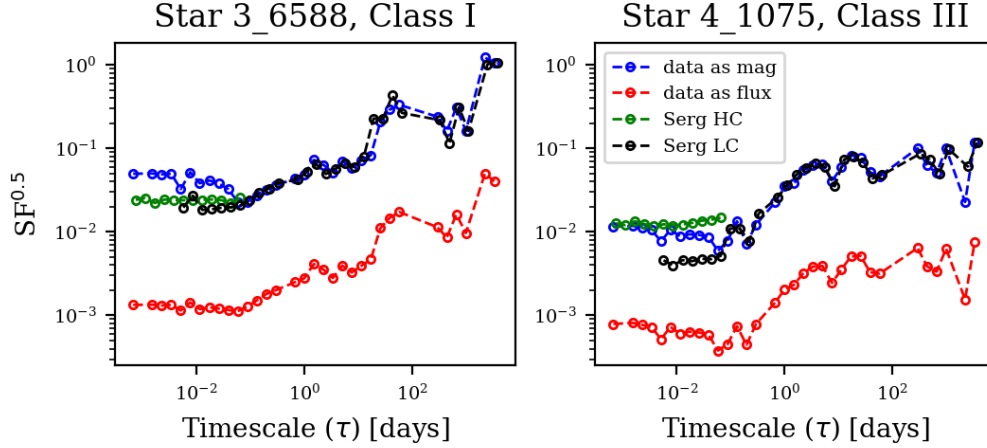


Figure 4: A reproduction of Figure 11 from [Sergison et al. \(2020\)](#) (data obtained from Naylor, T., private communication) with calculated data plotted over the Sergison data, both including the flux-to-mag function, and after removing it. It can be seen that removing this function causes the calculated data to match with the Sergison data much more closely.

The calculation of the points of the structure functions for each light-curve was performed by a FORTRAN code (obtained from Naylor, T., private communication). The first step was to validate the code by feeding data obtained from the [Sergison et al. \(2020\)](#) electronic tables through the code to confirm the same results are obtained. The objects chosen for this were two objects with structure functions explicitly presented in Sergison, to allow for direct comparison. The initial calculated structure functions are represented by the red line in Figure 4, and can be seen to be consistently around a factor 10 less variable than those presented in Sergison. The source of this discrepancy was due to the fact that the FORTRAN code expected an input in flux but was receiving an input in magnitudes. By commenting out the flux-to-mag function in the FORTRAN code, the discrepancy was eliminated, and the calculated structure function matched more closely to those presented in Sergison, as evidenced by the blue line in Figure 4. There is still a slight deviation between the calculated structure function and those presented in Sergison, because the data from Sergison has had photon and instrumental noise subtracted prior to plotting, whereas the calculated data is plotted from the raw lightcurve. Additionally, the Sergison structure functions are taken separately for the sample of high-cadence (low timescale) data and low-cadence (long timescale) data, whereas the lightcurves that lead

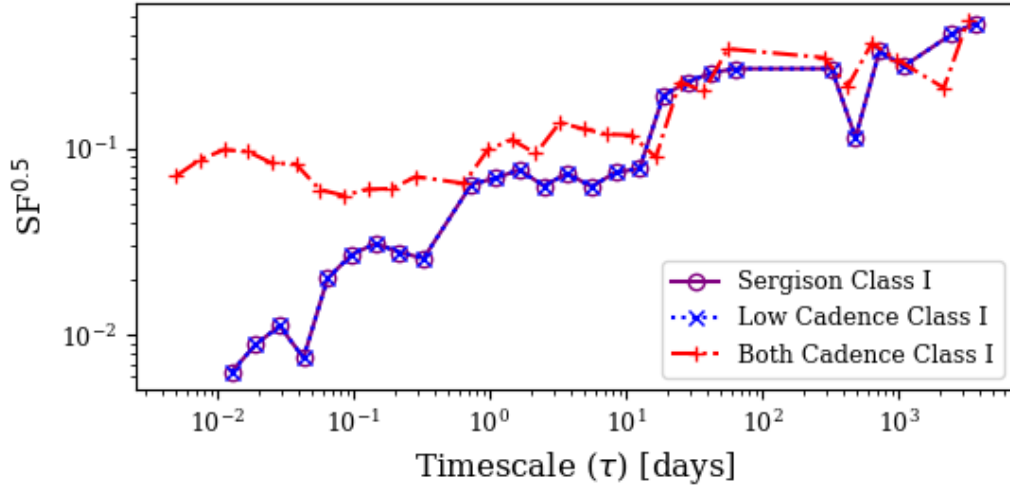


Figure 5: A recreation of Sergison Figure 12 (data obtained via Naylor, Private Communication) showing presented median SFs for Class I, II and III YSOs, with calculated median SFs overlaid, using the full data set and the Low Cadence only data set. It can be seen that while the Median SF for the full data set is close to the Sergison Class I median, the Low Cadence only data set matches exactly. This confirms the accuracy of the Python code used to calculate median structure functions.

to the calculated structure function contained all observations in one data-set.

After verifying that the SF code was working as expected, the next step was to work towards calculating a median SF. In the Sergison paper, Figure 12 presents median structure functions for multiple classes of YSO, and as such the data from Sergison was again used to attempt to recreate this figure. The Sergison structure functions were downloaded from the Sergison electronic tables and then run through a custom Python code to calculate the median SF, by taking the median value of variability for each timescale present in the data. As can be seen in Figure 5 the data was close at long timescales, but differed at shorter timescales. This discrepancy arose because the Sergison data sample contained two types of observation, ‘High Cadence’ and ‘Low Cadence’ data, referring to the frequency of observations of each object. The Sergison data contained in part a set of 82 ~ 30 second exposures taken over one night on 10 November 2013, which make up the High Cadence data set. These observations were taken to investigate the short-timescale behaviour of YSOs, whereas the rest of the data - referred to as Low Cadence - consisted of ~ 450 long-exposure observations, taken over 16 nights in September 2004, and 68 nights in August to November 2005, and a few more observations from 2007 and 2013. The data downloaded from the Sergison electronic table and fed into the median SF code contained both High and Low Cadence data, whereas the data used to create Figure 12 from Sergison consisted of only Low Cadence data. Once only the Low Cadence data was fed into the median SF code, an output was obtained that exactly reproduced Figure 12 from Sergison, confirming that the method employed by the median SF Python code

generated correct results.

4.2 ADAPTING THE METHOD FOR COLLECTED DATA

After verifying that the structure functions and median structure functions were being calculated correctly, an attempt was made to apply the workflow to some data from other sources. The first sample of Likely Class I candidates observed via ZTF (the Marton sample, as outlined in Section 3) was used for the first test, which successfully ran through the structure function code, but when attempting to create a median SF for these Class I objects, the first obstacle to creating median SFs became apparent. The FORTRAN code calculates a minimum timescale bin to sort observation pairs into, and then calculates all following bins by multiplying the previous bin by 1.5. Due to this, depending on what the minimum bin size was, observations of different objects would have structure function points at different timescales, making the calculation of the median at a certain timescale difficult. For example, if the first timescale bin sat at 1 day, then the following bins would be 1.5 days, then 2.25 then 3.38 etc. If for the second star, the first timescale bin happened to be calculated as 0.5 days, then the second bin would be 0.75 days, then 1.13 then 1.69. It can be seen then, that finding a median variability for any timescale between these two stars would be difficult, as none of the timescales for which there is a measure of variability coincide. This problem was overcome by manually setting the minimum timescale bin to a set width of 0.005 days, ensuring that for all structure functions, the data series would have matching timescale values, making the median easier to calculate.

After solving the timescale problem, another problem was encountered where if some stars in a sample had no observations within a certain timescale bin - a situation that was very likely as observations were being taken from multiple different locations in the sky, and hence many different sampling rates for the lightcurve - then that would affect the median structure function calculated as Python would attempt to median an array with empty elements - similar to `[1, 2, , 3, 4, ,]` - which would give an error. This problem was overcome by finding the structure function with the most data-points, and then replacing the 'empty' elements in all other structure functions up to that length using the Python `None` datatype, which would not affect the calculation of the median. This allowed for the calculation of median structure functions for an object type with observations taken from all over the sky, not just from one location. The `None` data type needed to be used instead of, for example, 0, as choosing a numerical value would impact the calculated median for that timescale bin.

The final task that needed to be completed was the subtraction of background noise from a structure function. As outlined in Section 2.3, at short timescales SFs are very susceptible to instrumental and background noise, but this can be combated by calculating the structure function of a known Non-Variable Star (NVS) of a similar magnitude to the

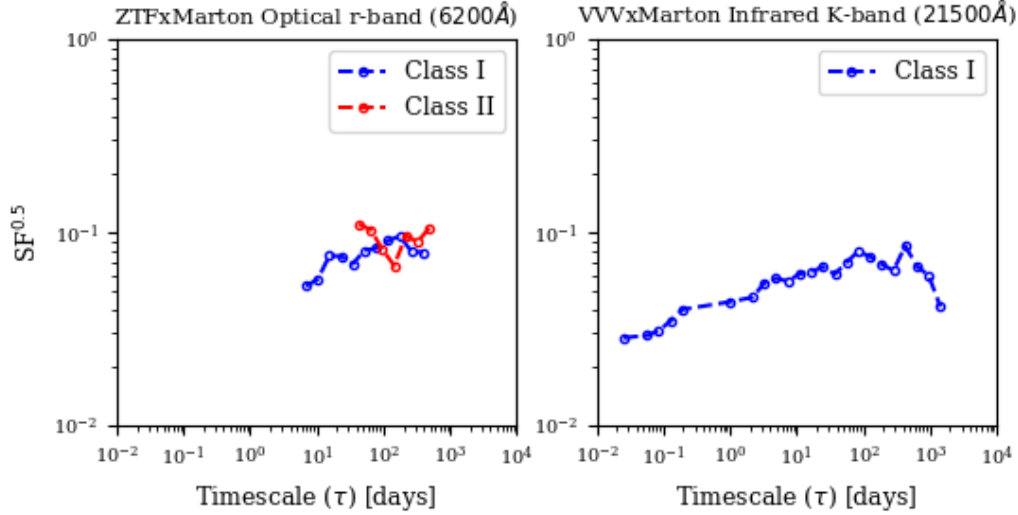


Figure 6: Median SFs for the Class I and II ZTFxMarton data in the optical r band, and the Class I VVVxMarton data in the infrared K band.

desired YSO, and subtracting the variability displayed by the non-variable star from that displayed by the YSO. The calculation of a ‘background subtracted structure function’ was initially attempted by calculating the mean magnitude of all YSOs in a Class; finding NVSs with mean magnitudes close to this magnitude; creating a median SF of these NVSs; and then for each timescale point, subtracting the Non-Variable-Median from the Class-Median. This method was later improved, by instead subtracting noise from each YSO individually, and then creating the final median SF from the selection of Background Subtracted Structure Functions (SSFs). To do this, it was useful to categorise all YSOs in a sample into bins of half-magnitude, so a Python code was written to calculate the mean magnitude of each object and sort them into their respective bins. The same code was also applied to the sample of suspected NVSs, obtained as described in Section 3. A structure function was then calculated for each NVS in each half-magnitude bin, and a median non-variable SF calculated for each half-magnitude bin. This would give a good representation of the average variation due to noise for an object with magnitude within that bin. Then each YSO in a half-magnitude bin had a SSF created by subtracting the median non-variable SF from its individual SF. A flow chart visually expressing the steps of this process is given in Appendix B, which may help conceptualising the order of steps taken. The final median structure function was then calculated by taking a median of all subtracted structure functions within a sample, giving a Median Background Subtracted Structure Function for a sample of objects.

4.3 APPLYING THE METHOD

After verifying that the methods to be used successfully replicate Sergison data, they were applied to the data sets to be investigated. The first data sample used was [Marton et al. \(2016\)](#), for both the VVV and ZTF observations. These data sets will be referred to as ZTFxMarton and VVVxMarton respectively. The ZTFxMarton sample contained 65 likely Class I candidates¹ and 110 likely Class II candidates, and the VVVxMarton sample contained 3227 likely Class I candidates. The median SFs produced by these data sets is shown in Figure 6. The data suggested that for the ZTF data Class I objects showed no more variability than Class II objects, and for VVV the structure functions did not display the expected power law increase in variability up to timescales of around 1-10 days, which called into question the validity of the inputted sample, as the expected shape for a YSO structure function as shown in Sergison would be a rise upward towards a timescale of ~ 10 days, followed by a levelling out. It can also be seen that there is no presented structure function for VVV Class II objects, as many objects in the sample displayed less variability than the non-variable objects, resulting in subtracted structure functions with negative values for $S(\tau)$. Both the lack of variability for Class II objects and the shape of the structure functions for Class Is suggested that the sample was heavily contaminated by NVSs. It was decided that this sample required further investigation to ensure that all objects were the objects they had been classified as in [Marton et al. \(2016\)](#).

Due to the scepticism of the accuracy of the Marton survey, the Broos survey MYStIX project was also used to separately identify YSO candidates from the ZTF data sets, as explained in Section 3. The ZTFxBroos sample contained 68 likely Class I candidates, however when a median SF was created for this sample it maintained the flat shape seen for the Marton sample, suggesting that it still was subject to contamination. Inspection of the SFs of individual objects within the sample confirmed this, as while some objects showed the expected SF shape, many others had the flat SF expected for a non-variable star. This led to further scepticism of the accuracy of the sample, and further investigation was performed on the accuracy of the data. After decreasing the acceptable error in magnitude for an object ($\text{magerr} < 0.05$) to be included in the sample, and switching to the ZTF Data Release 3 which was released during the work on this project, a sample of ZTFxBroos objects was obtained which contained 88 likely Class 1 candidates, which when plotted showed a shape more similar to what was expected, as shown in Figure 7. This sample was then used to calculate a ‘noise subtracted median SF’, or ‘median SSF’, using the method outlined in Section 4.2 and Appendix B.

When creating the median SSF a sample of non-variable stars is needed. The ‘Golden Region’ stars were used, as outlined in Section 3. After removing Class I YSOs without

¹For clarity, the term ‘likely Class N Candidates’ refers to stars that, at the time of processing, were being treated as the object stated, rather than a judgement on the actual likelihood an object was a certain class. The accuracy of these classifications is discussed in more depth later in this section.

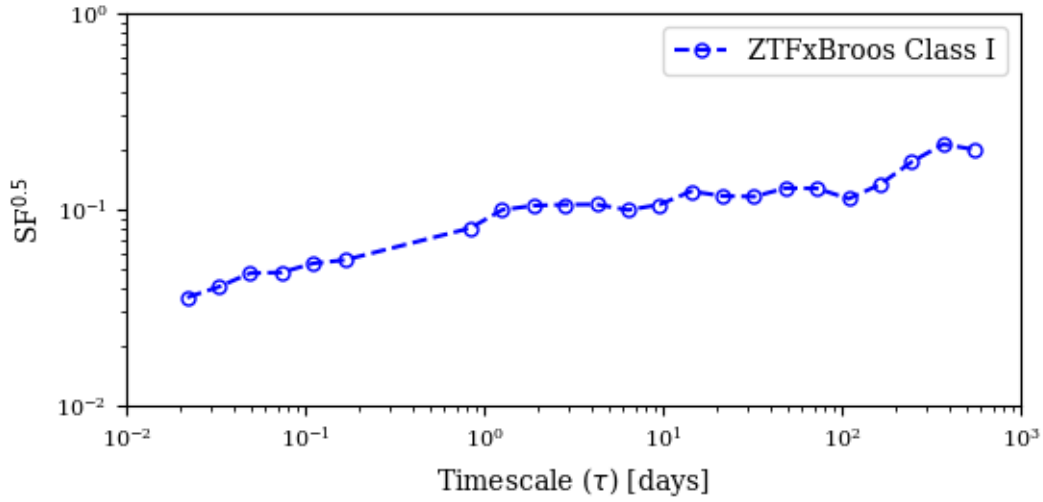


Figure 7: A median SF for Class I candidates identified within the ZTFxBroos sample.

a NVS equivalent magnitude bin, a median SSF of the remaining 75 Class I YSOs was created, and was observed to have the expected shape. This was taken as evidence that the MYStIX catalogue was providing a more reliable sample of stars, although it was decided that going forward a minimum number of observations would be required for a star to be included in a sample, to ensure that data was only being collected for objects that could be turned into useful structure functions.

Since a median SF of Class I objects in optical wavelengths had been obtained, it was required to obtain a sample of Class II objects. A sample was extracted from MYStIX containing 768 stars with numbers of observations varying between 1 observation and 281. It was chosen that any objects with less than 199 observations would be discarded, as this roughly halved the sample and ensured that behaviour seen in a median SSF would be more likely to be driven by physical behaviour - this brought the sample size down to 381. However when attempting to create a median SSF for this sample, many of the outputted median SSF values for each timescale ended up being negative, suggesting that the variability of the YSOs was less than the variability of the NVSs used to estimate background noise, which suggested that the sample used for Class II objects was contaminated heavily with NVSs. By plotting a median SSF for each half-magnitude bin, as shown in Figure 8, it was concluded that the dimmer magnitude bins were most likely to be contaminated. A new sample of YSOs was obtained applying the additional Povich et al. (2014) criteria on top of the MYStIX Broos et al. (2013) criteria, as described in Section 3. The new sample was investigated using a new metric of variability before returning to structure functions - cumulative distribution graphs.

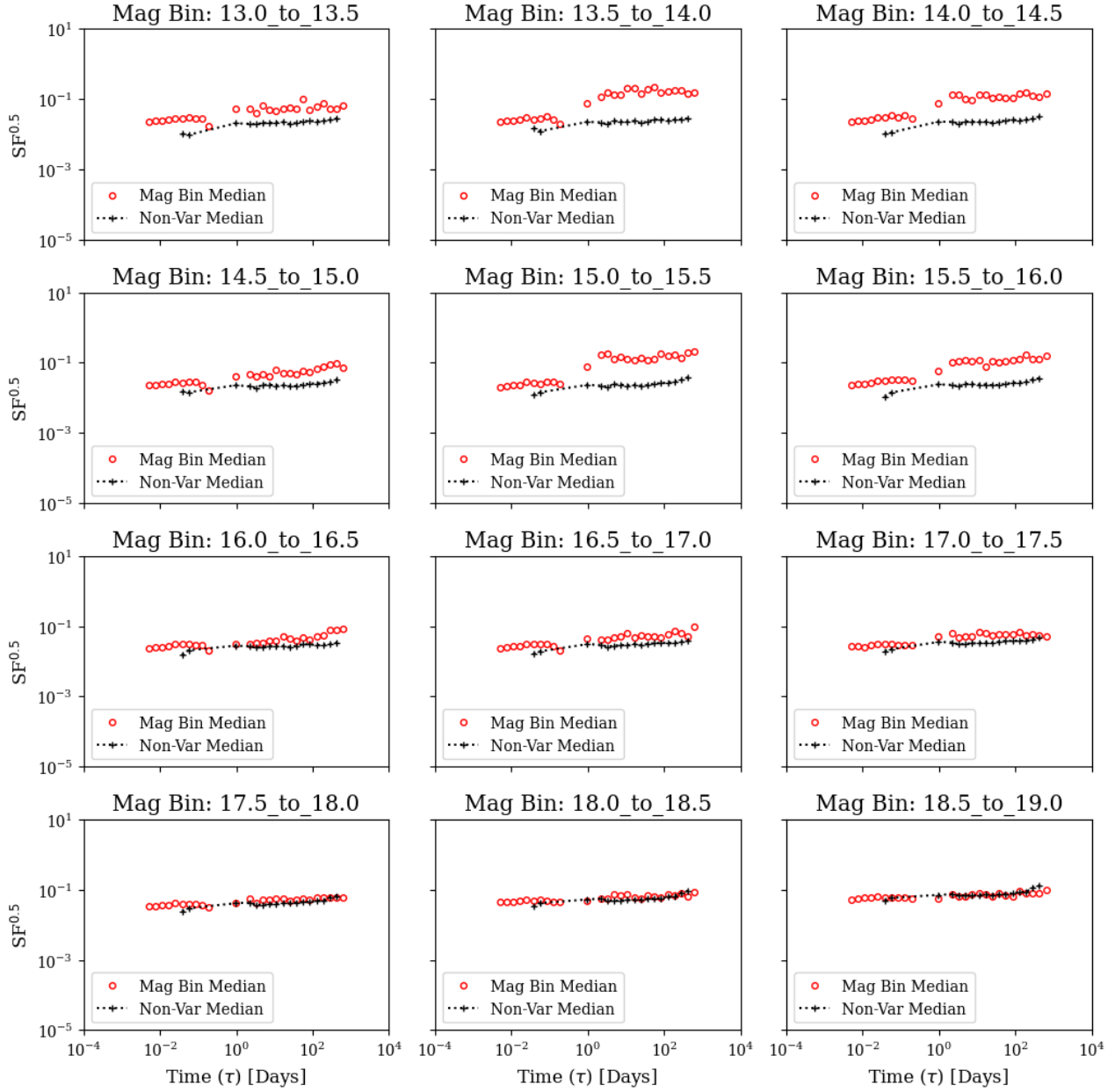


Figure 8: The median SF for the Class II ZTFxBroos sample plotted per half-magnitude bin, alongside the NVS median for the same magnitude bin. This suggests the dimmest magnitude bins (which are also the most populated) are the most likely to be contaminating the sample.

4.4 CUMULATIVE DISTRIBUTION GRAPHS

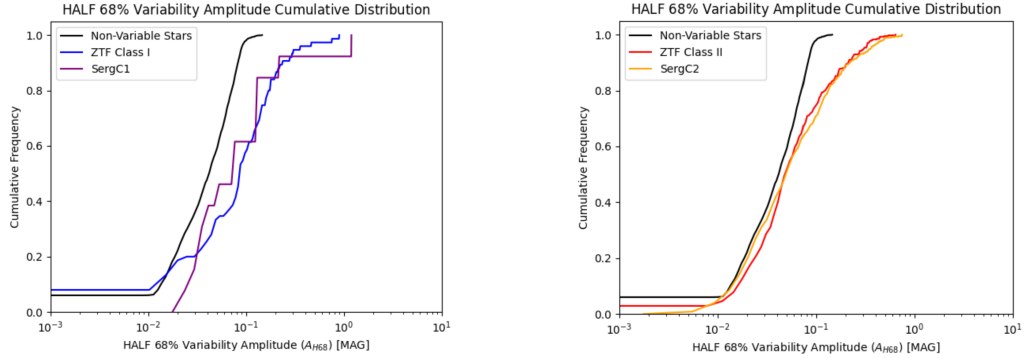


Figure 9: A_{H68} cumulative frequency graphs comparing ZTF Class I and Class II observations with the variability amplitudes presented in [Sergison et al. \(2020\)](#).

A simple way to determine the amplitude of variability is to take the difference between the brightest and dimmest observations of an object. However, such a metric is dependent on how many observations are available, and increases as more data is obtained and the full range of variability is explored. If instead the difference between the 16th and 84th percentile points in the distribution of magnitudes is taken (or the range enclosing 68% of the data) then the variability amplitude will not depend systematically on the number of observations. By taking the half-68% variability amplitude (A_{H68}) as the metric, it will tend towards a fixed value as the number of datapoints increases. This method was employed in [Sergison et al. \(2020\)](#), and also applied to the ZTF and VVV data sets to see if the same behaviour is observed. In Figure 9 it can be seen that the half-amplitude variability for ZTF observations closely follows the same shape presented in Sergison, supporting the conclusions drawn in that paper. The further right on the x-axis the line on the cumulative frequency graph, the more variable the object is. The non-variable stars are presented on both graphs as a baseline, so that it can be seen how much more variable the different classes of YSO are. In Figure 10 it can be seen that for both ZTF data and VVV data the Class I objects are more variable than Class II objects, and the VVV infrared objects appear to display more variability than the Class I objects. Structure functions were then applied to the new [Povich et al. \(2014\)](#) sample to attempt to verify these results.

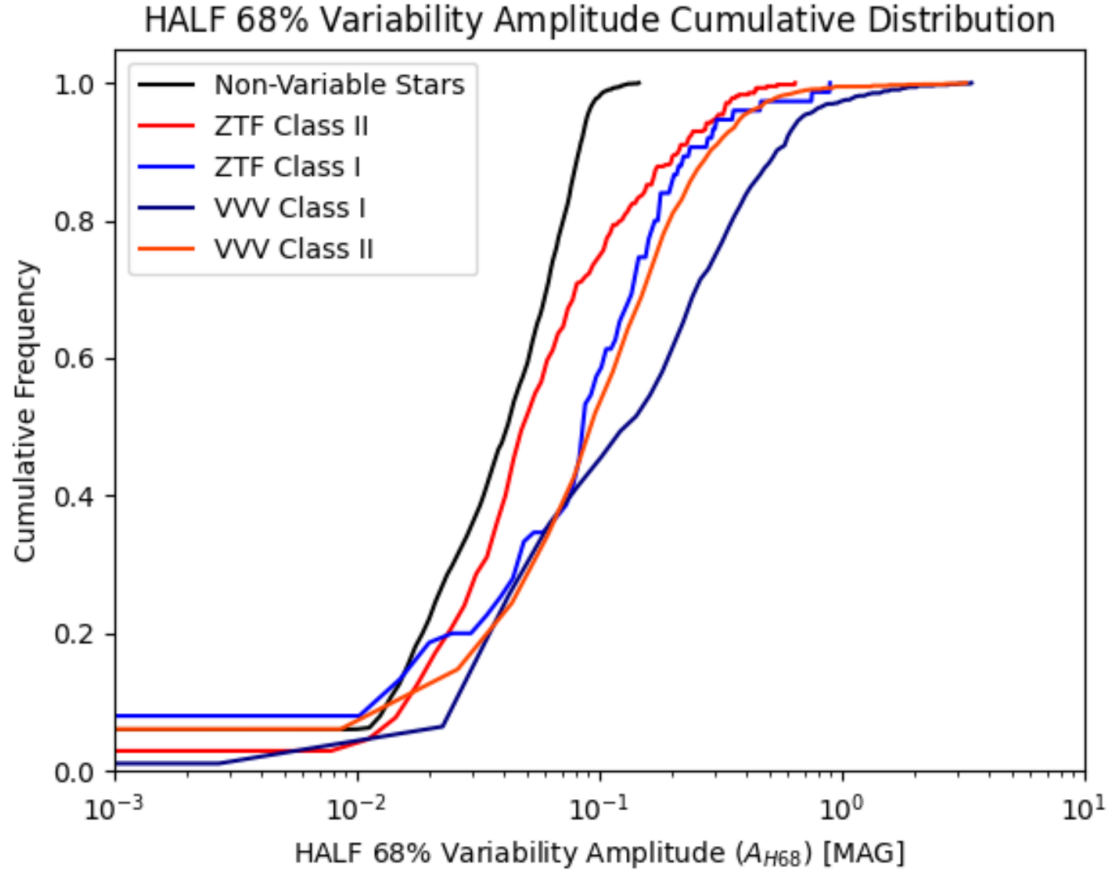


Figure 10: The cumulative distribution graphs for the different classes of object in ZTF and VVV. It can be seen that for both ZTF and VVV the Class I objects vary more than the Class II objects, and that the variability of the objects in the infrared exceeds that in the optical. Figure obtained from Reid, C., private communication.

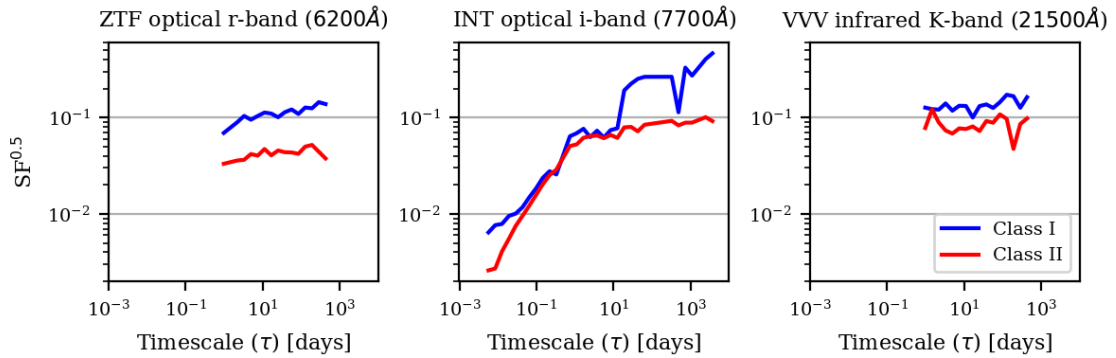


Figure 11: A comparison between the structure functions for the different classes of object in the differing wavebands. It can be seen that in each waveband, the Class I objects display more variability than the Class II objects.

5 FINAL RESULTS

Figure 10 supports the conclusion drawn in [Sergison et al. \(2020\)](#), showing that for both VVV and ZTF observations the Class I objects display more variability than the Class IIs. This figure also shows that the observations of the YSOs in infrared display more variability than in the optical. Structure functions were created from the same sample to see if similar results were obtained, and are presented in Figure 11. This figure shows median structure functions for 67 Class I and 396 Class II objects observed in optical wavelengths by ZTF, and 226 Class I and 272 Class II objects observed in infrared by VVV, all as classified by [Povich et al. \(2014\)](#), compared against the 12 Class I and 500 Class II objects observed in near-infrared and analysed in [Sergison et al. \(2020\)](#). These structure functions support those points, showing that in the optical wavelengths, the Class I objects are twice as variable on timescales of 1 day, up to 2.6 times more variable at 100 days, and for the infrared objects, the Class Is are about 1.5 times more variable at all timescales. An additional result supported by both Figure 10 and Figure 11 is that the Class II objects show more variability in infrared than in optical, the structure functions show roughly twice as much variability in infrared than in optical.

6 CONCLUSIONS

The first conclusion - that Class I objects display more long-timescale variability than Class II objects - is consistent with the results published in [Sergison et al. \(2020\)](#). Since long-timescale variability is most likely to be driven by accretion (Section 2.2), this result suggests that the accretion of matter onto the Class I objects occurs at a much more variable rate than the accretion of matter onto Class II objects, implying that the dynamical processes occurring in the envelope and disc are more volatile time-dependent than the dynamics surrounding Class II objects. This is a useful result for understanding planetary formation, where clearer understanding of the dynamic processes that drive the formation of planets and proto-planets is constantly being sought.

The second conclusion, that Class I and specifically Class II YSOs display more variability in the infrared than the optical, is more puzzling. Due to the dusty envelope and circumstellar disc surrounding Class I objects which would be expected to extinct optical light more than infrared light, it could be expected that optical wavelengths would show more variability than the infrared. A suggested explanation for this behaviour is that, since the infrared wavelengths probe the variability of the cooler accretion disc more than the hotter proto-star, then we could be seeing a geometrically driven variation arising from a variability in height of the accretion disc. It is possible that, since a large source of infrared flux is heat from the star that is absorbed and re-radiated by the disc, as the disc increases in height the surface area that radiates infrared flux increases, leading to more flux being emitted, and then as the disc thins the flux decreases. If this is the cause of the variability, it would suggest that the height of the accretion disc changes irregularly over the course of the YSOs evolution, which could provide dynamical instabilities that impact on proto-planetary formation.

7 FUTURE WORK

There are multiple avenues down which further research could be taken. One such avenue is to use the GAIA variability information found in Data Release 3 upon its release (currently scheduled as late 2021) to obtain more accurate light-curve data for potential YSO candidates, which will lead to even more accurate structure functions, especially using light-curves with low-cadence data to allow further investigation into the short-timescale behaviours of YSO, which were largely ignored in this report.

It is also possible that the age of a cluster in which the Young Stars are forming may have an impact on the variability, so by creating median structure functions for objects within separate clusters, and then estimating the age of the clusters through the use of isochrone data ([Bell et al. 2014](#)) to see if a correlation can be found between the age of a cluster and the variability of the YSOs within it.

It may also be useful to take i-band observations of multiple separate star forming regions, to confirm whether the i-band behaviour reported in [Sergison et al. \(2020\)](#) is general behaviour of i-band observations of all YSOs, as it has been treated in this paper, or behaviour specific to the young cluster Cep OB3b. This could also be combined with the previous avenue to confirm whether the i-band behaviour is related to the age of Cep OB3b.

8 ACKNOWLEDGEMENTS

The author would like to thank Andy Wilson, graduate student at the University of Exeter for sharing the location of the ‘Golden Region’ for non-variable stars.

This report is based in part on observations obtained with the Samuel Oschin 48-inch Telescope at the Palomar Observatory as part of the Zwicky Transient Facility project. ZTF is supported by the National Science Foundation under Grant No. AST-1440341 and a collaboration including Caltech, IPAC, the Weizmann Institute for Science, the Oskar Klein Center at Stockholm University, the University of Maryland, the University of Washington, Deutsches Elektronen-Synchrotron and Humboldt University, Los Alamos National Laboratories, the TANGO Consortium of Taiwan, the University of Wisconsin at Milwaukee, and Lawrence Berkeley National Laboratories. Operations are conducted by COO, IPAC, and UW.

It is also based in part on data products from VVV Survey observations made with the VISTA telescope at the ESO Paranal Observatory under programme ID 179.B-2002.

The Isaac Newton Telescope is operated on the island of La Palma by the Isaac Newton

Group of Telescopes in the Spanish Observatorio del Roque de los Muchachos of the Instituto de Astrofísica de Canarias. The author wishes to recognize and acknowledge the very significant cultural role and reverence that the summit of Maunakea has always had within the indigenous Hawaiian community.

BIBLIOGRAPHY

- Alencar S. H. P., Bouvier J., Catala C., Fernandez M., Gameiro J. F., Fonseca N., Guimarães M. M., Hetem J. G., McGinnis P. T., Moraux E., Teixeira P., 2010, *Highlights of Astronomy*, 15, 735
- Allen T. S., Gutermuth R. A., Kryukova E., Megeath S. T., Pipher J. L., Naylor T., Jeffries R. D., Wolk S. J., Spitzbart B., Muzerolle J., 2012, *ApJ*, 750, 125
- Audard M., Ábrahám P., Dunham M. M., Green J. D., Grosso N., Hamaguchi K., Kastner J. H., Kóspál Á., Lodato G., Romanova M. M., Skinner S. L., Vorobyov E. I., Zhu Z., 2014, in Beuther H., Klessen R. S., Dullemond C. P., Henning T., eds, *Protostars and Planets VI Episodic Accretion in Young Stars*. p. 387
- Bell C. P. M., Rees J. M., Naylor T., Mayne N. J., Jeffries R. D., Mamajek E. E., Rowe J., 2014, *MNRAS*, 445, 3496
- Breger M., 1979, *PASP*, 91, 5
- Broos P. S., Getman K. V., Povich M. S., Feigelson E. D., Townsley L. K., Naylor T., Kuhn M. A., King R. R., Busk H. A., 2013, *ApJS*, 209, 32
- Cody A. M., Stauffer J., Baglin A., Micela G., Rebull L. M., Flaccomio E., Morales-Calderón M., Aigrain S., Bouvier J., Hillenbrand L. A., Gutermuth R., Song I., Turner N., Alencar S. H. P., Zwintz K., Plavchan P., Carpenter J., Findeisen K., Carey S., Terebey S., Hartmann L., Calvet N., Teixeira P., Vrba F. J., Wolk S., Covey K., Poppenhaeger 2014, *AJ*, 147, 82
- Contreras Peña C., Lucas P. W., Kurtev R., Minniti D., Caratti o Garatti A., Marocco F., Thompson M. A., Froebrich D., Kumar M. S. N., Stimson W., Navarro Molina C., Borissova J., Gledhill T., Terzi R., 2017, *MNRAS*, 465, 3039
- Contreras Peña C., Naylor T., Morrell S., 2019, *MNRAS*, 486, 4590
- D’Angelo C. R., Spruit H. C., 2012, *MNRAS*, 420, 416
- de Vries W. H., Becker R. H., White R. L., 2003, *AJ*, 126, 1217
- Feigelson E., Townsley L., Broos P., Busk H., Getman K., King R., Kuhn M., Naylor T., Povich M., Baddeley A., Bate M., Indebetouw R., Luhman K., McCaughrean M., Pittard J., Pudritz R., Sills A., Song Y., Wadsley J., 2013, *The Astrophysical Journal Supplement Series*, 209

- Freedman W. L., Hughes S. M., Madore B. F., Mould J. R., Lee M. G., Stetson P., Kenicutt R. C., Turner A., Ferrarese L., Ford H., Graham J. A., Hill R., Hoessel J. G., Huchra J., Illingworth G. D., 1994, *ApJ*, 427, 628
- Guenthner K., 2009, Master's thesis, Universität Leipzig
- Herbst W., Bailer-Jones C. A. L., Mundt R., Meisenheimer K., Wackermann R., 2002, *A&A*, 396, 513
- Herbst W., Herbst D. K., Grossman E. J., Weinstein D., 1994, *AJ*, 108, 1906
- Koenig X. P., Leisawitz D. T., 2014, *ApJ*, 791, 131
- Lada C. J., Wilking B. A., 1984, *ApJ*, 287, 610
- Lafler J., Kinman T. D., 1965, *ApJS*, 11, 216
- Littlefair S. P., Naylor T., Mayne N. J., Saunders E. S., Jeffries R. D., 2010, *MNRAS*, 403, 545
- Maddison S., 2017, Astronomy Online - Classifying YSOs, <http://astronomy.swin.edu.au/~smaddiso/teaching/yso.html>, [Online; accessed 05-April-2021]
- Marton G., Toth L. V., Paladini R., Kun M., Zahorecz S., McGehee P., Kiss C., 2016, *VizieR Online Data Catalog*, p. J/MNRAS/458/3479
- Masci F. J., Laher R. R., Rusholme B., Shupe D. L., Groom S., Surace J., Jackson E., Monkewitz S., Beck R., Flynn D., Terek S., Landry W., Hacopians E., Desai V., Howell J., Brooke T., Imel D., Wachter S., Ye Q.-Z., Lin H.-W., Cenko S. B., Cunningham V., Rebbapragada U., Bue B., Miller A. A., Mahabal A., Bellm E. C., Patterson M. T., 2019, *PASP*, 131, 018003
- McBride A., Lingg R., Kounkel M., Covey K., Hutchinson B., 2020, arXiv e-prints, p. arXiv:2012.10463
- Morbidelli A., Raymond S. N., 2016, *Journal of Geophysical Research (Planets)*, 121, 1962
- Povich M. S., Kuhn M. A., Getman K. V., Busk H. A., Feigelson E. D., Broos P. S., Townsley L. K., King R. R., Naylor T., 2014, *VizieR Online Data Catalog*, p. J/ApJS/209/31
- Reipurth B., 1990, in Mirzoian L. V., Pettersen B. R., Tsvetkov M. K., eds, *Flare Stars in Star Clusters, Associations and the Solar Vicinity Vol. 137, FU-Orionis Eruptions and Early Stellar Evolution*. p. 229
- Rice T. S., Wolk S. J., Aspin C., 2012, *ApJ*, 755, 65

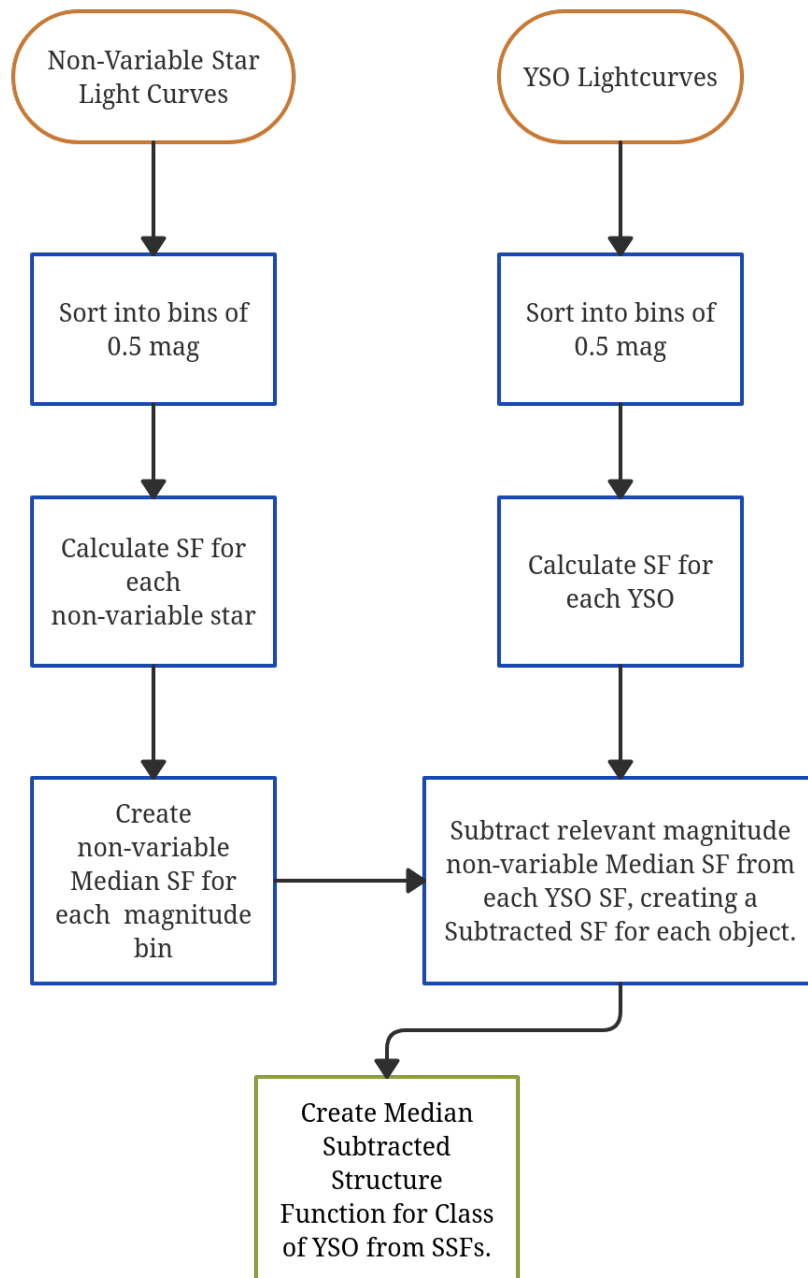
- Sacco G. G., Argiroffi C., Orlando S., Maggio A., Peres G., Reale F., 2008, *A&A*, 491, L17
- Saito R. K., Hempel M., Minniti D., Lucas P. W., Rejkuba M., Toledo I., Gonzalez O. A., Alonso-Garcia J., Irwin M. J., Gonzalez-Solares E., Hodgkin S. T., Lewis J. R., Cross N., Ivanov V. D., Kerins E., Emerson J. P., Soto M., Amores E. B., Gurovich S., Dekany I., Angeloni R., Beamin J. C., Catelan M., Padilla N., Zoccali M., Pietrukowicz P., Moni Bidin C., Mauro F., Geisler D., Folkes S. L., Sale S. E., Borissova J., Kurtev R., Ahumada A. V., 2015, *VizieR Online Data Catalog*, p. II/337
- Scholz A., Froebrich D., Wood K., 2013, *MNRAS*, 430, 2910
- Scholz A., Xu X., Jayawardhana R., Wood K., Eislöffel J., Quinn C., 2009, *MNRAS*, 398, 873
- Sergison D. J., Naylor T., Littlefair S. P., Bell C. P. M., Williams C. D. H., 2020, *MNRAS*, 491, 5035
- Venuti L., Bouvier J., Flaccomio E., Alencar S. H. P., Irwin J., Stauffer J. R., Cody A. M., Teixeira P. S., Sousa A. P., Micela G., Cuillandre J. C., Peres G., 2014, *A&A*, 570, A82
- Vorobyov E. I., Basu S., 2005, *ApJ*, 633, L137
- Wolk S. J., Rice T. S., Aspin C., 2013, *ApJ*, 773, 145

A VVV DATA SQL QUERY

The SQL query used to obtain the VVV observations used, presented to allow easier recreation of the data

```
1.  SELECT v.sourceID, d.mjd, d.aperMag5, d.aperMag5Err,
    b.flag, d.ppErrBits
2.  FROM vvvSourceXDetectionBestMatch AS b.vvvDetection
    AS d.vvvVariability AS v
3.  WHERE b.multiframeID = d.multiframeID
4.  AND b.extNum = d.extNum
5.  AND b.seqNum = d.seqNum
6.  AND d.filterID = 5
7.  AND v.sourceID = b.sourceID
8.  AND mjd > 0
9.  AND aperMag5 > 11
10. AND ppErrBits < 256
11. AND v.sourceID IN (
12. [list of all required sourceIDs],
13. [single ID per line],
14. [line ending with comma],
15. [except the last line]
16. )
17. ORDER BY b.sourceID, d.mjd
```

B MEDIAN SF FLOW CHART



A flow chart to show the steps taken to calculate the Median SF for each class, with background noise subtracted.