University of Victoria
Department of Physics and Astronomy

Analyzing the Time Variability of Deeply Embedded Protostars

Spencer Plovie

In partial fulfillment of the requirements of
the Physics and Astronomy Co-op Program

Summer 2020

Work Term #1

Protostar Analysis Intern

National Research Council of Canada – Dominion Astrophysical Observatory

Dr. Doug Johnstone, Physics and Astronomy

Contents

Abstract

The intrinsic variability of deeply embedded protostars can be analyzed with the structure function formalism. The JCMT Transient Survey sub-millimeter light curve data in eight star-forming regions of the sky was used to search for variability timescales with the structure function and other time-based methods. The Serpens Main region was the primary focus of the analysis thus far, with several protostars having significant excess power on longer timescales and prestellar cores showing little to no intrinsic variability.

I. Introduction

Star-forming regions are dense regions within a molecular cloud where localized gravitational instabilities are seeding forming stars (Stahler, 2004). In their early stages, these stellar nurseries are host to a variety of young stellar objects, the earliest stage being protostars. After a localized dense core within the molecular cloud has begun to collapse in on itself, the material falls into the centre where the protostar and protoplanetary disk begin to form. Given that the protostar is not yet hot enough for fusion to take place, the protostar continues to accrete material, primarily through the protostellar disk surrounding it. This accretion rate is expected to be unstable, making it highly variable, resulting in a detected accretion-driven variability in the light emitted from the protostar and its disk (Yoo et al., 2017).

This highly variable stage in stellar formation is what we are interested in observing. However, the surrounding natal molecular cloud core from which the protostar is forming is optically thick, providing a barrier to observations at visible and near infrared wavelengths. When light is emitted from the protostar at these wavelengths, it gets absorbed by the envelope and re-emitted at a 'redder' wavelength – hence why we probe in the sub-millimetre (Herczeg et al., 2017).

Data from eight regions was used in the analysis – IC348 and NGC1333 (star-forming region and reflection nebula, respectively; in constellation Perseus), NGC2024 (Flame Nebula; in constellation Orion), NGC2068 (reflection nebula; constellation Orion), OMC23 (star-forming region; part of the Orion Molecular Cloud Complex), Ophiuchus Core (cloud complex), and the Serpens Main and Serpens South star-forming regions. For much of the analysis performed thus far, we focused on Serpens Main, with the most significant and robust source in this region being EC 53, a well-researched and known protostar with strong intrinsic variability timescales that lie far above the expected noise levels (Yoo et al., 2018).

These variability timescales were analyzed with the structure function, a time-based method of analysis. The structure function encapsulates the degree of variability as a function of timescale (Sergison et al., 2019). In the present case, we search for stochastic variability timescales – which are found through interpretation of the structure functions created for individual sources – to help us interpret characteristics of the JCMT data set.

II. Observations and Data Reduction

Several steps for data reduction and calibration were taken with the data before it was ready to analyze. The Transient Survey data was taken with the James Clerk Maxwell Telescope (JCMT) using the SCUBA-2 camera at a wavelength of 850 μm ('microns') and 450 microns at an approximately monthly cadence – times between observations varied from 14 days to upwards of 2 months for a given region. Within the camera, there are 4 sub-arrays, each consisting of 40x32 bolometers for each of the 450 and 850 micron wavelengths (Mairs et al., 2020). For the measurement of a given bolometer, there are several components that contribute to the overall signal – astronomical signal, atmospheric extinction correction (correcting for the signal that does not reach the telescope through the atmosphere), and several sources of noise (white noise, common mode noise, excess noise at low frequencies). From here, the data was reduced by applying five main models that separated noise from astronomical signal. This included a general cleanup, flat-fielding the data, removing high-frequency spikes and cosmic rays, and smoothing out the beginning and end of the time series (Mairs et al., 2020).

The 850 micron reduced and calibrated data were uploaded to the CFHT VOSpace, a storage space on CANFAR (Canadian Advanced Network for Astronomical Research). For a given region, metadata, containing the details of each observed epoch, and source info, containing the measured brightness of all sources, was downloaded and further processed in a Python 3.7 Jupyter Notebook before being used in the analysis. The fluxes were stored in a two-dimensional array, a set of fluxes for each epoch for each source. Each flux value had a corresponding time, which had been converted from Julian dates to a zero-reference point. The number of time pairs were found by taking the difference in time between all possible pairings – obtaining these values was a crucial step in the procedure and is discussed in the *Analysis and Results* section. The flux differences across these times were also obtained, used for a different method of analysis. Samples of code illustrating this basic data reduction can be found in Appendix A.

Another part of the reduction process was categorizing the sources as bright protostars, prestellar cores, and accretion disk sources. 'Bright' was defined as a source with a maximum flux of at least 150 mJy. From here, the bright sources were categorized by their distance, in arcseconds, from the nearest identified protostar or disk source. If the source's angular distance to the nearest protostar was both less than 10" and less than the distance to the nearest disk, it was classified as a protostar. If the latter condition was not met, and the angular distance to the nearest disk was less than the distance to the nearest protostar, it was classified as a disk source – and if neither condition was met, it was considered a prestellar source.

III. Analysis and Results

The structure function plot is a time-based method of analysis used to search for intrinsic variability on different timescales for a given source. The most important aspect of the structure function is that it considers all possible time pairs, then takes the difference in the corresponding fluxes in the light curve. Upper- and lower-time bounds are provided for the pairs to fall within distinct, logarithmically spaced bins. A value for the structure function is then calculated as:

$$(Eq.1) \quad SF(\tau_1, \tau_2)_{fid} = \frac{1}{p(\tau_1, \tau_2)} \sum \frac{(F_i - F_j)^2}{SD_{fid}}$$

where (τ_1, τ_2) are the time bounds and *p* is the number of pairs found within the time bounds. As seen in the equation above, a flux pair within the time constraint of a particular bin is squared and summed up with all other squared pairs (Sergison et al., 2019). Instead of dividing by a median flux value, the fiducial structure function has been divided by a standard noise model, SD_{fid}. This provides a baseline where we would expect the structure function to lie if it were just noise with no significant intrinsic variability timescales. For a given source, i, the following formula was used to calculate the fiducial noise:

$$(Eq.2) \quad SD_{fid}(i) = \left[(0.014)^2 + \left(0.02 \times f_m(i)\right)^2\right]^{\frac{1}{2}} \text{ Jy}$$

where f_m is the mean peak brightness, and 0.014 Jy represents the estimated uniform noise in faint sources (Johnstone et al., 2018).

The structure function is cumulative in nature – the power on longer timescales should almost always be greater than it is on shorter timescales. In general, most of the plot has a certain positive slope, then flattens out after the timescales of intrinsic variability of the source have been surpassed. For an ideal example of how the structure function works and how it can be interpreted, please refer to Appendix B.

The histogram showing the frequency of the timescale pairs for a given region plays an important role in understanding the corresponding structure function. Figure C1 shows the distribution of timescale pairs for the Serpens Main region. The bins are chosen to be 1.5 times larger than the previous, thus encompassing a broader range for each successive bin. The distribution of these pairs depends on how the data was taken in terms of its cadence and number of time pairs – for a given region, there are 40-55 data points to create pairs with, using the following formula:

$$(Eq.3) \quad pairs = \frac{N(N-1)}{2}$$

where *N* is the number of epochs observed for that region. Due to the nature of the structure function, gaps in the data collection have little effect on the resulting plot, so long as there are enough time pairs within a given bin. This is important to recognize for bins with very few data points – the corresponding points on the structure function plot are, as a result, 'too noisy' to be considered reliable, and are neglected in the analysis. This can be seen with the vertical red dashed lines in Figure C1, which will appear in structure function plots discussed later in this section.

An example of a structure function plot for a given source can be seen in Figure C2 on the left, where EC 53's fiducial structure function is shown in black. This source sets an example within our data set for what we would expect to see in a protostar with strong variability timescales. The power increases until a peak five times greater than the expected noise is observed at the 1-year mark, followed by a characteristic dip at the 1.5-year mark – this dip can be attributed to the well-known period of EC 53.

The blue line in this plot was calculated by randomizing the light curve data 100 times for each source and re-creating the structure function for each randomization, then taking the mean value of

these randomizations. Given that all timescales are treated equally upon randomization, this removes any time dependence in the power. The error bars seen here represent the standard deviation of the 'spread' in randomized values. Note that values lying outside of the vertical red dashed lines for both the fiducial structure function and its mean randomization are not considered when interpreting the plot. The large error bars on these shortest and longest timescales indicate very few timescale pairs contributing to these values.

An example of a 'flattened' structure function can be seen in Figure C2 on the right. This plot was obtained by first fitting a linear regression to the light curve of the source and subtracting it out of the data – this removed any linear secular variability present in a source, allowing for investigation of other intrinsic variability timescales. While other sources in the Serpens Main region experienced significant changes to their corresponding structure functions after being 'flattened', EC 53 maintains a very similar shape to the plot on the left in the regions of the plot we are interested in – this is due to EC 53's period being shorter than the transient survey's observing window, resulting in no long-term linear trend being subtracted out.

Another important part of the analysis involved inspecting multiple sources at once for a given classification. Figure C3 shows the structure functions for the six brightest protostars in Serpens Main. Several sources have power increasing significantly past the baseline: in particular, source0, source2 (EC 53), and source3. For sources 0 and 3, we see a distinct change after the linear secular variability has been removed on the right panel. Since they still lie far above the fiducial expectation, along with EC 53, these structure functions show evidence of strong intrinsic, short-timescale, variability. Sources that stay within the horizontal red dashed lines do not show strong variability on any timescales. Figure C4 shows something similar, but for the six brightest prestellar cores in this region. Given that these sources do not have a variable source in their centre yet, the structure function should reflect this by staying close to the fiducial noise levels. This can be seen with all six sources staying primarily within the horizontal red dashed lines. It is worth noting that there is evidence of more variability on long timescales, which could suggest issues with the long-term calibration of the data.

Another method of analysis currently being developed is using contour plots to search for these intrinsic variability timescales. Figures C5 and C6 show contour plots for the bright protostars and bright prestellar cores (respectively) in the Serpens Main region. The plots represent, at a given time lag, the fraction of the total number of sources for a given type that produce a particular deviation value. For the deviation from the expected (plots on the left), the following formula was used:

$$(Eq.4) \quad fraction_{fid} = \frac{SF_{rand} - SF_{fid}}{SF_{fid}}$$

where SF_{rand} is the mean random value seen as the blue line in Figure C2 and SF_{fid} is the fiducial structure function value (as seen in plots C2-C4). This generated a value on the y-axis with a range [-1,∞). For the deviation from the standard error (plots on the right), the following formula was used:

$$(Eq.5) \quad fraction_{sig} = \frac{SF_{rand} - SF_{fid}}{\sigma_{fid}}$$

Where σ_{fid} is the standard deviation error, seen with the blue error bars in Figure C2. This generated a value on the y-axis with a range (-∞,∞). Both types of plots are still under interpretation and are being

compared with an ideal toy model to help understand the shapes of the contours. These plots will not be included in the Discussion section of this report.

IV. Discussion

The structure function plots created for this analysis have been shown to be difficult to interpret for individual sources or for sources where the characteristics are not well-understood beforehand. EC 53 is a robust source in this region and its corresponding structure function confirmed its strong variability timescales. The plots created for the prestellar cores were also able to reaffirm our understanding that without a source to create stochastic variability timescales, the resulting structure function does not show significant power increase on longer timescales when compared with an expected fiducial noise as a reference point.

One thing we hope to find through the structure function analysis is a timescale for the envelope surrounding the protostar and its accretion disk to respond to variability. For example, when a bright flash occurs, we only see it after the envelope has re-emitted this flash at a longer wavelength. The amount of time it takes for the envelope to 'respond' to these changes in brightness should be affecting the timescales we are able to see and would provide an important and unique physical constraint on these systems. If the envelope takes time to respond to intrinsic short timescale variations due to the physics of accretion onto the protostar, these intrinsic variations will be 'smoothed out'. We hope to recover these envelope timescales through further analysis of Serpens Main and the other seven regions in the survey.

V. Conclusions

The intrinsic variability timescales for protostars, disk sources, and prestellar cores were searched for and analyzed in the Serpens Main star-forming region with the structure function, a time-based method of analysis. The structure functions for several protostars in this region showed strong variability timescales compared to their expected noise levels, particularly EC 53, a protostar with well-known variability timescales. The structure functions for several bright prestellar cores showed little to no power over longer timescales and stayed close to their expected noise levels, indicating there were no intrinsic variability timescales, which was expected for these types of sources. Individual sources will continue to be interpreted using the structure function and the contour plots discussed to search for other notable characteristics in the data set.

VI. Acknowledgements

I would like to give a special thanks to my research supervisor, Doug Johnstone, for his guidance and support throughout my co-op work term. I am beyond grateful for his endless patience in helping me fix my coding errors, answering all my questions, and keeping me connected with other researchers in the field of study. He genuinely cares about his research students and it shows through the effort he puts in to help them succeed. I would also like to thank Logan Francis and Colton Broughton for their help in

answering my Python-related questions. Both were always more than happy to provide suggestions and resources to improve my code.

VII. References

Herczeg, Gregory J. et al. *How Do Stars Gain Their Mass? A JCMT/SCUBA-2 Transient Survey of Protostars in Nearby Star-forming Regions*. The Astrophysical Journal, Volume 849, Issue 1, article id. 43, 14 pp. (2017). Crossref. Web.

Historical Weather Data, Government of Canada. Accessed June 2020: https://climate.weather.gc.ca/historical_data/search_historic_data_e.html

Johnstone, Doug et al. *The JCMT Transient Survey: Stochastic and Secular Variability of Protostars and Disks In the Submillimeter Region Observed over 18 Months.* The Astrophysical Journal 854.1 (2018): 31. Crossref. Web.

Mairs, Steve. *Data Reduction Calibration Information*, May 27th, 2020. Zoom-recorded team meeting.

Sergison, Darryl J et al. *Characterizing the i-Band Variability of YSOs over Six Orders of Magnitude in Time-Scale.* Monthly Notices of the Royal Astronomical Society 491.4 (2019): 5035–5055. Crossref. Web.

Stahler, S. W. & Palla, F. (2004). *The Formation of Stars.* Weinheim: Wiley-VCH. ISBN 3-527-40559-3.

Yoo, Hyunju et al. *The JCMT Transient Survey: Detection of Submillimeter Variability in a Class I Protostar EC 53 in Serpens Main*. The Astrophysical Journal, Volume 849, Issue 1, article id. 69, 10 pp. (2017). Crossref. Web.

VIII. Appendices

Appendix A: Data reduction

```
## Grabbing the source flux values from the 'source_info'.txt file (data1)##
## Grabbing the Julian dates from the 'metadata'.txt file (data2)##

data1.readline() #Reads past titles
sources = data1.readlines()
#Note that every source is ONE LINE. Another .readlines() will just start skipping sources
#Data is NOT ready to use yet, needs each token split up into variables in an array or list

data2.readline() #Reads past titles
sources_metadata = data2.readlines()

sources_list = [] #Creating a list to gather ALL data (for now)
metadata_list  = []

for line in sources:
    sources_list.append(line.split()) #Splits each line into a variable in a list

for line in sources_metadata:
    metadata_list.append(line.split())

sources_array = np.array(sources_list) #Convert to array
```

```
metadata_array = np.array(metadata_list)

num_fluxes = int(len(metadata_array)) #Number of flux values for each source
num_sources = int(len(sources_array)) #Number of sources

flux = np.zeros((num_sources,num_fluxes))
#Creating a 2D array of floats for flux data from each source (source zero = brightest source)

#Creating a variable for all of the Julian dates
time = np.zeros(num_fluxes)

for i in range(num_sources):
    for j in range(num_fluxes):
        flux[i][j] = sources_array[i][j+15] #Appending all flux values for each source
        #Assuming the fluxes start at the 15th index in the source_info file

for i in range(num_fluxes):
    time[i] = metadata_list[i][4] #Grabbing all the Julian times

start = time[0]
for i in range(len(time)):
    time[i] = time[i] - start #Starting the time at 0 days

time_years = np.zeros(len(time))
for i in range(len(time)):
    time_years[i] = time[i]/365 #Converting from days to years

data1.close()
data2.close()
```

Appendix B: Victoria Weather Data

Five years worth of weather data for the Victoria Harbour region were obtained from the Government of Canada Historical Climate Data website (Historical Weather Data, 2020). With 1800 data points for the mean daily temperature, a regular cadence, and well-recognized characteristics, this data was used as an ideal example of a structure function. Referring to Figure A1, we see regular variability with a sinusoidal shape observed on longer timescales, and the greatest temperature variation of up to 20˚C on ½-year timescales. This translates as the peak in the structure function seen in Figure A2, followed by a characteristic dip at approximately the 1-year mark. This dip is due to an aliasing effect, where the frequency at which we measure is on the same timescale as the frequency of the source we are observing. On longer timescales, the structure function flattens out – this is past the intrinsic variability timescales of the source (one can imagine that if the data continued on forever and we 'zoomed out' on Figure A1, the structure would be too small to resolve and we would see a flat line). A particularly interesting characteristic of the weather data's structure function is that on short timescales it appears to be steeper than expected, with a notable change in concavity seen around the 1-month mark. This implies that we are observing more variability than expected from a one-year sinusoid alone, likely due to weather patterns (which add to the 'noisy' look of the sinusoid in Figure A1) – demonstrating multiple timescales at play in the data contributing to the overall shape of the corresponding structure function.

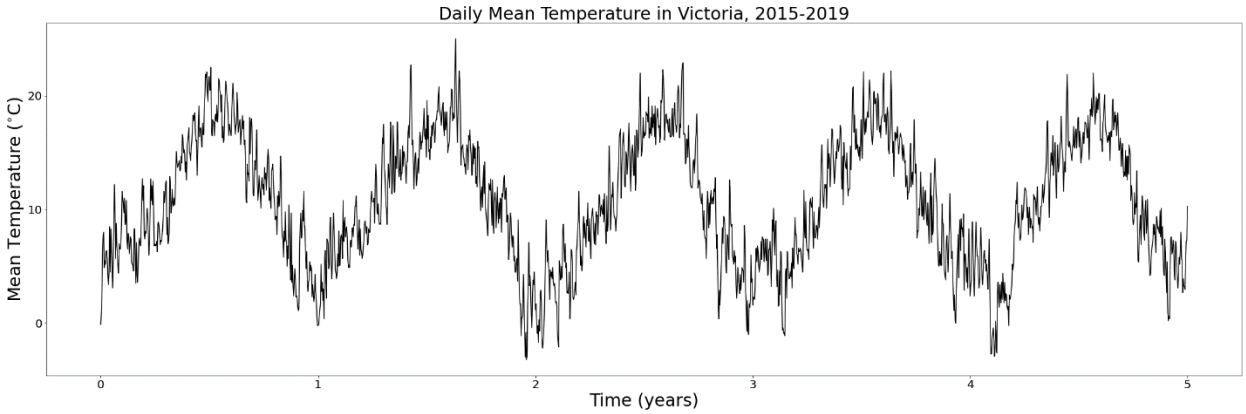

Figure B1. Change in temperature over the 5-year period of data collected for the Victoria Harbour weather station.

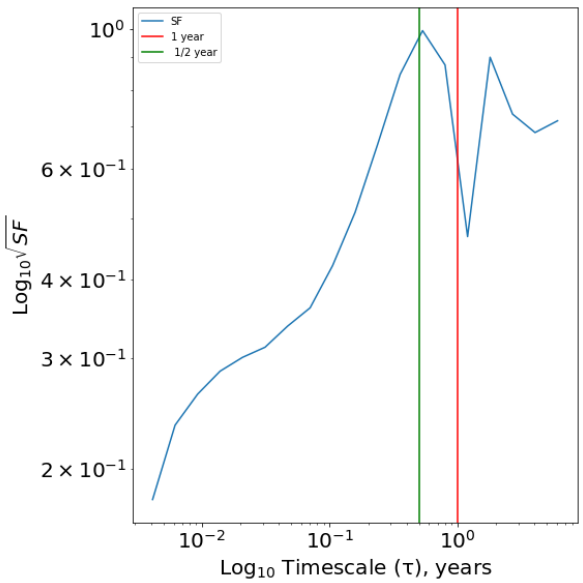

Figure B2. The resulting structure function plot for the Victoria weather data (see text for detail).

Appendix C: Results

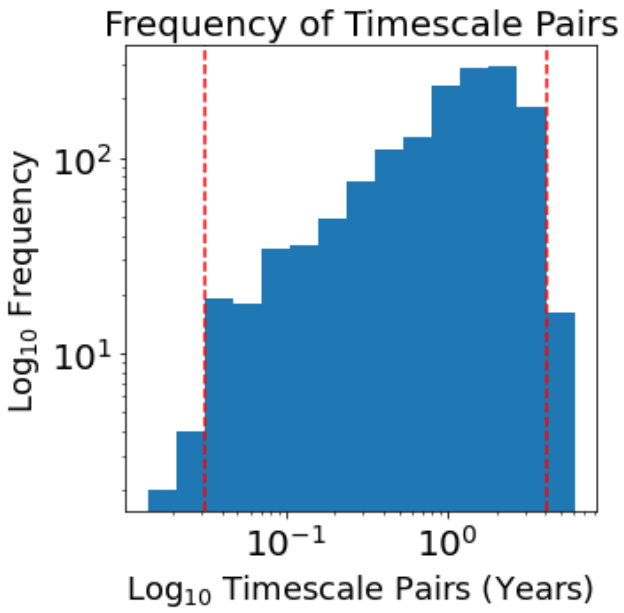

Figure C1: Histogram plot illustrating the frequency distribution of timescale pairs.

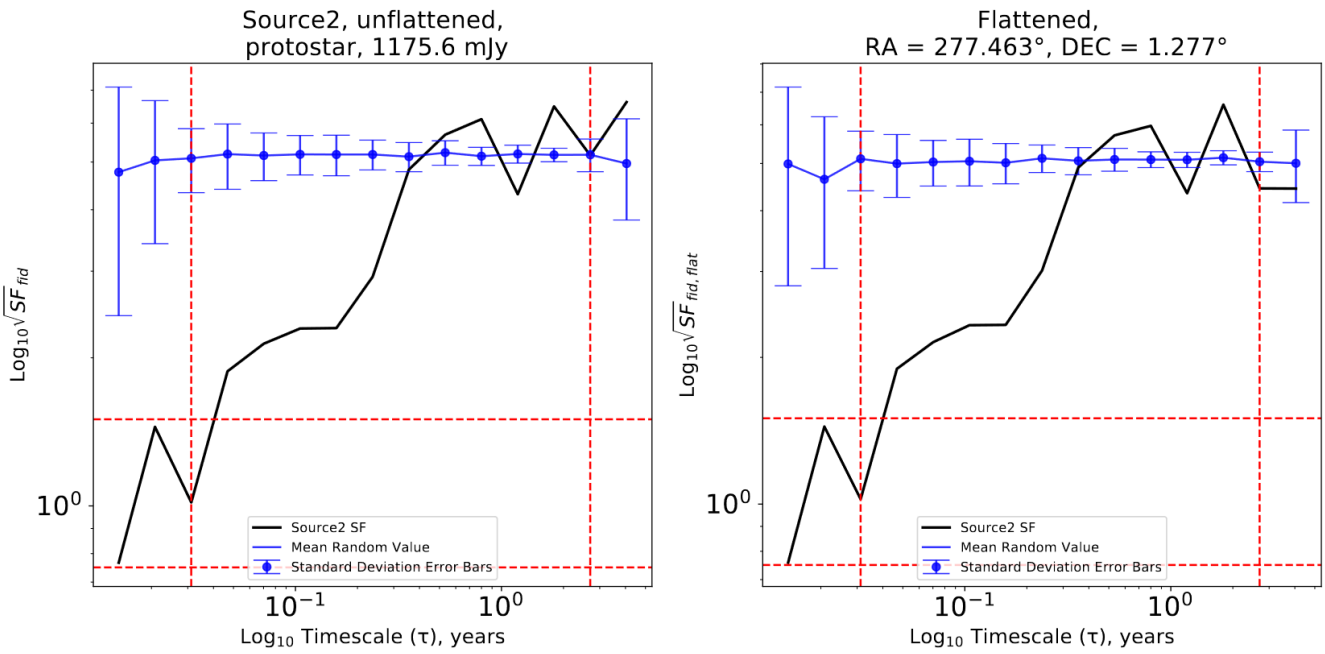

Figure C2. The structure function plot for source 2, commonly known as EC 53, in Serpens Main (black) and its mean randomization value with error bars (blue). The mean peak brightness (1175.6 mJy) and the right ascension and declination coordinates (RA=277.463˚, DEC=1.277˚) have been provided for this source in the titles.

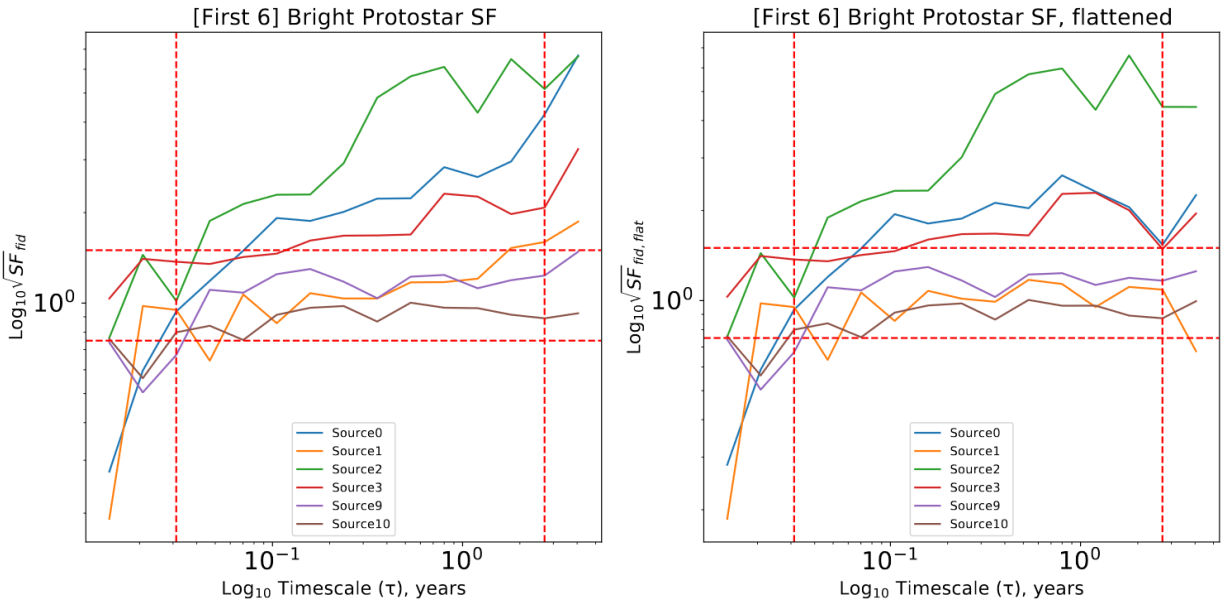

Figure C3. Fiducial structure functions for the six brightest protostars in the Serpens Main region.

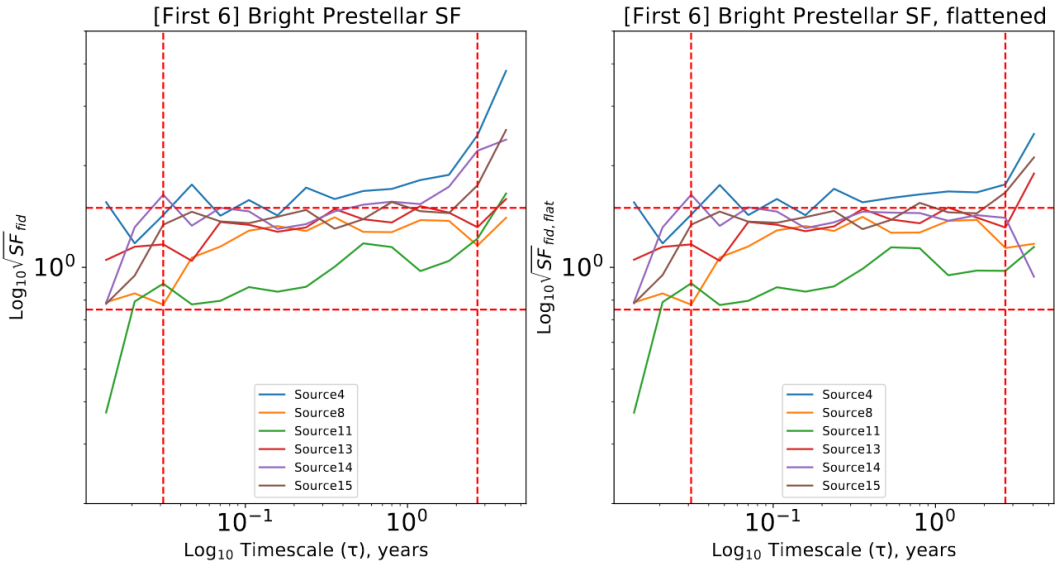

Figure C4. Fiducial structure functions for the six brightest prestellar cores in the Serpens Main region.

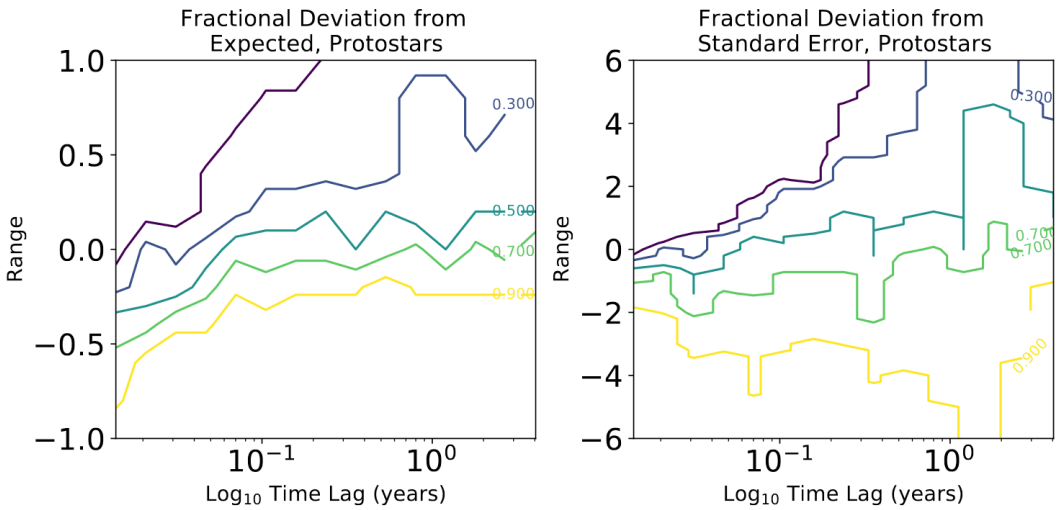

Figure C5. Contour plots for the bright protostars in Serpens Main. On the left, the fractional deviation from the mean structure function value compared to the fiducial. On the right, the fractional deviation from the mean structure function value compared to the fiducial but normalized by the standard error bars.

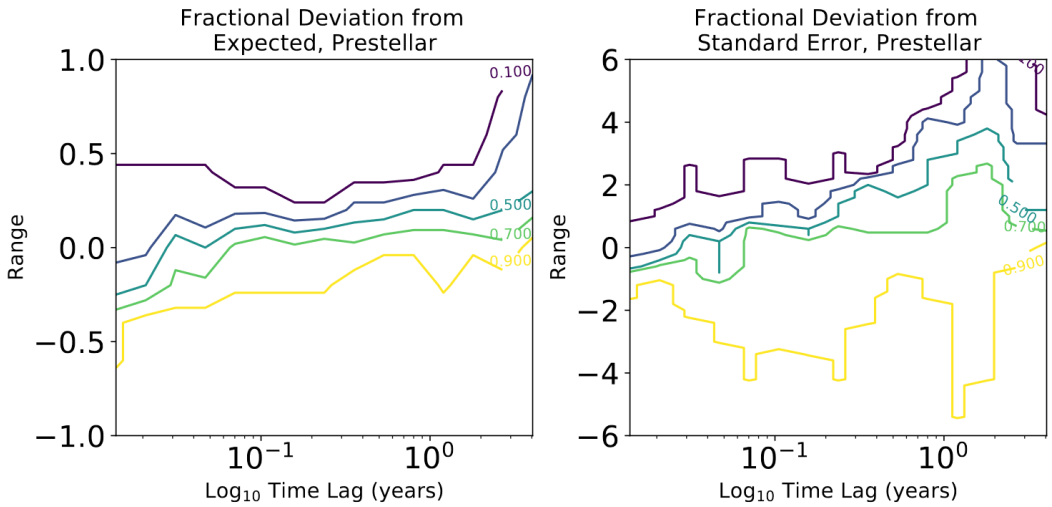

Figure C6. Contour plots for the bright prestellar cores in Serpens Main. On the left, the fractional deviation from the mean structure function value compared to the fiducial. On the right, the fractional deviation from the mean structure function value compared to the fiducial but normalized by the standard error bars.