Application of Machine Learning Techniques To Young Stellar Object Classification

by

Breanna Lydia Crompvoets
B.Sc., University of Regina, 2021

A Thesis Submitted in Partial Fulfillment of the
Requirements for the Degree of

MASTER OF SCIENCE

in the Department of Physics and Astronomy

University of Victoria

Application of Machine Learning Techniques To Young Stellar Object Classification

by

Breanna Lydia Crompvoets
B.Sc., University of Regina, 2021

Supervisory Committee

Dr. J. Di Francesco, Co-Supervisor
(Department of Physics and Astronomy)

Dr. J. Willis, Co-Supervisor
(Department of Physics and Astronomy)

Abstract

Among the first observations released to the public from the James Webb Space Telescope (JWST) was a section of the star-forming region NGC 3324 known colloquially as the "Cosmic Cliffs." We build a photometric catalog of the region and analyze these data using the Probabilistic Random Forest machine learning method. We find 496 YSOs out of 19 497 total objects within the field, 474 of which have not been found in previous works. Using the obtained probabilities of objects being YSOs, we employ a Monte Carlo approach to determine a local star formation rate of 1×10^{-4} $M_\odot$/yr, for the region. We also find that the surface density of YSOs in the Cosmic Cliffs is largely coincident with column densities derived from Herschel data, up to a column density of 1.37×10^{22} cm^{-2}. The newly determined number and spatial distribution of YSOs in the Cosmic Cliffs demonstrate that JWST is far more capable of detecting YSOs in dusty regions than Spitzer.

Table of Contents

List of Tables

List of Figures

Acknowledgements

First and foremost I would like to thank **Dr. James Di Francesco**. His patience, kindness, and encouragement have made this journey so much lighter. His breadth of knowledge and excitement to share it have helped me enormously throughout these two years, and I am truly grateful for his mentorship.

I would also like to thank two of the members of HAA for their help. Firstly, thank you to **Dr. Hossen Teimoorinia** for fascinating discussions on the possibilities of machine learning, and how we can push their applications. His guidance and discussion has helped to make this idea a reality. Secondly, thank you to **Dr. Peter Stetson** for the in-depth explanations on photometry, and how powerful the methods can be even on their own. His knowledge and support greatly changed the scope of this project.

Finally, I would also like to thank my office-mates and roommates for all of their support and feedback throughout this degree. From teaching me what a CMD was, to discussions on the validity of my method and statistics, they have acted as sounding board and friends.

I acknowledge and respect the Lekwungen-speaking peoples on whose traditional territories
I work and live, and the Songhees, Esquimalt and WSANEC peoples whose historical
relationships with the land continue to this day.

Dedication

Dedicated to my grandfather Cornelius Koot, who was excited to have a second master in the family.

Chapter 1

Introduction

Young Stellar Objects (or YSOs) are centralized gas condensations of stellar mass and size that may have circumstellar dust envelopes and disks. The central objects, i.e., the protostars, have not yet ignited hydrogen fusion in their cores and entered the main sequence of stellar evolution. With them, we can study processes of star formation and, by extension, planet formation. To probe these processes, we must begin with an accurate identification of YSOs at different stages of their evolution, in a variety of environments.

This work is concerned with the rapid and robust identification of YSOs using data from the James Webb Space Telescope through the use of machine learning. It consists of a submitted paper, Chapter 2, bookended by a more detailed introduction than provided in the paper, this Chapter, and an explanation of what future work is possible now that this study is completed, Chapter 3. As a proof-of-concept, we utilize the machine learning Probabilistic Random Forest method (Reis et al., 2019) to identify YSOs in NGC 3324, a star-forming region observed early on by JWST (Pontoppidan et al., 2022). We further analyze its star formation rates and efficiency based on the available data, and compare these to an analysis of other regions in Pokhrel et al. (2021).

This Chapter provides an introduction to the concepts relevant for the main body of this work. In Section 1.1, we summarize the different stages of star formation along with their importance for determining accurate statistics for star formation rate calculations. Section 1.2 describes two infra-red space telescopes and their respective abilities to identify the different stages of star formation. Lastly, section 1.3 describes the machine learning tools available to us and why we chose one in particular to implement.

1.1 Star Formation

Star formation is a complex and highly dynamic process, with many ongoing avenues of research. For this thesis, we focus on the identification of YSOs in star-forming regions. Figure 1.1 shows the progression of a low-mass YSO through the stages of its evolution towards the main sequence in relation to both its spectral energy distribution (SED) and the physical attributes it possesses at each stage. In general, the central object grows in mass while the mass surrounding it, in the form of a surrounding disk or envelope, is accreted or dispersed. Observationally, the YSO proceeds from being a cold modified black body visible in the sub-mm to a stellar black body with some possible disk emission in excess of that from the central object visible in the optical and infra-red (IR) over 10^7 years or less.

There are three main phases of low-mass star formation, beginning with the prestellar phase. At this point, there is no compact central object, only a dense core of gas within a parental molecular cloud. After cores have condensed enough, they collapse, forming a protostar surrounded by a flat disk and a larger spheroidal envelope of gas; this phase is known as the protostallar phase. Once the envelope is accreted or dispersed, the pre-main sequence phase begins, and is marked by the accretion or dispersal of the disk. These latter two phases can each be further split into two distinct stages of YSO classification: Class 0+Class I and Class II+Class III, respectively. Each stage is summarized below, along with the approximate length of time each lasts for a low-mass star, i.e., one with mass < 8 $M_\odot$. It is important to note that these lifetimes are statistically determined, and should be treated as half-lives.

The first stage of YSO evolution is represented by Class 0 YSOs. Here the central object, or "protostar", is actively accreting gas and dust, and its exterior envelope is as massive or more massive than the protostar within it, $M_* \leq M_{\mathrm{envelope}}$ (Barsony, 1994; André, 2015; Ward-Thompson & Whitworth, 2011). Due to this stage of YSO being deeply embedded, the associated high dust extinction makes them not visible in the optical or near-IR (NIR). These objects are often traceable by their outflows and shock emission (Barsony, 1994; Reiter et al., 2022) in the mid-IR (MIR), or identifiable by their sharp peaks in the sub-mm with no optical/IR counterparts. This stage only lasts for $\sim 1 - 2 \times 10^5$ years (Dunham et al., 2015). In addition to being difficult to detect even in the MIR, the short lifespan leads to fewer Class 0 YSOs, by relative abundance to older classes, in clouds with a constant star formation rate.

The next stage of YSO evolution is represented by Class I YSOs. While such a YSO is still actively accreting gas and dust, its envelope has been diminished, leading to a protostar that

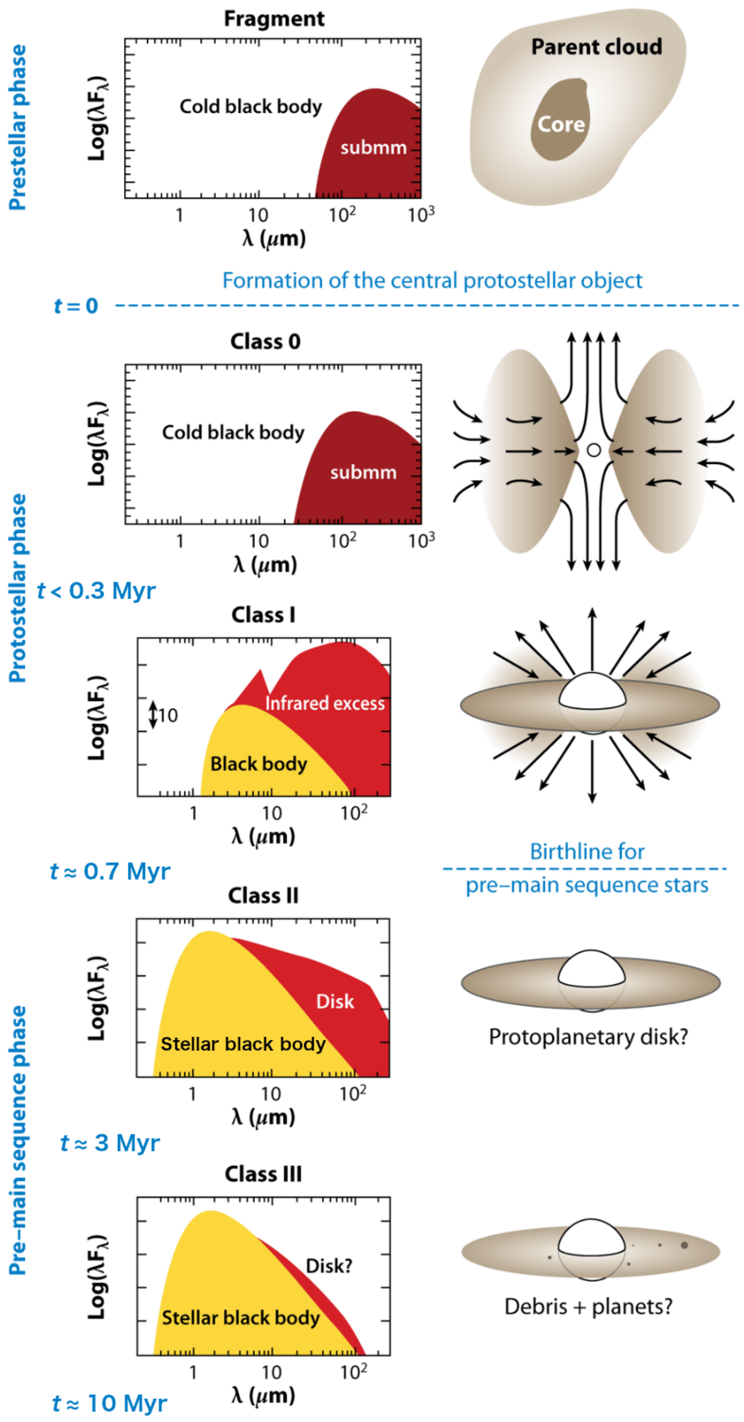

Figure 1.1: The different stages of low-mass star formation. Each stage is labelled above its SED, and a cartoon showing the envelope, disk, and inflows/outflows is next to it. The approximate time since the formation of the central protostellar object is written on the left, beneath each subsequent stage. Figure from Dauphas & Chaussidon (2011).

is more massive than the envelope, $M_* > M_{envelope}$. Because it is not so deeply embedded, the Class I YSO may be visible in the NIR, but with significant MIR and far-IR (FIR) excess. Although YSOs in this stage still possess in-falling envelopes, the subsequent shocks and outflows associated with these are much less extreme than those with Class 0 objects, and are often far less collimated as well (André, 2015). This stage of star formation lasts $\sim 3-5 \times 10^5$ years (or about twice as long as for Class 0 YSOs; Dunham et al., 2015), and, thus, they generally are more plentiful than Class 0 YSOs. As such, these YSOs are often used when deriving the star formation rate and star-forming efficiency from IR surveys, (e.g., Pokhrel et al., 2021).

Next, another stage sometimes mentioned is the Flat-Spectrum YSO stage, which bridges the Class I and Class II YSO stages. These have only remnant envelopes and mark the transition from the protostellar phase to the pre-main sequence phase.

The third stage consists of Class II YSOs. Here the YSO is accreting solely through its disk and has no circumstellar envelope. These stars are also often referred to as Classical T-Tauri stars, and can be identified by the modest NIR excess and strong emission lines in their spectra (Ward-Thompson & Whitworth, 2011), the latter caused by hot gas accreting onto the central protostar. The dense gas and dust in the disk makes these stars ideal candidates for planet formation studies. This stage lasts for about 2×10^6 years (Evans et al., 2009), and, hence, are even more plentiful than Class I YSOs.

Class III YSOs represent the final stage of star formation before the main sequence. These objects have now largely dissipated their disks, are contracting as they near the point of hydrogen ignition in their cores, and their SEDs are now largely that of a reddened blackbody (Adams et al., 1987). They are also called Weak-line T-Tauri stars, as they have much weaker emission lines than Class II YSOs (Ward-Thompson & Whitworth, 2011). Lastly, this stage of YSO is the most difficult to detect in the IR, with minimal excesses and signatures similar to both regular field stars and AGB objects. Thus, IR surveys do not tend to characterize these objects well (e.g., Dunham et al., 2015).

Identification of YSOs is important for the determination of star formation rate (SFR; the mass of stars formed per year) and efficiency (SFE; the amount of gas transformed into stars compared to the total amount of gas available). Indeed, the formation rates of stars can be related to the gas density within their natal clouds. This idea was first proposed by Schmidt (1959), who postulated that the SFR surface density (Σ_{SFR}) must be proportional to the gas surface density (Σ_{gas}) to some power, Equation 1.1.

$$\Sigma_{SFR} \propto (\Sigma_{gas})^n \tag{1.1}$$

Kennicutt (1998) improved upon this basic power law by specifying that the power had to be 1.4 ± 0.15 in nearby galaxies, using surveys of high-mass YSOs and both HI and H_2 measurements for the total gas. More recent works have sought to better define this relation in the Galactic context, and have found a wide variety of powers from different dense gas tracers. In particular, several works have found a more linear relation (e.g., Gao & Solomon, 2004; Wu et al., 2005; Evans et al., 2014a; Lada et al., 2010). Theoretical works have also postulated the need for free-fall time to be included in this relationship (e.g., Krumholz & McKee, 2005; Padoan et al., 2012; Krumholz et al., 2019).

Pokhrel et al. (2021) recently produced a new study of the Kennicutt-Schmidt relation, where they analyzed the SFR surface density to gas column density in 12 nearby molecular clouds based on data from the Spitzer Space Telescope. For their analysis, they used Class 0/I YSO surface densities, which are more reflective of the current gas density production capacity, i.e., within the last 0.5 Myr. From this, they found a slope of n=2 for Equation 1.1. Furthermore, they normalized the gas column density by the mean local free fall time and constrained the slope of the new relation to be 1. Hence, any differences in the relationships are likely due to differing SFE in the individual clouds, Equation 1.2. This slope of 1 is further supported as the mean of the slopes across all 12 clouds is 0.94 and the median is 0.99, with a standard deviation of 0.11.

$$\log(\Sigma_{\mathrm{SFR}}) \propto \log(\Sigma_{\mathrm{gas}}/t_{\mathrm{ff}}) + \log(\varepsilon_{\mathrm{ff}}) \tag{1.2}$$

By doing such, they were were able to find a clear relation with a smaller standard deviation in Σ_{SFR} than from Equation 1.1, see Figure 1.2 for a comparison of the two relations.

These relations fit well up to a column density of 1.4×10^{22} cm^{-2}, after which the SFE drops off with different rates in different clouds. This drop-off may be caused by an increase of turbulence due to feedback in the dense and clustered locations found predominantly at the highest column densities in molecular clouds. Conversely, it may instead be an artifact of Spitzer's limitations in probing dense and clustered locations. This latter caveat is explored in the next Section, § 1.2.

1.2 YSO Identification

The infrared is a thriving ground for the identification of YSOs with space telescopes, as evidenced in Figure 1.1. Indeed, the NIR (0.78 - 2.5 μm) is even useful for the separation of Class 0 YSOs from older sources, as members of that class will not be detectable in this range. The MIR (2.5 - 25 μm) covers the majority of the emission from Class I, II, and III

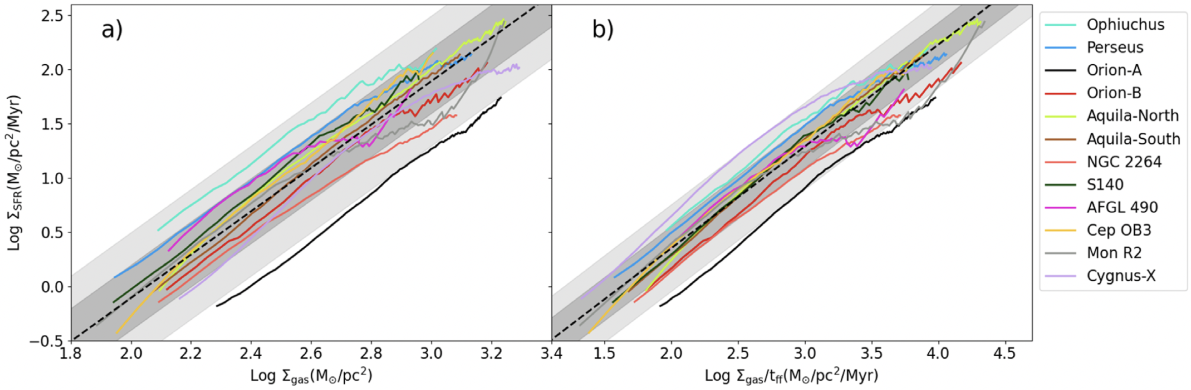

Figure 1.2: Star formation relation fits for a number of different nearby clouds, with (a) comparison to both gas column density only; and (b) gas column density normalized by free fall time. Figure from Pokhrel et al. (2021).

sources, and possibly the shortest wavelength emission from Class 0 sources. Finally, the far-IR (FIR; 25 - 1000 µm) is useful for the identification of all YSO stages, and in particular Class 0. In this work we will focus on two telescopes which operate in the NIR and MIR: the Spitzer Space Telescope, and the James Webb Space Telescope (henceforth referred to as JWST).

The Spitzer Space Telescope operated over 2003-2020, with two cameras along with an Infra-Red Spectrometer (IRS). The first camera was the InfraRed Array Camera (IRAC) which operated in four bands: 3.6 µm, 4.5 µm, 5.8 µm, and 8.0 µm. The second was the Multiband Imaging Photometer for Spitzer (MIPS) which operated in three bands: 24 µm, 70 µm, and 160 µm. Because of this wide wavelength range and its large area mapping capabilities, Spitzer was used for several surveys of star-forming regions which have contributed much to our understanding of star formation. Unfortunately, the 160 µm data were not used extensively due to saturation and coverage limitations, and the large beam.

For instance, the Spitzer Cores to Disks, or c2d, survey (Evans et al., 2009, 2014b) imaged five large nearby (within 300 pc) clouds and identified 1024 YSOs, and the Gould Belt Survey (Dunham et al., 2015) imaged 23 further clouds in the Gould Belt and identified 2966 YSO. These YSOs were identified based on their colours between the four Spitzer IRAC bands, the MIPS 24 and 70 µm bands, and 2MASS data, when available. 2MASS (Skrutskie et al., 2006) is a ground-based all-sky survey that sampled short IR wavelengths, between 1.2 µm and 2.2 µm.

YSOs can be identified by their red-excess, caused by the extended emission of dusty envelopes and disks in addition to their modified black body spectra, as in Figure 1.1. Thus cuts in colours and magnitudes can effectively separate YSOs with significant dust in their envelopes or disks from the majority of the evolutionary sequence, with the caveat of losing some of the more evolved YSO population. YSOs, however, are not the only dusty objects in the sky. Distant reddened galaxies and evolved red stars are known contaminants of these criteria.

The YSO-identifying colour criteria themselves, as used by Allen et al. (2004); Gutermuth et al. (2008), etc., include 4-6 bands across the Spitzer and 2MASS datasets. For example, one series of colour cuts used to distinguish YSOs from sources dominated by shock emission must satisfy all three of these criteria:

- $[3.6] - [4.5] > \frac{1.2}{0.55}([[4.5] - [5.8]] - 0.3) + 0.8$
- $[4.5] - [5.8] \leq 0.85$
- $[3.6] - [4.5] > 1.05$

Each contaminant source and YSO stage have a different set of criteria that must be satisfied, and a specific order must be taken to remove one set of contaminants at a time. This method is effective for separating YSOs from many of the contaminant objects, at the risk of losing some real YSOs. A major caveat is that these criteria require that objects must have data in all the relevant bands to be classified. This caveat will be returned to in Section 1.3.

Spitzer had two mission phases during its lifetime: the first, cold, mission ran over 2003-2009, after which the liquid helium coolant ran out. The second, warm, mission, running from 2009 until decommissioning in 2020, had only the 3.6 µm and 4.5 µm bands available, as the instruments were then too warm to obtain sensitive data at longer IR wavelengths. Spitzer was succeeded by JWST, which began operation in 2022. JWST also operates two cameras: the Near-InfraRed Camera (NIRCam) with 29 filters between 0.7 µm and 4.8 µm; and the Mid-InfraRed Imager (MIRI) with 9 filters between 5.6 µm and 25.5 µm. Not only does JWST have several more filters available than Spitzer, it also surpasses the sensitivity of Spitzer by two orders of magnitude (Rigby et al., 2023), with seven times the resolution in wavelengths in common. The sensitivity and resolution performance improvements allow JWST to probe sources to much lower luminosities, and resolve many of the previously identified Spitzer sources into several individual objects (e.g. Lenkić et al., 2023).

The 3.6 µm and 4.5 µm filters from Spitzer, IRAC1 and IRAC2, have counterparts in JWST, F356W and F444W, which have very similar transmission curves, as seen in Figure 1.3. These matched filters are useful when comparing Spitzer sources to JWST sources. Indeed, in Chapter 2, we compare the 4.5 µm Spitzer flux and the 4.44 µm JWST flux to check that we are indeed measuring the same sources, and find a 1:1 correlation between the two.

JWST is not relegated to using the exact same filters as Spitzer. In Chapter 2, we analyze a region imaged with the Early Release Observations, which included five additional NIRCam bands without direct analogy to Spitzer's bands. Hence, we cannot apply the same colour criteria used on Spitzer data to JWST data. Instead, a new formulation must be determined. To accomplish this goal rapidly, we choose to turn to machine learning. Machine learning shows promise as an alternative to the use of CMD criteria as it will leverage new relationships from the YSO fluxes than were previously identified with Spitzer. Furthermore, machine learning has the additional benefit of being able to assign probabilities of an object being a YSO.

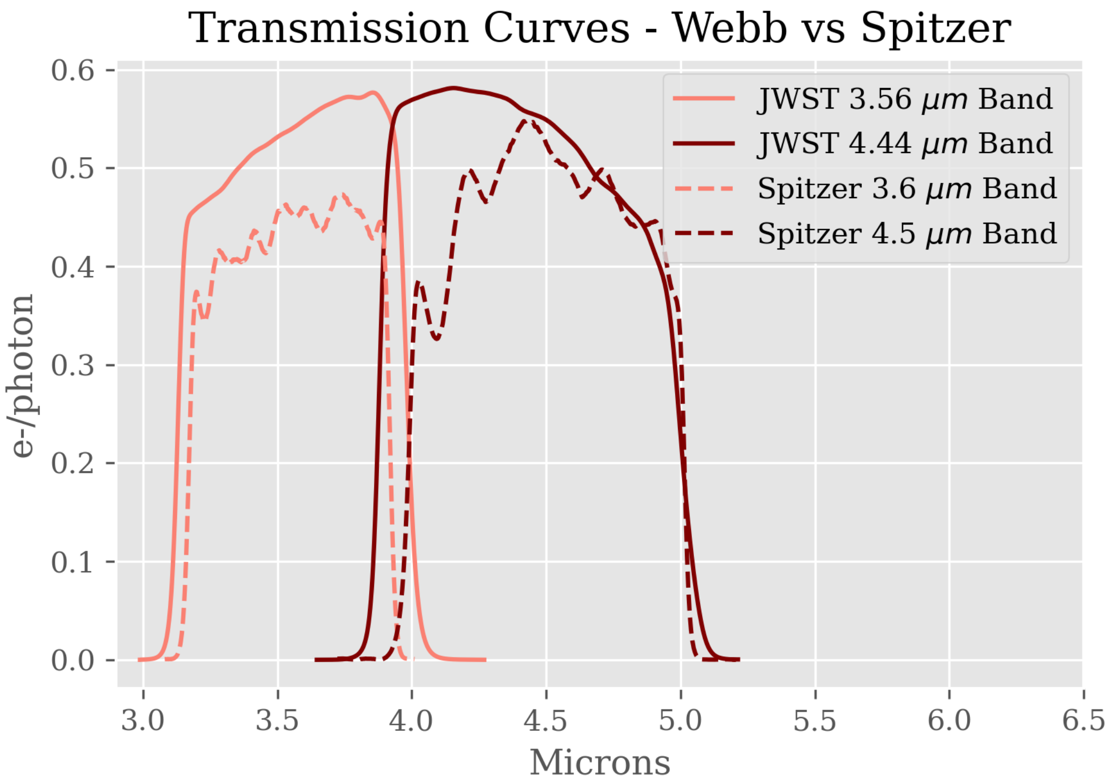

Figure 1.3: Transmission curves of both JWST (solid lines) and Spitzer (dashed lines) for the JWST NIRCam bands that match the Spitzer bands. JWST has an additional 27 filters in NIRCam that are not shown in this comparison.

1.3 Machine Learning Tools

In this section, we will outline briefly the different types of machine learning (ML) methods available, (Section 1.3.1), give an in-depth explanation of the method we chose and how it compares to its more standard form (Section 1.3.2), and finally summarize the previous works of ML applied to YSO identification (Section 1.3.3).

1.3.1 The Various Methods

There are many different types of machine learning, and in this work we will only touch on those mentioned here. Machine learning can be summarized as a set of decisions based on algorithms suited to error minimization that are built to predict a class, a value, or to group data in n-dimensional feature space. These decisions and how they are carried out vary significantly from algorithm to algorithm, and a comprehensive review of the error minimization and optimization are beyond the scope of this thesis. Instead, we will summarize those applied to YSO identification in recent years. In particular, we will look at neural network models and classical models.

Neural Networks

The term "neural network" is used broadly to describe any machine learning model which functions on the premise of "nodes" and "neurons." These networks are analogous to the brain; when an animal experiences something, it sets off a particular pattern of neurons through their sensory synapses, which send the signal to the brain, and this distinct pattern explains to the brain what it just experienced. So too with neural networks. We will briefly discuss two types of neural networks - standard Artificial Neural Networks (ANNs) and newer Convolutional Neural Networks (CNNs).

ANNs (Robbins & Monro, 1951; Rosenblatt, 1957) are built out of layers of nodes, where each node in one layer is connected to every node in the following layer by the neurons. When training, each neuron and node pair are given a certain weight, such that if a certain pattern of neurons is fired, it is linked to the final classification (or regression). The values of these weights are determined through a number of different methods depending on the type of ANN being implemented.

CNNs (Fukushima, 1980) are a more complicated version of ANNs built for the express purpose of analyzing images. They are constructed by taking the image, splitting it into however many filters are present, and convolving these down into a simpler parameter space, where the algorithm learns how to bring important features of the data to the surface to aid in classification. For instance, if it were trained to decide a cat from a dog, an important feature to "learn" would be the shape of the ears. CNNs are powerful tools for the classification of images on their own, and can also be used in conjunction with other networks, either classical or neural. In these cases, the final class is determined based on the returned probabilities from both networks.

Classical Methods

Classical methods are based on a variety of mathematical algorithms, with the idea that these are generally easier to interpret than neural networks. If a neural network is the student who "just knows" the answer, than a classical method is one which "shows their work." There are several classical methods, which can be separated into two further groups: those that work with decision-trees and those that make use of altering the dimensions of the data. Furthermore, they can also be distinguished by whether or not they are supervised (whether or not the result is already known to the algorithm). Supervised methods aim to produce results we expect to see, while unsupervised methods allow us to explore relationships within the data that have not been previously considered.

Decision tree-based classical (supervised) methods include random forest (Ho, 1995; Breiman, 2001), gradient boosting (Friedman, 2001, 2002), XGBoost, etc. For classification, Random Forest (RF) models will take the mode of the result from a number of different trees. For regression (numerical) tasks, RF takes the mean of the results from all trees. The number of decisions made in a single tree, as well as the number of trees defined can greatly impact the performance of these models. Although they can be very good at dealing with imbalanced training sets, if the model overfits[1] the data it may lead to inaccurate results. Gradient Boosting (GB) machines also use decision trees but in this case each tree is performed consecutively so that the next tree can learn from the mistakes of the previous. This learning is done by minimizing the gradient of the difference between the expected outcome and the true outcome at each step. A step beyond GB is eXtreme Gradient Boosting (XGBoost or XGB), where the model additionally performs error optimization, forcing the next tree to solve more aggressively where the previous tree went wrong. These methods are also prone to over-fitting, though with classifications they can still be helpful to separate out imbalanced classes from a very large sample.

---

[1] Overfitting is when an algorithm is only able to classify the data it was trained upon and is unable to generalize to larger or new datasets.

Classical methods which alter the feature space of the data include Support Vector Machines (SVM, Cortes & Vapnik, 1995), Logistic Regressors, K-nearest neighbours (K-NN, Fix & Hodges, 1951) (all supervised), and t-distributed stochastic neighbour embedding (t-SNE, Hinton & Roweis, 2002) (unsupervised), among others. Dimensionality here refers to the number of features (in the case of tabular data, this would be the columns of the table) or the number of extracted features (mathematical convolutions of the input data). SVM models take the input data and separate them into higher dimensional space such that classes can be assigned. Similarly, a Logisitic Regressor (LR) attempts to define a function to linearly separate classes in the n-dimensional space of the feature input. K-NN defines a parameter space in which all the data reside. The class of a new source is then determined by the classes of its k nearest neighbours, i.e., if 3 out of 5 of its nearest neighbours are Class 1, the new data-point will be assigned Class 1. Lastly, t-SNE is an unsupervised method which takes the n-dimensional data and convolves it down into two dimensions. By colouring the data based on class, we can often see the classes separate themselves. This approach is not able to classify data themselves, but it can help to find outliers or strange objects within different classes. These methods, in theory, are most similar to the empirical criteria performed on IR data for identifying YSOs by Gutermuth et al. (2008), but instead these methods perform their cuts in a higher dimensional space.

Finally, different ML methods can be used in conjunction with one another by utilizing the Stacking method, where the results of one method are fed into a second (or third, fourth, etc.) method and together they are used to classify (or regress) the data. Each method has benefits and drawbacks, and the best tool for the job is rarely clear.

1.3.2 The Probabilistic Random Forest Method

Though powerful, ML algorithms can have pitfalls. First, imbalances in the numbers of objects in different classes can make it difficult to train an ML algorithm to classify objects equally well, particularly in the case of parametric models such as ANNs and SVM. As YSOs are much less numerous than regular field stars, such imbalances are a relevant issue when identifying YSOs via ML. RF models are better at handling imbalanced datasets by individually classifying each object based off a training set (Breiman, 2001). Second, most ML algorithms cannot handle missing data. There are several different methods that have been developed to avoid this pitfall. For instance, the missing data can be imputed through the use of copulas, i.e., functions that connect the probability distributions of different features to each other. When using copulas, the joint probability distributions of the data are broken down into their base components, and the copulas combine these probabilities together (Nelsen, 2007). Using copulas to fill missing data, however, assumes that data are missing because an object is not within the field observed by the particular filter.

A different approach used to fill missing data from photometric catalogs assumes objects in question are indeed present but not seen as point sources (Chiu et al., 2021). For example, a YSO may be heavily obscured in certain filters, or its SED may drop below the sensitivity limit of the instrument. The solution provided by Chiu et al. (2021) was to fill the data in missing bands with 1% of the smallest flux obtained in that band as a reasonable estimate for the thermal noise of the detector. This method hence accounts for the clarity of an object at different wavelengths, which is important for the determination of class.

An alternative solution to the issue of missing data is provided by the Probabilistic Random Forest (PRF) method released by Reis et al. (2019). The PRF method uses both the values and errors of each filter to create probability distributions for each data point, where the expectation value is the data point's flux, and the standard deviation is the error on this flux, assuming a Gaussian distribution. An RF-like algorithm is hence trained, and when an object is sent through the network, it is no longer sent along one branch of the tree. Instead, at every decision node, the probability of the object being on either side of the node is propagated, with probability determined by the Gaussian distribution. For a full prescription, see Reis et al. (2019). This method has the benefit of not assuming what the missing data may be while still accounting for them by passing any node that relies upon the data with equal probability to either side.

1.3.3 Previous works

ML has been used for YSO identification several times over the last few years, and these works are summarized here.

Marton et al. (2016) used an SVM model to identify YSOs from all-sky WISE data. They sought to identify Class I and II sources from the total field. To do this, they used a series of SVM models, where at each step they aimed to separate out a different type of contaminant, such as extra-galactic objects, main-sequence stars, evolved stars, interstellar medium related objects like shocks, AGB candidates, and Class III YSO candidates. This approach had the added benefit of tuning the model to smaller changes in the sample with each iteration. Their object classes were determined by matches to SIMBAD objects, and with those they were able to get a large enough training sample to ensure statistical reliability. They estimated that they have a false positive rate of 1.9%, but do not provide further statistics such as recall, precision, or any other standard metric for comparison to other works. Refer to Chapter 2.3.2 for definitions of these metrics.

Miettinen (2018) used a compilation of various ML methods to classify YSOs into three categories: Class 0, Class I, and Flat-Spectrum. They tested eight different methods, seven classical methods, and one ANN. They found that the best method to separate the classes from one another was a gradient boosting model. As they used supervised models, they had to have a classification pre-applied to their data, and this was done by using the data-set from the Herschel Orion Protostar Survey, as classified by Furlan et al. (2016). Although providing fair results, the major shortcoming of this work is that they only consider protostars, and thus they could only classify objects into their YSO stage if they already knew the object was a YSO.

A more realistic approach is to create a method which can classify YSOs based on the unsorted input from a telescope. This route was taken by Cornu & Montillaud (2021), who used an ANN to classify objects from Spitzer as either Class I YSOs, Class II YSOs, or contaminant objects. The reason for using only those three classes is that (1) Class 0 objects are not easily visible in Spitzer data (the only data used in this survey); and (2) Class III objects have a signature very similar to regular main sequence stars at these wavelengths, and thus the neural network would lose performance in trying to correctly classify Class III. As the most important YSOs to star formation rate calculations are the youngest of the stages, it is then a reasonable sacrifice to allow any Class III YSOs to be classified as contaminants. They tested their network on the full Spitzer Orion Survey (Megeath et al., 2012), as well as on a survey of NGC 2264 (Rapson et al., 2014). These surveys include a number of different object types, and the training and validation sets were specifically chosen to increase model performance while still being representative of the night sky. In choosing such active star-forming regions, however, they put themselves at risk of training a network which is not applicable to more quiescent regions.

Chiu et al. (2021) performed still another classification using the full c2d survey mentioned previously, (Evans et al., 2014a) which analyzed more quiescent regions than the Orion and NGC 2264 star-forming regions. They also used an ANN as well as testing random forest, XGBoost, KNN, and SVM classical methods to classify objects into one of three categories: YSOs, extra-galactic sources (EGs), and stars. They found that the ANN had the highest recall and F1-Score, refer to Chapter 2.3.2 for metric definitions. Furthermore, they tested the benefits of using all bands, bands plus errors, bands where the fluxes of each object are normalized such that distance effects are removed, and using only three bands. With full bands, they utilized the four IRAC bands, the MIPS 24 μm band, and the JHK bands from 2MASS. They obtain the highest recall and precision of all other networks explored in this section. They also provide a website to label Spitzer data that is available for anyone to use.

Kuhn et al. (2021) classified YSOs using data from the Spitzer GLIMPSE surveys across the entire Galactic Plane. One subsection of these data is the GLIMPSE Vela-Carina survey (Majewski et al., 2007) which imaged all of the Carina Nebula region, which will be referred back to in Chapter 2. They utilized a RF trained on an imbalanced mix of YSOs and contaminant (non-YSO) field objects where the number of YSOs accounted for less than 25% of the full training set. They used a standard ML metric called "Area Under the Curve" (AUC), where the curve in question is the Receiver-Operator Curve (ROC), to determine the best fits. A ROC curve shows a binary classifier's ability to distinguish true from false classifications, as the discrimination threshold changes. As RF models output the probabilities of objects being in the positive class, Kuhn et al. (2021) defined an object as being a YSO if it had at least a 50% probability of being so determined by their RF model. After YSOs were identified as such, Kuhn et al. (2021) further used cuts based on spectral index to determine their respective stages within pre-main-sequence evolution, thereby labelling them as either Class I, Flat-Spectrum, or Class II YSOs. They published their catalog as the Spitzer-IRAC Catalog for YSOs, henceforth referred to as SPICY.

Kinson et al. (2021, 2022) recently used the PRF method to catalog massive YSOs based on Spitzer data in two galaxies: NGC 6822 and M 33. They classified NIR and FIR data of these regions into several classes: YSOs, oxygen- and carbon-rich AGB stars, RGB and red supergiant stars, AGN, massive main-sequence stars, Wolf-Rayet stars, and Galactic foreground stars. In both cases, they used accuracy as a metric, and located hundreds of new YSOs and several new star-forming regions within these galaxies. In Kinson et al. (2021), they additionally use t-SNE to check to see how the two methods compared. They specified 20 groups and found that t-SNE was unable to put YSOs in their own group, hence implying that t-SNE alone is insufficient for the identification of massive YSOs in galaxies.

Chapter 2

Finding YSOs Within Webb Data, a Case Study of NGC 3324

This chapter is based off of a paper submitted to the Astrophysical Journal. All data analysis in this paper was done by the author of this thesis, Breanna Crompvoets, bar the production of column density maps that were provided via private communication with Dr. Thomas Preibisch. The JWST FITS images were accessed as data products from the Mikulski Archive for Space Telescopes, and photometry on these data was conducted by Breanna Crompvoets with advice from Dr. Peter Stetson on how to use his software. A detailed set of instructions to obtain photometry from Webb data, as determined for this work, is located in Appendix A, and a sample of the code used to classify the objects is located in Appendix B. Co-authors on the submitted paper include Dr. James Di Francesco, Dr. Hossen Teimoorinia, and Dr. Thomas Preibisch.

2.1 Introduction

Stars form within molecular clouds throughout the Galaxy. The efficiency of such clouds in making stars can be measured by taking the ratio of the total mass of stars produced over a given time to the total mass of the cloud that produced them. The former quantity can be obtained through censuses of the Young Stellar Objects (YSOs) associated with clouds. Given that many clouds enshroud their YSOs behind significant amounts of dust intermixed with their gas, sensitive multi-band infrared imaging of clouds in the infrared is critical for detecting YSO populations. Moreover, YSOs may be spatially crowded, especially in distant clouds, requiring high resolution observations to identify them individually. In this paper, we determine the YSO population revealed by infrared images of the “Cosmic Cliffs,” hereafter CC, part of the NGC 3324 star-forming region observed early on by the James Webb Space Telescope (JWST) utilizing the NIRCam instrument, and whose data were released publicly in July 2022.

Prior to Webb, the infrared Spitzer Space Telescope contributed much to the field of star formation, providing in-depth photometric surveys of several Galactic star-forming regions (e.g., Majewski et al., 2007; Evans et al., 2014a). These surveys have themselves been extensively studied and YSOs in different stages have been identified through a variety of techniques (e.g., Gutermuth et al., 2009; Dunham et al., 2015). Even after Spitzer ceased operations, its data continue to be reviewed and more YSOs identified, with recent work focusing on using machine learning techniques (Kuhn et al., 2021).

Despite the continued usefulness of its data, Spitzer nonetheless had intrinsic sensitivity and resolution limitations in its ability to detect embedded YSOs. Fortunately, JWST now surpasses the sensitivity of Spitzer by approximately two orders of magnitude (Rigby et al., 2023) in the same field, at a resolution over seven times that of Spitzer at wavelengths in common. With JWST operating in both the near- and mid-infrared, it is poised to provide sensitive observations of previously undetected YSOs buried within molecular clouds.

Astronomers have recently begun turning to machine learning (ML) as a tool for analyzing large amounts of astronomical data, and indeed ML has been used for the classification of YSOs (e.g., see Cornu & Montillaud, 2021; Kuhn et al., 2021). With JWST providing images with far more sources in small fields than previous telescopes, ML is a logical choice for the classifications of those sources. In particular, we explore classifying YSOs from other objects in the JWST fields using a Probabilistic Random Forest approach.

This paper will serve as a proof of concept to identify the improvement of using JWST NIRCam data over Spitzer data for the purpose of identifying YSOs via the use of ML. Further JWST data becoming public in the near future will allow for these approaches to be expanded and refined.

This paper is split into the following sections. Section 2.2 summarizes background information relevant to topics described in the paper. Section 2.3 describes how we created the catalog of JWST data for the CC, as well as a description of the ML model used. Section 2.4 provides the results of this ML model, including the number of candidate YSOs detected and a comparison with those found in previous works. Section 2.5 discusses the accuracy of our classifications, and provides an analysis of the capabilities of JWST for YSO detection in comparison to those of Spitzer. A summary in Section 2.6 concludes the paper.

Figure 2.1: The region of NGC 3324 imaged by JWST. Stiff peaks along the Gum 31 H-II region are formed by stellar winds originating from the massive stars in NGC 3324. This colour image is made from six NIRCam filters. Red: F444W, Orange: F335M, Yellow: F470N, Green: F200W, Cyan: F187N, Blue: F090W. Image credit: NASA, ESA, CSA, and STScI.

2.2 Background

2.2.1 NGC 3324 and Gum 31

NGC 3324 is a young stellar cluster located on the North-Eastern corner of the Carina Nebula Complex, inside of the Gum 31 HII region. The diameter of Gum 31 is $\sim 15'$ (Cappa et al., 2008), while NGC 3324 is $11'$ across (Bisht et al., 2021). Recent studies of the young cluster have placed it at a distance of 2.35 kpc (Göppl & Preibisch, 2022), while the HII region is attached via filamentary structures to the larger Carina Nebula Complex (Preibisch et al., 2012; Roccatagliata et al., 2013). NGC 3324 includes at least three massive stars, one of which (HD 92207, spectral type A0Ia) has a very strong stellar wind (Kudritzki et al., 1999). This cluster excites the Gum 31 nebula, and causes sharp features to form along the edges of the bubble, as seen in the JWST/NIRCam image shown in Figure 2.1.

Previously, the Gum 31 region was examined with Spitzer, WISE, and Herschel data by Ohlendorf et al. (2013) to search for YSOs. They found a total of 661 candidate YSOs within the WISE data, as well as 91 protostar candidates within the Herschel data. Of these, 17 sources were fit to SED models, and a completeness limit of ~ 1 M$_\odot$ was determined. Based on their results, they extrapolated the initial mass function to determine a total population of 5000 YSOs within the Gum 31 nebula whose mass is greater than ~ 0.1 M$_\odot$. From spatial distributions of the candidate YSOs, they concluded that both collect-and-collapse and radiative triggering are the prevailing avenues of star formation actively ongoing within Gum 31.

Further Herschel data of the greater Carina Nebula Complex were analyzed by Gaczkowski et al. (2013), which connects to and hence is likely the larger portion of the nebula which forms Gum 31 (Gaczkowski et al., 2013). They find 144 protostars with detections in the 70 μm band, and estimated the total protostellar population to have 2880 objects, based on the Kroupa initial mass function. From this estimate, they derived a star formation rate within the full Carina Nebula Complex of ~ 0.017 M$_\odot$/yr. A spatial analysis of all 642 sources detected revealed that the Class 0 and Class I sources are located along the cloud boundaries, while the more-evolved YSOs identified in previous surveys are located along the cloud edges. Lastly, they found a non-statistical deficit of massive young stars, leading to the conclusion that the latest round of star formation is no longer forming massive stars in the Carina Nebula Complex.

Gum 31 has been examined in the x-ray by Preibisch et al. (2014), who found 679 x-ray emitting objects, 500 of which are matched to objects detectable in the IR. Although x-rays trace YSOs, they are also known to trace other sources such as fore- and background field stars and extragalactic sources. As such, Preibisch et al. (2014) determined that the contamination within their sample is likely to be 25-30%. They found, however, that those objects with mid-IR counterparts, found using VISTA, make up approximately 75% of the catalog, and propose that these objects are true YSOs and not contaminants. The x-ray emitting sources with VISTA counterparts are likely made of mainly Class III YSOs, as x-ray selection most efficiently captures this class (Feigelson, 2018).

Preibisch et al. (2014) also analyzed the x-ray luminosity function and determined there should be approximately 4000 YSOs of mass greater than ~ 0.1 M$_\odot$ within the associated emission nebula of Gum 31, which falls within the estimate of Ohlendorf et al. (2013). They found that 30% of YSOs are concentrated in two areas, one of which is NGC 3324, and that the remaining 70% are spread homogeneously across the full bubble. By comparing the sizes of Gum 31 and the CC and using the above distribution, their results suggest approximately 630 YSOs should exist within the CC field alone, including many Class III objects.

Filter	t_{exp}(s)	Use
F090W	25768.32	dust and background stellar field
F187N	46382.88	ionized gas via the bright Pa-α
F200W	25768.32	dust and background stellar field
F335M	6442.08	3.3 μm PAH emission
F444W	6442.08	dust scattering from large grains
F470N	11595.72	H_2 from embedded jets/outflows
F770W	6771.08	PAH emission
F1130W	6771.08	PAH emission
F1280W	6993.12	12.81 μm [Ne II] line emission
F1800W	5994.08	cool dust and proplyds

Table 2.1: JWST filters used by ERO to image NGC 3324, their exposure times, and their uses as described in Pontoppidan et al. (2022). The first block lists NIRCam filters and the second lists MIRI filters.

2.2.2 JWST Data

On July 11th, 2022, the first observations from JWST were released (Pontoppidan et al., 2022). These data included observations of four different astrophysical objects, one of which was the NGC 3324 star-forming region located adjacent to the Carina Nebula Complex. JWST imaged this region with two instruments: the Near Infra-Red Camera (NIRCam) to observe its dust and look for emission lines of H_2, Poly-Aromatic Hydrocarbons (PAHs), and Pa-α; and the Mid Infra-Red Instrument (MIRI) which was able to probe for objects hidden within the dust that may have been rendered undetectable at shorter wavelengths.

The Early Release Observations (ERO, Pontoppidan et al., 2022) of NGC 3324 focused on the CC, a $\sim 7'.4 \times 4'.4$ area with NIRCam, and a $\sim 6'.4 \times 2'.2$ area within the NIRCam field with MIRI, located on the edge of the Gum 31 bubble. The data were collected in six NIRCam bands and four MIRI bands; see Table 2.1 for details on the bands used, exposure times, and uses. The exposure times varied for each filter, and the `FULLBOX` 10-point dither pattern was used for NIRCam imaging and 8-point dither for MIRI imaging (Pontoppidan et al., 2022). We will henceforth focus on the NIRCam data for this paper as, due to its much wider field of view, NIRCam will be used more extensively of the two, and indeed often on its own.

Reduced FITS images as well as a source catalog generated for each filter by the JWST pipeline, were made publicly available through the Mikulski Archive for Space Telescopes (MAST)[1]. For our use, we accessed the available FITS images and made our own source

---

[1] doi:10.17909/67ft-nb86

catalogs from them, as those provided by MAST were deemed inadequate for our intended purpose. These data products have been reprocessed several times since July 2022 as better JWST flux calibrations became available; for this paper, the data products for NGC 3324 were last accessed on June 13th, 2023.

The JWST data of the CC have already been probed to understand the capabilities of JWST to detect jets and outflows. Previously, Reiter et al. (2022) looked at data from the narrowband 1.87 µm filter and the difference between data from the narrowband 4.7 µm and wideband 4.44 µm filters from JWST. In combination with archival Hubble data, they used these datasets to identify 31 outflows within the CC field of view, including seven Herbig-Haro objects only visible in the infrared (IR). Along with their identifications of outflows, they provided a list of 24 outflow driving source candidates, i.e., IR-excess sources located along the estimated arcs of travel determined by tracing the observed outflows back in time. As not all outflows were visible in the earlier Hubble data, they could not determine proper motions for all outflows, and as a result straight-line estimation was used when appropriate. Reiter et al. (2022) also checked to see if any of their identified YSO candidates had been previously identified with Spitzer data via comparison with SPICY and found matches to six of the 24 possible outflow driving source candidates.

More recently, an analysis of JWST data of the nearby star-forming dwarf galaxy NGC 6822 has illustrated the importance of using JWST data for YSO identification. Lenkić et al. (2023) found that YSOs selected using Spitzer-based criteria could be matched to as many as six different objects in JWST images that had been previously blended together in lower-resolution Spitzer data. They additionally were able to identify nearly three times as many YSOs as in previous surveys, and clarified that several "YSOs" identified in previous surveys were instead either reddened stars or galaxies, further emphasizing the improvement of JWST over Spitzer.

2.3 Methodology

In this section, we describe the method of creating and using photometric data of the CC. We explain how the catalog was created (§2.3.1), how the Probabilistic Random Forest was verified as a reasonable and indeed preferred machine learning model (§2.3.2), and finally how the model was executed (§2.3.3).

2.3.1 Catalog Creation

As stated above, we used pipeline-generated JWST images of the CC fields that were made

Band	Number
F090W	19061
F200W	19490
F335M	19458
F444W	19486
F470N	19427

Table 2.2: The number of objects with photometry in each band.

publicly available as FITS files by MAST. We did not use the photometric data also provided by MAST for two reasons: (1) a catalog of all identified sources was not accessible, and, more importantly, (2) many objects were found to have photometric data in only a subset of the bands for which they are visible. For example, an object's fluxes could be provided for the F090W and F200W bands but not the F444W band despite the object being seen in that image. Thus, we applied the photometric software DAOPHOT (Stetson, 1987, 2011, version as of May 30th, 2023, accessed via private communication) to the NIRCam data, keeping only objects that were seen in at least three bands that had measurement errors less than 0.2 mags. With these criteria, we retrieved a total of 19 497 individual sources. The number of objects in each NIRCam band is written in Table 2.2.

The DAOPHOT program obtains photometry by first identifying all point and extended sources within the field, and running synthetic aperture photometry on them. From this photometry, a local sky brightness and magnitude is computed for each object. Finally an empirical point-spread function (PSF) is determined for each frame and used to compute PSF photometry for each object. Importantly, our PSF is defined from 200 user-identified sources in each band. These sources must (a) not include any bad or saturated pixels; (b) be free from the bright dusty fields and hence from any stray emission which could confuse the PSF algorithm; and (c) must be within 2-sigma of the mean PSF from all identified PSF stars. Each source is fitted to the PSF using least-squares profile fits.

At this point, we excluded the F187N data from our analysis due to a positional misalignment relative to other images at approximately x = 6000 pixels. When we attempted to use those data regardless, spurious YSO detections were obtained within 500 pixels to either side of the misalignment. These no longer occurred when this band was removed from the ensemble. As this band does not contain signatures necessary for YSO identification, we do not attempt to correct this image.

Next, we chose 253 Spitzer-detected objects within the CC field, which could be securely matched to a bright and relatively singular counterpart in the JWST data. Appropriate sources were hand-chosen based on their magnitude, where only JWST objects with bright fluxes that were coincident with Spitzer sources (i.e., within the latters' FWHM) were chosen. We removed any sources that were blended or which we were unable to identify their JWST counterpart confidently. We next used the `Astropy` (The Astropy Collaboration et al., 2018) `match_coordinates_sky_` task to match the Spitzer detections from the GLIMPSE catalog of the region (Majewski et al., 2007)[2], available SPICY targets (Kuhn et al., 2021), and appropriate YSOs from Ohlendorf et al. (2013), resulting in a total of 23 YSOs: 18 from SPICY, 11 from Ohlendorf et al. (2013), with six objects overlapping. The locations of these objects were then updated to match their hand-selected JWST counterparts. This process resulted in 208 objects detected by Spitzer being matched to JWST sources, of which 185 are contaminant objects, and 23 are YSOs.

Once the catalog was created, we utilized data augmentation to create a simulated dataset for training purposes. For this approach, we first removed any identified sources with missing data, leaving us with 22 YSOs and 181 contaminant objects from which to build an augmented dataset. We then withheld 10% of the data to confirm the generalization of our network, and so the augmented dataset was made from the data of 19 YSOs and 152 contaminants. Next, we determined an error in the flux for each band that is representative for all objects. We then randomly sampled and varied the data from objects in each class by adding or subtracting this error, efficiently producing new data that were still, within error, effectively the same as the old data and with which we could thus securely apply classifications and probabilities of classifications. This strategy allowed us to over-sample the smaller YSO class while still producing unique data with which the algorithm could be trained.

Finally, we split our data into training and validation sets. The training set is made of 20 000 objects obtained via the aforementioned data augmentation for each class. The validation set uses all 208 objects with classifications (23 YSOs and 185 contaminant objects). We maintained this imbalanced set to ensure that the metrics we obtained were representative and that we did not produce misleading metrics resulting from under-sampling the contaminant class or over-sampling the YSO class. In addition to having data that are, by design, similar to the augmented data, this validation set also has 19 objects which are completely unknown to the augmented data.

2.3.2 Testing Validity of Probabilistic Random Forest method

We compare here the results of the regular RF to the PRF on a test dataset. Reis et al.

---

[2] doi:10.26131/IRSA213

(2019) provided a comparison with the regular RF that shows that when all labels are correct, the PRF and RF perform at the same accuracy. When purposefully introducing incorrect target labels, however, they found that the PRF greatly outperforms the regular RF. As an additional test, we demonstrate the relative performance of a PRF vs. a regular RF approach, using both copulas and thermal noise to fill the missing data. For this test, we use data from the Cores to Disks (c2d, Evans et al., 2014a) survey, which contains Spitzer data with both completely filled and missing values. We first use 10 000 objects with all bands available, then 9000 objects with all bands available and an additional 1000 (randomly chosen) objects with data missing in at least one band to obtain a case where 90% of data are filled. Similarly, we also obtain data-sets where 80%, 70%, 60%, and 50% of data are filled. In all cases, the data are real observations, and no data are artificially removed. We use YSOs as our positive class and all others as contaminants, where YSOs make up approximately a third of the sample.

We focus on four metrics that can be used to assess the performance of an ML model: accuracy, recall, precision, and F1-Score. Each of these requires some combination of the numbers of True Positives (TP), False Positives (FP), True Negatives (TN), and False Negatives (FN). For our purposes, TP is the number of objects correctly classified as YSOs, TN is the number of objects correctly classified as contaminants, FP is the number of contaminant objects incorrectly classified as YSOs, and FN is the number of YSOs incorrectly classified as contaminants. Accuracy, $\mathrm{A} = (\mathrm{TP} + \mathrm{TN})/(\mathrm{TP} + \mathrm{FP} + \mathrm{FN} + \mathrm{TN})$, is a measure of the total number of correct identifications but can be easily made misleadingly high as a result of having a relatively populous negative class. As such, we do not use it here. The F1-score, $\mathrm{F1} = 2\mathrm{R} \times \mathrm{P}/(\mathrm{R} + \mathrm{P})$, however, is a metric defined as the harmonic balance between recall $\mathrm{R} = \mathrm{TP}/(\mathrm{TP} + \mathrm{FN})$ and precision $\mathrm{P} = \mathrm{TP}/(\mathrm{TP} + \mathrm{FP})$. We use it as our metric of choice because we are aiming to configure a network with low contamination by much more numerous contaminants (high precision) while still maintaining a high recovery of YSOs (high recall).

Figure 2.2 shows the F1-Score performances for the RF and PRF models when the amount of data available is decreased. Two RF methods are evaluated, one with copulas and the other with thermal noise used to fill missing data. We used the `Python` package `copulas` for our calculations. A slight decrease in F1-Score is seen with the increase in objects that are missing flux measurements. Nevertheless, we find that all three cases perform well ($> 95\%$) and within a few percent of each other, though the PRF F1-Scores are slightly but systematically lower than the RF scores. In general, filling the data with noise or using copulas performs equally well.

Despite the high proficiency of using copulas or thermal noise, these methods come with

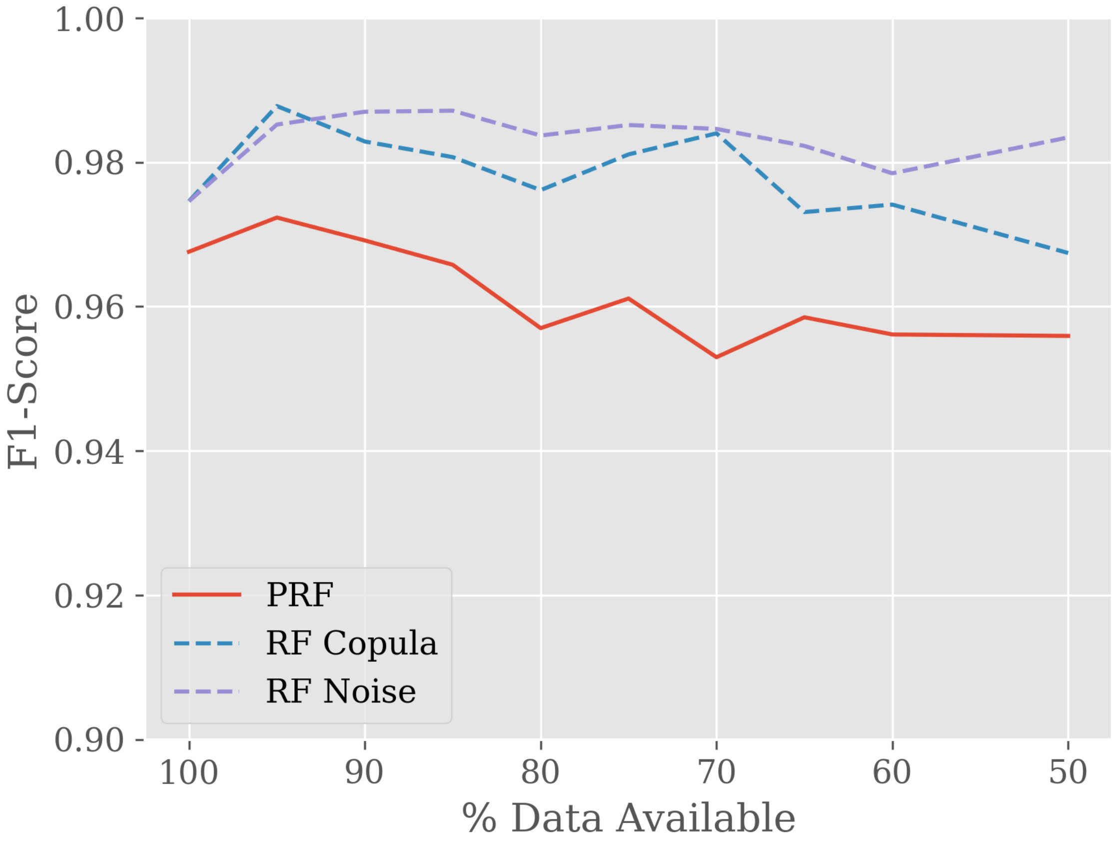

Figure 2.2: A comparison of the F1-Scores for the validation set of the PRF (solid red line), an RF filled via copulas (blue dashed line), and an RF filled with thermal noise (purple dashed line) as a function of the amount of missing data.

their own caveats. The use of copulas assumes that the data were not imaged and that the environment that generated the copula is reflective of that in which the point source in question is missing data, which is not necessarily the case for point sources missing JWST data. Similarly, filling the data points with thermal noise assumes that the object was not detectable because it was too weak at a particular wavelength, and does not take into account that an object may be saturated, or that the source was not imaged at all. Both cases rely on certain assumptions, which may or may not be true for any given point source. The PRF, however, does not make any assumptions about why data are missing, and instead works around them. Although this method is much simpler than the above methods for filling missing data, it still performs quite well in the above test. Indeed, this paper serves as a proof of concept for applying the PRF method to JWST data.

For this paper, we choose not to fill the data with either noise or copulas when comparing to the regular RF model, as in some cases the objects are not measured in the band due to over-saturation, and others because they are not visible due to intervening dust or not detected at that wavelength. In this way, we end up with one algorithm which we allow to make its own assumptions about the missing data (PRF) and one algorithm which makes no assumptions about the missing data and instead ignores those objects (RF).

2.3.3 Applying the Probabilistic Random Forest method

To classify the full JWST dataset, we must first have predetermined classifications for some portion of the data. We chose to train and test on IR-identified sources, for consistency with previous works. Between the SPICY catalog (Kuhn et al., 2021) and the YSOs identified by Ohlendorf et al. (2013), we obtain 23 YSOs and 185 objects identified as contaminants, which are securely matched to JWST data; see §2.3.1 above. By training on data within this field, the possible effects of extinction on the measured fluxes are minimized. Furthermore, absolute flux calibrations are less important than the shapes of the spectral energy distributions as the latter are what the model learns. To emphasize this point, and to avoid biasing our results to bright objects, we do not train on magnitudes, but rather on features that could be extracted from the magnitudes, such as colors.

The PRF requires three input parameters for supervised classification: the input data, errors on the input data, and either the targets for the input data, or the probability of the object being the positive or negative label. After some testing, we determined that using the targets rather than the probabilities was appropriate for our uses. We tested several different input data structures: including all five NIRCam bands available, and removing one band at a time to explore their relative impact on the classifications. The scores remain similar with these tests, and so we decided to take a conservative approach to avoid prejudging the merit of including one band over another and use all bands for the models.

We analyzed data from the ERO of the CC region in five NIRCam filters, those corresponding to 0.90 µm (F090W), 2.00 µm (F200W), 3.35 µm (F335M), 4.44 µm (F444W), and 4.70 µm (F470N). These bands are able to penetrate to moderate depth with sensitivity limits of 28.6 mag, 26.2 mag, 22.1 mag, 22.8 mag, and 22.3 mag, respectively. Conversely, the photometry of objects is able to be retrieved to magnitudes of 12.9 mag, 10.5 mag, 8.2 mag, 7.7 mag, and 6.7 mag, respectively, where objects beyond these limits have too many saturated pixels for accurate photometry to be retrieved. Indeed, the saturation limits of the longer wavelength range are counterproductive to the identification of YSOs, as that range is where the YSOs will be at their brightest in the NIRCam data. As such, the saturation limits of the longer wavelengths may lead to a slight underestimation of YSO numbers. Fortunately, this saturation limit is only relevant for just 60 of 19 497 total objects within the CC field, and so should not greatly impact the results. Hence, previous IR studies with lower sensitivity can still be complementary to those with JWST.

Figure 2.3 shows the distribution of the number of candidate YSOs (cYSOs) identified vs. F1-Score based on 100 runs of the PRF using all five NIRCam bands available. The number of identified objects varies from run to run, despite the F1-Scores remaining the same. Hence, we take the classifications for all objects over 100 runs of the model and take the probability of a given object being a YSO as the average probability that it is identified as a YSO. With all bands included in the ensemble, 496 objects are identified in the CC as having a $> 50\%$ probability on average of being a YSO. The error on the assigned probabilities is $\sim 1\%$.

Figure 2.4 shows various metrics for the validation set of the cYSO classification as a function of the cut on the probability of a given object being a cYSO. The higher the cut, the fewer objects are classified as cYSOs, leading to a general decrease in recall and increase in precision, though the F1-Score remains relatively even. There are no cYSOs with exactly 100% probability with the RF, leading to the sharp decrease in all metrics at this extreme cut.

Figure 2.5 shows the numbers of YSOs that are identified via the RF and PRF when the minimum threshold probability of being a YSO is increased. The threshold ultimately selected is arbitrary, and so we set it to 50% to require that the probability of the object being a YSO is greater than the probability it is not. This threshold is standard for ML purposes, and is further justified in our case because our analysis is based on YSOs identified previously in the CC by Kuhn et al. (2021) with the same threshold. We estimated earlier that the

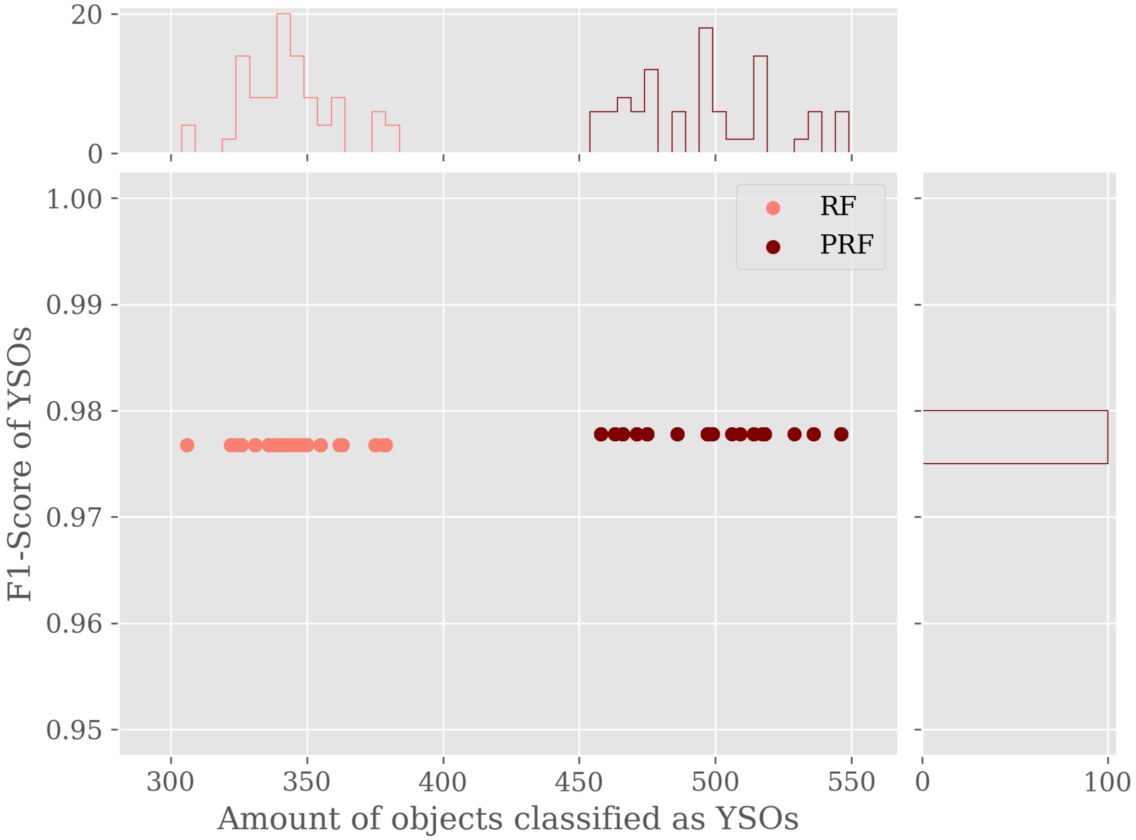

Figure 2.3: A plot comparing the number of cYSOs found vs the F1-Score on the validation sets for 100 runs of the PRF using all JWST bands available.

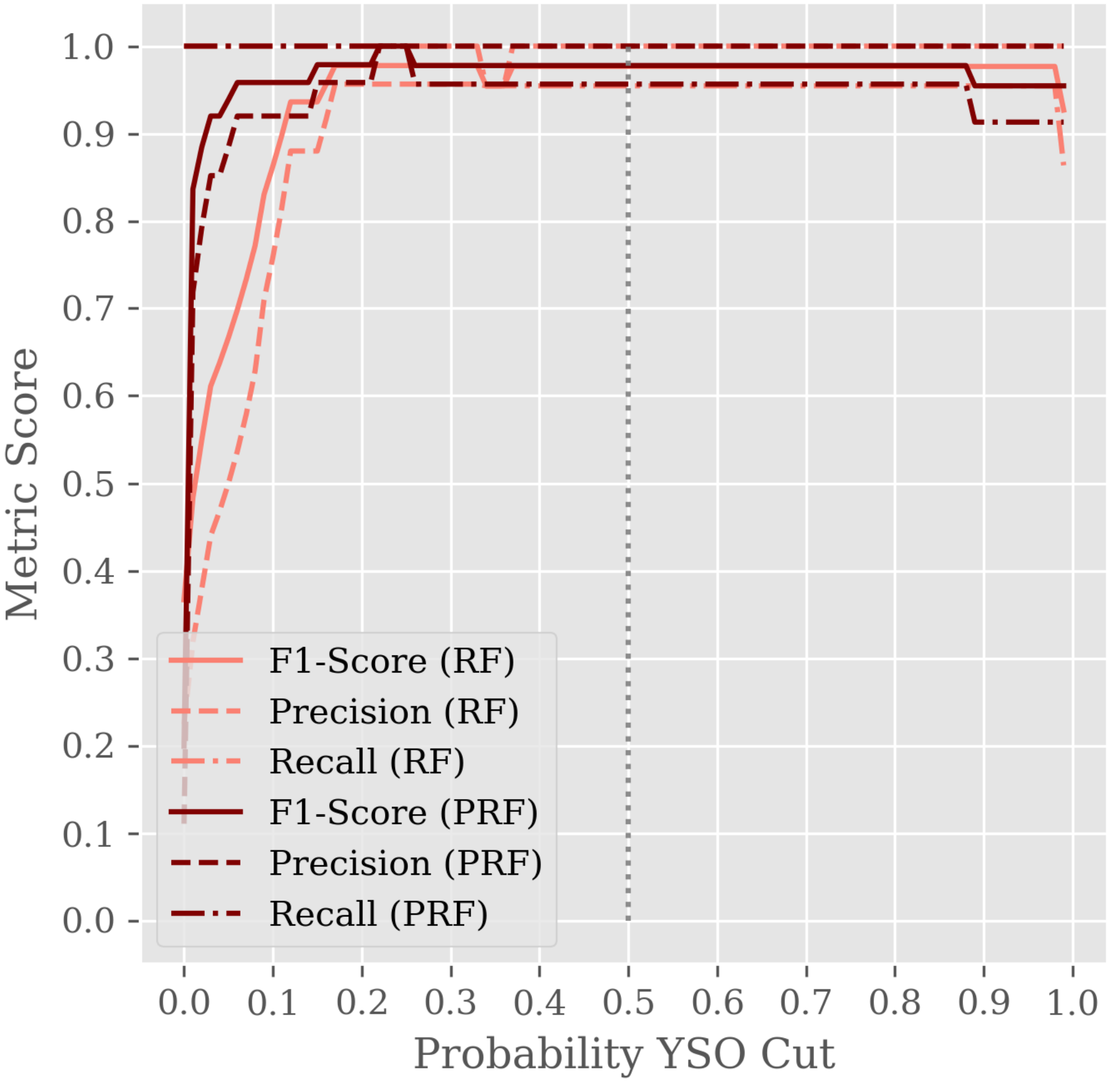

Figure 2.4: The trends of various metrics with a variation of the cut in probability that determines if a given object is a YSO. There are no YSOs with 100% probability.

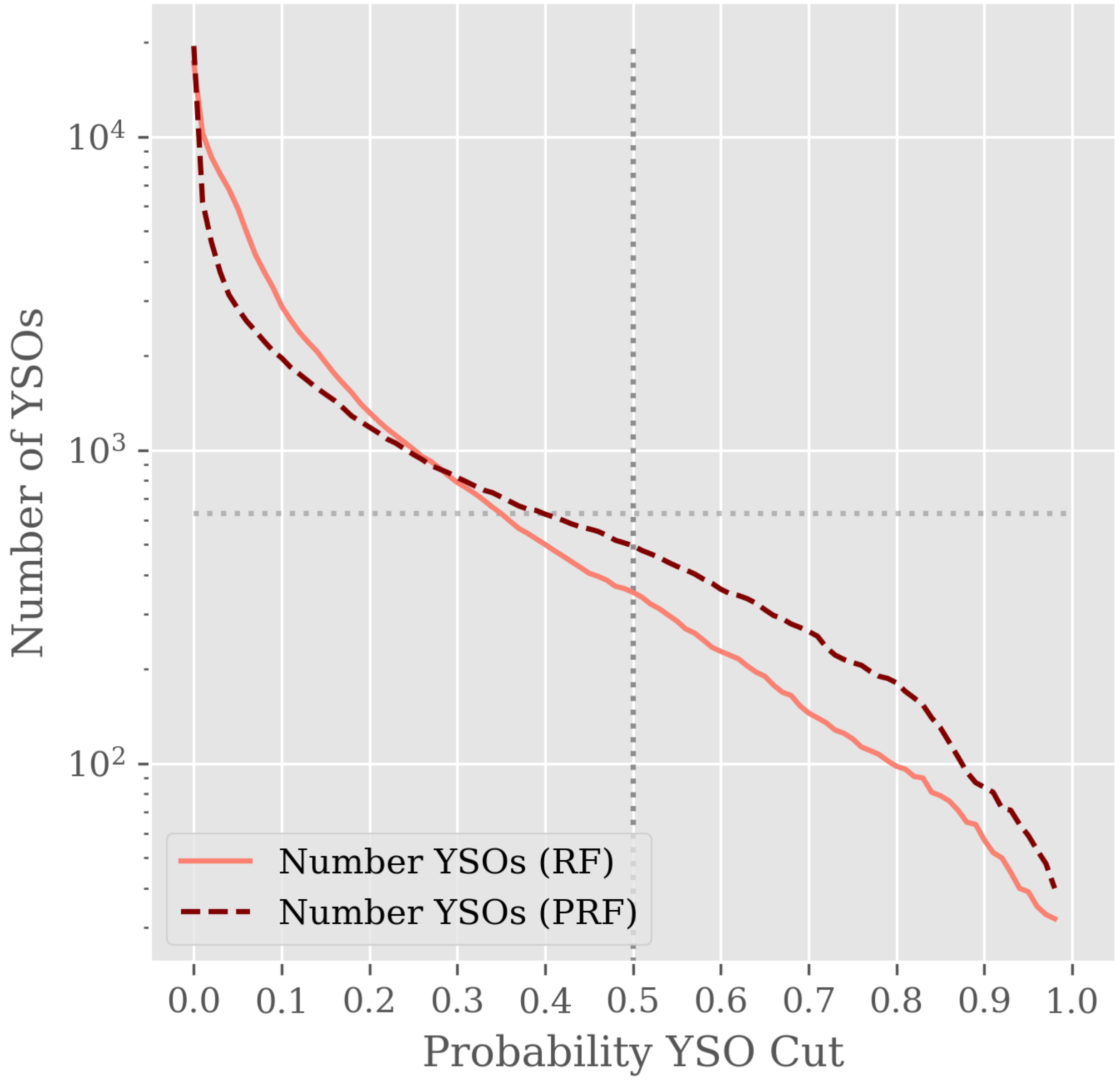

Figure 2.5: The trend of number of YSOs with the cut in probability that determines if a given object is a YSO. There are no cYSOs with exactly 100% probability. The gray horizontal line marks the upper limit of YSOs we expect within the region from previous works, and the gray vertical line marks the 50% threshold.

number of YSOs likely to be within the CC region is approximately 630, using estimates from Preibisch et al. (2014). Indeed, the number of cYSOs with probability greater than 50% is 496, well within this estimate. All PRF algorithms output the probability of an object being a YSO or contaminant. Again, we eliminate the bias of the random seed on the classification of objects by taking the average probability over 100 runs.

Finally, we find that the PRF is able to outperform the RF when looking for True Positives. Figure 2.6 shows the relative performance based on the number of True Positives returned as a function of precision. This metric can be determined from rearrangement of the precision equation to find Equation 2.1, which is what we plot as "Relative Performance" (RP).

$$\mathrm{RP} = \frac{\mathrm{TP}_{\mathrm{PRF}}}{\mathrm{TP}_{\mathrm{RF}}} = \frac{\mathrm{P}_{\mathrm{PRF}}}{\mathrm{P}_{\mathrm{RF}}}\frac{\mathrm{N}_{\mathrm{PRF}}}{\mathrm{N}_{\mathrm{RF}}}. \tag{2.1}$$

Hence, the higher number of cYSOs found with the PRF is due to a higher number of True Positives. Indeed, at a threshold of 50%, RP indicates that we find ~ 1.4 times as many True Positives with the PRF as the RF.

2.4 Results

From our photometry of the JWST CC field, we retrieve a total of 19 497 objects. Of these, 496 are cYSOs, i.e., they have a probability greater than 50% of being so based on the PRF model. This total includes 474 new cYSOs, of which 133 were identified only by the PRF as they contained missing data and hence could not be classified by the RF model. 310 of the remaining cYSOs are found to be cYSOs with the regular RF with a probability $> 50\%$, and 31 are not found to be cYSOs with the RF to this threshold. In our catalog, we mark the latter 31 as being insecurely classified.

To evaluate the goodness of these results, we address standard machine learning metrics in addition to previously discussed metrics. Figure 2.7 shows the Receiver-Operator Curves (ROC), a standard metric for general machine learning applications, for the identification of cYSOs with a threshold determined by the probability of an object being a YSO. The Area Under the Curve (AUC) here provides the goodness of fit, with an AUC value approaching 1 being an excellent fit. We include this Figure, however, to demonstrate the deceptive nature of the AUC when applied to imbalanced data sets. For example, if almost all objects are placed in a large class, the AUC value will approach 1 regardless of the performance of the much smaller class.

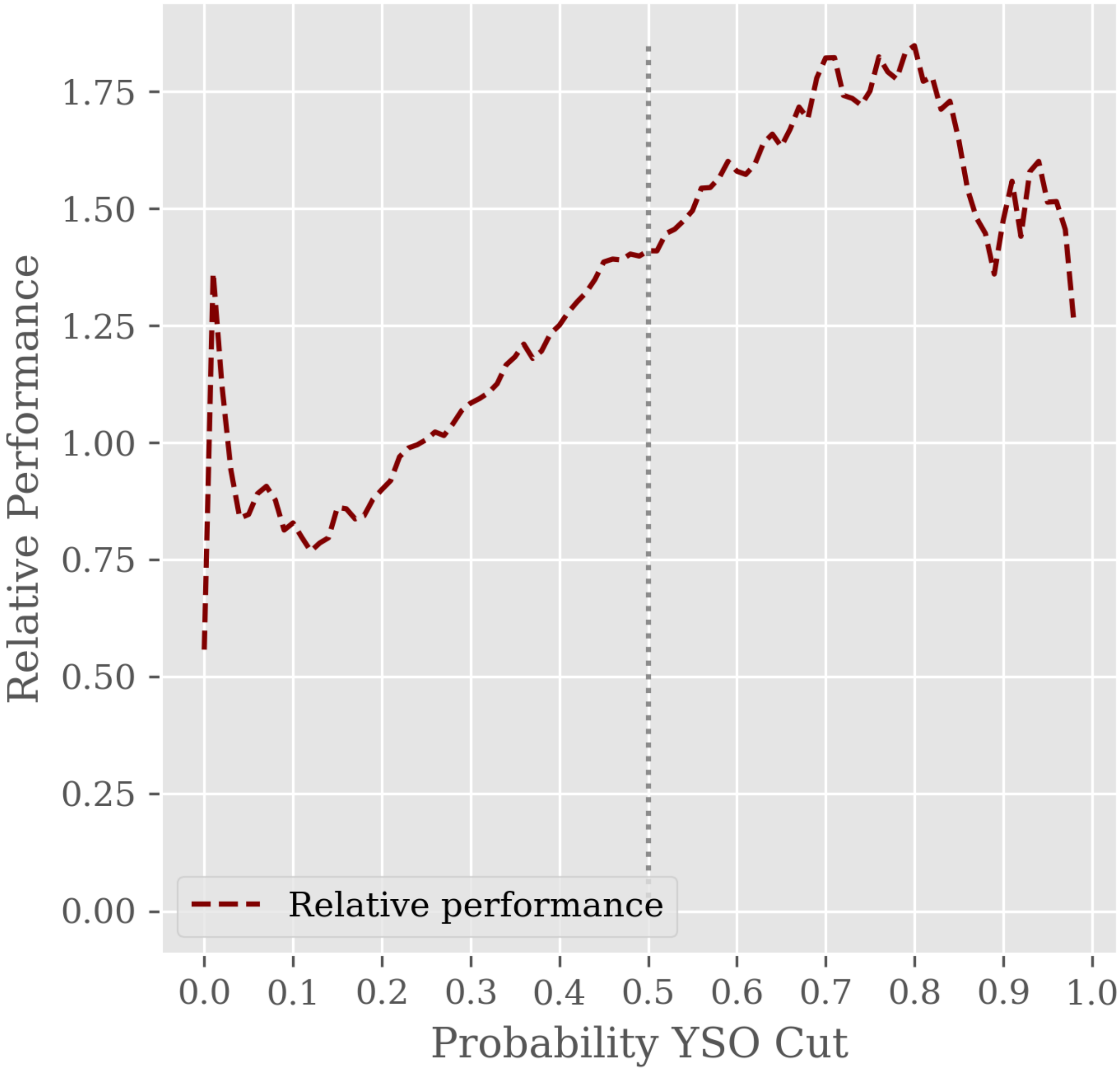

Figure 2.6: The relative performance of the PRF vs the RF models, see text for details on this metric. The vertical gray line marks the 50% threshold line.

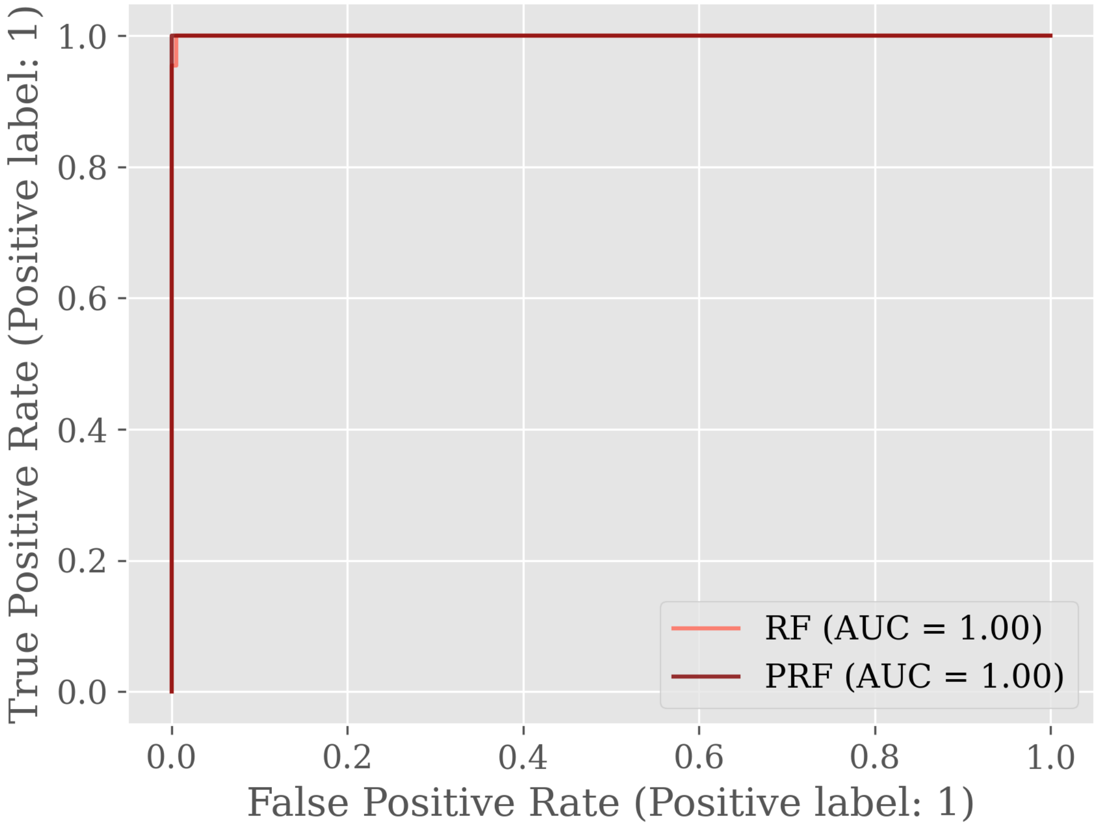

Figure 2.7: ROC curves for each model with threshold based on the probability of an object being a YSO.

Figure 2.8: The confusion matrices for the validation set using a cut of 50% probability for an object to be a YSO based on the RF results (top) and PRF result (bottom).

Figure 2.8 shows the confusion matrices for the validation set using both RF (top) and PRF (bottom) models. The metrics are indeed identical when only considering objects found in all bands, i.e., the PRF and RF perform at the same level when tested on the same data. The original dataset used 19 YSOs and 152 contaminants to build out an augmented dataset. Therefore, both algorithms are correctly classifying the same new data (3 YSOs and 16 contaminants) to the same extent, while the PRF is additionally classifying 5 objects with missing data (1 YSO and 4 contaminants). Both of the RF and PRF methods produce high scores, and so we instead refer to Figure 2.6 to judge the relative performance of these algorithms.

2.4.1 Comparison to Previous Works

We summarize here the similarities and differences between our work and previous studies. In particular, we look at previously identified YSOs to determine if the PRF model was indeed unable to generalize well enough to locate them, or if they were not bona fide YSOs. The former case occurs in three different instances: (1) the YSO is identified by x-ray emission and is likely a Class III YSO, which can be similar to regular field stars in the IR; (2) the YSO is identified by Herschel data in the FIR, and is not identifiable in a significant enough number of bands in the NIR to warrant classification; and (3) the YSO has too many saturated pixels, yielding poor photometric measurements of its flux. The case of a previously identified object not being a bona fide YSO, after all, is caused by one of three options: (1) the source found previously was actually a blending of several sources, resulting in inaccurate SEDs; (2) the source was a reddened galaxy or evolved star rather than a YSO; or (3) the wrong source was identified as a YSO – this last possibility is relevant only for the suggested driving sources of outflows identified by Reiter et al. (2022).

cYSO	Previous Works	Prob	Nearby
J103652.3-583809	HH 1219, SPICY 7434	1.00	-
J103701.5-583751	HH c-3	0.86	-
J103702.1-583658	HH c-4, SPICY 7467	1.00	-
J103653.7-583632	HH c-5	-	0.26
J103647.3-583810	MHO 1634, SPICY 7423, OHL 1	0.85	-
J103651.4-583753	MHO 1636	0.15	-
J103650.5-583752	MHO 1637, OHL 3	1.00	-
J103652.3-583809	MHO 1638	0.02	1.0
J103653.8-583748	MHO 1639, HH 1221, HH 1003 Aa	0.99	-

cYSO	Previous Works	Prob	Nearby
J103651.3-583709	MHO 1640	0.14	0.56
J103654.4-583618	MHO 1645, MHO 1646	0.79	-
J103654.0-583720	MHO 1647, HH 1002a, SPICY 7441	1.00	-
J103653.1-583737	MHO 1650	0.01	0.84
J103653.3-583754	MHO 1651, HH 1003B/Ca, SPICY 7438, OHL 4	1.00	-
J103636.8-583615	OHL 8	1.00	-
J103649.2-583821	OHL 9	1.00	-
J103700.1-583424	OHL	0.94	-
J103648.9-583805	SPICY 7428, OHL 2, X-RAY 27	1.00	-
J103652.5-583725	SPICY 7435, OHL 3	1.00	-
J103656.6-583659	SPICY 7444	1.00	-
J103657.5-583637	SPICY 7447	0.57	-
J103658.4-583619	SPICY 7448	1.00	-
J103659.1-583525	SPICY 7454	1.00	-
J103700.1-583528	SPICY 7461	1.00	-
J103700.3-583830	SPICY 7462	0.99	-
J103700.7-583545	SPICY 7464	1.00	-
J103700.9-583623	SPICY 7465	1.00	-
J103705.7-583419	SPICY 7473	1.00	-
J103706.4-583518	SPICY 7475, OHL 6	1.00	-
J103706.7-583420	SPICY 7476	1.00	-
J103706.9-583655	SPICY 7477, X-RAY 105	1.00	-
J103708.4-583655	SPICY 7479	1.00	-
J103711.3-583445	SPICY 7481, OHL 7	1.00	-
J103711.7-583424	SPICY 7482	1.00	-
J103652.4-583627	X-RAY 34	-	0.99
J103653.4-584028	X-RAY 35	0.18	0.51
J103653.6-583634	X-RAY 36	0.00	0.91
J103653.9-584055	X-RAY 39	0.53	-
J103654.8-583733	X-RAY 40	-	0.84
J103658.5-583500	X-RAY 51	0.02	0.76
J103658.6-583516	X-RAY 53	0.14	0.81
J103658.7-583518	X-RAY 56	-	0.62
J103700.3-583830	X-RAY 64	0.00	0.99

cYSO	Previous Works	Prob	Nearby
J103704.1-583500	X-RAY 85	-	0.62
J103705.9-583518	X-RAY 95	-	0.94
J103711.3-583445	X-RAY 134	-	1.00
J103653.9-583629	HH 1223a, SPICY 7440, OHL 5	0.01	-
J103642.3-583804	MHO 1632	0.05	-
J103648.0-583819	MHO 1633	0.00	-
J103646.7-583805	MHO 1635	-	-
J103654.2-583626	MHO 1643	-	0.09
J103653.6-583520	MHO 1649	0.20	-
J103652.7-583805	MHO 1652	0.25	-
J103706.9-583710	OHL 4	0.36	-
J103640.4-584031	X-RAY 11	0.01	-
J103654.0-584032	X-RAY 38	0.01	-
J103659.6-583651	X-RAY 58	0.38	-
J103701.6-583512	X-RAY 73	0.00	-
J103704.8-583625	X-RAY 91	0.00	-
J103706.3-583634	X-RAY 100	0.00	-
J103706.8-583450	X-RAY 104	0.00	-
J103706.9-583710	X-RAY 106	0.36	-
J103708.2-583706	X-RAY 111	0.11	-
J103711.0-583528	X-RAY 129	0.00	-
J103711.1-583508	X-RAY 130	0.00	-

Table 2.3: Candidate YSOs from other works matched to our candidate YSOs. The different catalogue (Ohlendorf et al., 2013; Preibisch et al., 2014; Kuhn et al., 2021) numbers and outflow (Reiter et al., 2022) numbers associated with each object are labelled, along with the probability of the object originally selected as the cYSO (within 0.0001 degrees of the position specified in previous works or matched by eye). If this probability is less than 50%, we also check for nearby (within 0.005 degrees) sources of higher probability due to resolution differences in previous surveys. The J-2000 designation is marked as the object with greater probability in these cases.

In total, 68 YSOs in the CC identified in previous works (Ohlendorf et al., 2013; Preibisch et al., 2014; Kuhn et al., 2021; Reiter et al., 2022) were matched to JWST objects within the FWHMs of their respective observations. These include 28 x-ray emitting sources, of which we expect some overlap, see Section § 2.4.1 for more detail. Of the 68 YSOs, 46 were matched to our cYSOs, leaving 22 objects that require a closer look to understand why the algorithm did not categorize them as cYSOs. Table 2.3 lists all the cYSOs found in this work with matches to YSOs identified previously, along with their PRF-based probabilities.

Given differences in the angular resolution and sensitivity between the JWST data and previous works, we chose as the most likely corresponding source to be the one within the FWHM of the lower resolution instrument. If multiple sources are located within this FWHM, however, the brightest source was nominally selected. After classifying the data, however, we found that nine of the 68 matched sources did not have high probability of being YSOs, though other sources within the original FWHM did. In such cases, we marked the object with greater probability as a 'nearby' possible match in Table 2.3, and updated the J2000 designation of the match to that of the most YSO-like object within the FWHM that we then categorize as a cYSO. After this change, we now have matches to 55 out of 68 cYSOs.

X-ray sources

Preibisch et al. (2014) found 76 x-ray objects within the CC, of which only 26 were matched to JWST objects within the FWHMs of the x-ray sources, including nine cYSOs. A further five objects have cYSOs as their secondary match. This small quantity of matches is unsurprising as x-rays most effectively sample Class III YSOs, which cannot be well identified by NIR colours (e.g., Dunham et al., 2015). These x-ray observations also had a detection limit of $\sim$ 0.8 $M_\odot$ (Preibisch et al., 2014), likely leaving much of the YSO population probed by JWST undetected. Furthermore, x-ray emission also traces other astronomical objects such as fore- and background Galactic field stars, and extragalactic sources (Broos et al., 2011; Preibisch et al., 2014; Feigelson, 2018). Having a low recovery of x-ray objects therefore arguably provides further support that any contamination by such non-YSOs into our selection of cYSOs is low, but also that we are likely under-sampling the more-evolved YSO population with less infrared excess.

IR sources

Of the 22 Spitzer sources found to be YSOs in Ohlendorf et al. (2013), 14 were matched to a JWST object. Most of the YSOs were spread above the H-II region layer or near the H-II region edge, though five were within the H-II region. Of these 14, two were identified as apparent clumps of nebulosity, and we labelled them as contaminants to help our model differentiate clumped gas from YSOs. Eleven out of twelve of the remaining sources were found by the PRF to be cYSOs with at least 50% probability. The one source not found to be a YSO was badly saturated.

All six YSOs in the CC identified by Gaczkowski et al. (2013) from Herschel data were located in the depths of the H-II region, and no point sources were visible in the NIR for matching. Of the nearby point sources available, none were found to match cYSOs with our RF or PRF models. The lack of matches to Herschel data is not unexpected since Herschel sources may simply not have near-infrared counterparts or are hidden by foreground infrared nebulosity. As such, this lack of matches with Herschel-identified YSOs moreover illustrates that we have successfully avoided labelling such dense knots of gas as cYSOs, though the inclusion of Herschel data to the ensemble in future remains a possibility.

For their SPICY catalog, Kuhn et al. (2022) estimated that the contamination rate of YSOs, as defined as the number of False Positives divided by the number of True Positives, is 10%. We ended up removing eight out of 26 objects identified in SPICY from our training set due to blending, or $\sim 30\%$. Of the remaining sources, all were retrieved as cYSOs with our model.

Outflow Driving Source Candidates

Of the 20 outflow driving source candidates identified in Reiter et al. (2022), ten were found to be cYSOs in our PRF-based classifications. Of these, four were marked in the training set, and an additional six were identified by the PRF. Of the remaining 10 candidate driving sources, four did not have satisfactory photometry due to either saturation or sensitivity limits and six were not identified as candidate YSOs in our models to any reasonable threshold.

Our low degree of cYSO matches to outflow driving source candidates found by Reiter et al. (2022) may be due to a handful of factors. Indeed, tracing outflow driving sources may require a more nuanced approach. For example, all Molecular Hydrogen Objects (MHOs) without associated Herbig-Haro (HH) objects identified by Reiter et al. (2022) do not have proper motions, and so the selected driving source candidate is identified as an object with IR-excess lying along the outflow axes. Also, the five sources that we do not identify as cYSOs may be false negatives, i.e., objects that should have been marked as a YSO but were not possibly due to over-fitting. Meanwhile, four other driving source candidates were found to be unlikely to be YSOs, and we have suggested alternatives. We describe the possible driving sources for each case in the Appendix C.

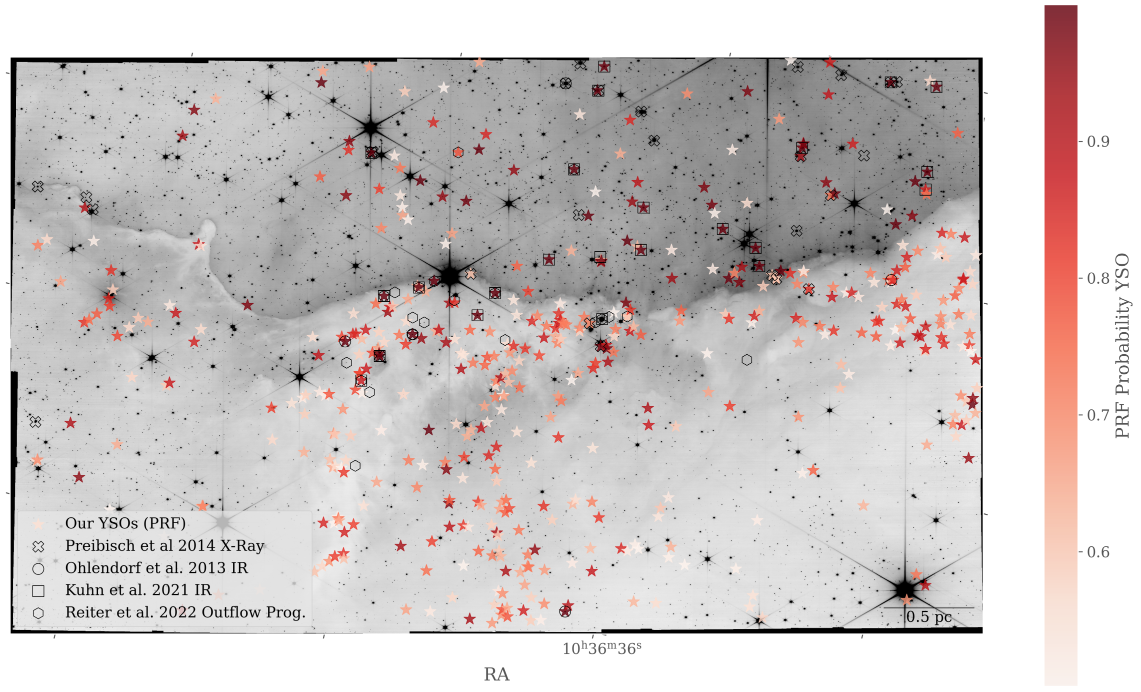

Figure 2.9: All cYSOs in this region matched to Webb sources from Preibisch et al. (2014) (black X's), Ohlendorf et al. (2013) (black circles), Kuhn et al. (2021) (black squares), Reiter et al. (2022) (black hexagons), and with our own cYSOs marked in varying shads of red based on their probability of being a YSO.

2.5 Discussion

This work aims to provide classifications of cYSOs within the ERO JWST data of NGC 3324. There are 19 497 objects within the JWST fields of $\sim 7.4' \times 4.4'$ by NIRCam. Of these, 496 were identified as cYSOs, and the remainder as contaminants. We find that the PRF and RF models generally identify the same objects as YSOs when data in all bands are available. The PRF was additionally able to retrieve 133 further cYSOs that did not have data in all bands, for a total of 474 newly identified cYSOs.

2.5.1 CMDs and CCDs

YSOs have often been found based on their colour-excess as determined from color-magnitude diagrams (CMDs) (e.g., Gutermuth et al., 2009). As a check to see if the YSOs we find show excess, we plot several different CMDs as well as color-color diagrams (CCDs) in Figure 2.10. In general, the F090W-F444W color on the x-axes serves as a proxy for the temperature of the source. The marked cYSOs show populations that remain on the red edge of the CMD

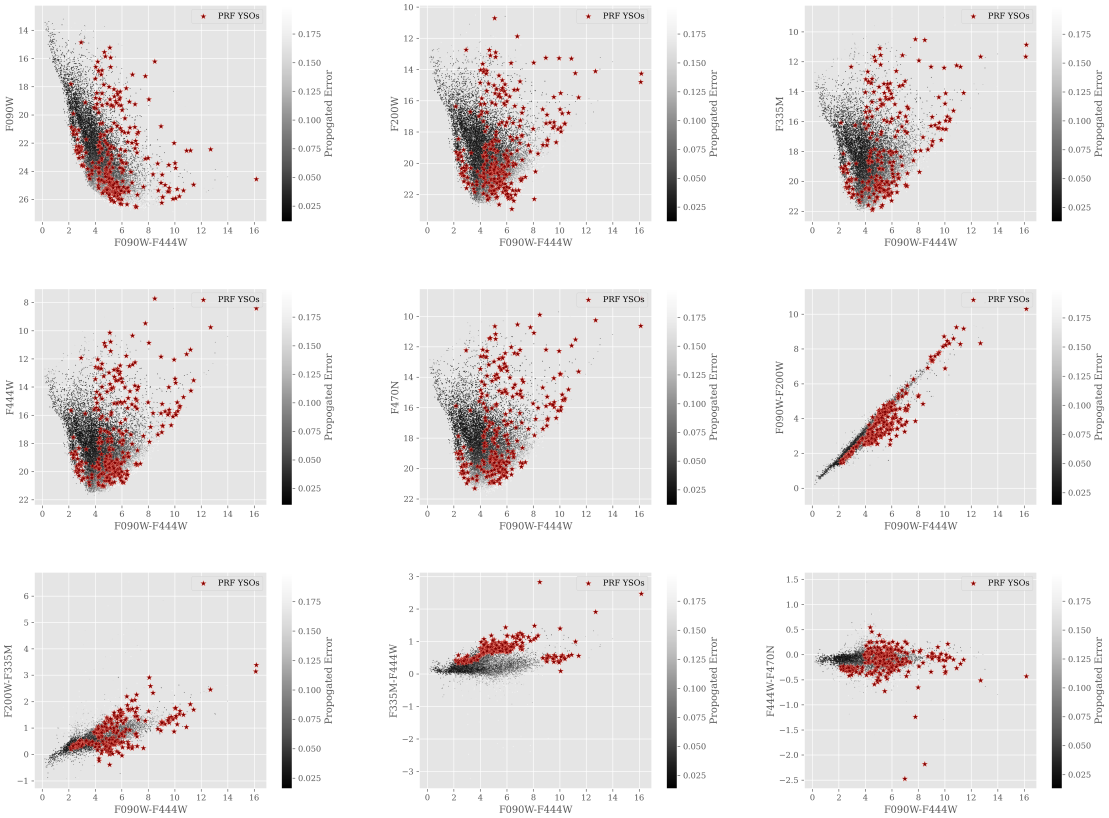

Figure 2.10: Color-magnitude and color-color diagrams of cYSOs (maroon stars) vs. contaminants (gray circles with color gradient specifying the error of the photometry).

distributions, while often times following a sharp angle from the main distribution in the CCD distributions. Spitzer-detected sources are located in the brighter regions of the CMDs ($\lessapprox$ 14 mag), due to the lower sensitivity of Spitzer. The removal of magnitudes and the use of only colors in our training sample was hence necessary to avoid biasing our models to only the brightest objects.

From the CMDs we see that nearly all the cYSOs occur to the red side of [F090W]-[F444W] = 4. Indeed the Spitzer objects with measurements in both F090W and F444W also generally fall on this side, though two objects fall to the blue side of this dividing line. Because YSOs are generally classified as the reddened population, we can consider those objects to the blue side of this reddening line to be possible contaminants. There are 38 objects, or 7.7% of the total number of cYSOs on this side, which we take as the lower bound of our contamination rate.

We also utilize the PARSEC evolutionary models (Nguyen et al., 2022) to plot the expected pre-main sequence isochrones for YSOs in the NIRCam filters over a range of foreground extinctions and provide context to the masses of the cYSOs in the CC field. We assume an age of 2 Myr for the pre-main sequence age and solar metallicity and plot in Figure 2.11 isochrones for a range of foreground extinctions from $A_V = 0.5$ to $A_V = 12.5$. We use the extinction law from Cardelli et al. (1989) with $R_V = 5.5$ for dusty regions to determine extinction at different wavelengths. Additionally, it is worth noting that the predicted colours and magnitudes from the PARSEC models are for the central source, and do not include a disk contribution, which can further affect their star-by-star extinction.

The PARSEC pre-main-sequence isochrones only reach down to a mass of 0.09 $M_\odot$ and we find about 73% of cYSOs are fainter than this cutoff over a wide range of extinctions, insinuating that we are seeing a substantial sub-stellar population in the CC field. We note that, although this regime of the CMD is likely to contain extra-galactic sources, these cYSOs are located within the dusty regions of the Cosmic Cliffs, and so are unlikely to be extra-galactic contaminants. Scholz et al. (2022) estimate that below 15 M_{JUP}, brown dwarfs formed by core collapse make up only 0.25% of the population of extremely low-mass objects, while rogue planets make up a more substantial fraction. If so, we may have also detected a substantial number of young rogue planets within our sample. Further observations, however, are necessary to ascertain the nature of the fainter sources in the CC.

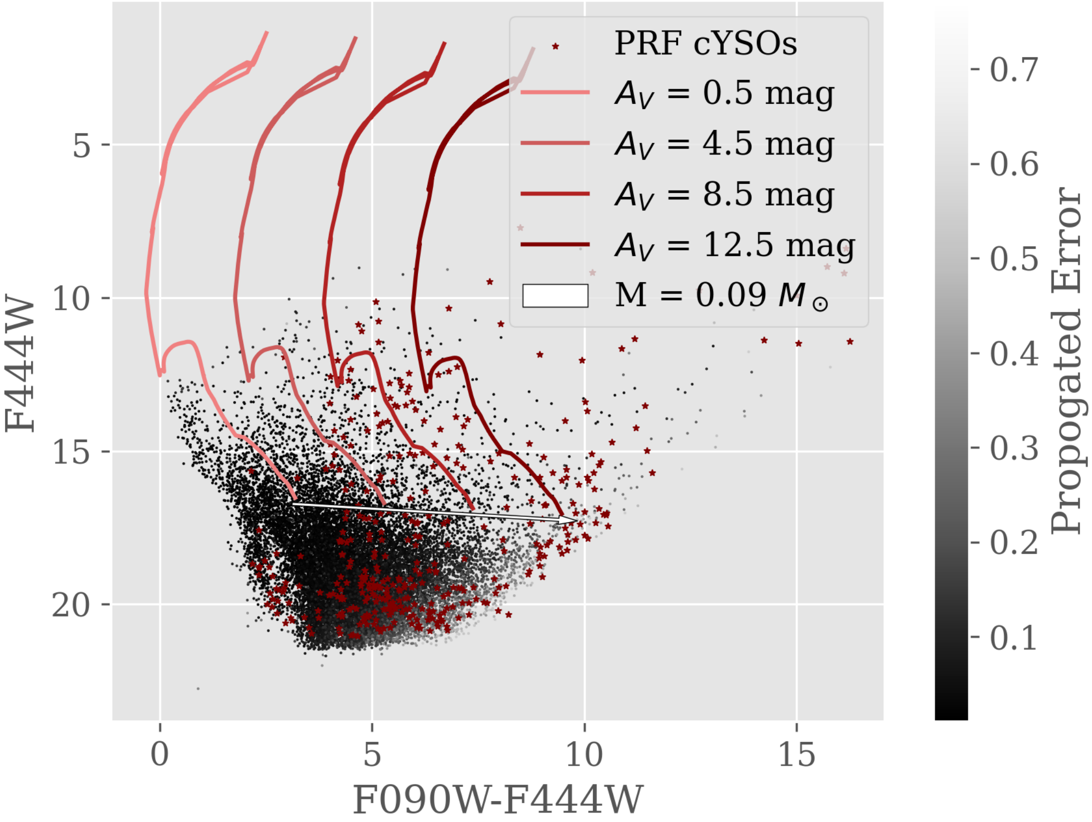

Figure 2.11: Diagram of [F090W-F444W] color vs. F444W magnitude for cYSOs and contaminants (symbols are as in Figure 2.10) overlaid with pre-main-sequence object isochrones from PARSEC models at an age of 2 Myr and various foreground extinctions.

2.5.2 YSO Surface Density

Figure 2.9 shows the locations of all cYSOs in the CC found with our PRF network, as well as all the YSOs found in other works which are matched to JWST sources (Ohlendorf et al., 2013; Preibisch et al., 2014; Kuhn et al., 2021; Reiter et al., 2022). The overall distribution of cYSOs in the field is similar to those found in previous works, with a surprising relative lack of cYSOs found to the south, i.e., on the left-hand side of the image. As may be expected, clustering occurs along the H-II region boundary, with fewer cYSOs within the cleared zone at the top of the image. These latter objects are indicative of a more-evolved YSO population that was left behind as the HII bubble pushed outwards. Table D.1 lists the J2000 coordinates and JWST magnitudes for the identified cYSOs shown in Figure 2.9.

The SPICY catalog lists 14 Class II cYSOs ($\sim 54\%$), five Class I cYSOs ($\sim 19\%$), six flat-spectrum cYSOs ($\sim 23\%$), and one YSO of uncertain type ($\sim 4\%$) within the CC. As the more evolved cYSOs are likely to be found to the west of the edge of the HII bubble (e.g. Gaczkowski et al., 2013), we count those found in our method. Of our 496 cYSOs found, 110, or approximately 21%, are located above the line of the HII bubble, a reasonable lower bound for the number of Class II objects in the sample. Further work will identify the spectral indices of the cYSOs found and allow us to specify their stages similarly. Indeed, the addition of MIRI data may be able to identify more of the dusty, and hence younger, YSO population.

Figure 2.12 shows the surface column density of our cYSOs with respect to the H_2 column density of gas in this region, as obtained with Herschel continuum data using techniques described in Preibisch et al. (2012). We place contours at levels from the minimum column density to the maximum column density that still includes a cYSO. The contour levels increase in steps of $5 \times 10^{20}\ \mathrm{cm^{-2}} \equiv \mathrm{A_V} = 0.5$ mag (e.g., Bohlin et al., 1978). We determine the average nearest neighbour surface density of YSOs in each grid cell, where the grid cell size is matched to the resolution of Herschel at $\sim 9.1 \times 10^{-4}\ \mathrm{pc^2}$ at this distance. We use a Monte Carlo method to incorporate the PRF probability of each of the 19 497 sources in the field being a YSO. The surface density is computed via Equation 2.2, where we choose the 11th nearest object, similar to Pokhrel et al. (2020, 2021). Figure 2.12 then shows the resulting surface density of YSOs, normalized to the total number of objects in each grid cell to highlight regions of YSO enhancement. Indeed we find an enhancement in the lower central region of the CC, uncorrelated with local enhancements of total objects.

$$\Sigma_* = \frac{\mathrm{n}-1}{\pi \mathrm{d}_\mathrm{n}^2}, \qquad \mathrm{n} = 11 \tag{2.2}$$

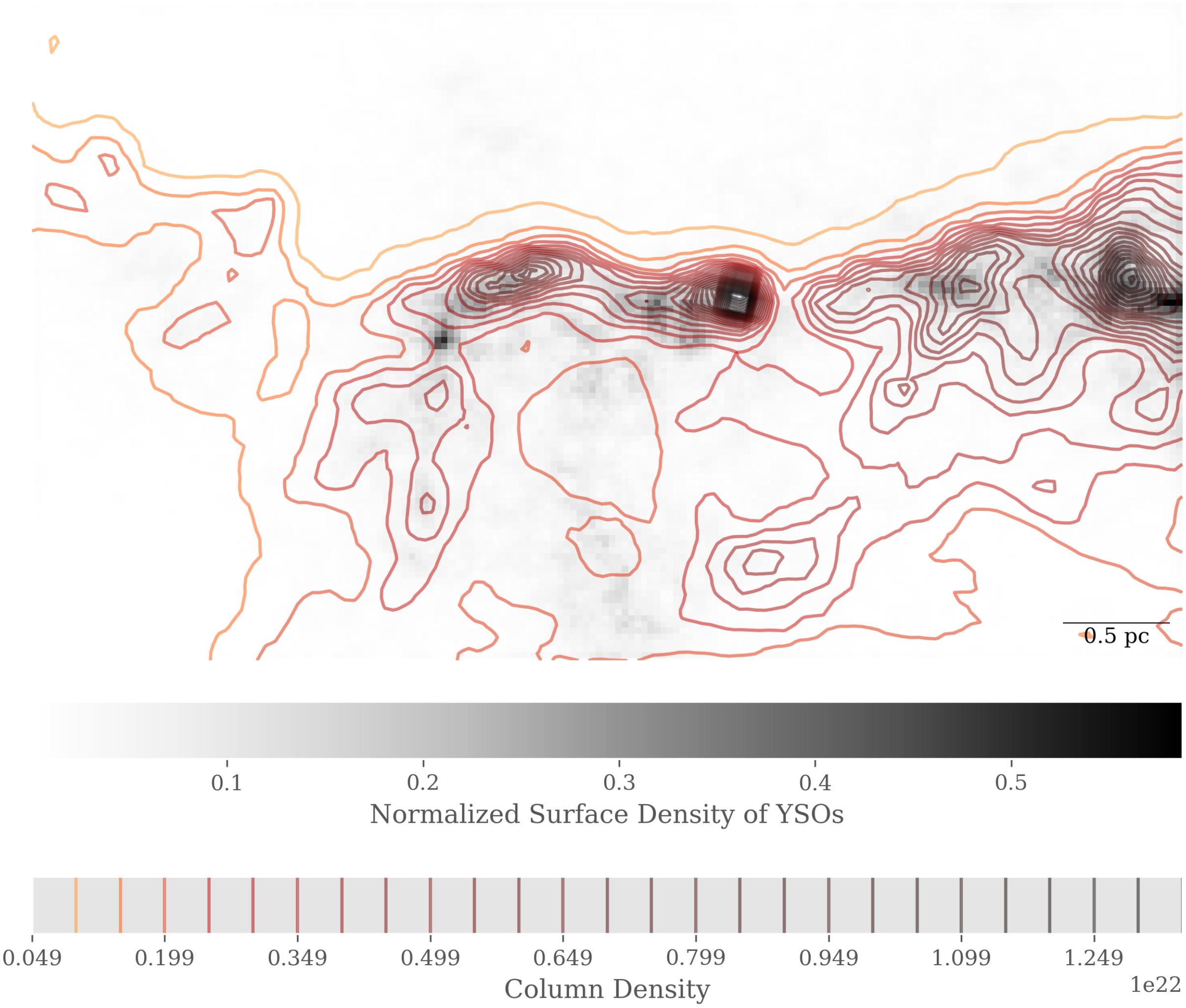

Figure 2.12: Binned surface density with column density contours overlaid. The surface density is normalized to the total number of objects in that bin to highlight regions of YSO enhancement. The column density contours are laid in steps of $A_V = 0.5$ mag.

The identified cYSOs surface density clearly traces out the areas of high column density along the condensed H-II region edge, while the Spitzer-detected YSOs (Ohlendorf et al., 2013; Kuhn et al., 2021) are relegated to the less dense areas within the bubble or along the outer edges of the H-II region. These results are reminiscent of Gaczkowski et al. (2013), where Herschel data identified the youngest sources along the H-II region boundaries, while the slightly more-evolved YSOs, perhaps from a previous star formation episode, are located along the H-II region edges. The strong correlation between our YSO surface density and column density provides further support that the PRF model has successfully found YSOs in general. The location of these sources relative to column density also allows us to claim that JWST performs much better at detecting dusty YSOs than Spitzer could, despite only having access to the NIRCam data (0.9-4.7 μm) at this time.

From the surface density of YSOs, we determine a star formation rate (SFR). Due to the sub-stellar population, we approach this in three ways. First, we assume the average mass of a YSO is 0.5 $M_\odot$ (Evans et al., 2009), and remove the sub-stellar component from the calculation. Second, we assume an average mass of 0.5 $M_\odot$ for those objects above the sub-stellar line, and include the sub-stellar component assuming 10% of this mass, i.e. 0.05 $M_\odot$ for the sub-stellar component. Third, we assume a constant average mass of 0.5 $M_\odot$ across the whole population, including the sub-stellar component. This will likely lead to an overestimate of the SFR. In all cases we assume a mean lifetime of 2 Myr for a Class II YSO (Evans et al., 2009). We assume that the majority of our cYSOs will be Class II for this calculation as that stage has a longer duration relative to other, earlier stages and as yet we do not know the specific stage of each object. Due to the low recall of x-ray sources in the JWST data, we assume that we do not heavily sample Class III YSOs in our population. Furthermore, we determine a star-forming efficiency, ε_{SF}, for the CC region via Equation 2.3, i.e., by dividing the estimated mass of cYSOs by the total gas mass in the CC field. N_{PS} is the number of protostars, M_{PS} is the mass of a protostar, m_H is the mass of a hydrogen atom, and $X = 0.71$ is the fraction of hydrogen in the local interstellar medium (Nieva & Przybilla, 2012). We tabulate our SFR and SFE values in each of the aforementioned cases in Table 2.4.

$$\varepsilon_{SF} = \frac{M_*}{M_{gas}} = \frac{N_{PS} \times M_{PS}}{(2\ m_H/X)\ N(H_2)} \tag{2.3}$$

We find that the SFR and SFE values vary significantly in each of the three cases. The inclusion of the sub-stellar component greatly impacts the SFR and SFE, while the SFE is closer to that expected from previous surveys when the sub-stellar component is ignored. This is expected as previous instruments have been unable to identify the sub-stellar

Case	SFR ($M_\odot$/yr)	SFE
No sub-stellar	2.6-3.1 $\times 10^{-5}$	0.08-0.09
Mass correction	5.1-6.7 $\times 10^{-5}$	0.15-0.20
No mass correction	9.3-12.4 $\times$ 10^{-5}	0.28-0.37

Table 2.4: The star formation rates and efficiencies for three separate cases. In the first case, the sub-stellar population is removed from the sample. The second assumes that the sub-stellar component has a mass 10% of the regular YSOs. Finally, in the third the sub-stellar population is included but the average mass is assumed to be 0.5 $M_\odot$ over the whole population. The range in each case is given by the surface density calculations (lower bound) and the number of YSOs identified with the PRF (upper bound).

Relation	Slope	Intercept
Pokhrel	2.00	-4.11
Surface Density	0.62±0.09	–0.75 ± 0.17
PRF YSO count	1.52±0.28	–2.22 ± 0.55

Table 2.5: The slope values for the relations shown in Figure 2.13.

components of star-forming region due to their lower resolution and sensitivity. Furthermore, these SFRs are much lower than previous estimates of the entire Carina Nebula Complex: 1.0-1.7 × 10^{-2} $M_{\odot}$/yr of the central region (Povich et al., 2011), or 1.7 × 10^{-2} $M_{\odot}$/yr of the full region (Gaczkowski et al., 2013). This difference is not surprising as we are only sampling a small population over a relatively minor star-forming area, away from the more active star-forming locations in the Carina Nebula Complex.

Finally, we can analyze the relationship between star formation rate surface density (SFR per unit area) and gas surface density (M_{gas} per unit area). We choose to look at the case with no sub-stellar component for comparison to previous works, and find that this relation is slightly shallower than the relationship found in nearby clouds based solely on Class 0/I sources identified with Spitzer (Pokhrel et al., 2021). Inclusion of the sub-stellar component further decreases the slope. Such flattening of the slope can be expected with a sampled population of primarily Class II YSOs that have migrated over their ∼2 Myr lifetimes from their high column density birth sites to adjacent regions of lower column density. The slope fits for the plotted relations are in Table 2.5. Errors on the SFR surface density are determined by the error on the number of YSOs used in the SFR calculation. Errors on the assumed lifetime and mass of the YSOs (leading to about a 50% error; Dunham et al., 2015) and on the mass of the gas (about a 10% error; Preibisch et al., 2012; Könyves et al., 2015) are ignored as they are systematic effects and as such would not affect the slope.

2.6 Conclusions

The PRF model, though simple in its methodology for dealing with missing data, has proven robust under several different metrics for testing (AUC, Confusion Matrices, F1-Score, Recall Score, Precision Score, etc.). Additional benefits of the PRF are that it is both fast (in comparison to traditional reddening cuts and fits to SEDs), and easy to implement (by being written in the popular coding language of Python). Indeed, this latter benefit was part of the reason for choosing the PRF instead of more involved ML methods that impute the data, such as Ishwaran & Kogalur (2023) available in R. Furthermore, since the PRF method

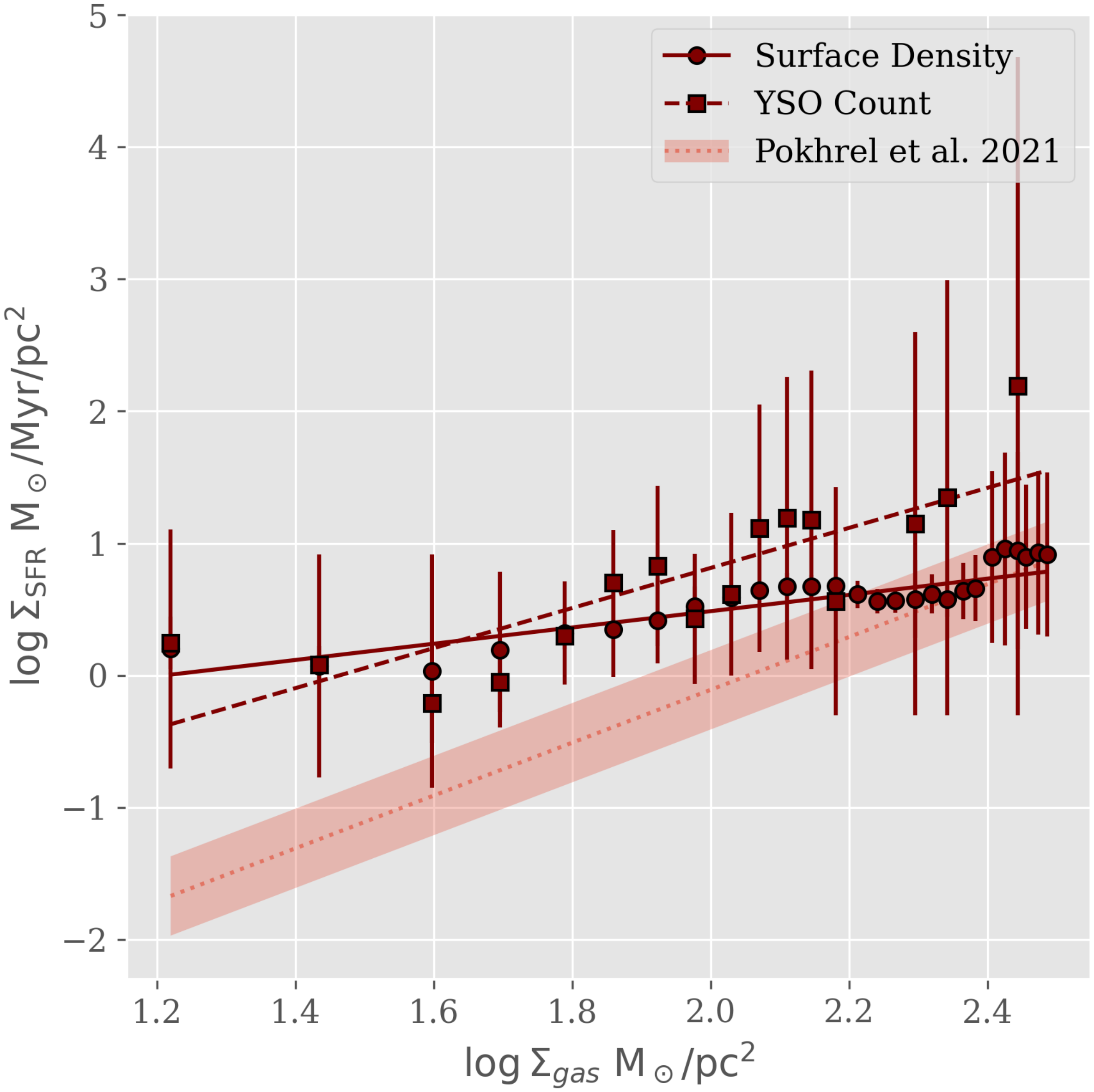

Figure 2.13: SFR surface density vs column density for each contour level in Figure 2.12. SFR measurements are calculated both by nearest neighbours surface density and by the number of cYSOs with probability greater than 50% found with the PRF. We add in the relationship of Pokhrel et al. (2021) for comparison. These relationships exclude the substellar component.

outputs the probabilities of an object being a YSO, it allows for Monte-Carlo sampling of the population of objects, and leaves open the option of changing the cYSO probability threshold as needed.

We find that the PRF method is a promising approach for accurate YSO identification from JWST data, especially as we wait for more data. The method is particularly promising as it is able to determine relationships well based only on NIRCam data, which, based on the field of views of the relevant detectors, will often cover much more of the field than MIRI data. The PRF was able to recover previously identified YSOs to 98% F1-Score and retrieved 496 cYSOs from the entire set.

Chapter 3

Future Work

This Chapter outlines future work that will build off the previous chapter, some of which will become the basis of my PhD dissertation. Foremost is an analysis of the MIRI data of the Cosmic Cliffs field, Section 3.1, after which the methodology provided here can be applied to data from other regions, summarized in Section 3.2. This analysis, though useful, would be subject to whatever fields are selected to be observed by JWST through a number of unrelated projects. For a more targeted approach, we will submit a proposal with JWST to image Monoceros R2, Section 3.3, which has the further potential to serve as the pilot for a larger survey of compact star-forming regions. Finally, we summarize other compelling science that could be gleaned from the data of the CC that have not yet been explored, Section 3.4.

3.1 Analyzing MIRI Data

MIRI data cover the wavelength range of 5-28 μm, the MIR region useful for distinguishing YSO stage. Indeed, these wavelengths also allow MIRI to probe high column densities where NIRCam's sensitivity drops off. For the CC, the MIRI data are complementary to the NIRCam data, though they have a smaller field of view, observing in particular the regions of highest column density along the ridge. For comparison, NIRCam observed a $129'' \times 129''$ field of view (2.2 pc^2 at the distance of NGC 3324), while MIRI observed a $74'' \times 133''$ field of view (1.3 pc^2 at the distance of NGC 3324).

The MIRI data of the Cosmic Cliffs were not included in the earlier analysis due to misalignments in the final data products produced by MAST. These misalignments resulted in incorrect photometric calibration and source matching, which forced us to abandon the data due to time constraints on the submission of this thesis. Future work will begin with reprocessing of the raw data, with the Space Telescope Help Desk suggesting the answer may lie in the `tweakreg` step of the JWST pipeline. Once NIRCam objects are matched to their MIRI counterparts we can find out what stage of star formation they are in. This further classification can be determined one of two ways: (1) via training a new ML model to determine the specific classes, much in the way that Miettinen (2018) did; or (2) via CMD criteria (see Section 1.1 for details) and the determination of spectral indices.

For the first option, Kuhn et al. (2021) provided the YSO stage classifications with their network, where the stages were determined via cuts in spectral index after YSO classification. We could use the classifications mapped onto our training set to then further classify all of the YSOs into their stages. We could implement a multi-class classification model, as in Miettinen (2018), or a binary classification where first the star formation phases are split, i.e., Class II/III YSOs (pre-main sequence phase) vs. Class 0/I YSOs (protostellar phase), and then where those classes are further split into their individual stages, akin to Marton et al. (2016). This option does not rely on us using spectral indices similar to those derived from Spitzer data for further classification.

The second option, using cuts in colour, is a simpler technique. In CMDs, YSOs can be identified by their red-excess caused by the dust in their disks and envelopes absorbing shorter wavelength light and re-emitting at longer wavelengths. Hence, cuts in red colour on CMDs can separate out the YSO population, and further cuts can separate out the distinct stages, (e.g., Gutermuth et al., 2008, 2009), though accurate class identification can be affected by the inclination angle of the disk which may cause some Class II sources to be misclassified as Class I sources. Colours too can be combined with the spectral index, Equation 3.1, to determine the precise stage, and is the basis for the classification of SPICY YSOs into stages. This metric is used because YSOs do not radiate at a precise temperature, but rather over the range defined by their circumstellar dust distributions (Lada & Wilking, 1984; Lada, 1991; Barsony, 1994). The caveat of this latter method is that extinction effects must be known and corrected for, and so additional data are needed to separate the stages accurately. In particular, the stages and associated spectral indices (corrected for extinction) used in Dunham et al. (2015) are:

Class 0+I: $\alpha \geq 0.3$

Class FS: $\alpha < 0.3$, $\alpha \geq -0.3$

Class II: $\alpha < -0.3$, $\alpha \geq -1.6$

Class III: $\alpha < -1.6$

Region	Release Date
Cosmic Cliffs/NGC 3324	July 12th, 2022
Tarantula Nebula/NGC 2070	September 6th, 2022
Orion Bar (Segment of Trapezium)	October 7th, 2022
Eagle Nebula/M16	October 19th, 2022
The Brick/G0.253+0.015	September 20th, 2023
Orion Trapezium Cluster	October 3rd, 2023
NGC 346	October 11th, 2023
Oph A	August 25th, 2024

Table 3.1: Notable star-forming regions imaged by JWST known as of September 6th, 2023 along with their public release date. If some data were released earlier but more data are set to be released, then the latter date is specified.

where

$$\alpha = \frac{\mathrm{d}\log(F_\lambda \lambda)}{\mathrm{d}\log(\lambda)}, \qquad 2.2\ \mu\mathrm{m} < \lambda < 20\ \mu\mathrm{m} \tag{3.1}$$

Given that the spectral index requires a wide range of wavelengths to most appropriately identify YSO stage, MIRI data are again needed for this analysis. Note: for Spitzer this index was often calculated between the IRAC and MIPS bands over 4.5-24 μm, or $4.5 - 8$ μm if the 24 μm MIPS band data were not available. For the CC, the MIRI data include the 7.7 μm and 18 μm bands as the nearest analogues to those Spitzer bands.

These identifications are important for the precise quantification of the star formation relation to gas surface density. In particular, Class 0 and Class I sources are best for constraining the measurements of the star formation rate to within 0.2 Myr, as done in Pokhrel et al. (2021). The SFR we calculated in the previous chapter was averaged over a longer period (2 Myr) due to the inclusion of Class II sources as we could not yet distinguish them from younger classes. Hence, obtaining stage classifications of our cYSOs will allow for accurate identification of the “instantaneous” star formation relation in the CC.

3.2 Other Star-Forming Regions

NGC 3324 provided an intriguing subject for this proof of concept study, but it was originally imaged more for the aesthetics of its sharp ionization boundary and its favourable alignment with the aperture angle of NIRCam and MIRI at the time than specifically for its star formation potential (Pontoppidan et al., 2022). Although there are a significant number of YSOs in the NIRCam field, it is neither the most active star-forming location in the Carina Nebula Complex, nor is it the densest. Several other star-forming regions have been more recently imaged by JWST during its first year of observations with their respective data becoming publicly available soon, see Table 3.1.

Due to the algorithmic methodology of this work, we can apply the PRF model and analyze each of these JWST-observed regions now reasonably quickly by following the same steps. Indeed, photometry can be similarly performed following the steps in Appendix A, after which training YSOs can be identified from previous works, and the rest of the analysis is automated based on the available bands. For instance, M16, specifically the "Pillar of Creation", Figure 3.1, is another star-forming region imaged with the same JWST NIRCam bands as in the CC and it has a similar number of previously classified cYSOs in SPICY. Hence, the analysis of this region merely needs photometry before it can be completed, and a cross comparison between it and the Cosmic Cliffs can be made. Indeed, Cornu & Montillaud (2021) similarly show how training on one region and testing on another can aid in network generalization.

We utilized colours as input for the PRF in the classification of the CC as the magnitude of an object in a specific band is less important than the change in brightness between two bands. This approach allowed us to use Spitzer-detected objects to build our training set without biasing our results to the brightest objects. Furthermore, we can generalize our results from one region to the next by using the NIRCam bands used only in the analysis of NGC 3324. With the addition of more data in similar bands across different fields, we can expand our training set as the YSOs should occupy similar regions of colour-colour space. In particular, NGC 3324, NGC 2070, M16, and NGC 346 all have JWST imaging data in the F090W, F200W, F335M, F444W, F770W, and F1130W bands, where the first four bands were utilized in the analysis of NGC 3324 in Chapter 2. We can also expand on the colours used for different regions, as each region has slightly different bands available. From this variety, we can eventually determine which bands are the most important for classification from our network, where importance is defined as those bands which help the most to distinguish contaminants from YSOs.

Overall, we are interested in the link between the production of YSOs in molecular clouds and their physical environments, particularly in dense gas clouds. We therefore need accurate YSO counts across a variety of fields to probe how environment impacts star formation.

3.3 Monoceros R2 Giant Molecular Cloud

The analysis of the publicly available JWST datasets will allow for a more general analysis of the SFR and efficiency within the Galaxy. Only small clumps within these regions, however,

Figure 3.1: The publicly released image of M16 as taken by JWST. This image is the composition of six NIRCam filters. Purple: F090W, Blue: F187N, Cyan: F200W, Yellow: F335M, Orange: F444W, Red: F470N. Image credit: NASA, ESA, CSA, and STScI.

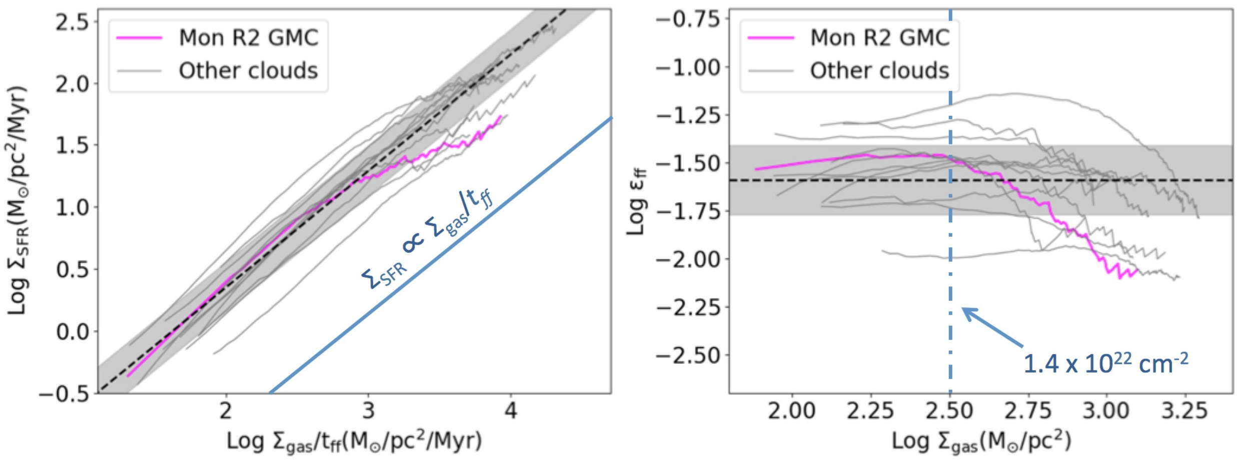

Figure 3.2: The SFR – Σ_{gas}/t_{ff} relations from Pokhrel et al. (2021), highlighting Mon R2. This cloud has the largest departure from the relations, with a clear dip in SFE after a gas column density of 1.4×10^{22} cm^{-2}.

have column density greater than 1.4×10^{22} cm^{-2} (Tremblin et al., 2013; Tatton et al., 2013), where the SFR – Σ_{gas}/t_{ff} relation from Pokhrel et al. (2021) falls off. Indeed, it is in regions denser than this that JWST is primed to provide the most interesting insights, as it is where Spitzer's sensitivity and resolution were no longer able to easily identify the most embedded, clustered Class 0/I YSOs.

To determine if the drop-off of star-forming efficiency at higher column densities is an artifact of Spitzer's observing capabilities in high column density or clustered regions, or if instead it is caused by feedback from closely spaced YSOs suppressing further star formation, we must obtain higher sensitivity and higher resolution IR observations of YSOs in dense and clustered regions. As such, we are in the process of preparing a Cycle 3 JWST proposal to image the active star-forming region known as the Monoceros R2 Giant Molecular Cloud. Monoceros R2, or Mon R2 for short, was chosen due to its high gas surface density and extremely clustered nature. It was also analyzed in the work by Pokhrel et al. (2021), where it had the greatest departure from their "universal" star formation relation, the relation between SFR and gas surface density, of all 12 nearby ($<$ 300 pc) clouds they studied, Figure 3.2. Furthermore, Mon R2 has been surveyed previously using several instruments, including Spitzer, Herschel, and ALMA, of which archival data are available for adjacent analyses.

The main caveat of JWST instruments is their relatively small field of view. This drawback can be mitigated by looking at the most compelling and richest parts of Mon R2. We will thus propose to look at the central $6' \times 6'$ region of the Mon R2 cluster, where three

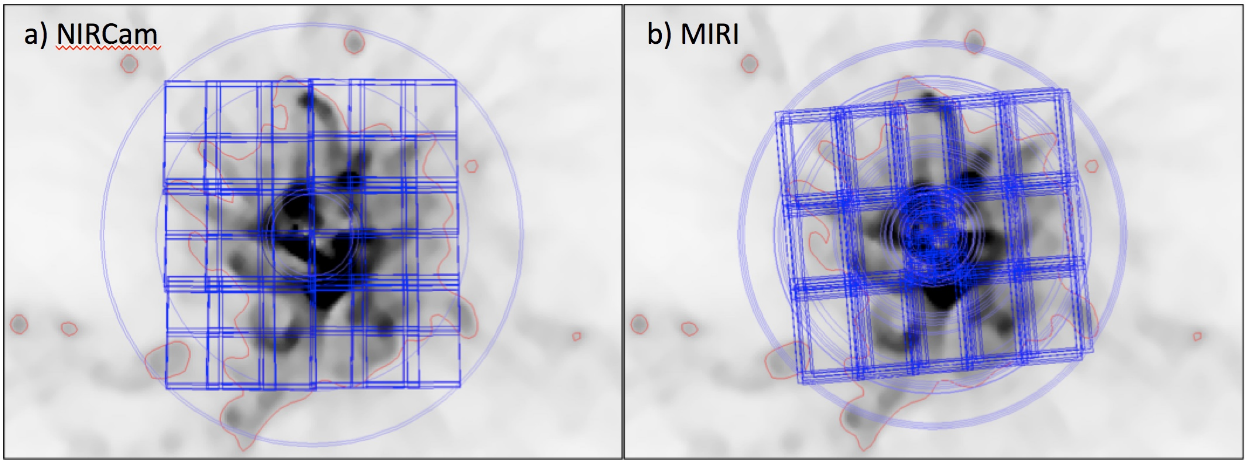

Figure 3.3: The central $6' \times 6'$ region of the Mon R2 along with the NIRCam (a) and MIRI (b) mosaics. The greyscale shows the column density data (recieved via private communication from Dr. Philippe André), and the red contour marks $1.4 \times 10^{22}\mathrm{cm}^{-2}$.

separate filaments conjoin. Indeed, this region contains 42% of the total known number of YSOs in the full region, including 53 Class I objects and 304 Class II objects, a factor of ~ 14 greater than the number of YSOs of previously known class in NGC 3324. The central region also has densities in the range of $\sim 10^{22-23}$ cm^{-2}, i.e. where the star-forming efficiencies appear to drop off in the study by Pokhrel et al. (2021), making it an optimal region for this study. With a large number of objects of known class, we should be able to identify objects of different class using machine learning approaches, and not just determine if they are or are not a YSO. Figure 3.3 shows the proposed mosaics in both NIRCam and MIRI, as well as the column densities. In particular, our proposed region primarily covers the densest part of the central region, with column density above 1.4×10^{22} cm^{-2}.

The JWST Mon R2 project can also function as a pilot study for the targeted identification of YSOs at high column densities, which will allow for the "universality" of the Kennicutt-Schmidt law proposed in Pokhrel et al. (2021) to be tested at high column densities in many other Galactic star-forming regions.

3.4 Further Compelling Avenues

Chapter 2 described the acquisition of a population of YSOs within the Cosmic Cliffs, while thus far the current Chapter has considered what more we can learn with additional MIRI data, and moreover how that analysis can be applied to future datasets. We now seek to consider what else can been explored given the data of the Cosmic Cliffs.

First, an in-depth analysis could reveal how well JWST can detect YSOs in comparison to Spitzer. From our case study, we learned that JWST is able to probe down to sub-stellar luminosities, and from the spatial distribution we inferred that it is able to detect embedded objects much more easily than Spitzer could. A quantification of this ability would be useful for estimating how many objects we could expect to see within dense regions on a cloud-by-cloud basis. Furthermore, the percentage of sub-stellar objects must be checked against models of extra-galactic sources to ensure the sample is robust and not a byproduct of contamination.

Once we have the MIRI data, we will also be able to look for red objects detected by MIRI without NIRCam counterparts. If these are co-located with high column density locations, these could be embedded Class 0 or Class I objects. Because our network focused on NIR data (from both Spitzer and JWST), we are likely biasing our detections to Class II sources, which are more readily identified from NIR data. Furthermore, a complementary survey of NGC 3324 with ALMA would be able to detect the presence of disks and compact envelopes to confirm the identification of some of our cYSO population determined in Chapter 2. ALMA will also be able to provide us information about the dense gas within the environment of the cYSOs, and thereby establish further identification of the cYSOs.

We can also ask ourselves some questions on how to best utilize JWST. For instance, is there a combination of bands that work best for detecting YSOs? Of the works mentioned in Section 3.2, several have the same set of NIRCam bands that were used for NGC 3324, though others have a completely different set. These were chosen to show extinction colours from the background stellar field, to image scattered dust, and to look for specific emission features from outflows which trace YSOs (Pontoppidan et al., 2022). Furthermore, what is the best strategy for imaging with JWST? As we found in writing our proposal for Mon R2 and in the analysis of the Cosmic Cliffs, the sensitivity of JWST is remarkable, and yet it comes with the caveat that the instruments onboard saturate quickly, which in turn means the photometric matches to data from previous surveys may shrink due to a smaller overlap in the objects which are not saturated in JWST, and are not too faint, small, or otherwise unresolved in previous surveys. This innate sensitivity leads us to having to choose more exposures and shorter integration times to mitigate the saturation, at the cost of greater observing overhead time, a limiting factor for data acquisition.

This work has utilized the Probabilistic Random Forest Method to identify over an order of magnitude new YSOs with the CC data from JWST, a substantial improvement over results from Spitzer. Furthermore, machine learning is a constantly growing and adapting field, and new methods may be able to constrain the YSO population even better. Machine learning too will allow us to incorporate data from previous studies with other telescopes such as Herschel, ALMA, SKA, etc., further increasing the diversity of the data available to aid in classification. This work, however, has served as the casting off point into a promising sea of future star formation studies with JWST.

Bibliography

Adams, F. C., Lada, C. J., & Shu, F. H. 1987, ApJ, 312, 788, doi: 10.1086/164924

Allen, L. E., Calvet, N., D'Alessio, P., et al. 2004, ApJS, 154, 363, doi: 10.1086/422715

André, P. 2015, in Encyclopedia of Astrobiology, ed. M. Gargaud, W. M. Irvine, R. Amils, I. Cleaves, Henderson James (Jim), D. L. Pinti, J. C. Quintanilla, D. Rouan, T. Spohn, S. Tirard, & M. Viso, 2308–2313, doi: 10.1007/978-3-662-44185-5_504

Barsony, M. 1994, in Astronomical Society of the Pacific Conference Series, Vol. 65, Clouds, Cores, and Low Mass Stars, ed. D. P. Clemens & R. Barvainis, 197

Bisht, D., Zhu, Q., Yadav, R. K. S., et al. 2021, MNRAS, 503, 5929, doi: 10.1093/mnras/stab691

Bohlin, R. C., Savage, B. D., & Drake, J. F. 1978, ApJ, 224, 132, doi: 10.1086/156357

Breiman, L. 2001, Machine Learning, 45, 5, doi: 10.1023/A:1010933404324

Broos, P. S., Getman, K. V., Povich, M. S., et al. 2011, ApJS, 194, 4, doi: 10.1088/0067-0049/194/1/4

Cappa, C., Niemela, V. S., Amorín, R., & Vasquez, J. 2008, A&A, 477, 173, doi: 10.1051/0004-6361:20067028

Cardelli, J. A., Clayton, G. C., & Mathis, J. S. 1989, ApJ, 345, 245, doi: 10.1086/167900

Chiu, Y. L., Ho, C. T., Wang, D. W., & Lai, S. P. 2021, Astronomy and Computing, 36, 100470, doi: 10.1016/j.ascom.2021.100470

Cornu, D., & Montillaud, J. 2021, A&A, 647, A116, doi: 10.1051/0004-6361/202038516

Cortes, C., & Vapnik, V. 1995, Machine Learning, 20, 273, doi: 10.1007/BF00994018

Dauphas, N., & Chaussidon, M. 2011, Annual Review of Earth and Planetary Sciences, 39, 351, doi: 10.1146/annurev-earth-040610-133428

Dunham, M. M., Allen, L. E., Evans, Neal J., I., et al. 2015, ApJS, 220, 11, doi: 10.1088/0067-0049/220/1/11

Evans, N. J., I., Allen, L. E., Blake, G. A., et al. 2014a, VizieR Online Data Catalog, II/332

—. 2014b, VizieR Online Data Catalog, II/332

Evans, Neal J., I., Dunham, M. M., Jørgensen, J. K., et al. 2009, ApJS, 181, 321, doi: 10.1088/0067-0049/181/2/321

Feigelson, E. D. 2018, Multiwavelength Studies of Young OB Associations, ed. S. Stahler (Cham: Springer International Publishing), 119–141, doi: 10.1007/978-3-319-22801-3_5

Fix, E., & Hodges, J. L. 1951, Non-Parametric Discrimination: Consistency Properties

Friedman, J. 2002, Computational Statistics & Data Analysis, 38, 367, doi: 10.1016/S0167-9473(01)00065-2

Friedman, J. H. 2001, Annals of statistics, 1189

Fukushima, K. 1980, Biological cybernetics, 36, 193

Furlan, E., Fischer, W. J., Ali, B., et al. 2016, ApJS, 224, 5, doi: 10.3847/0067-0049/224/1/5

Gaczkowski, B., Preibisch, T., Ratzka, T., et al. 2013, A&A, 549, A67, doi: 10.1051/0004-6361/201219836

Gao, Y., & Solomon, P. M. 2004, ApJ, 606, 271, doi: 10.1086/382999

Göppl, C., & Preibisch, T. 2022, A&A, 660, A11, doi: 10.1051/0004-6361/202142576

Gutermuth, R. A., Megeath, S. T., Myers, P. C., et al. 2009, ApJS, 184, 18, doi: 10.1088/0067-0049/184/1/18

Gutermuth, R. A., Myers, P. C., Megeath, S. T., et al. 2008, ApJ, 674, 336, doi: 10.1086/524722

Hinton, G. E., & Roweis, S. 2002, Advances in neural information processing systems, 15

Ho, T. K. 1995, in Proceedings of 3rd international conference on document analysis and recognition, Vol. 1, IEEE, 278–282

Ishwaran, H., & Kogalur, U. 2023, Fast Unified Random Forests for Survival, Regression, and Classification (RF-SRC). https://cran.r-project.org/package=randomForestSRC

Kennicutt, Robert C., J. 1998, ApJ, 498, 541, doi: 10.1086/305588

Kinson, D. A., Oliveira, J. M., & van Loon, J. T. 2021, MNRAS, 507, 5106, doi: 10.1093/mnras/stab2386

—. 2022, MNRAS, 517, 140, doi: 10.1093/mnras/stac2692

Könyves, V., André, P., Men'Shchikov, A., et al. 2015, Astronomy & Astrophysics, 584, A91

Krumholz, M. R., & McKee, C. F. 2005, ApJ, 630, 250, doi: 10.1086/431734

Krumholz, M. R., McKee, C. F., & Bland-Hawthorn, J. 2019, ARA&A, 57, 227, doi: 10.1146/annurev-astro-091918-104430

Kudritzki, R. P., Puls, J., Lennon, D. J., et al. 1999, A&A, 350, 970, doi: 10.48550/arXiv.astro-ph/9910449

Kuhn, M. A., de Souza, R. S., Krone-Martins, A., et al. 2021, ApJS, 254, 33, doi: 10.3847/1538-4365/abe465

Kuhn, M. A., Saber, R., Povich, M. S., et al. 2022, AJ, 165, 3, doi: 10.3847/1538-3881/ac9314

Lada, C. J. 1991, in NATO Advanced Study Institute (ASI) Series C, Vol. 342, The Physics of Star Formation and Early Stellar Evolution, ed. C. J. Lada & N. D. Kylafis, 329, doi: 10.1007/978-94-011-3642-6_9

Lada, C. J., Lombardi, M., & Alves, J. F. 2010, ApJ, 724, 687, doi: 10.1088/0004-637X/724/1/687

Lada, C. J., & Wilking, B. A. 1984, ApJ, 287, 610, doi: 10.1086/162719

Lenkić, L., Nally, C., Jones, O. C., et al. 2023, arXiv e-prints, arXiv:2307.15704, doi: 10.48550/arXiv.2307.15704

Majewski, S., Babler, B., Churchwell, E., et al. 2007, Galactic Structure and Star Formation in Vela-Carina, Spitzer Proposal ID 40791

Marton, G., Tóth, L. V., Paladini, R., et al. 2016, MNRAS, 458, 3479, doi: 10.1093/mnras/stw398

Megeath, S. T., Gutermuth, R., Muzerolle, J., et al. 2012, AJ, 144, 192, doi: 10.1088/0004-6256/144/6/192

Miettinen, O. 2018, Ap&SS, 363, 197, doi: 10.1007/s10509-018-3418-7

Nelsen, R. B. 2007, An introduction to copulas (Springer Science & Business Media)

Nguyen, C. T., Costa, G., Girardi, L., et al. 2022, A&A, 665, A126, doi: 10.1051/0004-6361/202244166

Nieva, M.-F., & Przybilla, N. 2012, Astronomy & Astrophysics, 539, A143

Ohlendorf, H., Preibisch, T., Gaczkowski, B., et al. 2013, A&A, 552, A14, doi: 10.1051/0004-6361/201220218

Padoan, P., Haugbølle, T., & Nordlund, Å. 2012, ApJ, 759, L27, doi: 10.1088/2041-8205/759/2/L27

Pokhrel, R., Gutermuth, R. A., Betti, S. K., et al. 2020, ApJ, 896, 60, doi: 10.3847/1538-4357/ab92a2

Pokhrel, R., Gutermuth, R. A., Krumholz, M. R., et al. 2021, ApJ, 912, L19, doi: 10.3847/2041-8213/abf564

Pontoppidan, K. M., Barrientes, J., Blome, C., et al. 2022, ApJ, 936, L14, doi: 10.17909/67ft-nb86

Povich, M. S., Smith, N., Majewski, S. R., et al. 2011, ApJS, 194, 14, doi: 10.1088/0067-0049/194/1/14

Preibisch, T., Mehlhorn, M., Townsley, L., Broos, P., & Ratzka, T. 2014, A&A, 564, A120, doi: 10.1051/0004-6361/201323133

Preibisch, T., Roccatagliata, V., Gaczkowski, B., & Ratzka, T. 2012, A&A, 541, A132, doi: 10.1051/0004-6361/201218851

Rapson, V. A., Pipher, J. L., Gutermuth, R. A., et al. 2014, ApJ, 794, 124, doi: 10.1088/0004-637X/794/2/124

Reis, I., Baron, D., & Shahaf, S. 2019, AJ, 157, 16, doi: 10.3847/1538-3881/aaf101

Reiter, M., Morse, J. A., Smith, N., et al. 2022, MNRAS, doi: 10.1093/mnras/stac2820

Rigby, J., Perrin, M., McElwain, M., et al. 2023, PASP, 135, 048001, doi: 10.1088/1538-3873/acb293

Robbins, H., & Monro, S. 1951, The annals of mathematical statistics, 400

Roccatagliata, V., Preibisch, T., Ratzka, T., & Gaczkowski, B. 2013, A&A, 554, A6, doi: 10.1051/0004-6361/201321081

Rosenblatt, F. 1957, The perceptron, a perceiving and recognizing automaton Project Para (Cornell Aeronautical Laboratory)

Schmidt, M. 1959, ApJ, 129, 243, doi: 10.1086/146614

Scholz, A., Muzic, K., Jayawardhana, R., Quinlan, L., & Wurster, J. 2022, PASP, 134, 104401, doi: 10.1088/1538-3873/ac9431

Skrutskie, M. F., Cutri, R. M., Stiening, R., et al. 2006, AJ, 131, 1163, doi: 10.1086/498708

Stetson, P. 2006, User's Manual for DAOPHOT II. https://www.ucolick.org/~bolte/AY257/HMWK3_2015/daophotii.pdf

Stetson, P. B. 1987, PASP, 99, 191, doi: 10.1086/131977

—. 2011, DAOPHOT: Crowded-field Stellar Photometry Package, Astrophysics Source Code Library, record ascl:1104.011. http://ascl.net/1104.011

Tatton, B. L., van Loon, J. T., Cioni, M. R., et al. 2013, A&A, 554, A33, doi: 10.1051/0004-6361/201321209

The Astropy Collaboration, Price-Whelan, A. M., Sipőcz, B. M., et al. 2018, AJ, 156, 123, doi: 10.3847/1538-3881/aabc4f

Tremblin, P., Minier, V., Schneider, N., et al. 2013, A&A, 560, A19, doi: 10.1051/0004-6361/201322233

Ward-Thompson, D., & Whitworth, A. P. 2011, An Introduction to Star Formation

Wu, J., Evans, Neal J., I., Gao, Y., et al. 2005, ApJ, 635, L173, doi: 10.1086/499623

Appendix A

Photometry

This appendix serves as a step-by-step guide to recreating the photometry used in this thesis with the photometry package `DAOPHOT` (Stetson, 1987, 2011). It is intended to be a step-by-step procedure for others to create meaningful photometry from the different files and functions available with the `DAOPHOT` package, but it is not a full manual to the software, which can be accessed via Stetson (2006). The software itself was obtained via private communication with Dr. Peter Stetson.

A.1 Preparing the data (from JWST MAST files)

1. Cut down FITS files from MAST such that they only contain the first image layer. If this is not done, the program will be confused and will be unable to open the image.
2. Adjust the image values such that they are in ADU instead of MJy/sr. This is done by taking each pixel value and multiplying it by the appropriate factors (available in the headers of the FITS files). Divide by some factor to set the saturation point at 32767. Change the pixel values to be short ints at this step to minimize space usage.
3. Take any bad pixels (e.g. pixels with value 0 or NAN), such as saturated pixels or the border pixels, and set to 32767 (max for short int).
4. Obtain the read noise from a piece of relatively blank sky. Cut out a region in CARTA, analyze, and use the root mean square (RMS) of this area as the `READ NOISE` estimate. This should be the same for all filters given that they are taken by the same instrument. Add this value into daophot.opt.
5. Set `GAIN` to be that of the instrument, about 1.8 for NIRCam, in daophot.opt.

6. Set `HIGH GOOD DATUM` to 32 000 in daophot.opt to avoid finding overly saturated sources.

7. Set `LOW GOOD DATUM` to something about ten times background the first time to avoid noisy pixels, this will be adjusted later once a catalogue of the good point sources in all filters has been found.

8. Create .als files for about ten stars that are the same in each image for use in `daomatch`.

A.2 Running DAOPHOT

1. Run `./daophot`.

   (a) `at filt`, where *filt* is whatever name you have given the first FITS file. I recommend shortening this to just the filter name itself. This step attaches the FITS file so you can work on it.

   (b) `fi`, and specify `1, 1` when asked for number of frames and images. This finds all of the point sources within the image that are between the LOW and HIGH GOOD DATUM values and outputs them to a .coo file. For JWST FITS images the calibration and creation of the final image has been done for you, which is why we choose `1, 1`. These numbers would change if you were dealing with multiple exposures in a mosaic.

   (c) Check the .coo files to ensure that low and high *sharp*/*round* are set properly. Remove any sources with *sharp* or *round* = ±9.999. There should be a clear tight *sharp* line between 0 and 2, which defines how sharply the identified object peaks. Very high *sharp* values tend to be bad pixels, while very low tend to be nebulous objects. There should also be a clear *round* line between -1 and 1, however this depends on the population you expect to see. Stars should all sit around 0, but galaxies will have greater absolute *round* values due to being stretched in the x/y directions.

   (d) `ph` will provide aperture photometry for all the sources found with the .coo files. The aperture radii must be specified in a file named `photo.opt`. Because of the resolution of JWST, you can start the aperture value at 1.5 pixels. You can specify up to 12 different apertures, which must increase monotonically. For NGC 3324 we went from 1.5 to 10.3 pixels for the 12 different apertures.

   (e) `pi` is used to define how many and to what limiting magnitude to search for stars to make the PSF. Choose `1600, 15.5` when prompted to increase the number of fainter stars (which will be able to build a better PSF) over the number of bright stars. 50 good stars is plenty to make the PSF, but the more the merrier.
   (f) `ps` will perform PSF fitting on the 'picked' stars from the previous step. Run this step several times over and over until there are no asterisks printed to the screen. The stars with asterisks will be removed from the .nei file, and this file will be used when `ps` is run again. Asterisks specify that a source is three sigma away from the mean calculated psf and hence should be removed from the list.
   (g) Review .nei stars in CARTA or other FITS program to make sure that they aren't in extremely nebulous zones, on the edge of saturated stars, and aren't blended, choose about 100 really good candidates,and repeat the previous step until satisfied.
   (h) `ex` to exit

2. Run `./allstar`. If the image is too large, you may need to update the relevant parameters in `allstar.f` and remake/recompile this file.

3. Repeat the above procedure for all bands.

4. Run `./daomatch`.
   (a) Make master input file the first band for consistency (use the .als file you made in the data prep step of stars visible in all bands), and output file name the name of the region.
   (b) Input each band when prompted (use .als file made in data prep stage). The output will show a matrix of the brightest stars in the two bands being compared and how many times they were matched to each other in different ways when performing different rotations, scalings, etc. Ideally one should see a straight diagonal with large numbers. This should not be an issue if you have used the .als files made in the data prep stage as you are ensuring the exact same stars are being compared.

5. Run `./daomaster`.
   (a) Input the filename (name of region) defined in the previous step.

(b) Choose the minimum number of images a source must be detected in in order to be a real source (we choose `3` to have at least three colours and three magnitudes possible to be derived), as well as the minimum fraction of images that the object should be in (this is more important for if the images are mosaics of different parts of the field. For our purpose, the fields are on top of one another and so `0` is sufficient), and the number of frames after which the source is automatically included in the catalogue (again more relevant when dealing with mosaics across the field, we choose `3` for consistency).

(c) Set the number of degrees of freedom allowed. This accounts for translations (2), rotations and scales (4), linear aberrations (6), and more complicated aberrations (12, 20).

(d) Set match radius large (we chose `9`) and make increasingly smaller when prompted until you reach about `3`. This step is applying slight changes to try to place each source properly. The 1st and 2nd columns should hopefully fall below 1. The 5th and 8th columns should approach the scaling between the original filter and all other filters. The 11th column shows the number of stars matched in that band to the full catalogue.

(e) Keep the radius at `3` when prompted (you may click this several times) until you see that the final number of YSOs has stabilized.

(f) Say `yes` to `Assign new star IDs?`. This will give you identical star IDs across all of the individual files for easy comparison later.

(g) Say `yes` to `mean mags and scatter`.

(h) Say `no` to `corrected mags and scatter` as this is only for if all of your images were in the same filter and you want to correct them so that the matched stars all have the same magnitude.

(i) Say `yes` to `raw mags and scatter`, this will give you a catalogue with all of the bands matched and their photometric errors!

(j) Say `yes` to `new transformations`, this will update the transformation table from `daomatch`.

(k) Say `no` to `transfer table`.

(l) Say `yes` to `individual .coo files` and overwrite old ones.

(m) `e`.

From this, a catalogue is created, however it has several missing data-values for each point source. To fill in the holes in the data, photometry must be taken at that source location for each filter.

6. Repeat `ph` in `./daophot` for each filter, using the new .coo files, and lowering the `LOW GOOD` in the .coo file header if needed.
7. Repeat `./allstar` for each filter.
8. Repeat `./daomaster`.
9. At this stage you will want to check to see if there are any objects which you can clearly see and have been missed. These can be manually added to the bottom of the `.mag` file. You need the `Star ID` (start it at the next thousands place above the last `Star ID` for simplicity), the `x` and `y` coordinates, and some example `mag`. Choose a magnitude that is in the middle of the range. The rest of the columns can be left blank.
10. Run `./allframe`, inputting in the mag file you got from the latest run of `daomaster`. This will produce `.alf` files. `Allframe` obtains photometry from all the images at once, which leads to more consistent photometry across filters.
11. Update the `.mch` file to use the `.alf` extensions and rerun `daomaster`.
12. Congratulations, you now have a catalog in the latest `.raw` file!

A.3 Analyzing Photometry

1. To obtain zero-points (zps): run `ph` on the original MAST images for the same objects that you used to make the PSF, as well as on those same objects in the ADU versions of the images. Set the apertures to increase to a radius of about 30 pixels, and compare the `.ap` results as magnitude vs aperture radius. After a certain increase in radius, the measured magnitude will no longer change. Take the measurements in both instrumental mags and MJy/sr and find the constant which will allow you to convert between the two, Equation A.1. Because mags are logs of the fluxes, the conversion factors can be separated out and absorbed into the constant. This will give you a constant that can be added to the ADU mags to place them in the same units as the output from

MAST, accounting for both conversion factors and additional offsets. The additional 25 in the equation is a constant added by `DAOPHOT`.

$$\mathrm{k} = \mathrm{mag}_{\mathrm{ADU}} - (\mathrm{mag}_{\mathrm{MJy/sr}} - 25 + 2.5 \log \mathrm{PHOTMJSR}) \tag{A.1}$$

From here, you can use the conversions provided by Space Telescope to convert to Vega mags via Equation A.2.

$$\mathrm{mag}_{\mathrm{Vega}} = \mathrm{ZP}_{\mathrm{Vega}} + (\mathrm{mag}_{\mathrm{ADU}} - \mathrm{k}) \tag{A.2}$$

2. To ensure the photometry is good and science can be done with it, look at a variety of colour-colour diagrams. If the photometry is good, you should mostly see clear correlations or anti-correlations. In general, your x axis should be the two wide bands which are furthest apart (for instance F090W and F444W from the six NIRCam filters we had available), as this functions as an approximation for temperature. For example, see Figure A.1.

3. You can compare the RMS of the residuals to the propagated error, Equation A.3.

$$\sqrt{\frac{1}{\mathrm{n}}\Sigma_{\mathrm{i}}(\mathrm{residual})_{\mathrm{i}}^{2}} = \sqrt{\sigma_{\mathrm{col}_1}^{2} + \sigma_{\mathrm{col}_2}^{2}} \tag{A.3}$$

   If the RMS is approximately equal to the error, there is one stellar population. If it is much larger, than this is a sure indicator of different populations.

4. You can make $\delta - \delta$ plots to try to separate out the different populations. To make a $\delta - \delta$ plot, start with finding the δ values of each colour for each object. This is done by fitting a line to the distribution, and then subtracting the objects colour vertically from the best fit line, i.e., the δ value for each object is the residual from the best fit line. The δ values of the colour with the greatest spread should then act as your x-axis, while the δ values of the remaining colours are plotted on the y-axis to find correlations and anti-correlations. Finally, add the correlations and subtract the anti-correlations to separate out the data into distinct populations. You may need to apply cuts on the allowed errors of objects in order to see the populations distinguish themselves. The population of field stars will likely be contained in a tight sequence, while any YSOs and galaxies will separate themselves. Figure A.2 shows the summing of correlations and anti-correlations and a notable spreading from the tight sequence. Without further

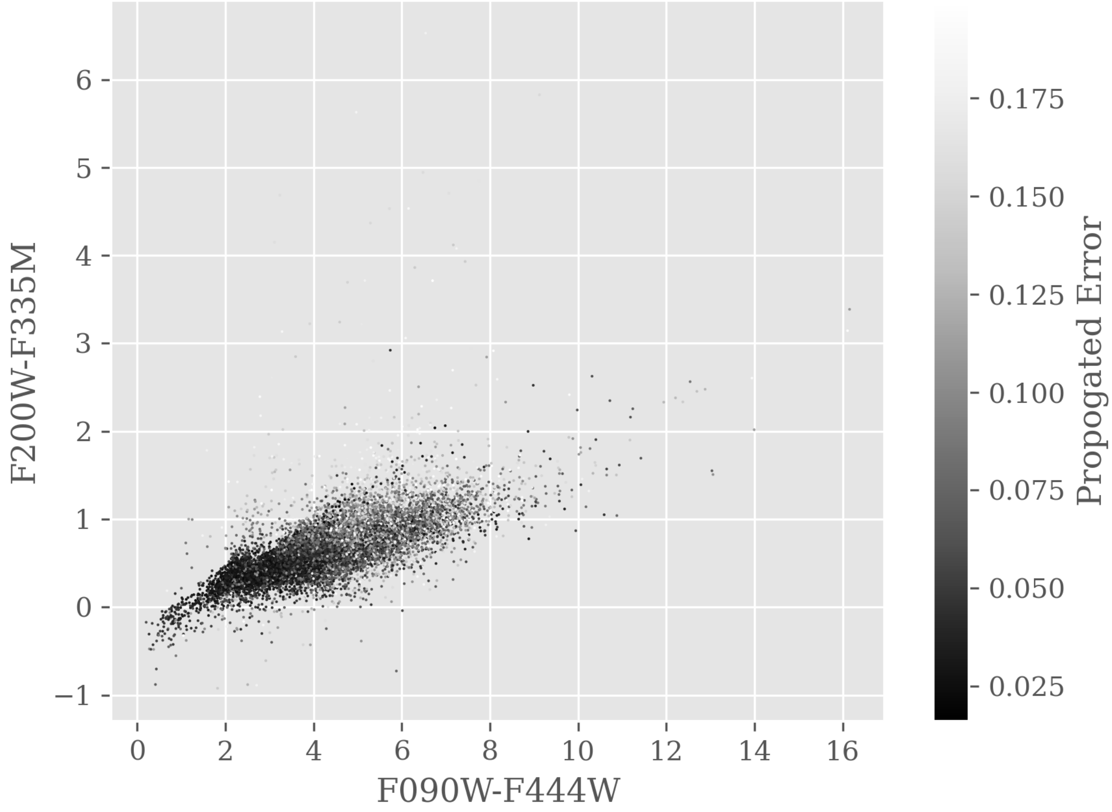

Figure A.1: An example of a colour-colour diagram which shows the tight linear correlation between F200W-F335M, and F090W-F444W (which traces temperature). This clear correlation is how we validate that our photometry is science-worthy.

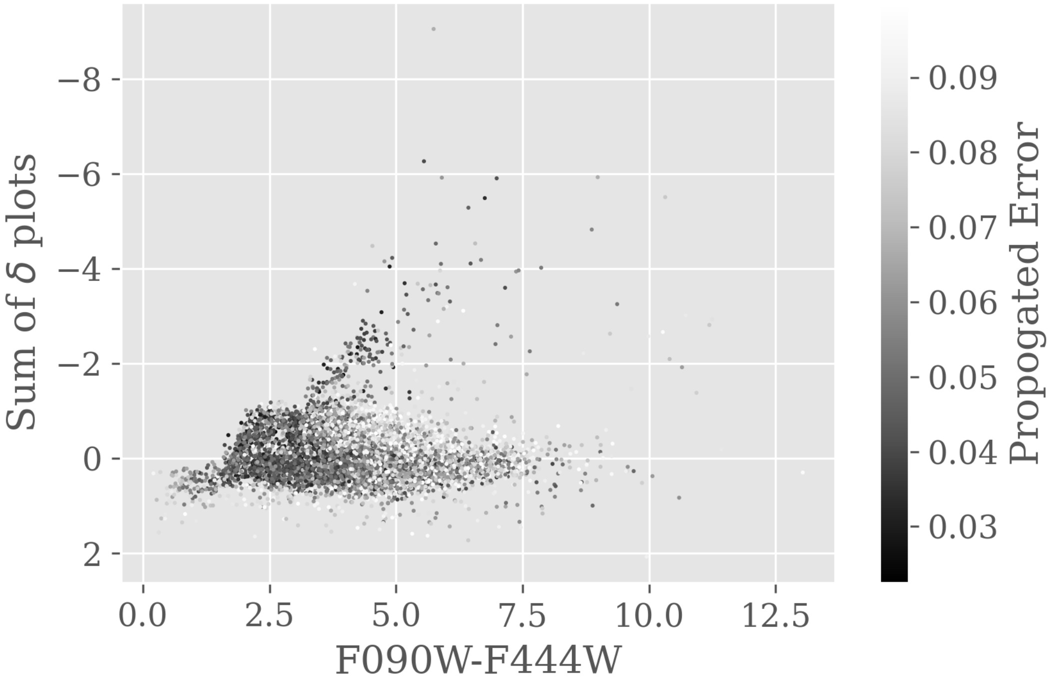

Figure A.2: The sum of several δ-values based on correlation and anti-correlation, vs F090W-F444W. The bottom of the distribution appears to be a dense line, likely the star population, and the more diffuse populations separating themselves from this trend may be other classes such as outflows, YSOs, or galaxies. See text for full discussion.

> information this cannot be used alone to identify the YSO population, as certain classes may look distinctly different. In Figure A.2 we can see as many as four populations, and we found that our YSOs tend to follow the sharp diagonal line as well as the upper more horizontal clustering above the main population, perhaps linked to the different phases of star formation. This analysis is circumstantial and so was not included in the paper.

This procedure may be improved by an update to the `FIND` method. `DAOPHOT` and many other photometry packages like it use increases in intensity above the background to identify point-like sources (the `FIND` algorithm), which works in premise. Unfortunately, when point sources are faint and not above the background to a significant threshold, or when the background varies significantly, this algorithm can no longer retrieve all point sources in a field of view. A possible direction to rectify these limitations would explore the use of a CNN, see Section 1.3.1 for a description. This approach would need to be done by using kernels of small size to iterate over the image and identify whenever there is a point source, and then to retrieve the x and y coordinates of this point source. Although a fairly straightforward work, a CNN-guided model could provide much improvement in completeness over the `FIND` algorithm in `DAOPHOT`.

Appendix B

PRF Code

The full code for this work can be accessed on Github. It includes changing the output of DAOPHOT from instrumental magnitudes to apparent Vega magnitudes (daophot_to_csv.py), augmenting the data to produce a new training set (Augment_Data.py), the code to run both the PRF and RF for 100 trials and, hence, obtain classifications (PRF_RF_mp.py), and creating all figures in Chapter 2 (PRF_DAOPHOT_IR.py). The full code for all files is over 1500 lines. As such, we attach only the Python code that is used to run the PRF once and return probabilities to demonstrate the ease of classification.

```
def get_best_prf(tr, va, te):
# Create a random seed - a new one is chosen for each iteration.
   seed = int(np.random.default_rng().random()/
      *np.random.default_rng().random()*10000)
# Shuffle data based on the random seed for both the training and validation sets.
   inp_tr, tar_tr = replicate_data_single(tr[:,:-1],tr[:,-1],/
      amounts=[len(tr[:,-1][tr[:,-1]==0]),len(tr[:,-1][tr[:,-1]==1])],seed=seed)
   inp_va, tar_va = replicate_data_single(va[:,:-1],va[:,-1],/
      amounts=[len(va[:,-1][va[:,-1]==0]),len(va[:,-1][va[:,-1]==1])],seed=seed)
# Specify which indices correspond to the magnitudes and which correspond to the
errors.
   inds = range(0,int(np.shape(inp_tr)[1]/2))
   e_inds = range(int(np.shape(inp_tr)[1]/2),np.shape(inp_tr)[1])
# Initiate the PRF, with 100 trees, requiring that tree branches be cut if the
probability of passing the node is less than 50%.
    prf_cls = prf(n_estimators=100, bootstrap=False, keep_proba=0.5)
# Fit the PRF model to the training set (from augmented data).
```

```
   prf_cls.fit(X=inp_tr[:,inds], dX=inp_tr[:,e_inds], y=tar_tr)
# Return the predicted probabilities for the validation set and the full data set.
   pred_va = prf_cls.predict_proba(X=inp_va[:,inds],dX=inp_va[:,e_inds])
   pred_te = prf_cls.predict_proba(X=te[:,inds], dX=te[:,e_inds])
# Return the predicted probabilities of all objects
   return np.transpose(pred_te)[0]
```

Appendix C

Outflow driving sources in the Cosmic Cliffs

In the following, we briefly describe the outflow driving source candidates determined in the Cosmic Cliffs field (Reiter et al., 2022) matched to JWST objects we identify, give the PRF-based probabilities of these being YSOs, and suggest alternatives if we find these have less than 50% probability of being a YSO. A label is assigned to each driving source candidate as being either FN = False Negative – the object is very likely a YSO but the PRF algorithm classifies it as a contaminant, TP = True Positive – the object is very likely a YSO and the PRF classifies it as such, MC = Misclassified by Reiter et al. (2022) – we find a candidate with higher probability along the outflow axis, NP = New Positive – a cYSO identified in our work that was not previously found, e.g., with Spitzer, or ND = No Detection – the candidate driving source did not have sufficient good photometry to be included in the catalog.

MHO 1632: An arced outflow seen to either side of the suggested driving source, J103642.3-583804, which is located deepest within the cloud and whose SED does not broach 20 mag in the NIRCam bands. With MIRI data, this YSO may become more identifiable. It has a probability of being a YSO of 5%. No other likely candidates are found along the outflow axis, and as such this candidate is likely a false negative. (FN)

MHO 1633: An arced sharp J-shaped outflow originating from J103648.0-583819. This object lies about midway within the dust cloud. This object has probability $< 1\%$, with no other possible driving sources along the outflow axis, and as such this candidate is likely a false negative. The object is most likely not identifiable as a YSO due to its faint SED. (FN)

MHO 1634: Several bow shocks are identified on either side of the driving source candidate, J103646.7-583805. This object has a probability of 85% of being a YSO, with no other objects along the axis having greater than 10% probability of being a YSO. It was previously identified as a YSO in SPICY. (TP)

MHO 1635: An arced outflow on either side of the axis. This object is only visible in two bands (F444W and F470N) and as such we did not have enough data to classify it as a YSO or not. No other nearby objects along the outflow axis have greater than 10% probability of being a YSO. Reiter et al. (2022) suggest that this object is only visible in MIRI data. Further studies will determine if we can retrieve it using the PRF method. (ND)

MHO 1636: a series of three separate outflows, which can be traced back towards an object which lies midway within the dust cloud, candidate driving source J103651.5-583754. The PRF yields a probability of 15%, likely due to the objects faint nature. (FN)

MHO 1637: Bipolar outflows originating from J103650.5-583752, an object with ∼ 100% probability of being a YSO, which was previously identified in Spitzer data and was part of the training set. (TP)

MHO 1638: A series of several outflows, tracing a gentle c-shape with the driving source candidate in the middle of the c appearing to originate from J103651.4-583748. This object has only a 2% probability of being a YSO, and all others along the outflow axis have probability less than that. The nearest YSO candidate along this axis is J103652.3-583809, which is also associated with HH 1219, has a probability of ∼ 100%, and was previously identified in Spitzer data. By visual inspection, the HH outflow also appears to be along the same axis. (MC)

MHO 1639, HH 1221, HH 1003 Aa: Driving source J103653.8-583748 has a probability of being a YSO of 99%, but was not previously identified as a YSO in Spitzer. The outflows trace a bright object located on the cusp of the cloud edge. This object is hidden in the shorter wavelengths, and appears to light up the edge of a nearby cavity, see Reiter et al. (2022) for details. (NP)

MHO 1640: Asymmetric outflow appearing to originate from J103651.5-583710, which has a probability of being a YSO of only 15%. Slightly further along the outflow axis, which would lead to a more symmetric outflow, is a fainter driving source candidate with a probability of being a YSO of 56%, J103651.3-583709. (MC, NP)

MHO 1643: The clear outflow driving source here J103654.2-583626 has twin bow shocks but was unable to be classified by us as a YSO due to saturation at long wavelengths as well as being too faint in short wavelengths. Indicative of the shortcomings of retrieving photometry from JWST images. (ND)

MHO 1645, MHO 1646: Bipolar outflow and H_2 knots trace the possible driving source J103654.4-583618, which we find has 79% probability of being a YSO. (NP)

MHO 1647, HH 1002a: Two bow shocks with proper motion trace driving source J103654.0-583720; this object has a ∼ 100% probability of being a YSO but was not found with Spitzer. (NP)

MHO 1649: A series of four nearby shocks trace J103653.6-583520 which has probability of being a YSO of 20%. No other possible driving sources are seen along the outflow axis. (FN)

MHO 1650: Single outflow with suggested driving source J103653.1-583737 (PRF probability <1% of being a YSO), which is coincident with another bright spot of H_2 emission. Another possible driving source, J103652.83-583739, however, lies along the outflow axis with a probability of being a YSO of 84%. (MC, NP)

MHO 1651, HH 1003 B, Ca: Two knots originate from driving source J103653.3-583754, which has a probability of being a YSO of $\sim$ 100% and was previously identified in Spitzer data and hence was part of the training data. (TP)

MHO 1652: Two knots trace driving source J103652.7-583805, which has a probability of being a YSO of 25%. No other sources lie along the outflow axis. (FN)

All HH objects had proper motions and could be traced back to their source.

HH c-3: Driving source J103701.5-583751 has probability 86% and was not previously identified in Spitzer data and hence not part of the training set. (NP)

HH c-4: Driving source J103702.1-583658 has probability $\sim$ 100% and was previously identified in Spitzer data and hence was part of the training set. (TP)

HH c-5: Possible driving source J103653.9-583632 does not have sufficient photometric data available ($>$ 2 bands). Nearby source J103653.7-583632 has probability 27% (PRF) or 41% (RF), respectively. (ND)

HH 1223a: Clear driving source is too saturated to obtain classification, but was previously identified with Spitzer. (ND)

We note that false negatives generally seem to be attributed to faint sources, while non-detections are attributed to over-saturated sources. With MIRI data included in the ensemble, we may be able to recover many as cYSOs. In total, we find 5 False Negatives, 4 True Positives, 3 Misclassifiactions, 6 New Positives, and 4 Non-Detections.

Appendix D

Sample of Magnitudes and Probabilities for some cYSOs

In the table below, we provide the full catalogue of magnitudes and probabilities of objects returned as candidate YSOs in the Cosmic Cliffs field imaged with JWST from either the PRF or the RF.

R.A.	decl.	F090W	F200W	F335M	F444W	F470N	Prob PRF	Prob RF
159.2530	-58.5958	18.18	14.39	13.47	12.78	12.80	1.00	1.00
159.2987	-58.5735	19.32	15.59	14.14	13.38	13.29	1.00	1.00
159.2462	-58.5902	16.84	14.07	13.16	12.35	12.29	1.00	1.00
159.2780	-58.5721	15.93	12.90	11.56	10.77	11.17	1.00	0.98
159.2588	-58.6160	19.29	16.15	14.50	13.51	13.38	1.00	1.00
159.2789	-58.6154	14.86	12.73	12.34	11.93	12.23	1.00	0.99
159.2536	-58.6063	16.88	14.16	13.31	12.64	12.62	0.99	1.00
159.2434	-58.6053	19.07	15.30	13.63	12.51	12.42	1.00	1.00
159.2503	-58.5912	16.60	13.20	12.20	11.45	11.48	1.00	1.00
159.2188	-58.6237	16.88	14.00	13.38	12.62	12.61	1.00	1.00
159.2514	-58.6416	16.60	14.03	13.17	12.57	12.63	0.99	1.00
159.2221	-58.6317	17.14	11.87	11.57	10.35	10.54	1.00	0.99
159.2360	-58.6164	18.36	14.68	13.49	12.71	12.69	1.00	1.00
159.2768	-58.5883	15.55	12.74	11.68	10.87	11.46	1.00	1.00
159.2036	-58.6346	15.22	10.71	11.10	10.14	10.66	1.00	0.99
159.2850	-58.6152	16.23	13.63	12.66	12.05	12.19	1.00	1.00
159.2972	-58.5793	15.85	13.09	11.93	11.10	11.42	1.00	1.00
159.2051	-58.6393	20.80	13.25	12.34	11.85	12.19	1.00	1.00
159.2106	-58.6312	19.24	14.76	13.31	12.25	12.20	1.00	1.00

R.A.	decl.	F090W	F200W	F335M	F444W	F470N	Prob PRF	Prob RF
159.2181	-58.6358	24.56	14.26	10.87	8.40	8.84	1.00	0.99
159.2250	-58.6221	22.44	14.11	11.66	9.75	10.26	1.00	0.99
159.2086	-58.6101	26.12	21.58	21.38	20.60	20.69	0.50	0.49
159.1549	-58.6040	26.66	21.73	21.43	20.22	20.79	0.90	0.87
159.1533	-58.6042	17.25	12.86	10.50	9.48	10.72	1.00	0.97
159.2031	-58.6108	25.43	21.27	21.03	20.17	21.10	0.73	0.68
159.1804	-58.6061	28.10	20.83	19.60	19.10	19.25	0.57	0.56
159.2262	-58.6138	27.93	19.94	18.39	17.85	17.92	0.53	0.46
159.1942	-58.6101	26.29	21.73	21.22	20.42	20.77	0.83	0.77
159.2107	-58.6095	19.54	16.87	16.47	15.48	15.30	0.65	0.56
159.2040	-58.6113	24.38	20.69	20.39	19.61	20.37	0.67	0.63
159.2206	-58.6106	23.79	19.76	18.95	17.77	17.78	0.58	0.58
159.2486	-58.6128	23.74	20.20	18.74	17.90	17.84	1.00	1.00
159.2231	-58.6096	23.43	16.55	14.81	13.41	13.77	0.71	0.69
159.1652	-58.6047	27.14	21.93	21.42	20.55	20.62	0.53	0.50
159.1793	-58.6064	27.55	21.46	20.44	19.64	19.91	0.63	0.59
159.1998	-58.6808	25.50	22.31	21.72	21.04	21.31	0.50	0.53
159.1644	-58.6020	21.10	19.42	19.16	18.75	19.02	0.88	0.83
159.2236	-58.6043	19.21	16.10	15.16	14.33	14.24	0.87	0.84
159.1711	-58.5982	28.00	19.84	17.92	17.45	17.50	0.53	0.45
159.2267	-58.6051	23.46	16.55	16.37	14.75	14.69	0.60	0.59
159.2269	-58.6051	22.52	14.24	12.34	11.34	11.52	0.63	0.55
159.2448	-58.6073	23.39	19.61	18.08	17.32	17.07	1.00	0.99
159.2651	-58.6107	21.17	18.61	17.21	16.30	16.05	0.65	0.62
159.2232	-58.6062	22.68	18.21	16.76	15.53	15.31	0.72	0.67
159.2177	-58.6057	17.86	14.90	14.02	13.23	13.19	0.73	0.69
159.2150	-58.6062	23.21	19.49	18.21	17.28	17.13	0.92	0.84
159.1697	-58.6029	25.43	17.64	16.43	15.85	15.97	0.62	0.58
159.2181	-58.6067	19.84	17.31	16.25	15.61	15.72	0.97	0.87
159.2386	-58.6099	25.06	21.58	19.89	19.16	19.49	0.99	0.95
159.2384	-58.6099	24.75	21.11	20.05	19.14	19.15	0.91	0.83
159.2183	-58.6075	19.40	15.71	14.85	14.03	14.02	0.98	0.95
159.2547	-58.6127	21.73	19.33	18.81	18.42	18.76	0.51	0.51
159.2140	-58.6077	24.57	21.46	21.53	20.36	21.00	0.81	0.75

R.A.	decl.	F090W	F200W	F335M	F444W	F470N	Prob PRF	Prob RF
159.2126	-58.6078	26.18	22.64	21.42	20.75	20.69	0.84	0.73
159.2166	-58.6130	26.47	20.12	17.72	16.24	16.63	0.71	0.72
159.1936	-58.6055	25.51	21.55	20.88	20.13	20.23	0.55	0.53
159.1808	-58.6040	25.99	21.52	20.74	20.00	20.25	0.71	0.64
159.2189	-58.6069	18.90	15.53	14.54	13.79	13.91	0.86	0.79
159.2181	-58.6134	20.90	16.33	15.41	14.29	14.49	0.65	0.59
159.1987	-58.6116	24.78	20.99	20.43	19.72	20.08	0.66	0.64
159.1678	-58.6078	25.41	21.33	20.72	20.04	20.27	0.45	0.52
159.2117	-58.6186	25.96	21.57	20.89	20.17	20.47	0.68	0.63
159.1746	-58.6141	27.26	22.19	21.53	20.72	20.89	0.79	0.75
159.2046	-58.6178	24.50	20.46	20.23	19.46	19.62	0.62	0.63
159.1488	-58.6111	25.45	20.17	19.80	18.95	19.26	0.78	0.73
159.1508	-58.6114	24.07	19.18	19.22	18.05	18.14	0.65	0.66
159.1764	-58.6147	23.85	20.27	19.90	19.09	19.26	0.74	0.69
159.1595	-58.6127	25.63	20.70	20.78	19.79	20.05	0.92	0.87
159.2064	-58.6185	24.52	20.60	19.97	19.26	19.22	0.62	0.67
159.1735	-58.6145	24.70	19.77	19.51	18.64	18.85	0.76	0.72
159.2010	-58.6180	24.43	20.40	20.25	19.49	19.73	0.67	0.57
159.1557	-58.6125	25.99	21.79	20.95	20.27	20.19	0.49	0.55
159.2012	-58.6182	24.51	20.24	20.15	19.39	19.41	0.50	0.52
159.1541	-58.6124	25.73	21.28	20.78	20.06	20.42	0.69	0.64
159.1502	-58.6120	24.38	20.61	20.41	19.36	19.32	0.78	0.75
159.1758	-58.6152	23.76	20.42	20.57	19.50	19.57	0.67	0.61
159.1977	-58.6180	24.02	20.85	20.44	19.74	19.71	0.54	0.54
159.1489	-58.6120	26.77	22.15	21.39	20.67	21.00	0.68	0.65
159.2052	-58.6189	25.31	21.50	21.79	20.37	20.92	0.92	0.87
159.1627	-58.6138	23.49	20.27	20.11	19.32	19.48	0.63	0.60
159.1716	-58.6149	25.30	21.38	20.68	20.02	20.47	0.47	0.51
159.1740	-58.6153	26.54	21.28	20.99	20.20	20.41	0.66	0.61
159.1640	-58.6141	24.66	20.38	20.30	19.47	19.45	0.60	0.58
159.1543	-58.6130	26.69	20.98	20.29	19.54	19.67	0.54	0.51
159.2075	-58.6178	22.38	18.48	18.05	17.30	17.47	0.70	0.65
159.1743	-58.6084	25.55	20.96	20.82	20.08	20.25	0.59	0.57
159.2135	-58.6183	21.94	19.01	18.34	17.63	17.58	0.83	0.83

R.A.	decl.	F090W	F200W	F335M	F444W	F470N	Prob PRF	Prob RF
159.1933	-58.6154	24.77	21.22	20.83	20.14	20.49	0.55	0.56
159.2737	-58.6105	23.48	20.39	19.16	18.44	18.05	0.85	0.77
159.1620	-58.6079	25.18	20.99	20.61	19.84	20.05	0.67	0.63
159.1539	-58.6070	23.72	19.21	19.19	18.42	18.56	0.70	0.66
159.1771	-58.6099	24.62	21.22	21.07	20.17	20.56	0.83	0.78
159.1499	-58.6069	25.50	21.11	20.57	19.86	20.02	0.52	0.53
159.1662	-58.6097	27.22	21.57	21.15	19.94	20.22	0.95	0.90
159.1547	-58.6088	25.74	21.44	21.06	20.33	21.04	0.59	0.58
159.1867	-58.6127	27.22	22.31	21.56	20.86	21.16	0.55	0.56
159.1492	-58.6081	26.57	21.95	21.43	20.51	20.49	0.59	0.59
159.1618	-58.6100	26.54	21.10	20.31	19.46	19.88	0.73	0.69
159.2272	-58.6181	18.44	15.85	15.03	14.34	14.28	0.81	0.71
159.1647	-58.6104	25.68	21.74	21.44	20.65	20.94	0.72	0.69
159.1619	-58.6101	24.85	20.56	20.19	19.37	19.52	0.70	0.69
159.2609	-58.5637	26.98	18.78	17.29	16.98	16.96	0.54	0.45
159.2135	-58.6168	19.14	15.30	14.49	13.55	13.52	0.82	0.75
159.1573	-58.6099	26.06	22.14	21.59	20.78	20.83	0.66	0.65
159.1511	-58.6093	27.48	21.89	21.60	20.45	21.06	0.86	0.77
159.1601	-58.6109	25.36	19.67	17.78	16.62	18.28	0.65	0.66
159.2039	-58.6164	25.21	21.44	21.12	20.33	20.84	0.74	0.71
159.2103	-58.6172	24.09	20.94	20.53	19.83	20.19	0.58	0.59
159.1700	-58.6123	25.89	21.92	21.56	20.77	21.05	0.72	0.69
159.2327	-58.6202	21.82	19.07	17.84	17.11	16.96	0.76	0.68
159.1693	-58.6124	24.84	21.03	20.61	19.92	20.25	0.47	0.50
159.2553	-58.6234	25.07	21.64	20.50	19.82	19.82	0.95	0.90
159.2286	-58.6049	21.43	18.78	18.19	16.77	17.21	0.82	0.74
159.2593	-58.6077	21.11	18.19	16.84	15.94	15.84	0.90	0.83
159.2248	-58.6040	23.44	19.39	18.03	17.11	16.95	0.63	0.62
159.2592	-58.5730	22.55	18.31	17.01	16.29	16.19	0.56	0.54
159.2397	-58.5708	22.33	20.48	19.96	19.52	19.84	0.63	0.60
159.2638	-58.5745	19.66	16.03	14.42	13.65	13.58	1.00	0.99
159.2387	-58.5715	27.01	19.38	17.82	17.33	17.33	0.50	0.44
159.2504	-58.5732	26.47	18.56	16.48	15.00	15.41	0.84	0.79
159.2508	-58.5734	27.29	24.16	19.81	15.70	17.62	0.68	0.64

R.A.	decl.	F090W	F200W	F335M	F444W	F470N	Prob PRF	Prob RF
159.2503	-58.5734	25.61	15.86	13.41	11.39	11.98	0.93	0.89
159.2347	-58.5716	24.24	16.89	15.69	15.21	15.28	0.52	0.52
159.2501	-58.5738	25.32	14.80	11.66	9.21	10.63	0.76	0.76
159.2668	-58.5760	24.26	21.64	21.25	20.75	21.05	0.51	0.55
159.2433	-58.5738	23.55	15.76	14.38	14.01	14.07	0.58	0.52
159.2488	-58.5755	26.54	15.69	12.85	11.49	11.98	0.54	0.57
159.2149	-58.5722	21.04	19.00	18.41	18.04	18.30	0.44	0.55
159.2444	-58.5770	27.60	19.54	17.72	17.34	17.45	0.54	0.45
159.2805	-58.5822	22.38	19.52	18.69	17.89	17.43	0.87	0.73
159.2255	-58.5759	26.42	18.65	17.23	16.70	16.68	0.53	0.47
159.2448	-58.5787	27.21	19.09	17.39	16.87	16.93	0.53	0.47
159.2218	-58.5766	27.41	20.92	19.28	18.49	18.73	0.61	0.56
159.2671	-58.5834	19.20	15.29	13.56	12.39	12.48	0.73	0.65
159.2986	-58.5874	23.08	19.83	19.67	18.59	19.14	0.86	0.81
159.2908	-58.5866	23.36	20.64	20.00	19.10	18.85	0.89	0.78
159.2422	-58.5700	26.94	23.32	21.09	20.35	20.42	0.77	0.65
159.2667	-58.5727	25.16	21.57	20.57	19.76	19.61	0.96	0.89
159.2750	-58.5734	20.95	17.59	16.03	15.07	14.94	1.00	0.99
159.2482	-58.5699	27.06	22.62	19.19	17.50	17.45	0.59	0.59
159.2234	-58.5592	27.13	19.52	18.37	17.94	17.99	0.64	0.63
159.2268	-58.5599	26.86	22.77	20.94	20.11	20.06	0.96	0.92
159.2253	-58.5598	26.00	22.41	20.48	19.47	19.46	1.00	0.95
159.2464	-58.5625	26.37	22.41	19.49	18.26	18.16	0.57	0.55
159.2378	-58.5615	26.83	18.99	17.42	17.03	17.10	0.50	0.45
159.2567	-58.5639	25.73	17.46	15.81	15.35	15.40	0.57	0.52
159.2391	-58.5619	27.27	22.66	20.73	19.82	20.18	0.81	0.78
159.2123	-58.5587	22.22	20.28	19.88	19.48	19.74	0.95	0.94
159.2212	-58.5600	25.97	22.29	19.38	17.89	17.83	0.59	0.57
159.2547	-58.5645	26.43	22.97	20.61	19.65	19.52	0.88	0.81
159.2698	-58.5844	21.01	17.99	16.69	15.81	15.77	0.96	0.92
159.2386	-58.5626	27.47	20.12	18.88	18.42	18.50	0.54	0.52
159.2521	-58.5648	27.04	23.11	20.17	19.17	19.09	0.71	0.67
159.2889	-58.5695	22.44	19.67	17.97	16.54	16.43	0.76	0.72
159.2430	-58.5648	26.88	22.53	19.32	18.02	17.83	0.56	0.52

R.A.	decl.	F090W	F200W	F335M	F444W	F470N	Prob PRF	Prob RF
159.2588	-58.5674	25.37	16.77	15.27	14.72	14.81	0.59	0.54
159.2605	-58.5683	26.09	18.83	17.87	17.37	17.40	0.61	0.62
159.2590	-58.5685	25.43	17.73	16.43	15.96	15.81	0.53	0.48
159.2997	-58.5744	19.48	16.98	16.13	15.58	15.58	0.57	0.45
159.2727	-58.5716	23.90	19.88	18.70	17.79	18.37	0.74	0.68
159.2737	-58.5718	16.21	12.13	10.55	7.72	9.90	0.86	0.78
159.1818	-58.5617	25.54	22.32	21.83	20.99	21.61	0.73	0.67
159.2544	-58.5646	26.41	23.06	20.49	19.73	19.57	0.83	0.80
159.2337	-58.5801	24.95	15.78	14.08	13.53	13.63	0.53	0.46
159.2354	-58.5810	25.00	20.53	17.94	16.84	16.75	0.73	0.70
159.2007	-58.5776	25.54	21.16	21.06	20.04	19.88	0.76	0.71
159.2416	-58.5936	22.87	19.44	18.03	17.09	16.86	0.99	0.98
159.2563	-58.5973	23.39	20.52	20.04	19.61	19.86	0.51	0.48
159.2429	-58.5959	22.66	20.06	19.23	18.36	18.34	0.53	0.47
159.2581	-58.5979	20.58	17.49	15.91	15.16	14.90	0.98	0.96
159.2134	-58.5928	25.52	20.75	20.87	20.13	20.48	0.66	0.63
159.2618	-58.5993	21.26	17.42	15.70	14.79	14.69	0.98	0.96
159.2832	-58.6042	23.73	20.65	19.64	18.90	18.87	0.69	0.64
159.1711	-58.5907	26.23	18.84	17.60	17.21	17.18	0.51	0.47
159.2384	-58.6007	21.98	18.37	16.33	15.23	15.09	0.94	0.92
159.2451	-58.6016	22.20	19.58	18.62	17.96	17.84	0.53	0.29
159.2723	-58.5969	21.38	18.80	17.95	16.98	16.85	0.53	0.48
159.1746	-58.5930	27.69	19.65	18.45	17.90	17.90	0.64	0.59
159.1862	-58.5947	21.14	19.48	19.16	18.81	19.04	0.53	0.51
159.1811	-58.5954	27.62	20.25	19.16	18.92	19.08	0.54	0.57
159.2022	-58.5993	25.37	20.70	20.35	19.44	19.77	0.87	0.81
159.2062	-58.5999	24.98	20.61	20.72	19.64	20.13	0.92	0.90
159.2233	-58.6029	21.14	17.38	15.81	15.14	15.07	0.77	0.63
159.1723	-58.5969	27.81	20.13	18.65	18.22	18.32	0.51	0.45
159.1675	-58.6147	24.05	20.08	19.74	18.92	19.03	0.59	0.58
159.2440	-58.6059	25.31	21.95	21.05	20.42	20.41	0.63	0.62
159.1731	-58.5975	27.53	19.06	17.26	17.08	17.12	0.52	0.46
159.1595	-58.5960	27.52	19.80	18.57	18.12	18.14	0.63	0.60
159.1719	-58.5929	27.74	19.58	17.98	17.74	17.81	0.52	0.46

R.A.	decl.	F090W	F200W	F335M	F444W	F470N	Prob PRF	Prob RF
159.2544	-58.6078	24.05	20.06	18.35	17.37	17.31	0.95	0.93
159.2432	-58.5931	21.57	18.95	18.12	17.20	16.66	0.71	0.58
159.2413	-58.5922	20.78	17.37	16.01	15.15	15.01	0.97	0.93
159.2435	-58.5834	18.47	15.38	14.69	13.97	13.91	0.83	0.81
159.2778	-58.5883	18.53	15.18	13.74	12.85	12.74	1.00	1.00
159.2476	-58.5846	21.85	17.26	15.65	14.82	14.89	0.54	0.44
159.2471	-58.5846	18.11	13.59	12.63	11.79	12.03	0.68	0.58
159.2777	-58.5883	19.32	16.73	15.66	14.82	15.20	0.86	0.83
159.1994	-58.5788	25.42	21.14	20.64	19.93	19.76	0.52	0.53
159.2746	-58.5883	21.92	18.50	16.64	15.49	15.27	0.95	0.93
159.2126	-58.5810	22.26	20.15	19.83	19.30	19.71	0.69	0.63
159.2463	-58.5874	18.34	14.94	13.46	12.47	12.42	1.00	0.99
159.2823	-58.5920	22.98	20.90	20.51	19.95	20.26	0.67	0.67
159.1730	-58.5843	26.98	21.73	20.99	20.08	20.13	0.51	0.56
159.2442	-58.5877	19.31	15.25	14.84	14.08	13.82	0.60	0.61
159.2448	-58.5884	18.01	15.00	13.69	12.98	12.97	0.61	0.57
159.2416	-58.5891	20.24	15.72	14.85	14.01	14.03	0.59	0.60
159.1629	-58.5795	24.05	20.68	20.04	19.32	19.39	0.58	0.52
159.2356	-58.5889	27.66	23.91	21.21	19.95	19.75	0.64	0.60
159.1857	-58.5830	27.10	19.74	18.63	18.13	18.18	0.52	0.54
159.1677	-58.5814	25.91	17.70	15.94	15.85	16.18	0.55	0.50
159.2324	-58.5895	27.29	19.40	17.83	17.39	17.41	0.51	0.47
159.2871	-58.5967	23.12	19.94	18.76	17.89	17.62	0.98	0.93
159.2605	-58.5935	22.12	19.58	18.23	17.20	17.05	0.56	0.52
159.2450	-58.5923	23.06	21.15	20.76	20.29	20.61	0.93	0.90
159.2441	-58.5877	17.14	14.26	13.44	12.69	12.68	0.63	0.55
159.1681	-58.6149	26.33	21.75	21.29	20.36	20.58	0.92	0.87
159.1476	-58.6086	27.30	21.89	21.17	20.35	20.26	0.48	0.55
159.1885	-58.6175	24.66	20.58	21.17	20.11	20.24	0.85	0.82
159.1937	-58.6357	25.83	18.55	17.32	16.77	16.82	0.52	0.52
159.2066	-58.6373	28.17	22.58	21.01	20.21	20.40	0.63	0.57
159.2028	-58.6372	27.59	19.06	17.59	17.08	17.11	0.59	0.54
159.2122	-58.6391	25.10	18.84	18.61	17.47	17.40	0.57	0.56
159.1804	-58.6164	25.96	21.85	20.95	20.26	20.19	0.59	0.61

R.A.	decl.	F090W	F200W	F335M	F444W	F470N	Prob PRF	Prob RF
159.1830	-58.6359	27.56	19.75	18.48	17.95	17.94	0.59	0.53
159.2529	-58.6446	18.70	15.56	14.09	13.27	13.31	0.96	0.89
159.2508	-58.6445	20.62	17.95	17.07	16.25	16.00	0.83	0.68
159.2399	-58.6435	22.21	19.47	18.76	17.93	17.92	0.96	0.91
159.1671	-58.6322	27.29	19.94	18.79	18.35	18.32	0.56	0.53
159.1768	-58.6357	25.78	17.47	15.95	15.45	15.50	0.58	0.52
159.2390	-58.6439	24.95	21.02	20.88	20.14	20.14	0.48	0.50
159.1621	-58.6348	28.55	23.71	21.24	20.34	20.52	0.65	0.67
159.1626	-58.6350	26.35	22.93	20.73	19.97	20.04	0.73	0.65
159.1750	-58.6369	24.98	17.39	16.25	15.72	15.75	0.56	0.54
159.1708	-58.6364	21.46	19.68	19.52	19.10	19.42	0.50	0.57
159.1612	-58.6354	27.79	20.20	18.80	18.33	18.36	0.51	0.46
159.1825	-58.6380	25.27	16.81	15.67	15.09	15.15	0.67	0.63
159.1917	-58.6398	24.89	21.38	19.62	18.87	18.94	0.78	0.69
159.2633	-58.6490	23.88	21.37	20.87	20.42	20.66	0.49	0.53
159.1534	-58.6331	25.95	17.82	16.49	16.09	16.20	0.61	0.57
159.1911	-58.6402	27.69	20.34	19.17	18.75	18.82	0.53	0.51
159.2717	-58.6443	23.63	21.32	20.90	20.52	20.79	0.66	0.70
159.1734	-58.6313	22.00	13.27	12.40	12.05	12.29	0.66	0.53
159.1448	-58.6239	25.20	21.45	21.65	20.42	20.48	0.66	0.61
159.2078	-58.6317	26.57	22.01	21.41	20.63	20.85	0.81	0.76
159.2679	-58.6392	22.50	20.39	19.94	19.60	19.83	0.59	0.47
159.1958	-58.6305	21.74	19.86	19.42	19.09	19.33	0.57	0.56
159.2383	-58.6359	24.02	20.99	20.39	19.20	19.17	0.66	0.64
159.2402	-58.6361	24.18	21.38	21.00	20.44	20.77	0.52	0.52
159.1596	-58.6264	21.04	19.33	18.97	18.57	18.86	0.93	0.90
159.2429	-58.6367	23.69	20.79	20.50	19.71	19.86	0.54	0.51
159.2494	-58.6376	23.22	19.78	19.73	18.76	18.51	0.73	0.70
159.2609	-58.6426	22.18	19.66	19.21	18.83	19.09	0.48	0.54
159.2097	-58.6328	26.48	21.00	20.24	19.46	19.63	0.73	0.67
159.2161	-58.6344	27.68	23.19	20.76	19.01	18.85	0.61	0.59
159.2530	-58.6392	23.48	21.08	20.56	20.06	20.31	0.45	0.51
159.2145	-58.6349	24.04	20.44	18.35	17.27	17.27	0.94	0.90
159.2071	-58.6345	27.56	21.84	20.69	19.94	20.20	0.57	0.53

R.A.	decl.	F090W	F200W	F335M	F444W	F470N	Prob PRF	Prob RF
159.2340	-58.6381	22.68	20.31	19.86	19.40	19.70	0.71	0.69
159.1860	-58.6322	25.83	18.50	17.31	16.74	16.72	0.52	0.49
159.2433	-58.6395	18.86	16.20	15.32	14.54	14.50	0.81	0.68
159.2041	-58.6348	22.07	18.25	15.99	14.95	16.99	0.98	0.93
159.2092	-58.6354	27.53	19.21	17.56	17.01	17.04	0.58	0.52
159.2520	-58.6386	25.36	21.74	21.33	20.65	20.96	0.53	0.56
159.2133	-58.6323	26.83	20.87	20.24	19.44	19.60	0.57	0.55
159.2681	-58.6500	23.52	21.24	20.83	20.44	20.70	0.66	0.70
159.2429	-58.6472	24.56	21.35	21.07	20.27	20.20	0.77	0.73
159.1846	-58.6647	22.72	20.76	20.46	20.08	20.34	0.58	0.59
159.1963	-58.6665	23.90	19.69	19.54	18.79	18.92	0.57	0.60
159.1677	-58.6645	25.20	21.94	21.68	20.95	21.42	0.65	0.63
159.2043	-58.6696	25.47	22.39	17.70	14.26	16.13	0.59	0.52
159.2088	-58.6704	22.21	20.34	20.08	19.63	19.89	0.83	0.79
159.1518	-58.6635	22.38	19.31	18.64	17.99	18.33	0.50	0.52
159.2003	-58.6695	23.69	21.05	20.21	19.29	19.24	0.81	0.72
159.2020	-58.6699	22.34	18.57	17.93	16.95	17.36	0.84	0.75
159.1977	-58.6709	25.00	21.92	21.59	20.78	21.16	0.77	0.74
159.1990	-58.6656	25.19	20.66	20.42	19.54	19.58	0.60	0.59
159.2148	-58.6734	24.15	20.85	20.49	19.82	20.18	0.52	0.56
159.2223	-58.6755	24.54	20.61	20.79	19.18	19.38	0.91	0.87
159.1950	-58.6722	24.62	20.41	20.38	19.44	20.45	0.82	0.79
159.1578	-58.6683	26.35	22.13	20.20	19.44	19.72	0.96	0.90
159.1702	-58.6709	22.51	20.68	20.39	19.98	20.31	0.92	0.87
159.2033	-58.6763	19.12	16.71	16.25	15.88	16.18	0.66	0.69
159.1591	-58.6729	19.45	17.94	17.48	17.17	17.44	0.41	0.50
159.2107	-58.6802	25.70	22.23	21.84	21.04	21.00	0.71	0.73
159.1596	-58.6740	21.25	17.12	16.44	15.05	17.00	0.69	0.64
159.1591	-58.6741	19.37	15.51	10.54	9.18	13.16	0.56	0.48
159.1959	-58.6717	24.69	21.35	21.27	20.39	21.01	0.74	0.71
159.2654	-58.6498	25.01	21.70	20.69	19.92	19.93	1.00	1.00
159.1987	-58.6643	26.95	21.73	21.58	20.27	19.78	0.65	0.66
159.2023	-58.6611	24.65	20.71	20.77	19.75	20.12	0.87	0.83
159.1917	-58.6410	25.20	20.54	20.18	19.44	19.58	0.53	0.56

R.A.	decl.	F090W	F200W	F335M	F444W	F470N	Prob PRF	Prob RF
159.2003	-58.6422	22.64	18.84	18.63	17.70	18.15	0.89	0.86
159.1486	-58.6364	27.56	22.23	20.26	19.41	19.58	0.61	0.59
159.2043	-58.6435	22.54	13.29	12.25	11.66	11.93	0.56	0.57
159.1901	-58.6426	25.18	17.37	16.09	15.61	15.64	0.59	0.55
159.1855	-58.6422	23.77	15.51	14.12	13.70	13.82	0.59	0.54
159.1893	-58.6429	24.82	17.44	16.29	15.87	15.90	0.53	0.51
159.1874	-58.6441	27.32	19.35	17.81	17.72	17.82	0.55	0.50
159.2229	-58.6509	23.34	20.94	20.38	20.00	20.33	0.47	0.52
159.1854	-58.6599	27.47	22.16	21.66	20.71	21.18	0.88	0.80
159.1776	-58.6455	27.61	22.04	21.09	20.34	20.48	0.57	0.57
159.2365	-58.6593	23.11	21.20	20.74	20.47	20.70	0.39	0.56
159.1377	-58.6476	21.36	19.75	19.37	19.03	19.34	0.67	0.65
159.1579	-58.6521	27.18	21.72	19.52	18.06	18.57	0.71	0.72
159.1577	-58.6522	25.24	22.23	20.26	18.89	18.99	0.65	0.55
159.1998	-58.6575	24.17	20.02	19.89	19.16	19.37	0.57	0.55
159.2517	-58.6648	21.34	16.77	15.02	14.18	14.16	0.97	0.94
159.1965	-58.6580	26.68	21.61	21.54	20.34	21.24	0.86	0.82
159.2449	-58.6654	17.43	14.85	14.05	13.43	13.46	0.93	0.86
159.1319	-58.6516	25.64	21.13	21.12	20.11	20.26	0.91	0.86
159.2044	-58.6529	24.65	21.47	20.87	20.23	20.19	0.40	0.52
159.2376	-58.6351	23.94	21.17	20.70	20.22	20.50	0.51	0.55
159.2294	-58.5597	26.06	17.96	16.51	15.97	15.97	0.59	0.53
159.1804	-58.6280	27.12	21.97	21.37	20.59	20.81	0.72	0.65
159.1679	-58.6186	24.72	21.03	20.58	19.89	20.15	0.54	0.58
159.1894	-58.6214	26.01	21.32	20.92	20.15	20.32	0.64	0.63
159.2453	-58.6284	18.89	15.38	13.92	13.08	13.10	0.99	0.95
159.1843	-58.6212	17.80	16.36	16.09	15.65	15.89	0.80	0.66
159.1584	-58.6181	27.18	22.01	21.24	20.54	20.73	0.51	0.55
159.2085	-58.6243	25.36	20.79	20.07	19.37	19.54	0.54	0.49
159.1680	-58.6193	25.10	20.27	20.21	19.34	19.61	0.82	0.77
159.1792	-58.6207	22.80	18.70	18.58	17.86	18.03	0.54	0.48
159.2088	-58.6246	27.42	20.80	20.23	19.43	19.72	0.56	0.51
159.2051	-58.6241	26.80	22.05	21.45	20.49	20.72	0.91	0.86
159.1461	-58.6157	21.36	19.57	19.19	18.78	19.04	0.93	0.91

R.A.	decl.	F090W	F200W	F335M	F444W	F470N	Prob PRF	Prob RF
159.2564	-58.6308	22.37	18.71	16.65	15.38	15.12	0.78	0.74
159.2118	-58.6258	25.47	20.68	20.00	19.30	19.60	0.56	0.53
159.1503	-58.6185	23.57	21.60	21.06	20.62	20.98	0.57	0.55
159.2368	-58.6292	19.23	16.68	15.73	15.12	15.08	0.66	0.42
159.1679	-58.6208	25.19	21.31	20.59	19.94	20.34	0.47	0.51
159.2201	-58.6275	21.48	18.38	16.72	15.99	16.02	0.93	0.85
159.1732	-58.6217	25.53	22.06	21.91	21.02	21.72	0.78	0.71
159.1669	-58.6210	21.07	19.14	18.81	18.33	18.60	0.90	0.84
159.1724	-58.6217	18.80	16.22	15.36	14.78	14.84	0.97	0.87
159.1803	-58.6230	24.84	20.35	20.44	19.49	19.80	0.97	0.92
159.2105	-58.6268	26.32	21.57	21.15	20.15	20.51	0.92	0.88
159.1917	-58.6231	20.58	17.67	16.76	16.07	16.21	0.57	0.50
159.2148	-58.6242	27.08	21.45	20.84	19.95	20.08	0.75	0.69
159.2282	-58.6257	21.22	16.46	15.14	13.98	14.31	0.62	0.58
159.1912	-58.6211	25.29	21.07	20.98	20.24	20.80	0.64	0.65
159.2042	-58.6194	25.58	21.20	21.42	20.49	20.36	0.72	0.70
159.2030	-58.6194	23.36	19.11	19.11	18.30	18.34	0.55	0.55
159.2003	-58.6191	23.93	20.04	19.91	19.09	19.35	0.72	0.64
159.1832	-58.6170	22.67	19.03	19.27	18.38	18.53	0.75	0.70
159.1509	-58.6131	22.85	18.55	18.37	17.65	17.81	0.59	0.52
159.1624	-58.6148	24.14	20.46	20.25	19.40	19.58	0.72	0.64
159.1464	-58.6129	22.66	20.80	20.35	19.89	20.31	0.68	0.63
159.1697	-58.6158	25.58	20.95	20.50	19.79	20.18	0.59	0.56
159.2155	-58.6217	25.41	20.01	19.75	18.91	19.05	0.69	0.66
159.1860	-58.6182	22.48	20.63	20.33	19.83	20.16	0.85	0.80
159.1939	-58.6194	20.87	19.28	19.08	18.69	18.98	0.54	0.56
159.2093	-58.6213	25.67	19.78	19.46	18.50	18.64	0.78	0.75
159.2622	-58.6279	22.15	19.50	18.65	17.92	17.92	0.85	0.58
159.2073	-58.6213	24.81	19.18	18.94	18.03	18.18	0.76	0.72
159.1975	-58.6202	24.21	19.37	17.04	15.86	16.02	0.66	0.65
159.2299	-58.6244	25.86	22.17	21.60	20.95	20.56	0.39	0.51
159.1922	-58.6199	24.96	20.60	21.01	19.79	20.44	0.89	0.83
159.1738	-58.6177	25.71	21.42	21.02	20.26	20.98	0.65	0.65
159.2664	-58.6292	23.92	21.42	20.89	20.42	20.72	0.48	0.51

R.A.	decl.	F090W	F200W	F335M	F444W	F470N	Prob PRF	Prob RF
159.2582	-58.6283	19.03	15.77	14.37	13.50	13.46	1.00	0.99
159.1628	-58.6167	26.08	22.22	21.71	20.88	21.46	0.76	0.71
159.1744	-58.6181	25.69	21.44	21.00	20.15	20.59	0.81	0.76
159.2072	-58.6228	25.16	19.98	19.51	18.69	18.88	0.73	0.67
159.2339	-58.6297	23.82	21.36	20.87	20.42	20.66	0.51	0.58
159.1820	-58.6234	24.49	20.30	20.20	19.43	19.80	0.71	0.69
159.2443	-58.6288	25.64	22.15	21.73	21.05	20.97	0.39	0.52
159.1552	-58.6225	26.91	22.26	21.58	20.78	21.20	0.74	0.72
159.1887	-58.6263	22.38	20.26	19.86	19.45	19.71	1.00	0.99
159.2450	-58.6315	19.33	16.24	14.95	14.17	14.14	0.95	0.90
159.2215	-58.6308	19.91	18.44	18.15	17.57	17.88	0.64	0.62
159.1801	-58.6261	25.11	20.16	19.70	18.99	19.29	0.52	0.54
159.1457	-58.6209	25.41	21.19	20.70	19.99	20.18	0.53	0.50
159.1624	-58.6227	26.81	22.14	21.57	20.76	21.07	0.81	0.76
159.2626	-58.6349	24.19	21.76	21.27	20.88	21.30	0.49	0.50
159.2131	-58.6280	25.51	21.77	21.89	20.64	20.82	0.89	0.85
159.2245	-58.6301	24.91	15.19	12.27	9.91	10.44	1.00	0.98
159.2101	-58.6300	25.13	20.39	20.52	19.50	19.98	0.95	0.92
159.2126	-58.6283	24.85	20.09	19.98	19.22	19.47	0.65	0.61
159.2621	-58.6348	24.26	21.05	20.06	19.34	19.39	0.85	0.81
159.1744	-58.6261	23.74	20.86	20.43	19.66	20.07	0.71	0.69
159.2223	-58.6320	18.89	13.56	11.92	10.86	11.09	0.64	0.65
159.2709	-58.6345	24.52	21.85	21.38	20.94	21.32	0.51	0.53
159.2475	-58.6347	22.24	19.29	17.81	17.05	17.51	1.00	0.97
159.2677	-58.5830	23.93	21.14	20.55	19.75	nan	0.99	-99.999
159.2214	-58.6061	23.10	19.25	17.97	15.32	nan	0.57	-99.999
159.1936	-58.6677	26.88	21.96	21.68	20.46	nan	0.80	-99.999
159.2270	-58.6069	25.94	22.45	nan	15.71	15.50	0.51	-99.999
159.1846	-58.6602	27.61	22.23	nan	20.84	21.30	0.60	-99.999
159.1851	-58.6465	18.78	16.26	nan	12.91	14.82	0.81	-99.999
159.2038	-58.6621	25.33	21.45	nan	20.54	20.88	0.54	-99.999
159.1832	-58.5954	27.11	21.85	nan	20.49	20.96	0.61	-99.999
159.2082	-58.6525	23.41	20.01	nan	17.37	17.28	0.97	-99.999
159.2126	-58.6792	24.66	20.24	nan	19.66	19.92	0.57	-99.999

R.A.	decl.	F090W	F200W	F335M	F444W	F470N	Prob PRF	Prob RF
159.2313	-58.5986	23.06	17.79	nan	15.08	15.73	0.76	-99.999
159.1798	-58.5603	20.86	17.62	nan	13.89	16.37	0.91	-99.999
159.2969	-58.5923	24.70	20.81	nan	8.98	6.71	0.51	-99.999
159.2300	-58.6057	22.00	19.08	nan	17.19	17.50	0.97	-99.999
159.1967	-58.6362	27.68	nan	14.59	11.42	11.56	0.88	-99.999
159.2358	-58.5672	nan	21.83	20.82	19.91	20.25	0.86	-99.999
159.2354	-58.5673	nan	21.01	19.47	18.75	18.83	0.61	-99.999
159.2385	-58.5682	nan	20.75	18.62	17.52	17.53	0.88	-99.999
159.1620	-58.6329	nan	23.81	21.10	20.19	20.23	0.81	-99.999
159.2386	-58.5686	nan	21.44	19.47	18.55	18.53	0.89	-99.999
159.2158	-58.6348	nan	20.52	17.73	16.64	16.59	0.86	-99.999
159.2566	-58.5722	nan	20.78	19.49	18.60	18.56	0.85	-99.999
159.1752	-58.5622	nan	20.67	20.41	19.60	19.10	0.68	-99.999
159.2448	-58.5709	nan	21.37	19.33	18.56	18.49	0.86	-99.999
159.2541	-58.5721	nan	18.02	15.19	13.99	13.86	0.76	-99.999
159.2393	-58.6139	nan	24.34	21.40	20.65	20.72	0.64	-99.999
159.2462	-58.5716	nan	18.38	15.45	14.25	14.11	0.73	-99.999
159.2359	-58.5705	nan	20.46	19.35	18.52	18.52	0.86	-99.999
159.2529	-58.5730	nan	18.01	16.53	15.78	15.58	0.86	-99.999
159.2449	-58.5721	nan	20.47	18.67	17.94	17.88	0.84	-99.999
159.2215	-58.6124	nan	23.05	19.75	17.82	17.80	0.72	-99.999
159.2587	-58.5710	nan	20.98	17.74	16.28	16.00	0.71	-99.999
159.2136	-58.6191	nan	21.75	20.42	19.73	19.41	0.56	-99.999
159.2444	-58.5663	nan	22.09	19.12	17.82	17.84	0.70	-99.999
159.2132	-58.5642	nan	22.62	21.82	21.09	20.83	0.65	-99.999
159.2362	-58.5607	nan	19.82	17.75	16.72	16.76	0.86	-99.999
159.2538	-58.5632	nan	22.52	19.94	19.23	19.34	0.58	-99.999
159.1775	-58.6027	nan	22.56	21.20	20.45	20.74	0.81	-99.999
159.2243	-58.6104	nan	22.37	18.96	16.62	16.56	0.73	-99.999
159.2419	-58.5622	nan	19.96	16.69	15.20	15.15	0.72	-99.999
159.2099	-58.6204	nan	21.09	20.30	19.53	19.48	0.71	-99.999
159.2722	-58.6169	nan	23.00	21.81	21.12	20.85	0.54	-99.999
159.2257	-58.5606	nan	20.72	18.54	17.84	17.84	0.61	-99.999
159.2645	-58.5656	nan	19.49	17.47	16.62	16.49	0.88	-99.999

R.A.	decl.	F090W	F200W	F335M	F444W	F470N	Prob PRF	Prob RF
159.2215	-58.6107	nan	17.85	15.11	13.68	13.56	0.71	-99.999
159.2199	-58.5605	nan	22.29	20.31	19.50	19.45	0.87	-99.999
159.2268	-58.5614	nan	23.14	21.39	20.53	20.60	0.69	-99.999
159.2459	-58.5641	nan	23.19	20.32	18.72	18.70	0.72	-99.999
159.2168	-58.5609	nan	22.73	21.19	20.51	20.68	0.58	-99.999
159.2217	-58.5616	nan	20.13	18.46	17.77	17.75	0.62	-99.999
159.2502	-58.5651	nan	20.83	17.75	16.56	16.45	0.77	-99.999
159.2195	-58.5616	nan	21.48	19.69	18.99	18.98	0.71	-99.999
159.2646	-58.5674	nan	19.80	16.78	15.48	15.39	0.71	-99.999
159.2381	-58.5651	nan	21.69	20.45	19.56	19.58	0.84	-99.999
159.2484	-58.5664	nan	20.54	17.33	16.05	16.10	0.68	-99.999
159.2643	-58.5685	nan	20.62	18.38	17.56	17.38	0.85	-99.999
159.2395	-58.5656	nan	21.17	19.25	18.41	18.83	0.83	-99.999
159.2426	-58.5660	nan	24.01	20.57	18.83	18.72	0.71	-99.999
159.2244	-58.6196	nan	20.67	19.01	18.35	18.27	0.54	-99.999
159.1714	-58.5991	nan	20.06	17.72	16.88	16.77	0.85	-99.999
159.2412	-58.5678	nan	19.45	16.88	15.80	15.69	0.86	-99.999
159.2199	-58.6603	nan	21.52	20.07	19.07	19.08	0.88	-99.999
159.2218	-58.6129	nan	19.85	16.05	13.81	13.69	0.71	-99.999
159.2303	-58.5831	nan	21.01	19.45	18.61	18.70	0.69	-99.999
159.2204	-58.6168	nan	23.08	20.39	19.57	19.52	0.85	-99.999
159.2441	-58.5850	nan	19.30	15.22	12.69	12.55	0.71	-99.999
159.2359	-58.5841	nan	21.70	19.17	18.20	18.22	0.85	-99.999
159.2462	-58.5862	nan	20.35	18.37	17.29	17.48	0.83	-99.999
159.1982	-58.6365	nan	21.64	19.84	19.09	19.27	0.79	-99.999
159.1519	-58.6349	nan	20.75	18.84	17.99	18.05	0.69	-99.999
159.1762	-58.6377	nan	20.84	19.16	18.47	18.49	0.62	-99.999
159.1745	-58.6372	nan	21.67	18.84	17.97	18.07	0.65	-99.999
159.2220	-58.6162	nan	21.39	17.37	16.01	16.02	0.71	-99.999
159.2126	-58.6148	nan	22.74	21.71	20.56	20.61	0.79	-99.999
159.1824	-58.6379	nan	18.60	17.04	16.33	16.41	0.60	-99.999
159.1534	-58.6343	nan	22.06	19.90	18.99	19.17	0.82	-99.999
159.1984	-58.6366	nan	22.46	20.85	20.04	20.24	0.81	-99.999
159.1629	-58.6351	nan	21.62	19.45	18.76	18.82	0.53	-99.999

R.A.	decl.	F090W	F200W	F335M	F444W	F470N	Prob PRF	Prob RF
159.2197	-58.6600	nan	20.01	18.78	18.09	17.97	0.56	-99.999
159.2341	-58.5894	nan	22.20	19.89	19.01	18.95	0.86	-99.999
159.2075	-58.6382	nan	20.09	18.71	17.80	17.75	0.85	-99.999
159.1648	-58.5814	nan	19.62	18.64	18.02	18.03	0.56	-99.999
159.1748	-58.6361	nan	23.34	21.18	20.57	20.74	0.51	-99.999
159.2060	-58.6399	nan	20.33	19.01	18.19	18.36	0.80	-99.999
159.1942	-58.6366	nan	21.45	18.65	17.15	17.30	0.79	-99.999
159.1889	-58.6376	nan	23.00	20.53	19.40	19.47	0.68	-99.999
159.1546	-58.6333	nan	22.60	20.39	19.37	19.54	0.83	-99.999
159.1739	-58.6357	nan	20.87	18.81	17.74	17.88	0.82	-99.999
159.2201	-58.6148	nan	21.02	18.44	16.28	16.23	0.73	-99.999
159.2113	-58.5898	nan	21.71	20.07	19.30	19.75	0.81	-99.999
159.2320	-58.5924	nan	22.64	20.79	20.07	19.92	0.82	-99.999
159.1597	-58.6338	nan	20.04	18.30	17.54	17.53	0.86	-99.999
159.2288	-58.5985	nan	23.70	20.50	19.00	19.05	0.68	-99.999
159.2201	-58.6136	nan	20.35	17.91	17.04	17.07	0.83	-99.999
159.2171	-58.6148	nan	20.07	18.64	17.83	17.79	0.88	-99.999
159.1607	-58.6367	nan	20.75	18.98	18.20	18.22	0.85	-99.999
159.2226	-58.5941	nan	20.15	17.95	17.29	17.15	0.51	-99.999
159.2353	-58.5832	nan	23.35	20.23	18.86	18.47	0.71	-99.999
159.2216	-58.6127	nan	20.04	16.90	14.83	14.65	0.71	-99.999
159.2210	-58.6188	nan	22.59	17.84	14.72	14.79	0.54	-99.999
159.2073	-58.6386	nan	20.43	17.70	16.82	16.79	0.85	-99.999
159.2399	-58.5738	nan	21.67	20.71	19.94	19.83	0.83	-99.999
159.2396	-58.5741	nan	21.49	20.17	19.39	19.60	0.81	-99.999
159.2360	-58.5737	nan	21.78	20.99	20.09	20.41	0.77	-99.999
159.2053	-58.6350	nan	21.92	20.40	19.59	19.70	0.67	-99.999
159.2508	-58.5762	nan	22.80	18.84	16.97	16.72	0.71	-99.999
159.1952	-58.5703	nan	22.62	21.72	21.11	21.24	0.63	-99.999
159.2361	-58.5756	nan	20.34	19.28	18.57	18.72	0.76	-99.999
159.1413	-58.6426	nan	22.86	21.64	21.00	21.47	0.56	-99.999
159.2205	-58.6269	nan	21.21	19.84	19.06	19.07	0.83	-99.999
159.2158	-58.6366	nan	22.84	20.77	19.77	19.83	0.75	-99.999
159.2274	-58.5772	nan	18.95	17.15	16.44	16.48	0.62	-99.999

R.A.	decl.	F090W	F200W	F335M	F444W	F470N	Prob PRF	Prob RF
159.2451	-58.5780	nan	20.04	18.99	18.11	18.09	0.85	-99.999
159.2090	-58.6371	nan	20.78	18.63	17.48	17.51	0.87	-99.999
159.2039	-58.6356	nan	20.76	19.12	18.26	18.38	0.80	-99.999
159.2351	-58.5791	nan	21.68	19.86	18.76	18.90	0.82	-99.999
159.2453	-58.5804	nan	20.68	18.23	16.41	16.26	0.71	-99.999
159.2429	-58.5803	nan	20.32	18.80	17.99	17.90	0.88	-99.999
159.2222	-58.6133	nan	22.39	19.51	18.35	18.25	0.85	-99.999
159.1504	-58.6356	nan	21.45	18.57	17.18	17.26	0.58	-99.999
159.1604	-58.5932	nan	22.56	21.38	20.75	21.18	0.57	-99.999
159.1765	-58.6412	nan	21.73	20.79	20.18	20.31	0.64	-99.999
159.1998	-58.6359	nan	21.45	19.36	18.53	18.49	0.87	-99.999
159.2237	-58.6137	nan	24.11	21.00	20.16	20.49	0.83	-99.999
159.2186	-58.6327	nan	24.90	20.09	17.87	17.78	0.66	-99.999
159.2215	-58.6133	nan	21.72	19.70	18.86	18.80	0.88	-99.999
159.2791	-58.6308	22.82	19.64	15.28	nan	16.87	0.65	-99.999
159.1486	-58.6560	24.64	20.61	19.62	nan	19.71	0.53	-99.999
159.2388	-58.6100	22.57	19.82	17.84	nan	18.57	0.64	-99.999
159.2556	-58.6624	22.23	18.37	17.01	nan	17.69	0.64	-99.999
159.2497	-58.5734	23.54	13.19	10.24	nan	9.62	0.54	-99.999
159.2329	-58.5930	nan	22.59	nan	18.39	18.18	0.60	-99.999
159.2034	-58.5959	nan	22.68	nan	21.20	20.97	0.60	-99.999
159.2178	-58.6127	nan	24.51	nan	20.38	19.90	0.60	-99.999

Table D.1: The fluxes and J2000 co-ordinates for sources classified as cYSOs in our models along with the probability of an object being a YSO based on the different models. A probability less than 0.5 meant the object was not classified as a YSO in that algorithm, and a probability of -99.999 meant the object had missing data and so was unable to be classified by the RF model.