



HST Survey of the Orion Nebula Cluster in ACS/Visible and WFC3/IR Bands. IV. A Bayesian Multiwavelength Study of Stellar Parameters in the Orion Nebula Cluster

Giovanni M. Strampelli^{1,6} , Massimo Robberto^{1,2} , Laurent Pueyo¹, Mario Gennaro¹ , Carlo F. Manara³ , Elena Sabbi¹ , and Antonio Aparicio^{4,5}

¹ Space Telescope Science Institute, 3700 San Martin Dr., Baltimore, MD 21218, USA; gstrampelli@stsci.edu

² Johns Hopkins University, 3400 N. Charles St., Baltimore, MD 21218, USA

³ European Southern Observatory, Garching bei München, Germany

⁴ Department of Astrophysics, University of La Laguna, Av. Astrofísico Francisco Sánchez, 38200 San Cristóbal de La Laguna, Tenerife, Canary Islands, Spain

⁵ Instituto de Astrofísica de Canarias, C. Vía Láctea, 38200, San Cristóbal de La Laguna, Tenerife, Canary Islands, Spain

⁶ University of California Santa Barbara, Santa Barbara, CA 93106, USA

Received 2023 February 3; revised 2024 March 9; accepted 2024 March 11; published 2024 May 20

Abstract

We have performed a comprehensive study of the Orion Nebula Cluster (ONC) combining the photometric data obtained by the two Hubble Space Telescope Treasury programs that targeted this region. To consistently analyze the rich data set obtained in a wide variety of filters, we adopted a Bayesian approach to fit the spectral energy distribution of the sources, deriving mass, age, extinction, distance, and accretion for each source in the region. The three-dimensional study of mass distribution for bona fide cluster members shows that mass segregation in the ONC extends to subsolar masses, while the age distribution strongly supports the idea that star formation in the ONC is best described by a major episode of star formation that happened ~ 1 Myr ago. For masses $\gtrsim 0.1 M_{\odot}$, our derived empirical initial mass function (IMF) is in good agreement with a Chabrier system IMF. Both the accretion luminosity (L_{acc}) and mass accretion rates (\dot{M}_{acc}) are best described by broken power-law relations. This suggests that for the majority of young circumstellar disks in this cluster the excess emission may be dominated by X-ray-driven photoevaporation by the central star rather than external photoevaporation. If this is the case, the slopes of the power-law relations may be largely determined by the initial conditions set at the onset of the star formation process, which may be quite similar between regions that eventually form clusters of different sizes.

Unified Astronomy Thesaurus concepts: Star forming regions (1565); Star formation (1569); Low mass stars (2050); Brown dwarfs (185); Bayesian statistics (1900); Markov chain Monte Carlo (1889)

Supporting material: figure set, interactive figure, machine-readable tables

1. Introduction

The Orion Nebula Cluster (ONC) represents one of the richest (~ 2000 members in the inner 2 pc radius; McBride & Kounkel 2019) and youngest (~ 2 Myr; Jeffries et al. 2011; Reggiani et al. 2011; Jerabkova et al. 2019) star-forming regions (SFRs) within 2 kpc (~ 402 pc; Kuhn et al. 2019) from the Sun (Lada & Lada 2003; Portegies Zwart et al. 2010). Due to its proximity and modest foreground extinction ($A_V \sim 1$; Scandariato et al. 2011), the ONC has been thoroughly studied at visible and near-infrared wavelengths and is therefore regarded as an ideal laboratory in which to investigate, down to very low mass objects, critical aspects of star and planetary formation, such as the initial mass function (IMF), mass segregation and early dynamical evolution, radiative feedback, and protoplanetary disk photoevaporation and evolution.

In this paper, we combine for the first time the data coming from the two Hubble Space Telescope (HST) Treasury programs (GO-10246 and GO-13826; P.I. M. Robberto) that targeted the ONC. Taking advantage of the extremely accurate photometry provided by the HST in a wide selection of filters, we adopt a Bayesian approach with a Markov Chain Monte Carlo (MCMC) strategy to estimate the main properties of each star in the cluster (e.g., mass, extinction, age, distance, and accretion) according to

the BT-Settl family of isochrones. A Bayesian approach allows incorporating as priors other available information on the source effective temperature, distance, reddening, membership, etc., obtaining a most accurate and consistent estimate of the stellar parameters.

Our first aim is to assess the viability of this technique, considering that our data set is largely based on archival data, taken at different epochs, and the number of filters available for each source is not uniform across the sample. Concentrating on the most reliable sources, we then perform a deeper analysis of the star formation and evolution history of the ONC.

The paper is organized as follows: The observations are presented in Section 2, including a new Advanced Camera for Surveys (ACS) photometric catalog based on the most recent instrument calibration data. In Section 3 we introduce our Bayesian code to perform spectral energy distribution (SED) fitting. In Section 4 we present our final catalog of bona fide cluster members, the derived stellar parameters, and the mass accretion luminosities and rates. In Section 5 we discuss the implication of our findings. Our conclusions are summarized in Section 6.

2. Data Sets and New ACS Photometry

The HST program GO-10246, executed between 2004 October and 2005 March, observed the ONC with the F435W, F555W, F658N, F775W, and F850LP filters of ACS/WFC; the F336W, F439W, F656N, and F814W filters of WFPC2; and the



Original content from this work may be used under the terms of the [Creative Commons Attribution 4.0 licence](#). Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.

F110W and F160W filters of NICMOS (Robberto et al. 2013). The richest data set is the one provided by ACS, which includes 3399 unique sources, most of them observed two or more times in different HST visits, separated by time intervals ranging from a few hours to several months. Robberto et al. (2013) list all individual detections, a grand total of 8185 entries, each with multicolor observations. In 2015 the HST program GO-13826 returned to the ONC with the F130N and F139M filters of WFC3 (Robberto et al. 2020). As shown by Figure 1, the areal coverage is similar for the ACS, WFPC2, and WFC3 data, while the NICMOS observations sampled about 1/4 of the field owing to the small size of the detector.

In Table 1 we show the number of sources detected in each filter. We will use this data set as our reference catalog.

Given both the advances in the ACS calibration and the possibility of obtaining more accurate data on close pairs, we reanalyzed the ACS data set to derive updated aperture photometry. In Table 2 we list the main parameters of the photometric system for the five ACS bandpasses including the most recent assessment of the zero-points at the epoch of observations, as provided by the *ACS Zero-points Calculator*.⁷ Compared to the original zero-points implemented by Robberto et al. (2013), the new values show a difference between -0.002 and 0.019 mag, with the smallest difference in the F555W filter and the highest one for both the F775W and F850LP filters.

Aperture photometry was obtained using the *aperture photometry* package from the *StraKLIP* pipeline (Strampelli et al. 2022). This tool allows the detection and subtraction of the signal from sources very close to the primary target to perform accurate photometry on both components. In this paper we shall focus on the main population of isolated stars, leaving the analysis of close pairs to a future paper.

The individual measures obtained in different visits were finally averaged after rejecting those with high uncertainty due to the residual detector defects or the presence of cosmic rays in the immediate vicinity.

In Figure 2 we plot our final averaged magnitudes with their estimated uncertainties for all sources in the catalog in the five ACS filters. Saturation starts at about $m_{435} = 16$, $m_{555} = 15.75$, $m_{658} = 12.25$, $m_{775} = 15.25$, and $m_{850} = 14.25$, corresponding to 0.7 , 0.5 , 1.2 , 0.14 , and $0.13 M_{\odot}$, respectively, assuming the BT-Settl 1 Myr isochrone at 400 pc, without extinction or accretion. At the other extreme, the 5σ sensitivity limits ($\sigma_{\text{mag}} \simeq 0.2$) are at $m_{435} = 23.77$, $m_{555} = 24.07$, $m_{658} = 20.15$, $m_{775} = 23.84$, and $m_{850} = 23.01$, roughly corresponding to 0.03 , 0.02 , 0.04 , 0.005 , and $0.003 M_{\odot}$, respectively, under the same assumptions. Table 7 provides the saturation/sensitivity limits as a function of extinction and detectable mass. All magnitudes are in the Vega system. Apart from ACS, for the other instruments and passbands we directly adopted the catalogs presented by Robberto et al. (2013, 2020).

Since most observations were not carried out simultaneously, variability represents a source of uncertainty. According to Herbst et al. (2002), about half of the stars brighter than $I \simeq 16$ show peak-to-peak variations of ~ 0.2 mag or more. While this is typically not the range of magnitudes probed by our deeper HST observations, it suggests that variability randomly affects the quality of the fit. In principle, systematic uncertainties could be reduced using colors combining measures obtained with the same instrument within

the same HST orbit, i.e., within about 45 minutes. Testing this approach, we found that in most cases having fewer data points with uncertainties added in quadrature precludes achieving any significant gain compared to the case where the SED fit is performed using single magnitudes. We have therefore carried out our analysis ignoring variability, at least initially, returning later to the sources that may not pass our tests on the quality of the fit because of spurious photometric data for a specific epoch of observation. Table 3 shows a sample of our final photometric catalog for all 3399 sources, fully available in electronic format. The table is organized as follows: the first column reports the ID as provided from the *StraKLIP* pipeline for cross-identification, while the following columns report the average number of saturated pixels, number of identified bad pixels, magnitude, and uncertainty for each filter.

3. Analysis

3.1. Bayesian Analysis

For our analysis we assume that the main unknown variables that characterize each source are mass, extinction, age, distance, and the accretion parameter (SP_{acc}), representing the fraction of stellar bolometric luminosity that can be attributed to accretion. To determine them, we perform SED fitting comparing the observed photometry to synthetic photometry derived from theoretical models of the BT-Settl family (see Section 3.1.1). In order to obtain a probability distribution for each parameter, our fitting procedure adopts a Bayesian approach with an MCMC algorithm, represented schematically in Figure 3 and based on the following main steps:

1. At the start of the MCMC run, five parent distributions are generated uniformly spanning the parameter space presented in Table 4.
2. For each star, a user's defined number of *walkers* is generated, limited in this work to a maximum of 100. Then, the following loop is iterated for each walker (referred to as the “walker loop” in Figure 3):
 - (a) From each distribution discussed above (see also Table 4), a random value is extracted to provide the initial state for the walker in that variable.
 - (b) The *observed magnitudes* of the star under consideration are pulled from the catalog and parsed to the fitting routine as fixed ingredients for the simulation. If one or more filters show signs of saturation, they are discarded from the fit.
 - (c) The five parameters randomly extracted from their parent distributions are combined with the models (see Section 3.1.1) to obtain synthetic magnitudes.
 - (d) The synthetic and observed magnitudes are iteratively compared to find the set of parameters that better reproduces the observations. This optimization requires (1) priors on the fitting parameters (see Section 3.1.3), (2) a likelihood function to compare observations and simulations (see Sections 3.1.2), and (3) an algorithm for sampling the posterior distribution of the fit parameters (see Section 3.1.4).
 - (e) The walkers are allowed to evolve until convergence is reached or a maximum number of steps is executed (2000 in our case). The resulting posterior distribution of the parameters is saved for analysis, and the routine moves to the next source.

⁷ <https://acszeropoints.stsci.edu/>

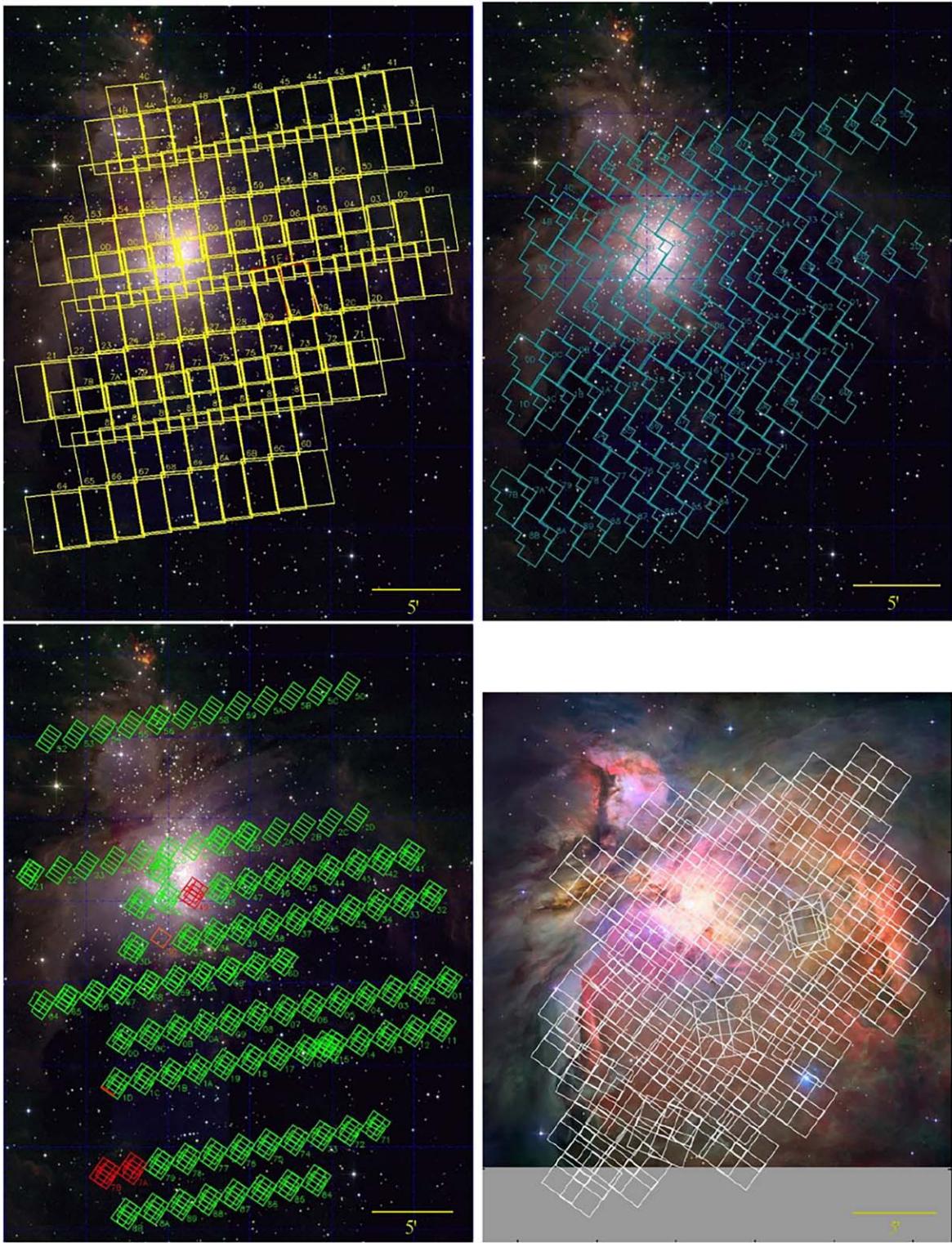


Figure 1. Total area coverage of the different instruments adopted in this work. From top to bottom (left to right): ACS, WFPC2, NICMOS (superimposed onto a color-composite JHK image of the Orion Nebula from the Two Micron All Sky Survey; Robberto et al. 2013), and WFC3 (superimposed onto a mosaic obtained from both space and ground telescope data; Robberto et al. 2020). For reference, the ACS mosaic covers $\sim 627 \text{ arcmin}^2$, while WFC3-IR covers $\sim 486 \text{ arcmin}^2$.

In the following subsections, we detail the main steps of the procedure.

3.1.1. Models

For our analysis, we use the evolutionary models and spectra of the BT-Settl family (AGSS2009 models using the Asplund

et al. 2009 solar abundances)⁸ spanning the age range between 1 Myr and 10 Gyr and masses from 0.005 to $1.4 M_{\odot}$. These values encompass the range of masses and ages appropriate for our (saturation-limited) ONC sample and noncluster

⁸ Downloaded from <http://svo2.cab.inta-csic.es/theory/newov2/>.

Table 1
Number of Sources Detected in Each Filter

F435W	F555W	F658N	F775W	F850LP	F130N	F139M	F110W	F160W	F336W	F439W	F656N	F814W
1264	1525	1316	2466	2725	2728	2897	425	518	549	627	1018	1365

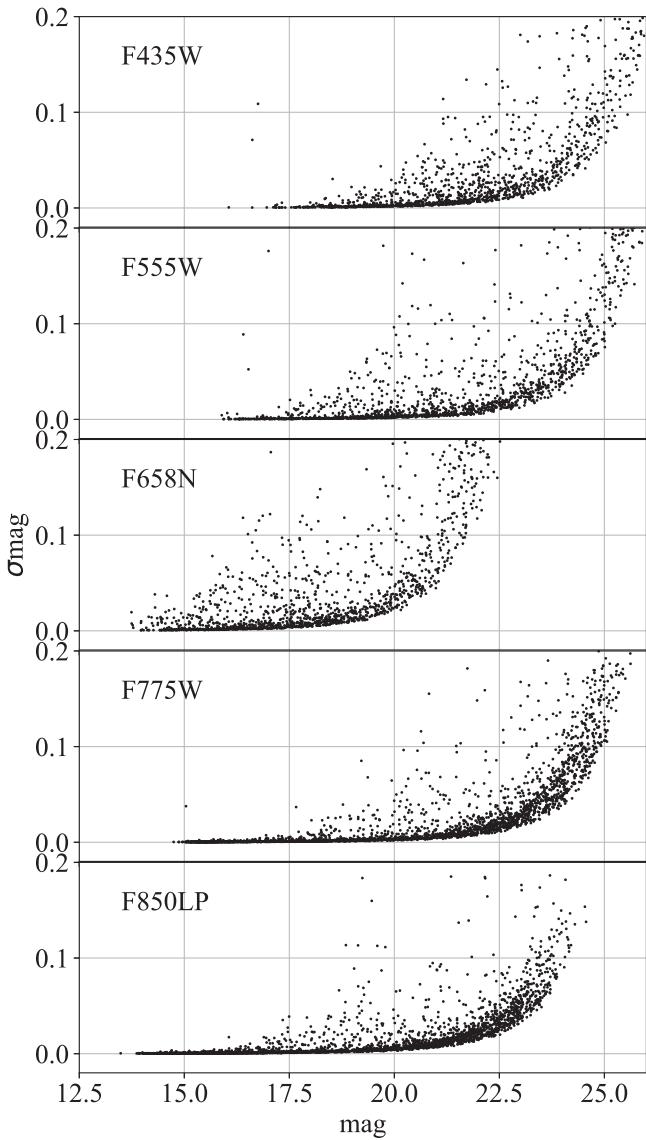


Figure 2. Photometric errors as a function of magnitude for the five ACS filters.

contaminants, based on the Besançon model (see below). The fitting routine requires a uniform grid of synthetic photometry in our filters, which we have created using Synphot (STScI Development Team 2018) and stsynphot (STScI Development Team 2020). In order to estimate the possible presence of accretion luminosity and derive mass accretion rates, we use a slab model described by Manara et al. (2012) where the combination of optically thin emission generated in the preshock region and optically thick emission generated by the heated photosphere is reproduced by combining a CLOUDY spectrum for a standard HII region with an 8000 K blackbody, respectively contributing about 1/4 and

3/4 of the total accretion luminosity (Gullbring et al. 2000). For the far-UV part of the accretion spectrum ($\lambda \lesssim 3000 \text{ \AA}$), following France et al. (2011), we prescribe a linear decrease of flux at wavelengths shorter than the Balmer jump.

To redden the synthetic photometry, we adopt the Cardelli et al. (1989) reddening law parameterized with $R_V = 3.1$.

3.1.2. Likelihood

To compare observations and simulations, we use a likelihood defined as

$$\mathcal{L}(\text{obs}|\theta) = \prod_{i=0}^{N_{\text{mag}}} e^{-0.5 \times \left(\frac{\text{mag}_{i,\text{obs}} - \text{mag}_{i,\text{mod}}(\theta)}{\text{emag}_{i,\text{obs}}} \right)^2}, \quad (1)$$

where i runs through the different magnitudes available for each star and the labels obs and mod represent the observed and model magnitudes (mag) with their uncertainties (emag). The parameter θ instead represents the whole set of fitting parameters corresponding to the different $\text{mag}_{i,\text{mod}}$.

3.1.3. Priors

We divide our sources into two classes of objects, each with its own specific priors:

1. Class A: sources having both a spectral type from Hillenbrand (1997) or Hillenbrand et al. (2013) and a distance estimate from Gaia DR3 (Gaia Collaboration et al. 2021). This class has strong priors that allow pinpointing with high accuracy both luminosity and mass. In particular:
 - (a) For each star, a prior on the mass is derived from the Hillenbrand et al. (2013) temperature when available, or converting the spectral type from Hillenbrand (1997) to effective temperatures using the Luhman et al. (2003) relation. The prior is assumed to be Gaussian, centered on the mass derived from the effective temperature using the model isochrones and with standard deviation given by the mass spread corresponding to two spectroscopic subclasses (i.e., the typical uncertainty of the spectral type reported by Hillenbrand et al. 2013).
 - (b) Similarly, the prior on the distance is a Gaussian centered on the reported parallax from Gaia DR3 (Gaia Collaboration et al. 2021) and standard deviation given by the corresponding parallax uncertainty.
2. Class B: these sources, typically fainter, miss either an independent estimate of the spectral type, a Gaia parallax, or both. This is a less robust sample owing to the much larger uncertainties on the temperature and/or distance. Still, also for this class, we can define priors on the basis of the results for Class A as follows:
 - (a) For the mass prior, if the source has no independent spectral type, we start combining the overall mass distribution determined for Class A objects. Since this sample is heavily biased toward the ONC, we

Table 2
ACS Visit Strategy and Photometric System from Robberto et al. (2013)

FILTER	PHOTPLAM (Å)	PHOTFLAM (erg cm ⁻² s ⁻¹ Å ⁻¹)	Ground Equivalent	Exposure	Integration Time (s)	Zero-point (Vega mag)
F435W	4329.2	3.113E-19	Johnson <i>B</i>	6	420	25.776
F555W	5360.2	1.948E-19	Johnson <i>V</i>	9	385	25.722
F658N	6584.0	1.977E-18	Broad H α	1	340	22.378
F775W	7693.9	9.928E-20	Sloan <i>i</i>	8	385	25.275
F850LP	9034.6	1.503E-19	Sloan <i>z</i>	7	385	24.345

Note. The zero-points for each filter are obtained from the STScI ACS *Zero-points Calculator*.

complement it by adding the distribution of masses obtained from the Besançon model of stellar population synthesis of the Galaxy in the direction of the ONC.⁹

- (b) For the distance prior, if the distance is not provided in Gaia DR3, we take a similar approach to the one adopted for the mass. We combine the distributions of parallaxes obtained for Class A objects with the distribution of distance values obtained from the Besançon model (converted to parallax).

Given the uncertainty on the membership, ONC or Galaxy, for Class B objects, we also define two other priors: one on the age, and another on the extinction.

3. To create the prior on the age, we start combining the distributions obtained for Class A objects and the distribution of age values obtained from the Besançon model.
4. For the prior on the extinction, we take only the distributions obtained for Class A objects. We do not adopt the extinction distribution provided by the Besançon model, as it does not account for the extinction caused by the OMC-1 molecular cloud, which can be both very large and nonuniform over the ONC field.

To obtain the final prior probability distribution (i.e., the probability distribution that would link the a priori knowledge about a variable before evidence is taken into account), we applied a Gaussian kernel density estimator¹⁰ to all distributions for both Class A and Class B. From now on we will refer to these resulting probability density functions as the priors, for short.

3.1.4. MCMC and Posterior Probabilities

To derive the distributions of parameters that best describe our observations, we use *emcee*, the Foreman-Mackey (2016) Python implementation of the Goodman & Weare (2010) affine-invariant sampler. For each given set of parameters θ , we evaluate the corresponding value of the magnitudes from the models and then use the likelihood and prior discussed above to obtain the posterior probability distribution for θ , ensuring that the individual chains converge and are thinned to retain uncorrelated samples. Basically, this approach consists in picking different points from the sample, in every n th step. As we are dividing these points from the overall Markov chain, the dependence becomes smaller and we can achieve a mostly independent sample.

⁹ https://model.obs-besancon.fr/modele_disci.php

¹⁰ <https://scikit-learn.org/stable/modules/generated/sklearn.neighbors.KernelDensity.html>

The following convergence criteria are checked every 100 iterations:

1. The chain must be longer than 100 times the estimated autocorrelation time, τ_f , i.e., the number of steps needed before the chain loses information about its starting properties.
2. The estimate of τ_f has changed by less than 1% since the previous check.

As long as these criteria are met, the routine keeps running until a total of 2000 post-convergence iterations are performed, or the maximum number of steps is elapsed. Then, the routine exits the iteration loop and saves the posterior distribution of the parameters to a file before moving to the next source, until the end of the catalog.

3.1.5. Bayesian Analysis Products

Starting from a catalog of 3399 unique sources, we perform Bayesian SED fitting using the MCMC algorithm to evaluate the star's principal parameters, i.e., mass, extinction, age, distance, and the SP_{acc} (see Section 3.1).

Figure 4 show examples of the SED fitting (left) and corner plots (right) representing the posterior probabilities for the parameters of two random stars. The red line in the left panel shows the original spectrum evaluated for the star, while the blue line shows the slab model adopted in this work scaled by the $\log_{10} SP_{acc}$ parameter. The black line instead represents the spectrum of the star corrected by the slab model. The filters adopted for each fit are shown as circles color-coded by instrument. On average, our tests show that we are able to recover the spectral type determined by Hillenbrand (1997) and Hillenbrand et al. (2013) within two subclasses (i.e., ± 200 K).

Figure 5, instead, shows examples of problematic fitting, generally due to the presence of a bimodal solution in one or more parameter posterior distributions, too few filters available to properly pinpoint the right spectrum to the photometry, the presence of an excess of flux in the filters most sensitive to accretion (e.g., F336W, F656N, F658N) not reflected in the other filters, or any combination of the above. This generally produces an over- or underestimate of the spectrum for the source.

Each corner plot shows the median value, which we adopted as the best-fit estimator, as well as the 68% credible intervals of each fitted parameter's marginal posterior distribution. We define this interval so that the 68% credible interval contains 68% of the total probability, with $(100 - 68)/2 = 16\%$ of the remaining probability on either side. This definition coincides with the usual 1σ interval for a Gaussian distribution. Note that the posterior distributions of our parameters are generally not

Table 3
Sample of the Photometric Data Available for All the Sources Included in This Work

ID	spx435	bpx435	m435 (mag)	spx555	bpx555	m555 (mag)	spx658	bpx658	m658 (mag)	spx775	bpx775	m775 (mag)	spx850	bpx850	m850 (mag)
1				0	2	$24.591^{+0.106}_{-0.106}$			0	2	$21.836^{+0.013}_{-0.013}$	0	2	$20.827^{+0.011}_{-0.011}$	
8	0	4	$21.583^{+0.008}_{-0.008}$	0	4	$19.725^{+0.003}_{-0.003}$	0	4	$17.609^{+0.007}_{-0.007}$	20	4	$16.387^{+0.001}_{-0.001}$	18	4	$15.274^{+0.001}_{-0.001}$

(This table is available in its entirety in machine-readable form.)

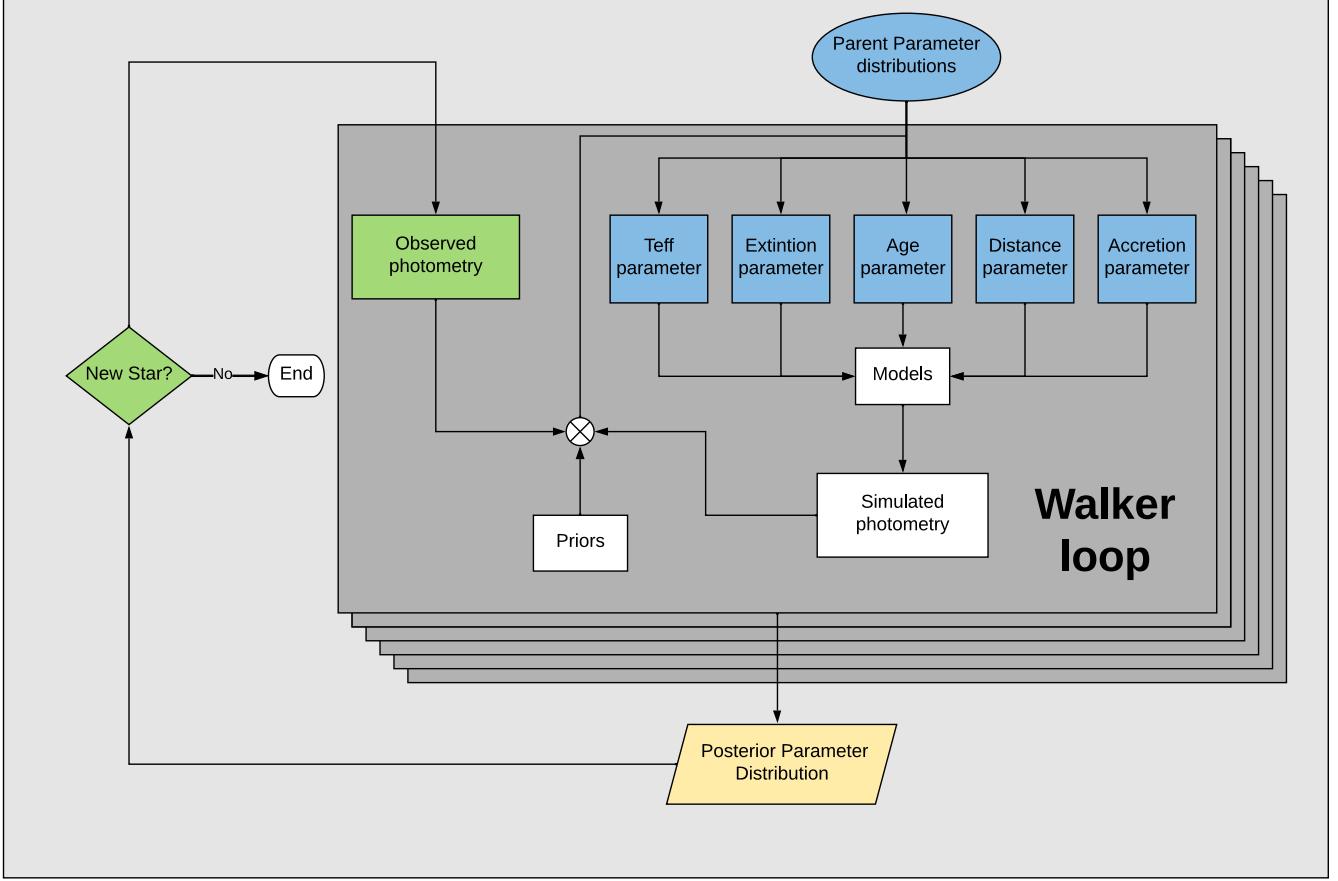


Figure 3. Flowchart for the ONC source parameter estimations. The light-gray area represents the whole MCMC run, while the dark-gray area represents each single walker loop. The ensemble of all walker loops represents the simulation for one single star. The green blocks represent fixed ingredients for the simulation, while the fitting parameters are shown in blue and the key steps of the Bayesian approach are shown in white. The yellow block shows the stored final output of each iteration. Note that all stars are treated completely independently until the endpoint.

Table 4

Range of Values Explored for Each Parameter during Our MCMC Analysis

$\log_{10}(\text{Mass})$	$\log_{10}(A_V)$	$\log_{10}(\text{Age})$	$\log_{10}(\text{SP}_{\text{acc}})$	Parallax
[-2, 1]	[-1, 2]	[0, 4]	[-5, 1]	[0.02, 6]

Note. From left to right: the stellar mass in solar units, visible extinction, stellar age in Myr, the rescale factor for the slab model we use to model accretion, and the parallax (corresponding to an interval between 167 pc and 50 kpc).

Gaussian or symmetric in most cases. The marginal distributions for the single parameters are still useful for defining the parameter credible intervals but do not capture the entire information available in the full posterior distributions, with their correlation. For detailed studies of individual sources, a complete gallery is available.

4. Results

4.1. Membership Selection

We will discuss in the following our approach to cluster membership selection. On the basis of our analysis, we determine ONC membership as follows. First, we eliminate all the sources where the MCMC fit could not converge, removing 468 sources from our sample and leaving us with a total of 2931 candidates. Then, to be an ONC member, the following is required:

1. The derived parallax of a candidate member must be compatible with 2.5 ± 0.3 mas (402^{+57}_{-45} pc), i.e., 1σ from the peak of the Gaia DR3 parallax distribution for all the sources falling in our field of view. This leaves us with 1278 candidate sources from the initial 2931. Taking full advantage of the posterior distribution of the parallax solution (PDPS) of each source, we evaluate the ratio of the area under the curve (AOC) of PDPS that lies within the same 2.5 ± 0.3 mas interval over the total AOC of the PDPS of each source. This ratio (which we will call *area ratio* or AR) provides us with an estimate of the probability that the two parallaxes are compatible.
2. The derived age of a candidate member must be compatible with a recent star formation event, i.e., ≤ 10 Myr. This captures the age range of interest, as previous works (e.g., Jeffries et al. 2011; Da Rio et al. 2012; Beccari et al. 2017) show that the typical age for this cluster is ~ 2 Myr, with an age spread at most of a few megayears. In particular, we find that $\sim 75\%$ of our candidates have an age $\lesssim 2$ Myr, while if we cut the age distribution at $\lesssim 10$ Myr, $\sim 88\%$ of the selected sources are younger than $\lesssim 2$ Myr and $\sim 97\%$ are younger than $\lesssim 5$ Myr. The age selection leads to discarding another 199 sources from our candidate source catalog, leaving us with a candidate cluster catalog of 1079 sources.
3. After visually inspecting each SED fit of these 1079 cluster members, we exclude 30 candidates owing to the

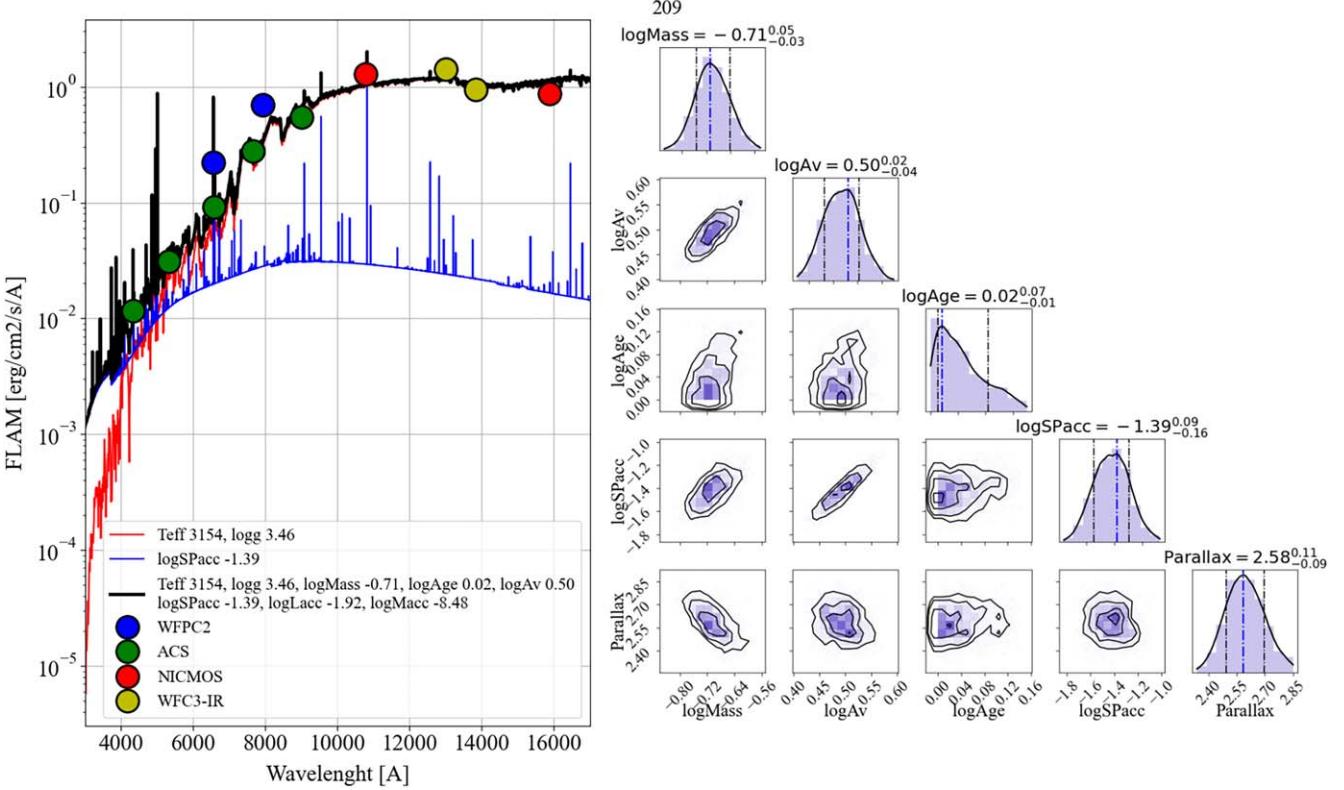


Figure 4. Sample of SED fitting (left) and corner plots (right) for cluster target sources in the catalog. The red line in the left panel shows the original spectrum for the star, while the blue line shows the slab model adopted in this work scaled by the $\log_{10} \text{SP}_{\text{acc}}$ parameter. The black line instead represents the spectrum of the star combined with the slab model. The filters utilized for each fit are shown as circles color-coded by their respective instrument. On the right panel, the peak (blue dotted line) and the limits of the 68% credible interval (black dotted lines) are reported for each parameter. The black line in the histograms shows the Gaussian kernel density estimate.

(The complete figure set (1049 images) is available.)

gross inconsistency between the observed photometry and the best fit (see examples in Figure 5), reducing the number of candidate cluster members to 1049.

This sample of 1049 sources distilled from the initial 3399 stars, i.e., $\sim 31\%$ of the initial sample, will represent our reference catalog of candidate cluster members (or reference catalog for short). Their AR distribution shows a peak at a median value $\text{AR} \sim 0.89^{+0.11}_{-0.17}$. This confirms that the majority of members in our reference catalog have high probability that their estimated distance distribution is compatible with the Gaia DR3 distance distribution for the ONC of 402^{+57}_{-45} pc. To assess the number of stars that may have been incorrectly classified as background objects, we generated a synthetic photometry catalog of background stars using Synphot. Starting from the stellar parameters T_{eff} , $\log Z$, $\log g$, and distance provided by a Besançon model of the Milky Way at the coordinates of the ONC, we determined the photometry of each source using *phoenix* spectra. For each filter, the magnitudes were extracted after reddening the spectrum using the sum of the extinction provided by the Besançon model and Scandariato's extinction map for the OMC (Scandariato et al. 2011), for a randomly generated position within our field of view. Comparing the number of background stars with the Besançon predictions, we are able to account for $\sim 90\%$ of the predicted background stars in the ACS redder filters and $\sim 80\%$ in the ACS blue and WFC3-IR filters. Overall, this test suggests that the significant fraction

of background sources determined by our approach is compatible with the predictions from the Besançon model, within the uncertainties on the extinction due to the nonuniform column density of the OMC.

As anticipated in Section 2, we cross-match the ACS (yellow) and WFC3 (red) surveys. Discussing the WFC3 data set, Robberto et al. (2020) present a detailed Bayesian analysis to disentangle bona fide low-mass stars and substellar cluster members from background sources based on the presence of water in absorption shown by F130N–F139M color.

Figure 6 shows the overlap between ACS and WFC3 surveys and their respective characterization as candidate cluster (green) and noncluster members (blue—either background, foreground, or where the sorting approach failed). It turns out that of 4504 WFC3 sources, only 2945 are also present in our ACS catalog (out of a total of 3399 ACS initial sources), or, in other words, $\sim 87\%$ of ACS sources overlap with WFC3. We must remind the reader that in the region of overlap between ACS and WFC3 some sources can be labeled cluster in one survey and noncluster in the other, and vice versa. Indeed, the total number of overlapping sources (2945) is not equal to the number of overlapping candidate clusters (757) plus the number of the overlapping noncandidate clusters (1237), as we need to include also the ACS candidates labeled as noncluster by WFC3 (225) and the WFC3 candidates labeled as noncluster by ACS (726).

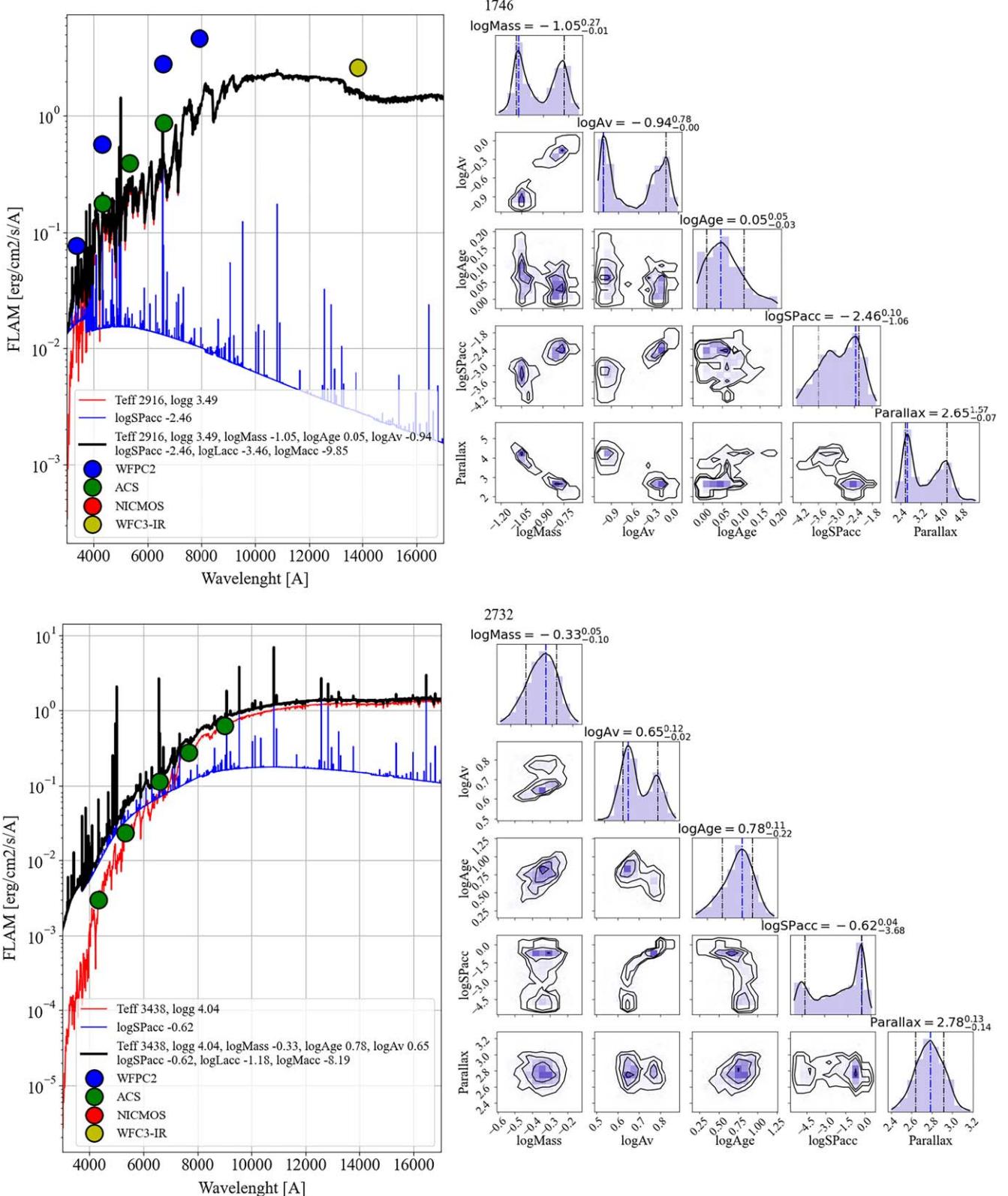


Figure 5. Similar to Figure 4, but showing two examples of problematic outcomes from our Bayesian fit. The top row shows a source detected in multiple filters, whose solution appears bimodal for almost all parameters and overall fails to match the observations. The bottom row instead shows the case of a source detected in only the ACS filters that appears to be bimodal in the A_v and SP_{acc} parameters and that produced an overestimated spectrum of accretion (blue line) that cannot be accounted for by the other filters.

From the analysis of the ACS reference catalog we find 67 sources not listed in the WFC3 catalog, while 225 sources are instead discarded by Roberto et al. (2020) on the basis of their

position on the color-magnitude diagram (CMD). Those are recovered by our MCMC analysis thanks to a compatible solution with the Gaia DR3 estimated distance and age of the cluster.

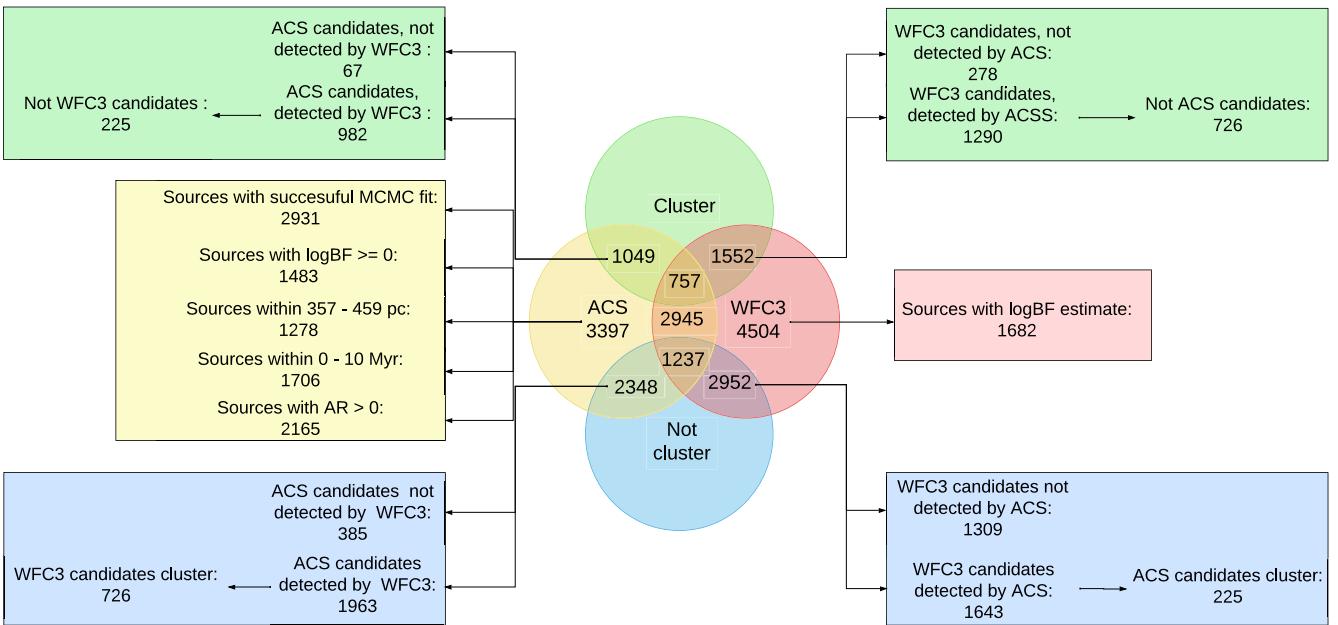


Figure 6. ACS/WFC3 ONC Venn diagram showing the overlap between ACS and WFC3 surveys and their respective characterization as candidate cluster/noncluster members.

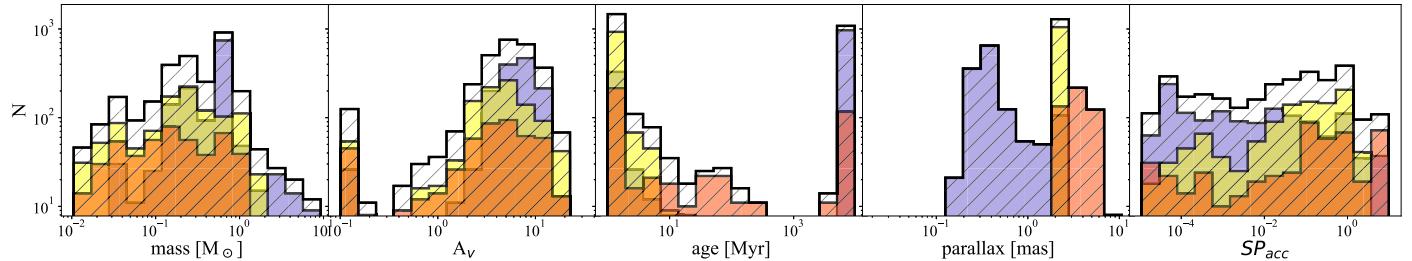


Figure 7. Histogram distributions for the five parameters (i.e., mass, extinction, age, parallax, and SP_{acc}) obtained through the Bayesian analysis described in Section 3.1 for all the sources in the catalog (white), cluster (*reference catalog*—yellow), background (blue), and foreground (red) members. Note: larger parallaxes mean closer sources to the Sun.

Analyzing the WFC3 sample, out of 4504 sources, we identify 1552 candidate cluster members on the basis of the Bayes factor $\log(BF) >= 0$, corresponding to a $\sim 34\%$ ratio, very similar to the $\sim 31\%$ we reported for the ACS survey. Of these 1552 sources, 278 are not listed in our initial ACS catalog. Of the remaining 1290 WFC3 candidate sources, 726 are not identified as ACS candidate sources by our method, either because the MCMC fit puts them out of the cluster (418 cases) or because it fails to converge to a solution.

Since we cannot exclude that both the 308 sources missing the MCMC fit and the 278 sources not detected by our survey (but still identified as candidate clusters by Robberto et al. 2020) are actually part of the cluster, we will add them back in, creating a second *expanded catalog* of candidate cluster members. Lacking the full distributions provided by the MCMC analysis, we will adopt the values for the mass, extinction, and photometry from Robberto et al. (2013) for these extra sources, but since only 497 of the 586 missing sources have an estimate of the mass and extinction, we will add only this last sample to our expanded reference catalog, counting 1546 sources in total.

Figure 7 shows the histograms of the estimated values of each parameter for all the sources in our catalog. The mass, age, and distance panels, in particular, show a mix of low-mass ($\lesssim 0.3 M_\odot$) and more massive ($\sim 1 M_\odot$) stars, of very young

($\lesssim 10$ Myr) and very old ($\gtrsim 1$ Gyr) stars, as well as close ones ($\log(\text{parallax}) \sim 0.4$ mas corresponds to ~ 400 pc) and far away ($\log(\text{parallax}) \gg 0.4$ mas). All these distributions are compatible with two distinct populations: one younger and closer to us, the ONC (yellow), and one older and far away (the background stars—blue). Moreover, the parallax distribution shows a small population of sources with a parallax much smaller than the one expected for ONC cluster members (red— $\log(\text{parallax}) > 0.4$ mas) compatible with the presence of an additional population of foreground objects that are indeed expected on the line of sight.

For the candidate cluster distributions from the *reference catalog*, the median values of each distribution are mass = $0.21 M_\odot$, $A_V = 4.35$ mag, age = 1.07 Myr, parallax = 2.52 mas, and $SP_{acc} = 0.08$.

Figure 8 shows the CMDs for eight combinations of simultaneously observed bandpasses, i.e., for ACS, m435 versus m435–m555, m555 versus m555–m775, m775 versus m775–m850, and m850 versus m775–m850, on the top row; for WFPC2, m336 versus m336–m439, m439 versus m439–m814, m656 versus m656–m814; for NICMOS, m110 versus m110–m130; and for WFC3-IR, m130 versus m130–m139, on the bottom row. All sources with photometry in each given pair of filters are plotted, together with our model 1 and 10 Myr isochrones and the $A_V = 1$ reddening vector, for comparison.

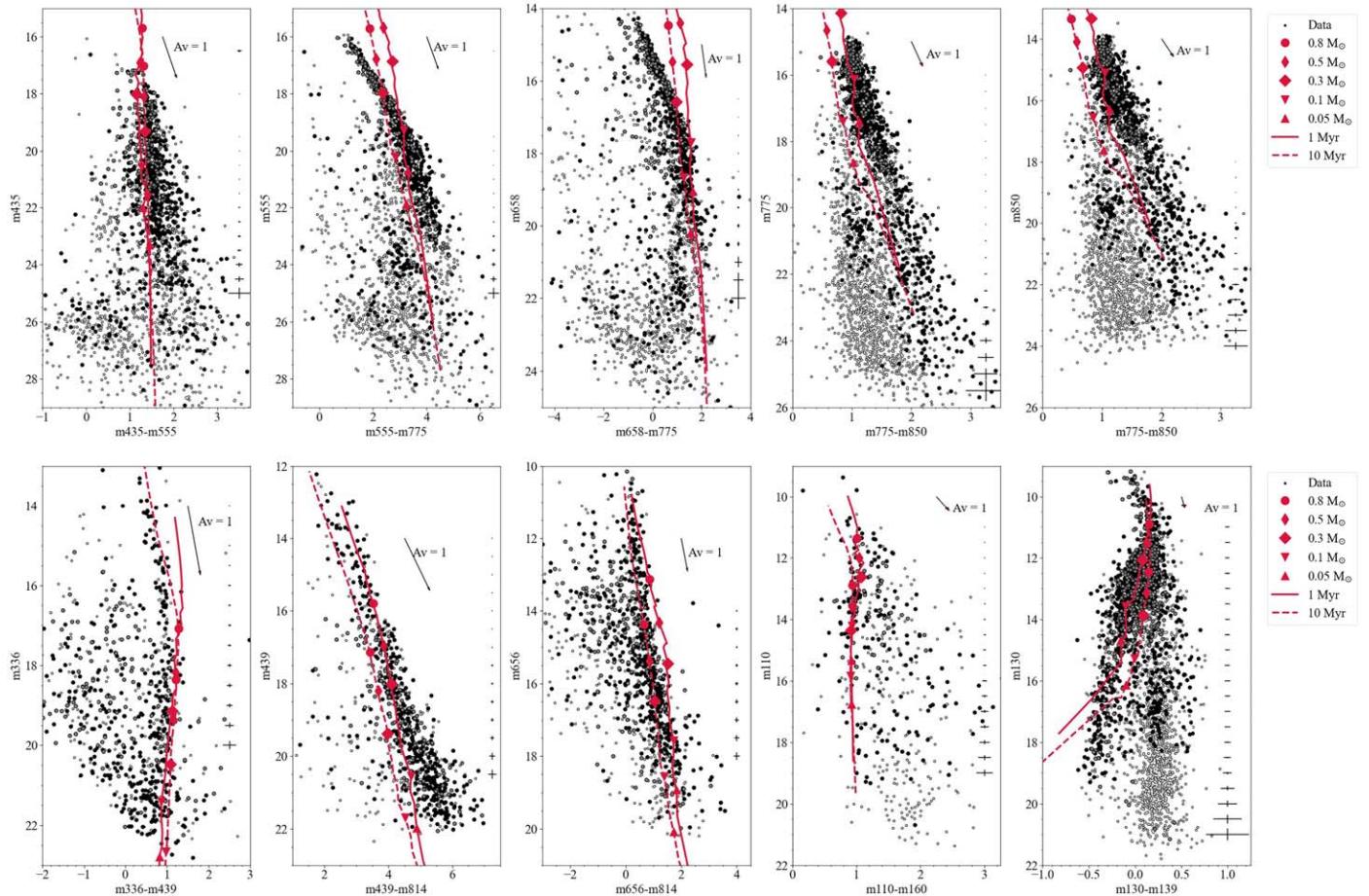


Figure 8. CMDs for ACS, WFPC2, NICMOS, and WFC3-IR filters for the whole sample of unique sources in the catalog. The black circles mark the reference catalog sources, while the gray ones mark the new sources added to create the expanded catalog. The red solid and dashed-dotted lines show the 1 and 10 Myr isochrone, respectively, for each CMD. Typical error bars averaged for bins of magnitude are shown on the right of each plot.

Table 5
Sample of the Physical Parameters of All the Sources Included in This Work

ID	Membership	R.A. (deg)	Decl. (deg)	Mass (M_{\odot})	A_V (mag)	Age (Myr)	Dist (pc)	T (K)	$\log L$ (L_{\odot})	R (Gcm)	$\log g$ (cm s^{-2})
1	<i>n</i>	83.541204	-5.373961	$0.6^{+0.1}_{-0.3}$	$4.2^{+0.7}_{-1.6}$	$7014.6^{+1197.5}_{-2224.4}$	$2995.3^{+1454.2}_{-1473.4}$	$3808.3^{+501.5}_{-446.2}$	$-1.3^{+0.6}_{-0.1}$	$36.8^{+11.7}_{-19.8}$	$4.7^{+0.2}_{-0.2}$
8	<i>y</i>	83.550104	-5.405411	$0.4^{+0.1}_{-0.1}$	$3.0^{+0.3}_{-0.6}$	$5.1^{+1.2}_{-1.3}$	$361.0^{+12.5}_{-12.6}$	$3387.8^{+76.5}_{-87.6}$	$-0.9^{+0.3}_{-1.0}$	$72.2^{+0.5}_{-1.8}$	$4.0^{+0.3}_{-0.7}$

(This table is available in its entirety in machine-readable form.)

Cluster members from the reference catalog are plotted as filled black circles, while the additional sources from the expanded catalog are marked in gray. Nonmembers (either because they can be tagged as field stars or due to unreliable estimates of the source's properties) are open circles. Gray circles mark instead the new sources added to create the expanded catalog as discussed above. In all these diagrams the positions of the cluster members selected using the criteria explained above are generally compatible with the locus of young stellar objects (once extinction and accretion are taken into account). The systematic departures at the bright end are due to the different saturation thresholds in the various filters.

From the masses and ages derived by the fitting routine, we have also retrieved the corresponding stellar luminosity (L_{\star}) and effective temperature (T_{eff}) interpolating over our family of BT-Settl isochrones. The physical parameters estimated

through our Bayesian analysis for all 3399 sources of our main catalog are provided in Table 5. The breakdown of the table is as follows: the first two columns show the ID as provided from the *StraKLIP* pipeline for cross-identification and the membership flag. The following columns show R.A. and decl., followed by the estimated parameters Mass, A_V , Age, distance of the source from the Sun, and derived parameters T_{eff} , $\log L_{\star}$, R , and $\log g$ with relative upper and lower error limits.

4.2. Accretion Luminosity and Mass Accretion Rate

In this subsection we will discuss the process adopted to evaluate L_{acc} and M_{acc} .

As mentioned earlier in Section 3.1, to model the intensity of the accretion spectrum in our Bayesian SED fitting, we introduce an accretion parameter, SP_{acc} , that represents the

Table 6
Sample of the Accretion Properties of All the Sources Included in This Work

ID	$\log L_{\text{acc}}$ (L_{\odot})	$\log \dot{M}_{\text{acc}}$ ($M_{\odot} \text{ yr}^{-1}$)
1	$-5.051^{+4.27}_{-0.054}$	$-12.468^{+4.319}_{-0.059}$
8	$-2.178^{+0.43}_{-0.022}$	$-9.137^{+0.388}_{-0.029}$

(This table is available in its entirety in machine-readable form.)

fraction of stellar bolometric luminosity that can be attributed to accretion. Having derived L_{\star} , one immediately obtains the accretion luminosity (L_{acc}) and therefore the mass accretion rate (\dot{M}_{acc}) by inverting the freefall equation that links the luminosity released when the accretion flow impacts the stellar surface with the rate of mass accretion:

$$L_{\text{acc}} \simeq \frac{GM_{\star}\dot{M}_{\text{acc}}}{R_{\star}} \left(1 - \frac{R_{\star}}{R_{\text{in}}}\right), \quad (2)$$

where R_{\star} and $R_{\text{in}} = 5R_{\star}$ are the star and inner-disk radius, respectively (Gullbring et al. 1998; Hartmann et al. 1998).

In Table 6 we list the accretion properties determined for our catalog. The breakdown of this table is as follows: the first column shows the ID as provided from the *StraKLIP* pipeline for cross-identification. The next columns list the derived accretion parameters $\log L_{\text{acc}}$ and $\log \dot{M}_{\text{acc}}$ and relative errors for each of our sources.

Focusing on the cluster members, in Figure 9 we show the distributions of L_{acc} and \dot{M}_{acc} . In solar units, the median values are $\log L_{\text{acc}} = -1.62 \pm 1.43$ and $\log \dot{M}_{\text{acc}} = -8.29 \pm 1.39$, respectively. On the same plot, using a twin axis as a reference, we also show the corresponding L_{\star} and M_{\star} distribution.

Our estimates of L_{acc} and \dot{M}_{acc} are probably affected by some selection bias inherent to the nature of the data set we are analyzing. While it is generally true that the number of filters plays a major role in the goodness of the fit, this is particularly true for the estimate of the accretion luminosity, in particular, if the missing filters are those most sensitive to the specific characteristics of the accretion spectrum, i.e., F336W (an HST equivalent to the Johnson U band) and the F656N and F658N filters centered on the $H\alpha$ line. Photometry in those bands may be missing either because the source is too faint or even because it is too bright, with strong accretion even leading to saturation. In this last case, the other bandpasses may provide an indication of accretion, but inaccurate or underestimated values. In addition, and especially in the inner regions of the ONC, the presence of circumstellar emission (e.g., proplyds) can influence the photometry in these same filters, leading to an overestimate of the accretion luminosity. Indeed, even though we removed from this analysis the known, spatially resolved proplyds, we cannot completely exclude the presence of circumstellar emission in unresolved stars.

5. Discussion

5.1. Three-dimensional Spatial Distribution

The distances determined by our MCMC Bayesian routine allow reconstructing the three-dimensional spatial distribution of the sources in our reference catalog. The left panel of Figure 10 shows the three main populations that can be discerned: the candidate background sources (red circles—1303 sources) to which the fitting procedure assigns an average

distance, $d \sim 3$ Kpc ($\log d(\text{pc}) \sim 3.6$), the candidate cluster members at about 400 pc ($\log d(\text{pc}) \sim 2.6$; black circles—1044 sources), and the population of candidate foreground sources (blue circles—350 sources) at $d \sim 250$ pc ($\log d(\text{pc}) \sim 2.4$). We include in the cluster sample the 30 sources with bimodal MCMC solutions discussed in Section 4.1, for which we have adopted the distances given by the highest maxima. The right panel of the same figure shows the spatial maps of the three populations, together with the sources with bimodal solutions. For the candidate cluster population the 2D spatial distribution shows a remarkable concentration toward the center (identified with θ^1 Ori C). The foreground population appears to be more uniformly spread, with a residual clustering toward the center of the cluster. One may notice that a higher concentration of foreground sources toward the nebular core is also visible in the spatial map of the Gaia foreground stars. In fact, 77% (189 out of 248) of our foreground sources match with Gaia foreground stars, indicating that the anomaly can be mostly driven by the Gaia prior, while 6% (16) of them are assigned to the background by Gaia, and 17% (43) are assigned to the cluster. For these 43 sources that Gaia identifies as clusters, $\sim 40\%$ of them still have a distance compatible with the cluster within the errors provided by our fit.

The distribution of the background population shows instead a clear drop of source density along a north–south lane, a feature tracing the high extinction caused by the main filament of the OMC-1 molecular cloud (Scandariato et al. 2011).

To assess how the main parameters of the low-mass objects probed by our survey vary with the radial distance from the center, we present in Figure 11 the cumulative distribution functions (CDFs) of the number of candidate cluster members versus distance, grouped according to their masses (left) and ages (right). The left panel of Figure 11 indicates that the more massive sources ($M > 0.7 M_{\odot}$, 185 sources; green line) have a stronger tendency to cluster toward the central region than intermediate-mass sources ($0.2 M_{\odot} < M < 0.7 M_{\odot}$, 383 sources; blue line), and these in turn are more clustered than the low-mass stars ($M < 0.2 M_{\odot}$, 481 sources; gray line), which tend to be the most dispersed. In particular, 50% of the more massive sources are within about 200'' (0.4 pc), versus 290'' (0.56 pc) for the intermediate-mass sources and 390'' (0.75 pc) for the lowest-mass objects. This trend indicates that mass segregation in the ONC extends to subsolar masses, possibly down to $0.2 M_{\odot}$, with a degree of concentration decreasing with decreasing mass, extending into the subsolar and possibly brown dwarf (BD) mass range, similar to what was found for masses $> 1-2 M_{\odot}$ (not probed by our survey) by Hillenbrand & Hartmann (1998).

Similarly, the right panel of Figure 11 presents the same CDF for three bins of ages. While old (> 5 Myr; green—25 sources) and intermediate-age sources (between 2 and 5 Myr; blue—89 sources) appear similarly dispersed through the cluster, the large majority of young sources (age < 2 Myr; gray—935 sources) show a higher degree of concentration toward the central region. This result sets a rather strong upper limit to the age of the cluster.

5.2. Hertzsprung–Russell Diagram

The presence of an isochronal age spread in the Hertzsprung–Russell (H-R) diagram of the ONC has been previously reported and analyzed by several authors, with evidence of contamination from older populations (e.g., Prosser et al. 1994;

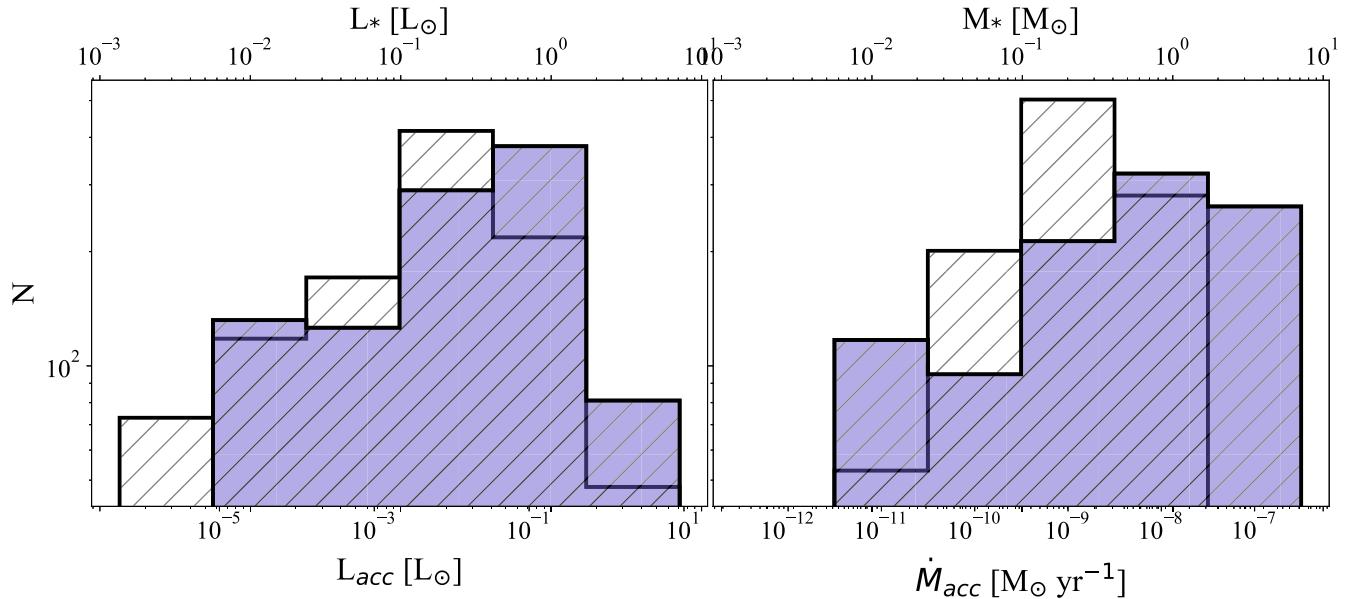


Figure 9. Histograms of the accretion luminosity L_{acc} (right) and mass accretion rate \dot{M}_{acc} (left) derived by the stellar parameters obtained from our Bayesian analysis (blue) with L_* and M_* overplotted on a twin axis, respectively (white).

Hillenbrand 1997; Slesnick et al. 2004; Alves & Bouy 2012; Beccari et al. 2017). In particular, Bouy et al. (2014) determined the presence of a large foreground population toward the Orion A molecular cloud, containing several distinct subgroups with an age spread in the range 5–10 Myr. Still, the fraction of foreground stars contaminating our survey area centered on the Trapezium should be small, with only $\sim 15\%$ of the WFC3-IR sources lying in the region of significant overlap between the cluster and noncluster members, where the Bayesian’s membership analysis of Robberto et al. (2020) has larger uncertainty. The same should be true for the ACS survey, given the very similar footprint. Of the 1049 sources that are part of the reference catalog, only $\sim 3\%$ of them have an age estimated between 5 and 10 Myr and $\sim 80\%$ have a distance $\lesssim 400$ pc, so it is quite possible that a number of them belong to a foreground population even if photometry shows that they are compatible with real cluster members.

In any case, regardless of the level of contamination by other populations, uncertainties in the measurements, presence and orientation of disk, spots, variability, anomalous extinction, etc., can all contribute to a spread in luminosity and in turn to the observed age spread (Reggiani et al. 2011). Some of these effects may be associated with photometric variability. Starspots, in particular, modulate the luminosity of a star (Gully-Santiago et al. 2017; Gangi et al. 2022), and low-mass, young stars show variability that can be attributed to large spots created by strong magnetic fields (Johns-Krull & Valenti 1996). Besides variability, models show that stars in the mass range 0.1 – $1.12 M_\odot$ with $\gtrsim 50\%$ of their surface covered by spots have radii inflated by a factor of 10% during the pre-main-sequence stage (Somers & Pinsonneault 2015), and between heavily spotted and spotted-free models the luminosity can change by a factor of 2.

With all these caveats, we observe a peak in the age distribution with a median value at $\sim 1.1 \pm 0.1$ Myr. About 88% of the sources are younger than 2 Myr, while only $\sim 3\%$ of the total sources have an age estimate above 5 Myr. Our results are consistent with a single major event of star formation in the ONC.

In Figure 12 we present the H-R diagram for the reference catalog of 1049 cluster sources in the form of a Hess diagram. We bin them in boxes of ~ 57 K by $\sim 0.076 \log_{10} L_*$ and color-code their number according to the scheme presented in the figure. The 1, 3, and 10 Myr isochrones are also shown for comparison (see Section 3.1.1).

In comparison to the previous H-R diagrams of the ONC (e.g., Da Rio et al. 2012), our new H-R diagram appears well constrained by the 1–3 Myr isochrones, with very few sources up to 10 Myr. There is still a scatter in luminosity for effective temperatures in the range of 4000–3000 K (i.e., the region dominated by our Class A sources). When T_{eff} decreases below ~ 3000 K, our solutions rely on fewer filters, and thus the priors have an increasingly stronger weight. Besides, the range of luminosities spanned by the different isochrones becomes narrower for $T_{\text{eff}} \lesssim 3000$ K. Both factors, the first observational and the second theoretical, contribute to the strong concentration of the sources around the 1–3 Myr isochrones.

5.3. Empirical Initial Mass Function

Deriving an accurate IMF for a rich, young stellar cluster like the ONC allows us in principle to discern its variations versus, e.g., radial distance from the center, isochronal age, and binarity. To achieve this goal, one must obtain a complete, extinction-limited list of reliable mass values. In the case of our data set, the stellar masses have been estimated combining different filter sets, and within each of them one will always find two bandpasses constraining the lowest and highest mass probed, for any given exposure time, extinction, and assumed isochrone. Due to the spread of extinction values toward the region (see Figure 7) and the nonuniformity of exposure times even within a single filter due, e.g., to the overlaps between adjacent frames, it is nearly impossible to derive homogeneous values of the limiting mass and therefore an extinction-limited sample from our master catalog. In practice, our data analysis strategy aimed at combining a diversity of information, both priors and data sets, in order to determine the most accurate parameters of each source, is less than optimal if one wants to

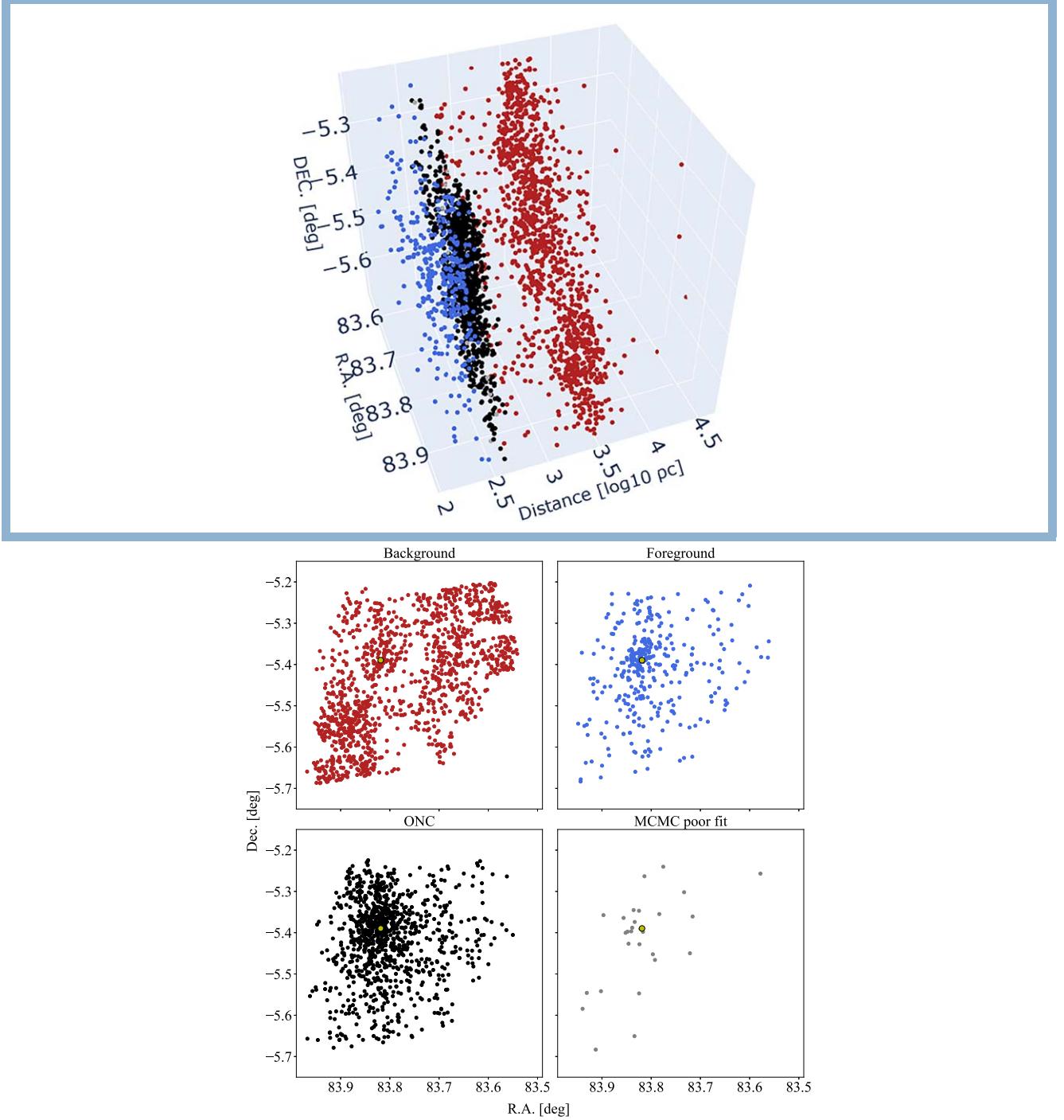


Figure 10. Spatial distribution of sources in the catalog divided by membership. Left: 3D spatial representation of foreground (blue), candidate ONC members (black), and candidate ONC members with nonconverging MCMC fit (gray) and background (red). The 3D spatial representation of sources in the catalog is divided by the membership: foreground (blue), candidate ONC members (black), and candidate ONC members with nonconverging MCMC fit (gray) and background (red). In the upper left portion of the interactive figure, functionalities to zoom, pan, or rotate the image have been added to interact with this figure. Each sample can be also hidden or shown by clicking through the labels in the legend. Right: 2D view for the same groups. A yellow marker shows the position of θ^1 Ori C in each plot.

derive global parameters, such as the IMF, that require a homogeneous sample free from any selection bias. On the other hand, it is worth assessing our mass distribution since our method allows retrieving sources that may be systematically rejected in more selective, controlled catalogs. To illustrate the range of masses probed by the different filter combinations, we show in the top part of Table 7 how the minimum mass reached in each filter depends on the extinction, and similarly the lower

portion shows the corresponding mass at the saturation limit. Building this table, we have adopted our reference isochrone and the typical exposure times of the surveys, adopted a $\sigma \sim 0.1$ uncertainty for the faint magnitude limits, and neglected the narrowband H α passbands of WFPC3 and ACS, as they may be strongly affected by accretion or circumstellar emission. The table shows how high extinction values push low-mass sources increasingly out of reach, while

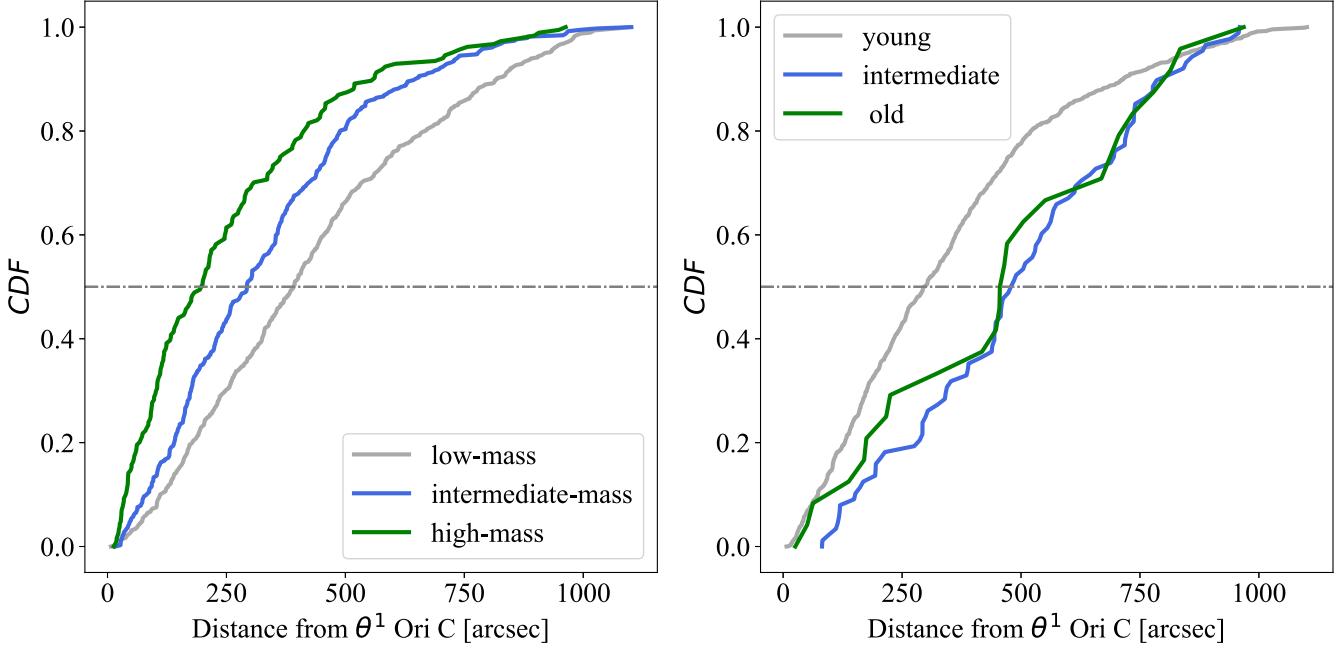


Figure 11. Cumulative distribution functions for sources in our reference catalog of cluster members as a function of distance from θ^1 Ori C (first three plots). Left: high-mass (mass $> 0.7 M_{\odot}$ —green), intermediate-mass (mass between 0.2 and $0.7 M_{\odot}$ —blue) and low-mass (mass $< 0.2 M_{\odot}$ —gray) sources. Right: old (age > 5 Myr—green), intermediate (age between 2 and 5 Myr—blue), and young (age < 2 Myr—gray) sources.

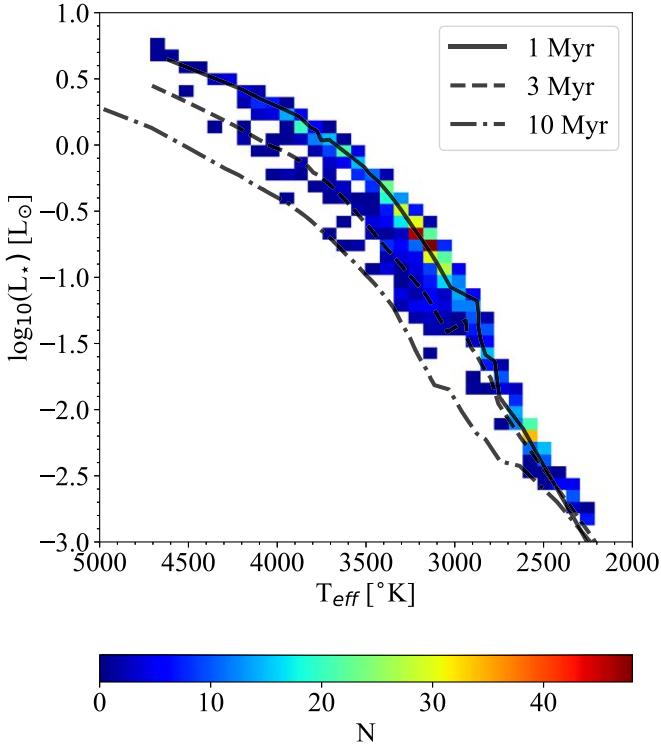


Figure 12. Hess diagram for the H-R diagram for ONC candidate cluster sources. Luminosity and temperature are derived from the best fit of the models. Three isochrones (1, 3, 10 Myr) are shown for comparison. The colors represent the number of stars present in each box.

high-mass stars become increasingly measurable as they fall below the saturation limits.

Therefore, a filter set composed of our ACS bandpasses will be generally limited at the low mass limit by the bluest filter, F435W, as it is unable to reliably detect masses $M < 0.025 M_{\odot}$.

even when $A_V = 0$. Vice versa, the upper mass limit will be determined by the wide-band F805LP filter, cutting off masses $M > 1.345 M_{\odot}$ even with $A_V = 9$.

Considering all these caveats on the mass, we down-selected the *expanded catalog* of cluster members to 1428 sources, from the initial 1546 sources (see Section 4.1), and derived an ONC mass distribution as follows. Taking full advantage of our MCMC calculations, instead of adopting just the final solution for the mass of each source, we consider the last 100 values determined by each walker as it approaches convergence. This provides us with a rather rich and robust statistical distribution of all parameters, mass in particular, compatible with the data and the priors. Then, extracting randomly for each source one value from its mass distribution, we produce a particular realization of the mass distribution in the ONC. We repeat this process 100 times, each iteration providing us with a slightly different IMF that, once averaged over the same constant bins, produces the mass distribution represented by the black markers in Figure 13. The error bars indicate the 1σ spread of the data in each bin, the only anomaly being the dip observed in the bin at about $0.07 M_{\odot}$. This is most probably a selection effect. Stars falling in this bin are typically at the sensitivity limit of the bluer ACS and WFP2 filters, yet massive enough to be poorly classified using their shallow WFC3 1.4 μ m absorption band (see Section 2 for more details). The two effects combined may cause MCMC to provide degenerate or erroneous solutions. The aforementioned clustering of source foreground objects toward the center of the cluster confirms that we do not account for all cluster members in the central region. On the other hand, our estimates—both mass and distance, i.e., membership—are more robust for higher masses, where we generally have a more complete SED, and for lower masses, where the WFC3 1.4 μ m absorption band becomes prominent. The deficit of stars can be reconciled by including the stars that Da Rio et al. (2012) classified as

Table 7
Minimum (top) and Maximum (bottom) Detectable Mass, in Solar Units, as a Function of Filters and Extinction

A_v	m435	m555	m775	m850	m130	m139	m110	m160	m336	m439	m814
0	0.025	0.016	0.004	0.002	0.003	0.003	0.009	0.003	0.152	0.099	0.027
1	0.034	0.022	0.005	0.002	0.003	0.003	0.011	0.003	0.468	0.144	0.034
2	0.046	0.030	0.006	0.003	0.003	0.003	0.013	0.003	0.942	0.328	0.040
3	0.081	0.037	0.008	0.003	0.004	0.004	0.016	0.004	1.367	0.612	0.049
4	0.118	0.048	0.010	0.004	0.005	0.004	0.020	0.004	1.367	1.050	0.068
5	0.237	0.076	0.013	0.005	0.005	0.005	0.024	0.005	1.367	1.367	0.095
6	0.494	0.108	0.017	0.007	0.006	0.006	0.029	0.005	1.367	1.367	0.114
7	0.826	0.158	0.022	0.008	0.007	0.006	0.034	0.006	1.367	1.367	0.153
8	1.262	0.319	0.028	0.010	0.008	0.007	0.041	0.006	1.367	1.367	0.250
9	1.367	0.510	0.034	0.013	0.009	0.008	0.046	0.007	1.367	1.367	0.405
0	0.296	0.104	0.102	0.102	0.501	0.456	0.502	0.201	1.196	1.087	0.532
1	0.548	0.143	0.129	0.124	0.615	0.557	0.656	0.245	1.367	1.367	0.748
2	0.929	0.278	0.187	0.160	0.765	0.657	0.873	0.294	1.367	1.367	1.057
3	1.337	0.474	0.312	0.241	0.930	0.792	1.112	0.357	1.367	1.367	1.337
4	1.367	0.676	0.478	0.359	1.115	0.932	1.318	0.421	1.367	1.367	1.354
5	1.367	1.014	0.657	0.509	1.283	1.094	1.345	0.518	1.367	1.367	1.354
6	1.367	1.348	0.946	0.684	1.343	1.248	1.345	0.605	1.367	1.367	1.354
7	1.367	1.365	1.261	0.943	1.343	1.345	1.345	0.721	1.367	1.367	1.354
8	1.367	1.365	1.355	1.222	1.343	1.345	1.345	0.863	1.367	1.367	1.354
9	1.367	1.365	1.355	1.349	1.343	1.345	1.345	0.973	1.367	1.367	1.354

Note. Filters are separated into four blocks for the ACS, WFC3/IR, NICMOS-3, and WFPC2 cameras, in that order.

belonging to this mass bin. This operation adds back 39 sources in the mass bin between 0.049 and $0.095 M_{\odot}$, obtaining the point represented by the red triangle in Figure 13 and a final catalog of 1467 total sources.

It is well established that for the low-mass stars and BDs (which is the region of the mass spectrum most relevant for our study), the IMF can be generally approximated with a shallower power law ($\xi(\log m) \propto m^{-\alpha}$; Kroupa 2001) or a lognormal function ($\xi(\log m) \propto e^{-(\log m - \log m_c)^2/2\sigma^2}$; Chabrier 2003). Adopting a Bayesian approach and an MCMC algorithm, we use both forms to fit our measured mass distribution in the mass range 0.025 – $1.34 M_{\odot}$. The resulting distributions are shown in Figure 14, with the best-fit parameters presented in Tables 8 and 9. For reference, also reported are the values from the canonical IMFs (Kroupa—Kroupa 2001; Chabrier singles and systems—Chabrier 2003), as well as the ONC IMFs obtained by Da Rio et al. (2012) and Gennaro & Robberto (2020).

When we compare the results for the broken power-law case from Table 8, our low-mass slope, $\alpha_0 = -0.17$, is compatible within 1σ with the Da Rio et al. (2012) measurement. The discrepancy versus Gennaro & Robberto (2020) is due to our data not accounting for unresolved multiple systems with separation closer than $0.^{\circ}1$ (or ~ 40 au at the distance of the ONC; G. M. Strampelli et al. 2023, in preparation), while Gennaro & Robberto (2020) leave the binary fraction free to vary in their Bayesian model. The transition mass between the low and mid ranges, $m_{lm} = 0.20 M_{\odot}$, falls between the value predicted by Kroupa (2001) and the one observed by Da Rio et al. (2012) and is also in very good agreement with the prediction from Gennaro & Robberto (2020). For the intermediate-mass slope, $\alpha_1 = -1.34$, we are in very good agreement with both predictions from Kroupa (2001) and Gennaro & Robberto (2020). On the other hand, we observe a shallower α_1 slope compared to the one observed by Da Rio et al. (2012), but their large uncertainty makes the results still compatible.

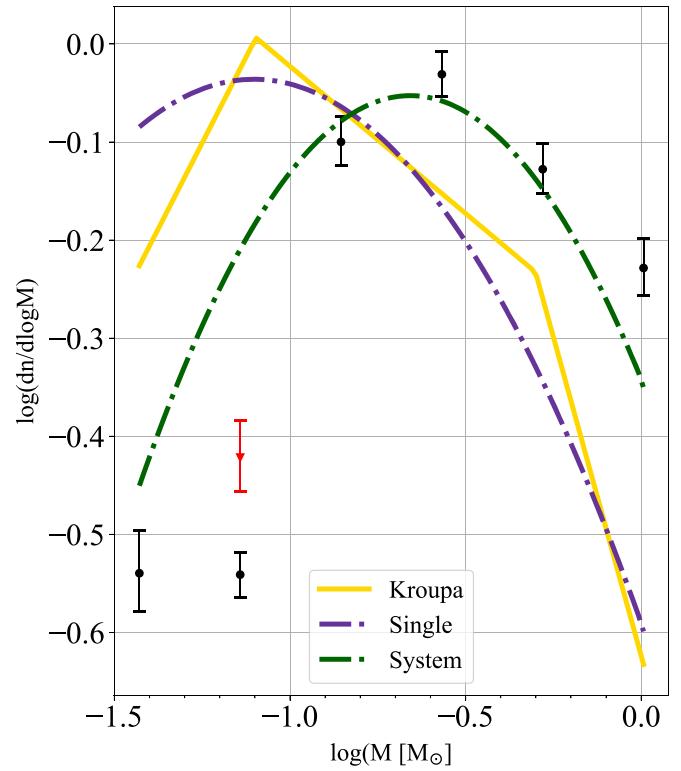


Figure 13. Comparison between the derived ONC mass distribution (black) and the canonical IMFs: Kroupa (yellow), and Chabrier singles (blue) and systems (green). Each black circle shows the average number of stars in each bin obtained from all the mass distribution histograms, with the error bars including the 1σ spread in each bin and the relative Poisson noise. The red marker shows the correction obtained adding Da Rio et al. (2012) sources in the mass bin 0.049 – $0.095 M_{\odot}$.

For the lognormal case (Table 9), we determine a peak mass compatible with the Gennaro & Robberto (2020) IMF, but with smaller σ (our distribution is narrower). Similar to the power-

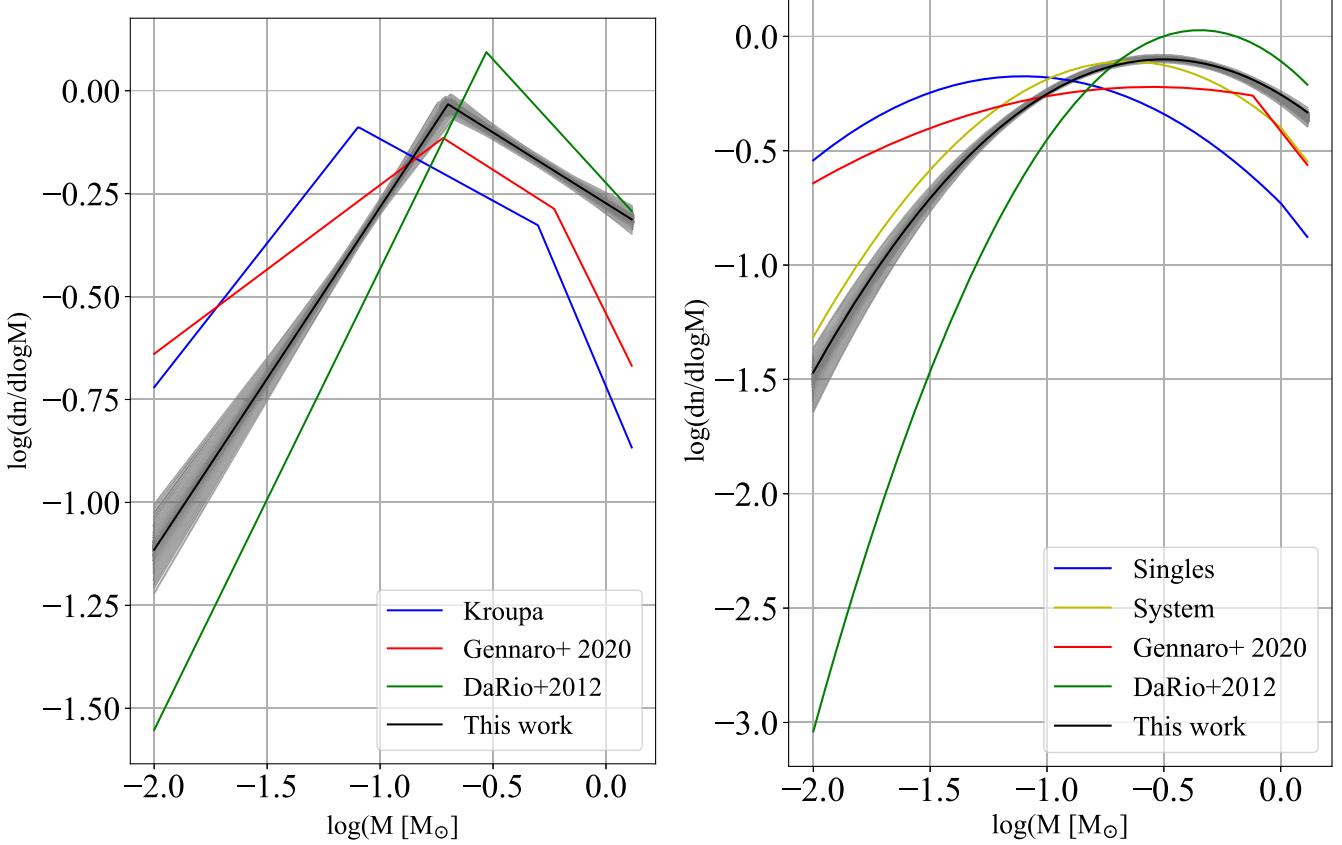


Figure 14. Comparison between the empirical ONC IMF (black) and 500 draws from the posteriors (gray) and the canonical IMFs: Kroupa (left), and Chabrier singles and systems (right). Also shown as reference in both plots are the IMFs obtained by Da Rio et al. (2012) and Gennaro & Roberto (2020).

law case, our derived mass function is indeed expected to be more similar to a “system” mass function, as we do not try to resolve close-in companions in our data. Compared to Da Rio et al. (2012), we observe a smaller peak mass and a larger distribution (higher σ value).

Overall, for masses $\gtrsim 0.1 M_\odot$ our empirical IMF is in good agreement with a Chabrier (2003) system IMF.

5.4. Accretion Rates in the ONC

In the left panel of Figure 15 we present the relation between L_{acc} and L_* for our selected cluster sources. Similar to Section 5.2, we apply a stronger selection to the reference catalog, including only the sources with $\text{AR} > 50\%$. The upper solid red line limits the area above which the majority of sources coincide with proplyds, according to the catalog by Ricci et al. (2008), making the estimate of L_{acc} unreliable. The bottom dashed-dotted red curve instead marks the stellar photospheric limit (see Equation (1) in Manara et al. 2017a) for 1 Myr sources, below which the chromosphere is as bright as or brighter than the accretion excess and therefore prevents disentangling the L_{acc} contribution.

Selecting only the reference catalog between these two lines and applying the stronger cluster selection discussed above, we can build a plot showing the $\dot{M}_{\text{acc}}-M_*$ relation for 631 sources (Figure 15, right panel). Similar to the cases of Chameleon I and Lupus (Manara et al. 2017b; Alcalá et al. 2017), \dot{M}_{acc} appear to increase with M_* . The scatter is large, but the relation appears steeper for low-mass objects, suggesting that one may consider either a single or a broken power-law relation. We tested which approach provides a better description of our data

(see the Appendix), finding that a segmented linear relation (corresponding to a broken power law in the linear scale) is preferable over a linear fit. With the already-mentioned caveats about saturation and biases, we can provide for $0.01 M_\odot \lesssim M_* \lesssim 0.7 M_\odot$ the following relations:

$$\begin{aligned} \log L_{\text{acc}} &= -(0.40 \pm 0.09) + (1.46 \pm 0.06) \log L_* & \text{if } \log L_* \leq -0.63 \\ \log L_{\text{acc}} &= -(0.99 \pm 0.09) + (0.42 \pm 0.40) \log L_* & \text{if } \log L_* > -0.63, \end{aligned} \quad (3)$$

while for \dot{M}_{acc} we obtain

$$\begin{aligned} \log \dot{M}_{\text{acc}} &= -(6.78 \pm 0.12) + (1.98 \pm 0.10) \log M_* & \text{if } \log M_* \leq -0.68 \\ \log \dot{M}_{\text{acc}} &= -(7.92 \pm 0.24) + (0.13 \pm 0.41) \log M_* & \text{if } \log M_* > -0.68. \end{aligned} \quad (4)$$

These relations are shown in Figure 15 as a blue dashed line. The filled area around each fit provides a visualization of the uncertainties associated with each fit.

A broken power-law relation for both L_{acc} and \dot{M}_{acc} distributions is a new result compared to previous studies of accretion in young SFRs (e.g., Manara et al. 2017a), and in particular for this cluster (Manara et al. 2012), where a linear relation was generally preferred, as there was not enough evidence to distinguish between them. For both parameters, the distributions appear strongly dependent on M_* (or L_*) below $M_* \lesssim 0.21 M_\odot$ (or $L_* \lesssim 0.23 L_\odot$). Afterward, the trend becomes flatter, indicating that the \dot{M}_{acc} (L_{acc}) remains more or

Table 8
Summary of the Results for the Broken Power-law Model

Authors	α_0	α_1	α_2	m_{lm}	m_{mh}
This work	$-0.17^{+0.04}_{-0.04}$	$-1.34^{+0.03}_{-0.02}$...	$0.20^{+0.01}_{-0.01}$...
Kroupa (2001)	$-0.3^{+0.7}_{-0.7}$	$-1.3^{+0.5}_{-0.5}$	$-2.3^{+0.3}_{-0.3}$	0.08	0.5
Da Rio et al. (2012) (BCAH98)	$-0.12^{+0.90}_{-0.90}$	$-1.6^{+0.33}_{-0.33}$...	0.29	...
Gennaro & Robberto (2020)	$-0.59^{+0.06}_{-0.06}$	$-1.35^{+0.33}_{-0.35}$	$-2.11^{+0.27}_{-0.37}$	$0.19^{+0.06}_{-0.05}$	$0.59^{+0.24}_{-0.17}$

Note. α_0 , α_1 , and α_2 represent the low-, intermediate-, and high-mass slopes, while m_{lm} and m_{mh} represent the low-to-mid and mid-to-high transition masses.

Table 9
Summary of the Results for the Lognormal Model

Authors	m_c	σ	m_{hm}	α_{hm}
This work	$0.30^{+0.01}_{-0.01}$	$0.57^{+0.01}_{-0.01}$
Chabrier (2003) (Single)	$0.079^{+0.0016}_{-0.021}$	$0.69^{+0.01}_{-0.05}$	1	-2.3
Chabrier (2003) (System)	0.22	0.57	1	-2.3
Da Rio et al. (2012) (BCAH98)	$0.45^{+0.02}_{-0.02}$	$0.44^{+0.05}_{-0.05}$
Gennaro & Robberto (2020)	$0.28^{+0.12}_{-0.08}$	$1.04^{+0.15}_{-0.13}$	$0.76^{+0.24}_{-0.20}$	$-2.30^{+0.31}_{-0.32}$

Note. m_c represents the characteristic mass, while σ is the width of the lognormal, m_{hm} is the transition mass, and α_{hm} is the slope of the power law for the high-mass regime.

less constant as M_\star (L_\star) increases. However, there is a large spread of values (about 2 orders of magnitudes) for both distributions.

Our results on the $L_{\text{acc}} - L_\star$ can be compared with the relations obtained in other SFRs like ρ Ophiuchi (Natta et al. 2006), σ Orionis (Rigliaco et al. 2011), Chameleon I (Manara et al. 2016a, 2017b), and Lupus (Alcalá et al. 2014, 2017). If we only consider the region where $L_\star \lesssim 0.23 L_\odot$ (before the braking point), we find a good agreement between our portion of the fit and the one obtained for Chameleon I, while we find a looser match (within 3σ) with the other SFRs. Similarly, the slope of the $\dot{M}_{\text{acc}} - M_\star$ relationship is in broad agreement with other SFRs (e.g., Chameleon I, Lupus, and NGC 2264; Venuti et al. 2014; Manara et al. 2017b; Alcalá et al. 2017), where the data seem to better support a double power-law fit, with lower-mass stars having a more rapid decrease of \dot{M}_{acc} compared to higher-mass stars (even though the authors could not completely rule out the linear dependence). In general, when comparing results between different SFRs, age differences should be taken into account. Both the small variation observed with the ages and the high uncertainties related to age estimation itself (e.g., Soderblom et al. 2014) will make any detailed analysis very difficult to perform.

The slope of the \dot{M}_{acc} relation can also be compared with theoretical models and be used as a proxy to evaluate the main processes behind the dispersal of protoplanetary disks (Clarke & Pringle 2006). Typical values suggest power laws with exponent $\sim 1.6-2$ and spread of $\sim 1-2$ dex (Alcalá et al. 2014; Manara et al. 2016b; Hartmann et al. 2016; Manara et al. 2017a; Alcalá et al. 2017). Recent works, however, show that accretion variability alone is not enough to explain the spread observed in \dot{M}_{acc} measurements (Fischer et al. 2022;

Flischlen et al. 2022) that must therefore be also related to physical processes such as disk evolution. It is still not clear in what measure the $\dot{M}_{\text{acc}} \propto M_\star^2$ relation is due to the disk evolutionary process or if it reflects how the initial conditions scale with stellar mass (Manara et al. 2023, and references therein). For example, Dullemond et al. (2006) were able to produce a steep $\dot{M}_{\text{acc}} - M_\star$ relation resulting from the imprint of the initial conditions for the formation of the star-disk system. A simple model of disk formation and evolution from collapsing cores provides a $\dot{M}_{\text{acc}} \propto M_\star^{1.8}$ relation, provided that cores of all mass have a similar distribution of rotation periods.

The ONC environment is characterized by strong UV flux that affects the structure and evolution of the disks closer to the central OB stars (Winter & Haworth 2022). However, from our findings, at least for low-mass stars, the $\dot{M}_{\text{acc}} - M_\star$ relation is remarkably similar to the one observed in low-mass SFRs, suggesting that external photoevaporation does not play a significant role in the majority of young circumstellar disks. The slope of the relation, therefore, may be largely determined by the initial conditions set at the onset of the star formation process, which may be quite similar between regions that eventually form clusters of different sizes (Manara et al. 2023).

Photoevaporation by the disk central star can also play a role in the $\dot{M}_{\text{acc}} - M_\star$ slope. X-ray-driven photoevaporation produces a slope between 1.6 and 1.9 (Ercolano et al. 2014), while UV-driven photoevaporation provides a value of ~ 1.35 (Clarke & Pringle 2006). In this context, the result we obtain for $M_\star \lesssim 0.21 M_\odot$ is probably more in line with the X-ray-driven scenario rather than the UV one, even though both the X-ray-driven and collapsing core scenarios provide shallower slopes compared to ours.

A consequence of a bimodal distribution on the $\dot{M}_{\text{acc}} - M_\star$ plane is a different evolutionary timescale for disk accretion around stars with different masses, i.e., disks around stars with $M_\star \lesssim 0.2 M_\odot$ will evolve faster toward lower values of \dot{M}_{acc} . The surveys in σ Orionis (Rigliaco et al. 2011), in the same ONC (Manara et al. 2012), and the spectroscopic survey of L1641 (Fang et al. 2013) arrive at similar conclusions, even though we need to point out that a faster evolution for lower-mass stars is opposite to predictions by Alexander & Armitage (2006), who postulated that the viscous timescale increases with stellar mass to explain the $\dot{M}_{\text{acc}} \propto M_\star^{-2}$ relation.

A reason behind the bimodal distribution can be found in two different physical regimes at play in the early stages of disk formation and evolution (Vorobyov & Basu 2009). These authors suggested that disk self-gravity will drive large accretion torques for stars more massive than $\sim 0.2 M_\odot$ soon after formation. This in turn will leave less material available in the Class II phase and therefore a lower accretion rate, which will flatten the relationship. For lower-mass stars instead, self-gravity is less important and disks evolve more viscously. In

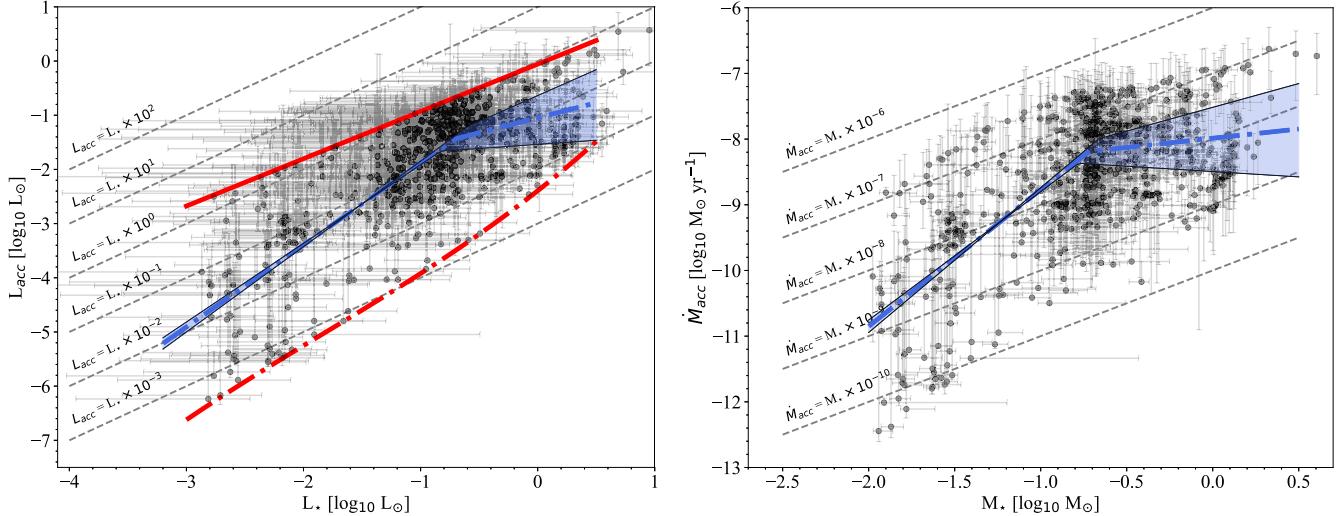


Figure 15. Accretion luminosity (L_{acc}) as a function of stellar luminosity (L_*) for ONC sources with relative error bars. Also shown are the limits derived owing to chromospheric emission for a 1 Myr old object (lower dashed-dotted red curve) and the limit above which we expect the majority of sources to be proplyd (solid red line), as explained in the text. Right: similar to the left panel, but for the mass accretion rate (\dot{M}_{acc}) as a function of stellar mass (M_*) for ONC sources. The thick blue dashed line represents the best fit obtained from the data using a piecewise linear function (Equation (A7)). The colored area provides a relative visualization of the errors on the fit.

particular, the authors provide the following value for the exponent as best fit to their model:

$$\begin{aligned} n &= 2.9 \pm 0.5 \text{ if } M_* < 0.2M_\odot \\ n &= 1.5 \pm 0.1 \text{ if } 0.2M_\odot \leq M_* < 3M_\odot, \end{aligned} \quad (5)$$

while the corresponding fits to the observation from Muzerolle et al. (2005, and references therein) for T Tauri stars and BDs of age 0.5–3.0 Myr produce

$$\begin{aligned} n &= 2.3 \pm 0.6 \text{ if } M_* < 0.2M_\odot \\ n &= 1.3 \pm 0.3 \text{ if } 0.2M_\odot \leq M_* < 3M_\odot. \end{aligned} \quad (6)$$

Remarkably, the Vorobyov & Basu (2009) model predicts a similar relation for $M_* \lesssim 0.21M_\odot$ compared to ours, with our estimate positioning itself between the model and the observational data relations. For $M_* > 0.21M_\odot$ instead, our fit is much shallower, even though we have to argue here that our selected sample is limited to M_* higher masses owing to saturation of one or more filters (only $\sim 15\%$ of sources in our selected catalog have higher masses). On the other hand, their data reach up to $3M_\odot$, so we are not covering quite the same mass range.

6. Conclusion

We have revisited the Roberto et al. (2013) HST/ACS/WFC ONC catalog in filters F435W, F555W, F658N, F775W, and F850LP to provide updated estimates of the average magnitudes in each filter. We then combined the photometry from ACS and WFPC2 with the more recent HST/WFC3-IR measures from Roberto et al. (2020) in filters F130N and F139M. The resulting data set contains the photometric data derived from the two HST Treasury programs dedicated to the ONC, spanning almost 15 yr. We adopted a Bayesian SED fitting to derive reliable estimates of the principal stellar parameters (i.e., temperature, extinction, age, accretion, and

distance) for the majority of the sources in the catalog and estimate general properties for the cluster. In particular, the three-dimensional study of mass distribution for bona fide cluster members shows that mass segregation in the ONC extends to subsolar masses, possibly down to $\sim 0.2M_\odot$. From the age distribution we unveil a star formation history heavily peaked at $\sim 1.1 \pm 0.1$ Myr, with a small tail extending up to 10 Myr. Our estimates, model dependent, do not support the idea of multiple episodes of significant star formation in the ONC. Training our data set on the group of cluster members with known parallax, we evaluate an average distance for the cluster of 397 ± 17 pc, and similarly for the extinction distribution we estimate a median extinction of 4.35 ± 4.10 mag.

The inclusion of accretion in our fitting procedure allows us to test the relationship between L_{acc} and \dot{M}_{acc} versus the stellar parameters. If we consider only the region where $L_* \leq 0.16L_\odot$, we also find good agreement between the $L_{\text{acc}}-L_*$ relation for this cluster and a less massive system like Chameleon I, while we find a looser agreement (within 3σ) with σ Orionis, ρ Ophiuchi, and Lupus SFRs. Overall, we find evidence for a bimodal relation for L_{acc} and \dot{M}_{acc} for this cluster. Both can be parameterized as broken power laws with slopes $\log L_{\text{acc}} 1.46 \pm 0.06$ for $\log L_* \leq -0.63$ and 0.42 ± 0.40 otherwise, and $\log \dot{M}_{\text{acc}} 1.98 \pm 0.10$ for $\log M_* \leq -0.68$ and 0.13 ± 0.41 otherwise.

The results we obtain for $M_* \lesssim 0.2M_\odot$ support a scenario where the excess emission is dominated by X-ray-emission from the central star rather than the UV one, while external photoevaporation does not look like it plays a significant role in the majority of young circumstellar disks. The slope of the relation, therefore, may be largely determined by the properties of the central stars, and therefore by the initial conditions set at the onset of the star formation process, rather than by the dynamical evolution of the cluster. These conditions may be quite similar between regions that eventually form clusters of different sizes (Manara et al. 2023).

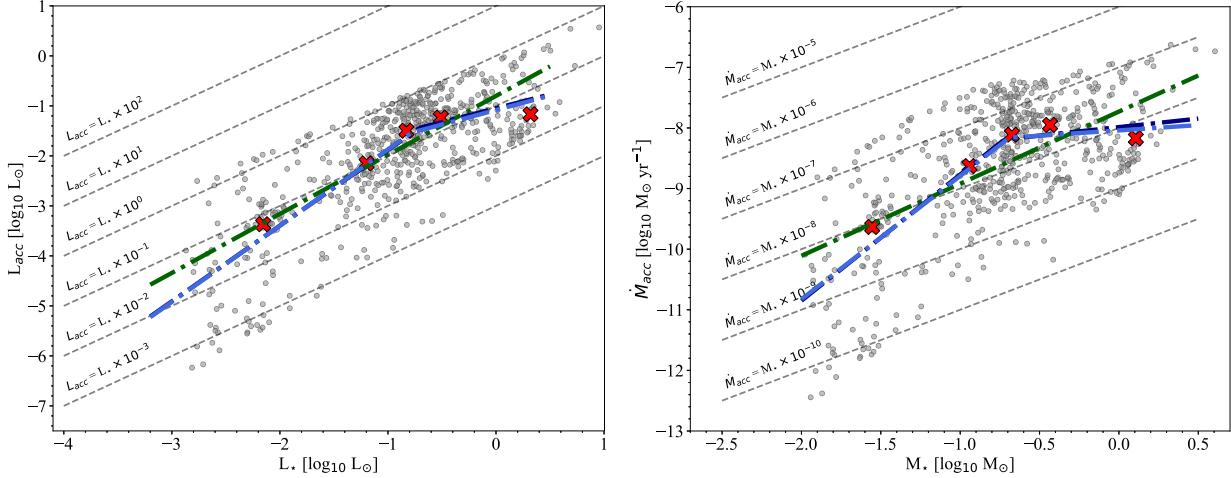


Figure A1. Left: accretion luminosity (L_{acc}) as a function of stellar luminosity (L_*) for ONC sources. Also shown is the fit for the linear relation (green) and bimodal relation (dark blue: selected; light blue: not selected). Right: similar to the left panel, but for the mass accretion rate (\dot{M}_{acc}) as a function of stellar mass (M_*) for ONC sources.

The shape of the IMF turns out to depend on the particular data set, as the type and number of filters used to derive the source parameters are not homogeneous. This introduces selection biases that need to be carefully assessed when deriving global system parameters such as the IMF. Overall, for masses $\gtrsim 0.1 M_\odot$, we find our results to be compatible with a canonical Chabrier system IMF.

This work underlines the potential and limitations of multicolor photometric surveys to characterize the main physical properties of SFRs.

Acknowledgments

The authors thank the anonymous referees for their useful comments on the manuscript. G.M.S. wants to thank the University of California Santa Barbara and the Instituto de Astrofísica de Canarias for its hospitality. The authors thank Gaspard Duchene for the invaluable suggestions and comments. Support for program No. GO-10246 was provided by NASA through a grant from the Space Telescope Science Institute, which is operated by the Association of Universities for Research in Astronomy, Incorporated, under NASA contract NASS-26555. J.A. was supported in part by a grant from the National Physical Science Consortium. G.M.S. and A.A. acknowledge support from the Spanish Agencia Estatal de Investigación del Ministerio de Ciencia e Innovación (AEI-MICINN) under grant ‘‘Proyectos de I+D+i’’ with references AYA2017-89841-P and PID2020-115981GB-I00. C.F.M. is funded by the European Union under the European Union’s Horizon Europe Research & Innovation Programme 101039452 (WANDA). Views and opinions expressed are, however, those of the author(s) only and do not necessarily reflect those of the European Union or the European Research Council. Neither the European Union nor the granting authority can be held responsible for them.

The data presented in this paper were obtained from the Mikulski Archive for Space Telescopes (MAST) at the Space Telescope Science Institute. The specific observations analyzed can be accessed via doi:[10.17909/rr5s-js47](https://doi.org/10.17909/rr5s-js47).

Facility: HST (ACS, WFC3).

Software: Numpy (van der Walt et al. 2011), Astropy (Astropy Collaboration et al. 2013; Price-Whelan et al. 2018),

Scipy (Virtanen et al. 2020), Matplotlib (Hunter 2007), PyKLIP (Wang et al. 2015), Pandas (McKinney et al. 2010).

Appendix

Power-law versus Broken Power-law Fit

Here we report on additional tests we performed to study the dependence of the accretion on the stellar parameters as presented in Section 4.2. To test whether a single power law provides a better model for our data compared to a broken power law (corresponding to a line or segmented line in logarithmic space), we adopted an approach similar to the one presented in Manara et al. (2017b). We started dividing both $\log L_{\text{acc}}$ and $\log \dot{M}_{\text{acc}}$ distributions in bins of an equal number of sources, and we evaluate the median value for each bin (represented by the red crosses in Figure A1). Both relations show a bending in their trend when approaching higher values of L_* or M_* .

Then, we fitted the data with a linear relation,

$$y = \theta_0 + \theta_1 \cdot x, \quad (A1)$$

or a segmented line,

$$\begin{aligned} y &= \theta_0 + \theta_1 \cdot x && \text{if } x \leq x_c \\ y &= \theta_0 + \theta_1 \cdot x + \theta_2 \cdot [x - x_c] && \text{if } x > x_c, \end{aligned} \quad (A2)$$

where θ_i ($i = 0, 1, 2$) and x_c are free parameters for the fit, y is either $\log(L_{\text{acc}})$ or $\log(\dot{M}_{\text{acc}})$, and x is either $\log(L_*)$ or $\log(M_*)$, depending on the case.

We perform the linear fit using the RANSAC (Fischler & Bolles 1981) scikit-learn Python module (Pedregosa et al. 2011). The fit has been performed 100 times, where during each iteration 100 trials were performed by the RANSAC routine to establish the best-fit parameters, and a final median and standard deviation have been extracted from the overall output distributions of parameters, providing the following best fit:

$$\log L_{\text{acc}} = -(0.79 \pm 0.13) + (1.17 \pm 0.10) \log L_*. \quad (A3)$$

The same linear fit applied to the $\dot{M}_{\text{acc}}-M_*$ relation provides instead the following best fit:

$$\log \dot{M}_{\text{acc}} = -(-7.68 \pm 0.14) + (1.21 \pm 0.14) \log M_*. \quad (A4)$$

Table A1

Summary of the Results for the Selected Information Criteria Adopted in This Work to Compare the Single and Broken Power-law Models Obtained from Our Data

	R^2	AIC	BIC	p	Case
$\log L_{\text{acc}}$	0.59	-353	-335	...	Single power law
$\log L_{\text{acc}}$	0.62	-409	-391	4.6×10^{-7}	Broken power law
$\log \dot{M}_{\text{acc}}$	0.39	-258	-240	...	Single power law
$\log \dot{M}_{\text{acc}}$	0.47	-331	-349	5.6×10^{-13}	Broken power law

Table A2

Summary of the Results for the Selected Information Criteria Adopted in This Work to Compare the Single and Broken Power-law Models Obtained from Our Selected Data ($\log M_* \lesssim -0.15$)

	R^2	AIC	BIC	p	Case
$\log L_{\text{acc}}$	0.60	-311	-293	...	Single power law
$\log L_{\text{acc}}$	0.61	-325	-308	5.8×10^{-2}	Broken power law
$\log \dot{M}_{\text{acc}}$	0.44	-238	-221	...	Single power law
$\log \dot{M}_{\text{acc}}$	0.48	-273	-256	5.5×10^{-5}	Broken power law

Both relations are shown as a green dashed line in Figure A1.

The segmented line instead has been fitted using piecewise regression (Pilgrim 2021), which fits simultaneously the break-point positions and the linear models using an iterative method. Similar to the RANSAC case, 1000 trials were performed by the routine to establish the best-fit parameters, and a final median and standard deviation have been extracted from the overall distributions of parameters. For this case, the best-fit parameters for L_{acc} are $\theta_0 = -0.39 \pm 0.08$, $\theta_1 = 1.47 \pm 0.06$, $\theta_2 = -0.95 \pm 0.15$, and $x_c = -0.70 \pm 0.1$, which can be translated to

$$\begin{aligned} \log L_{\text{acc}} &= -(0.39 \pm 0.08) + (1.47 \pm 0.06) \log L_* \\ \text{if } \log L_* &\leq -0.70 \\ \log L_{\text{acc}} &= -(1.11 \pm 0.08) + (0.51 \pm 0.14) \log L_* \\ \text{if } \log L_* &> -0.70, \end{aligned} \quad (\text{A5})$$

while for \dot{M}_{acc} we obtain $\theta_0 = -6.77 \pm 0.14$, $\theta_1 = 1.98 \pm 0.11$, $\theta_2 = -1.84 \pm 0.19$, and $x_c = -0.68 \pm 0.05$, or

$$\begin{aligned} \log \dot{M}_{\text{acc}} &= -(6.77 \pm 0.14) + (1.98 \pm 0.11) \log M_* \\ \text{if } \log M_* &\leq -0.68 \\ \log \dot{M}_{\text{acc}} &= -(8.11 \pm 0.14) + (0.14 \pm 0.15) \log M_* \\ \text{if } \log M_* &> -0.68. \end{aligned} \quad (\text{A6})$$

Both relations are shown as a blue dashed line in Figure A1.

To compare the results of the linear and segmented models, we selected three information criteria: R^2 , the Akaike information criteria (AIC), and the Bayesian information criteria (BIC). All of them provide a statistical measure of how well the model approximates the real data, but the BIC and AIC also take into account the number of free parameters included in the model, penalizing the results accordingly. Overall, we can identify the better model by looking for the one that maximizes R^2 and at the same time minimize both the AIC and BIC. We assume that a difference in AIC between two models of 14 or more is enough to exclude the model with

higher AIC (Murtaugh 2014), while a difference in BIC of 10 or more is sufficient to exclude the model with higher BIC (Riviere-Marichalar et al. 2016). For the power law only, the piecewise regression algorithm also performs the Davies test (p) for the existence of at least one break point. In general, if $p < 0.05$, this means we can safely reject the null hypothesis of no break points at 5% significance (Davies 1987, 2002).

The results obtained for both models are presented in Table A1, showing that the broken power law maximizes the R^2 while minimizing both the AIC and the BIC for both the $\log L_{\text{acc}}$ and $\log \dot{M}_{\text{acc}}$ distributions (with a difference greater than the threshold provided above), and it can be assumed as a better descriptor of these distributions. Moreover, the broken power law always has $p \ll 0.05$, so the presence of a braking point is very likely. Because the second portion of the broken power-law fit (in particular for the $\dot{M}_{\text{acc}}-M_*$ relation) seems to be driven by a cluster of massive stars with lower \dot{M}_{acc} closer to the right end of our distribution (i.e., sources with $\log M_* \gtrsim -0.15$ or $\sim 0.7 M_\odot$), we decide to test whether the predilection for a broken power-law relation still stands if we eliminate these sources.

In this case, we obtain $\theta_0 = -0.4 \pm 0.09$, $\theta_1 = 1.46 \pm 0.06$, $\theta_2 = -1.04 \pm 0.4$, and $x_c = -0.63 \pm 0.11$, which can be translated to

$$\begin{aligned} \log L_{\text{acc}} &= -(0.40 \pm 0.09) + (1.46 \pm 0.06) \log L_* \\ \text{if } \log L_* &\leq -0.63 \\ \log L_{\text{acc}} &= -(0.99 \pm 0.09) + (0.42 \pm 0.40) \log L_* \\ \text{if } \log L_* &> -0.63, \end{aligned} \quad (\text{A7})$$

while for \dot{M}_{acc} we obtain $\theta_0 = -6.78 \pm 0.12$, $\theta_1 = 1.98 \pm 0.10$, $\theta_2 = -1.84 \pm 0.10$, and $x_c = -0.68 \pm 0.07$, or

$$\begin{aligned} \log \dot{M}_{\text{acc}} &= -(6.78 \pm 0.12) + (1.98 \pm 0.10) \log M_* \\ \text{if } \log M_* &\leq -0.68 \\ \log \dot{M}_{\text{acc}} &= -(7.92 \pm 0.24) + (0.13 \pm 0.41) \log M_* \\ \text{if } \log M_* &> -0.68. \end{aligned} \quad (\text{A8})$$

Both relations are shown as light-blue dashed lines in Figure A1.

Table A2 reports the values we evaluate for the three information criteria for this elected case. Also in this case, the three information criteria prefer the broken power law over the linear model (even if with a weakened signal, in particular for R^2), while still keeping $p \ll 0.05$, confirming that the change in the slope is real and cannot be ascribed only to the presence of the sources with $M_* \gtrsim 0.7$.

ORCID iDs

- Giovanni M. Strampelli  <https://orcid.org/0000-0002-1652-420X>
 Massimo Robberto  <https://orcid.org/0000-0002-9573-3199>
 Mario Gennaro  <https://orcid.org/0000-0002-5581-2896>
 Carlo F. Manara  <https://orcid.org/0000-0003-3562-262X>
 Elena Sabbi  <https://orcid.org/0000-0003-2954-7643>
 Antonio Aparicio  <https://orcid.org/0000-0002-6054-0004>

References

- Alcalá, J. M., Manara, C. F., Natta, A., et al. 2017, *A&A*, 600, A20
 Alcalá, J. M., Natta, A., Manara, C. F., et al. 2014, *A&A*, 561, A2

- Alexander, R. D., & Armitage, P. J. 2006, *ApJL*, **639**, L83
 Alves, J., & Bouy, H. 2012, *A&A*, **547**, A97
 Asplund, M., Grevesse, N., Sauval, A. J., & Scott, P. 2009, *ARA&A*, **47**, 481
 Astropy Collaboration, Robitaille, T. P., Tollerud, E. J., et al. 2013, *A&A*, **558**, A33
 Beccari, G., Petr-Gotzens, M. G., Boffin, H. M. J., et al. 2017, *A&A*, **604**, A22
 Bouy, H., Alves, J., Bertin, E., Sarro, L. M., & Barrado, D. 2014, *A&A*, **564**, A29
 Cardelli, J. A., Clayton, G. C., & Mathis, J. S. 1989, *ApJ*, **345**, 245
 Chabrier, G. 2003, *PASP*, **115**, 763
 Clarke, C. J., & Pringle, J. E. 2006, *MNRAS*, **370**, L10
 Da Rio, N., Robberto, M., Hillenbrand, L. A., Henning, T., & Stassun, K. G. 2012, *ApJ*, **748**, 14
 Davies, R. B. 1987, *Biometrika*, **74**, 33
 Davies, R. B. 2002, *Biometrika*, **89**, 484
 Dullemond, C. P., Natta, A., & Testi, L. 2006, *ApJL*, **645**, L69
 Ercolano, B., Mayr, D., Owen, J. E., Rosotti, G., & Manara, C. F. 2014, *MNRAS*, **439**, 256
 Fang, M., Kim, J. S., van Boekel, R., et al. 2013, *ApJS*, **207**, 5
 Fischer, W. J., Hillenbrand, L. A., Herczeg, G. J., et al. 2023, in ASP Conf. Ser. 534, Protostars and Planets VII, ed. S.-I. Inutsuka et al. (San Francisco, CA: ASP), 355
 Fischler, M. A., & Bolles, R. C. 1981, *Commun. ACM*, **24**, 381
 Flaischlen, S., Preibisch, T., Kluge, M., Manara, C. F., & Ercolano, B. 2022, *A&A*, **666**, A55
 Foreman-Mackey, D. 2016, *JOSS*, **1**, 24
 France, K., Yang, H., & Linsky, J. L. 2011, *ApJ*, **729**, 7
 Gaia Collaboration, Brown, A. G. A., Vallenari, A., et al. 2021, *A&A*, **649**, A1
 Gangi, M., Antoniucci, S., Biazzo, K., et al. 2022, *A&A*, **667**, A124
 Gennaro, M., & Robberto, M. 2020, *ApJ*, **896**, 80
 Goodman, J., & Weare, J. 2010, *CAMCS*, **5**, 65
 Gullbring, E., Calvet, N., Muñoz, J., & Hartmann, L. 2000, *ApJ*, **544**, 927
 Gullbring, E., Hartmann, L., Briceño, C., & Calvet, N. 1998, *ApJ*, **492**, 323
 Gully-Santiago, M. A., Herczeg, G. J., Czekala, I., et al. 2017, *ApJ*, **836**, 200
 Hartmann, L., Calvet, N., Gullbring, E., & D'Alessio, P. 1998, *ApJ*, **495**, 385
 Hartmann, L., Herczeg, G., & Calvet, N. 2016, *ARA&A*, **54**, 135
 Herbst, W., Bailer-Jones, C. A. L., Mundt, R., Meisenheimer, K., & Wacker, R. 2002, *A&A*, **396**, 513
 Hillenbrand, L. A. 1997, *AJ*, **113**, 1733
 Hillenbrand, L. A., & Hartmann, L. W. 1998, *ApJ*, **492**, 540
 Hillenbrand, L. A., Hoffer, A. S., & Herczeg, G. J. 2013, *AJ*, **146**, 85
 Hunter, J. D. 2007, *CSE*, **9**, 90
 Jeffries, R. D., Littlefair, S. P., Naylor, T., & Mayne, N. J. 2011, *MNRAS*, **418**, 1948
 Jerabkova, T., Beccari, G., Boffin, H. M. J., et al. 2019, *A&A*, **627**, A57
 Johns-Krull, C. M., & Valenti, J. A. 1996, *ApJL*, **459**, L95
 Kroupa, P. 2001, *MNRAS*, **322**, 231
 Kuhn, M. A., Hillenbrand, L. A., Sills, A., Feigelson, E. D., & Getman, K. V. 2019, *ApJ*, **870**, 32
 Lada, C. J., & Lada, E. A. 2003, *ARA&A*, **41**, 57
 Luhman, K. L., Stauffer, J. R., Muench, A. A., et al. 2003, *ApJ*, **593**, 1093
 Manara, C. F., Ansdell, M., Rosotti, G. P., et al. 2023, in ASP Conf. Ser. 534, Protostars and Planets VII, ed. S.-I. Inutsuka (San Francisco, CA: ASP), 539
 Manara, C. F., Fedele, D., Herczeg, G. J., & Teixeira, P. S. 2016a, *A&A*, **585**, A136
 Manara, C. F., Frasca, A., Alcalá, J. M., et al. 2017a, *A&A*, **605**, A86
 Manara, C. F., Robberto, M., Da Rio, N., et al. 2012, *ApJ*, **755**, 154
 Manara, C. F., Rosotti, G., Testi, L., et al. 2016b, *A&A*, **591**, L3
 Manara, C. F., Testi, L., Herczeg, G. J., et al. 2017b, *A&A*, **604**, A127
 McBride, A., & Kounkel, M. 2019, *ApJ*, **884**, 6
 McKinney, W., et al. 2010, in Proc. 9th Python in Science Conf., 56
 Murtaugh, P. A. 2014, *Ecol*, **95**, 611
 Muñoz, J., Luhman, K. L., Briceño, C., Hartmann, L., & Calvet, N. 2005, *ApJ*, **625**, 906
 Natta, A., Testi, L., & Randich, S. 2006, *A&A*, **452**, 245
 Pedregosa, F., Varoquaux, G., Gramfort, A., et al. 2011, *Journal of Machine Learning Research*, **12**, 2825
 Pilgrim, C. 2021, *JOSS*, **6**, 3859
 Portegies Zwart, S. F., McMillan, S. L., & Gieles, M. 2010, *ARA&A*, **48**, 431
 Price-Whelan, A. M., Sipőcz, B. M., Günther, H. M., et al. 2018, *AJ*, **156**, 123
 Prosser, C. F., Stauffer, J. R., Hartmann, L., et al. 1994, *ApJ*, **421**, 517
 Reggiani, M., Robberto, M., Da Rio, N., et al. 2011, *A&A*, **534**, A83
 Ricci, L., Robberto, M., & Soderblom, D. R. 2008, *AJ*, **136**, 2136
 Rigliaco, E., Natta, A., Randich, S., Testi, L., & Biazzo, K. 2011, *A&A*, **525**, A47
 Riviere-Marichalar, P., Merín, B., Kamp, I., Eiroa, C., & Montesinos, B. 2016, *A&A*, **594**, A59
 Robberto, M., Gennaro, M., Ubeira Gabellini, M. G., et al. 2020, *ApJ*, **896**, 79
 Robberto, M., Soderblom, D. R., Bergeron, E., et al. 2013, *ApJS*, **207**, 10
 Scandariato, G., Robberto, M., Pagano, I., & Hillenbrand, L. A. 2011, *A&A*, **533**, A38
 Slesnick, C. L., Hillenbrand, L. A., & Carpenter, J. M. 2004, *ApJ*, **610**, 1045
 Soderblom, D. R., Hillenbrand, L. A., Jeffries, R. D., Mamajek, E. E., & Taylor, T. 2014, in Protostars and Planets VI, ed. H. Beuther et al. (Tucson, AZ: Univ. of Arizona Press), 219
 Somers, G., & Pinsonneault, M. H. 2015, *ApJ*, **807**, 174
 Strampelli, G. M., Pueyo, L., Aguilar, J., et al. 2022, *AJ*, **164**, 147
 STScI Development Team 2018, synphot: Synthetic photometry using Astropy, Astrophysics Source Code Library, ascl:1811.001
 STScI Development Team 2020, stsynphot: synphot for HST and JWST, Astrophysics Source Code Library, ascl:2010.003
 van der Walt, S., Colbert, S. C., & Varoquaux, G. 2011, *CSE*, **13**, 22
 Venuti, L., Bouvier, J., Flaccomio, E., et al. 2014, *A&A*, **570**, A82
 Virtanen, P., Gommers, R., Oliphant, T. E., et al. 2020, *NatMe*, **17**, 261
 Vorobyov, E. I., & Basu, S. 2009, *ApJ*, **703**, 922
 Wang, J. J., Ruffio, J.-B., De Rosa, R. J., et al. 2015, pyKLIP: PSF Subtraction for Exoplanets and Disks, Astrophysics Source Code Library, ascl:1506.001
 Winter, A. J., & Haworth, T. J. 2022, *EPJP*, **137**, 1132