Monthly Notices
of the
ROYAL ASTRONOMICAL SOCIETY

MNRAS **438,** 2355–2370 (2014)
Advance Access publication 2014 January 10

doi:10.1093/mnras/stt2351

An analytic method to compute star cluster luminosity statistics

Robert L. da Silva,[1⋆†] Mark R. Krumholz,[1⋆] Michele Fumagalli[2,3⋆‡] and S. Michael Fall[4⋆]

[1]*Department of Astronomy and Astrophysics, University of California, 1156 High Street, Santa Cruz, CA 95064, USA*
[2]*Carnegie Observatories, 813 Santa Barbara Street, Pasadena, CA 91101, USA*
[3]*Department of Astrophysics, Princeton University, Princeton, NJ 08544-1001, USA*
[4]*Space Telescope Science Institute, 3700 San Martin Drive, Baltimore, MD 21218, USA*

Accepted 2013 December 3. Received 2013 October 25

ABSTRACT

The luminosity distribution of the brightest star clusters in a population of galaxies encodes critical pieces of information about how clusters form, evolve and disperse, and whether and how these processes depend on the large-scale galactic environment. However, extracting constraints on models from these data is challenging, in part because comparisons between theory and observation have traditionally required computationally intensive Monte Carlo methods to generate mock data that can be compared to observations. We introduce a new method that circumvents this limitation by allowing analytic computation of cluster order statistics, i.e. the luminosity distribution of the *N*th most luminous cluster in a population. Our method is flexible and requires few assumptions, allowing for parametrized variations in the initial cluster mass function and its upper and lower cutoffs, variations in the cluster age distribution, stellar evolution and dust extinction, as well as observational uncertainties in both the properties of star clusters and their underlying host galaxies. The method is fast enough to make it feasible for the first time to use Markov chain Monte Carlo methods to search parameter space to find best-fitting values for the parameters describing cluster formation and disruption, and to obtain rigorous confidence intervals on the inferred values. We implement our method in a software package called the Cluster Luminosity Order-Statistic Code, which we have made publicly available.

Key words: methods: data analysis – methods: numerical – methods: statistical – techniques: photometric – galaxies: star clusters: general.

1 INTRODUCTION

Stars do not generally form in isolation, in either space or time. Instead, they form in a spatially and temporally clustered fashion (e.g. Lada & Lada 2003; Bressert et al. 2010; Gutermuth et al. 2011), at a density far above the background field stellar density in their host galaxy. This clustering has profound effects on the observable properties of galaxies (Fumagalli, da Silva & Krumholz 2011b; da Silva, Fumagalli & Krumholz 2012), and it also provides an important clue to the physical mechanisms that govern the process of star formation. If we could confidently measure the fraction of stars that form in star clusters, the mass distribution of those clusters (including its upper and lower limits), and the rate at which clusters dissolve into the field stellar population, we would learn a great deal about how stars form.

Unfortunately, extracting all of these quantities from observations is far from trivial. When high spatial resolution multicolor photometry is available, the standard approach is to use stellar population synthesis models to assign masses and ages to each cluster, then measure the distributions of these or other quantities of interest. These observations generally indicate that the mass distribution can be approximated as a (possibly truncated) power law $\mathrm{d}N/\mathrm{d}M \propto M^{\beta}$ with $\beta \approx -2$ over a wide mass range (e.g. Zhang & Fall 1999; Bik et al. 2003; Boutloukos & Lamers 2003; Fall 2006; Fall, Chandar & Whitmore 2009; Chandar, Fall & Whitmore 2010; Bastian et al. 2012a,b; Fall & Chandar 2012). There is more controversy over the age distribution, mostly arising from issues of how the samples are selected. More inclusive cluster catalogues constructed to include all objects above a surface brightness threshold tend to show power-law age distributions $\mathrm{d}N/\mathrm{d}t \propto t^{\gamma}$ with $\gamma \approx -0.9$ (Fall, Chandar & Whitmore 2005; Fall et al. 2009; Chandar et al. 2010; Fall & Chandar 2012). If one imposes additional selection

⋆ E-mail: rdasilva.astro@gmail.com (RLd); mkrumhol@ucsc.edu (MRK); fumagalli.astro@gmail.com (MF); fall@stsci.edu (SMF)
† NSF Graduate Research Fellow.
‡ Hubble Fellow.

criteria based on morphology or crowding, this removes many young clusters from the sample, yielding a distribution that can still be approximated by a power law, but with a significantly shallower index, $\gamma \approx 0$ (Gieles, Lamers & Portegies Zwart 2007; Bastian et al. 2011, 2012a,b). While measuring the mass and age distribution by assigning masses and ages to all clusters has the virtue of being conceptually direct, the data required to use this approach are available only for a relatively modest number of galaxies. Single-band photometry capable of resolving the brightest few clusters is available for a much larger sample of galaxies (e.g. Larsen & Richtler 1999; Larsen 2002; Bastian 2008), and exploiting such large but lower quality data sets is the only feasible means to detect whether cluster mass or age distributions deviate from power-law behaviour at the very high mass end, where the number of clusters in an individual galaxy is necessarily very small, and data from many galaxies must therefore be combined to yield a statistically meaningful result.

The primary method of using these data to study the tip of the cluster mass function has traditionally been to use Monte Carlo methods to compute the luminosity distribution that would be expected from a given theoretical model, and compare that to the observations (e.g. Bastian 2008; Larsen 2009; Fouesneau et al. 2012). This has the advantage that it allows one to handle observational errors properly, and to include 'nuisance' parameters such as dust extinction that limit the information that can be extracted from data. Unfortunately, Monte Carlo methods can be forbiddingly expensive to employ. The Lada & Lada (2003) compilation of clusters just within 2 kpc of the Sun includes $\sim$100 entries, and this survey covers only $\sim$1–3 per cent of the Milky Way's star-forming disc, and a significantly smaller fraction of the Milky Way's total star formation budget. Thus, a single Monte Carlo realization of the star clusters in a Milky Way-like galaxy, including the effects of cluster disruption, might require that $\sim 10^6$ clusters be drawn, and determining the order statistics of this distribution (i.e. the distributions of luminosities of the most luminous cluster, second most luminous cluster, etc.) might then require $\sim 10^3$ realizations, for a total of $\sim 10^9$ total draws from the cluster luminosity distribution. The problem is far worse if one considers more rapidly star-forming galaxies like the Antennae, which have larger cluster populations. Since our knowledge of the various processes that influence cluster luminosity distributions is limited, and the set of parameters describing them is therefore relatively large, ideally one would like to be able to search the parameter space for models that fit observations using standard Markov chain Monte Carlo (MCMC) methods. However, this is not feasible if one requires $\sim 10^9$ draws from the cluster luminosity distribution at every point in this parameter space. As a result, many authors have resorted to fixing many of the parameters that describe cluster formation, and varying only a single one (e.g. the upper mass cutoff – Bastian 2008) in an attempt to fit observations. Clearly, this approach is not ideal.

In this paper, we introduce a method to solve this problem. We show that it is possible to calculate the cluster luminosity distribution and its order statistics *analytically*, even including parametrized treatments of processes such as cluster disruption, stellar evolution and dust extinction. While our method is not quite as general in the types of distributions that one can handle as a full Monte Carlo method, it retains the vast majority of Monte Carlo's flexibility and requires only a tiny fraction of the computational time. Moreover, using our method the computational time is close to independent of the number of clusters present and it is much more expensive to compute order statistics than it is to compute the luminosity function itself. This makes our method particularly advantageous for calculations involving large galaxies, and those seeking to explore the tip of the luminosity distribution. In a companion paper (da Silva et al., in preparation), we use this method to revisit the question of whether the observed relationship between star formation rate (SFR) of luminosity of the most luminous cluster provides strong constraints on the upper limit to the cluster mass function. We have developed a software tool called the Cluster Luminosity Order-Statistic Code (CLOC) to perform these analytic calculations, and made it publicly available under the terms of the GNU General Public License.

The remainder of this paper is organized as follows. Section 2 describes our model and its derivation. In Section 3, we describe the publicly available code that implements this model, and present comparisons between it and a full Monte Carlo method. In Section 4, we use our formalism to explore how the various parameters that go into the cluster luminosity function affects its shape in order to gain insight about what sorts of observations can be used to constrain star cluster formation.

2 THE MODEL

2.1 Cluster order statistics

Our overall goal is to derive an analytic expression for the probability distribution function (PDF) and cumulative distribution function (CDF) of the kth order statistic of star cluster luminosities, or any other property. Formally, we define $\phi_k(L)$ as the PDF of the kth most luminous cluster in a region of interest, either a galaxy or some specified subgalactic volume.[1] We normalize this and all other PDFs in this paper to unity, i.e. $\int \phi_k(L)\,\mathrm{d}L = 1$. We define $\Phi_k(L) = \int_0^L \phi_k(L')\,\mathrm{d}L'$ as the corresponding CDF. Thus, $\Phi_k(L)$ is the probability that the kth most luminous cluster in a population has a luminosity of L or less, while $\phi_k(L)\,\mathrm{d}L$ is the probability that the kth most luminous cluster has a luminosity in the infinitesimal range L to $L + \mathrm{d}L$.

We will perform this calculation in several steps. In this section, we will derive $\phi_k(L)$ and $\Phi_k(L)$ under the assumption that we know both the PDF $\phi(L)$ for the luminosity of a single cluster and the expected number of clusters $\langle N \rangle$ in the region of interest. In subsequent sections, we will derive these two quantities from parametrized versions of the cluster mass and age distributions. Deriving $\phi_k(L)$ and $\Phi_k(L)$ from $\phi(L)$ and $\langle N \rangle$ is most straightforward if we assume that cluster formation is a Poisson process, so that clusters are created fully independently of one another, and the number of clusters in a given region is Poisson distributed. We note that this cannot be precisely true, simply due to mass conservation: for a purely Poisson distribution, there is a finite, non-zero probability for any number or total mass of clusters, whereas in reality the probability that the total mass of star clusters in a given region exceeds the total baryonic mass of the region is identically zero. Nonetheless, when the total mass of star clusters is large compared to the mass of any individual cluster (as is often the case in practice), then the Poisson assumption should be reasonable, and so we will adopt it. In Appendix A,

[1] Note that our convention in defining the first order statistic as describing the distribution of the most massive or luminous cluster, while sensible for astronomy (where samples are usually mass- or luminosity-limited and thus the least luminous cluster is generally not observed), is the opposite of the standard statistics convention whereby the first order statistic describes the distribution of the smallest member of a sample, not the largest. The usual convention may be recovered by replacing k by $N + 1 - k$ in all the expressions below, where N is the size of the sample.

we provide a more detailed derivation of the PDF and CDF that shows how to generalize to the non-Poisson case.

There is one more subtlety with which we must reckon before proceeding to calculate. For any Poisson process, and for most non-Poisson ones, there is a finite probability that a region of interest will contain a number of clusters N that is smaller than the order statistic k in which we are interested. For example, we might be interested in the luminosity distribution of the second most luminous cluster ($k = 2$), but some of the regions we are examining will contain only 0 or 1 clusters. We must therefore decide between two possible ways of handling this case: we could either say that the kth most luminous cluster has a luminosity of 0 if there are fewer than k clusters present, or we could restrict our calculation of the PDF of the kth most luminous cluster to the case where k or more clusters are present. We argue in Appendix A that the former approach is preferable, and we will therefore say that, if the number of clusters N is smaller than the order statistic k we are computing, then the luminosity of the kth most luminous cluster is 0.

With this choice, we are now prepared to derive $\phi_k(L)$ and $\Phi_k(L)$. For a Poisson process, the expected number of clusters with luminosity $>L$ is

$$\langle N(>L)\rangle = \langle N\rangle(1-\Phi(L)), \tag{1}$$

where $\Phi(L) = \int_0^L \phi(L')\,\mathrm{d}L'$ is the CDF of luminosity for a single cluster. The probability that there are exactly m clusters with luminosity $>L$ is given by the Poisson formula

$$P_m(>L) = \frac{1}{m!}\langle N(>L)\rangle^m e^{-\langle N(>L)\rangle} \tag{2}$$

$$= \frac{e^{-\langle N\rangle}}{m!}\langle N\rangle^m[1-\Phi(L)]^m e^{\langle N\rangle\Phi(L)}. \tag{3}$$

When the number of clusters present, N, is larger than the order k, the PDF $\phi_k(L)$ should be proportional to the probability that a single cluster is in the luminosity range L to $L + \mathrm{d}L$, multiplied by the probability that exactly $k-1$ clusters are more luminous than L, i.e. we should have $\phi_k(L) \propto \phi(L)P_{k-1}(>L)$. When $N < k$, the luminosity of the kth most luminous cluster is zero. Combining these two terms, the complete PDF is

$$\phi_k(L) = \frac{\Gamma(k,\langle N\rangle)}{\Gamma(k)}\delta(L) + \langle N\rangle P_{k-1}(>L)\phi(L) \tag{4}$$

$$= \frac{\Gamma(k,\langle N\rangle)}{\Gamma(k)}\delta(L) + \langle N\rangle\frac{\langle N\rangle^{k-1}[1-\Phi(L)]^{k-1}}{(k-1)!}\phi(L), \tag{5}$$

where $\Gamma(x)$ is the usual (complete) Γ function and $\Gamma(x, s)$ is the incomplete Γ function. The coefficient of the δ function is the probability that, for a Poisson process, the number of clusters is smaller than k, while the second term represents the product $\phi_k(L)P_{k-1}(>L)$. The coefficient on this term is chosen to ensure that $\int\phi_k(L)\,\mathrm{d}L = 1$. We derive both coefficients by alternative means in Appendix A. Note that the coefficient of the δ function, $\Gamma(k, \langle N\rangle)/\Gamma(k)$, goes to zero extremely rapidly as $k/\langle N\rangle \to 0$. Thus, this term is significant only when $\langle N\rangle \lesssim k$. There term would vanish entirely if we adopted the alternative approach to defining order statistics by excluding the case $N < k$, but in that case the other term would have to modified as well.

The CDF $\Phi_k(L)$ is the probability that the kth most luminous cluster has a luminosity $\leq L$, but this must be equal to the probability that at most $k-1$ clusters have luminosities $\geq L$. For example, the probability that the second brightest cluster has a luminosity $\leq L$, which is $\Phi_2(L)$, must be equal to the probability that there are either 0 or 1 clusters brighter than L, which is $P_0(>L) + P_1(>L)$. Thus, in general we have

$$\Phi_k(L) = \sum_{m=0}^{k-1} P_m(>L) \tag{6}$$

$$= e^{-\langle N\rangle}e^{\langle N\rangle\Phi(L)}\sum_{m=0}^{k-1}\frac{\langle N\rangle^m[1-\Phi(L)]^m}{m!}. \tag{7}$$

Note that $\Phi_k(L)$ remains finite in the limit $L \to 0$, even if $\Phi(L)$ is identically zero below some finite minimum L. This behaviour occurs because, even if there is zero probability that any individual cluster has a luminosity $L = 0$, we can still find a luminosity of exactly 0 for the kth most luminous cluster if there are fewer than k clusters present in the region of interest, and the probability of this occurring is finite.

2.2 Calculation of the expected number of clusters

The second step in our derivation is to calculate the expected number of clusters $\langle N\rangle$. This is a function of the SFR in the region under study $\dot{M}_*$, the star cluster mass function $\psi(M)$, the cluster age distribution $\chi(t)$ defined over the full range of cluster masses,[2] the minimum and maximum cluster ages $t_{\rm min}$ and $t_{\rm max}$ used to define the sample, and the fraction of stars in clusters at birth, which we denote $f_{\rm c}$.[3] We normalize the mass and age distributions such that $\int\psi(M)\,\mathrm{d}M = \int\chi(t)\,\mathrm{d}t = 1$, where the integrals are taken over all possible masses and ages, respectively.

Given these definitions, the expected number of clusters *formed* during the time interval of interest is

$$\langle N_{\rm form}\rangle = \frac{\dot{M}_*\Delta t}{\langle M\rangle}f_{\rm c}, \tag{8}$$

where $\langle M\rangle = \int M\psi(M)\,\mathrm{d}M$ is the expectation value of the cluster mass and $\Delta t = t_{\rm max} - t_{\rm min}$ is the age range in the observed sample. If the cluster age distribution is not flat, indicating that not all clusters that form survive to indefinite ages, the expected number of clusters

[2] There are two subtle points to be made here. First, the age distribution $\chi(t)$ must be that for all clusters, not, as is sometimes reported in the literature, the age distribution for a luminosity-limited sample. Secondly, in principle the mass and age distributions might not be independent, in which case we would need to consider the joint distribution $g(M, t)$. There is a dispute on this point in the observational literature – e.g. see Bastian et al. (2011, 2012a,b) versus Fall et al. (2009), Chandar et al. (2010) and Fall & Chandar (2012). Fortunately, even in those papers where the authors do report that the mass and age distributions are not independent, the covariance is very weak, at least at the large masses with which we will be concerned. Similarly, some theoretical models also predict that cluster disruption will be mass dependent (e.g. Kruijssen et al. 2012), but the predicted dependence is again weak. For these reasons, we will assume that the mass and age distributions are independent.

[3] Note that the quantity $f_{\rm c}$ is subtly different from the cluster formation efficiency Γ defined by some authors (e.g. Bastian 2008), because Γ refers to the fraction of stars formed as part of gravitationally bound clusters. In contrast, $f_{\rm c}$ depends only on the observational criteria used to define clusters when selecting them in an observed galaxy. Thus, $f_{\rm c}$ and Γ are identical only if the observational selection criteria pick out all gravitationally bound structures, and only such structures.

that survive long enough to be observed will be reduced. Let $P_{\rm surv}(t)$ be the probability that a cluster survives to age t, in which case

$$\langle N\rangle = \frac{\dot{M}\Delta t}{\langle M\rangle} f_{\rm c}\left(\frac{1}{\Delta t}\int_{t_{\rm min}}^{t_{\rm max}} P_{\rm surv}(t)\,{\rm d}t\right) \equiv \frac{\dot{M}_*\Delta t}{\langle M\rangle}\mathcal{F}_{\rm c} \tag{9}$$

is the expected number of surviving clusters within the age interval of interest. The quantity in parentheses is the time-averaged fraction of surviving clusters, and the quantity $\mathcal{F}_{\rm c}$ that we have defined is the fraction of all stars in clusters, averaged over the stellar age range under consideration. For a constant SFR, the survival probability is proportional to the cluster age distribution, renormalized so that the survival probability is unity at time $t = 0$, i.e. $P_{\rm surv}(t) = \chi(t)/\chi(0)$. Thus,

$$\mathcal{F}_{\rm c} = \frac{f_{\rm c}}{\chi(0)\Delta t}\int_{t_{\rm min}}^{t_{\rm max}} \chi(t)\,{\rm d}t. \tag{10}$$

Note that although for simplicity we have assumed constant $\dot{M}_*$, our results in the end depend only on the cluster age distribution $\chi(t)$ and the expected number of clusters $\langle N\rangle$. Thus, the formulae throughout this paper are equally valid for other combinations of $P_{\rm surv}(t)$ and $\dot{M}_*(t)$ that give the same $\chi(t)$ and $\langle N\rangle$.

To proceed further we must specify functional forms for $\psi(M)$ and $\chi(t)$. To render the problem analytically tractable, we will assume that both of these can be described by truncated power laws. Specifically, we adopt

$$\psi(M) = \begin{cases} AM^{\beta}, & M_{\rm min} < M < M_{\rm max} \\ 0, & \text{otherwise} \end{cases} \tag{11}$$

and

$$\chi(t) = \begin{cases} B, & t < t_0 \\ B(t/t_0)^{\gamma}, & t_0 \le t < t_1 \\ 0, & t \ge t_1. \end{cases} \tag{12}$$

Here t_0 may be understood as the age at which clusters begin to disappear and t_1 is the maximum possible age of any cluster. The normalization factors appearing in these equations are

$$\begin{aligned}\frac{1}{A} &= \int_{M_{\rm min}}^{M_{\rm max}} M^{\beta}\,{\rm d}M \\ &= \begin{cases} (M_{\rm max}^{\beta+1} - M_{\rm min}^{\beta+1})/(\beta+1), & \beta \ne -1 \\ \ln(M_{\rm max}/M_{\rm min}), & \beta = -1. \end{cases}\end{aligned} \tag{13}$$

and

$$\begin{aligned}\frac{1}{B} &= t_0 + \int_{t_0}^{t_1}\left(\frac{t}{t_0}\right)^{\gamma}{\rm d}t \\ &= t_0 + \begin{cases} (t_1^{\gamma+1} - t_0^{\gamma+1})/(\gamma+1), & \gamma \ne -1 \\ \ln(t_1/t_0), & \gamma = -1. \end{cases}\end{aligned} \tag{14}$$

The functional forms for both the mass and age distributions are well motivated by observations. As discussed in Section 1, there is an observational consensus that the mass function is well fitted by a (possibly truncated) power law with index $\beta \approx -2$. There is dispute in the observational community about the age distribution, but all groups agree that a power law is a good fit to the data. The dispute is whether the index $\gamma \approx -0.9$ or ≈ 0, with most of the disagreement stemming from how the cluster sample is selected.

With these definitions, we can write out $\langle M\rangle$ explicitly as

$$\langle M\rangle = A \times \begin{cases} \left(M_{\rm max}^{\beta+2} - M_{\rm min}^{\beta+2}\right)/(\beta+2), & \beta \ne -2 \\ \ln(M_{\rm max}/M_{\rm min}), & \beta = -2. \end{cases} \tag{15}$$

We can similarly write out $\mathcal{F}_{\rm c}$ explicitly. For simplicity, we will assume that $t_{\rm min} \ge t_0$ and $t_{\rm max} \le t_1$, so that the age distribution $\chi(t)$ over the observed age range can be represented by a pure power law. Given that all observed open cluster samples satisfy this condition, this is not a significant limitation. With this assumption, we have

$$\begin{aligned}\mathcal{F}_{\rm c} &= f_{\rm c}\frac{t_0}{\Delta t} \\ &\times \begin{cases} \left[(t_{\rm max}/t_0)^{\gamma+1} - (t_{\rm min}/t_0)^{\gamma+1}\right]/(\gamma+1), & \gamma \ne -1 \\ \ln(t_{\rm max}/t_{\rm min}), & \gamma = -1. \end{cases}\end{aligned} \tag{16}$$

We could substitute this into equation (9) to obtain an explicit form for $\langle N\rangle$, but this would simply replace $\mathcal{F}_{\rm c}$ with $f_{\rm c}$ and t_0 as the variables that must be specified to compute cluster luminosity order statistics. Since these two quantities enter the problem only through the combination $\mathcal{F}_{\rm c}$, we will use $\mathcal{F}_{\rm c}$ as the variable of interest through the rest of this work, keeping in mind that it is related to the physical quantities $f_{\rm c}$ and t_0 via equation (16).

2.3 The cluster luminosity function: dust and stellar evolution

The final step in our calculation is to derive the PDF $\phi(L)$ for the luminosity of a single cluster. This quantity depends on three factors. The first is the cluster mass distribution $\psi(M)$, since more massive clusters are more luminous, all other things being equal. The second is the cluster age distribution $\chi(t)$, since at fixed mass there will be a range of cluster ages, and the mass-to-light ratio depends on the cluster age. The third factor is the distribution of dust optical depths, which we denote $\eta(\tau)$. The amount of extinction may vary from cluster to cluster, and this will create a scatter in the observed luminosity even at fixed mass and age. A fourth possible factor, which we will not include in our formalism, is stochastic variation in luminosity from cluster to cluster at fixed mass, age and extinction due to the effects of incomplete initial mass function (IMF) sampling. While this is significant for clusters with masses below $\sim 10^{3.5}$ $M_\odot$ (Cerviño & Luridiana 2004; Fouesneau et al. 2012), we focus in this work on the PDF of luminous and massive clusters, and in particular on the PDF of the most luminous cluster, which minimizes the importance of this effect. Below, we verify via Monte Carlo calculation that this effect is indeed negligible for $\phi_1(L)$ except at the very lowest SFRs. Thus, we are left with age-dependent mass-to-light ratio and dust extinction as the two effects we must include.

To handle the age dependence, we define $\Upsilon(t)$ as the mass-to-light ratio for a cluster of age t, so that the luminosity $L = M/\Upsilon(t)$; note that $\Upsilon(t)$ must be defined relative to a particular waveband. For ages t in the range 10 Myr to 1 Gyr, and many wavebands in the visible part of the spectrum, it is approximately the case that $\Upsilon(t) \propto t^{\zeta}$, where both the index ζ and the constant of proportionality depend on the choice of waveband. In Appendix B we fit for $\Upsilon(t)$ in V band, and obtain

$$\Upsilon(t) = \Upsilon_*\left(\frac{t}{10\ {\rm Myr}}\right)^{\zeta}, \tag{17}$$

with $\zeta = 0.688$ and $\Upsilon_* = 8.3 \times 10^{-21}\ M_\odot\ ({\rm ergs^{-1}Hz^{-1}})^{-1}$.

Since $\Upsilon(t)$ is a deterministic one-to-one function of t, the distribution of mass-to-light ratios for a cluster population can be computed from the distribution of cluster ages via

$$\theta(\Upsilon) \propto \chi(t)\left|\frac{{\rm d}\Upsilon}{{\rm d}t}\right|^{-1}. \tag{18}$$

The intrinsic luminosity of a cluster (i.e. before dust extinction is applied) is $L_{\rm in} = M/\Upsilon$, and so the distribution of intrinsic luminosities is (Fall 2006)

$$\phi_{\rm in}(L_{\rm in}) = \int_0^\infty \psi(\Upsilon L_{\rm in})\theta(\Upsilon)\Upsilon \,{\rm d}\Upsilon. \quad (19)$$

For the purposes of algebraic evaluation, it is most convenient to transform to logarithmic variables, which allows us to write the integral as a convolution. We define $\xi([1/\Upsilon]) = \Upsilon^2\theta(\Upsilon)$ as the distribution of light-to-mass (instead of mass-to-light) ratios, and compute the PDF $\phi_{\rm in}(\log L_{\rm in}) = L_{\rm in}\phi_{\rm in}(L_{\rm in})$ via

$$\begin{aligned}\phi_{\rm in}(\log L_{\rm in}) &\propto \psi(\log M) * \xi(-\log\Upsilon)\\ &\equiv \int_{-\infty}^\infty \psi(\log L_{\rm in} - \log\Upsilon)\xi(-\log\Upsilon)\,{\rm d}\log\Upsilon, \quad (20)\end{aligned}$$

where $*$ denotes convolution. We defer actual calculation of this convolution to Appendix C, since it is conceptually straightforward but algebraically tedious. The result for $\phi_{\rm in}(\log L_{\rm in})$ is given by equation (C14), and the corresponding CDF $\Phi_{\rm in}(\log L_{\rm in})$ by equations (C15) and (C20).

We model dust as providing a distribution of optical depths $\eta(\tau)$ (in the appropriate waveband) that is uniform in the range from τ_0 to τ_1, such that the luminosity of a given cluster is reduced by a factor of $e^{-\tau}$. Our choice of distribution is motivated by a compromise between realism and analytic tractability. The simplest approach would be to adopt a single dust optical depth for all clusters, which corresponds to decreasing the luminosity of each cluster by a constant factor. This is trivial to include, but would miss the potentially important effect that differential extinction can broaden the luminosity distribution. In order to capture this effect while still retaining a distribution that can be calculated analytically, we adopt the next most complicated approach, which is a step function distribution. Should it be desirable, it is straightforward to mix distributions with different step functions to create essentially arbitrary dust distributions. From this distribution of dust extinctions, and the distribution of intrinsic luminosities computed above, the distribution of observed luminosities can again be obtained via convolution,

$$\begin{aligned}\phi(\log L) &= \phi_{\rm in}(\log L_{\rm in}) * \eta(-\tau)\\ &= \int_{-\infty}^\infty \phi_{\rm in}(\log L_{\rm in})\eta(\log L_{\rm in} - \log L)\,{\rm d}\log L_{\rm in}, \quad (21)\end{aligned}$$

where $L = L_{\rm in}e^{-\tau}$. As with the computation required to compute the PDF of intrinsic luminosities, the convolution is straightforward but algebraically tedious to compute. We give the result in Appendix D; the final expressions for $\phi(\log L)$ and $\Phi(\log L)$ are given by equations (D11) and (D14), respectively.

With this step complete, we now have a full analytic description of the order statistics of cluster luminosities, including the effects of age-dependent mass-to-light ratios and a range of dust extinctions. Specifically, we can compute the PDF and CDF of an arbitrary order statistic from equations (5) and (7), using the expected number of clusters $\langle N\rangle$ given by equation (9) and the PDF and CDF of luminosity for individual clusters given by equations (D11) and (D14).

2.4 Observational uncertainties

We now add one final element to our model, which is that neither SFRs nor cluster luminosities can be measured perfectly. There are several options for how to treat this issue, depending on the application one has in mind. One might choose simply to use the formalism above to generate theoretical distributions of cluster luminosity, and then compare these to observations using a statistical technique that accounts for the observational errors. In this case, one can simply use the formalism as we have described it thus far, without accounting for observational error. However, an alternative and often preferable approach is to fold reasonable estimates of the errors into the theoretical model, and then to compare the model including these error estimates with the observed data. This makes it possible to use non-parametric tests (e.g. the Kolmogorov–Smirnov test) that do not naturally handle observational errors.

To fold observational errors into our model, we define $\epsilon_{\rm S}$ and $\epsilon_{\rm L}$ as the uncertainties on the SFR and cluster luminosities, respectively. Both of these errors are dominated by systematic effects that are highly uncertain. For cluster luminosities, the dominant errors arise from the need to extrapolate the cluster profile to large radii in order to assign a total luminosity (Larsen & Richtler 1999). These can lead to an approximately half magnitude of error. Similarly, depending on the choice of star formation tracer, observational estimates of the SFR are subject to uncertainties arising from dust extinction, ionizing photon escape, the choice of stellar IMF and ambiguities in the choice of time-scale over which the SFR is averaged, among others – see Kennicutt & Evans (2012) for a recent review. Typical errors are again $\sim$0.5 dex.

Despite the fact that these errors are systematic and non-Gaussian, we make a simplistic assumption that they can nevertheless be at least roughly approximated as a simple Gaussian blur applied to both the log cluster luminosities and log SFRs. Under this assumption, we can write the distribution of observed luminosities $L_{\rm obs}$ in a galaxy with observed SFR $\dot{M}_{*,\rm obs}$ as simply

$$\begin{aligned}\phi_{k,\rm obs}(\log L_{\rm obs} \mid \log\dot{M}_{*,\rm obs}) = \frac{1}{2\pi\epsilon_{\rm S}\epsilon_{\rm L}}\iint\phi_k(\log L \mid \log\dot{M}_*)\\ \times\exp\left\{-\frac{[\log(\dot{M}_*/\dot{M}_{*,\rm obs})]^2}{2\epsilon_{\rm S}^2} - \frac{\left[\log(L/L_{\rm obs})\right]^2}{2\epsilon_{\rm L}^2}\right\}\\ \times\,{\rm d}\log\dot{M}_*\,{\rm d}\log L, \quad (22)\end{aligned}$$

where $\phi_k(\log L \mid \log\dot{M}_*)$ is computed as described in the preceding sections. The expression for $\Phi_{k,\rm obs}(\log L_{\rm obs})$ is analogous.

2.5 The importance of variable mass-to-light ratios

To demonstrate the effects of variable mass-to-light ratios, dust extinction, and observational uncertainties, and to understand why it is crucial to include these effects in any realistic model, it is helpful to compare the results of our formalism that includes them to simplified formalisms in which these complications are ignored. To this end, we use three different approaches to compute the PDF and CDF of the first order statistics of cluster luminosity, $\Phi_1(L)$ and $\phi_1(L)$, as a function of SFR $\dot{M}_*$. The first calculation uses the full formalism we have just derived. The second uses a simplified formalism in which we adopt a fixed mass-to-light ratio $\Upsilon_{\rm fit}$, and set the luminosity distribution to

$$\phi(L) = \psi(M/\Upsilon_{\rm fit})\Upsilon_{\rm fit} \quad (23)$$

before using equations (5) and (7) to compute $\phi_1(L)$ and $\Phi_1(L)$. We determine the value of $\Upsilon_{\rm fit}$ by performing a least-squares fit to minimize the difference between the median values of L and $\dot{M}_*$ as computed via the two formalisms. This approach amounts to ignoring the scatter in the relationship between cluster luminosity

statistics and SFR induced by the presence of a range of cluster ages, dust and observational uncertainties. The third approach we use is even simpler, but is common in the literature. This is to assert that the expected mass of the most massive cluster $\langle M_1 \rangle$ is such that the expectation value of the mass being in the interval $M_1 - M_{\rm max}$ is unity, i.e.

$$1 = N \int_{\langle M_1 \rangle}^{M_{\rm max}} A M^{\beta} \, {\rm d}M. \tag{24}$$

In this case, we have

$$\langle M_1 \rangle = \left(M_{\rm max}^{\beta+1} - \frac{1}{\langle N \rangle A} \right)^{1/(\beta+1)}, \tag{25}$$

and the expected luminosity of the most massive cluster is then $\langle L \rangle = \langle M_1 \rangle / \Upsilon_{\rm fit}$.

We show the results we obtain from these three methods for two example sets of parameters in Fig. 1; the parameters used for the model in the lower panel are those given in Table 1, while the upper panel is identical except for the value of $M_{\rm max}$. There are several noteworthy points about this figure. First, the 5–95 percentile confidence interval we obtain from the full formalism completely encompasses the confidence interval we obtain by assuming a fixed mass-to-light ratio. This is expected, as the full formalism has other degrees of freedom to explore in the mass-to-light ratio, thus allowing more scatter. On the other hand, the median luminosities we obtain in all three models are very similar except at the highest SFRs. At these large SFRs, models with a fixed mass-to-light ratio predict a stark flattening, while the full model produces a more gradual tapering. This is due to the additional variability from the scatter in mass-to-light ratios that allows a continually increasing range of luminosities. This effect is particularly important for efforts to constrain $M_{\rm max}$ using the observed relationship between $\dot{M}_*$ and the luminosity of the most luminous cluster (Weidner, Kroupa & Larsen 2004; Bastian 2008), and it shows that any such attempt is likely to fail if it does not properly account for variations in mass-to-light ratio. Our conclusion on this point is consistent with that of Bastian (2008), who found using Monte Carlo simulations that adopting a fixed mass-to-light ratio is a poor approximation.

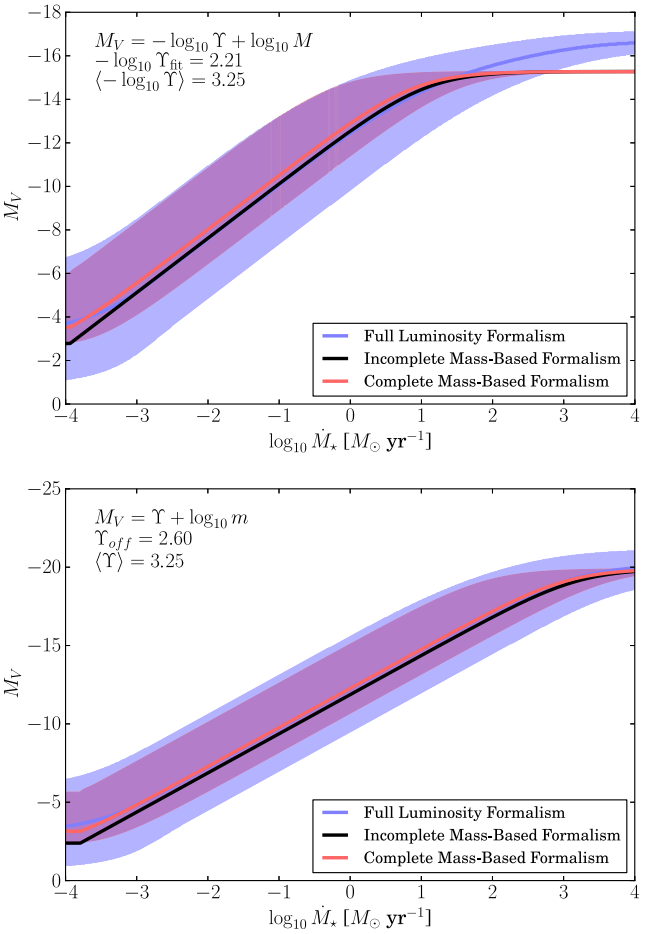

Figure 1. Comparison of the first order statistic $\phi_1(L)$ as a function of SFR $\dot{M}_*$ for differing levels of model complexity. The black lines ('incomplete mass-based formalism') represent the expectation value of the most luminous cluster determined by computing the maximum mass from equation (25) and then applying a fixed mass-to-light ratio $\Upsilon_{\rm fit}$. The red lines ('complete mass-based formalism') show the median, and red bands the 5–95 percentile range, for the first order statistic computed using equations (5) and (7) for $\phi_1(L)$ and $\Phi_1(L)$, but using a fixed mass-to-light ratio to compute the luminosity distribution for individual clusters (equation 23). The blue lines and bands ('full luminosity formalism') show the median and 5–95 percentile range, for $\phi_1(L)$ and $\Phi_1(L)$ computed from the full formalism, using equations (5), (7) and (D11). The parameters used for the computation are given in Table 1 for the lower panel; the upper panel is identical except that it uses $M_{\rm max} = 10^7$ M$_\odot$ rather than 10^9 M$_\odot$. The best-fitting mass-to-light ratio $\Upsilon_{\rm fit}$ is shown in each panel.

3 SOFTWARE IMPLEMENTATION AND VALIDATION

We have implemented the analytic formalism for computing cluster luminosity statistics in a software package called Cluster Luminosity Order-Statistic Code (CLOC), which we have released under the GNU General Public License. The code is available for download at https://code.google.com/p/cluster-cloc/. CLOC takes as inputs the parameters required to compute the cluster luminosity function and its order statistics. To remind the reader, these are the overall SFR $\dot{M}_*$ (which may be given as a single value or, more commonly, a range with the computation to be performed on a grid of $\dot{M}_*$ values), the minimum and maximum cluster ages ($t_{\rm min}$ and $t_{\rm max}$), the parameters of the initial cluster mass function (ICMF; $M_{\rm min}$, $M_{\rm max}$, and β), the minimum and maximum amounts of dust extinction to use (τ_0 and τ_1), the parameters describing the cluster age distribution ($\mathcal{F}_{\rm c}$ and γ), the parameters describing the observational error ($\epsilon_{\rm S}$ and $\epsilon_{\rm L}$) and the parameters describing the time evolution of the light to mass ratio ($\Upsilon_\star$ and ζ). The last two of these depend only on the choice of observational filter and stellar evolution, and so should not be regarded as free parameters. Given these inputs, the code uses the algorithm described in this paper to produce a set of default outputs described in Table 2. The code is implemented in C++, with PYTHON wrappers to call the program and parse the output files. In the remainder of this section, we verify the accuracy of CLOC via comparison to two different Monte Carlo methods, making slightly different assumptions.

3.1 Monte Carlo verification

Our first comparison is to a Monte Carlo calculation that, like CLOC, assumes that cluster formation is a Poisson process, and also computes the mass-to-light ratios of clusters using the same approximate relationship (equation 17). This allows us to verify the accuracy of our analytically calculated PDFs, and our software implementation thereof. Luckily, the Monte Carlo implementation of the model is

Table 1. Fiducial parameter values.

Parameter	Description	Fiducial value
$M_{\rm max}$	Maximum cluster mass	$10^9\,{\rm M}_\odot$
$M_{\rm min}$	Minimum cluster mass	$100\,{\rm M}_\odot$
β	ICMF power law index	-2
γ	Cluster age distribution power law index	-0.9
$t_{\rm min}$	Minimum age of sample	10^7 yr
$t_{\rm max}$	Maximum age of sample	10^9 yr
τ_0	Minimum dust optical depth	0
τ_1	Maximum dust optical depth	1
$\mathcal{F}_{\rm c}$	Fraction of stars in clusters at time of observation	0.01

Table 2. Description of software outputs.

Variable Name	Description
x	the x-array for luminosity arrays in units of ln (ergs^{-1} Hz^{-1})
pdf_l	the PDF of the luminosity of a single cluster $\phi(L)$ before observational uncertainty convolution
pdf_l_obs	the PDF of a single cluster $\phi(L)$ after observational uncertainty convolution
sfr_x	the x-axis for the $\dot{M}_*$–L_1 relation in ${\rm M}_\odot$ yr^{-1}
q5, q50, q95	the 5th, 50th and 95th percentile of the $\Phi_1(L)$ distribution corresponding to sfr_x in the same units as x

Note that the full distributions at each SFR are output and easily obtainable, but these are the only numbers output to this summary data structure.

simple compared with the analytic derivation of its results (although of course *much* slower to run). The process is as follows: we adopt a set of parameters $M_{\rm min} = 500\,{\rm M}_\odot$, $M_{\rm max} = 10^9\ {\rm M}_\odot$, $\beta = -2$, $t_{\rm min} = 10^7$ yr, $t_{\rm max} = 10^9$ yr, $\mathcal{F}_{\rm c} = 0.01$, $\gamma = -0.9$, $\tau_0 = 0$ $\tau_1 = 1$ and $\dot{M}_* = 0.1\ {\rm M}_\odot\ {\rm yr}^{-1}$. From these parameters, we compute the mean cluster mass $\langle M\rangle = 7254\,{\rm M}_\odot$ and the expected number of clusters $\langle N\rangle = 136.5$ from equations (9) and (15). We then create a sample of clusters via the following algorithm.

(i) Draw an actual number of clusters N from a Poisson distribution with expectation value $\langle N\rangle$.

(ii) For each cluster, draw a mass M from the ICMF $\psi(M)$, a light-to-mass ratio from the distribution $\xi([1/\Upsilon])$, and a dust optical depth τ from $\eta(\tau)$. Each of these distributions is determined fully from the input parameters.

(iii) From the drawn values, compute the observed luminosity $L = (M/\Upsilon)e^{-\tau}$.

We repeat this process 10^6 times, to create 10^6 independent cluster samples. From each sample, we record the luminosity L of the most luminous cluster. The code that performs these tasks is included for download with CLOC.

We then run CLOC with the same input parameters, using $\epsilon_{\rm S} = \epsilon_{\rm L} = 0$, i.e. assuming that there is no observational uncertainty on either the SFR or cluster luminosities. We compare the analytically predicted PDF $\phi_1(L)$ of the most luminous cluster as computed by CLOC with the results of the Monte Carlo code in Fig. 2. We see that CLOC exactly predicts the PDF of the Monte Carlo realizations, performing as desired, and with a run time that is far smaller than that of the Monte Carlo code.

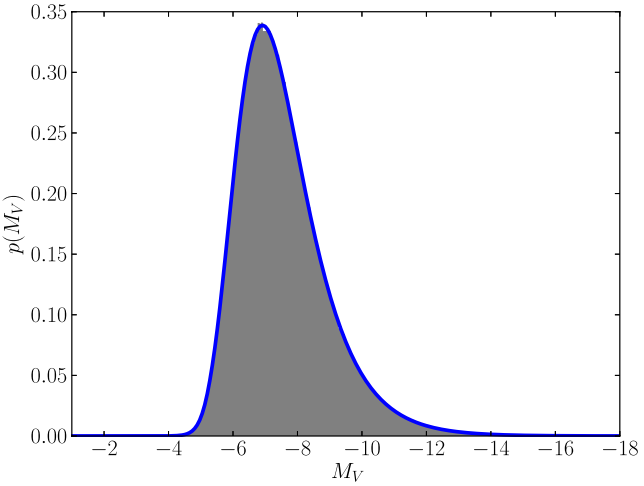

Figure 2. Comparison of the analytic prediction of the PDF of the most luminous cluster as computed by CLOC (blue curve) and the result of 10^6 Monte Carlo realizations of a cluster population (grey histogram). The input parameters for this test are $M_{\rm min} = 500\,{\rm M}_\odot$, $M_{\rm max} = 10^9\ {\rm M}_\odot$, $\beta = -2$, $t_{\rm min} = 10^7$ yr, $t_{\rm max} = 10^9$ yr, $\mathcal{F}_{\rm c} = 0.01$, $\gamma = -0.9$, $\tau_0 = 0$ $\tau_1 = 1$ and $\dot{M}_* = 0.1\ {\rm M}_\odot\ {\rm yr}^{-1}$.

3.2 Comparison to SLUG

We next compare CLOC to the Monte Carlo code SLUG (Fumagalli et al. 2011a; da Silva, Fumagalli & Krumholz 2012). This test is interesting because SLUG and CLOC treat star cluster formation in somewhat different ways, and this difference allows us to check the sensitivity of our predictions to some of the assumptions we made along the way. SLUG is like CLOC in that it produces a population of clusters with a specified ICMF and following a cluster disruption law that produces a specified cluster age distribution, but it differs in two ways. First, SLUG does not assign clusters a fixed, deterministic mass-to-light ratio. Instead, it populates the clusters with individual stars, each of which has an individual mass-to-light ratio determined by stellar evolution models. Thus, SLUG uses the full numerical evolution of the mean mass-to-light ratio as computed from stellar evolution codes, rather than our power-law approximation to it, and also correctly handles the case where the IMF is not fully sampled. The comparison to SLUG enables us to determine where we can no longer safely assume that stochastic variations in mass-to-light

ratio due to imperfect sampling of the IMF is negligible for order statistics at the bright end of the cluster luminosity function.

The second difference between CLOC and SLUG is that SLUG uses a mass-constrained method to sample the cluster mass function, and this method is not precisely described by Poisson statistics. Specifically, when given a time interval $\Delta t_{\rm SLUG}$, an SFR $\dot{M}_*$ and an ICMF, SLUG draws clusters from the ICMF until the total mass of clusters drawn exceeds the target mass $\dot{M}_* \Delta t_{\rm SLUG}$; it keeps the last cluster drawn if the result of doing so is closer to the target mass than the result of omitting this cluster. When the expected mass of stars $\dot{M}_* \Delta t_{\rm SLUG}$ is much larger than the maximum cluster mass $M_{\rm max}$, the distribution of number of clusters should converge to the Poisson distribution we have assumed. At the other extreme, $\dot{M}_* \Delta t_{\rm SLUG} \ll \langle M \rangle$, the SFR ceases to be a well-defined concept. Since stars form in (approximately) discrete events of finite mass, one can only define a meaningful SFR by averaging over time-scales that are long compared to the mean time between events. The behaviour in the intermediate regime, where $\langle M \rangle \ll \dot{M}_* \Delta t_{\rm SLUG} \ll M_{\rm max}$ is more complex, and the distribution of number of clusters, and of star formation history, begin to depend on how one samples from the ICMF. The different assumptions made by SLUG and CLOC in this case will produce somewhat different results. We emphasize that neither code's prescription is necessarily physically correct in the intermediate regime, as the results depend on the real physical details of how galaxies form clusters. Both mass-limited sampling and Poisson sampling are at best reasonable guesses at the right answer. The differences between these two approaches can therefore provide some measure of how accurate *any* method of producing synthetic cluster catalogues can hope to be in the regime where the ICMF is not well sampled.

With this discussion in mind, our procedure for comparing SLUG to CLOC is as follows. As in our previous test, we consider clusters in the age range $t_{\rm min} = 10\,{\rm Myr}$ and $t_{\rm max} = 1\,{\rm Gyr}$. We use the default SLUG prescription for cluster formation disruption, which amounts to $f_{\rm c} = 1$, $t_0 = 1\,{\rm Myr}$ and $\gamma = -1$, and from these values, we compute $\mathcal{F}_{\rm c} = 0.0047$ using equation (16). We then run both CLOC and SLUG for six cases: we use SFRs $\dot{M}_* = 10^{-3}$, 10^{-2} and $10^{-1}\,{\rm M}_\odot\,{\rm yr}^{-1}$, and maximum cluster masses $M_{\rm max} = 10^5$ and $10^9\,{\rm M}_\odot$. We do not use any dust extinction for this test, as SLUG does not include any. All other parameters are as specified in Table 1. For the SLUG runs, we perform 10^3 realizations of the cluster population for each case. We output the cluster population at an age of 100 Myr for the runs with $M_{\rm max} = 10^9\,{\rm M}_\odot$, and at an age of 1 Gyr for the $M_{\rm max} = 10^5\,{\rm M}_\odot$ runs. Note that we do not consider SFRs higher than $10^{-1}\,{\rm M}_\odot\,{\rm yr}^{-1}$ due to issues of computational cost: performing even 1000 SLUG runs at $\dot{M}_* = 10^{-1}\,{\rm M}_\odot\,{\rm yr}^{-1}$ requires of the order of a CPU day, and the computational cost is linear in both the number of realizations and the SFR. However, the vast majority of the observational sample is at higher SFRs, illustrating the difficulty of using Monte Carlo methods to analyse the observations.

We show the results of a comparison between CLOC and SLUG in Fig. 3. We see that the agreement between the two codes is generally quite good, but that there are some important differences. First focus in the left-hand column, showing the models with $M_{\rm max} = 10^5\,{\rm M}_\odot$ and $\Delta t_{\rm SLUG} = 1\,{\rm Gyr}$. These runs are in the regime where $\dot{M}_* \Delta t_{\rm SLUG} \gg M_{\rm max}$, so our assumption that N is Poisson distributed should be safe. Thus, differences between CLOC and SLUG in this column are entirely due to the treatment of mass-to-light ratio in CLOC. At $\dot{M}_* = 10^{-1}$ and $10^{-2}\,{\rm M}_\odot\,{\rm yr}^{-1}$, the difference between the two codes is minimal. However, at $\dot{M}_* = 10^{-3}\,{\rm M}_\odot\,{\rm yr}^{-1}$ we see that the distribution produced by SLUG is noticeably broader than the one computed by CLOC. This difference occurs because CLOC's value for the mass-to-light ratio assumes that each cluster fully samples the IMF, but at very low SFRs the maximum cluster mass is likely to be well below the $\sim 10^{3.5}\,{\rm M}_\odot$ value required for full sampling (Cerviño & Luridiana 2004; Fouesneau et al. 2012). This induces an additional scatter in mass-to-light ratio that is not included in CLOC, and that broadens the distribution. This indicates that CLOC's results for the luminosity distribution of the brightest cluster should not be considered reliable at SFRs below $\sim 10^{-3}$–$10^{-2}\,{\rm M}_\odot\,{\rm yr}^{-1}$, due to its incomplete treatment of IMF sampling effects. Fortunately, at such low SFRs, codes like SLUG are fairly fast to run, since the number of stars involved is small.

Now consider the right-hand column, which uses $M_{\rm max} = 10^9\,{\rm M}_\odot$ and $\Delta t_{\rm SLUG} = 100\,{\rm Myr}$. These runs are in the regime where $\dot{M}_* \Delta t_{\rm SLUG} \gg M_{\rm max}$, and so differences in how the ICMF is sampled begin to be important, on top of IMF sampling effects within clusters. In particular, note that, because the sampling is mass constrained, even though we have set $M_{\rm max} = 10^9\,{\rm M}_\odot$, no cluster of that mass can ever be created in the SLUG runs, because $\dot{M}_* \Delta t_{\rm SLUG} = 10^6$–$10^8\,{\rm M}_\odot$. Thus, even if SLUG does draw a cluster close to $10^9\,{\rm M}_\odot$ from the ICMF, it will reject it on the grounds that a mass of 0 is closer to the target mass than a mass of $10^9\,{\rm M}_\odot$. We have chosen this extreme case intentionally, to show the importance of ICMF sampling effects. At $\dot{M}_* = 10^{-3}\,{\rm M}_\odot$, we find that the SLUG distribution is not only broader than the CLOC one, it is systematically shifted to higher luminosity. The broadening is almost certainly a result of the same effect as in the upper-left panel, i.e. extra scatter in the mass-to-light ratio in SLUG due to incomplete IMF sampling. At $\dot{M}_* = 10^{-2}\,{\rm M}_\odot\,{\rm yr}^{-1}$, the broadening effect has vanished, but the SLUG distribution remains shifted to higher luminosity than the one predicted by CLOC by $\sim$1–2 mag. Only once the SFR reaches $\dot{M}_* = 10^{-1}\,{\rm M}_\odot\,{\rm yr}^{-1}$ do CLOC and SLUG agree well. Clearly, at low SFRs, CLOC's assumption that the number of clusters is Poisson distributed produces fewer massive, luminous clusters than SLUG's mass-limited sampling method, leading to a systematic offset in the first order statistic PDF.

The most important point to take from this comparison exercise is that, as long as one avoids the regime $\dot{M}_* \Delta t_{\rm SLUG} \ll M_{\rm max}$ or $\dot{M}_* \ll 10^{-2}\,{\rm M}_\odot\,{\rm yr}^{-1}$, differences between CLOC and SLUG (and presumably between CLOC and other Monte Carlo codes that behave similarly to SLUG) are negligible. At SFRs below $\sim 10^{-2}\,{\rm M}_\odot\,{\rm yr}^{-1}$, CLOC systematically underestimates the breadth of the cluster luminosity distribution due to its omission of IMF sampling effects. If one wishes to consider models with $\dot{M}_* \Delta t_{\rm SLUG} \ll M_{\rm max}$ the situation is considerably more complicated. In this case, differing choices of exactly how to handle the incompletely sampled ICMF can result in an SFR-dependent offset of $\sim$1–2 mag level in the predicted luminosity PDF. The correct physical answer in this regime is unclear.

4 DISCUSSION OF PARAMETER EFFECTS

Having presented the basic outline of the derivation and the parameters that determine cluster luminosities, we now turn to a study of the effects of varying these parameters, with particular attention to the first order statistic. Our goal is both to demonstrate the power of the analytic formalism, and also to build some intuition to help us interpret observations of the relationship between SFR and brightest cluster luminosity, which we refer to for simplicity as the 'SFR–L_1' relation. This has traditionally been used in an attempt to deduce parameters describing star cluster formation (e.g. Bastian 2008). To that end, we consider a fiducial model whose parameters are given in Table 1, and we then systematically vary the parameters.

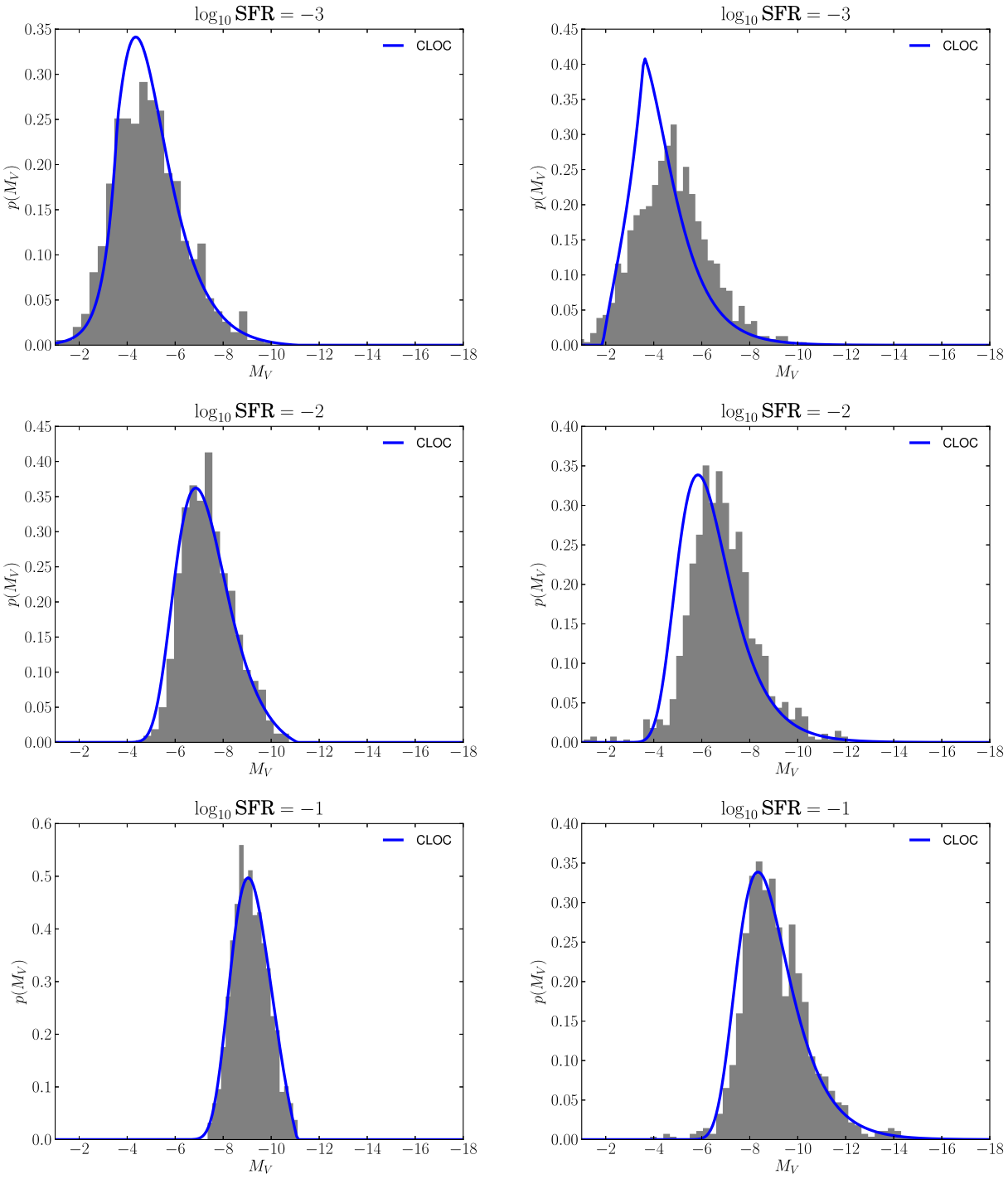

Figure 3. Comparison of the PDFs of the most luminous cluster as computed analytically by CLOC (blue curve) and via Monte Carlo sampling by SLUG (grey histogram), following the procedure outlined in the main text. The left-hand column uses $M_{\rm max} = 10^5$ M$_\odot$ and a run time of 1 Gyr in SLUG, while the right-hand column uses $M_{\rm max} = 10^9$ M$_\odot$ and a run time of 100 Myr. The SFRs used are 10^{-3}, 10^{-2} and 10^{-1} M$_\odot$ yr^{-1}, in the top, middle and bottom rows, respectively.

The results of this experiment are shown in Figs 4 and 5 which we discuss below. Although we focus on the first order statistic here, we note that many of the phenomena we identify are generic, and will affect higher order statistics as well.

4.1 Clustering parameters

Ignoring essentially random effects of dust and light-to-mass ratio, the dominant input shaping the SFR–L_1 diagram is the cluster mass function, which in turn is set by the fraction of stars in clusters $\mathcal{F}_{\rm c}$, and the ICMF parameters $M_{\rm min}$, $M_{\rm max}$ and β. The effects of the first factor, $\mathcal{F}_{\rm c}$, are simple. (We remind the reader again that $\mathcal{F}_{\rm c}$ is not the same as the mass fraction of stars that *form* in clusters or in other gravitationally bound structures; it is the fraction of stars in the observationally selected age range that are in observationally identified clusters today.) Varying $\mathcal{F}_{\rm c}$ simply translates the observed relation left or right, such that $\mathcal{F}_{\rm c} = 0.1$ with a given SFR $\dot{M}_*$ is fully equivalent to having $\mathcal{F}_{\rm c} = 1$ and an SFR of $0.1\dot{M}_*$. This can

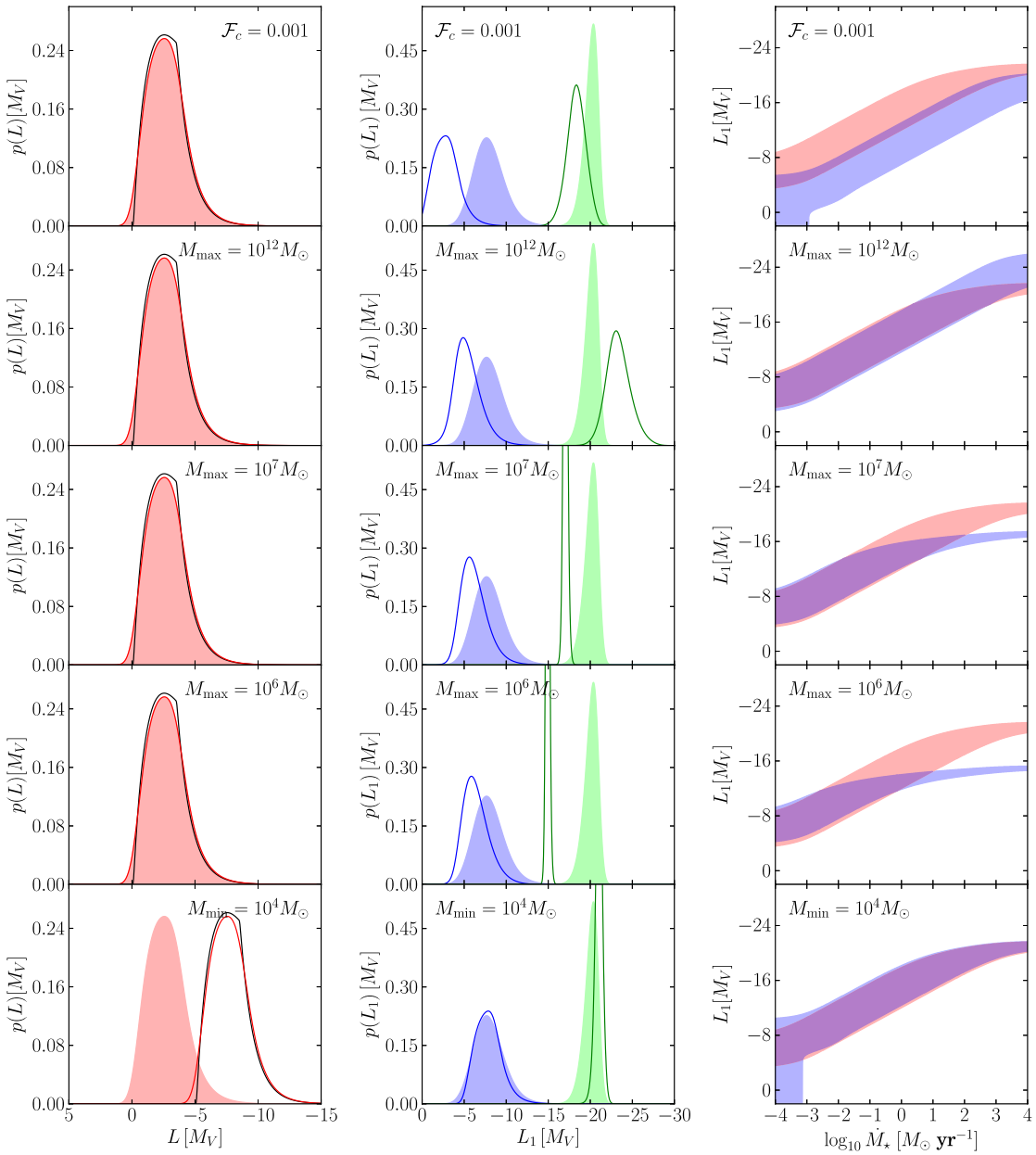

Figure 4. (Left) the PDF of the luminosity of a single cluster before observational error is applied. The filled region corresponds to the fiducial model, while the solid red line denotes the PDF for that result when the change for each corresponding row is applied. The black line corresponds to the model before observational errors are included. (Middle) PDFs of the most luminous cluster for SFRs $\log_{10} \dot{M}_\star = -2.75$ (blue) and 2.75 (green). The filled regions are for the fiducial values and solid lines are for when the change for each corresponding row is applied. (Right) The 5–95 percentile confidence range for the luminosity of the brightest cluster as a function of $\dot{M}_*$. The red region is for the fiducial model and blue is the altered model.

be seen in the first row of Fig. 4. The value of $\mathcal{F}_{\rm c}$ also has another effect. In the upper panel of Fig. 4, notice that 5 per cent confidence contour in the low $\mathcal{F}_{\rm c}$ model extends all the way to the bottom of the plot (and in fact all the way to $-\infty$) at the lowest SFRs. This occurs because, when $\dot{M}_*$ and $\mathcal{F}_{\rm c}$ are both very low, $\langle N \rangle$ is low as well, and there is a reasonable chance that there will be no clusters present at the time of the observation. Thus, for these models the PDF has a significant component at zero luminosity, corresponding to the δ function term in equation (5). When the prefactor on this term, $\Gamma(1, \langle N \rangle)/\Gamma(1)$, exceeds 0.05, the 5 per cent confidence contour must encompass zero luminosity. The value of $\mathcal{F}_{\rm c}$, to which $\langle N \rangle$ is proportional, determines at what SFR this happens.

Behaviour as a result of changing the upper and lower limits on the ICMF is less trivial than the simple translation that results from modifying $\mathcal{F}_{\rm c}$, and is illustrated in the bottom four panels of Fig. 4. For qualitative trends, when all other variables are held constant it is reasonable to treat the $\dot{M}_*$–L_1 relation as behaving like the $\dot{M}_*$–M_1 relation, i.e. the relationship between SFR and the mass of the most massive cluster. This only is reasonable when applied over a relatively small time window Δt where any clusters older than this age are likely to have faded too dramatically to be candidates to be the most luminous. Thus, changing the upper mass cutoff $M_{\rm max}$ has little to no effect until there is a sufficiently high cluster formation rate to make the probability of a cluster near the maximal mass

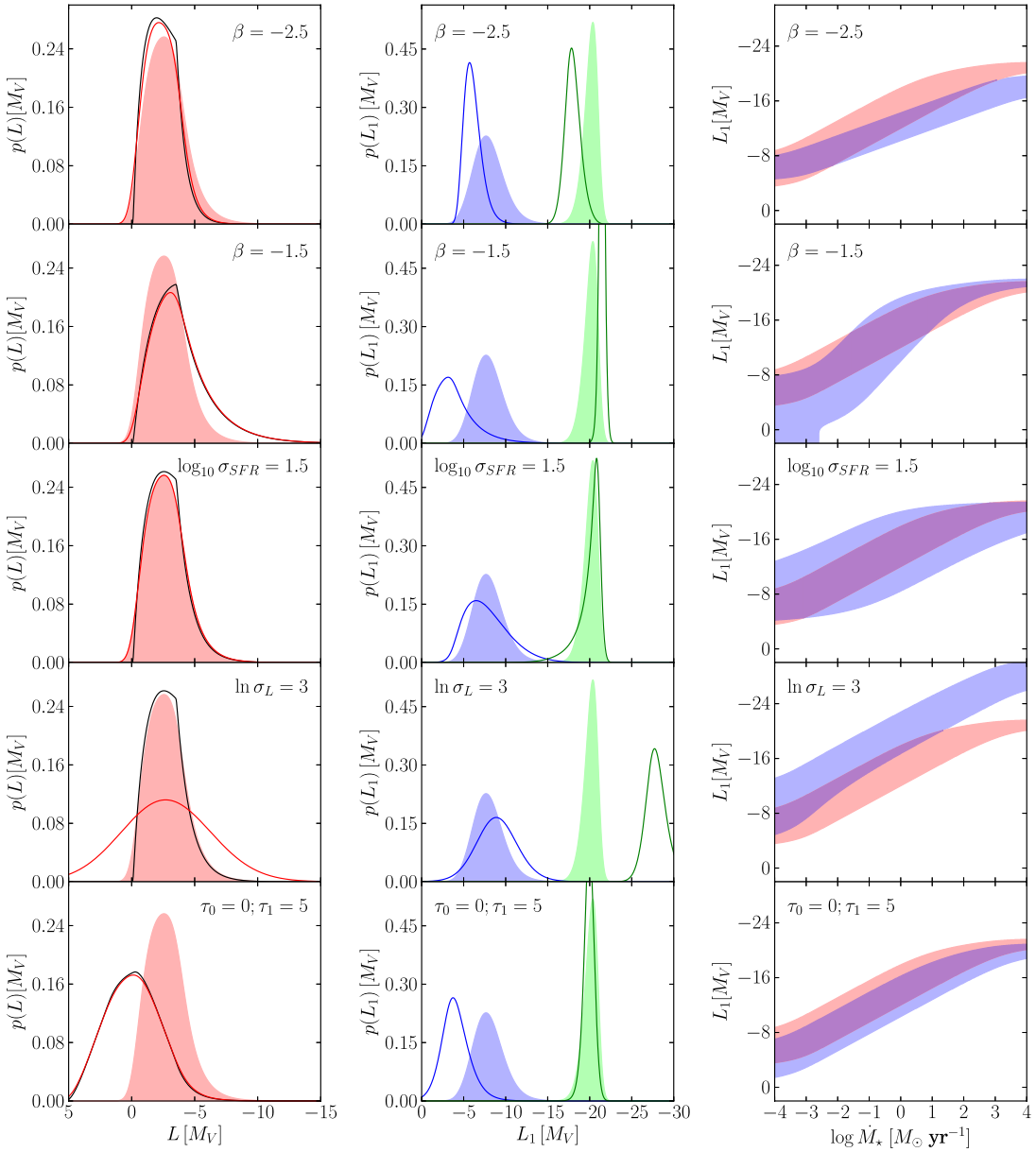

Figure 5. Same as Fig. 4 but with different parameters varied.

forming over a time Δt relatively high. Thus, the effect of changing the maximum cluster mass from $M_{\rm max}$ to a value $M'_{\rm max} < M_{\rm max}$ is to first order to leave the distribution unchanged until the SFR reaches a critical value $\dot{M}_{*,\rm lim}$. We can find this limiting SFR by first noting that we need to expect to find at least one cluster in the range $M'_{\rm max} - M_{\rm max}$ with an age $< \Delta t$ for the change to have a large effect. This condition can be roughly estimated using equation (25) to compute the expected mass of the most massive cluster and equation (9) to compute the expected number of clusters. Combining these two results gives the SFR at which we expect to produce a cluster with a mass of at least $CM_{\rm max}$. This is roughly

$$\dot{M}_{\star,\rm lim} \approx \frac{\langle M\rangle}{\mathcal{F}_{\rm c}\Delta t(\beta+1)}\left[\frac{(M_{\rm min}/M_{\rm max})^{\beta+1}-1}{(M'_{\rm max}/M_{\rm max})^{\beta+1}-1}\right] \approx \frac{\langle M\rangle}{\mathcal{F}_{\rm c}\Delta t(\beta+1)}\left(\frac{M_{\rm min}}{M'_{\rm max}}\right)^{\beta+1}, \quad (26)$$

where in the second step we have assumed that $(M_{\rm min}/M_{\rm max})^{\beta+1} \gg 1$ and $(M'_{\rm max}/M_{\rm max})^{\beta+1} \gg 1$, as is the case for any realistic values of $M_{\rm min}/M_{\rm max}$ and β. Above $\dot{M}_{\star,\rm lim}$, the relation between $\dot{M}_*$ and the luminosity of the brightest cluster is dramatically flattened. One can see an example of this by comparing the second and fourth rows of Fig. 4, which differ only in their values of $M_{\rm max}$.

The effect of the lower mass cutoff $M_{\rm min}$ is twofold (see equation 25). Consider the use of a higher minimum mass $M'_{\rm min}$ compared to a fiducial case with a lower limit of $M_{\rm min}$. First, note that the observed maximum mass cannot be less than $M'_{\rm min}$ and thus in any regime where M_1 is in the range of $M_{\rm min} - M'_{\rm min}$ will have its value set to a floor of approximately $M'_{\rm min}$. This is of course assuming that there is at least one cluster. The other effect of raising $M_{\rm min}$ is to raise the mean cluster mass $\langle M\rangle$. This both decreases the allowed range of cluster masses and thus increases M_1 and L_1, but also decreases the expected number of clusters $\langle N\rangle$, making $N = 0$ a more likely outcome. The net effect is that, at higher SFRs, increasing $M_{\rm min}$ very

slightly increases the expected luminosity of the brightest cluster, but it also raises the SFR at which the 5 per cent confidence contour extends all the way down to zero.

Changing the value of the slope β will affect the observed relation in several ways. The dominant effects are on the overall slope and dispersion of the relation (see the first two rows of Fig. 5). A flatter mass function corresponds to a broader distribution of cluster masses at fixed number of clusters N. As a result, β closer to 0 results in the most massive cluster spanning a wider range of luminosities at fixed SFR. A steeper mass function will result in a less dispersed distribution. The ICMF slope also directly controls how the most luminous cluster varies with N and thus will affect the overall slope of the relation.

Finally, we have already seen that the cluster age distribution plays a key role in setting the total number of extant clusters at the time of observation. However, this role is entirely encoded in the parameter $\mathcal{F}_c$, which provides the mapping between the mass fraction of stars observed to be in clusters at the time of the observation ($\mathcal{F}_c$) and the mass fraction of stars formed in clusters (f_c). At fixed Δt and f_c, changing the cluster age distribution slope γ changes the value of $\mathcal{F}_c$, and thus the expected number of clusters. However, changes in γ are degenerate with changes in f_c that leave the overall value of $\mathcal{F}_c$ the same.

4.2 Dust extinction and observational uncertainties

In addition to the parameters discussed above that characterize the physical way clusters form and evolve, there are additional 'nuisance' parameters that affect the observed luminosity function, and that must be accounted for if we are to have confidence in any deductions we make about the physical parameters. Dust is one such nuisance parameter, though one that is often ignored. As illustrated in the bottom panel of Fig. 5, higher mean dust extinctions lower the median luminosity expected at fixed SFR. Equally importantly, dispersion in the dust optical depth distribution broadens the distribution. In the extreme case of a highly extinct galaxy, the most luminous cluster might well be the one with the lower extinction, rather than the one with the highest intrinsic luminosity. More generally, variations in the mean or width of this distribution can mimic the effects of many other parameters, and the problem is even worse if the amount of dust extinction is systematically correlated with the SFR. Should this be the case, there is little that one can do short of attempting to estimate the extinction of each cluster individually.

A final, also commonly neglected nuisance effect is the uncertainties in the measurements of SFR and cluster luminosity themselves. While the photometric errors are often quite small, significant errors can arise from uncertainties in how to extrapolate the cluster surface brightness distribution to large radii. These effects can introduce scatter of 0.5−1.5 magnitudes (Larsen & Richtler 1999). The SFR also remains significantly uncertain due to scatter in the SFR calibration and even stochastic variations in the SFR indicators [see Kennicutt & Evans (2012) for a recent review]. Robust SFR measurements are best achieved by combining two SFR indicators to capture both obscured and unobscured populations (e.g. UV and IR, or Hα and IR), and in this case the error is likely 0.5 dex or less, but many studies of cluster statistics are based on less accurate SFR measurements. These uncertainties have the effect of broadening the distribution along both the $\dot{M}_*$ and L_1 axes, and must be correctly accounted for when interpreting observations. The effect of varying the assumed uncertainty can be dramatic, as evidenced by rows 3 and 4 of Fig. 5.

5 CONCLUSIONS

In this paper, we present a new analytic method to compute luminosity order statistics of star clusters from theoretical models of the cluster formation process, including realistic parametrized treatments of cluster aging, cluster disruption, dust extinction and observational uncertainties in the determination of both cluster luminosities and galaxy SFRs. We have implemented this analytic method in a new software package, the CLOCcode, which is released under the terms of the GNU General Public License, and we have verified that this package produces results consistent with the full Monte Carlo stellar and cluster population synthesis code SLUG (da Silva et al. 2012) in the regime where the SFR is large enough that the initial stellar and cluster mass functions are well sampled.

The primary advantage of our method compared to previous work is its speed. Monte Carlo methods of computing order statistics of cluster luminosity (i.e. the probability distribution of the most luminous cluster in a population, second most luminous, third most luminous, etc.) are extremely expensive, requiring vast numbers of trials to produce converged distributions. In contrast, because our method is analytic, we are able to obtain the same results in a tiny fraction of the time – for some of the examples we present, the difference in computation time is a matter of days versus milliseconds. The reduction in computational cost that we achieve is such that we can, for the first time, use MCMC methods to explore the full, multidimensional parameter space characterizing the way star clusters form, fade and disrupt, as well the a variety of observational uncertainties that affect measurements of star cluster luminosities and galaxy SFRs. We can therefore conduct statistically rigorous analyses of what can be inferred about the properties of star cluster formation and evolution from observed cluster luminosity distributions, the order statistics thereof, and the dependence of both of these quantities on the large-scale properties of galaxies. The freedom to explore the ways in which nuisance variables confound our attempts to constrain the relevant cluster parameters opens the door for an unprecedented analysis of the relationship between galaxy SFRs and brightest cluster luminosities, which is the subject of the companion paper (da Silva et al., in preparation).

ACKNOWLEDGEMENTS

RLdS and MRK acknowledge support from NASA through Hubble Award Number 13256 issued by the Space Telescope Science Institute, which is operated by the Association of Universities for Research in Astronomy, Inc., under NASA contract NAS 5-26555. The Work of RLdS was supported by the National Science Foundation Graduate Research Fellowship. Support for MF was provided by NASA through Hubble Fellowship grant HF-51305.01-A awarded by the Space Telescope Science Institute, which is operated by the Association of Universities for Research in Astronomy, Inc., for NASA, under contract NAS 5-26555. MRK acknowledges support from an Alfred P. Sloan Fellowship, from NSF grant AST-0955300 and from NASA ATP grant NNX13AB84G. MRK and SMF acknowledge the hospitality of the Aspen Center for Physics, which is supported by NSF grant PHY-1066293. This research has made use of the NASA/IPAC Extragalactic Database (NED) which is operated by the Jet Propulsion Laboratory, California Institute

of Technology, under contract with the National Aeronautics and Space Administration.

REFERENCES

Bastian N., 2008, MNRAS, 390, 759
Bastian N. et al., 2011, MNRAS, 417, L6
Bastian N. et al., 2012a, MNRAS, 419, 2606
Bastian N., Konstantopoulos I. S., Trancho G., Weisz D. R., Larsen S. S., Fouesneau M., Kaschinski C. B., Gieles M., 2012b, A&A, 541, A25
Bik A., Lamers H. J. G. L. M., Bastian N., Panagia N., Romaniello M., 2003, A&A, 397, 473
Boutloukos S. G., Lamers H. J. G. L. M., 2003, MNRAS, 338, 717
Bressert E. et al., 2010, MNRAS, 409, L54
Cerviño M., Luridiana V., 2004, A&A, 413, 145
Chandar R., Fall S. M., Whitmore B. C., 2010, ApJ, 711, 1263
da Silva R. L., Fumagalli M., Krumholz M., 2012, ApJ, 745, 145
Fall S. M., 2006, ApJ, 652, 1129
Fall S. M., Chandar R., 2012, ApJ, 752, 96
Fall S. M., Chandar R., Whitmore B. C., 2005, ApJ, 631, L133
Fall S. M., Chandar R., Whitmore B. C., 2009, ApJ, 704, 453
Fouesneau M., Lançon A., Chandar R., Whitmore B. C., 2012, ApJ, 750, 60
Fumagalli M., da Silva R., Krumholz M., Bigiel F., 2011a, in Treyer M., Wyder T., Neill J., Seibert M., Lee J., eds, ASP Conf. Ser. Vol. 440, UP2010: Have Observations Revealed a Variable Upper End of the Initial Mass Function? Astron. Soc. Pac., San Francisco, p. 155
Fumagalli M., da Silva R. L., Krumholz M. R., 2011b, ApJ, 741, L26
Gieles M., Lamers H. J. G. L. M., Portegies Zwart S. F., 2007, ApJ, 668, 268
Gutermuth R. A., Pipher J. L., Megeath S. T., Myers P. C., Allen L. E., Allen T. S., 2011, ApJ, 739, 84
Haas M. R., Anders P., 2010, A&A, 512, A79
Kennicutt R. C., Evans N. J., 2012, ARA&A, 50, 531
Kruijssen J. M. D., Pelupessy F. I., Lamers H. J. G. L. M., Portegies Zwart S. F., Bastian N., Icke V., 2012, MNRAS, 421, 1927
Lada C. J., Lada E. A., 2003, ARA&A, 41, 57
Larsen S. S., 2002, AJ, 124, 1393
Larsen S. S., 2009, A&A, 494, 539
Larsen S. S., Richtler T., 1999, A&A, 345, 59
Rose C., Smith M. D., 2002, Mathematical Statistics with Mathematica. Springer-Verlag, New York
Weidner C., Kroupa P., Larsen S. S., 2004, MNRAS, 350, 1503
Zhang Q., Fall S. M., 1999, ApJ, 527, L81

APPENDIX A: GENERAL DERIVATION OF CLUSTER ORDER STATISTICS

Here we derive the order statistics for clusters using a more general method that can be extended to the case where cluster luminosities are independent of one another, but where the full cluster formation process does not obey Poisson statistics. Let $\phi(L)$ be the luminosity PDF of a single cluster and $\Phi(L) = \int_0^L \phi(L')\,\mathrm{d}L'$ be the corresponding CDF. Note that, although we use L as the variable, our derivation applies equally well to any other quantity that is defined for a star cluster, for example mass. First consider a region of study containing exactly N clusters. For independently drawn cluster luminosities, the probability that any single cluster has a luminosity $>L$ is $1-\Phi(L)$, and for $m \leq N$, the probability that exactly m clusters have luminosities $>L$ is simply given by the binomial distribution. Thus, we have

$$P_m(>L) = \begin{cases} \frac{N!}{(N-m)!m!}[1-\Phi(L)]^m \Phi(L)^{N-m}, & m \leq N \\ 0, & m > N. \end{cases} \quad \text{(A1)}$$

To obtain the CDF $\Phi_k(L)$, recall that $\Phi_k(L)$ is the probability that the kth most luminous cluster has a luminosity $\leq L$. If $N \geq k$, this probability must be equal to the probability that there are between 0 and $k-1$ clusters that have luminosities $>L$, and therefore

$$\Phi_k(L \mid N) = \sum_{m=0}^{k-1} P_m(>L). \quad \text{(A2)}$$

Note that for $N < k$, this evaluates to $\Phi_k(L|N) = 1$ for any luminosity where $\Phi(L) \neq 0$. This amounts to asserting that, in a region with $N < k$ clusters, the CDF of the kth most luminous cluster is 1 for any non-zero value of L. For $N \geq k$, the corresponding PDF is

$$\phi_k(L \mid N) = \frac{\mathrm{d}}{\mathrm{d}L}\Phi_k(L \mid N) \quad \text{(A3)}$$

$$= \frac{N!}{(N-k)!(k-1)!}\Phi(L)^{N-k}[1-\Phi(L)]^{k-1}\phi(L). \quad \text{(A4)}$$

Note that the second equality is not immediately obvious, but is a standard result in statistics that can be proven by a variety of arguments (e.g. Rose & Smith 2002, section 9.4).

The case $N < k$ is more subtle, since this amounts to asking what we mean by the PDF $\phi_k(L|N)$ when $N < k$. To put the question in words: what is the probability that the kth most luminous cluster has a luminosity in the range L to $L + \mathrm{d}L$, if we are considering a region where there are fewer than k clusters present? We must answer this question if we are to define a meaningful PDF, because in any sample of galaxies or subgalactic regions, there are likely to be regions that contain no or only a small number of clusters. There are two possible approaches. One could simply exclude cases where $N < k$, and compute statistics in the remaining cases. This would amount to changing the summations below (equations A6 and A7) to run from $N = k$ to ∞ rather than $N = 0$ to ∞. The other option is to assign a luminosity of 0 to the kth most luminous cluster in regions where $N < k$.

While both options are equally valid from the standpoint of statistics, from a practical standpoint the second one is preferable. The difficulty with excluding the case $N < k$ is that, in order to compare a model of this form to observations, we would be required to construct an observational sample in which we exclude regions that contain too few clusters. However, finite observational sensitivity means that we can never count clusters with certainty. In particular, we cannot easily distinguish between the possibilities that there are no clusters present and that there are clusters present, but below our detection limit. For this reason, we could never be certain of successfully constructing an observational sample that is appropriately cleaned of cluster-free regions. In contrast, if we simply assign a luminosity of zero in our formalism when $N < k$, we avoid this complication. In this case, we need make no effort to sort our observational sample into galaxies with and without a large enough number of clusters, and can instead handle cases of non-detections by folding the observational upper limits into our analysis. For this reason, we choose to formally extend the definition of $\phi_k(N)$ to

$$\phi_k(L \mid N) = \begin{cases} \frac{N!}{(N-k)!(k-1)!}\Phi(L)^{N-k}[1-\Phi(L)]^{k-1}\phi(L), & k \leq N \\ \delta(L), & k > N. \end{cases} \quad \text{(A5)}$$

This choice is also consistent with the CDF for the case $N < k$. As noted above, $\Phi_k(L)$ evaluates to unity for $N < k$ and $\Phi(L) \neq 0$, so $(\mathrm{d}/\mathrm{d}L)\Phi_k(L \mid N) = 0$ for any L such that $\Phi(L) \neq 0$. However, for $\phi_k(L|N)$ to be properly normalized, it must have an integral of unity overall L. The definition given by equation (A5) meets these requirements, as for any $N < k$ it gives a zero derivative for any

luminosity L that it is possible for a cluster to have, but also has unit integral over all luminosities.

We are now in a position to compute the PDF $\phi_k(L)$ and CDF $\Phi_k(L)$ for a population of regions with varying numbers of clusters. This is given by the sum of $\phi_k(L|N)$ and $\Phi_k(L|N)$ weighted by the probability $P(N)$ that a given region contains exactly N clusters:

$$\Phi_k(L) = \sum_{N=0}^{\infty} P(N)\Phi_k(L \mid N) \tag{A6}$$

$$\phi_k(L) = \sum_{N=0}^{\infty} P(N)\phi_k(L \mid N). \tag{A7}$$

To proceed further one requires the discrete probability distribution $P(N)$ for the number of clusters. One might guess that $P(N)$ is Poisson distributed, but, as noted in the main text, this cannot be strictly true due to mass conservation. As has been discussed in the context of sampling from the IMF (e.g. Haas & Anders 2010), many other choices are possible that enforce mass conservation to varying degrees. For example, the SLUG code to which we compare in Section 3.2 uses a 'stop-nearest' approach in which clusters are drawn from the cluster mass function until the total mass exceeds the specified mass budget, and then one keeps or does not keep the last cluster drawn based on which choice puts the total mass closest to the target value. Alternately, one could always or never keep the last cluster, which corresponds to ensuring that one always overshoots or undershoots the mass budget, one could produce a list of clusters but then sort them by mass, or any number of other approaches. Clearly, each of these approaches will generate a different distribution of $P(N)$ values. To calculate the order statistics for a given method of sampling the cluster mass function, one must derive $P(N)$ for that approach (either analytically or numerically) and then use that distribution in equations (A6) and (A7).[4]

Without a real physical theory of star cluster formation there is no obvious reason to favour one method of mass-limited sampling over another. However, in the limit where the mean cluster mass is much less than the total gas mass, it is probably reasonable to approximate that clusters form independently of one another, in which case $P(N)$ will be Poisson distributed. In this case, the PDF and CDF of cluster luminosities are

$$\Phi_k(L) = \sum_{N=0}^{\infty} \frac{\langle N\rangle^N e^{-\langle N\rangle}}{N!}\Phi_k(L \mid N) \tag{A8}$$

$$\phi_k(L) = \sum_{N=0}^{\infty} \frac{\langle N\rangle^N e^{-\langle N\rangle}}{N!}\phi_k(L \mid N), \tag{A9}$$

where $\langle N\rangle$ is the expected number of clusters in the region under consideration.

Evaluating equation (A8), we have

$$\Phi_k(L) = \sum_{N=0}^{\infty} \frac{\langle N\rangle^N e^{-\langle N\rangle}}{N!} \sum_{m=0}^{k-1} P_m(>L) \tag{A10}$$

$$= \sum_{m=0}^{k-1}\sum_{N=0}^{\infty} \frac{\langle N\rangle^N e^{-\langle N\rangle}}{N!} P_m(>L) \tag{A11}$$

$$= \sum_{m=0}^{k-1} \frac{\langle N\rangle^m [1-\Phi(L)]^m e^{-\langle N\rangle}}{m!} \sum_{N=m}^{\infty} \frac{\langle N\rangle^{N-m}\Phi(L)^{N-m}}{(N-m)!} \tag{A12}$$

$$= e^{-\langle N\rangle} e^{\langle N\rangle\Phi(L)} \sum_{m=0}^{k-1} \frac{\{\langle N\rangle[1-\Phi(L)]\}^m}{m!}. \tag{A13}$$

Note that in the second line we exchanged the order of summation, which is possible because the sums involved are absolutely convergent. The third line is simply a substitution using equation (A1), and the fourth line follows from the definition of the exponential function. The final line is equation (7) of the main text.

Similarly, evaluating equation (A9) gives

$$\phi_k(L) = \sum_{N=0}^{k-1} \frac{\langle N\rangle^N e^{-\langle N\rangle}}{N!}\delta(L) + \sum_{N=k}^{\infty} \frac{\langle N\rangle^N e^{-\langle N\rangle}}{(N-k)!(k-1)!} \times \Phi(L)^{N-k}[1-\Phi(L)]^{k-1}\phi(L) \tag{A14}$$

$$= \frac{\Gamma(k,\langle N\rangle)}{\Gamma(k)}\delta(L) + \frac{\langle N\rangle^k e^{-\langle N\rangle}}{(k-1)!}[1-\Phi(L)]^{k-1}\phi(L) \times \sum_{N=k}^{\infty} \frac{\langle N\rangle^{N-k}\Phi(L)^{N-k}}{(N-k)!} \tag{A15}$$

$$= \frac{\Gamma(k,\langle N\rangle)}{\Gamma(k)}\delta(L) + \langle N\rangle \frac{\{\langle N\rangle[(1-\Phi(L)]\}^{k-1}}{(k-1)!} e^{-\langle N\rangle} e^{\langle N\rangle\Phi(L)}. \tag{A16}$$

The first line follows from substituting equation (A5) into equation (A9), the second is just an algebraic re-arrangement, and the final line uses the definition of the exponential function. This is equation (5) of the main text.

APPENDIX B: FIT FOR Υ

In this appendix, we compute an approximate power-law fit to the age-dependent cluster mass-to-light ratio $\Upsilon(t)$ for V band. We run a STARBURST99 simulation of a simple stellar population with a Kroupa IMF, Padova+AGB stellar tracks, Lej+SMI stellar atmospheres and solar metallicity. From this simulation, we find that at ages $t = 10$ Myr-1 Gyr, the light-to-mass ratio in V bands is well approximated by $\Upsilon(t) = \Upsilon_*(t/10\mathrm{Myr})^{\zeta}$, with best-fitting parameters

$$\zeta = 0.688 \tag{B1}$$

$$\Upsilon_* = 8.3\times10^{-21}\,\mathrm{M}_\odot\,\left(\mathrm{erg\,s^{-1}\,Hz^{-1}}\right)^{-1}. \tag{B2}$$

Fig. B1 shows both the STARBURST99 result and our best-fitting function. The maximum deviation between the fit and the numerical result is 0.07 dex.

APPENDIX C: THE LUMINOSITY FUNCTION FOR VARIABLE AGES

Here, we evaluate the convolution

$$\phi_{\rm in}(\log L_{\rm in}) \propto \psi(\log M) * \xi(-\log\Upsilon) \tag{C1}$$

[4] The stop-nearest method and some others like it present a further complication. For this method, the masses and luminosities of individual clusters are not independent, because the last cluster drawn is much more likely to be kept if its mass is smaller than if it is large. Thus, clusters drawn late in the selection process are not independent of those drawn early. The formalism given here cannot be used in this case.

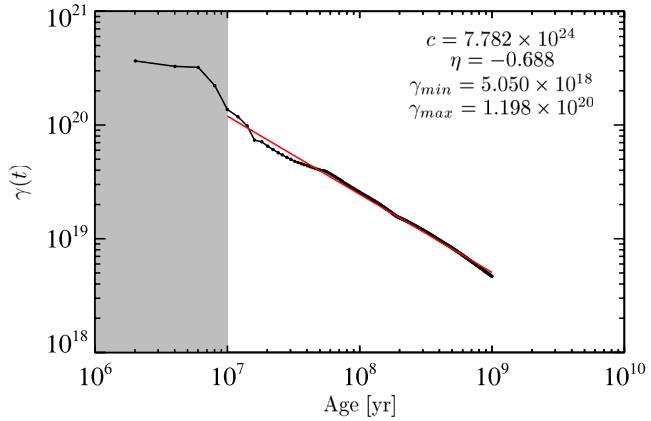

Figure B1. Fit of a simple power-law approximation ($1/\Upsilon \propto t^{-\zeta}$, red) for the light to mass ratio $1/\Upsilon(t)$, compared to the results of a STARBURST99 calculation of Υ with a Kroupa IMF, Padova+AGB stellar tracks, Lej+SMI stellar atmospheres and solar metallicity (black). The greyed out region corresponds to populations younger than 10 Myr, for which the fit is poor.

for the PDF of intrinsic luminosities including age-dependent mass-to-light ratios, where $L_{\rm in} = M/\Upsilon$. First we evaluate $\xi([1/\Upsilon])$ using equation (18), which gives

$$\xi([1/\Upsilon]) \propto \begin{cases} \Upsilon^{(\gamma-1-\zeta)/\zeta}, & 1/\Upsilon_0 < 1/\Upsilon < 1/\Upsilon_1 \\ 0, & \text{otherwise}, \end{cases} \tag{C2}$$

where for convenience we have defined $\Upsilon_0 = \Upsilon_*(t_{\rm min}/10{\rm Myr})^\zeta$ and $\Upsilon_1 = \Upsilon_*(t_{\rm max}/10{\rm Myr})^\zeta$. The calculation can be done most easily by transforming to logarithmic variables. Given the above equation for $\xi([1/\Upsilon])$, the PDF of $-\log\Upsilon$ is given by

$$f_1(-\log\Upsilon) \propto \begin{cases} \Upsilon^{(\gamma-1)/\zeta}, & 1/\Upsilon_0 < 1/\Upsilon < 1/\Upsilon_1 \\ 0, & \text{otherwise}. \end{cases} \tag{C3}$$

We can similarly transform the ICMF $\psi(M)$ to a logarithmic variable as

$$f_2(\log M) \propto \begin{cases} M^{\beta+1}, & M_{\rm min} < M < M_{\rm max} \\ 0 & \text{otherwise}. \end{cases} \tag{C4}$$

Since the intrinsic luminosity $L_{\rm in}$ obeys $\log L_{\rm in} = \log M - \log\Upsilon$, we can now find the PDF of $\log L_{\rm in}$ via the substitution $z = -\log\Upsilon$, giving

$$\phi_{\rm in}(\log L_{\rm in}) \propto \int_{-\infty}^{\infty} f_1(z) f_2(\log L_{\rm min} - z)\,{\rm d}z \tag{C5}$$

$$= \int_{-\infty}^{\infty} [\exp(z)]^{(\gamma-1)/\zeta} \left[\exp(\log L_{\rm in} - z)\right]^{\beta+1}\,{\rm d}z \tag{C6}$$

$$\propto \begin{cases} 0, & L_{\rm in} > \frac{M_{\rm max}}{\Upsilon_0} \\ \int_{z_0}^{z_1} G(z)\,{\rm d}z, & \frac{M_{\rm min}}{\Upsilon_1} < L_{\rm in} < \frac{M_{\rm max}}{\Upsilon_0} \\ 0, & L_{\rm in} < \frac{M_{\rm min}}{\Upsilon_1}, \end{cases} \tag{C7}$$

where for convenience we have defined

$$z_0 = \log\max[1/\Upsilon_1, L_{\rm in}/M_{\rm max}] \tag{C8}$$

$$z_1 = \log\min[1/\Upsilon_0, L_{\rm in}/M_{\rm min}] \tag{C9}$$

$$G(z) = [\exp(z)]^{(\gamma-1)/\zeta} \left[\exp(\log L_{\rm in} - z)\right]^{\beta+1} \tag{C10}$$

$$= L_{\rm in}^{\beta+1} \exp(\omega z), \tag{C11}$$

where

$$\omega = \frac{\gamma-1}{\zeta} - \beta - 1. \tag{C12}$$

The integral $\int G(z)\,{\rm d}z$ therefore trivially evaluates to $(L_{\rm in}^{\beta+1}/\omega)\exp(\omega z)$, up to the constant of integration.

We now limit ourselves to considering the case $\log(M_{\rm max}/M_{\rm min}) > \log(\Upsilon_1/\Upsilon_0)$, which amounts to saying that the mass distribution is broad enough that a cluster with the minimum possible mass at the youngest possible age is still dimmer than a cluster at the maximum possible mass and the oldest possible age. Using our fit to V band, $t_{\rm min} = 10\,{\rm Myr}$ and $t_{\rm max} = 1\,{\rm Gyr}$, this is true as long as $\log_{10}(M_{\rm max}/M_{\rm min}) > 1.38$, which is a fairly unrestrictive requirement given the observations imply a far broader range of cluster masses exists. With this assumption $\phi(\log L_{\rm in})$ reduces to

$$\phi(\log L_{\rm in}) \propto \begin{cases} 0, & L_{\rm in} > \frac{M_{\rm max}}{\Upsilon_0} \\ \int_{\log L_{\rm in}/M_{\rm max}}^{\log 1/\Upsilon_0} G(z)\,{\rm d}z, & \frac{M_{\rm max}}{\Upsilon_1} < L_{\rm in} < \frac{M_{\rm max}}{\Upsilon_0} \\ \int_{\log 1/\Upsilon_1}^{\log 1/\Upsilon_0} G(z)\,{\rm d}z, & \frac{M_{\rm min}}{\Upsilon_0} \le L_{\rm in} \le \frac{M_{\rm max}}{\Upsilon_1} \\ \int_{\log 1/\Upsilon_1}^{\log L_{\rm in}/M_{\rm min}} G(z)\,{\rm d}z, & \frac{M_{\rm min}}{\Upsilon_1} < L_{\rm in} < \frac{M_{\rm min}}{\Upsilon_0} \\ 0, & L_{\rm in} < \frac{M_{\rm min}}{\Upsilon_1} \end{cases} \tag{C13}$$

$$\propto L_{\rm in}^{\beta+1} \begin{cases} 0, & L_{\rm in} > \frac{M_{\rm max}}{\Upsilon_0} \\ \Upsilon_0^{-\omega} - \frac{M_{\rm max}^{-\omega}}{L_{\rm in}^{-\omega}}, & \frac{M_{\rm max}}{\Upsilon_1} < L_{\rm in} < \frac{M_{\rm max}}{\Upsilon_0} \\ \Upsilon_0^{-\omega} - \Upsilon_1^{-\omega}, & \frac{M_{\rm min}}{\Upsilon_0} \le L_{\rm in} \le \frac{M_{\rm max}}{\Upsilon_1} \\ \frac{M_{\rm min}^{-\omega}}{L_{\rm in}^{-\omega}} - \Upsilon_1^{-\omega}, & \frac{M_{\rm min}}{\Upsilon_1} < L_{\rm in} < \frac{M_{\rm min}}{\Upsilon_0} \\ 0, & L_{\rm in} < \frac{M_{\rm min}}{\Upsilon_1}. \end{cases} \tag{C14}$$

The corresponding CDF for $\gamma \neq 1$ is

$$\Phi_{\rm in}(\log L_{\rm in}) = B \begin{cases} 1/B, & \left[\frac{M_{\rm max}}{\Upsilon_0}, \infty\right) \\ \frac{L_{\rm in}^{\beta+1} - M_{\rm max}^{\beta+1}/\Upsilon_1^{\beta+1}}{(\beta+1)\Upsilon_0^\omega} - & \\ \quad \frac{L_{\rm in}^{\nu} - M_{\rm max}^{\nu}/\Upsilon_1^{\nu}}{\nu M_{\rm max}^{\omega}} & \\ \quad + B_1 + B_2, & \left[\frac{M_{\rm max}}{\Upsilon_1}, \frac{M_{\rm max}}{\Upsilon_0}\right) \\ \left(\frac{\Upsilon_0^{-\omega} - \Upsilon_1^{-\omega}}{\beta+1}\right) \times & \\ \quad \left(L_{\rm in}^{\beta+1} - M_{\rm min}^{\beta+1}/\Upsilon_0^{\beta+1}\right) & \\ \quad + B_1, & \left[\frac{M_{\rm min}}{\Upsilon_0}, \frac{M_{\rm max}}{\Upsilon_1}\right) \\ \frac{L_{\rm in}^{\nu} - M_{\rm min}^{\nu}/\Upsilon_1^{\nu}}{\nu M_{\rm min}^{\omega}} - & \\ \quad \frac{L_{\rm in}^{\beta+1} - M_{\rm min}^{\beta+1}/\Upsilon_1^{\beta+1}}{(\beta+1)\Upsilon_1^\omega}, & \left[\frac{M_{\rm min}}{\Upsilon_1}, \frac{M_{\rm min}}{\Upsilon_0}\right) \\ 0, & \left[0, \frac{M_{\rm min}}{\Upsilon_1}\right), \end{cases} \tag{C15}$$

where the interval in each case specifies the range in $L_{\rm in}$ over which it applies, and we have defined

$$B_1 = \frac{M_{\rm min}^{\nu}/\Upsilon_0^{\nu} - M_{\rm min}^{\nu}/\Upsilon_1^{\nu}}{\nu M_{\rm min}^{\omega}} - \frac{M_{\rm min}^{\beta+1}/\Upsilon_0^{\beta+1} - M_{\rm min}^{\beta+1}/\Upsilon_1^{\beta+1}}{(\beta+1)\Upsilon_1^{\omega}} \tag{C16}$$

$$B_2 = \left(\frac{\Upsilon_0^{-\omega} - \Upsilon_1^{-\omega}}{\beta+1}\right)\left(M_{\rm max}^{\beta+1}/\Upsilon_1^{\beta+1} - M_{\rm min}^{\beta+1}/\Upsilon_0^{\beta+1}\right) \tag{C17}$$

$$1/B = B_1 + B_2 + \frac{M_{\rm max}^{\beta+1}/\Upsilon_0^{\beta+1} - M_{\rm max}^{\beta+1}/\Upsilon_1^{\beta+1}}{(\beta+1)\Upsilon_0^{\omega}} - \frac{M_{\rm max}^{\nu}/\Upsilon_0^{\nu} - M_{\rm max}^{\nu}/\Upsilon_1^{\nu}}{\nu M_{\rm max}^{\omega}} \tag{C18}$$

$$\nu = \frac{\gamma-1}{\zeta}. \tag{C19}$$

For $\gamma = 1$, we instead have a CDF

$$\Phi_{\rm in}(\log L_{\rm in}) = B' \begin{cases} 1/B', & \left[\frac{M_{\rm max}}{\Upsilon_0}, \infty\right) \\ \frac{L_{\rm in}^{\beta+1} - M_{\rm max}^{\beta+1}/\Upsilon_1^{\beta+1}}{(\beta+1)\Upsilon_0^{\omega}} - \frac{\log(\Upsilon_1 L_{\rm in}/M_{\rm max})}{M_{\rm max}^{\omega}} + B_1' + B_2', & \left[\frac{M_{\rm max}}{\Upsilon_1}, \frac{M_{\rm max}}{\Upsilon_0}\right) \\ \left(\frac{\Upsilon_0^{-\omega} - \Upsilon_1^{-\omega}}{\beta+1}\right) \times \left(L_{\rm in}^{\beta+1} - M_{\rm min}^{\beta+1}/\Upsilon_0^{\beta+1}\right) + B_1' & \left[\frac{M_{\rm min}}{\Upsilon_0}, \frac{M_{\rm max}}{\Upsilon_1}\right) \\ \frac{1}{M_{\rm min}^{\omega}} \log \frac{L_{\rm in}}{M_{\rm min}/\Upsilon_1} - \frac{L_{\rm in}^{\beta+1} - M_{\rm min}^{\beta+1}/\Upsilon_1^{\beta+1}}{(\beta+1)\Upsilon_1^{\omega}}, & \left[\frac{M_{\rm min}}{\Upsilon_1}, \frac{M_{\rm min}}{\Upsilon_0}\right) \\ 0, & \left[0, \frac{M_{\rm min}}{\Upsilon_1}\right), \end{cases} \tag{C20}$$

where

$$B_1' = \frac{1}{M_{\rm min}^{\omega}} \log\frac{\Upsilon_0}{\Upsilon_1} - \left[\frac{M_{\rm min}^{\beta+1}}{(\beta+1)\Upsilon_1^{\omega}}\right]\left[\Upsilon_1^{-(\beta+1)} - \Upsilon_0^{-(\beta+1)}\right], \tag{C21}$$

$$B_2' = \frac{\Upsilon_0^{-\omega} - \Upsilon_1^{-\omega}}{\beta+1}\left(M_{\rm max}^{\beta+1}/\Upsilon_1^{\beta+1} - M_{\rm min}^{\beta+1}/\Upsilon_0^{\beta+1}\right) \tag{C22}$$

$$1/B' = B_1' + B_2' + \left[\frac{M_{\rm max}^{\beta+1}}{(\beta+1)\Upsilon_0^{\omega}}\right]\left[\Upsilon_0^{-(\beta+1)} - \Upsilon_1^{-(\beta+1)}\right] - \frac{1}{M_{\rm max}^{\omega}} \log\frac{\Upsilon_0}{\Upsilon_1}. \tag{C23}$$

APPENDIX D: THE LUMINOSITY FUNCTION FOR VARIABLE AGES AND DUST

In this appendix, we derive the PDF and CDF for clusters including the effects of both variable ages and dust, by evaluating the convolution

$$\phi(\log L) \propto \phi_{\rm in}(\log L_{\rm in}) * \eta(-\tau), \tag{D1}$$

where $\phi_{\rm in}(L_{\rm in})$ is the distribution of intrinsic luminosities given by equation (C14),

$$\eta(\tau) = \frac{\mathbf{1}_{(\tau_0,\tau_1)}(\tau)}{\tau_1 - \tau_0} = \begin{cases} \frac{1}{\tau_1-\tau_0}, & \tau_0 < \tau < \tau_1 \\ 0 & \text{otherwise} \end{cases} \tag{D2}$$

is the distribution of dust optical depths and $L = L_{\rm in}e^{-\tau}$. Here $\mathbf{1}_{(x_0,x_1)}$ is the indicator function, which is unity on the interval (x_0, x_1) and zero elsewhere.

As in Appendix C, we evaluate the PDF of the sum by transforming to logarithmic variables. To simplify the analysis, first note that equation (C14) for $\phi_{\rm in}(\log L_{\rm in})$ can be rewritten as a sum of power laws multiplied by indicator functions:

$$\phi_{\rm in}(\log L_{\rm in}) \propto \sum_{i=1}^{5} C_i L_{\rm in}^{p_i} \mathbf{1}_{(L_{{\rm in},0,i}, L_{{\rm in},1,i})}(L_{\rm in}), \tag{D3}$$

with

$$\boldsymbol{C} = (\Upsilon_0^{-\omega}, -M_{\rm max}^{-\omega}, \Upsilon_0^{\omega} - \Upsilon_1^{-\omega}, M_{\rm min}^{-\omega}, -\Upsilon_1^{\omega}) \tag{D4}$$

$$\boldsymbol{p} = (\beta+1, \beta+1+\omega, \beta+1, \beta+1+\omega, \beta+1) \tag{D5}$$

$$\boldsymbol{L}_{\rm in,0} = \left(\frac{M_{\rm max}}{\Upsilon_1}, \frac{M_{\rm max}}{\Upsilon_1}, \frac{M_{\rm min}}{\Upsilon_0}, \frac{M_{\rm min}}{\Upsilon_1}, \frac{M_{\rm min}}{\Upsilon_1}\right) \tag{D6}$$

$$\boldsymbol{L}_{\rm in,1} = \left(\frac{M_{\rm max}}{\Upsilon_0}, \frac{M_{\rm max}}{\Upsilon_0}, \frac{M_{\rm max}}{\Upsilon_1}, \frac{M_{\rm min}}{\Upsilon_0}, \frac{M_{\rm min}}{\Upsilon_0}\right). \tag{D7}$$

The intrinsic and observed luminosities are related by $\log L = \log L_{\rm in} - \tau$, and we let $z = \log L_{\rm in}$, so the convolution may be written as

$$\phi(\log L) \propto \int_{-\infty}^{\infty} \phi_{\rm in}(z)\eta(z - \log L)\,{\rm d}z \tag{D8}$$

$$= \frac{1}{\tau_1-\tau_0}\sum_{i=1}^{5} C_i \int_{-\infty}^{\infty} e^{p_i z}\mathbf{1}_{(L_{{\rm in},0,i}, L_{{\rm in},1,i})}(e^z)\mathbf{1}_{(\tau_0,\tau_1)}(z - \log L)\,{\rm d}z \tag{D9}$$

$$= \frac{1}{\tau_1-\tau_0}\sum_{i=1}^{5} C_i \int_{-\infty}^{\infty} e^{p_i z}\mathbf{1}_{(\log L_{{\rm in},0,i}, \log L_{{\rm in},1,i})}(z). \mathbf{1}_{(\tau_0+\log L, \tau_1+\log L)}(z)\,{\rm d}z \tag{D10}$$

$$= \frac{1}{\tau_1-\tau_0}\sum_{i=1}^{5} C_i \max\left(\frac{L_{i,1}^{p_i} - L_{i,2}^{p_i}}{p_i}, 0\right), \tag{D11}$$

where

$$L_{0,i} = \max\left(\log L_{{\rm in},0,i}, \tau_0 + \log L\right) \tag{D12}$$

$$L_{1,i} = \min\left(\log L_{{\rm in},1,i}, \tau_1 + \log L\right). \tag{D13}$$

The corresponding CDF is

$$\Phi(\log L) = \int_{-\infty}^{\log L} \phi(\log L')\,{\rm d}\log L'. \tag{D14}$$

We refrain from writing out the full result of this integration, because, although each term is trivial to evaluate, there are a very large number of them thanks to the multiple min and max operators involved. The full expression is included in the CLOC software package.

This paper has been typeset from a TEX/LATEX file prepared by the author.