UNIVERSITY OF CALIFORNIA
SANTA CRUZ

STOCHASTICALLY LIGHTING UP GALAXIES: STATISTICAL IMPLICATIONS OF STELLAR CLUSTERING

A dissertation submitted in partial satisfaction of the
requirements for the degree of

DOCTOR OF PHILOSOPHY

in

ASTRONOMY AND ASTROPHYSICS
with an emphasis in STATISTICS

by

Robert Louis da Silva

March 2014

The Dissertation of Robert Louis da Silva
is approved:

Professor J. Xavier Prochaska, Chair

Professor Mark Krumholz

Doctor Michele Fumagalli

Tyrus Miller
Vice Provost and Dean of Graduate Studies

Table of Contents

List of Figures

List of Tables

Abstract

Stochastically Lighting Up Galaxies:
Statistical Implications of Stellar Clustering

by

Robert Louis da Silva

Stars form discretely. At the very least, they form in units of individual stars. However, their discreteness likely extends to much larger spatially and temporally correlated structures known as star clusters. This discreteness has a profound impact on the light that a population of stars will produce even at fixed star formation rate. Ignoring the effects of this clustering when analyzing observations can lead to significant errors and biases.

This work presents an exploration of the effects of this clustering, the foundation of which is the construction of `SLUG`, a code which Stochastically Lights Up Galaxies. It accounts for the effects of clustering by populating composite stellar populations ("galaxies") one cluster at a time where each cluster is filled by individual stars whose evolution is tracked. This is the first code capable of exploring stochasticity for stellar populations composed of clusters and led to several significant insights in the field. Most notably, the scatter of luminosities due to stochastically placing clusters over the star formation history of a population greatly exceeds the effects of stochastically sampling a population with a stellar initial mass function. This has profound implications for interpretations of star formation rates, deriving initial mass functions, and the star formation rate distribution of the universe.

We also explore the statistics of luminosities of clusters themselves, deriving an analytical method (`CLOC`) for calculating the full distribution of cluster order statistics roughly one billion times faster than a suite of Monte Carlo simulations. This giant leap forward in speed provides the groundwork for a previously impossible robust exploration of the relevant parameter space (e.g. dust opacity distributions, cluster mass function shape and cutoffs, and cluster disruption parameters).

To my wonderful wife,

Melanie,

who kept me sane through grad school.

Acknowledgments

I would like to thank my advisors J. Xavier Prochaska and Mark Krumholz, who have been extremely supportive, welcoming, and (above all) patient. X, who took me as his student what seems like a lifetime ago. We worked together longer than most marriages last and still I feel that if I was his student for a decade longer I would still be learning from him. Mark, so generous with his time. He helped me so much with writing each of these papers. I knew when I decided to leave academia that I wasn't going to get any flak from either one of them. Also, hopefully they read this document. Thank you both for the wonderful opportunity you gave me.

Also a thank you to my first research advisor, David Schlegel, who showed me how awesome Astronomy was, taught me how to code, and put me on the path that led me to where I am today.

I am incredibly grateful to my family.

First my father, whose example of dedication and work ethic is an one I strive to follow. For decades, he woke well before the sun to teach with an enthusiasm and skill that his students (including myself for four years) could only marvel at as he left an indelible impact on our lives and community. Who, after retirement, works even more. Despite initial outward appearances, where I get all of my silliness. Whose personality, I sometimes too closely replicate. Through constant sacrifice, whether it be an hours long commute so we could grow up in a house by the beach instead of an apartment or spoiling us with luxuries he wasn't able to have even when money was tight, he has shown me the depth of what love for a family is.

My mother, who taught me the true meaning of kindness. Her unwavering support throughout my life has been a constant foundation upon which all of my success has been built. In moments of indecision or discouragement, I knew that support was just a phone call away. Waking early every morning to cook a meal for the family that her job prevented her from being able to join, her silent unending willingness to give continues to teach me generosity.

My brother, who is as good an older brother as one could hope for. Teaching me algebra and so many other things. Playing with an annoying and demanding little brother. Always good for a spirited discussion about football or life. And,

most importantly, putting me in my place when needed.

My friends. Brothers-in-arms. Groomsmen in my wedding. Michele Fumagalli, who is coauthor on all the work presented inside this book and member of my committee. Countless hours at the whiteboard talking about science loudly to the dismay of our officemates. Quite possibly the most brilliant man I know with an exceptional ability to synthesize a body of literature to ask interesting questions (mostly that we got excited about until we realized they wouldn't work until some actually did). Sharing an office and advisor for so many years, he is the standard that I always strove to match (yet never succeeded to). A true friend who I can always rely on for juicy stories, brilliant scientific ideas, and to make me smile. Valery Rashkov, who finally made it out! He graduated without his face exploding a single time despite his constant threats to the contrary. We were evicted together from our offices but always kept our sunny dispositions. Between his browsing of wikipedia and google maps, his guidance was invaluable in my transition outside of academia. Even though he hated everything, he found his way into being one of my favorite people. Jacob Arnold, a man so similar in humor, taste, and neuroses that I consider him a brother. Countless sighs shared in exile in Nat. Sci. II. Terrible jokes. Laughing so hard my ribs hurt for days. Organizing the perfect Settlers of Catan Bachelor's party. Being there for each other when one of us drew too much on the whiteboard. Drinking too much root beer. Helping me with my mulligan days. We all finally made it!

Lastly, my wife Melanie. No one knows better than her what these last 5 and a half years have been like. Years filled with stress, long hours, and indecision. She's seen me from 11th to 22nd grade. She supported me more than she should have had to. Of course this thesis is dedicated to you. It had to be. You saw every hour of it. Brought me food on observing runs. Let me buy board games or whatever else I happened to be collecting. Cheered me up when I was down. Motivated me when I was discouraged. Knew when I needed to be distracted and when to be pushed on task. Putting up with a delayed wedding, and all the other delays that a Ph.D. brings. There is no way I could have done this without you. I am so looking forward to the next chapter of our lives together.

I also have to thank the National Science Foundation Graduate Research Fellowship for funding a lot of this research along with the Hubble Space Telescope and the UCSC Cota-Robles Fellowship.

Chapter 1

Introduction

As astronomers, our fundamental tool for discovery and exploration is the telescope. This allows us to see celestial objects through the light that they emit. In the visible wavelengths of light, the universe is dominated by the radiation emitted by stars. Unfortunately, nearly all of these observations are unresolved, meaning that the light from many (sometimes millions, billions, or even trillions) of stars is blended together. In these jumbled systems, we cannot directly infer any physical properties (such as mass, star formation history, or stellar mass distribution). We are stuck having to tease information out of this ensemble of stars.

In order to interpret these systems, we turn to models. To begin, we start with the fundamental unit of galaxies: stars. Stars are believed to be nearly entirely determined by their mass which determines the size, temperature, lifetime, brightness, spectral energy distribution, evolution, and all other physical properties of the star. The metallicity of material that makes up the star has second order effects and, in rare cases, interacting multiples of stars can greatly influence the evolution of the star. However, one can largely think of stars as belonging to a main sequence, a one dimensional family of objects.

We have built up libraries of the trajectories of stars of different masses through time. These tracks follow the physical properties of stars as they evolve. When combined with models for the stellar surfaces that arise as a function of their stellar structure, we can determine the observational properties of these stars.

However, these models only account for the evolution of individual stars. In order to account for a population or stars, we need to know how the stars are

distributed in both mass and age. Surprisingly, the distribution of the initial masses of stars (or initial mass function: IMF) is remarkably consistent everywhere we are capable of directly checking (although many indirect methods report often conflicting variations in the IMF, these variations have never been widely accepted by the community). Massive stars thus have a well-quantified trend of being significantly less common than lighter stars. Once this distribution is chosen, one needs simply to fill a stellar population with stars accordingly. The distribution of ages is significantly more complex as the physical properties and environment of each population can have significant effects on the occurrence of stars being born. Some populations, such as globular clusters, were born nearly simultaneously while others continuously form at a constant rate and still others were marked by dramatic bursty episodes of greatly enhanced star formation normally attributed to considerable events in the dynamical history of these systems (such as gravitational instabilities or galaxy mergers).

Motivated by computational limitations, the standard approach to combine models of single stars to composite populations of an ensemble of stars has been one of approximation. By determining the properties of an average star (weighted by the choice of stellar initial mass function) as a function of its age, one needs simply to scale these luminosity by a factor corresponding to how many stars are present (i.e. the mass of the stellar population). Then one combines these average stars at each age weighted by the star formation history in order determine the luminosity of the population. Thus these methods spread fractional pieces of stars throughout time and create a one-to-one relation between a stellar population's star formation history and its luminosity. These techniques are known as point-mass approximations as they assume that all of the probability for the luminosity resides at a value for each mass.

The errors arise because stars do not form in fractional amounts. Stars form discretely. For example, a small enough population might not have any of the most massive stars. Since the luminosity of stars is extremely sensitive to their mass (with the proportionality of light to mass being highly superlinear), this means that even the lack of a few of the most massive stars can profoundly impact its luminosity. Thus some clusters will be brighter and some dimmer even if they have the same mass. The effects of deviating from a one-to-one relation are called stochastic. These

issues led several groups to develop Monte Carlo methods for building up composite stellar populations. This approach is much more computationally expensive as it requires that every population follows the evolution of each individual star, however it properly accounts for the true distribution of masses within a population. `SLUG`, which is the code that forms the backbone of this thesis, is one such code (presented in chapter 2).

Aside from being extremely fast and having dedicated significant resources to a useful user interface, `SLUG` did something novel that led to it being of significant scientific impact. It allowed for a clustering of stars in time. Observations and theoretical simulations of the turbulent environments in which stars formed strongly indicated that stars form in bursts that are collocated in space and time, but `SLUG` was the first to include these effects and bring attention to them. This clustering, leads to an enhanced stochastic populating of the star formation history with stars (rather than a smooth stream of stars, stars are formed in small episodic bursts). In fact, once clustering was accounted for, stochasticity in populating the star formation history is actually greatly dominant over traditionally considering IMF sampling stochasticity. These realizations have profound impact on the interpretation of stellar properties, most dramatically the star formation rates of observed galaxies (discussed in detail in chapter 3).

`SLUG`, while comparatively fast, was too slow to perform statistical studies of massive populations of stars with varying clustering parameters. Seeking to improve the capability to constrain the clustering properties that formed the key insight of `SLUG`, we turned one of the most well studied star cluster relations, the correlation between star formation rate and most luminous cluster (the $\dot{M}_\star - L_1$ relation). It was prohibitively expensive to consider running `SLUG` models for a scientifically interesting region of parameter space. This was because each realization of the relation required millions of galaxies, each with billions of stars.

Thus we turned to an analytic method that after a few reasonable assumptions, allowed us to leverage properties of the distributions of the luminosities to calculate distributions of the $\dot{M}_\star - L_1$ relation roughly a billion times faster than `SLUG` could do. This formed the basis of the `CLOC` tool that makes up the last chapter of this thesis (presented in chapter 4). This great speed increase, allows an unprecedented exploration of the relevant parameters that affect observations.

Chapter 2

SLUG - Stochastically Lighting Up Galaxies: Methods and Validating Tests

The effects of stochasticity on the luminosities of stellar populations are an often neglected but crucial element for understanding populations in the low mass or low star formation rate regime. To address this issue, we present SLUG, a new code to "Stochastically Light Up Galaxies". SLUG synthesizes stellar populations using a Monte Carlo technique that properly treats stochastic sampling including the effects of clustering, the stellar initial mass function, star formation history, stellar evolution, and cluster disruption. This code produces many useful outputs, such as i) catalogs of star clusters and their properties, such as their stellar initial mass distributions and their photometric properties in a variety of filters, ii) two dimensional histograms of color-magnitude diagrams of every star in the simulation, iii) and the photometric properties of field stars and the integrated photometry of the entire simulated galaxy. After presenting the SLUG algorithm in detail, we validate the code through comparisons with STARBURST99 in the well-sampled regime, and with observed photometry of Milky Way clusters. Finally, we demonstrate the SLUG's capabilities by presenting outputs in the stochastic regime. SLUG is publicly distributed through the website http://sites.google.com/site/runslug/.

2.1 Introduction

Fundamental progress in understanding the properties of galaxies, star clusters and stellar populations comes from the comparisons between the observed and synthetic photometry derived from stellar evolution codes. It has become common practice to infer properties such as star formation rate (SFR), star formation history (SFH), age, metallicity, redshift, and stellar mass from photometry. Despite the limits of theoretical modeling of stellar populations (such as uncertainties with dust, stellar evolution, and the stellar initial mass function [IMF]; see Conroy et al. 2009, 2010; Conroy & Gunn 2010) synthetic libraries have reached a degree of precision that allows accurate estimates of these parameters – although sometimes with degeneracy – in massive galaxies and clusters.

However, observations reveal a higher complexity in lower mass systems where scaling relations which apply to more massive systems cannot be trivially extrapolated (e.g., Lee et al., 2007; Weisz et al., 2008). Moreover, in lower mass systems, the limited number of stars that are present render these systems inconsistent with the predictions of most of the currently available codes for synthetic photometry (such as STARBURST99 [Leitherer et al. SB99; 1999]; PEGASE [Fioc & Rocca-Volmerange 1997]; or GALEV [Kotulla et al. 2009]). This is because these stellar population models predict only the mean luminosities in given bands, assuming many realizations of the populations. In reality, the individual realizations may have significant scatter about these mean luminosities. For this reason, it is safe to compare them to individual observations only when the coeval stellar populations being observed are quite large. Violation of this last condition leads to stochastic variations in the photometric properties, but these codes are not designed to fully capture them.

For example in globular clusters, some of the simplest observed stellar populations, failure to account for sampling effects can lead to a significant error in the estimated contributions of blue horizontal branch and AGB stars to the integrated light. As a result, correct estimates of globular cluster ages and metallicities based on their integrated light are possible only if one correctly accounts for this stochasticity (Brocato et al., 1999; Colucci et al., 2011).

Moreover, in weakly star forming regions, stochastic effects can mimic those

of a varying IMF. Indeed, recent observations in the low SFR regime have led to serious consideration of a varying IMF (Pflamm-Altenburg & Kroupa, 2008; Hoversten & Glazebrook, 2008; Meurer et al., 2009; Lee et al., 2009). However a fully self-consistent model of stochasticity, allowing for a range of parameters such as differing degrees of stellar clustering, metallicities, stellar tracks, input IMFs and initial cluster mass functions (ICMFs), and SFHs has not been available to test the null hypothesis of a non-varying but stochastically sampled IMF.

These considerations apply not only to the dwarf galaxies studied by Lee et al. (2009) but also to the outer regions of galaxies such as XUV disks (Boissier et al., 2007a; Thilker et al., 2007) and outlying HII regions (Werk et al., 2008; Gogarten et al., 2009) where the stochasticity becomes crucial in the interpretation of inferred SFRs and SFHs.

While the number of studies that use Monte Carlo approaches to address problems on scales of clusters and galaxies is growing (e.g., Raimondo et al., 2005; Popescu & Hanson, 2009a; Silva-Villa & Larsen, 2011; Fouesneau & Lançon, 2010; Eldridge, 2012), a general purpose tool to study photometry in clusters and galaxies has not previously been available. To fill this need, we have created SLUG, a code that allows proper study of the stochastic star formation regime at a range of scales from individual star clusters to entire galaxies. SLUG provides a variety of tools for studying the stochastic regime, such as the ability to create catalogs of clusters including their individual IMFs and photometric properties, color-magnitude diagrams (CMDs) of entire galaxies where we keep track of the photometry of every star, as well as integrated photometry of entire composite stellar populations.

This paper, the first of a series, focuses on the methods used in the code along with several tests to demonstrate that we are reliably reproducing observations and predictions from other codes for synthetic photometry. We then demonstrate the use of this code in the stochastic regime. In a companion paper (Fumagalli et al., 2011b), we use SLUG to show that, once random sampling is included, a stochastic non-varying IMF can reproduce the observed variation of the Hα/FUV ratio in dwarf galaxies, without resorting to modifications of the IMF. In a the second paper of the series (da Silva et al in prep.) we will explore in detail the implications of stochastic star formation with clustering. Further work will apply this code to a variety of astrophysical questions, such as understanding SFR calibrations in the

stochastic regime and further study of other claims of IMF variation.

The layout of the paper is as follows: §2.2 presents an introduction to stochasticity and its effects on the luminosity of stellar populations; §2.3 gives a detailed description of the SLUG algorithm; §2.4 discusses various tests of the code; §2.5 shows a presentation of the code's outputs in the stochastic regime; finally; §3.5 summarizes the results.

2.2 What is Stochasticity?

Many astrophysical studies require creation of synthetic photometry of galaxies and other collections of stars in order to compare with observations. In this section we present a discussion of the various effects of stochasticity and the regimes in which they are important.

2.2.1 Coeval Stellar Populations

The standard procedure for calculating the luminosity from a coeval population of stars used by the most popular implementations (such as SB99) is as follows (Tinsley, 1980; Scalo, 1986).

To find the mean luminosity of a coeval population of stars with initial mass M (i.e. mass at birth, before any mass loss due to stellar evolution) in some band β at a time t after formation ($L_{\beta,coeval}(t)$), one simply integrates the luminosity of each star in that band as a function of the initial stellar mass m and time ($L_\beta(m,t)$), weighting by the distribution of initial stellar masses (i.e. the IMF) dN/dm:

$$L_{\beta,coeval}(t) = \left(\frac{M}{M_\odot}\right) \int_{m_{min}}^{m_{max}} L_\beta(m,t) \frac{dN}{dm}\, dm, \tag{2.1}$$

where m_{min} and m_{max} are the minimum and maximum initial stellar masses allowed by the IMF, and we normalize the IMF such that $\int m(dN/dm)\, dm = 1\ M_\odot$. Note that the mean total luminosity simply scales linearly with the initial mass of the stellar population. However, for small stellar populations, any individual set of stars drawn from the given mass distribution may have a luminosity that deviates significantly from the mean. This is because each realization of a given mass M is built up with a different sampling of stellar masses which, due to the non-linear de-

pendence of luminosity on stellar mass, yields a different luminosity. For example, if one realization of a stellar population with a total mass of 20 $M_\odot$ consists of one 20 $M_\odot$ star, its total luminosity will be quite different than if the same population were composed of twenty 1 $M_\odot$ stars. We call this type of stochastic process *sampling* stochasticity. When stochastic sampling is important, the distribution of luminosities can be both very broad and highly asymmetric, and so it becomes important to know the shape of the luminosity distribution as well as its mean.

One of the more significant manifestations of sampling stochasticity is the apparent undersampling of the upper end of the IMF. Since the IMF is steeply declining with increasing stellar mass, it is improbable that a low mass population will contain a massive star. As a result, the IMF in a low mass population with few stars can often appear truncated and less luminous than a fully-sampled assumption would have predicted[1]. When considering young clusters, those that are not well-populated at the upper end of the IMF can appear much less luminous since the luminosity dependence on mass is much steeper than the slope of the IMF, resulting in the majority of the light being produced by the most massive stars.

One can roughly estimate the mass below which this effect is significant by calculating the expectation value of obtaining a star above a given mass. We do so following the formalism of Elmegreen (2000), who find that the total mass (M) required to expect a single star above a mass m is

$$M \sim 3 \times 10^3 \left(\frac{m}{100 M_\odot} \right)^{1.35}. \tag{2.2}$$

This statement is clearly dependent on one's choice of IMF. Elmegreen (2000) uses a Salpeter IMF with a lower limit of 0.3 $M_\odot$ and no upper limit. (However if one imposes an upper limit as done by Weidner & Kroupa 2004 the result does not change significantly.) This result implies that in order to reasonably expect even a single 120 $M_\odot$ star[2], one would need to sample approximately $10^4 M_\odot \equiv M_{trunc}$. Thus, one can ignore this apparent pseudo truncation and other sampling effects

[1]Extremely rare drawings of the IMF at low masses can also produce pseudo top-heavy IMFs which are overly luminous per unit mass.

[2]Due to limitations of stellar evolutionary tracks, this is the highest stellar mass SLUG can model and is a reasonable guess for the highly uncertain absolute stellar mass limit. While some (e.g., Figer, 2005) suggest a value of $\sim 120 - 150 M_\odot$ others (e.g., Crowther et al., 2010) suggest it may be as high as 300 $M_\odot$.

only for *coeval* populations with masses $\gg M_{trunc}$. (Additional reference on the limits of stochastic sampling including stellar evolution, can be found in Cerviño & Luridiana 2006, 2004. For specific considerations to Hα luminosity [one of the features of a stellar population most sensitive to stochasticity], see Cerviño et al. 2003.)

Another manner in which stochastic sampling can manifest in coeval populations is for stars going through particularly short-lived and luminous phases of evolution after they leave the main sequence (e.g., AGB and blue horizontal branch stars Buzzoni, 1989; Lançon & Mouhcine, 2000; Lançon, 2011; Fouesneau & Lançon, 2010). Since these phases are short, only a very narrow range of masses is undergoing one of them at any given time. Thus the exact sampling within that mass range can have a significant impact on the number of stars in that phase. As a result, a non-infinite population of stars can have additional random scatter in luminosity even if $M > M_{trunc}$. This effect is more important in populations with little ongoing star formation relative to their stellar mass (otherwise new stars dominate the photometric properties of the population), at specific ages when these post-main sequence populations contribute significantly to the luminosity of the population (Colucci et al., 2011).

2.2.2 Composite Stellar Populations

In order to characterize a more complicated star formation history, the next step is to integrate over the coeval populations discussed above to find the luminosity of all stars in a given band at a time τ,

$$L_{\beta,total}(\tau) = \int_{-\infty}^{\tau} \frac{\mathrm{SFR}(t)}{M_{\odot}} L_{\beta,coeval}(\tau - t)dt, \tag{2.3}$$

where SFR(t) is the star formation rate as a function of time.

In order to treat such models as represenative, two conditions must be met: (1) each of the summed coeval populations is large enough to ignore the effects of sampling stochasticity and (2) the SFR is continuously sampled as well. These conditions can quickly break down for sufficiently low SFRs.

To illustrate this point, consider a galaxy forming stars at a constant rate. In order for stochastic sampling effects to be negligible within some time interval

dt, there need to be at least M_{trunc} worth of stars formed in that interval. For the SFR to be considered reasonably well sampled, dt must be much smaller than the evolutionary timescales of any of the stars, which are $\approx 10^6$ yr for the massive stars that generally dominate the light in an actively star-forming system. Thus the condition for stochastic sampling effects to be negligible is that

$$dt = \frac{M_{trunc}}{SFR} \ll 10^6 \text{ yr}, \tag{2.4}$$

implying that this condition is met only for SFRs consistently $\gg 10^{-2} M_\odot$ yr$^{-1} \equiv$ SFR$_{temp}$. However, this *temporal* stochasticity is amplified when one considers that stars are believed to be formed in discrete collections known as clusters. As a result, the clumping in time of star formation in clusters can produce stochastic effects even in regions with SFRs higher than SFR$_{temp}$. In this case the characteristic mass in Equation **??** is replaced with a mass characteristic of the clusters being drawn (discussed further in da Silva in prep.; Fumagalli et al. 2011b).

The conditions required to ignore the effects of stochasticity break down in a variety of astrophysical environments such as dwarf galaxies (e.g., Lee et al., 2009), low star formation rate regions in the outskirts of galaxies (e.g., Boissier et al., 2007b; Fumagalli & Gavazzi, 2008; Bigiel et al., 2010), and low surface brightness galaxies (e.g., Boissier et al., 2008).

2.3 Technique

2.3.1 Overview

Here we present a brief overview of the code while we present each step in detail in the subsequent sections.

SLUG simulates star formation according to the scheme presented in Figure 2.1. We create collections of star clusters obeying a user-defined ICMF (which can include a given mass fraction of stars not formed in clusters), SFH, IMF, and choice of stellar evolutionary tracks. This collection defines a "galaxy". We do not currently include any effects of chemical evolution. A description of the parameters that users can vary is provided in Table 2.1 and will be described in the following sections.

These galaxies are built up (§2.3.2) by first drawing the age of the cluster from a distribution defined by the given SFH. The mass of the cluster is drawn from the ICMF. Next, the cluster's mass is then filled up with stars according to an IMF. Each of the stars within the cluster is evolved using a stellar evolutionary track combined with a model spectral energy distribution (SED) to determine a variety of integrated fluxes corresponding to commonly used photometric filters (§2.3.3).

At a given set of time steps, these fluxes are summed over each star cluster. The clusters are then disrupted according to the prescription of Fall et al. (2009a). Disrupted clusters have their fluxes added to a "field" population while surviving clusters have their properties stored individually.

The code repeats these operations until a stellar mass equal to the integral of the provided SFH is created. The run time of the code is roughly

$$\begin{aligned} t_{run} &\sim 4\mathrm{s} + (60\mathrm{s})\,(1 + 6.6[1 - f_c]) \left(\frac{\int_0^{t_{max}} \mathrm{SFR}(t)dt}{10^7 M_\odot}\right) \left(\frac{\mathrm{timesteps}}{5}\right) \\ &\quad \times \left(\frac{\mathrm{No.\ filters}}{2}\right) \left(\frac{\mathrm{cpu\ Speed}}{2.33\ \mathrm{GHz}}\right) \end{aligned} \tag{2.5}$$

where f_c is the fraction of stars in clusters and t_{max} is the maximum run time of the simulation.

The code outputs a variety of files that keep track of the properties of the stars, clusters, and total integrated stellar populations. Table 2.2 provides a short description of each available output file. All outputs are parsed and transformed into binary FITS tables.

The code is open source and written in C++ with wrapping and parsing routines written in IDL. This entire process can be controlled through an IDL graphical user interface (see Figure 2.2) or either the UNIX or IDL command lines. The IDL routines are also available wrapped in packages for use with the IDL virtual machine[3] for those without IDL licenses. SLUG is available for download with an up-to-date manual, visit the SLUG website at http://sites.google.com/site/runslug/.

[3]The IDL virtual machine is freely available from http://www.ittvis.com/language/en-us/productsservices/idl/idlmodules/idlvirtualmachine.aspx

2.3.2 Cluster Creation

Most stars are thought to be born in star clusters (Lada & Lada, 2003) and the distribution of star cluster masses appears to obey a power law distribution, where observations (e.g., Zhang & Fall, 1999a; Lada & Lada, 2003; Fall et al., 2009b; Chandar et al., 2010a) and theory (e.g., Fall et al., 2010) suggest that the index (β) of the power law $dN/dM \propto M^{-\beta}$ is approximately 2. SLUG allows for both clustered and unclustered star formation. The user can choose what fraction of the stellar mass is formed in star clusters. If the code is forming clusters ($f_c > 0$), the ICMF's power law slope as well as its upper and lower bounds can be varied. If unclustered star formation is desired ($f_c = 0$), the stars' masses are drawn individually from an IMF and treated as a disrupted "cluster" of one star for the remainder of the code.

The initial masses of stars are drawn from an IMF. Choices of IMF[4] currently are Chabrier (2003), Kroupa (2001), Salpeter (1955), a user-defined arbitrary power law, and the recently proposed IGIMF (Kroupa & Weidner, 2003; Pflamm-Altenburg & Kroupa, 2008). While the Chabrier, Kroupa, Salpeter, and power law IMFs are implemented as a standard probability density function of stellar masses, the IGIMF has additional features that require different treatment (see Appendix 2.8).

Regardless of the choice of IMF, we draw stars until the total mass of the star cluster is built up. Since the random distribution of stars never exactly equals the mass of the cluster, a question arises as to whether to keep the last star added. This last star increases the mass of the cluster above the cluster mass drawn from the ICMF. We determine whether or not to keep that star in the cluster based on whether keeping the star in makes the total mass of stars closer to the mass drawn from the ICMF than leaving it out[5].

Independent of its mass, the age of the cluster relative to the galaxy is assigned in a probabilistic manner weighted by the SFH (which can be arbitrary) such that the SFH is reproduced on average. This is analogous to how the full IMF is reproduced on average when one combines many realizations of clusters, but individual realizations can be top-heavy or appear truncated. Note that this

[4]IMFs are truncated to $0.08 - 120 M_\odot$ due to the lack of stellar tracks outside that range.

[5]The effects of different sampling methods and their dependence on the ICMF is studied in detail by Haas & Anders (2010a). Our method is identical to their 'stop-nearest' method.

method produces a scatter in the SFHs for even a given "constant" SFR. Thus SLUG's definition of a galaxy with a constant SFR is not a galaxy where the SFR is instantaneously constant at all times [6], but rather a galaxy that produces a mass of stars over a time dt equal to SFR$\times dt$ which is distributed in clusters whose ages are drawn from a uniform distribution. This interpretation of what a SFR is and its implications is discussed in more detail in da Silva et al. (in prep.).

Clusters are born until the total mass of stars formed is equal to the integral of the SFH. As with the problem of populating a cluster with stars, a galaxy will never be filled to exactly its given mass with an integer number of clusters. Therefore we apply the same condition for populating the galaxy as we do for the clusters: we add clusters until we exceed the galaxy mass (defined as the integral of the SFH) and keep the final cluster only if the updated total galaxy mass is closer to the desired value. As a result the average SFR over the entire simulation of a particular galaxy can be higher or lower than the input value. This effect is small for most regimes, but very rare drawings of the ICMF at low SFRs can produce mild departures. We emphasize that this is not the effect of any error associated with the code but rather is the necessary result of our interpretation of what a SFR means. This behavior is analogous to the situation that arises when drawing from the IMF: the total mass of stars drawn will never exactly match the target cluster mass, and for rare drawings it is possible that the actual cluster mass will differ from the target mass by a non-trivial fraction.

We demonstrate the results of this procedure in Figure 2.3. While lower average SFRs tend to produce larger fractional scatter in the instantaneous SFR, significant scatter remains until the SFRs exceed 10 $M_\odot$ yr^{-1}. This scatter is a direct result of the finite size of clusters. This type of "bursty" behavior may be responsible for the observed bursty SFHs of dwarf galaxies (Weisz et al., 2012a). To clarify with an example, consider that a $10^7 M_\odot$ cluster (when averaged over 1 Myr similarly to the curves shown in Figure 2.3), will appear as a deviant peak for all but the highest SFRs, where the contribution of that individual cluster is drowned out by enough other clusters. Of course averaging over a larger time interval (δt) reduces

[6] A constant SFR cannot be instantaneously constant because stars form in discrete units of mass. For example, when a star is born, the instantaneous SFR is infinite. Thus we must turn to a more probabilistic interpretation of the SFR.

the influence of any single cluster, and in the limit of averaging over arbitrarily large intervals the difference between the average and input SFR ($\Delta_{\rm avg-input}$) must approach zero. The rate at which $\Delta_{\rm avg-input}$ decreases as a function of δt is a function of the ICMF.

We note that in this release of the code all stars in a cluster are treated as having identically the same age, which may not be the case (e.g., see Bernasconi & Maeder, 1996). While observations suggest a scatter of a several Myr (Palla & Stahler, 1999; Jeffries, 2007; Hosokawa et al., 2011), the mass dependence of this scatter is unclear. Given these uncertainties, and that the intracluster age scatter is at most a few Myr (typically small compared to the cluster age distribution), we chose to neglect this effect for now.

Example of Cluster Creation Algorithm

To illustrate our procedure for forming stars in clusters, we now give an example. Suppose a user requests a SFH consisting of a constant SFR of 2 $M_\odot$ yr^{-1} for 10^6 yr, with an ICMF restricted to the mass range $10^5 - 10^7$ $M_\odot$. In response, the code starts by determining an age for the first cluster. This age is drawn from a distribution that is exactly equal to the normalized SFH. In this example, the SFR is constant so the SFH is flat and hence the distribution from which the the age is drawn is simply a uniform probability from 0 to 10^6 yr. Suppose the code draws an age of 5×10^5 yr for the first cluster. Once the age has been determined, the code then draws a cluster mass from the ICMF. Suppose this mass turns out to be 1.6×10^6 $M_\odot$. The code then populates that cluster with stars until the total mass of stars is greater than 1.6×10^6 $M_\odot$. Since the total mass of stars formed at this point does not exceed the integral of the SFH ($\int_0^{10^6 \text{ yr}} 2 M_\odot \text{ yr}^{-1}\, dt = 2 \times 10^6$ $M_\odot$), the code draws another cluster. Suppose that this time the draw results in a cluster age of 1×10^5 yr and a cluster mass of 5×10^5 $M_\odot$. At this point the code has created a total mass in clusters that is greater than the integral of the SFR ($1.6 \times 10^6 M_\odot + 5 \times 10^5 M_\odot = 2.1 \times 10^6 M_\odot > 2 \times 10^6$ $M_\odot$). Because the total stellar mass if the code keeps the last cluster ($2.1 \times 10^6 M_\odot$) is closer to the integral of the SFR ($2 \times 10^6 M_\odot$) than if it discards the last cluster (1.6×10^6 $M_\odot$), the code keeps the last cluster. It then fills that cluster from the IMF. At this point the code

terminates, having drawn two clusters of mass 1.6×10^6 $M_\odot$ and 5×10^5 $M_\odot$ and ages of 5×10^5 yr and 1×10^5 yr.

2.3.3 Stellar Tracks, SEDs, and Broad Band Photometry

Given the mass and age of each star, we need to determine its properties for a variety of observables. We use the same algorithms adopted by SB99 (Leitherer et al., 1999; Vázquez & Leitherer, 2005) to create a set of tables over which SLUG interpolates the stellar photometry. These tables are constructed in advance to reduce the run time. The available tracks and SEDs are listed in Table 2.3.

The first step in SLUG is to determine the physical properties of each star. To this end, we make use of a variety of stellar evolutionary models. Modifying the SB99 source code, we were able to obtain the full range of stellar tracks available to SB99 (Padova and Geneva; see Table 2.3). In the future we plan to implement a wider range of stellar tracks including those from Eldridge & Stanway (2009) and the BaSTI library (Pietrinferni et al., 2007; Cordier et al., 2007). We supplement the Geneva tracks with the Padova+AGB tracks for stars in the mass range 0.15-0.8 $M_\odot$. These models provide luminosities, gravities, chemical compositions, and effective temperatures at discrete intervals in the evolution of a discrete number of stellar masses. We then need to map these physical properties to stellar atmospheres in order to estimate the spectral energy distributions of the stars. SLUG allows users to choose from one of five possible SB99 algorithms for modeling the atmospheres. One possible model makes use of the Schmutz (1998) atmospheres which are dependent on the stellar wind model. For these models we implement all four prescriptions of stellar winds available in SB99 (see Table 2.3). It is important to note that the SB99 algorithms match SEDs to tracks with a nearest neighbor approach and not through interpolation. Therefore there can be some mild discreteness in the output SEDs. Future work will include removal of this effect.

With SEDs in hand, we can convolve with filters to determine the photometry of each point in our stellar tracks. For this step we include the effects of nebular continuum (free-free, free-bound, and 2 photon processes) as implemented in SB99, but neglect nebular line emission for this first release of the code. (For a discussion of the importance of nebular continuum for the SEDs, see Reines et al.

2010, Leitherer & Heckman 1995, and Mollá et al. 2009.) The full list of available filters is presented in Table 2.4. We also integrate the SED to determine the bolometric luminosity as well as to calculate $Q(H^0)$, the number of hydrogen ionizing photons emitted per second. One can convert $Q(H^0)$ to Hα luminosity with a simple conversion assuming case B recombination. Following the notation of Osterbrock & Ferland 2006,

$$L_{H\alpha} = (1 - f_{esc})(1 - f_{dust})Q(H^0) \left(\frac{\alpha_{H\alpha}^{eff}}{\alpha_B} \right) h\nu_{H\alpha} \quad (2.6)$$

$$\approx 1.37 \times 10^{-12}(1 - f_{esc})(1 - f_{dust})Q(H^0) \text{ ergs/s} \quad (2.7)$$

where f_{esc} is the escape fraction (poorly constrained but thought to be between 0.05 [Boselli et al. 2009] and 0.4 [Hirashita et al. 2003]) and f_{dust} represents the fraction of of ionizing photons absorbed by dust grains (e.g., see appendix of McKee & Williams, 1997, who suggest a value of 0.37). To better characterize the ionizing luminosity we also keep track of $Q(He^0)$ and $Q(He^1)$ which represent the numbers of ionizing photons in the HeI and HeII continua respectively.

The above steps allow us to create a discrete two-dimensional table for each flux band where one axis represents stellar mass, the other represents time, and the value of the table is the logarithm of the flux in that band at the appropriate mass and time. Our tables are created through use of the isochrone synthesis method such that our results are stable against the numerical issues that arise from a fixed mass approach (Charlot & Bruzual, 1991).

2.3.4 Evaluating the Stellar Properties

To determine the properties of a given star of any mass at any given time, we first determine if the star is still 'alive'. This is done by an interpolation in time to find the minimum mass of a dead star (m_{death}) at a given time according to our stellar evolution models (where we call a star "dead" if it no longer has entries in our stellar tracks). If the star is less massive than m_{death}, we interpolate our model tables to determine the flux in a given filter to a precision of 0.01 dex .

For computational speed, there are a variety of approximations and restrictions we are forced to implement. The current scheme only allows ages up to

1 Gyr for the stellar tracks (to be expanded in later releases of the code). We do not evolve stars less massive than 0.9 $M_\odot$ (a number which can be changed by the user). These stars do not evolve past the main sequence for the current maximum age of the code of 1 Gyr, so these stars are treated as having their zero-age main sequence (ZAMS) properties at all times. Due to limitations of the stellar tracks, we treat the photometric properties of all stars less massive than 0.156 $M_\odot$ identically to those of 0.156 $M_\odot$ stars. For many purposes, more massive stars dominate the light in the bands such that this approximation is reasonable (but caution is advised for redder bands dominated by older populations). The tracks also impose a $120 M_\odot$ upper mass limit on stars.

Currently, we neglect the effects of binary stellar evolution (see Eldridge & Stanway, 2009; van Bever & Vanbeveren, 1998; Dionne & Robert, 2006), which may have an impact on the derived results by producing a bluer population with a reduced number of red supergiants and increased age range of Wolf-Rayet stars.

2.3.5 Cluster Disruption

If the user chooses to form stars in star clusters, we randomly disrupt our clusters in a mass independent way such that $dN/d\tau \propto \tau^{-1}$ (following Fall et al., 2009a). We start cluster disruption 1 Myr after the cluster forms. This results in 90% of star clusters being disrupted for each factor of 10 in age after 1 Myr. We continue to calculate the photometry for stars in disrupted clusters, and we include their contribution in our calculations of the integrated properties of the galaxy, as well as in a set of "field" variables and outputs.

2.4 Validating Tests

In this section we present a variety of tests to validate the outputs of SLUG. For these tests we make use of a set of fiducial parameters presented in Table 2.5 unless otherwise noted [7]. To emphasize that SLUG can be applied at different regimes, we arrange these tests in order of scale starting with individual clusters and then considering integrated properties of entire galaxies in the well-sampled

[7]Since we aim to test SLUG rather than to perform a study of the effects that the multiple parameters have on the luminosity distributions, we choose widely adopted vlaues.

regime.

2.4.1 Photometry of Clusters

To demonstrate that SLUG reproduces properties of observed clusters, we turn to the catalog of young star clusters compiled in Larsen (1999). To reproduce the clusters' photometry we modify our fiducial IMF to extend down to 0.08 $M_{\odot}$ and run a SLUG model with a SFR of $1 M_{\odot}$ yr^{-1} for 500 Myr, evaluated every 10 Myr. Note that the SFR does not directly affect the ICMF or the properties of the clusters, only the number of clusters in existence at a given time. We show the results of this exercise in Figure 2.4 where we find remarkable agreement between the models and the data. As is clear from the figure, we are able to reproduce both the location and spread of most of the observed data. Clusters that fall outside of the locus of the SLUG models can easily be reproduced when one accounts for a modest amount of reddening (see reddening vector).

2.4.2 Cluster Birthline

Another test of the photometry of clusters is to compare their Hα luminosity to their bolometric luminosity. Work by Corbelli et al. (2009) has shown that newly born clusters lie along a birthline in this parameter space. They found that the distribution of star clusters was incompatible with a truncation of the IMF within clusters determined by the mass of the clusters are prescribed by Weidner & Kroupa (2006). In Fig. 2.5 we compare the same models as in Section 2.4.1 (assuming $f_{esc} = 0$ and $f_{dust} = 0$) with those of Corbelli et al. (2009), and find good agreement without altering the IMF. Our theoretical predictions differ slightly in the tilt of the locus of points from those by Corbelli et al. (2009), since we characterize the properties of our stars in a different manner (making use of stellar tracks rather than fitting formulae). To better demonstrate the origin of the birthline we also make use of SLUG's ability to keep track of the IMF of each individual cluster (see bottom panel of Figure 2.5). We group clusters by their location in this diagram and sum all of their IMFs together to produce a mean IMF for each region. Here we can see that the birthline (from left to right) is a sequence of clusters with progressively more well-sampled upper ends of the IMF. Extremely rare deviants exist below the

birthline where more extremely massive ($> 100 M_\odot$) stars are drawn than average. Note that these rare clusters consisting of essentially isolated O stars have been reported in the Milky Way (de Wit et al., 2004, 2005) and the SMC (Oey et al., 2004; Lamb et al., 2010) in numbers consistent with stochastic sampling of the IMF. This prediction of the preferential combinations of IMF realizations lying along the cluster birthline will be testable with future observations and is an example of the predictive power and insight provided by SLUG.

2.4.3 Comparison with SB99

A third obvious comparison for SLUG is SB99 itself. Being widely used, SB99 serves as a benchmark for our code. Indeed, one of the motivations for making use of the same tracks and SED algorithms as SB99 is that our code should be able to exactly reproduce SB99 in the well-sampled regime. To that end we now present a variety of tests where we compare to SB99 to demonstrate to demonstrate that we can reproduce their results in the regime where neither sampling nor temporal stochasticity are important.

To compare the outputs of both SB99 and SLUG, we choose an instantaneous burst of star formation to demonstrate the matching of the codes in both amplitude and time. We run a SB99 model similar to our fiducial model (i.e. IMF slope of -2.35 from 1-120 $M_\odot$, solar metallicity, Padova+AGB tracks, and Lej+Smi SEDs [see Table 2.3 for definitions of these parameters]). To meaningfully compare with SB99 we must choose SLUG input parameters such that we are evaluating a population where SB99's estimates of the mean are a valid model for the total luminosity of the system. We therefore draw a very large instantaneous population of $10^9 M_\odot$. To nullify any possible effects of random truncations due to populating the clusters, we ensure all clusters are very large by modifying the fiducial ICMF to a restricted range ($10^6 - 2 \times 10^6 M_\odot$). Similar results are obtained if we simply turn clustering off. We present the results in Figure 2.6. It is evident that we are accurately able to reproduce SB99 in the well-sampled regime for integrated "galaxy" properties. We match both the amplitude and time evolution in all photometric bands.

This can also be seen by looking at the full SEDs. In Figure 2.7, we present

photometry for all 15 of the flux bands available for SLUG and compare with the spectra and integrated photometry produced by SB99 at a variety of time steps. Again we are able to fully reproduce the photometric properties in the well-sampled regime from FUV to K-band.

In both these tests, SLUG matches SB99 within 0.026 dex for all fluxes at all times.

2.5 Stochasticity in Action

Having demonstrated that SLUG can reproduce realistic clusters as well as reproduce SB99's results, we now present outputs of SLUG in the stochastic regime.

2.5.1 Effects on Coeval Populations

Recent studies (e.g., Dalcanton et al., 2009) have demonstrated the wealth of information that can be obtained using resolved color-magntiude diagrams (CMDs) of stars within a galaxy. For comparison with such studies in the stochastic regime, SLUG produces two-dimensional histograms for the user's choice of filters. Such diagrams allow us to directly characterize the effects of stochasticity in a coeval population. In Figure 2.8, we compare two realizations of CMDs produced by SLUG for a $10^5 M_\odot$ instantaneous burst to the theoretical isochrones from which they are produced. Aside from demonstrating we accurately reproduce the tracks, we are able to see the effects of stochasticity in populating the rapid phases of evolution differently in the two realizations. Note that SLUG is capable of producing such diagrams for any given SFH.

2.5.2 Effects on Composite Populations

While individual clusters of stars can be treated as coeval, larger systems are intrinsically built of composite populations. One of the most basic composite populations one can consider is a galaxy forming stars at a constant star formation rate. As discussed in Section 2.2.2, the value of the SFR will have a significant impact on the effects of stochasticity.

To demonstrate the differences that stochasticity makes, we compare SLUG

realizations to those of a well-sampled SB99 model. In Figure 2.9, we first examine the luminosities for SFRs of 100, 500, and 1000 realizations of 1, 10^{-1}, and 10^{-2} $M_\odot$ yr^{-1} respectively. We use our fiducial values for the ICMF and cluster mass fraction. For each SFR, we show the mean and median of the SLUG runs along with the 5 and 95 percentiles.

First, we note that the agreement between the mean of the SLUG models and the SB99 prediction of the mean again exhibits our ability to reproduce SB99's outputs.

However, one can clearly see an increase in fractional scatter as the SFR decreases. This can be attributed to the more bursty SFHs which are the result of the grouping of age in massive clusters. This scatter appears at higher SFRs than predicted by our naive discussion in Section 2.2.2 as a direct result of the clustering. In fact, nearly all of the scatter seen in Figure 2.9 is a result of the clustering rather than sampling of the IMF in clusters. This is most clearly demonstrated by Figure 2.10 which shows similar simulations but with completely unclustered star formation. The figure presents the results of 100, 500, and 1000 realizations of SFRs of 10^{-2}, 10^{-3}, and 10^{-4} $M_\odot$ yr^{-1} respectively. Without clustering the 10^{-2} $M_\odot$ yr^{-1} models have approximately an order of magnitude less scatter in the log of the luminosity within the 5-95 percentile range. We see that the unclustered stochastic effects behave as predicted in Section 2.2.2 where the fractional scatter is small for SFRs $\sim 10^{-2}$ $M_\odot$ yr^{-1} and quickly increases as the SFR decreases (also discussed in Fumagalli et al., 2011b).

For a demonstration of the effects of clustering on the photometric properties of galaxies, we present the tracks of a random subset of individual stochastic realizations of clustered star formation in Figure 2.11. One can see that the Q(H^0) curves are less uniform than the R luminosity. This is a direct result of the sensitivity of Q(H^0) to the youngest, most massive stars. One can also see that the scatter increases with decreasing SFR as expected. Of note is the evolution of the 4th from the top simulation with SFR=$10^{-2} M_\odot \mathrm{yr}^{-1}$ as it is marked by several significantly large clusters which lead to a very "bursty" SFH. This is to be further discussed in da Silva et al. (in prep.) where we elaborate on the effects of stochastic star formation when one includes clusters.

2.6 Summary

We introduce SLUG, a new code that correctly accounts for the effects of stochasticity (with caveats discussed in the text) by populating galaxies with stars and clusters of stars and then following their evolution using stellar evolutionary tracks. Cluster disruption is taken into account and a variety of outputs are created.

We present a series of tests comparing SLUG to observations and other theoretical predictions. SLUG is able to reproduce the photometric properties of clusters from the Larsen (1999) catalog as well as the Corbelli et al. (2009) birthline. It can also reproduce the results of SB99 in the well-sampled regime.

Finally we present SLUG outputs in the stochastic regime and demonstrate the flexibility of the code to address a variety of astrophysical problems with its variety of possible outputs.

SLUG is a publicly available code, and can be found at http://sites.google.com/site/runslug/.

2.7 Acknowledgements

We would like to thank the anonymous referee for useful comments which improved the clarity of the paper. The work of R. L. dS. is supported by a National Science Foundation Graduate Research Fellowship. R. L. dS. is partially supported by an NSF CAREER grant (AST-0548180). MRK acknowledges support from: an Alfred P. Sloan Fellowship; the National Science Foundation through grants AST-0807739 and CAREER-0955300; and NASA through Astrophysics Theory and Fundamental Physics grant NNX09AK31G and a *Spitzer Space Telescope* theoretical research program grant. We would like to thank J. X. Prochaska for help in reading and providing input on the early stages of this manuscript. We would like to thank F. Bigiel for encouraging us to create SLUG and useful conversations with J. Eldridge, C. Weidner, R. Bernstein, and J. Colucci.

2.8 Appendix: Implementation of IGIMF

The IGIMF theory (Kroupa & Weidner, 2003; Weidner et al., 2010b) is a statement that (1) most (if not all) stars form in clusters, (2) the SFR controls the upper cutoff of the ICMF, and (3) that each cluster's mass changes the upper cutoff of the IMF in that cluster. Thus the distribution of stars in a galaxy is not simply drawn from the IMF, but is the result of of a joint distribution function of stars and star clusters.

We implement the IGIMF following Weidner et al. (2010b). We use the work of Pflamm-Altenburg & Kroupa (2008), Weidner & Kroupa (2005), and Weidner et al. (2004a) to define the maximum cluster mass as

$$M_{ecl,max} = 84793 \left(\frac{\langle \mathrm{SFR} \rangle}{M_\odot \ \mathrm{yr}^{-1}} \right)^{3/4}, \tag{2.8}$$

where $\langle \mathrm{SFR} \rangle$ is the time-average SFR. Thus the SFR affects the upper cut off of the ICMF. We determine the average star formation rate over a time interval defined by the user (fiducially 10^7 yr).

After a cluster mass has been drawn, we must adjust the upper cutoff of the IMF that we use to draw stars for that cluster. The relation between maximum stellar mass and cluster mass ($m_{max} - M_{ecl}$) has been studied by Weidner & Kroupa (2004) and Weidner et al. (2010a). Following their treatment, we solve a system of equations numerically for m_{max} as a function of M_{ecl}.

The first equation is simply a statement that the total cluster mass (M_{ecl}) is the integral of the distribution of masses ($\frac{dN}{dm}$) integrated from the lowest to highest mass star in the cluster:

$$M_{ecl} = \int_{m_{min}}^{m_{max}} m \frac{dN}{dm} dm. \tag{2.9}$$

The next constraint is derived based on the statement that there is only one star in the cluster with mass equal to m_{max}. Their choice of implementation of this

statement is as follows[8]:

$$1 = \int_{m_{max}}^{m_{max,\star}} \frac{dN}{dm} dm \tag{2.10}$$

where $m_{max,\star}$ is the maximum stellar mass possible.

In the specific case of a Kroupa (2001) IMF, these equations reduce to the following (taken from Weidner & Kroupa, 2004).

$$1 = k \left[\left(\frac{m_H}{m_0} \right)^{\alpha_1} \left(\frac{m_0}{m_1} \right)^{\alpha_2} m_1^{\alpha_3} \left(\frac{m_{max,*}^{1-\alpha^3}}{1-\alpha_3} - \frac{m_{max}^{1-\alpha^3}}{1-\alpha_3} \right) \right] \tag{2.11}$$

$$\begin{aligned} \frac{M_{cl}}{k} = & \frac{m_H^{\alpha_0}}{2-\alpha_0}(m_H^{2-\alpha_0} - m_{low}^{2-\alpha_0}) + \frac{m_H^{\alpha_1}}{2-\alpha_1}(m_0^{2-\alpha_1} - m_H^{2-\alpha_1}) \\ & + \frac{\left(\frac{m_H}{m_0}\right)^{\alpha_2} m_0^{\alpha_2}}{2-\alpha_1}(m_1^{2-\alpha_2} - m_0^{2-\alpha_2}) + \frac{\left(\frac{m_H}{m_0}\right)^{\alpha_2} \left(\frac{m_0}{m1}\right)^{\alpha_2} m_1^{\alpha_3}}{2-\alpha_3}(m_{max}^{2-\alpha_3} - m_0^{2-\alpha_3}) \end{aligned} \tag{2.12}$$

where

$$\begin{aligned} \alpha_0 &= +0.30, \quad m_{low} = 0.01 \\ \alpha_1 &= +1.30, \quad m_H = 0.08 \\ \alpha_2 &= +2.30, \quad m_0 = 1.00 \\ \alpha_3 &= +2.35, \quad m_{max,*} = 120 \end{aligned} \tag{2.13}$$

We fit a 6th order polynomial to the numerical solution to find:

$$\log_{10} m_{max} = \sum_{i=0}^{6} a_i (\log_{10} M_{cl})^i \tag{2.14}$$

where a=[1.449, -2.522, 2.055, -0.616, 0.0897, -0.00643, 0.000182].

We then use this upper mass limit to modify the standard Kroupa (2001) IMF to fill in the stars for each cluster. Figure 2.12 demonstrates the result. One can see that we are accurately applying the cutoff to the IMF in the IGIMF.

[8]Cerviño et al. (2010) have pointed out that this expression does not equate to the logical statement mentioned above — it under-predicts the maximum mass in 63% of cases. In fact, this formalism equates rather to the statement that the expectation value of stars in the interval $m_{max} - m_{max,\star}$ is equal to 1. However this is the standard formalism of the IGIMF, so it is the formalism we implement.

Table 2.1. Input Parameters

Parameter	Description
Controlling the Physics	
IMF	stellar initial mass function; can choose Kroupa, Salpeter, Chabrier, IGIMF, or an arbitrary slope
ICMF	initial cluster mass function, can change slope, minimum and maximum mass
Stellar Evolutionary Tracks	library of models used for stellar evolution
Metallicity	metallicity of the stellar population
Stellar Atmosphere	which scheme and models are used for SEDs
Stellar Wind Model[a]	which wind model is used for SEDs
Fraction of stars in clusters	mass fraction of stars formed in clusters
Controlling the Simulation	
Maximum time	how long the simulation is run
SFH	can be arbitrary
Seed	random seed used for simulation
Controlling Output	
Time step	time between code outputs
Fluxes	choose which fluxes to output
Colors	which colors to use for CMDs
CMD output parameters	choice of number of bins and

Table 2.1 (cont'd)

Parameter	Description
	range of color and luminosity for each CMD
Cluster output?	set to print output for each cluster
IMF output?	set to output IMF histograms for each cluster

[a] only applicable for Schmutz (1998) O star atmospheres in Lejeune+Sch models (see Table 2.3)

Table 2.2. SLUG Output Files

Name	Description
Histogram	a 2d histogram of the user's choice of color-magnitude diagram(s)
Cluster	of every star in the "galaxy" at each timestep mass, fluxes most massive star born in cluster, number of stars formed in cluster, and age of each undisrupted cluster at each timestep
IMF	a histogram of the IMF of each cluster that appears in the Cluster file
Integral	the total flux of the entire "galaxy" at each timestep
Miscellaneous	the total stellar mass actually formed, as well as the actual SFH and ICMF of the simulation

Table 2.3. Stellar Properties

Parameter	Allowed Values
Tracks	Geneva STD[a], Geneva High[b], Padova STD[c], Padova AGB[d]
Metallicity[e]	Geneva: 0.001, 0.004, 0.008, 0.020, 0.040
	Padova: 0.0004, 0.004, 0.008, 0.020, 0.50
SEDs	Planck[f], Lejeune[g], Lejeune+Sch[h], Lejeune+Smi[i], Pau+Smi[j]
Wind Models[k]	Maeder[l], Empirical[m], Theoretical[m], Elson[n]

[a]Charbonnel et al. (1996, 1999)

[b]same a, but for high mass stars use higher mass loss rate models from Meynet et al. (1994)

[c]Fagotto et al. (1994b,a) and references therein

[d]same as c except use Padova+AGB implementation from Vázquez & Leitherer (2005)

[e]solar is 0.20

[f]simple blackbody SED

[g]Lejeune et al. (1997, 1998)

[h]same as g, but for O stars use Schmutz (1998)

[i]same as g, but for O stars use the Smith et al. (2002) implementation of the Hillier & Miller (1998)

[j]same as i, but also include the Smith et al. (2002) implementation of Pauldrach et al. (2001) atmospheres

[k]only relevant when using Schmutz (1998) atmosheres with Lejeune+Sch models

[l]de Jager et al. (1988); Maeder & Meynet (1987); Maeder (1990)

[m]Leitherer et al. (1992)

[n]Elson et al. (1989)

Table 2.4. Broad Band Filters

Filter	Reference
NUV	1
FUV	1
u	2
g	2
r	2
i	2
z	2
J	3
H	3
K	3
U	4
B	4
V	4
R	4
I	4
$Q(H^0)$	5
$Q(He^0)$	5
$Q(He^1)$	5
L_{bol}	6

[1]GALEX; Morrissey et al. (2005)

[2]SDSS; Fukugita et al. (1996)

[3]Skrutskie et al. (2006)

[4]Johnson-Cousins; Appenzeller et al. (1998)

[5]Obtained by integrating SED blueward of 912, 504, and 208 Å for $Q(H^0)$, $Q(He^0)$, $Q(He^1)$ respectively.

[6]Given by stellar evolutionary tracks.

Table 2.5. Fiducial Inputs

Parameter	Fiducial Value
Time step	10^6 yr
Maximum time	10^9 yr
IMF	1-120$M_\odot$; slope=-2.35
ICMF	$20 - 10^7 M_\odot$; slope=-2
Stellar Evolutionary Tracks	Padova+AGB
Metallicity	Solar; $Z = 0.20$
Stellar Atmosphere	Lej+Smi[a]
Fraction of stars in clusters	100%

[a]While the preferred SEDs for SB99 are the Pau+Smi atmospheres, we find that the Pauldrach models are far too discrete. Therefore while we provide the Pau+Smi atmospheres, we recommend the Lej+Smi.

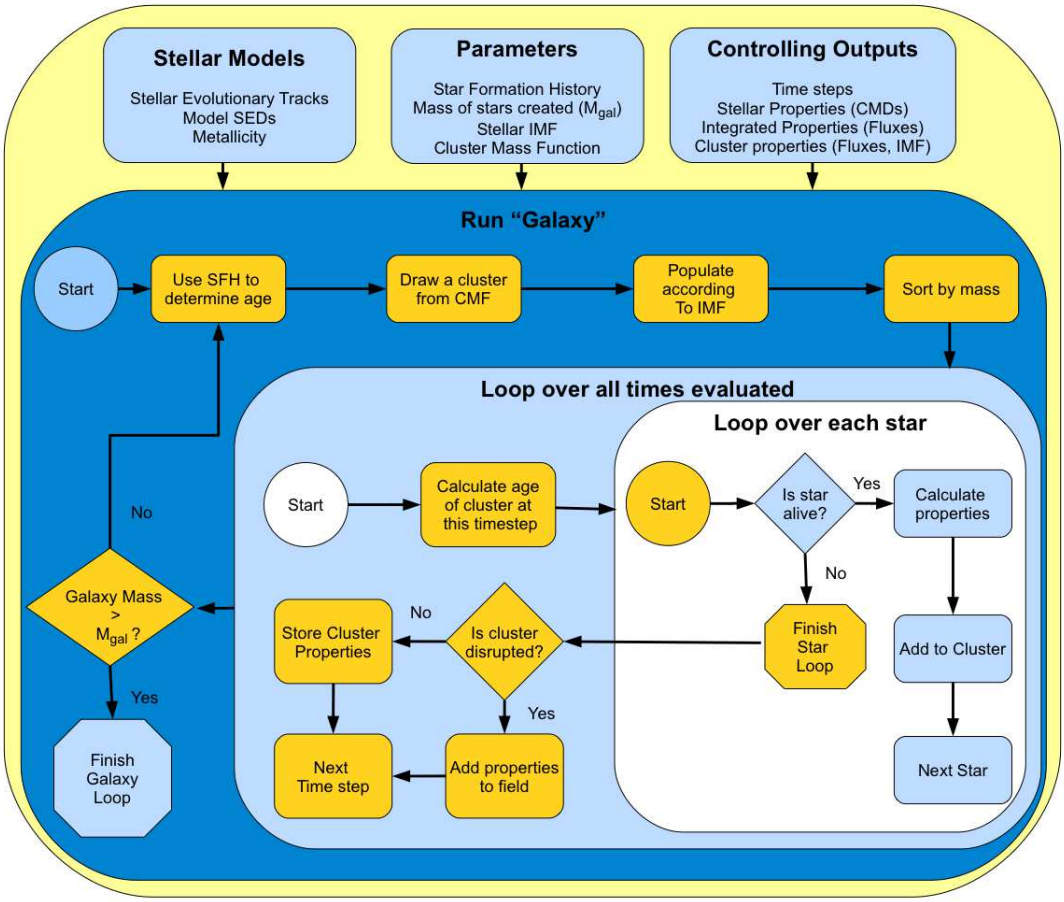

Figure 2.1 A schematic flow-chart describing the algorithm of the the SLUG code. Note that for the case of unclustered star formation, the cluster mass is drawn from the IMF and the population step is skipped as the single star is treated as part of a disrupted cluster for the remainder of the code. Note this is updated from Fumagalli et al. (2010).

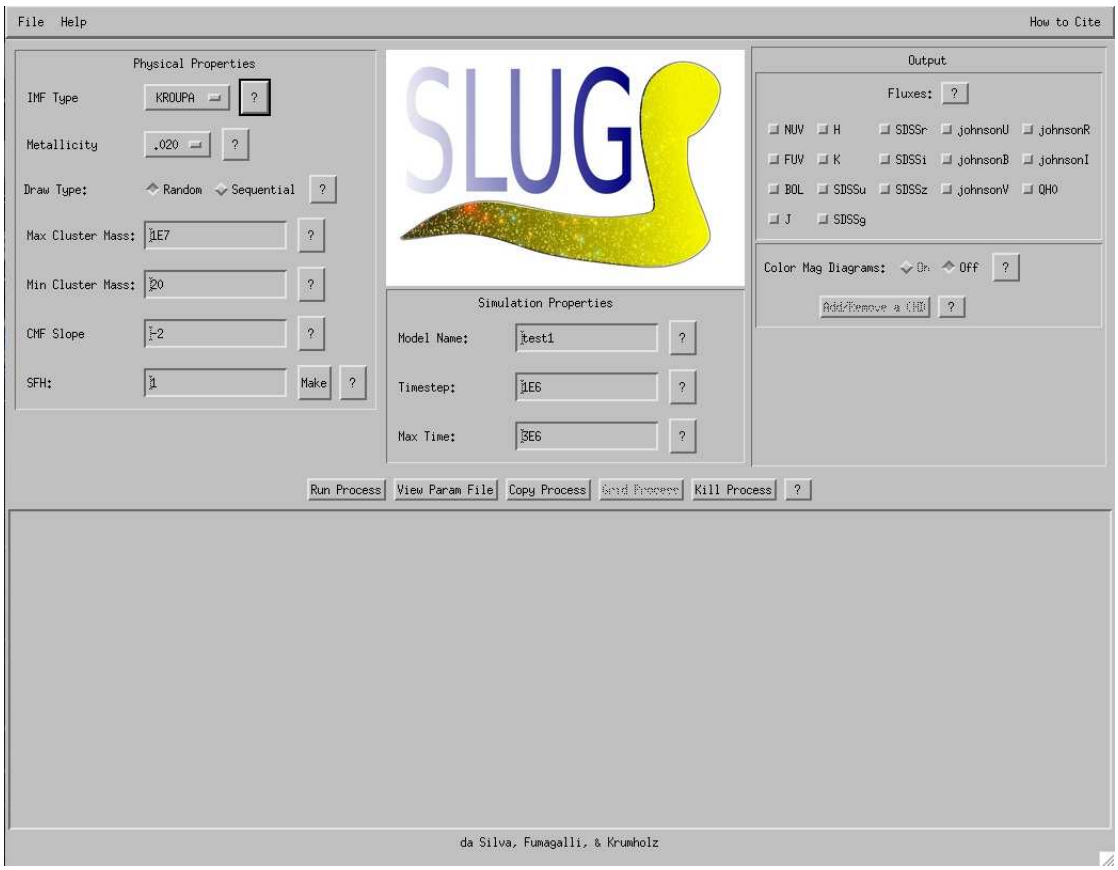

Figure 2.2 IDL GUI interface for running the code. The code may also be called via the UNIX or IDL command lines.

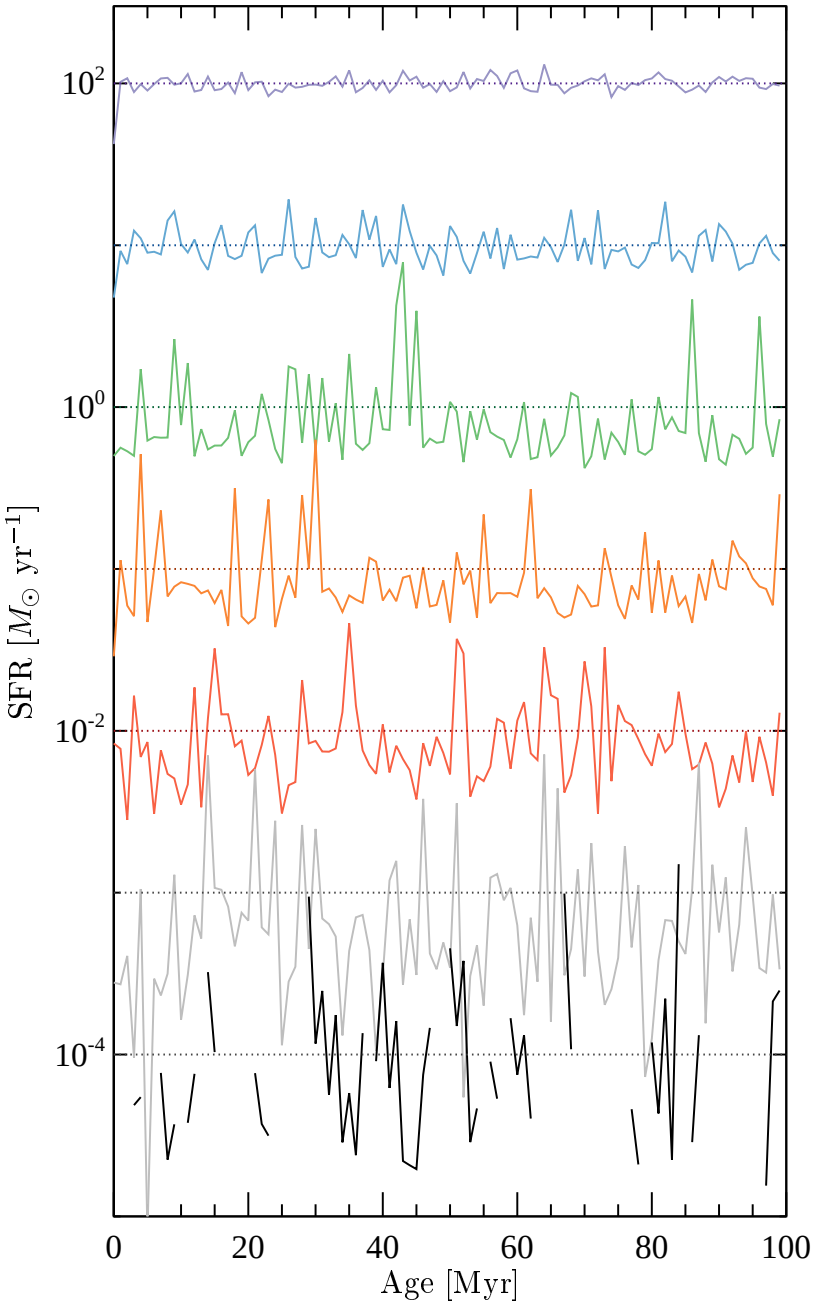

Figure 2.3 Examples of star formation histories averaged over 1 Myr bins for simulations with varying input constant SFRs of 0.0001–100 $M_\odot$ yr^{-1}. The dotted lines show the input SFR. The average SFR of the simulation in each case is within 2, 0.2, and <0.02 percent of the input for 10^{-4},10^{-3}, and $> 10^{-2}$ $M_\odot$ yr^{-1} respectively. SFRs of zero are masked.

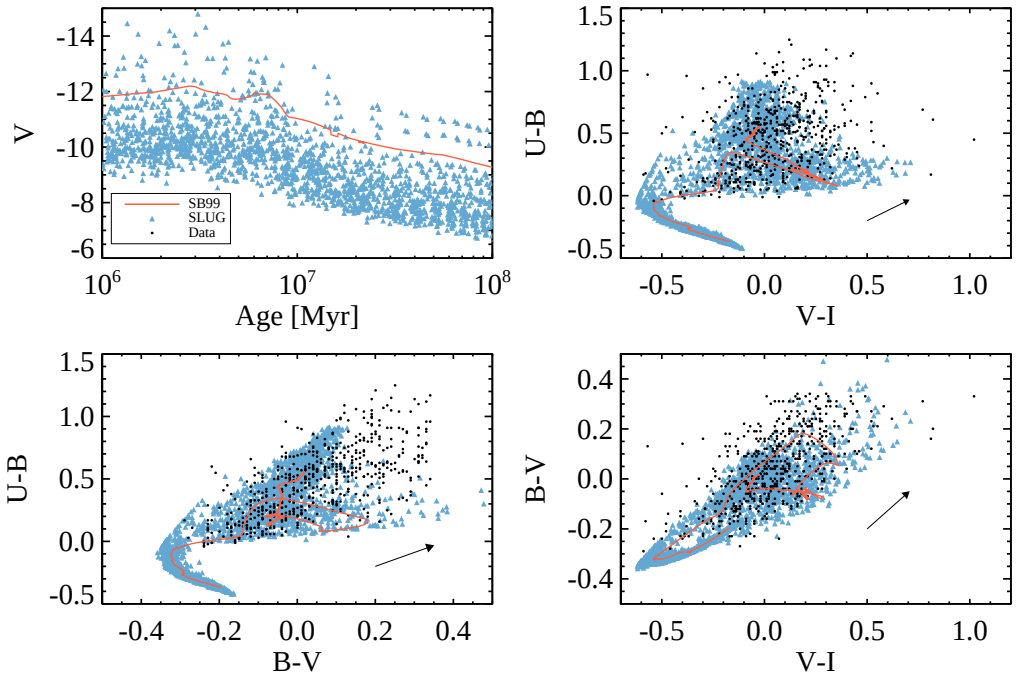

Figure 2.4 Comparison of observed young star clusters from Larsen (1999) (black points) to SLUG models of clusters $> 10^4 M_\odot$ (blue triangles). The orange curves show the trajectory of a SB99 $10^5 M_\odot$ cluster as a function of time. Data are omitted from upper left panel as the ages are not present in the Larsen (1999) catalog. Arrows denote the extinction vector for $A_V = 0.5$ mag (created following appendix B of Schlegel et al., 1998).

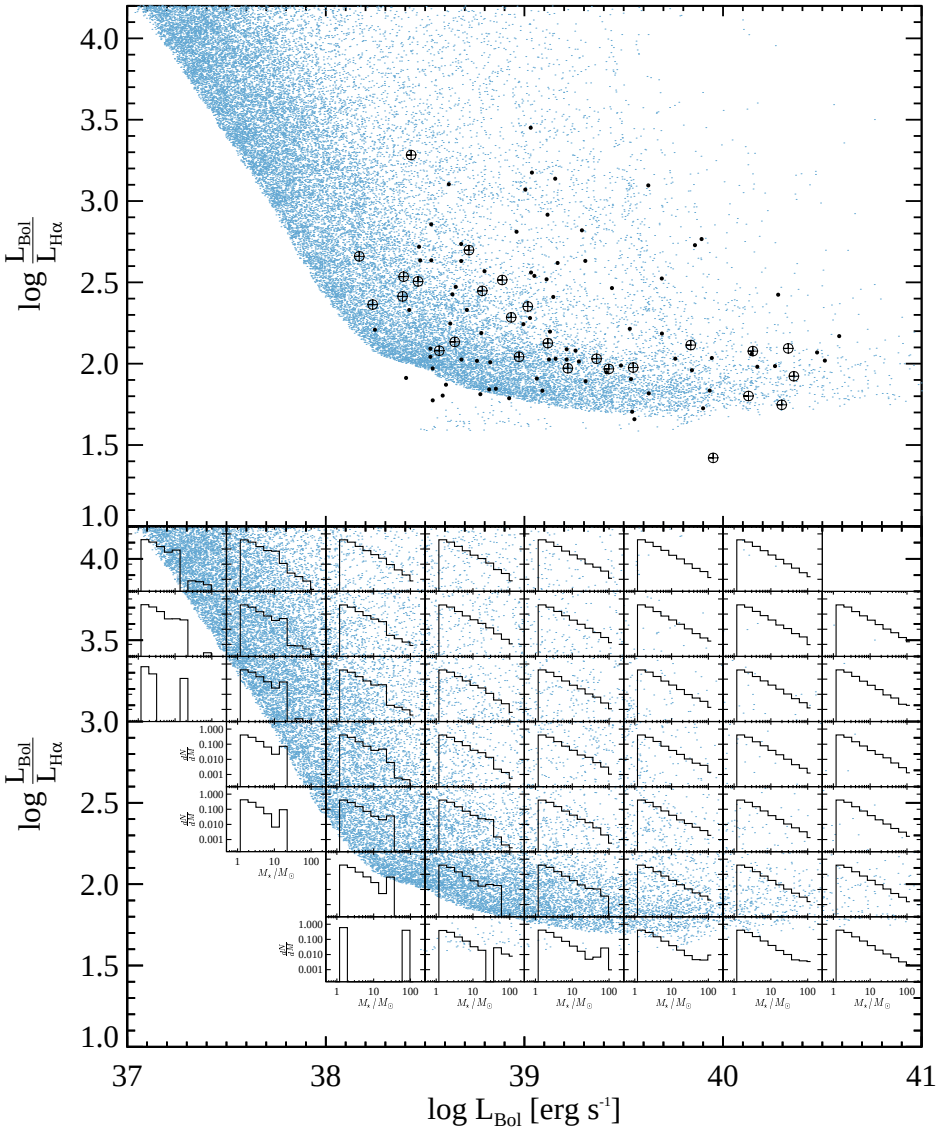

Figure 2.5 (*top*) Here we present the birthline as first discussed by Corbelli et al. (2009) whose original data are shown as black points and crosses (circle-crosses denote their 'clean' sample). Blue data points are clusters from SLUG. We see that our models are in relatively good agreement with observations. (*bottom*) We present overlays demonstrating the average IMFs in each region of the birthline plot. Note how the IMF becomes progressively more bottom heavy moving to lower luminosity.

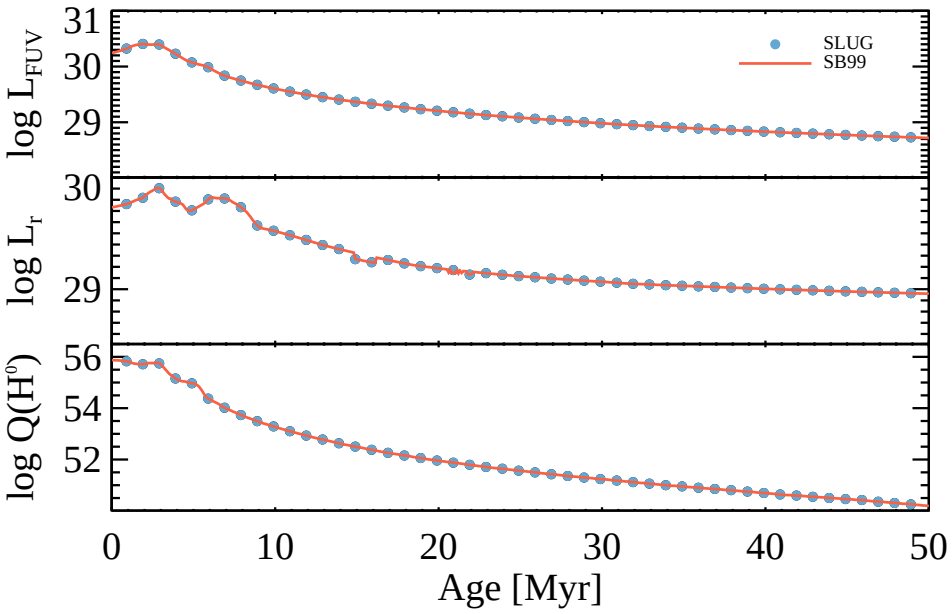

Figure 2.6 A comparison of SLUG and SB99 simulations of an instantaneous burst of $10^9 M_\odot$. We find good agreement between the two predictions in both the absolute normalization of the fluxes as well as the time-dependent behavior. FUV and r band fluxes are in units of ergs/s/Hz while $\mathrm{Q(H^0)}$ is in units of photons/s. In the L_r panel, in the age ranges of 12-18 Myr and 20-22 Myr, one can see minor imperfections in interpolating the stellar tracks as implemented by SB99.

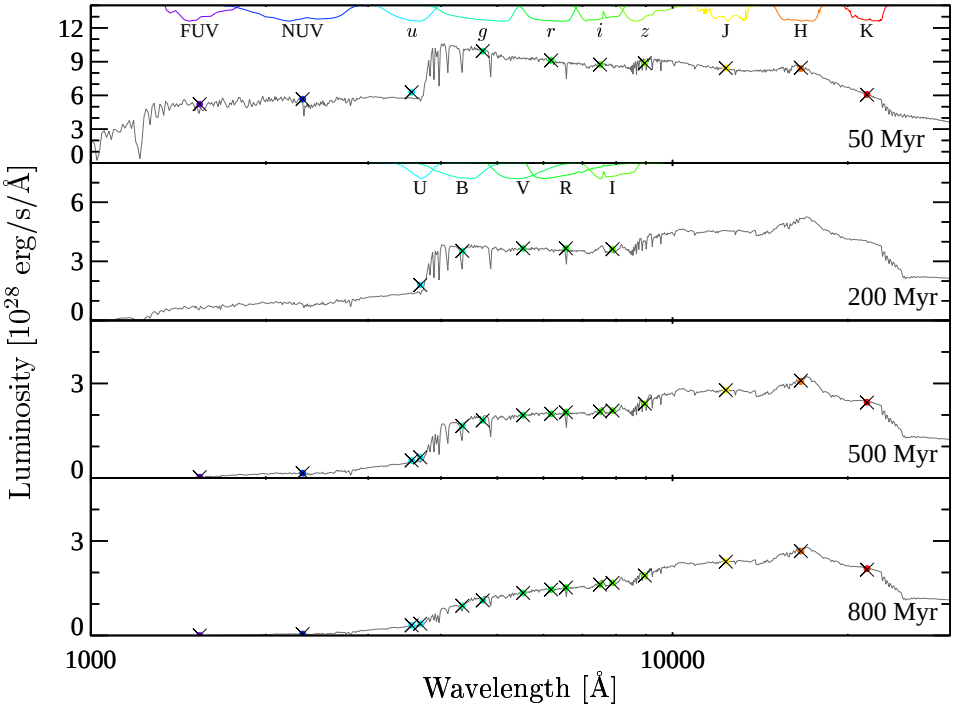

Figure 2.7 A comparison of SLUG and SB99 photometry for an instantaneous burst of $10^9 M_\odot$, evaluated at the ages indicated in each panel. The grey solid line represents the output spectrum from SB99 for such a population. The filled color circles show the SB99 integrated fluxes for the filters available to SLUG. The black $\times$'s mark the SLUG photometry for the well-sampled model described in section 2.4.3.

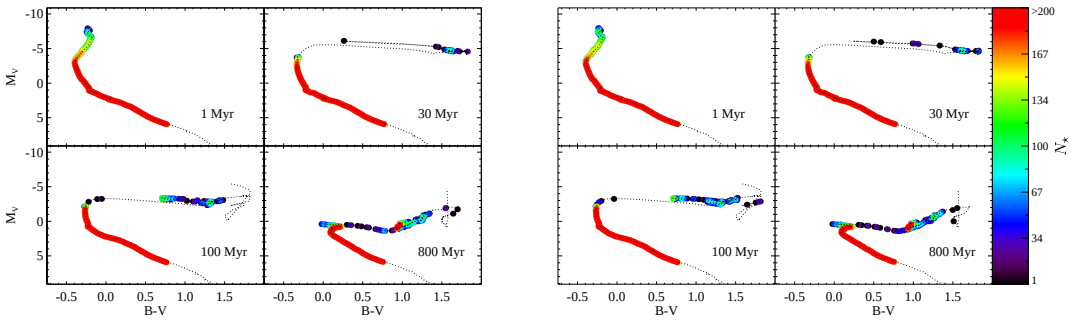

Figure 2.8 CMDs of two realization of an instantaneous $10^5 M_\odot$ burst population at the ages indicated in each panel. Only stars more massive than 0.9 $M_\odot$ are binned in the CMDs. The dotted lines show the corresponding theoretical isochrones. The SLUG CMD has been convolved with circular top-hat PSF solely to improve visibility. The color bar denotes the number of stars in that region of the diagram.

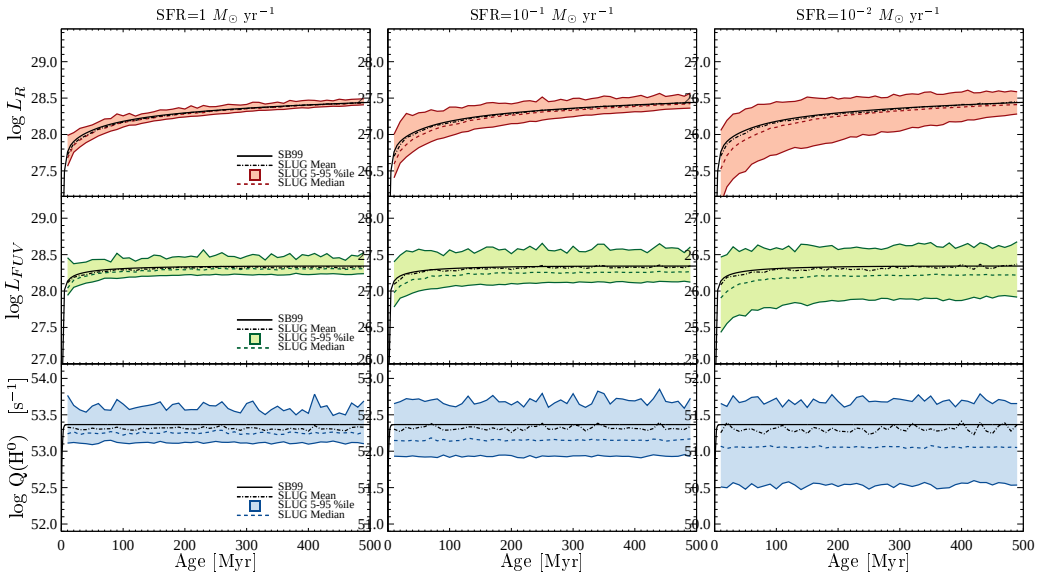

Figure 2.9 R-band, FUV, and ionizing photon luminosities vs. time for galaxies with constant SFRs of 1, 10^{-1}, and 10^{-2} $M_\odot$ yr^{-1} as indicated. R-band and FUV luminosities are in units of erg s^{-1} Hz^{-1}. We compare a fully sampled realization from SB99 (solid black lines) with 100, 500, and 1000 realizations from SLUG for SFRs of 1, 10^{-1}, and 10^{-2} $M_\odot$ yr^{-1} respectively. The SLUG models are represented by their arithmetic mean (black dash-dotted line), median (colored dashed line) and 5-95 percentile range (filled color region). The timestep in our SLUG models was set to 10 million years. Note that the y-axis in each panel has been chosen to match the SFR, but always spans the same logarithmic interval.

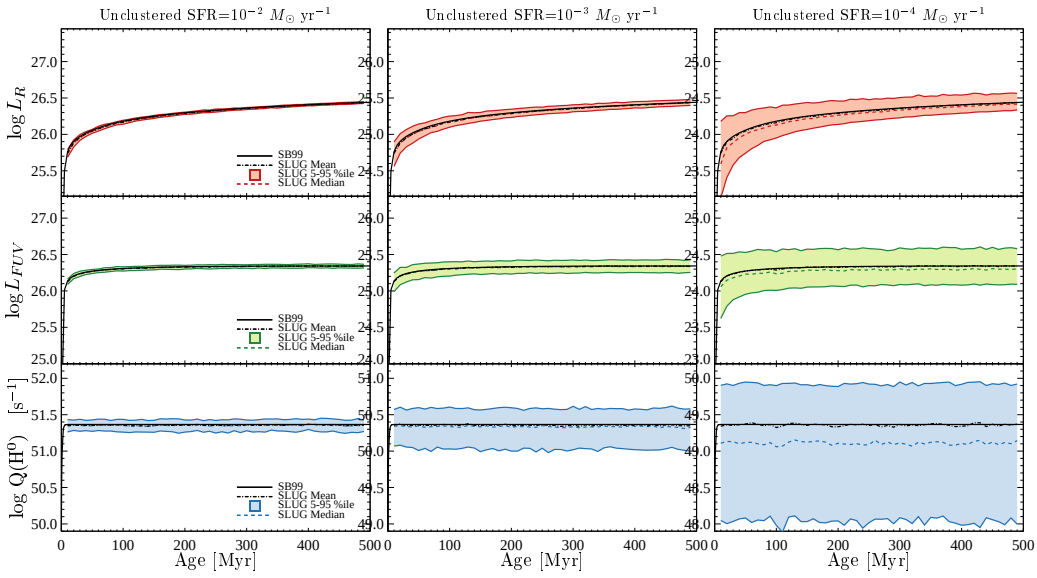

Figure 2.10 Same as Figure 2.9, but this time made with unclustered star formation, and using lower SFR. Note the third panel of Figure 2.9 is the same SFR as the first panel of this figure. These figures were constructed with 100, 500, and 1000 realizations at SFRs of 10^{-2}, 10^{-3}, and 10^{-4} $M_\odot$ yr^{-1} respectively.

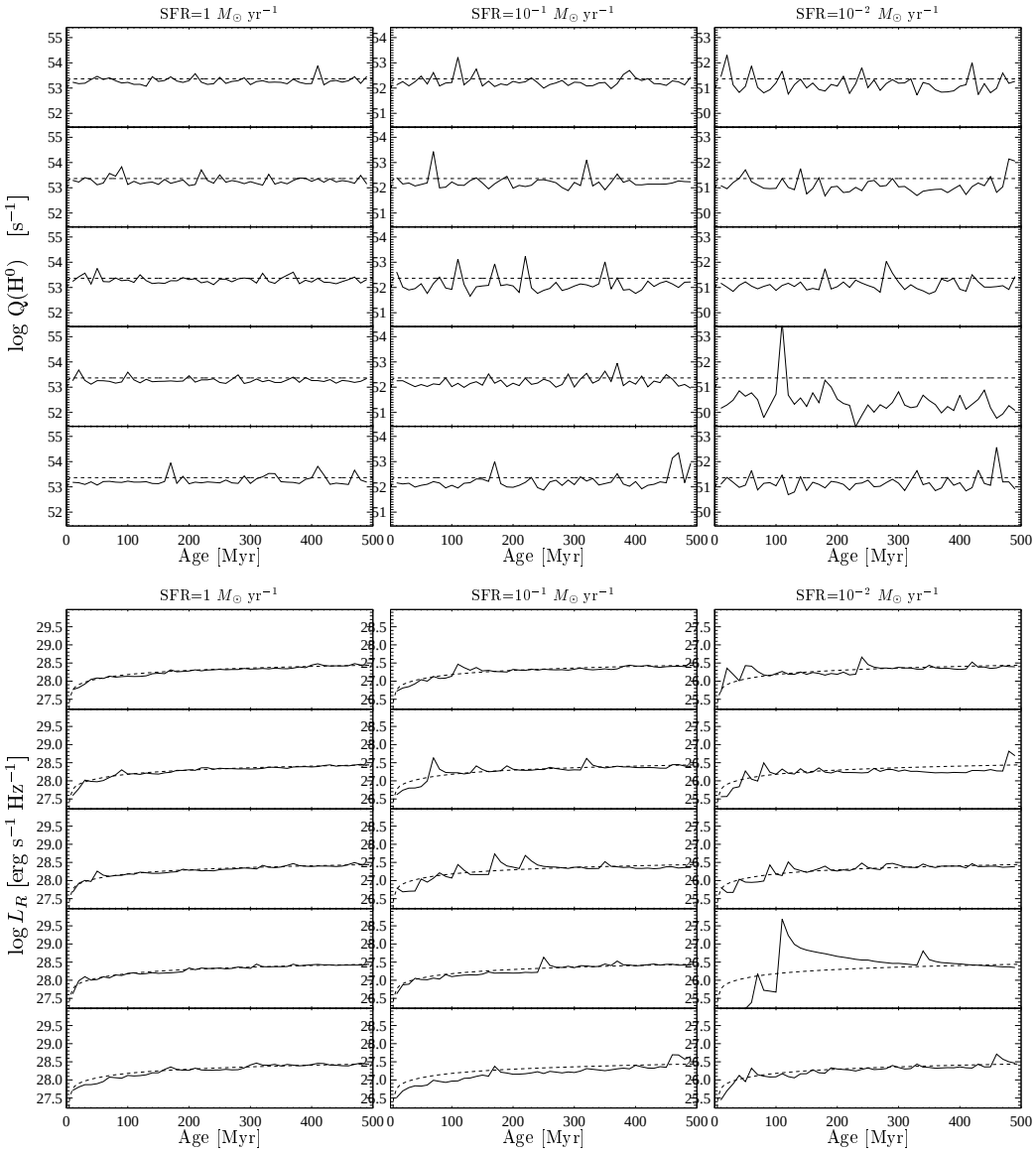

Figure 2.11 Solid lines show the evolution of $Q(H^0)$ and R band luminosity for individual simulations with clustered star formation with SFRs of 1, 10^{-1}, and 10^{-2} $M_\odot$ yr^{-1}. Dashed lines show the SB99 prediction. Note that the y-axis in each panel has been chosen to match the SFR, but always spans the same logarithmic interval.

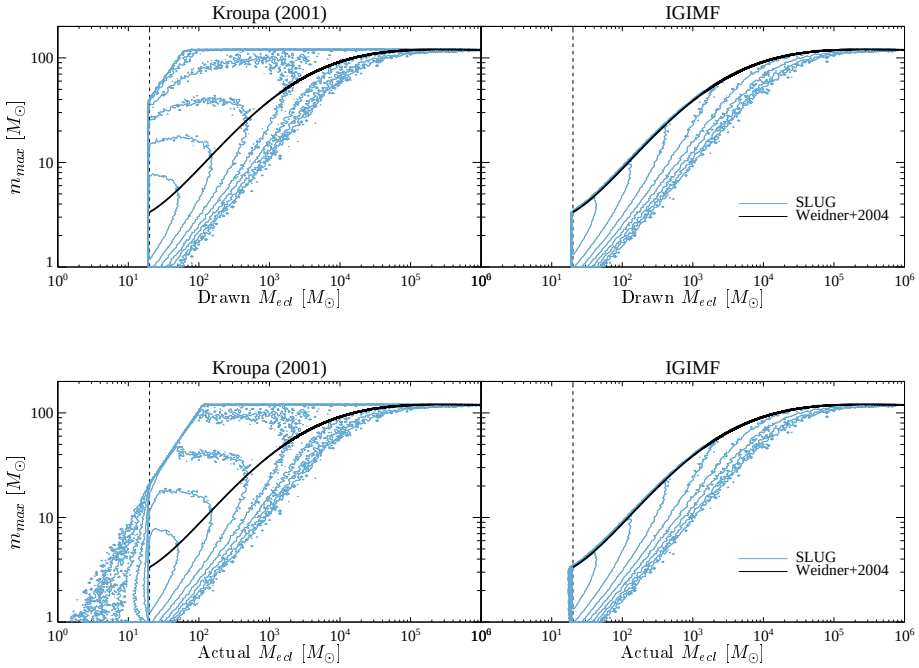

Figure 2.12 The mass of the largest star in a cluster vs. that cluster's mass for clusters created by SLUG for a Kroupa (2001) IMF (*left*) and the IGIMF (*right*). The black lines denote the analytic prediction of the maximum possible stellar mass in a cluster in the IGIMF model, the black dashed line notes the lower limit of the initial cluster mass function, and blue contours denote the location of SLUG models. Top panels show the maximum stellar mass as a function of the cluster mass drawn from the ICMF, while bottom panels show the same relation relative to the sum of the masses of all stars actually populating the clusters. These two differ slightly– see section 2.3.2.

Chapter 3

Quantifying SFR Uncertainties

The integrated light of a stellar population, measured through photometric filters that are sensitive to the presence of young stars, is often used to infer the star formation rate (SFR) for that population. However, these techniques rely on an assumption that star formation is a continuous process, whereas in reality stars form in discrete spatially- and temporally-correlated structures. This discreteness causes the light output to undergo significant time-dependent fluctuations, which, if not accounted for, introduce errors and biases in the inferred SFRs. We use `SLUG` (a code that Stochastically Lights Up Galaxies) to simulate galaxies undergoing stochastic star formation. We then use these simulations to present a quantitative analysis of these effects and provide tools for calculating probability distribution functions of SFRs given a set of observations. We show that, depending on the SFR tracer used, stochastic fluctuations can produce non-trivial errors at SFRs as high as 1 $M_\odot$ yr^{-1}, and we suggest methods by which future analyses that rely on integrated-light SFR indicators can properly account for stochastic effects. We emphasize that due to the stochastic behavior of blue tracers of SFR, one cannot assign a deterministic single value to the SFR of an individual galaxy.

3.1 Introduction

Stellar light is the primary observable in astronomy, and it provides most of our knowledge of the universe and its evolution. While for the nearest stellar populations we can observe individual stars, we are often restricted to measuring

the integrated photometric properties of stars, both spatially and spectrally. These integrated properties, when filtered through a model for stellar populations, can then yield estimates of the mass, star formation rate (SFR), star formation history (SFH), initial mass function (IMF), and numerous other properties for the observed stellar population.

Because the light produced by a star is a function of its mass and age, the stellar population synthesis (SPS) models required to map between observed luminosity and underlying physical properties involve calculating a sum over the mass and ages of all the stars that comprise the population. The most commonly-used approaches for evaluating this sum rely on several assumptions for computational efficiency. Most relevant to this paper, it is common to assume that the IMF and star formation history (SFH) are infinitely well populated (e.g., `Starburst99`: Leitherer et al. 1999; Vázquez & Leitherer 2005; `PEGASE`: Fioc & Rocca-Volmerange 1997; `GALEV`: Kotulla et al. 2009; `FSPS`: Conroy et al. 2009, 2010; Conroy & Gunn 2010). This approach is convenient because it replaces the sum with a separable double-integral: one first integrates over the IMF at fixed time to calculate the light per unit mass for a stellar population as a function of age, and then integrates this light to mass ratio weighted by the star formation history in order to arrive at an estimate of the integrated light produced by stars of all ages.

While this approach is convenient, it can also be dangerous. The potential pitfalls of assuming a fully-sampled IMF when analyzing a simple stellar population (i.e., a group of stars of uniform age) are well-known: if the IMF is not fully-sampled, the highly nonlinear dependence of luminosity on stellar mass causes the manner in which stars discretely fill a population's mass to have large consequences for the luminosity (e.g., Cerviño & Valls-Gabaud, 2003; Cerviño & Luridiana, 2004; Popescu & Hanson, 2009b, 2010a,b). In this case, there is no longer a deterministic relation between the total mass and age of the population to the total luminosity and color of its integrated light. The implication is that the inverse problem, that of determining the mass or age of a simple stellar population from its photometric properties, no longer has a unique solution. Nor can this non-uniqueness be described as a simple error symmetrically bracketing a central estimate. In a small stellar population, a single high mass star can dramatically increase (and at times dominate) the luminosity of a stellar population. Thus, this very high luminosity for rare realizations

skews the mean of the luminosity distribution well away from its median. As a result, mean relations for luminosities that are in the stochastic regime (where a single star can dramatically affect the luminosity) often greatly overpredict the luminosity of a randomly chosen realization. See Cerviño (2013) for a recent review of this topic, and a discussion of the implications of these uncertainties.

The hazards of assuming a well-sampled SFH, and thus the accuracy of SPS models that make this assumption, have received significantly less attention (e.g., see the recent review by Kennicutt & Evans 2012a). We know from observations of both the Milky Way and nearby galaxies that star formation is a highly clustered process (e.g., Lada & Lada, 2003), which more closely resembles a series of discrete bursts identifiable with the formation of individual clusters than the continuous creation of new stars at a constant rate. Only when the SFR is sufficiently high do the individual bursts blur together to create an approximately continuous SFH (see figures 3 and 11 of da Silva et al. 2012, hereafter Paper I). The question of how integrated light is affected by stellar clustering coupled to finite IMF sampling motivated us to create the Stochastically Lighting Up Galaxies (`SLUG`) code (da Silva et al., 2012, hereafter Paper I). This code hierarchically follows clusters drawn from a cluster mass function, each of which is individually populated on a star-by-star basis according to an IMF. Each star evolves following an individual evolutionary track, and contributes light calculated from an individual stellar atmosphere model. As a result of this approach, `SLUG` produces Monte Carlo realizations of stellar populations rather than simply the mean results, including stochasticity in both the IMF and the SFH. Our initial application of this code (Fumagalli et al., 2010; da Silva et al., 2012) showed that, for non-simple stellar populations, SFH sampling stochasticity turns out to affect the light output of stellar populations far more than IMF sampling stochasticity. Indeed, Fumagalli et al. (2011b) (also see Weisz et al. 2012b) show that this effect explains the low Hα to FUV ratios seen in dwarf galaxies (Lee et al., 2009; Boselli et al., 2009; Meurer et al., 2009), something that some earlier authors had erroneously attributed to variations in the IMF itself. Since this initial application, `SLUG` has been used to study these effects in a number of other contexts (Siana et al., 2010; Cook et al., 2012; Andrews et al., 2013; Forero-Romero & Dijkstra, 2013).

In this paper we extend the application of `SLUG` to the problem of inter-

preting SFR indicators (SFIs). These are, by construction, extremely sensitive to the properties of the most massive, shortest lived, brightest stars, and thus are very vulnerable to stochasticity. They are therefore subject to the same "inverse problem" that affects the determination of mass and age for simple stellar populations: at low SFRs, where IMF and SFH are sparsely sampled, there is no unique mapping between SFRs and SFIs, and thus no unique way to infer a SFR from a SFI in an individual galaxy[1]. Given these limitations, our goal in this paper is to provide the next-best possible solution: a full characterization of the probability distribution function (PDF) of SFR given a particular observed value of SFI. The layout of the paper is as follows: Section 3.2 describes a library of `SLUG` simulations that we have performed to solve the forward problem of characterizing the distribution of luminosities that result from stochastic sampling of the IMF, including the effects of clustering[2] and a discussion of the dependence on free parameters. Section 3.3 describes how we use these `SLUG` simulations to solve the inverse problem of determining the PDF of SFR given a set of observations, including the higher-dimensional correlations between the true underlying SFR and multiple SFIs. Finally, Section 3.4 discusses the implications of this work, and Section 3.5 summarizes our conclusions.

3.2 The Distribution of Luminosity at Fixed Star Formation Rate

3.2.1 `SLUG` Simulations

We first consider the problem of determining the distribution of luminosities of SFIs given an input SFR. This allows us to determine, for example, how much scatter is expected for a given stellar population and to characterize the types of errors one might incur if only using the mean properties. We approach this problem via `SLUG` simulations, which produce Monte Carlo realizations of photometric properties given a set of user inputs including the input SFH, IMF, the initial cluster

[1]The mean relations are still accurate. On average those SFRs produce that SFI luminosity. However, the interpretations that simply use the mean relation are not appropriate. The broad and highly skewed nature of the PDFs for SFR given a SFI mean that care must be taken to properly interpret observations.

[2]Stellar clustering is the dominant mechanism for SFH sampling stochasticity, thus in several places we use the terms "SFH stochasticity" and "effects of clustering" interchangeably.

Table 3.1. SLUG Simulation Parameters

	Fiducial
f_c	1
$t_{\rm sf}$ [Myr]	500
$m_{max}\,[M_\odot]$	10^7
$m_{min}\,[M_\odot]$	20
[Fe/H]	0
IMF	Kroupa

Note. — Here f_c is the clustering fraction, $t_{\rm sf}$ is the duration of star formation, $m_{\rm min}$ and $m_{\rm max}$ are the minimum and maximum of the ICMF, [Fe/H] is the metallicity used for the stellar evolution and atmosphere models, and IMF is the choice of stellar initial mass function. For a description of how each of these parameters is implemented in SLUG, see Paper I.

mass function (ICMF), the fraction of star formation occurring in clusters, and a set of stellar evolutionary tracks and atmosphere models. The code also takes parameters describing how clusters disrupt, but these affect only the properties of the cluster population, not the integrated light of a galaxy, and so we will not refer to them further. Unless otherwise noted, all our simulations make use of the default SLUG parameter choices described in Paper I, and summarized in Table 3.1. We also refer readers to da Silva et al. (2012) for a full description of SLUG's functionality.

For the purposes of this paper, we restrict ourselves to very simple input SFHs: those with constant SFR over a time of 500 Myr.[3] Our choice of time period

[3]It is important to note that, as discussed in Paper I, the input SFH does not match the actual

is long enough that we avoid any transient initial phases of the buildup of the stellar population. The primary output of each SLUG simulation is a realization of the PDF of luminosities given a SFR and other ancillary variables,

$$p(\ell \mid \log \mathrm{SFR}, \phi), \tag{3.1}$$

where ℓ is a vector of log luminosities in various photometric bands and ϕ denotes parameters that define the model. For simplicity, in the analysis that follows we will omit ϕ except where relevant.

While SLUG is capable of producing photometry in many bands, and the next release of the code will support full spectra, here we focus on the three most common indicators of the SFR: the FUV luminosity L_{FUV}, the bolometric luminosity L_{bol}, and the Hα luminosity $L_{\mathrm{H}\alpha}$. The last of these is a recombination line produced when the ionizing radiation of the stars interacts with the ISM, and SLUG does not report this directly. Instead, it reports the rate of hydrogen-ionizing photon emission $\mathrm{Q(H^0)}$, which we convert to Hα luminosity via

$$\begin{aligned} L_{\mathrm{H}\alpha} &= (1-f_{\mathrm{esc}})(1-f_{\mathrm{dust}})\mathrm{Q(H^0)}\alpha^{\mathrm{eff}}_{\mathrm{H}\alpha} h\nu_{\mathrm{H}\alpha} \\ &\approx 1.37\times 10^{-12}(1-f_{\mathrm{esc}})(1-f_{\mathrm{dust}})\mathrm{Q(H^0)\ erg}, \end{aligned} \tag{3.2}$$

where f_{esc} and f_{dust} are the fractions of ionizing photons that escape from the galaxy and that are absorbed by dust grains rather than hydrogen atoms, respectively, $\alpha^{\mathrm{eff}}_{\mathrm{H}\alpha}$ is the recombination rate coefficient for recombination routes that lead to emission of an Hα photon, and $h\nu_{\mathrm{H}\alpha} = 1.89$ eV is the energy of an Hα photon. For the purposes of the plots presented in this paper, we take $f_{\mathrm{esc}} = f_{\mathrm{dust}} = 0$, but adopting non-zero values for one or both of them would simply amount to applying a constant shift to our results[4]. Similarly, although we focus on Hα, the results will be identical up

realized SFH. In fact, due to stochastic sampling of the cluster mass function, the output SFH will differ from the input SFH as it will exhibit a series of bursts on small time scales (see figure 3 in Paper I). This is because there is no "constant" star formation rate. For example, consider a galaxy forming stars at 1 $M_\odot \mathrm{yr}^{-1}$. In one day, 1/365th of a solar mass of gas is not transformed into a star. Constant star formation rates (and star formation histories in general) can only be considered continuous when averaged over some time interval. In our case, the observations dictate their own averaging window and we investigate how well the continuous model matches reality.

[4]Note that care should be taken when correcting for dust with other star formation rate indicators since (as we will show in section 3.2.2) at fixed star formation rate, the ratios of SFIs can have considerable scatter.

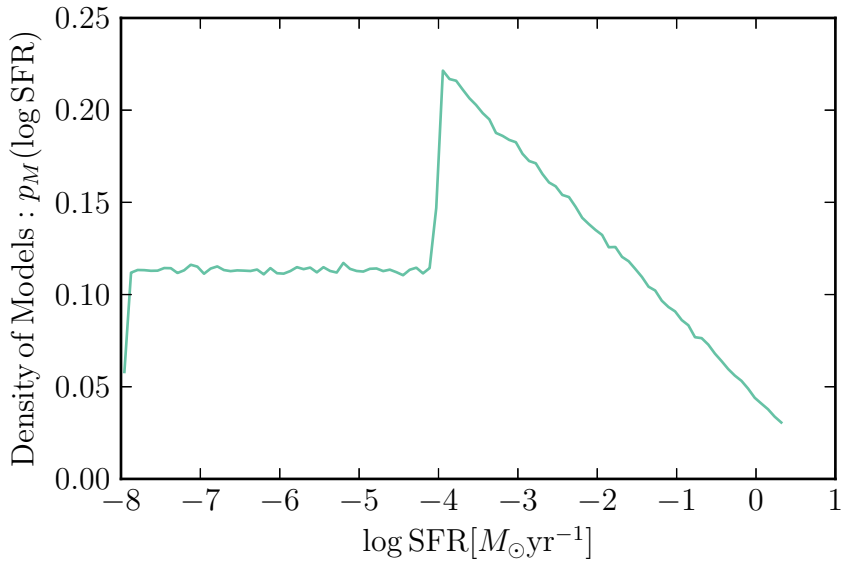

Figure 3.1 Distribution of input SFRs for `SLUG` simulations. The discontinuity at 10^{-4} occurs because of the change in simulation strategy.

to a constant shift for any other hydrogen recombination line, or any other source of emission (e.g., free-free emission) that is directly proportional to the ionizing luminosity. We leave for future work the discussion of other SFIs that have more complex, non-linear relationships with the ionizing photon production rate (e.g., [O II] 372.7 nm, [Ne II] 12.8 μm, [Ne III] 15.6 μm – Kennicutt & Evans 2012a).

In order to characterize the PDFs of our chosen SFIs, we run approximately 1.8×10^6 `SLUG` models. Of these models, we run 9.83×10^5 at input SFRs with a distribution of $\log \mathrm{SFR}$ that has a linear form with a slope of -1 over a range in $\log \mathrm{SFR}$ from -4 to 0.3, where SFRs here are measured in $M_\odot$ yr^{-1}. The remaining 0.8×10^6 models are uniformly distributed in $\log \mathrm{SFR}$ over a range from -8 to -4. The distribution of the model star formation rates $p_M(\log \mathrm{SFR})$ is shown in Figure 3.1. Our choice of distribution is motivated by the practical requirement that we need more simulations to adequately sample the PDFs at lower SFRs because the scatter is larger. As we will show in section 3.3.1, our results do not depend on the assumed distribution of models, $p_M(\log \mathrm{SFR})$.

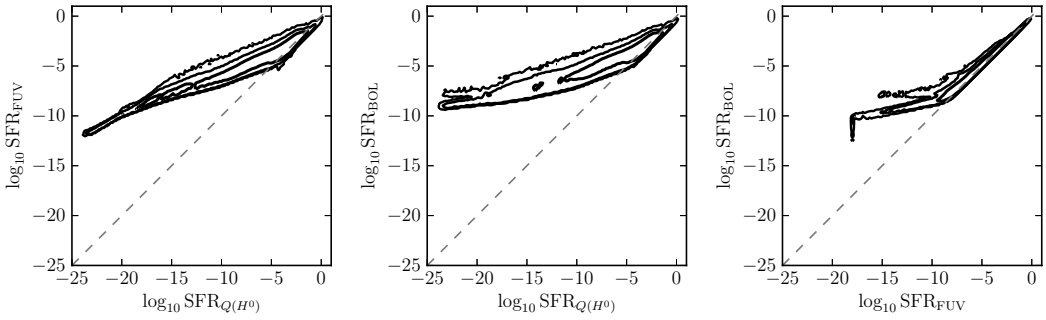

Figure 3.2 Contours representing $10^{[1,2,3]}$ models for the three different SFIs converted to SFRs using the point-mass approximations. Without stochastic effects, the galaxies would be forced to lie exactly on the dashed line.

3.2.2 Simulation Results

For convenience, we report the result of our simulations in SFR space, meaning that we report luminosities as the SFRs one would infer using the approximation of perfect IMF and SFH sampling, which we refer to as the "point mass approximation"[5]. For our fiducial IMF, stellar evolution tracks and atmosphere models, the conversions between these and the luminosities reported by `SLUG` are

$$\mathrm{SFR}_{Q(\mathrm{H}^0)} = 7.6383 \times 10^{-54} (M_\odot \mathrm{yr}^{-1}\ \mathrm{s}) Q(\mathrm{H}^0) \quad (3.3)$$

$$\mathrm{SFR}_{\mathrm{FUV}} = 9.6415 \times 10^{-29} (M_\odot \mathrm{yr}^{-1}\ \mathrm{erg}^{-1}\ \mathrm{s\ Hz}) L_{\mathrm{FUV}} \quad (3.4)$$

$$\mathrm{SFR}_{\mathrm{bol}} = 2.6607 \times 10^{-44} (M_\odot \mathrm{yr}^{-1}\ \mathrm{erg}^{-1}\ \mathrm{s}) L_{\mathrm{bol}}. \quad (3.5)$$

This approach allows us to report the results using the different SFIs on a common scale, making them easier to compare. It also allows us to separate the effects of stochastic sampling from the dependence of the results on the choice of stellar evolution and atmosphere models as these, to good approximation, simply cause changes in the conversion constants in equations (3.3) – (3.5).

Each `SLUG` model may be thought of as a point in a four-dimensional parameter space defined by these three luminosities and their corresponding intrinsic SFR. In Figure 3.2, we show the raw distribution of our models in three orthogonal projections of this parameter space. Figure 3.3 presents the distributions related to

[5]This is standard statistical terminology. It arises from the fact that one treats the posterior PDF as having all of its mass located at a single point.

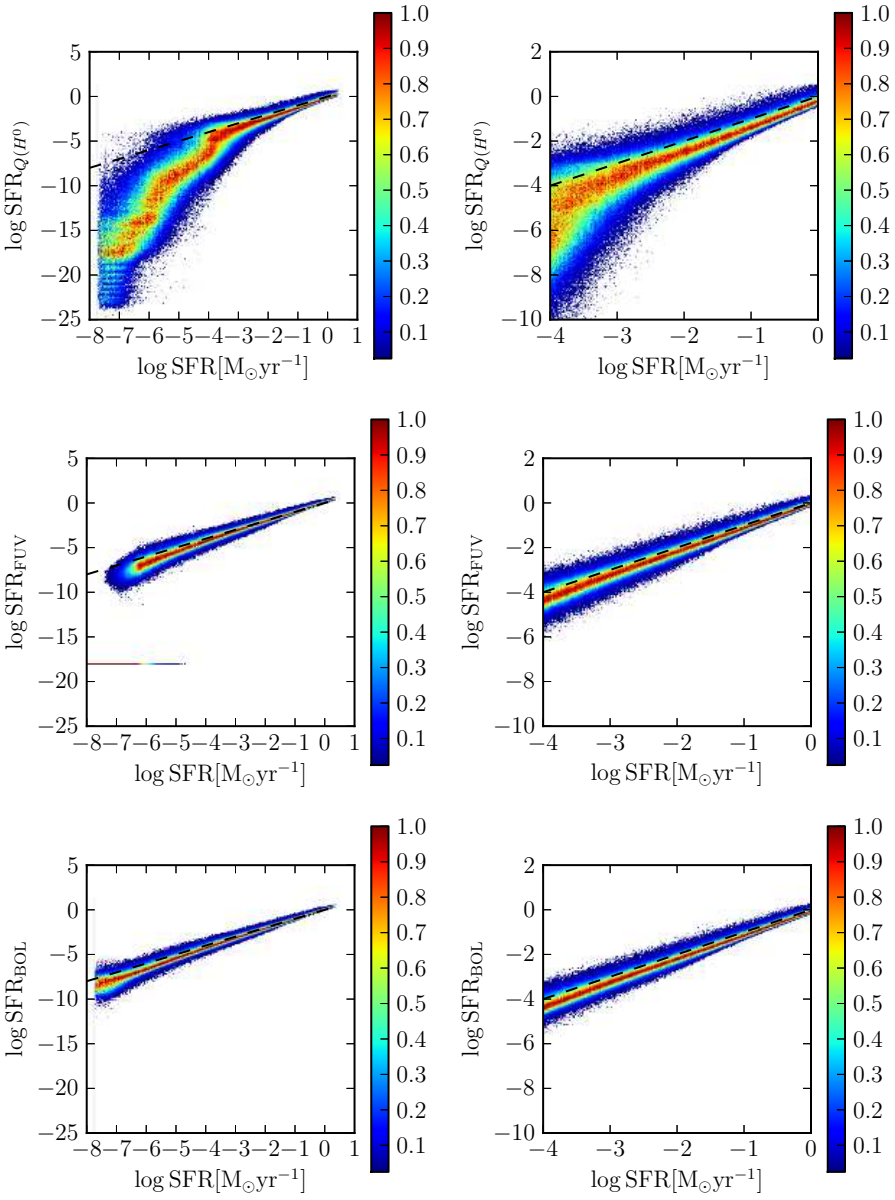

Figure 3.3 (*left*) PDFs of the SFIs vs. intrinsic SFR arising just from stochastic effects (presented as fraction of the maximum value in each intrinsic SFR bin). The dashed line represents the point-mass approximation. The hard cutoff at $\log \mathrm{SFR} = -8 + \log 2 [M_\odot \mathrm{yr}^{-1}]$ is the smallest SFR that can produce any clusters with a mass of 20 $M_\odot$, the minimum cluster mass we allow. The horizontal stripe for $\mathrm{SFR}_{\mathrm{FUV}}$ at -18 corresponds to the lower limit of FUV luminosity given by the `SLUG` models. (*right*) Zoomed in version of plots in left column.

their intrinsic SFRs. We can immediately see that there is significant mass of models well away from the line predicted by the point-mass approximation, confirming the necessity of the stochastic treatment and our assertions that full PDFs should be used in place of simple mean relations. We also see that, as expected, the deviation from the line is largest for $\mathrm{SFR}_{Q(\mathrm{H}^0)}$, and smaller for the other two dimensions. This was discussed in Fumagalli et al. (2011b), as tracers that are sensitive to stars with lifetimes shorter than a few Myr are most sensitive to the flickering in the SFH, while SFI that depend on longer lived stars average over longer time scales and are thus more stable in recovering the mean SFH.

While a clear picture of the ensemble of all the models is presented in figure 3.3 (which is critically useful in our subsequent analysis – see Section 3.3), explorations of the level of scatter can perhaps be better addressed by Figure 3.4, which shows the marginal distributions of $p(\ell \mid \log \mathrm{SFR})$. To emphasize the shape of the distribution over the actual values that are related to adopted point mass calibrations, we plot the distribution of the offsets between these inferred SFRs and the true SFR that was used in each simulation. It is again clear that $\mathrm{Q}(\mathrm{H}^0)$ has the largest scatter[6], in extreme cases producing estimates that differ from the true SFR by as much as eight orders of magnitude! Furthermore, these distributions are clearly not Gaussians centered on the true SFR. Instead, they are highly asymmetric. Finally, it is clear that as the SFR increases, the PDF gets narrower. This is the result being better sampled and the laws of statistics of large numbers.

[6]It is important to caution that, while large scatter is a real limitation of SFIs based on ionizing luminosity, it would be incorrect to conclude from this that alternates such as FUV or bolometric luminosity are always preferable. If the true SFR is stable on the $\sim 10 - 100$ Myr timescales to which these tracers are sensitive, as is the case in our models, then, all other factors equal, they are preferable. However, in a galaxy where the intrinsic SFR might be variable on shorter timescales (e.g., in a merging or interacting galaxy), the longer averaging interval of FUV or bolometric luminosity becomes a disadvantage, as it produces too coarse an estimate of the true SFR.

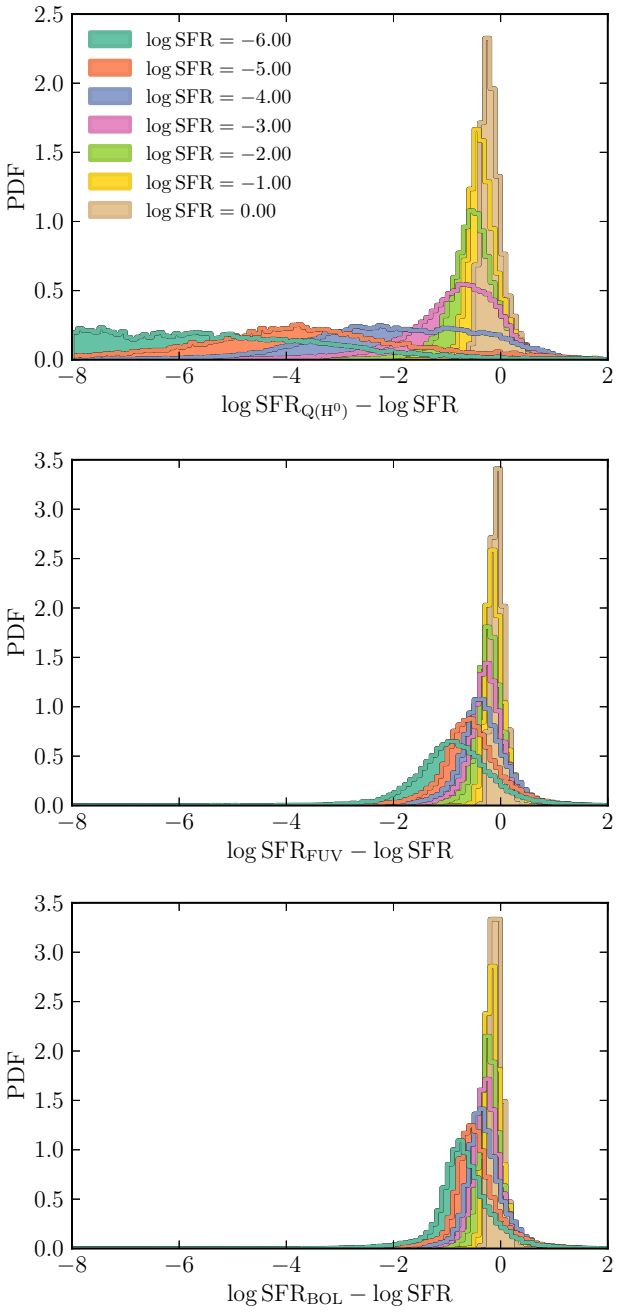

Figure 3.4 PDFs for individual components of ℓ normalized by the point-mass approximation for ease of comparison. Models are grouped by SFR into bins 0.25 dex wide, and are color-coded by input SFR as indicated in the legend.

3.3 The Distribution of Star Formation Rate at Fixed Luminosity

3.3.1 Derivation

Thus far we have shown how one can estimate the probability distribution of log luminosities ℓ given an intrinsic star formation rate, $p(\ell \mid \log \mathrm{SFR})$. However, we want to invert the problem and find the full distribution of SFRs given ℓ, i.e., $p(\log \mathrm{SFR} \mid \ell)$. We perform this inversion with a technique known as implied conditional regression. The idea behind this technique is simple. We start with the following decomposition:

$$p(\log \mathrm{SFR} \mid \ell) = \frac{p(\log \mathrm{SFR}, \ell)}{p(\ell)}. \tag{3.6}$$

Each `SLUG` model has a known SFR and produces an output ℓ, and thus represents a sample point in the multidimensional parameter space $(\log \mathrm{SFR}, \ell)$; we denote such a point as a vector $\mathbf{z}$, where the first component is $\log \mathrm{SFR}$, and the three log luminosities that comprise ℓ form the second through fourth components. This definition can obviously be generalized to an arbitrary number of components in ℓ. In this space, we define the distance between two points $\mathbf{z}_1$ and $\mathbf{z}_2$ by the usual Cartesian metric,

$$|\mathbf{z_1} - \mathbf{z}_2| = \sqrt{(\log \mathrm{SFR}_1 - \log \mathrm{SFR}_2)^2 + (\log \mathrm{SFR}_{Q(\mathrm{H}^0),1} - \log \mathrm{SFR}_{Q(\mathrm{H}^0),2})^2 + \cdots}. \tag{3.7}$$

The first task in computing $p(\log \mathrm{SFR} \mid \ell)$ is to use these sample points to estimate the underlying multidimensional probability distribution $p(\log \mathrm{SFR}, \ell)$ and its projection along the $\log \mathrm{SFR}$ direction

$$p(\ell) = \int p(\log \mathrm{SFR}, \ell)\, d \log \mathrm{SFR}. \tag{3.8}$$

To do this, we use a kernel density estimation technique which constructs the PDF as a sum of kernels centered on each multidimensional simulation point. Explicitly,

we approximate the value of the PDF at a position $\mathbf{z} = (\log \mathrm{SFR}, \ell)$ by

$$p(\log \mathrm{SFR}, \ell) = A \sum_i K(|\mathbf{z}_i - \mathbf{z}|; h), \tag{3.9}$$

where $\mathbf{z}_i$ is the position of the ith sample point, A is a normalization constant, and K is the kernel function, which has the bandwidth parameter h. For its compactness, we choose to use an Epanechnikov kernel, which is of the form

$$K(z; h) \propto \begin{cases} 1 - z^2/h^2, & z < h \\ 0, & z \geq h \end{cases}. \tag{3.10}$$

The parameter h must be chosen to balance the competing demands of smoothness, favoring larger h, and fidelity, favoring smaller h. We choose to set this parameter equal to 0.1 dex because exploration of histograms at various bin sizes indicates that there is little structure below this scale. We are thus washing out any features of this PDF on scales below 0.1 dex in any dimension. The result of this procedure is an estimate of the multidimensional probability density $p(\log \mathrm{SFR}, \ell)$ describing our raw `SLUG` data, and, by plugging into equation (3.6), an estimate of $p(\log \mathrm{SFR} \mid \ell)$.

The second step in computing $p(\log \mathrm{SFR} \mid \ell)$ is to the apply a proper weighting of the prior probability distribution of star formation rates. Simply applying equation (3.6) using our computed $p(\log \mathrm{SFR}, \ell)$ amounts to adopting a prior probability distribution of log star formation rates that follows the distribution of our `SLUG` simulations, shown in Figure 3.1. This is clearly not an ideal choice, as this distribution was chosen to ensure good sampling of the PDF, rather than to reflect a realistic prior distribution. Fortunately, it is trivial to rescale the results to an arbitrary prior probability distribution using Bayes's theorem,

$$p(\log \mathrm{SFR} \mid \ell) = \frac{p(\ell \mid \log \mathrm{SFR}) p(\log \mathrm{SFR})}{p(\ell)}, \tag{3.11}$$

where $p(\log \mathrm{SFR})$ is the prior probability distribution for the star formation rate.

Our input grid of models has a distribution of log SFR given by $p(\log \mathrm{SFR}) = p_M(\log \mathrm{SFR})$, where $p_M(\log \mathrm{SFR})$ is the distribution shown in Figure 3.1. Bayes's theorem tells us that we can use the results from one prior distribution $p_1(\log \mathrm{SFR})$) to find the results for a different prior distribution $p_2(\log \mathrm{SFR})$ by multiplying

$p(\log \mathrm{SFR} \mid \ell)$ by $p_2(\log \mathrm{SFR})/p_1 \log \mathrm{SFR})$.[7] For the case of transforming our SLUG simulations to a desired $p_2(\log \mathrm{SFR})$, we set $p_1(\log \mathrm{SFR}) = p_M(\log \mathrm{SFR})$. This is equivalent to assigning a different relative weighting of each of the models in the library such that the effective $p(\log \mathrm{SFR})$ matches whatever form is desired.

For the purposes of this analysis, we restrict ourselves to two possible priors. The first is the observed star formation rate distribution which Bothwell et al. (2011) parameterize as a Schecter function with slope -1.51 and characteristic SFR of $9.2\ M_\odot\ \mathrm{yr}^{-1}$. However, as a caveat, note that this observational determination was made ignoring the effects of stochasticity. This is unlikely to affect the characteristic SFR, since this is high enough that stochastic effects probably do not dominate the error budget for the FUV plus IR star formation rate indicated used in Bothwell et al.'s study (c.f. Figure 3.4). On the other hand, the slope at low SFR may be more problematic, a topic to which we return below.

The second prior we consider is a flat distribution of $\log \mathrm{SFR}$. This flat model is perhaps the most obvious prior, and is close to the distribution used in the SLUG simulations $p_M(\log \mathrm{SFR})$, but is in fact a relatively poor choice. The reality is that lower SFRs are more common and hence should be weighted more highly. Contrary to unfortunately common practice, assigning a uniform prior is neither "robust" nor prior-agnostic. It is in fact a very specific choice for a prior, which in this case is relatively poor. However, it does offer an interesting model to compare against to distinguish effects for the choice of prior. It also has the benefit that changing to another prior is perhaps easier to visualize since the term $p_1(\log \mathrm{SFR})$ is a constant. A flat prior on the linear scale of SFR (i.e., $p(\mathrm{SFR}) \propto 1$) is an extremely poor choice and should be avoided, since it is equivalent to assuming that higher values of $\log \mathrm{SFR}$ are more common.

Once a prior has been chosen, we are at last in a position to derive the final PDF of $\log \mathrm{SFR}$ given a set of observations. We can think of a given set of observational data as describing a PDF $p(\ell \mid \mathrm{data})$ of luminosities in one or more bands; the simplest case would be an observation of a single tracer which produces a central value of log luminosity with a Gaussian error distribution, in which case $p(\ell \mid \mathrm{data})$ is a Gaussian in one dimension (corresponding to the SFI measured)

[7] This operation requires calculation of a new normalization constant, which is simple to compute in the case of the one-dimensional SFR.

and is flat in the other dimensions (corresponding to SFIs that were not measured). Given the observations, and a choice of prior distribution $p(\log \mathrm{SFR})$ for the SFR, the final posterior distribution for the SFR is given by applying equation (3.6), rescaling by the chosen prior, and then integrating over the luminosity distribution implied by the data. The result is

$$p(\log \mathrm{SFR} \mid \mathrm{data}) = \int \frac{p(\log \mathrm{SFR}, \ell)}{p(\ell)} \frac{p(\log \mathrm{SFR})}{p_M(\log \mathrm{SFR})} p(\ell \mid \mathrm{data})\, d\ell, \tag{3.12}$$

where $p(\log \mathrm{SFR}, \ell)$ is given by equation (3.9), $p(\ell)$ is given by equation (3.8), and $p_M(\log \mathrm{SFR})$ is the PDF of SFRs in our `SLUG` simulations.

3.3.2 Results

To understand the results for the estimates of $p(\log \mathrm{SFR} \mid \ell)$, we begin by examining an example corresponding to the simplest case of a measurement for a single tracer. Consider an observation of Hα luminosity corresponding to $\log \mathrm{SFR}_{Q(H^0)} = -3$ with a Gaussian error bar of width σ. In Figure 3.5, we show the posterior PDF for the SFR given this measurement of Hα using both flat and Schecter function priors. If we had to assume point-mass conversion, we would infer $\log \mathrm{SFR} = -3$ for the galaxy SFR (the black dashed line). However, given the skewness in the flux distribution, the peak and mean of the true PDF[8] are significantly offset and neither corresponds to the point-mass estimate. We will characterize the difference between the point-mass estimate and the mean of the true PDF as the "bias". Note that this bias is *not* meant as a simple offset that one can blindly apply to the observational determination to get a "better" answer that fixes the stochastic issues. In practice, stochasticity fundamentally breaks the deterministic relationship between luminosity and SFR and thus the full PDF should be used whenever possible (or at least the first four moments of the distribution).

We can also see from Figure 3.5 that the posterior PDF of SFR has significant width. Thus even a perfect measurement of the luminosity, corresponding to $\sigma = 0$ in the Figure, retains a systematic uncertainty in the SFR with a standard deviation of approximately 0.5 dex and a significant negative tail. Indeed, in the

[8]Note that, as is always the case, this PDF is only true in so far as the prior is the correct prior to use and that other assumptions made are accurate as well regarding the IMF, stellar tracks and atmospheres, etc.

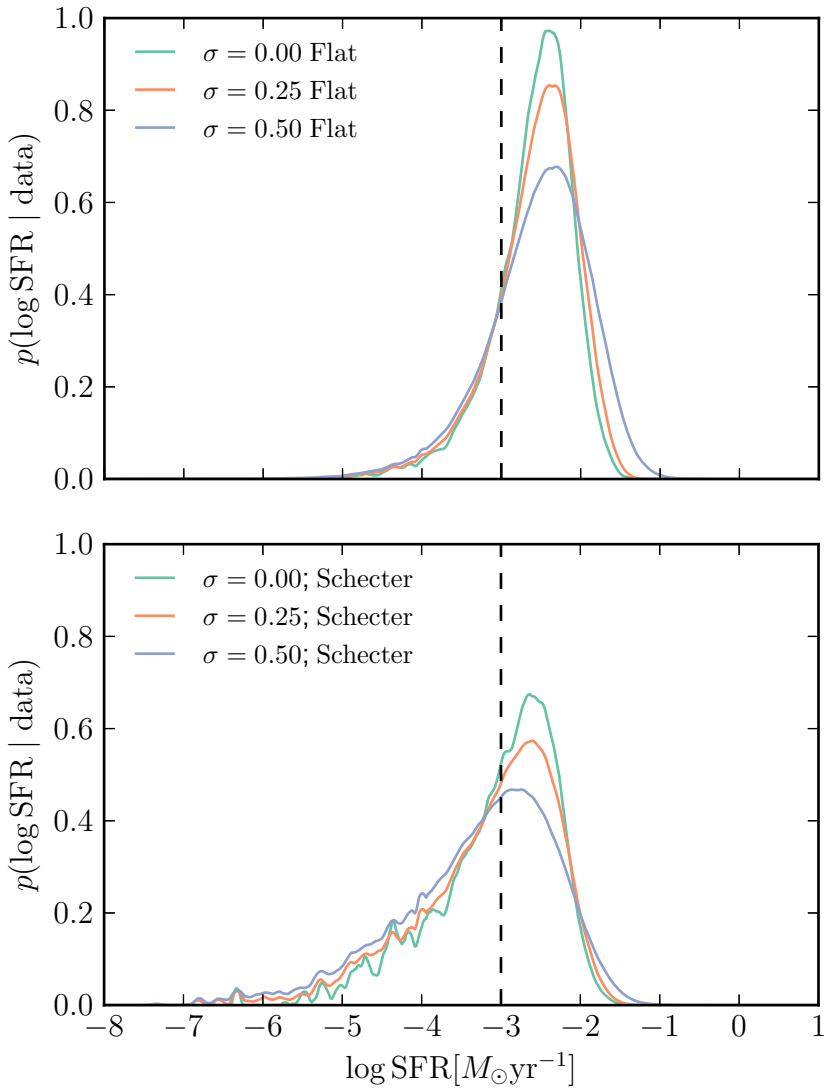

Figure 3.5 Posterior distributions for SFR given an observed Hα luminosity corresponding to a SFR centered at $\mathrm{SFR}_{Q(H^0)} = -3$. The observed log luminosity is taken to have a Gaussian-distributed uncertainty whose width σ (measured in dex) corresponds to the values shown in the legend; $\sigma = 0$ corresponds to a δ function distribution. The top panel shows results using a flat prior, and the bottom panel shows the results using a Schecter function prior (see Section 3.3.1). The curves get noisier at lower SFRs due to the smaller number of models and the more dispersed nature of the PDFs.

example shown, this stochastic uncertainty dominates the error budget, as is clear from the fact that the PDFs for observational errors of $\sigma = 0$, 0.25 dex, and 0.5 dex are only marginally different. Finally, we can see that the choice of prior does affect the results, but not significantly[9].

Given the results shown in Figure 3.5, it is obviously of interest to know how the bias and uncertainty depend on the observed value of a particular SFI. We formally define these quantities as follows. Consider an observation of a particular SFI I which returns an estimated log star formation rate $\log \mathrm{SFR}_I$ using the point-mass estimate (i.e., using equations 3.3 – 3.5), with a Gaussian error distribution σ on $\log \mathrm{SFR}_I$. The posterior probability distribution for the true star formation rate $p(\log \mathrm{SFR} \mid \log \mathrm{SFR}_I \pm \sigma)$ is then given by equation (3.12), treating the observed luminosity distribution $p(\ell \mid \mathrm{data})$ as a Gaussian of width σ centered at $\log \mathrm{SFR}_I$. The corresponding mean estimate of $\log \mathrm{SFR}$ is

$$\overline{\log \mathrm{SFR}} = \int p(\log \mathrm{SFR} \mid \log \mathrm{SFR}_I \pm \sigma) \log \mathrm{SFR}\, d \log \mathrm{SFR}. \tag{3.13}$$

We define the bias b and scatter s, respectively, as

$$b(\log \mathrm{SFR}_I) \equiv \overline{\log \mathrm{SFR}} - \log \mathrm{SFR}_I \tag{3.14}$$

$$s(\log \mathrm{SFR}_I)^2 \equiv \int p(\log \mathrm{SFR} \mid \log \mathrm{SFR}_I) \left(\log \mathrm{SFR} - \overline{\log \mathrm{SFR}}\right)^2 d \log \mathrm{SFR} \tag{3.15}$$

i.e., for a given observation of a single tracer, we define the bias as the difference between the mean value of $\log \mathrm{SFR}$ computed from the full PDF and the point-mass estimate, and the scatter as the second moment of the PDF of $\log \mathrm{SFR}$. Due to the nature of the distributions, normally the bias is positive.

Figure 3.6 shows the bias and scatter as a function of the observed luminosity of the three SFIs we consider in this paper, ionizing/Hα luminosity, FUV luminosity, and bolometric luminosity. As expected, we see that both the bias and scatter are reduced at high star formation rates, and that both are largest for ionizing luminosity-based SFRs, since they are the most sensitive to the most massive stars. Although it is not immediately apparent from the figure, ionization-based

[9]Given that the posteriors are so broad, this is the result of the fact that the priors are similar. Choosing a linearly flat $p(\mathrm{SFR}) \propto 1$ prior would produce significantly different results with a much higher weighting of higher SFRs.

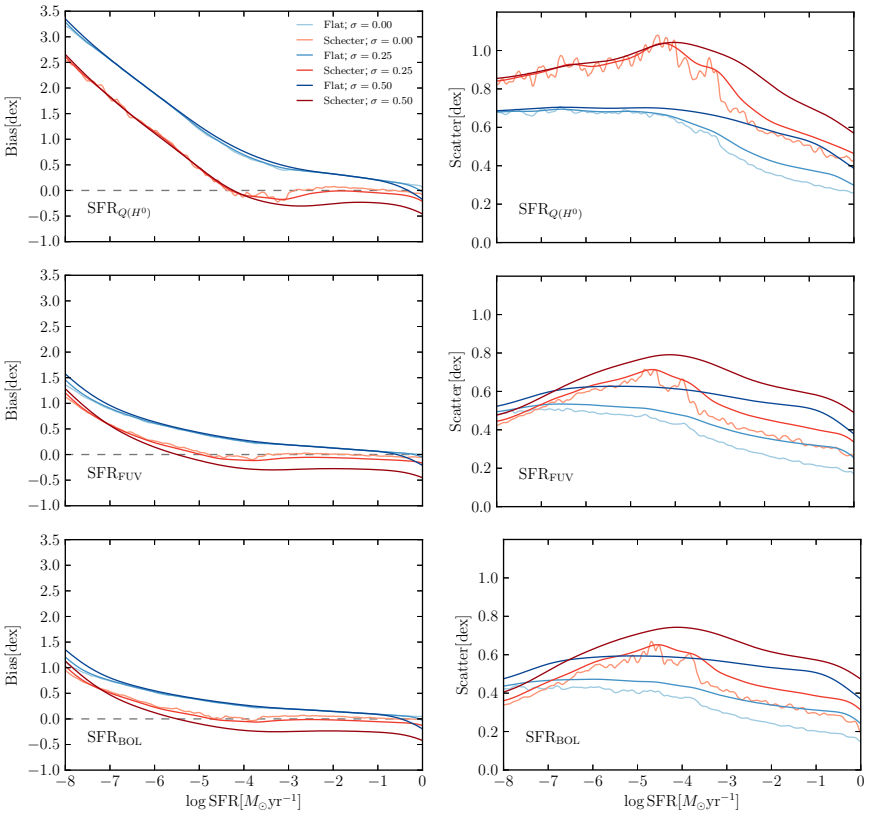

Figure 3.6 Bias (eqt. 3.14) and scatter (eqt. 3.15) due to stochasticity in SFR estimates using the SFIs indicated in each panel. The lower observational error models produce noisier curves because they are averaging over fewer `SLUG` models.

SFIs also have the longest tails (this produces the high value of the bias). We also see the choice of prior has a larger effect in the higher uncertainty observations. This is because there is a bigger dynamic range for the PDF to affect the result. As is always the case, the closer the PDF is to a δ function, the less a prior matters.

We also see that the uncertainty is characteristically largest at $\log \mathrm{SFR} \approx -4$. Two effects contribute to this peak. First, the luminosity, particularly the ionizing luminosity, is dominated by stars with masses $\gtrsim 20\ M_\odot$. For our adopted IMF, these contribute a fraction $f_N \sim 10^{-2.5}$ of stars by number. The expected number of such very massive stars present at any given time is $\langle N \rangle = f_N t_{\mathrm{life}} (\mathrm{SFR}/\langle M \rangle)$, where $\langle M \rangle \sim 1\ M_\odot$ is the mean stellar mass and $t_{\mathrm{life}} \sim 4$ Myr is the lifetime of the very massive stars with which we are concerned. Thus a star formation rate of $\sim 10^{-4}\ M_\odot\ \mathrm{yr}^{-1}$ is the value for which the expected number of very massive stars present at any given time transitions from being $\gtrsim 1$ to $\lesssim 1$, and thus represents something of a maximum in the amount of stochastic flickering.

The second effect is more subtle, and points to a fundamental limitation of our understanding. We adopt a minimum cluster mass of 20 $M_\odot$, and, as can be seen from Figure 3.3, this imposes a minimum star formation rate $\log \mathrm{SFR} \sim -8$ corresponding to the lowest star formation possible with a minimum cluster mass of 20 $M_\odot$. SFRs below this value always produce luminosities of zero in our model. However, this means that the range of possible SFRs for a given observed (non-zero) luminosity has a hard lower limit, and this has the effect of limiting the width of the SFR PDF, and thus the scatter, at the very lowest SFRs. Such a hard edge to star formation is obviously artificial, but it does point out the fact that, at very low SFRs, it is not possible to make a good estimate of the scatter without knowing exactly how star formation and stellar clustering works in regimes where the number of star clusters present at any given time is likely to be zero. Without this knowledge, one cannot calculate the probability that a galaxy with a SFR of, say, $10^{-5}\ M_\odot\ \mathrm{yr}^{-1}$ based on the point mass estimate is actually a galaxy with a true SFR of $10^{-8}\ M_\odot\ \mathrm{yr}^{-1}$ that has just formed a single O star and thus has a temporarily boosted luminosity.

A much more subtle version of this effect, is responsible for the very slight turn-down in bias and scatter that we observe as the SFR approaches 1 $M_\odot\ \mathrm{yr}^{-1}$. For reasons of numerical cost we have not been able to run models with $\log \mathrm{SFR} \gtrsim 0.3$,

and this slightly limits the bias and scatter at the highest SFRs we explore. As is apparent from Figure 3.6, however, the effect is very minor.

3.3.3 Publicly-Available Tools

We caution that, while the summary statistics discussed in the previous section are useful rules of thumb, those attempting a proper statistical analysis of their data should make use of the full PDFs and calculate posterior probability distributions from Equation 3.12. To facilitate such computations, we have made two tools publicly-available at *https://sites.google.com/site/runslug/plots*.

First, we have created an interactive visualization tool; Figure 3.7 shows a screenshot. Its operation is as follows. As discussed above, one may think of our simulations as populating a four-dimensional parameter space (SFR, $\mathrm{SFR}_{Q(H^0)}$, $\mathrm{SFR}_{\mathrm{FUV}}$, $\mathrm{SFR}_{\mathrm{BOL}}$). Either an input theoretical star formation rate SFR, or an observation of one or more of the star formation tracers, picks out a particular part of this parameter space, and therefore restricts the range of values available for the other tracers. The visualization tool allows users to see these effects by selecting a range of values in one more more of the four parameters. The tool then shows the corresponding range in the other parameters. For example, in the screen shot shown in Figure 3.7, a user has selected a range of intrinsic SFRs centered around $\log\ \mathrm{SFR} = -4$ (bottom panel), and the tools is displaying the corresponding range of values for $\mathrm{SFR}_{Q(H^0)}$, $\mathrm{SFR}_{\mathrm{FUV}}$, and $\mathrm{SFR}_{\mathrm{BOL}}$ (top three panels). Versions of the tool are available for both flat and Schecter function priors, and for different clustering fractions (see Section 3.4.3).

Second, we have made available both the full output of the `SLUG` simulations and a set of python scripts to parse them and use them evaluate Equation 3.12 for a specified set of observational constraints. The basic strategy implemented in the code for calculating a $p(\mathrm{SFR} \mid \mathrm{data})$ is

1. Run the script that loads in the 1.8 million galaxy simulations and performs the kernel density estimate.

2. Evaluate the density on a grid of SFI values, weighted by the appropriate prior.

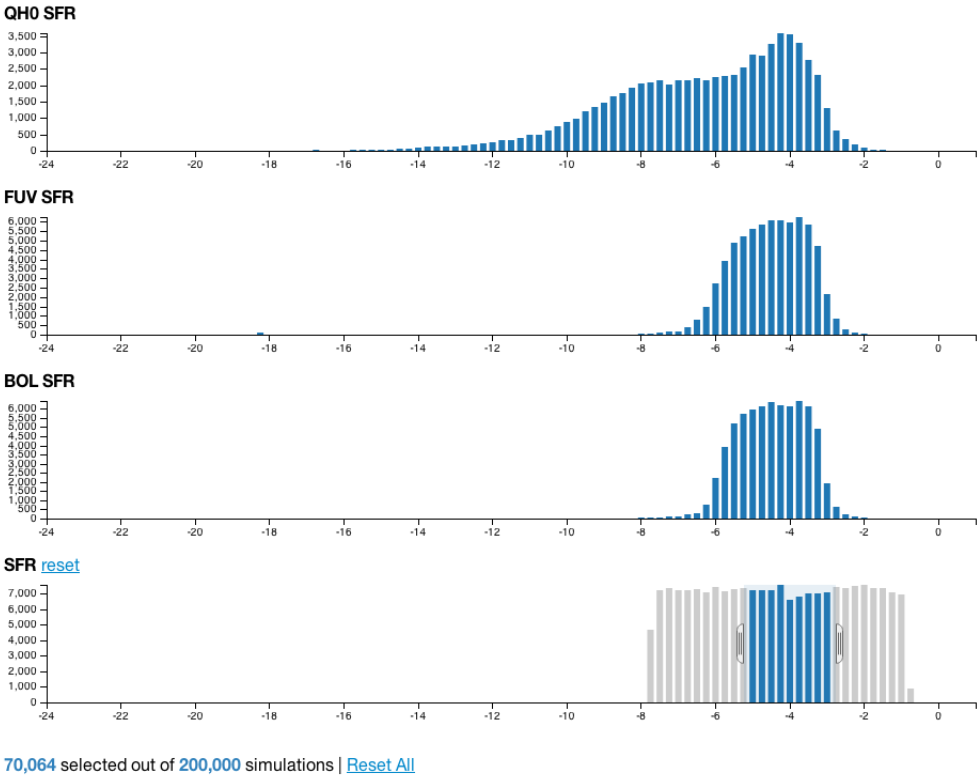

Figure 3.7 Screenshot of interactive data visualization tool for the 4-dimensional parameter space (SFR, $\mathrm{SFR}_{Q(H^0)}$, $\mathrm{SFR}_{\mathrm{FUV}}$, $\mathrm{SFR}_{\mathrm{BOL}}$). Available at *https://sites.google.com/site/runslug/plots*. Selections can be applied to any dimension(s) to show the effects on the others.

3. Weight each point in the above grid by the input observational PDF, $p(\ell \mid \text{data})$. As an example, the posted python code demonstrates how to do this for a Gaussian error bar.

The output is a PDF similar to the one plotted in Figure 3.5. The entire operation should take a few minutes at most, with most of the time spent in step 1, which only needs to occur once for evaluation of an entire dataset. We note that one of the benefits of our approach, and our code, is that we can easily extend to considering the distribution of SFR given a joint set of constraints. Nothing changes in the formalism since we have thus far always been treating ℓ as a vector.

3.4 Discussion

Having discussed at length the quantitative implications of stochasticity for the interpretation of SFIs, in this section we step back and consider some of the broader implications of our results. We also discuss some caveats and cautions.

3.4.1 Star Formation Rate Distributions and the Cosmic Star Formation Rate Budget

We have already alluded to one important implication of our results: because there is both a systematic bias and a scatter in SFR determinations, and because both of these quantities depend systematically on the observed value for the SFI, there is likely to be a similar systematic bias in observational determinations of the distribution of star formation rates in a galaxy population derived using point-mass calibrations. A number of authors have published such determinations based on a variety of SFIs in both the local and high-redshift Universes (to name but a few of many examples, Salim et al. 2007: FUV at $z \sim 0$; Bothwell et al. 2011: FUV plus infrared / bolometric at $z \sim 0$; Fontanot et al. 2012: FUV plus infrared / bolometric at $z \sim 0.4 - 1.2$; Ly et al. 2012: Hα at $z \sim 0.5$; Smit et al. 2012: FUV at $z \sim 4 - 7$; Bauer et al. 2013: Hα at $z \sim 0.05 - 0.3$). Our findings suggest that the results of these surveys may suffer from significant systematic errors, with the extent of the problem depending on the tracer used and on the range of SFR being studied. In particular, faint end slopes may need to be revised, as our results open up the possibility that there may be a non-negligible population of galaxies that have significant SFRs averaged over time, but that are missed in observational surveys simply because they happen to have relatively low UV or ionizing photon luminosities at the instant that the observation is made.

We note that, in setting the prior probability distribution used in our Bayesian analysis, we have relied on these potentially flawed measurements.[10] In principle the proper way to address this issue is via forward modeling. Given a parameterized functional form for the SFR distribution (e.g., a Schecter function), one could use $p(\ell \mid \log \mathrm{SFR})$ to calculate the observed SFI luminosity distribution

[10]This is not a deficiency of our method compared to others, as any non-trivial statistical analysis requires the use of some prior distribution for the star formation rate, either explicitly or implicitly.

that would be expected for a particular choice of parameters describing the SFR, and then adjust those parameters iteratively until the predict SFI luminosity distribution matches the observed one. However, such an approach is beyond the scope of this work, as an accurate forward model would need to be constructed on a survey-by-survey basis, as it would have to fold in uncertainties and errors arising from finite instrumental sensitivity, the color or other cuts used to define the sample, and similar effects.

This issue may also affect determinations of the cosmic star formation rate budget (e.g. Hopkins & Beacom, 2006). These measurements are somewhat less vulnerable to stochasticity than measurements of the SFR distribution, as they necessarily involve averaging over a large number of galaxies and thus averaging out stochasticity (though given the large scatter, the required number of galaxies may be large). If one could in fact observe every Hα photon, for example, emitted in a particular field in a given redshift range, there would be no error from stochasticity as long as the field were large enough to have a bulk SFR larger than $\sim 1\ M_{\odot}$ yr^{-1}. However, in practice measurements of the SFR budget are based on flux-limited samples, and stochasticity can interact with the flux limit by scattering some galaxies with low average SFRs into the sample, while scattering others with higher SFRs out of it. Which of these two effects dominates is a subtle question, since there are more low-luminosity galaxies that could potentially scatter above the flux cut, but the skewness of the PDF is such that galaxies are more likely to be under- than over-luminous for their SFR. Again, rigorous treatment of this issue requires that the study's selection function be analyzed properly with Monte Carlo simulations.

3.4.2 Kennicutt-Schmidt Relations

Another area where luminosity-dependent bias and scatter in SFIs can cause problems is in empirical determinations of the relationship between gas and star formation in galaxies, generically known as Kennicutt-Schmidt relations (Schmidt, 1959; Kennicutt, 1998). Prior to the past decade, such relationships were generally measured as integrated quantities over fairly large spiral galaxies. In the past decade, however, there has been a concerted effort to push these measurements to galaxies

with lower global SFRs (e.g., Lee et al., 2009; Boselli et al., 2009; Meurer et al., 2009), and to ever-smaller spatial scales within large galaxies (e.g., Wong & Blitz, 2002; Kennicutt et al., 2007; Bigiel et al., 2008, 2010; Schruba et al., 2010; Onodera et al., 2010; Bolatto et al., 2011; Calzetti et al., 2012; Momose et al., 2013; Leroy et al., 2013). These efforts have pushed the data into realms of ever-lower absolute SFR, and thus greater vulnerability to stochasticity.

To take one example, for the lowest gas surface density bin in the sample of Bigiel et al. (2010), the median SFR surface density is inferred to be a bit over 10^{-6} $M_\odot$ yr^{-1} kpc^{-2}. For the mean pixel size of 600 pc used in the study, this corresponds to $< 10^{-6}$ $M_\odot$ yr^{-1}. The study uses FUV as its SFI of choice, and consulting Figure 3.6, we see that, for a Schecter function prior and assuming negligible observational errors, we expect a scatter of ~ 0.5 dex from stochasticity alone. If we adopt a flat prior distribution of SFRs (perhaps reasonable inside a galaxy), we also expect a similar amount of bias. This will obviously affect the mean relation that one infers between gas and SFR, and it should be accounted for when fitting the observations. Qualitatively, the net effect of stochasticity is likely to be that the inferred relationship between SFR and gas surface density is too steep at the lowest SFRs (due to the bias) and that the inferred scatter will be larger than the true one (due to the extra scatter in the SFI-SFR relation imposed by the stochasticity).

3.4.3 Sensitivity to Parameter Choices

We end this discussion with a caution regarding the sensitivity of our results to some of the parameters we have chosen in our `SLUG` simulations. The results obviously depend to some extent on the choice of stellar evolutionary tracks and atmosphere models, but this is true even in the absence of stochasticity. The parameters that are unique to our stochastic models are those that describe how stars are clustered. A full analysis of the effects of varying the cluster mass function's minimum and maximum mass, as well as its power law index and the total fraction of stars formed in clusters, is well beyond the scope of this paper. However, to explore the effects of clustering to gain some intuition, we focus on a single parameter: the total fraction of stars formed in clusters f_c.[11] This is likely the single

[11] An important note on nomenclature: some authors whose interest lies primarily in stellar dynamics (e.g., Portegies Zwart et al., 2010) limit the definition of star clusters to include only those

most important parameter. Our default choice is $f_c = 1$. This is motivated by the observation that, in the Milky Way, most star formation occurs in clusters (Lada & Lada, 2003), and by the result that models with $f_c = 1$ provide an excellent match to the observed distribution of Hα to FUV ratios in local dwarf galaxies (Fumagalli et al., 2011b). However, to investigate how our results would change if we alter this parameter, we run roughly $15,000$ unclustered models ($f_c = 0$) and $25,000$ with $f_c = 0.5$. These models are uniformly distributed in $\log \mathrm{SFR}$ between -4 and -2.

Figure 3.8 shows the PDFs of offset between SFI and true SFR that we obtain from the unclustered and reduced clustering runs; it should be compared with Figure 3.4 for our fiducial case. The comparison indicates that reducing the clustering can significantly reduce the spread of SFI values produced at fixed SFR. This will correspondingly significantly decrease the scatter in the inferred SFR PDFs.

This result implies that, at least at low SFRs, it is crucial to understand the clustering properties of star formation in order to do something as simple as inferring a star formation rate. A more accurate determination of stellar clustering parameters, and whether they vary with galactic environment, is therefore urgently needed. Our fiducial parameters are reasonable first approximations based on empirical constraints from local galaxies, but if clustering parameters vary systematically with galaxy properties, the effects of stochasticity on inferences of the SFR may as well.

stellar structures that are gravitationally-bound and dynamically-relaxed. These are distinguished from associations – collections of stars that are born in spatial and temporal proximity to one another, but need not be bound or relaxed. Since we care only about the temporal correlation of star formation, and not about the dynamical evolution of the structures in which the stars form, we are interested in a much more expansive definition of clustering, one that includes both clusters and associations. Thus our f_c parameter is not directly comparable to the parameter Γ that is sometimes introduced to denote the fraction of star formation that occurs in structures the remain bound after the transition from gas-dominated to gas-free evolution (e.g., Bastian, 2008a).

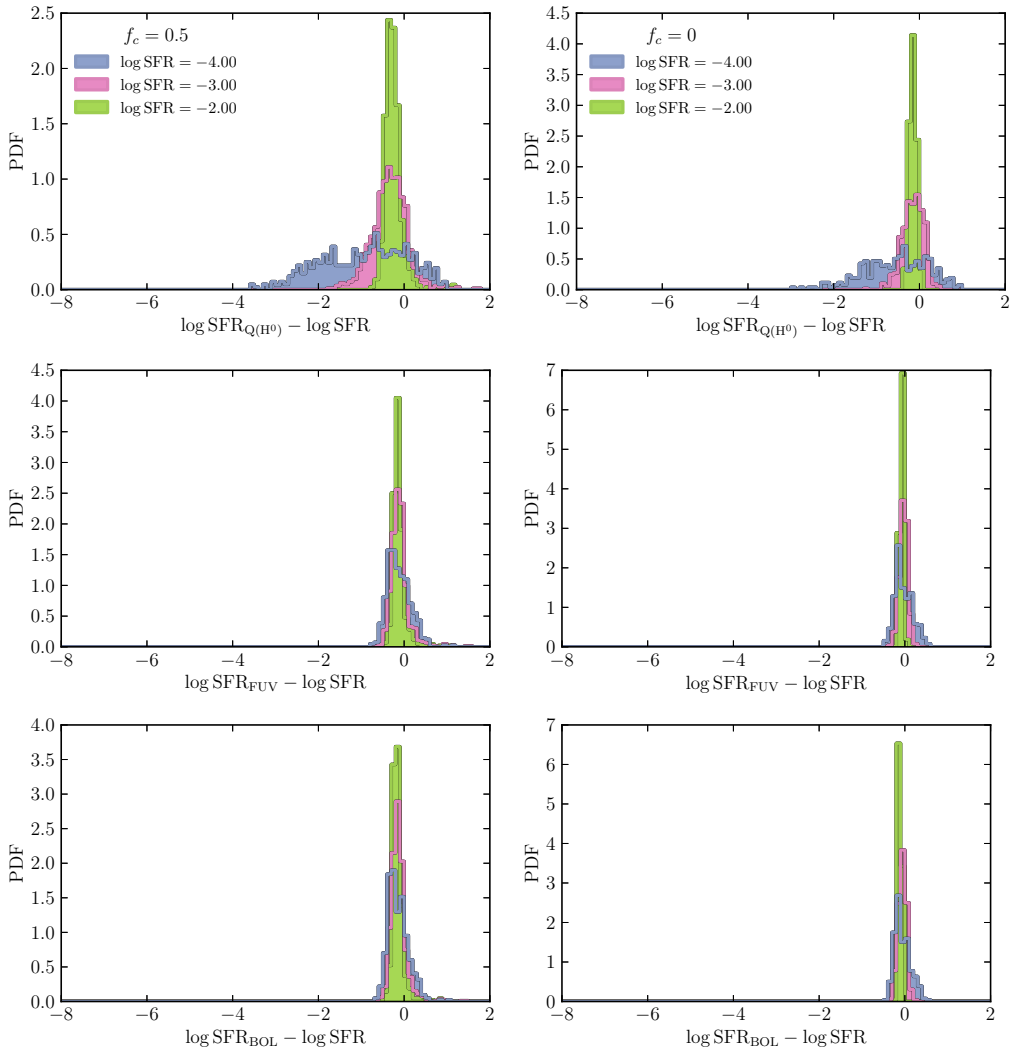

Figure 3.8 Same as Figure 3.4 but for $f_c = 0.5$ (left) and $f_c = 0$ (right).

3.5 Summary

While star formation in galaxies is often imagined as a continuous, ongoing process, observations tell us that the actual distribution of stellar ages is highly stochastic, with stars mostly forming in discrete bursts of finite size. At sufficiently high star formation rates (SFRs), the overall process of star formation in a galaxy consists of many such bursts, and the continuous approximation is reasonable. In this paper, we use the Stochastically Lighting Up Galaxies (SLUG) code to investigate what happens at lower SFRs when this approximation begins to break down, with particular attention to how this breakdown affects our ability to infer the underling SFR using a variety of star formation indicators (SFIs).

We show that the generic effect of stochasticity is to produce a broad probability distribution function (PDF) for SFI luminosity a fixed SFR. The breadth of the PDF depends on both the SFI being used and on the true SFR. We then devote the bulk of the paper to understanding the implications of this spread in SFI at fixed SFR for the inverse problem of inferring the true SFR given an observed SFI. We derive an analytic expression for the PDF of true SFR given a set of observational constraints, and provide software to evaluate this PDF using our simulation results and a set of user-specified observational constraints.

Using this formalism, we show that the process of inferring the SFR from an observed SFI is subject to scatter, and, more worryingly bias, meaning that the process of simply converting between SFI and SFR using the standard calibrations that apply at higher SFRs is likely to lead to systematic errors when used at low SFRs. The strength of the bias and scatter depend on both the observed values of the SFI and on its observational uncertainty, and on the choice of SFI. Ionization-based SFIs such as Hα emission in particularly can be problematic due to the very short timescales over which they average; for such indicators, a scatter of several tenths of a dex is expected even at inferred SFRs as high as $\sim 1\ M_\odot\ \mathrm{yr}^{-1}$. Even for indicators much less subject to scatter such as FUV luminosity, for measurements with non-trivial observational uncertainty, biases of up to ~ 0.5 dex are possible.

Finally, we discuss the implications of these results for efforts to construct "luminosity functions" of star formation rate, for estimates of the cosmic star formation rate budget, and for inferences of the Kennicutt-Schmidt Law relating gas

content to star formation rate. The Legacy Extragalactic UV Survey (LEGUS; Calzetti et al., 2014, in preparation) will provide a valuable data set for this type of analysis.

Acknowledgments

R.L.dS. and M.R.K. acknowledge support from NASA through Hubble Award Numbers HST-AR-13256.01-A and HST-GO-13364.26-A 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 R.L.dS was supported by the National Science Foundation Graduate Research Fellowship. Support for M.F. 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. M.R.K. acknowledges support from NSF grant AST-0955300, and from NASA ATP grant NNX13AB84G.

Chapter 4

Cluster Order Statistics

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 Nth most luminous cluster in a population. Our method is flexible and requires few assumptions, allowing for parameterized 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 (`CLOC`), which we have made publicly available.

4.1 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 et al., 2011c; da Silva et al., 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) powerlaw $dN/dM \propto M^{\beta}$ with $\beta \approx -2$ over a wide mass range (e.g. Zhang & Fall, 1999b; Bik et al., 2003; Boutloukos & Lamers, 2003; Fall, 2006; Fall et al., 2009c; Chandar et al., 2010b; Fall & Chandar, 2012; Bastian et al., 2012a,b). There is more controversy over the age distribution, mostly arising from issues of how the samples are selected. More inclusive cluster catalogs constructed to include all objects above a surface brightness threshold tend to show powerlaw age distributions $dN/dt \propto t^{\gamma}$ with $\gamma \approx -0.9$ (Fall et al., 2005, 2009c; Chandar et al., 2010b; Fall & Chandar, 2012). If one imposes additional selection criteria based on morphology or crowding, this removes many young clusters from the sample, yielding a distribution that can still be approximated by a powerlaw, but with a significantly shallower index, $\gamma \approx 0$ (Gieles et al., 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, 2008b), and exploiting such large but lower-quality data sets is the only feasible means to detect whether cluster mass or age distributions deviate from powerlaw behavior 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 this 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, 2008b; 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. Lada & Lada (2003)'s compilation of clusters just within 2 kpc of the Sun includes ~ 100 entries, and this survey covers only $\sim 1-3\%$ of the Milky Way's star-forming disk, 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 2008b) 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 parameterized 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, nor is it 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., 2013, in preparation), we use this method to revisit the question of whether the observed relationship between star formation rate 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 4.2 describes our model and its derivation. In Section 4.3 we describe the publicly available code that implements this model, and present comparisons between it and a full Monte Carlo method. In Section 4.4 we use our formalism to explore how the various parameters that go into the cluster luminosity function affects it shape in order to gain insight about what sorts of observations can be used to constrain star cluster formation.

4.2 The Model

4.2.1 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 sub-galactic volume.[1] We normalize this and all other PDFs in this paper to unity, i.e., $\int \phi_k(L)\, dL = 1$. We define $\Phi_k(L) = \int_0^L \phi_k(L')\, dL'$ 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 $\phi_k(L)\, dL$ is the probability that the kth most luminous cluster has a luminosity in the infinitesimal range L to $L + dL$.

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 parameterized 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 4.6, 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.

[1] Note that our convention in defining the 1st 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 1st 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 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 there k or more clusters present. We argue in Appendix 4.6 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

$$\langle N(>L)\rangle = \langle N\rangle(1-\Phi(L)), \tag{4.1}$$

where $\Phi(L) = \int_0^L \phi(L')\,dL'$ 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,

$$\begin{aligned} P_m(>L) &= \frac{1}{m!}\langle N(>L)\rangle^m e^{-\langle N(>L)\rangle} &(4.2)\\ &= \frac{e^{-\langle N\rangle}}{m!}\langle N\rangle^m[1-\Phi(L)]^m e^{\langle N\rangle\Phi(L)}. &(4.3)\end{aligned}$$

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 $L+dL$, 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

$$\begin{aligned} \phi_k(L) &= \frac{\Gamma(k,\langle N\rangle)}{\Gamma(k)}\delta(L) + \langle N\rangle P_{k-1}(>L)\phi(L) &(4.4)\\ &= \frac{\Gamma(k,\langle N\rangle)}{\Gamma(k)}\delta(L) \\ &\quad + \langle N\rangle\frac{\langle N\rangle^{k-1}[1-\Phi(L)]^{k-1}}{(k-1)!}\phi(L), &(4.5)\end{aligned}$$

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 $\int \phi_k(L)\, dL = 1$. We derive both coefficients by alternative means in Appendix 4.6. Note that the coefficient of the δ-function, $\Gamma(k, \langle N \rangle)/\Gamma(k)$, goes to zero extremely rapidly as $k/\langle N \rangle \rightarrow 0$. 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 that the 2nd 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

$$\Phi_k(L) = \sum_{m=0}^{k-1} P_m(> L) \tag{4.6}$$

$$= e^{-\langle N \rangle} e^{\langle N \rangle \Phi(L)} \sum_{m=0}^{k-1} \frac{\langle N \rangle^m [1 - \Phi(L)]^m}{m!}. \tag{4.7}$$

Note that $\Phi_k(L)$ remains finite in the limit $L \rightarrow 0$, even if $\Phi(L)$ is identically zero below some finite minimum L. This behavior 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.

4.2.2 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 star formation rate 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

[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-

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_c.[3] We normalize the mass and age distributions such that $\int \psi(M)\,dM = \int \chi(t)\,dt = 1$, 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

$$\langle N_{\rm form} \rangle = \frac{\dot{M}_* \Delta t}{\langle M \rangle} f_c, \tag{4.8}$$

where $\langle M \rangle = \int M \psi(M)\,dM$ 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 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

$$\langle N \rangle = \frac{\dot{M} \Delta t}{\langle M \rangle} f_c \left(\frac{1}{\Delta t} \int_{t_{\rm min}}^{t_{\rm max}} P_{\rm surv}(t)\,dt \right) \equiv \frac{\dot{M}_* \Delta t}{\langle M \rangle} \mathcal{F}_c \tag{4.9}$$

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 $\mathcal{F}_c$ that we have defined is the fraction of all stars in clusters, averaged over the stellar age range under consideration. For a constant star formation rate, 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

$$\mathcal{F}_c = \frac{f_c}{\chi(0)\Delta t} \int_{t_{\rm min}}^{t_{\rm max}} \chi(t)\,dt. \tag{4.10}$$

limited sample. Second, 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. (2009c), Chandar et al. (2010b), and Fall & Chandar (2012). Fortunately, even in those papers where the authors do report that the mass and age distributions are not independent, the co-variance 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_c is subtly different from the cluster formation efficiency Γ defined by some authors (e.g. Bastian, 2008b), because Γ refers to the fraction of stars formed as part of gravitationally-bound clusters. In contrast, f_c depends only on the observational criteria used to define clusters when selecting them in an observed galaxy. Thus f_c and Γ are identical only if the observational selection criteria pick out all gravitationally bound structures, and only such structures.

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 powerlaws. Specifically, we adopt

$$\psi(M) = \begin{cases} AM^{\beta}, & M_{\rm min} < M < M_{\rm max} \\ 0, & \text{otherwise} \end{cases} \tag{4.11}$$

and

$$\chi(t) = \begin{cases} B, & t < t_0 \\ B(t/t_0)^{\gamma}, & t_0 \leq t < t_1 \\ 0, & t \geq t_1. \end{cases} \tag{4.12}$$

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

$$\begin{aligned} \frac{1}{A} &= \int_{M_{\rm min}}^{M_{\rm max}} M^{\beta}\, dM \\ &= \begin{cases} (M_{\rm max}^{\beta+1} - M_{\rm min}^{\beta+1})/(\beta+1), & \beta \neq -1 \\ \ln(M_{\rm max}/M_{\rm min}), & \beta = -1 \end{cases}, \end{aligned} \tag{4.13}$$

and

$$\begin{aligned} \frac{1}{B} &= t_0 + \int_{t_0}^{t_1} \left(\frac{t}{t_0}\right)^{\gamma} dt \\ &= t_0 + \begin{cases} (t_1^{\gamma+1} - t_0^{\gamma+1})/(\gamma+1), & \gamma \neq -1 \\ \ln(t_1/t_0), & \gamma = -1 \end{cases}. \end{aligned} \tag{4.14}$$

The functional forms for both the mass and age distributions are well-motivated by observations. As discussed in the Introduction, there is an observational consensus

that the mass function is well fit by a (possibly truncated) powerlaw with index $\beta \approx -2$. There is dispute in the observational community about the age distribution, but all groups agree that a powerlaw 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

$$\langle M \rangle = A \times \begin{cases} \left(M_{\rm max}^{\beta+2} - M_{\rm min}^{\beta+2} \right) / (\beta+2), & \beta \neq -2 \\ \ln(M_{\rm max}/M_{\rm min}), & \beta = -2 \end{cases}. \tag{4.15}$$

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

$$\mathcal{F}_c = f_c \frac{t_0}{\Delta t} \times \begin{cases} \left[(t_{\rm max}/t_0)^{\gamma+1} - (t_{\rm min}/t_0)^{\gamma+1} \right] / (\gamma+1), & \gamma \neq -1 \\ \ln(t_{\rm max}/t_{\rm min}), & \gamma = -1 \end{cases}. \tag{4.16}$$

We could substitute this into equation (4.9) to obtain an explicit form for $\langle N \rangle$, but this would simply replace $\mathcal{F}_c$ with f_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 $\mathcal{F}_c$, we will use $\mathcal{F}_c$ as the variable of interest through the rest of this work, keeping in mind that it is related to the physical quantities f_c and t_0 via equation (4.16).

4.2.3 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 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 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 star formation rates. 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 4.7 we fit for $\Upsilon(t)$ in V band, and obtain

$$\Upsilon(t) = \Upsilon_* \left(\frac{t}{10\ \mathrm{Myr}}\right)^\zeta \tag{4.17}$$

with $\zeta = 0.688$ and $\Upsilon_* = 8.3 \times 10^{-21}\ M_\odot \left(\mathrm{erg\ s^{-1}\ Hz^{-1}}\right)^{-1}$.

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

$$\theta(\Upsilon) \propto \chi(t) \left|\frac{d\Upsilon}{dt}\right|^{-1}. \tag{4.18}$$

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

$$\phi_{\mathrm{in}}(L_{\mathrm{in}}) = \int_0^\infty \psi(\Upsilon L_{\mathrm{in}})\theta(\Upsilon)\Upsilon\, d\Upsilon. \tag{4.19}$$

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

$$\begin{aligned}\phi_{\rm in}(\log L_{\rm in}) &\propto \psi(\log M) * \xi(-\log\Upsilon) \\ &\equiv \int_{-\infty}^{\infty} \psi(\log L_{\rm in} - \log\Upsilon)\xi(-\log\Upsilon)\, d\log\Upsilon, \end{aligned} \tag{4.20}$$

where $*$ denotes convolution. We defer actual calculation of this convolution to Appendix 4.8, since it is conceptually straightforward but algebraically tedious. The result for $\phi_{\rm in}(\log L_{\rm in})$ is given by equation (4.58), and the corresponding CDF $\Phi_{\rm in}(\log L_{\rm in})$ by equations (4.59) and (4.64).

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 $e^{-\tau}$. 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,

$$\begin{aligned}\phi(\log L) &= \phi_{\rm in}(\log L_{\rm in}) * \eta(-\tau) \\ &= \int_{-\infty}^{\infty} \phi_{\rm in}(\log L_{\rm in})\eta(\log L_{\rm in} - \log L)\, d\log L_{\rm in} \end{aligned} \tag{4.21}$$

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 4.9; the final expressions for $\phi(\log L)$ and

$\Phi(\log L)$ are given by equations (4.78) and (4.81), 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 (4.5) and (4.7), using the expected number of clusters $\langle N \rangle$ given by equation (4.9) and the PDF and CDF of luminosity for individual clusters given by equations (4.78) and (4.81).

4.2.4 Observational Uncertainties

We now add one final element to our model, which is that neither star formation rates 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 ϵ_S and ϵ_L as the uncertainties on the star formation rate 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 initial mass function, and ambiguities in the choice of timescale over which the SFR is averaged, among others – see Kennicutt & Evans (2012b) for a recent review. Typical errors are again ~ 0.5 dex.

Table 4.1 Fiducial Parameter Values

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

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 star formation rates. Under this assumption, we can write the distribution of observed luminosities $L_{\rm obs}$ in a galaxy with observed star formation rate $\dot{M}_{*,\rm obs}$ as simply

$$\begin{aligned}\phi_{k,\rm obs}(\log L_{\rm obs} \mid \log \dot{M}_{*,\rm obs}) = \frac{1}{2\pi\epsilon_S\epsilon_L} \times \\ \iint \phi_k(\log L \mid \log \dot{M}_*) \times \\ \exp\left\{-\frac{[\log(\dot{M}_*/\dot{M}_{*,\rm obs})]^2}{2\epsilon_S^2} - \frac{[\log(L/L_{\rm obs})]^2}{2\epsilon_L^2}\right\} \\ d\log\dot{M}_*\, d\log L,\end{aligned} \tag{4.22}$$

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.

4.2.5 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 star

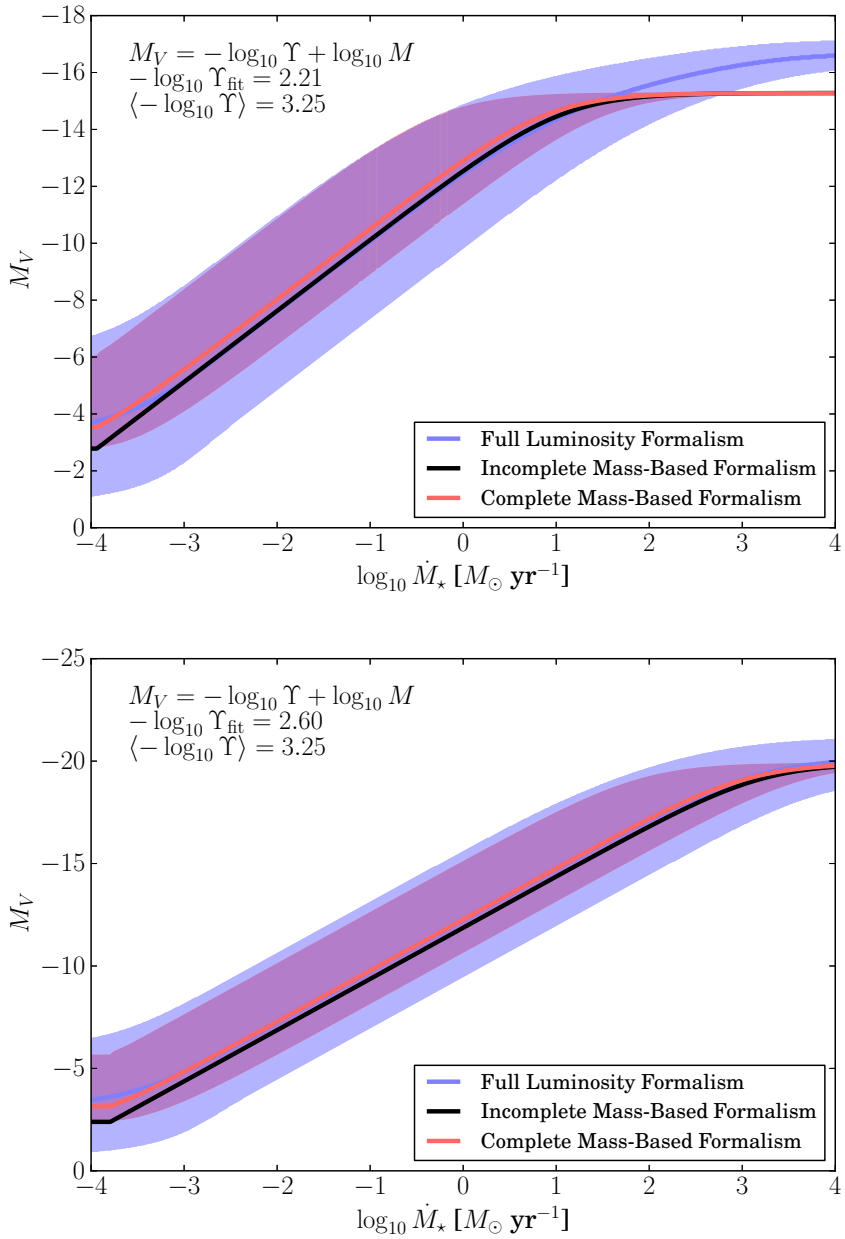

Figure 4.1 Comparison of the first order statistic $\phi_1(L)$ as a function of star formation rate $\dot{M}_*$ for differing levels of model complexity. Black lines ("incomplete mass-based formalism") represent the expectation value of the most luminous cluster determined by computing the maximum mass from equation (4.25) and then applying a fixed mass-to-light ratio $\Upsilon_{\rm fit}$. Red lines ("complete mass-based formalism") show the median, and red bands the $5-95$ percentile range, for the 1st order statistic computed using equations (4.5) and (4.7) 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 4.23). 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 (4.5), (4.7), and (4.78). The parameters used for the computation are given in Table 4.1 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.

formation rate $\dot{M}_*$. 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

$$\phi(L) = \psi(M/\Upsilon_{\rm fit})\Upsilon_{\rm fit} \tag{4.23}$$

before using equations (4.5) and (4.7) 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 versus $\dot{M}_*$ as computed via the two formalisms. This approach amounts to ignoring the scatter in the relationship between cluster luminosity statistics and star formation rate 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.,

$$1 = N \int_{\langle M_1 \rangle}^{M_{\rm max}} A M^\beta \, dM. \tag{4.24}$$

In this case we have

$$\langle M_1 \rangle = \left(M_{\rm max}^{\beta+1} - \frac{1}{\langle N \rangle A} \right)^{1/(\beta+1)}, \tag{4.25}$$

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 Figure 4.1; the parameters used for the model in the lower panel are those given in Table 4.1, 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 $M_{\max}$ using the observed relationship between $\dot{M}_*$ and the luminosity of the most luminous cluster (Weidner et al., 2004b; Bastian, 2008b), 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 (2008b), who found using Monte Carlo simulations that adopting a fixed mass-to-light ratio is a poor approximation.

4.3 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 star formation rate $\dot{M}_*$ (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 ($t_{\min}$ and $t_{\max}$), the parameters of the initial cluster mass function ($M_{\min}$, $M_{\max}$, and β), the minimum and maximum amounts of dust extinction to use (τ_0 and τ_1), the parameters describing the cluster age distribution ($\mathcal{F}_c$ and γ), the parameters describing the observational error (ϵ_S and ϵ_L), and the parameters describing the time evolution of the light to mass ratio (Υ_* 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 4.2. 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.

Table 4.2 Description of Software Outputs

Variable Name	Description
x	the x-array for luminosity arrays in units of $\ln\left(\text{erg s}^{-1}\text{ Hz}^{-1}\right)$
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 $M_\odot$ 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.

4.3.1 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 4.17). 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 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 $M_{\min} = 500\ M_\odot$, $M_{\max} = 10^9\ M_\odot$, $\beta = -2$, $t_{\min} = 10^7$ yr, $t_{\max} = 10^9$ yr, $\mathcal{F}_c = 0.01$, $\gamma = -0.9$, $\tau_0 = 0$ $\tau_1 = 1$, and $\dot{M}_* = 0.1\ M_\odot$ yr^{-1}. From these parameters, we compute the mean cluster mass $\langle M \rangle = 7254\ M_\odot$ and the expected number of clusters $\langle N \rangle = 136.5$ from equations (4.9) and (4.15). We then create a sample of clusters via the following algorithm:

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

2. For each cluster, draw a mass M from the initial cluster mass function $\psi(M)$, 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.

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

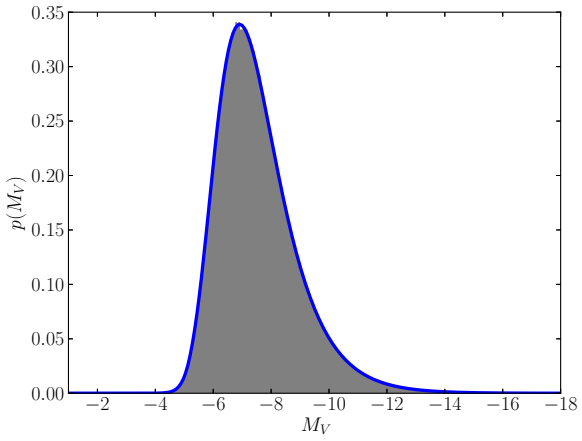

Figure 4.2 Comparison of the analytic prediction of the PDF of the most luminous cluster as computed by `CLOC` (blue curve) and the result of 10^6 Monte Carlo realizations of a cluster population (grey histogram). The input parameters for this test are $M_{\rm min} = 500\ M_\odot$, $M_{\rm max} = 10^9\ M_\odot$, $\beta = -2$, $t_{\rm min} = 10^7$ yr, $t_{\rm max} = 10^9$ yr, $\mathcal{F}_c = 0.01$, $\gamma = -0.9$, $\tau_0 = 0$ $\tau_1 = 1$, and $\dot{M}_* = 0.1\ M_\odot\ {\rm yr}^{-1}$.

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 $\epsilon_S = \epsilon_L = 0$, i.e., assuming that there is no observational uncertainty on either the star formation rate 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 Figure 4.2. 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.

4.3.2 Comparison to `SLUG`

We next compare `CLOC` to the Monte Carlo code `SLUG` (da Silva et al., 2012; Fumagalli et al., 2011a). 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

initial cluster mass function 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 powerlaw 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 $\Delta t_{\rm SLUG}$, a star formation rate $\dot{M}_*$, 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 star formation rate 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 timescales that are long compared to the mean time between events. The behavior 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 catalogs 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 $t_{\rm min} = 10$ Myr and $t_{\rm max} = 1$ Gyr. We use the default SLUG prescription for cluster formation disruption, which amounts to $f_c = 1$, $t_0 = 1$ Myr and $\gamma = -1$, and from these values, we compute $\mathcal{F}_c = 0.0047$ using equation (4.16). We then run both CLOC and SLUG for six cases: we use star formation rates $\dot{M}_* = 10^{-3}$, 10^{-2}, and 10^{-1} $M_\odot$ yr^{-1}, and maximum cluster masses $M_{\rm max} = 10^5$ and 10^9 $M_\odot$. We do not use any dust extinction for this test, as SLUG does not include any. All other parameters are as specified in Table 4.1. 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 $M_{\rm max} = 10^9$ $M_\odot$, and at an age of 1 Gyr for the $M_{\rm max} = 10^5$ $M_\odot$ runs. Note that we do not consider star formation rates higher that 10^{-1} $M_\odot$ yr^{-1} due to issues of computational cost: performing even 1000 SLUG runs at $\dot{M}_* = 10^{-1}$ $M_\odot$ yr^{-1} requires of order a CPU-day, and the computational cost is linear in both the number of realizations and the star formation rate. However, the vast majority of the observational sample is at higher SFRs, illustrating the difficulty of using Monte Carlo methods to analyze the observations.

We show the results of a comparison between CLOC and SLUG in Figure 4.3. We see that the agreement between the two codes is generally quite good, but that there are some important differences. First focus on the left column, showing the models with $M_{\rm max} = 10^5$ $M_\odot$ and $\Delta t_{\rm SLUG} = 1$ 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} $M_\odot$ yr^{-1}, the difference between the two codes is minimal. However, at $\dot{M}_* = 10^{-3}$ $M_\odot$ 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 $\sim 10^{3.5}$ $M_\odot$ 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 $\sim 10^{-3} - 10^{-2}$ $M_\odot$ 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 column, which uses $M_{\rm max} = 10^9$ $M_\odot$ and $\Delta t_{\rm SLUG} =$ 100 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 $M_{\rm max} = 10^9$ $M_\odot$, no cluster of that mass can ever be created in the `SLUG` runs, because $\dot{M}_* \Delta t_{\rm SLUG} = 10^6 - 10^8$ $M_\odot$. Thus, even if `SLUG` does draw a cluster close to 10^9 $M_\odot$ 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 $M_\odot$. We have chosen this extreme case intentionally, to show the importance of ICMF sampling effects. At $\dot{M}_* = 10^{-3}$ $M_\odot$, 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 $\dot{M}_* = 10^{-2}$ $M_\odot$ 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 star formation rate reaches $\dot{M}_* = 10^{-1}$ $M_\odot$ yr^{-1} do `CLOC` and `SLUG` agree well. Clearly at low star formation rates, `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 $\dot{M}_* \ll 10^{-2}$ $M_\odot$ 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 $\sim 10^{-2}$ $M_\odot$ 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 a star formation rate-dependent offset of $\sim 1 - 2$ mag level in

the predicted luminosity PDF. The correct physical answer in this regime is unclear.

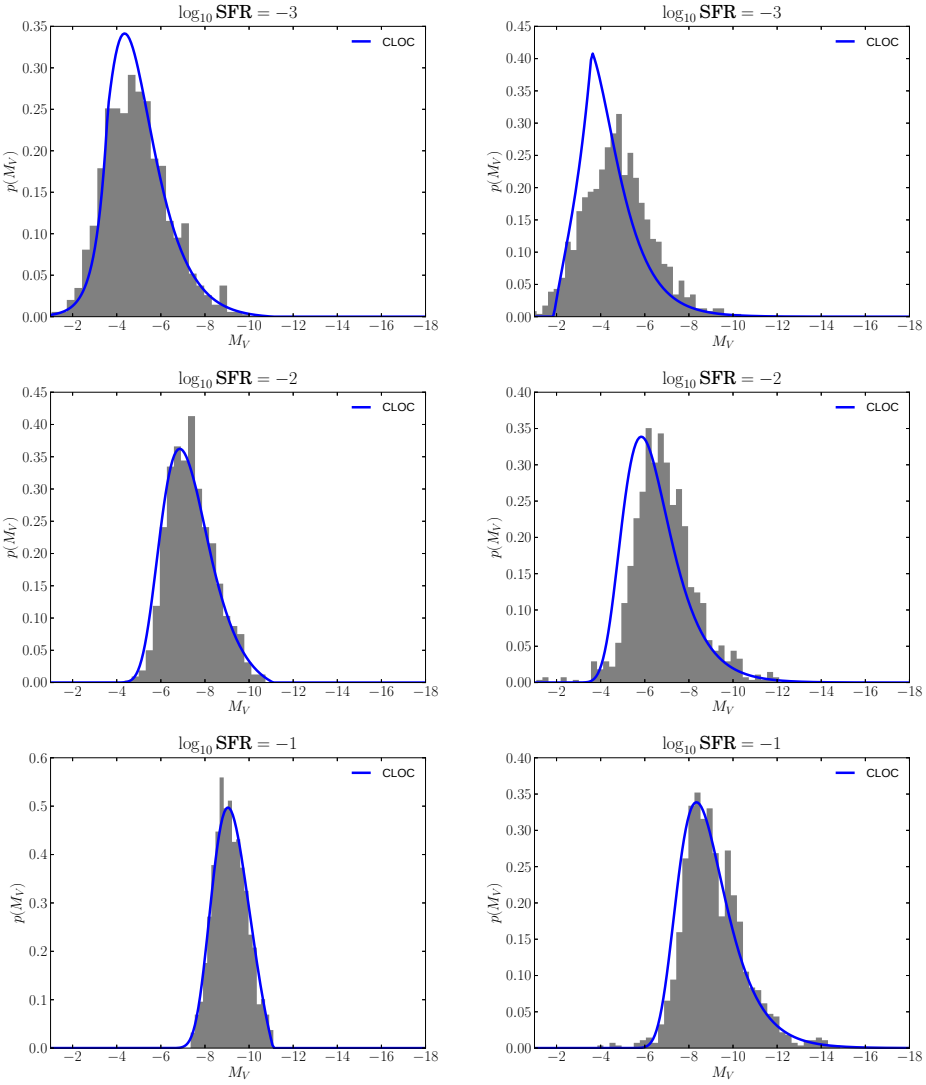

Figure 4.3 Comparison of the PDFs of the most luminous cluster as computed analytically by `CLOC` (blue curve) and via Monte Carlo sampling by `SLUG` (gray histogram), following the procedure outlined in the main text. The left column uses $M_{\rm max} = 10^5\ M_\odot$ and a run time of 1 Gyr in `SLUG`, while the right column uses $M_{\rm max} = 10^9\ M_\odot$ and a run time of 100 Myr. The star formation rates used are 10^{-3}, 10^{-2}, and $10^{-1}\ M_\odot\ {\rm yr}^{-1}$, in the top, middle, and bottom rows, respectively.

4.4 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 star formation rate 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, 2008b). To that end, we consider a fiducial model whose parameters are given in table 4.1, and we then systematically vary the parameters. The results of this experiment are shown in figures 4.4 and 4.5 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.

4.4.1 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 $\mathcal{F}_c$, and the initial cluster mass function parameters $M_{\rm min}$, $M_{\rm max}$, and β. The effects of the first factor, $\mathcal{F}_c$, are simple. (We remind the reader again that $\mathcal{F}_c$ 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 $\mathcal{F}_c$ simply translates the observed relation left or right, such that $\mathcal{F}_c = 0.1$ with a given star formation rate $\dot{M}_*$ is fully equivalent to having $\mathcal{F}_c = 1$ and a star formation rate of $0.1\dot{M}_*$. This can be seen in the first row of Figure 4.4. The value of $\mathcal{F}_c$ also has another effect. In the upper panel of Figure 4.4, notice that 5% confidence contour in the low $\mathcal{F}_c$ 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 $\mathcal{F}_c$ 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

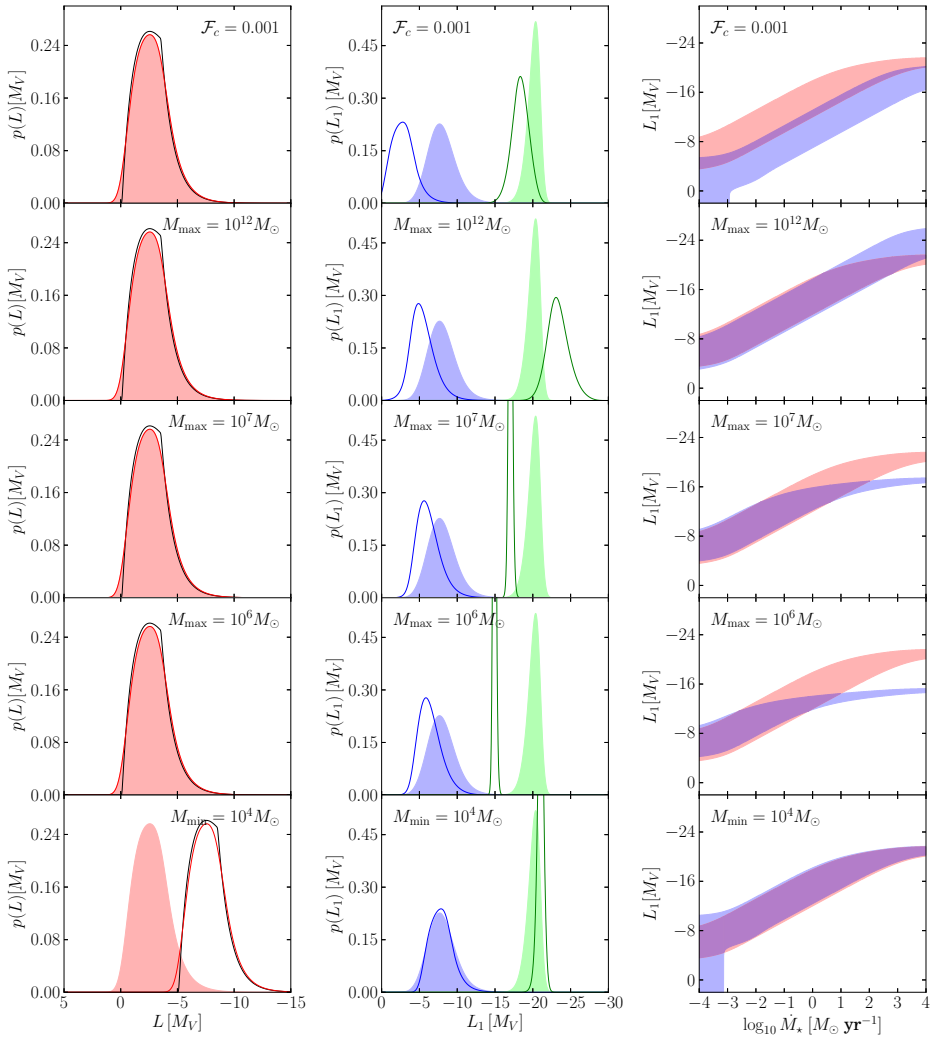

Figure 4.4 (*left*) the PDF of the luminosity of a single cluster before observational error is applied. The filled region corresponds to the fiducial model, while the solid red line denotes the PDF for that results when the change for each corresponding row is applied. The black line corresponds to the model before observational errors are included. (*middle*) PDFs of the most luminous cluster for star formation rates $\log_{10} \dot{M}_\star = -2.75$ (blue) and 2.75 (green). 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.

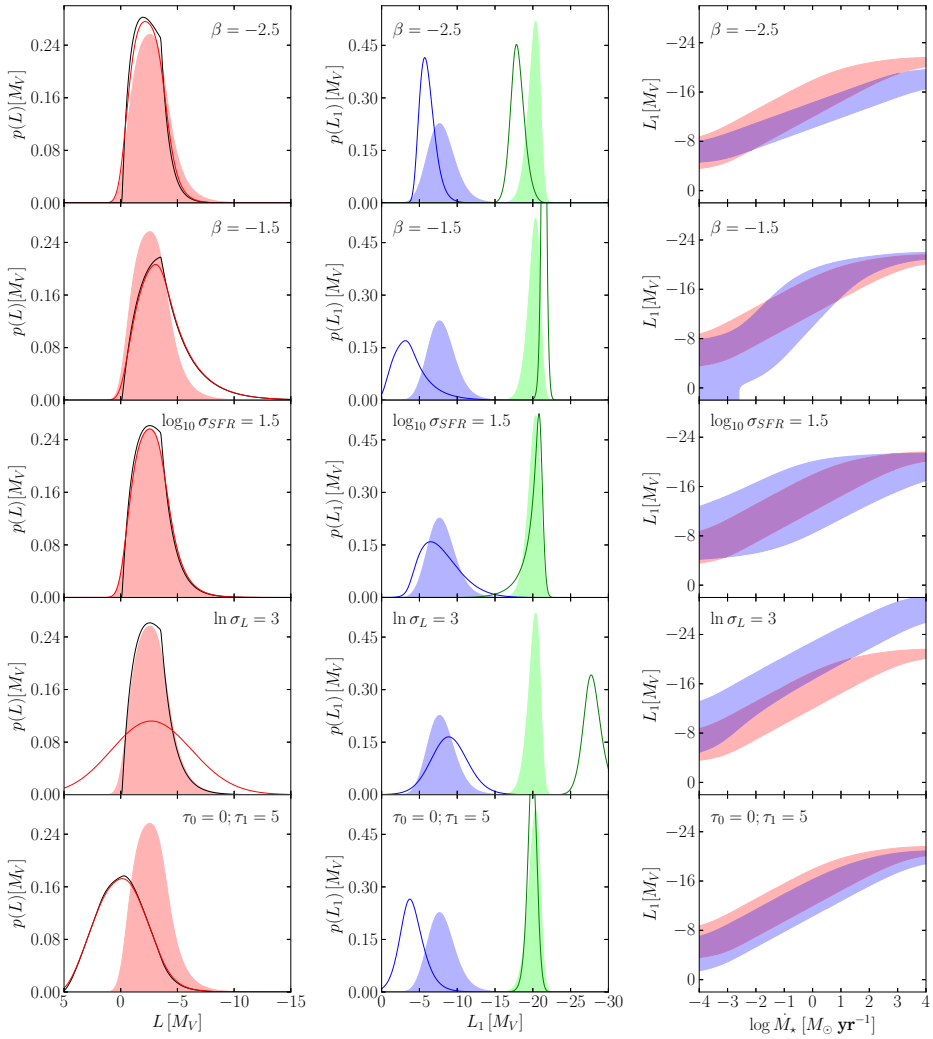

Figure 4.5 Same as figure 4.4 but with different parameters varied.

at zero luminosity, corresponding to the δ-function term in equation (4.5). When the prefactor on this term, $\Gamma(1, \langle N \rangle)/\Gamma(1)$, exceeds 0.05, the 5% confidence contour must encompass zero luminosity. The value of $\mathcal{F}_c$, to which $\langle N \rangle$ is proportional, determines at what SFR this happens.

Behavior as a result of changing the upper and lower limits on the ICMF is less trivial than the simple translation that results from modifying $\mathcal{F}_c$, and is illustrated in the bottom four panels of Figure 4.4. 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 star formation rate 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 (CFR) to make the probability of a cluster near the maximal mass 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 star formation rate 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 (4.25) to compute the expected mass of the most massive cluster and equation (4.9) to compute the expected number of clusters. Combining these two results gives the star formation rate at which we expect to produce a cluster with a mass of at least $CM_{\rm max}$. This is roughly

$$\begin{aligned} \dot{M}_{\star,\rm lim} &\approx \frac{\langle M \rangle}{\mathcal{F}_c \Delta t(\beta+1)} \left[\frac{(M_{\rm min}/M_{\rm max})^{\beta+1} - 1}{(M'_{\rm max}/M_{\rm max})^{\beta+1} - 1} \right] \\ &\approx \frac{\langle M \rangle}{\mathcal{F}_c \Delta t(\beta+1)} \left(\frac{M_{\rm min}}{M'_{\rm max}} \right)^{\beta+1}, \end{aligned} \tag{4.26}$$

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. 4.4, which differ only in their values of $M_{\rm max}$.

The effect of the lower mass cutoff $M_{\rm min}$ is twofold (see equation 4.25). 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% 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 of figure 4.5). 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.

4.4.2 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 Figure 4.5, higher mean dust extinctions lower the median luminosity at expected at fixed SFR. Equally importantly, dispersion in the dust optical depth distribution broadens the distribution. In the extreme case of a highly-extincted 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 star formation rate. 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 star formation rate 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 (2012b) for a recent review). Robust SFR measurements are best achieved by combining two star formation rate 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 Figure 4.5.

4.5 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 parameterized treatments of cluster aging, cluster disruption, dust extinction, and observational uncertainties in the determination of both cluster luminosities and galaxy star formation rates. We have implemented this analytic method in a new software package, the Cluster Luminosity Order-Statistic (`CLOC`) code, 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 star formation rate 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 Monte Carlo Markov Chain methods to explore the full, multi-dimensional 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 star formation rates. 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 star formation rates and brightest cluster luminosities, which is the subject of the companion paper (da Silva et al., in prep.).

Acknowledgements

R.L.dS. and M.R.K. 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 R.L.dS was supported by the National Science Foundation Graduate Research Fellowship. Support for M.F. 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. M.R.K. acknowledges support from an Alfred P. Sloan Fellowship, from NSF grant AST-0955300, and from NASA ATP grant NNX13AB84G. M.R.K. and S.M.F. 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.

4.6 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 $\Phi(L) = \int_0^L \phi(L')\,dL'$ 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

$$P_m(>L) = \begin{cases} \frac{N!}{(N-m)!m!}[1-\Phi(L)]^m \Phi(L)^{N-m}, & m \leq N \\ 0, & m > N \end{cases} . \tag{4.27}$$

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

$$\Phi_k(L \mid N) = \sum_{m=0}^{k-1} P_m(> L) \tag{4.28}$$

Note that for $N < k$, this evaluates to $\Phi_k(L \mid N) = 1$ 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

$$\begin{aligned} \phi_k(L \mid N) &= \frac{d}{dL}\Phi_k(L \mid N) & (4.29) \\ &= \frac{N!}{(N-k)!(k-1)!}\Phi(L)^{N-k}[1-\Phi(L)]^{k-1} \times \\ &\quad \phi(L) & (4.30) \end{aligned}$$

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 $\phi_k(L \mid N)$ 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 $L + dL$, 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 sub-galactic 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 4.32 and 4.33) 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

$$\phi_k(L \mid N) = \begin{cases} \frac{N!}{(N-k)!(k-1)!}\Phi(L)^{N-k}[1-\Phi(L)]^{k-1}\phi(L), & k \leq N \\ \delta(L), & k > N \end{cases} . \tag{4.31}$$

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 $(d/dL)\Phi_k(L \mid N) = 0$ for any L such that $\Phi(L) \neq 0$. However, for $\phi_k(L \mid N)$ to be properly normalized, it must have an integral of unity over all L. The definition given by equation (4.31) 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 $\phi_k(L \mid N)$ and $\Phi_k(L \mid N)$ weighted by the probability $P(N)$ that a given region contains exactly N clusters:

$$\Phi_k(L) = \sum_{N=0}^{\infty} P(N)\Phi_k(L \mid N) \tag{4.32}$$

$$\phi_k(L) = \sum_{N=0}^{\infty} P(N)\phi_k(L \mid N). \tag{4.33}$$

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, 2010b), many other choices are possible that enforce mass conservation to varying degrees. For example, the `SLUG` code to which we compare in Section 4.3.2 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 (4.32) and (4.33).[4]

Without a real physical theory of star cluster formation there is no obvious reason to favor 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

$$\Phi_k(L) = \sum_{N=0}^{\infty} \frac{\langle N\rangle^N e^{-\langle N\rangle}}{N!} \Phi_k(L \mid N) \tag{4.34}$$

$$\phi_k(L) = \sum_{N=0}^{\infty} \frac{\langle N\rangle^N e^{-\langle N\rangle}}{N!} \phi_k(L \mid N), \tag{4.35}$$

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

[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 small 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.

Evaluating equation (4.34), we have

$$\Phi_k(L) = \sum_{N=0}^{\infty} \frac{\langle N\rangle^N e^{-\langle N\rangle}}{N!} \sum_{m=0}^{k-1} P_m(>L) \tag{4.36}$$

$$= \sum_{m=0}^{k-1} \sum_{N=0}^{\infty} \frac{\langle N\rangle^N e^{-\langle N\rangle}}{N!} P_m(>L) \tag{4.37}$$

$$= \sum_{m=0}^{k-1} \frac{\langle N\rangle^m [1-\Phi(L)]^m e^{-\langle N\rangle}}{m!} \times \sum_{N=m}^{\infty} \frac{\langle N\rangle^{N-m}\Phi(L)^{N-m}}{(N-m)!} \tag{4.38}$$

$$= e^{-\langle N\rangle} e^{\langle N\rangle\Phi(L)} \sum_{m=0}^{k-1} \frac{\{\langle N\rangle[1-\Phi(L)]\}^m}{m!}. \tag{4.39}$$

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 (4.27), and the fourth line follows from the definition of the exponential function. The final line is equation (4.7) of the main text.

Similarly, evaluating equation (4.35) gives

$$\phi_k(L) = \sum_{N=0}^{k-1} \frac{\langle N\rangle^N e^{-\langle N\rangle}}{N!}\delta(L) + \sum_{N=k}^{\infty} \frac{\langle N\rangle^N e^{-\langle N\rangle}}{(N-k)!(k-1)!} \times \Phi(L)^{N-k}[1-\Phi(L)]^{k-1}\phi(L) \tag{4.40}$$

$$= \frac{\Gamma(k,\langle N\rangle)}{\Gamma(k)}\delta(L) + \frac{\langle N\rangle^k e^{-\langle N\rangle}}{(k-1)!}[1-\Phi(L)]^{k-1}\phi(L) \times \sum_{N=k}^{\infty} \frac{\langle N\rangle^{N-k}\Phi(L)^{N-k}}{(N-k)!} \tag{4.41}$$

$$= \frac{\Gamma(k,\langle N\rangle)}{\Gamma(k)}\delta(L) + \langle N\rangle \frac{\{\langle N\rangle[(1-\Phi(L)]\}^{k-1}}{(k-1)!} e^{-\langle N\rangle} e^{\langle N\rangle\Phi(L)}. \tag{4.42}$$

The first line follows from substituting equation (4.31) into equation (4.35), the second is just an algebraic re-arrangement, and the final line uses the definition of the exponential function. This is equation (4.5) of the main text.

4.7 Fit for Υ

In this Appendix we compute an approximate powerlaw 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-fit parameters

$$\zeta = 0.688 \tag{4.43}$$

$$\Upsilon_* = 8.3 \times 10^{-21}\, M_\odot \left(\mathrm{erg\ s^{-1}\ Hz^{-1}}\right)^{-1}. \tag{4.44}$$

Figure 4.6 shows both the STARBURST99 result and our best-fitting function. The maximum deviation between the fit and the numerical result is 0.07 dex.

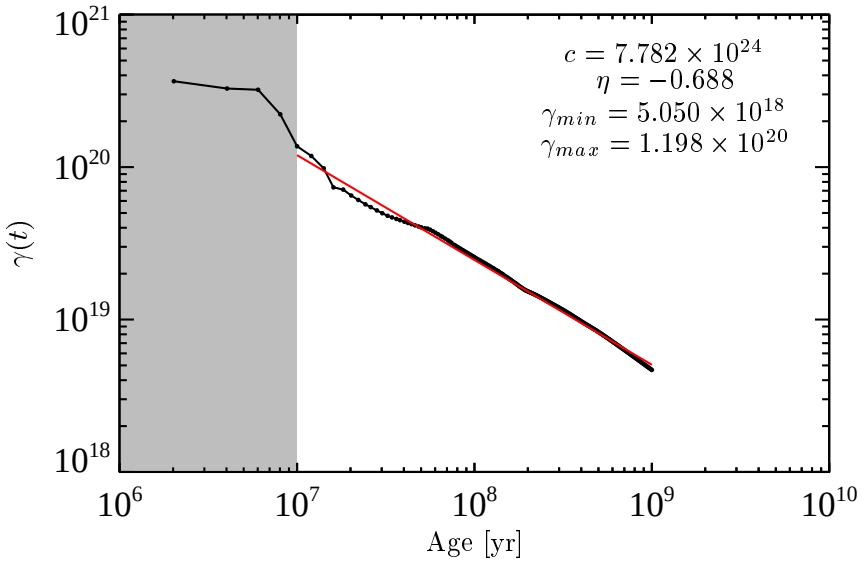

Figure 4.6 Fit of a simple power law approximation ($1/\Upsilon \propto t^{-\zeta}$, red) for the light to mass ratio $1/\Upsilon(t)$, compared to the results of a STARBURST99 calculation of Υ with a Kroupa IMF, Padova+AGB stellar tracks, Lej+SMI stellar atmospheres, and Solar metallicity (black). The grayed out region corresponds to populations younger than 10 Myr, for which the fit is poor.

4.8 The Luminosity Function for Variable Ages

Here we evaluate the convolution

$$\phi_{\rm in}(\log L_{\rm in}) \propto \psi(\log M) * \xi(-\log \Upsilon) \tag{4.45}$$

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 (4.18), which gives

$$\xi([1/\Upsilon]) \propto \begin{cases} \Upsilon^{(\gamma-1-\zeta)/\zeta}, & 1/\Upsilon_0 < 1/\Upsilon < 1/\Upsilon_1 \\ 0, & \text{otherwise} \end{cases}, \tag{4.46}$$

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

$$f_1(-\log\Upsilon) \propto \begin{cases} \Upsilon^{(\gamma-1)/\zeta}, & 1/\Upsilon_0 < 1/\Upsilon < 1/\Upsilon_1 \\ 0, & \text{otherwise} \end{cases}. \tag{4.47}$$

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

$$f_2(\log M) \propto \begin{cases} M^{\beta+1}, & M_{\rm min} < M < M_{\rm max} \\ 0 & \text{otherwise} \end{cases}. \tag{4.48}$$

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

$$\begin{aligned}
&\phi_{\rm in}(\log L_{\rm in}) \\
&\quad \propto \int_{-\infty}^{\infty} f_1(z) f_2(\log L_{\rm min} - z)\, dz && (4.49)\\
&\quad = \int_{-\infty}^{\infty} [\exp(z)]^{(\gamma-1)/\zeta} \left[\exp\left(\log L_{\rm in} - z\right)\right]^{\beta+1}\, dz && (4.50)\\
&\quad \propto \begin{cases} 0, & L_{\rm in} > \frac{M_{\rm max}}{\Upsilon_0} \\ \int_{z_0}^{z_1} G(z)\, dz, & \frac{M_{\rm min}}{\Upsilon_1} < L_{\rm in} < \frac{M_{\rm max}}{\Upsilon_0} \\ 0, & L_{\rm in} < \frac{M_{\rm min}}{\Upsilon_1} \end{cases} && (4.51)
\end{aligned}$$

where for convenience we have defined

$$\begin{aligned}
z_0 &= \log \max[1/\Upsilon_1, L_{\rm in}/M_{\rm max}] && (4.52)\\
z_1 &= \log \min[1/\Upsilon_0, L_{\rm in}/M_{\rm min}] && (4.53)\\
G(z) &= [\exp(z)]^{(\gamma-1)/\zeta} \left[\exp\left(\log L_{\rm in} - z\right)\right]^{\beta+1} && (4.54)\\
&= L_{\rm in}^{\beta+1} \exp(\omega z), && (4.55)
\end{aligned}$$

where

$$\omega = \frac{\gamma - 1}{\zeta} - \beta - 1. \qquad (4.56)$$

The integral $\int G(z)\, dz$ 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 the 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, $t_{\rm min} = 10$ Myr, and $t_{\rm max} = 1$ 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

$$\begin{aligned}
&\phi(\log L_{\rm in}) \\
&\quad\propto \begin{cases}
0, & L_{\rm in} > \frac{M_{\rm max}}{\Upsilon_0} \\
\int_{\log L_{\rm in}/M_{\rm max}}^{\log 1/\Upsilon_0} G(z)\,dz, & \frac{M_{\rm max}}{\Upsilon_1} < L_{\rm in} < \frac{M_{\rm max}}{\Upsilon_0} \\
\int_{\log 1/\Upsilon_1}^{\log 1/\Upsilon_0} G(z)\,dz, & \frac{M_{\rm min}}{\Upsilon_0} \le L_{\rm in} \le \frac{M_{\rm max}}{\Upsilon_1} \\
\int_{\log 1/\Upsilon_1}^{\log L_{\rm in}/M_{\rm min}} G(z)\,dz, & \frac{M_{\rm min}}{\Upsilon_1} < L_{\rm in} < \frac{M_{\rm min}}{\Upsilon_0} \\
0, & L_{\rm in} < \frac{M_{\rm min}}{\Upsilon_1}
\end{cases} \qquad (4.57) \\
&\quad\propto L_{\rm in}^{\beta+1} \begin{cases}
0, & L_{\rm in} > \frac{M_{\rm max}}{\Upsilon_0} \\
\Upsilon_0^{-\omega} - \frac{M_{\rm max}^{-\omega}}{L_{\rm in}^{-\omega}}, & \frac{M_{\rm max}}{\Upsilon_1} < L_{\rm in} < \frac{M_{\rm max}}{\Upsilon_0} \\
\Upsilon_0^{-\omega} - \Upsilon_1^{-\omega}, & \frac{M_{\rm min}}{\Upsilon_0} \le L_{\rm in} \le \frac{M_{\rm max}}{\Upsilon_1} \\
\frac{M_{\rm min}^{-\omega}}{L_{\rm in}^{-\omega}} - \Upsilon_1^{-\omega}, & \frac{M_{\rm min}}{\Upsilon_1} < L_{\rm in} < \frac{M_{\rm min}}{\Upsilon_0} \\
0, & L_{\rm in} < \frac{M_{\rm min}}{\Upsilon_1}
\end{cases} . \qquad (4.58)
\end{aligned}$$

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

$$\begin{aligned}
&\Phi_{\rm in}(\log L_{\rm in}) = \\
&\quad B \begin{cases}
1/B, & [\frac{M_{\rm max}}{\Upsilon_0}, \infty) \\
\frac{L_{\rm in}^{\beta+1} - M_{\rm max}^{\beta+1}/\Upsilon_1^{\beta+1}}{(\beta+1)\Upsilon_0^{\omega}} - & \\
\quad \frac{L_{\rm in}^{\nu} - M_{\rm max}^{\nu}/\Upsilon_1^{\nu}}{\nu M_{\rm max}^{\omega}} & \\
\quad + B_1 + B_2, & [\frac{M_{\rm max}}{\Upsilon_1}, \frac{M_{\rm max}}{\Upsilon_0}) \\
\left(\frac{\Upsilon_0^{-\omega} - \Upsilon_1^{-\omega}}{\beta+1}\right) \times & \\
\quad \left(L_{\rm in}^{\beta+1} - M_{\rm min}^{\beta+1}/\Upsilon_0^{\beta+1}\right) & \\
\quad + B_1, & [\frac{M_{\rm min}}{\Upsilon_0}, \frac{M_{\rm max}}{\Upsilon_1}) \\
\frac{L_{\rm in}^{\nu} - M_{\rm min}^{\nu}/\Upsilon_1^{\nu}}{\nu M_{\rm min}^{\omega}} - & \\
\quad \frac{L_{\rm in}^{\beta+1} - M_{\rm min}^{\beta+1}/\Upsilon_1^{\beta+1}}{(\beta+1)\Upsilon_1^{\omega}}, & [\frac{M_{\rm min}}{\Upsilon_1}, \frac{M_{\rm min}}{\Upsilon_0}) \\
0, & [0, \frac{M_{\rm min}}{\Upsilon_1})
\end{cases} . \qquad (4.59)
\end{aligned}$$

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

we have defined

$$
\begin{aligned}
B_1 &= \frac{M_{\min}^{\nu}/\Upsilon_0^{\nu} - M_{\min}^{\nu}/\Upsilon_1^{\nu}}{\nu M_{\min}^{\omega}} - \\
&\quad \frac{M_{\min}^{\beta+1}/\Upsilon_0^{\beta+1} - M_{\min}^{\beta+1}/\Upsilon_1^{\beta+1}}{(\beta+1)\Upsilon_1^{\omega}} \qquad (4.60)\\
B_2 &= \left(\frac{\Upsilon_0^{-\omega} - \Upsilon_1^{-\omega}}{\beta+1}\right) \times \\
&\quad \left(M_{\max}^{\beta+1}/\Upsilon_1^{\beta+1} - M_{\min}^{\beta+1}/\Upsilon_0^{\beta+1}\right) \qquad (4.61)\\
1/B &= B_1 + B_2 + \\
&\quad \frac{M_{\max}^{\beta+1}/\Upsilon_0^{\beta+1} - M_{\max}^{\beta+1}/\Upsilon_1^{\beta+1}}{(\beta+1)\Upsilon_0^{\omega}} - \\
&\quad \frac{M_{\max}^{\nu}/\Upsilon_0^{\nu} - M_{\max}^{\nu}/\Upsilon_1^{\nu}}{\nu M_{\max}^{\omega}} \qquad (4.62)\\
\nu &= \frac{\gamma-1}{\zeta}. \qquad (4.63)
\end{aligned}
$$

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

$$
\Phi_{\rm in}(\log L_{\rm in}) =
$$

$$
B' \begin{cases}
1/B', & [\frac{M_{\max}}{\Upsilon_0}, \infty) \\
\frac{L_{\rm in}^{\beta+1} - M_{\max}^{\beta+1}/\Upsilon_1^{\beta+1}}{(\beta+1)\Upsilon_0^{\omega}} & \\
\quad - \frac{\log(\Upsilon_1 L_{\rm in}/M_{\max})}{M_{\max}^{\omega}} & \\
\quad + B_1' + B_2', & [\frac{M_{\max}}{\Upsilon_1}, \frac{M_{\max}}{\Upsilon_0}) \\
\left(\frac{\Upsilon_0^{-\omega} - \Upsilon_1^{-\omega}}{\beta+1}\right) \times & \\
\quad \left(L_{\rm in}^{\beta+1} - M_{\min}^{\beta+1}/\Upsilon_0^{\beta+1}\right) & \\
\quad + B_1' & [\frac{M_{\min}}{\Upsilon_0}, \frac{M_{\max}}{\Upsilon_1}) \\
\frac{1}{M_{\min}^{\omega}} \log \frac{L_{\rm in}}{M_{\min}/\Upsilon_1} & \\
\quad - \frac{L_{\rm in}^{\beta+1} - M_{\min}^{\beta+1}/\Upsilon_1^{\beta+1}}{(\beta+1)\Upsilon_1^{\omega}}, & [\frac{M_{\min}}{\Upsilon_1}, \frac{M_{\min}}{\Upsilon_0}) \\
0, & [0, \frac{M_{\min}}{\Upsilon_1})
\end{cases} \qquad (4.64)
$$

where

$$
\begin{aligned}
B_1' &= \frac{1}{M_{\rm min}^{\omega}} \log \frac{\Upsilon_0}{\Upsilon_1} \\
&\quad - \left[\frac{M_{\rm min}^{\beta+1}}{(\beta+1)\Upsilon_1^{\omega}} \right] \left[\Upsilon_1^{-(\beta+1)} - \Upsilon_0^{-(\beta+1)} \right] \qquad (4.65) \\
B_2' &= \frac{\Upsilon_0^{-\omega} - \Upsilon_1^{-\omega}}{\beta+1} \times \\
&\quad \left(M_{\rm max}^{\beta+1}/\Upsilon_1^{\beta+1} - M_{\rm min}^{\beta+1}/\Upsilon_0^{\beta+1} \right) \qquad (4.66) \\
1/B' &= B_1' + B_2' + \\
&\quad \left[\frac{M_{\rm max}^{\beta+1}}{(\beta+1)\Upsilon_0^{\omega}} \right] \left[\Upsilon_0^{-(\beta+1)} - \Upsilon_1^{-(\beta+1)} \right] \\
&\quad - \frac{1}{M_{\rm max}^{\omega}} \log \frac{\Upsilon_0}{\Upsilon_1}. \qquad (4.67)
\end{aligned}
$$

4.9 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

$$\phi(\log L) \propto \phi_{\rm in}(\log L_{\rm in}) * \eta(-\tau), \qquad (4.68)$$

where $\phi_{\rm in}(L_{\rm in})$ is the distribution of intrinsic luminosities given by equation (4.58),

$$\eta(\tau) = \frac{\mathbf{1}_{(\tau_0,\tau_1)}(\tau)}{\tau_1 - \tau_0} = \begin{cases} \frac{1}{\tau_1 - \tau_0}, & \tau_0 < \tau < \tau_1 \\ 0 & \text{otherwise} \end{cases} \qquad (4.69)$$

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 4.8, we evaluate the PDF of the sum by transforming to logarithmic variables. To simplify the analysis, first note that equation (4.58) for $\phi_{\rm in}(\log L_{\rm in})$ can be rewritten as a sum of powerlaws multiplied by indicator functions:

$$\phi_{\rm in}(\log L_{\rm in}) \propto \sum_{i=1}^{5} C_i L_{\rm in}^{p_i} \mathbf{1}_{(L_{{\rm in},0,i}, L_{{\rm in},1,i})}(L_{\rm in}), \qquad (4.70)$$

with

$$\begin{aligned}
\mathbf{C} &= (\Upsilon_0^{-\omega}, -M_{\max}^{-\omega}, \Upsilon_0^{\omega} - \Upsilon_1^{-\omega}, M_{\min}^{-\omega}, -\Upsilon_1^{\omega}) && (4.71)\\
\mathbf{p} &= (\beta+1, \beta+1+\omega, \beta+1, \beta+1+\omega, \beta+1) && (4.72)\\
\mathbf{L}_{\text{in},0} &= \left(\frac{M_{\max}}{\Upsilon_1}, \frac{M_{\max}}{\Upsilon_1}, \frac{M_{\min}}{\Upsilon_0}, \frac{M_{\min}}{\Upsilon_1}, \frac{M_{\min}}{\Upsilon_1}\right) && (4.73)\\
\mathbf{L}_{\text{in},1} &= \left(\frac{M_{\max}}{\Upsilon_0}, \frac{M_{\max}}{\Upsilon_0}, \frac{M_{\max}}{\Upsilon_1}, \frac{M_{\min}}{\Upsilon_0}, \frac{M_{\min}}{\Upsilon_0}\right). && (4.74)
\end{aligned}$$

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

$$\begin{aligned}
&\phi(\log L)\\
&\quad\propto \int_{-\infty}^{\infty} \phi_{\text{in}}(z)\eta(z - \log L)\, dz && (4.75)\\
&\quad= \frac{1}{\tau_1 - \tau_0} \sum_{i=1}^{5} C_i \int_{-\infty}^{\infty} e^{p_i z} \mathbf{1}_{(L_{\text{in},0,i}, L_{\text{in},1,i})}(e^z) \cdot\\
&\qquad \mathbf{1}_{(\tau_0,\tau_1)}(z - \log L)\, dz && (4.76)\\
&\quad= \frac{1}{\tau_1 - \tau_0} \sum_{i=1}^{5} C_i \int_{-\infty}^{\infty} e^{p_i z} \mathbf{1}_{(\log L_{\text{in},0,i}, \log L_{\text{in},1,i})}(z) \cdot\\
&\qquad \mathbf{1}_{(\tau_0+\log L, \tau_1+\log L)}(z)\, dz && (4.77)\\
&\quad= \frac{1}{\tau_1 - \tau_0} \sum_{i=1}^{5} C_i \max\left(\frac{L_{i,1}^{p_i} - L_{i,2}^{p_i}}{p_i}, 0\right), && (4.78)
\end{aligned}$$

where

$$\begin{aligned}
L_{0,i} &= \max\left(\log L_{\text{in},0,i}, \tau_0 + \log L\right) && (4.79)\\
L_{1,i} &= \min\left(\log L_{\text{in},1,i}, \tau_1 + \log L\right). && (4.80)
\end{aligned}$$

The corresponding CDF is

$$\Phi(\log L) = \int_{-\infty}^{\log L} \phi(\log L')\, d\log L'. \tag{4.81}$$

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.

Chapter 5

Conclusions & Outlook

The main thrust of this work is to reevaluate the way in which astronomers do one of their most fundamental exercises: stellar population synthesis. While an ambitious and perhaps audacious goal, we have shown that standard approaches that make use of the "point-mass" assumption have serious problems in many astronomically interesting regimes. With `SLUG`, we have shown that this can lead to incorrect conclusions regrading the nature of the IMF and galaxy star formation rates. Future work will explore the effects on estimating population masses and ages as well, which even nature's simplest examples (i.e. single age star clusters) are affected by these same effects.

While other works had considered IMF sampling stochasticity, this work is the first to explore the effects of SFH sampling stochasticity. Remarkably, when working with composite stellar populations the properties of the stellar clustering appear to be at least as important as the properties of the IMF itself. Given its complete lack of literature attention thus far, this is an alarming call to reconsider how astronomers have interepreted their observations. For example, the artificial constructs of "star formation rates" need to be reexamined. There are no constant star formation rates in the universe. Instead, we have a collection of instantaneous bursts. Before considering stellar clustering, the large size of galaxies made it reasonable to neglect this time sensitive "flickering" as the large number of individual stars blurred together. However, the clumpy clustered nature of star formation introduces orders of magnitude more burstiness to the SFHs making such a treatment no longer viable.

With this newfound understanding of the importance of stellar clustering to interpreting even the most basic properties of the stellar populations we see, the need to understand the properties of that clustering are even more crucial. `CLOC` provides a framework to begin the next generation of studies in clustering properties that will hopefully allow greater constraints on these critical parameters.

Bibliography

Andrews, J. E., Calzetti, D., Chandar, R., et al. 2013, ApJ, 767, 51

Appenzeller, I., Fricke, K., Fürtig, W., et al. 1998, The Messenger, 94, 1

Bastian, N. 2008a, MNRAS, 390, 759

—. 2008b, MNRAS, 390, 759

Bastian, N., Konstantopoulos, I. S., Trancho, G., et al. 2012a, A&A, 541, A25

Bastian, N., Adamo, A., Gieles, M., et al. 2011, MNRAS, 417, L6

—. 2012b, MNRAS, 419, 2606

Bauer, A. E., Hopkins, A. M., Gunawardhana, M., et al. 2013, MNRAS, 434, 209

Bernasconi, P. A., & Maeder, A. 1996, A&A, 307, 829

Bigiel, F., Leroy, A., Walter, F., et al. 2010, AJ, 140, 1194

—. 2008, AJ, 136, 2846

Bik, A., Lamers, H. J. G. L. M., Bastian, N., Panagia, N., & Romaniello, M. 2003, A&A, 397, 473

Boissier, S., Gil de Paz, A., Boselli, A., et al. 2007a, ApJS, 173, 524

—. 2007b, ApJS, 173, 524

—. 2008, ApJ, 681, 244

Bolatto, A. D., Leroy, A. K., Jameson, K., et al. 2011, ApJ, 741, 12

Boselli, A., Boissier, S., Cortese, L., et al. 2009, ApJ, 706, 1527

Bothwell, M. S., Kenicutt, R. C., Johnson, B. D., et al. 2011, MNRAS, 415, 1815

Boutloukos, S. G., & Lamers, H. J. G. L. M. 2003, MNRAS, 338, 717

Bressert, E., Bastian, N., Gutermuth, R., et al. 2010, MNRAS, 409, L54

Brocato, E., Castellani, V., Raimondo, G., & Romaniello, M. 1999, A&AS, 136, 65

Buzzoni, A. 1989, ApJS, 71, 817

Calzetti, D., Liu, G., & Koda, J. 2012, ApJ, 752, 98

Cerviño, M. 2013, New A Rev., 57, 123

Cerviño, M., & Luridiana, V. 2004, A&A, 413, 145

—. 2006, A&A, 451, 475

Cerviño, M., Luridiana, V., Pérez, E., Vílchez, J. M., & Valls-Gabaud, D. 2003, A&A, 407, 177

Cerviño, M., Perez, E., Sanchez, N., Roman-Zuniga, C., & Valls-Gabaud, D. 2010, ArXiv e-prints, arXiv:1010.3022

Cerviño, M., & Valls-Gabaud, D. 2003, MNRAS, 338, 481

Chabrier, G. 2003, PASP, 115, 763

Chandar, R., Fall, S. M., & Whitmore, B. C. 2010a, ApJ, 711, 1263

—. 2010b, ApJ, 711, 1263

Charbonnel, C., Däppen, W., Schaerer, D., et al. 1999, A&AS, 135, 405

Charbonnel, C., Meynet, G., Maeder, A., & Schaerer, D. 1996, A&AS, 115, 339

Charlot, S., & Bruzual, A. G. 1991, ApJ, 367, 126

Colucci, J. E., Bernstein, R. A., Cameron, S. A., & McWilliam, A. 2011, ArXiv e-prints, arXiv:1104.3141

Conroy, C., & Gunn, J. E. 2010, ApJ, 712, 833

Conroy, C., Gunn, J. E., & White, M. 2009, ApJ, 699, 486

Conroy, C., White, M., & Gunn, J. E. 2010, ApJ, 708, 58

Cook, D. O., Seth, A. C., Dale, D. A., et al. 2012, ApJ, 751, 100

Corbelli, E., Verley, S., Elmegreen, B. G., & Giovanardi, C. 2009, A&A, 495, 479

Cordier, D., Pietrinferni, A., Cassisi, S., & Salaris, M. 2007, AJ, 133, 468

Crowther, P. A., Schnurr, O., Hirschi, R., et al. 2010, MNRAS, 408, 731

da Silva, R. L., Fumagalli, M., & Krumholz, M. 2012, ApJ, 745, 145

Dalcanton, J. J., Williams, B. F., Seth, A. C., et al. 2009, ApJS, 183, 67

de Jager, C., Nieuwenhuijzen, H., & van der Hucht, K. A. 1988, A&AS, 72, 259

de Wit, W. J., Testi, L., Palla, F., Vanzi, L., & Zinnecker, H. 2004, A&A, 425, 937

de Wit, W. J., Testi, L., Palla, F., & Zinnecker, H. 2005, A&A, 437, 247

Dionne, D., & Robert, C. 2006, ApJ, 641, 252

Eldridge, J. J. 2012, MNRAS, 422, 794

Eldridge, J. J., & Stanway, E. R. 2009, MNRAS, 400, 1019

Elmegreen, B. G. 2000, ApJ, 539, 342

Elson, R. A. W., Fall, S. M., & Freeman, K. C. 1989, ApJ, 336, 734

Fagotto, F., Bressan, A., Bertelli, G., & Chiosi, C. 1994a, A&AS, 104, 365

—. 1994b, A&AS, 105, 29

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

—. 2009a, ApJ, 704, 453

—. 2009b, ApJ, 704, 453

—. 2009c, ApJ, 704, 453

Fall, S. M., Krumholz, M. R., & Matzner, C. D. 2010, ApJ, 710, L142

Figer, D. F. 2005, Nature, 434, 192

Fioc, M., & Rocca-Volmerange, B. 1997, A&A, 326, 950

Fontanot, F., Cristiani, S., Santini, P., et al. 2012, MNRAS, 421, 241

Forero-Romero, J. E., & Dijkstra, M. 2013, MNRAS, 428, 2163

Fouesneau, M., & Lançon, A. 2010, A&A, 521, A22

Fouesneau, M., Lançon, A., Chandar, R., & Whitmore, B. C. 2012, ApJ, 750, 60

Fukugita, M., Ichikawa, T., Gunn, J. E., et al. 1996, AJ, 111, 1748

Fumagalli, M., da Silva, R., Krumholz, M., & Bigiel, F. 2010, ArXiv e-prints, arXiv:1009.0524

Fumagalli, M., da Silva, R., Krumholz, M., & Bigiel, F. 2011a, in Astronomical Society of the Pacific Conference Series, Vol. 440, UP2010: Have Observations Revealed a Variable Upper End of the Initial Mass Function?, ed. M. Treyer, T. Wyder, J. Neill, M. Seibert, & J. Lee, 155

Fumagalli, M., da Silva, R. L., & Krumholz, M. R. 2011b, ArXiv e-prints, arXiv:1105.6101

—. 2011c, ApJ, 741, L26

Fumagalli, M., & Gavazzi, G. 2008, A&A, 490, 571

Gieles, M., Lamers, H. J. G. L. M., & Portegies Zwart, S. F. 2007, ApJ, 668, 268

Gogarten, S. M., Dalcanton, J. J., Williams, B. F., et al. 2009, ApJ, 691, 115

Gutermuth, R. A., Pipher, J. L., Megeath, S. T., et al. 2011, ApJ, 739, 84

Haas, M. R., & Anders, P. 2010a, A&A, 512, A79+

—. 2010b, A&A, 512, A79

Hillier, D. J., & Miller, D. L. 1998, ApJ, 496, 407

Hirashita, H., Buat, V., & Inoue, A. K. 2003, A&A, 410, 83

Hopkins, A. M., & Beacom, J. F. 2006, ApJ, 651, 142

Hosokawa, T., Offner, S. S. R., & Krumholz, M. R. 2011, ArXiv e-prints, arXiv:1101.3599

Hoversten, E. A., & Glazebrook, K. 2008, ApJ, 675, 163

Jeffries, R. D. 2007, MNRAS, 381, 1169

Kennicutt, R. C., & Evans, N. J. 2012a, ARA&A, 50, 531

—. 2012b, ARA&A, 50, 531

Kennicutt, Jr., R. C. 1998, ARA&A, 36, 189

Kennicutt, Jr., R. C., Calzetti, D., Walter, F., et al. 2007, ApJ, 671, 333

Kotulla, R., Fritze, U., Weilbacher, P., & Anders, P. 2009, MNRAS, 396, 462

Kroupa, P. 2001, MNRAS, 322, 231

Kroupa, P., & Weidner, C. 2003, ApJ, 598, 1076

Kruijssen, J. M. D., Pelupessy, F. I., Lamers, H. J. G. L. M., et al. 2012, MNRAS, 421, 1927

Lada, C. J., & Lada, E. A. 2003, ARA&A, 41, 57

Lamb, J. B., Oey, M. S., Werk, J. K., & Ingleby, L. D. 2010, ApJ, 725, 1886

Lançon, A. 2011, in Astronomical Society of the Pacific Conference Series, Vol. 445, Why Galaxies Care about AGB Stars II: Shining Examples and Common Inhabitants, ed. F. Kerschbaum, T. Lebzelter, & R. F. Wing, 379

Lançon, A., & Mouhcine, M. 2000, in Astronomical Society of the Pacific Conference Series, Vol. 211, Massive Stellar Clusters, ed. A. Lançon & C. M. Boily, 34

Larsen, S. S. 1999, A&AS, 139, 393

—. 2002, AJ, 124, 1393

—. 2009, A&A, 494, 539

Larsen, S. S., & Richtler, T. 1999, A&A, 345, 59

Lee, J. C., Kennicutt, R. C., Funes, José G., S. J., Sakai, S., & Akiyama, S. 2007, ApJ, 671, L113

Lee, J. C., Gil de Paz, A., Tremonti, C., et al. 2009, ApJ, 706, 599

Leitherer, C., & Heckman, T. M. 1995, ApJS, 96, 9

Leitherer, C., Robert, C., & Drissen, L. 1992, ApJ, 401, 596

Leitherer, C., Schaerer, D., Goldader, J. D., et al. 1999, ApJS, 123, 3

Lejeune, T., Cuisinier, F., & Buser, R. 1997, A&AS, 125, 229

—. 1998, A&AS, 130, 65

Leroy, A. K., Walter, F., Sandstrom, K., et al. 2013, AJ, 146, 19

Ly, C., Malkan, M. A., Kashikawa, N., et al. 2012, ApJ, 747, L16

Maeder, A. 1990, A&AS, 84, 139

Maeder, A., & Meynet, G. 1987, A&A, 182, 243

McKee, C. F., & Williams, J. P. 1997, ApJ, 476, 144

Meurer, G. R., Wong, O. I., Kim, J. H., et al. 2009, ApJ, 695, 765

Meynet, G., Maeder, A., Schaller, G., Schaerer, D., & Charbonnel, C. 1994, A&AS, 103, 97

Mollá, M., García-Vargas, M. L., & Bressan, A. 2009, MNRAS, 398, 451

Momose, R., Koda, J., Kennicutt, Jr., R. C., et al. 2013, ApJ, 772, L13

Morrissey, P., Schiminovich, D., Barlow, T. A., et al. 2005, ApJ, 619, L7

Oey, M. S., King, N. L., & Parker, J. W. 2004, AJ, 127, 1632

Onodera, S., Kuno, N., Tosaki, T., et al. 2010, ApJ, 722, L127

Osterbrock, D. E., & Ferland, G. J. 2006,

Palla, F., & Stahler, S. W. 1999, ApJ, 525, 772

Pauldrach, A. W. A., Hoffmann, T. L., & Lennon, M. 2001, A&A, 375, 161

Pflamm-Altenburg, J., & Kroupa, P. 2008, Nature, 455, 641

Pietrinferni, A., Cassisi, S., Salaris, M., Cordier, D., & Castelli, F. 2007, in IAU Symposium, Vol. 241, IAU Symposium, ed. A. Vazdekis & R. F. Peletier, 39–40

Popescu, B., & Hanson, M. M. 2009a, AJ, 138, 1724

—. 2009b, AJ, 138, 1724

—. 2010a, ApJ, 724, 296

—. 2010b, ApJ, 713, L21

Portegies Zwart, S. F., McMillan, S. L. W., & Gieles, M. 2010, ARA&A, 48, 431

Raimondo, G., Brocato, E., Cantiello, M., & Capaccioli, M. 2005, AJ, 130, 2625

Reines, A. E., Nidever, D. L., Whelan, D. G., & Johnson, K. E. 2010, ApJ, 708, 26

Rose, C., & Smith, M. D. 2002, Mathematical Statistics with Mathematica

Salim, S., Rich, R. M., Charlot, S., et al. 2007, ApJS, 173, 267

Salpeter, E. E. 1955, ApJ, 121, 161

Scalo, J. M. 1986, Fund. Cosmic Phys., 11, 1

Schlegel, D. J., Finkbeiner, D. P., & Davis, M. 1998, ApJ, 500, 525

Schmidt, M. 1959, ApJ, 129, 243

Schmutz, W. 1998, in Astronomical Society of the Pacific Conference Series, Vol. 131, Properties of Hot Luminous Stars, ed. I. Howarth, 119–+

Schruba, A., Leroy, A. K., Walter, F., Sandstrom, K., & Rosolowsky, E. 2010, ApJ, 722, 1699

Siana, B., Teplitz, H. I., Ferguson, H. C., et al. 2010, ApJ, 723, 241

Silva-Villa, E., & Larsen, S. S. 2011, A&A, 529, A25

Skrutskie, M. F., Cutri, R. M., Stiening, R., et al. 2006, AJ, 131, 1163

Smit, R., Bouwens, R. J., Franx, M., et al. 2012, ApJ, 756, 14

Smith, L. J., Norris, R. P. F., & Crowther, P. A. 2002, MNRAS, 337, 1309

Thilker, D. A., Bianchi, L., Meurer, G., et al. 2007, ApJS, 173, 538

Tinsley, B. M. 1980, Fund. Cosmic Phys., 5, 287

van Bever, J., & Vanbeveren, D. 1998, A&A, 334, 21

Vázquez, G. A., & Leitherer, C. 2005, ApJ, 621, 695

Weidner, C., & Kroupa, P. 2004, MNRAS, 348, 187

—. 2005, ApJ, 625, 754

—. 2006, MNRAS, 365, 1333

Weidner, C., Kroupa, P., & Bonnell, I. A. D. 2010a, MNRAS, 401, 275

Weidner, C., Kroupa, P., & Larsen, S. S. 2004a, MNRAS, 350, 1503

—. 2004b, MNRAS, 350, 1503

Weidner, C., Pflamm-Altenburg, J., & Kroupa, P. 2010b, ArXiv e-prints, arXiv:1011.1905

Weisz, D. R., Skillman, E. D., Cannon, J. M., et al. 2008, ApJ, 689, 160

Weisz, D. R., Johnson, B. D., Johnson, L. C., et al. 2012a, ApJ, 744, 44

—. 2012b, ApJ, 744, 44

Werk, J. K., Putman, M. E., Meurer, G. R., et al. 2008, ApJ, 678, 888

Wong, T., & Blitz, L. 2002, ApJ, 569, 157

Zhang, Q., & Fall, S. M. 1999a, ApJ, 527, L81

—. 1999b, ApJ, 527, L81