How do bound star clusters form?

Mark R. Krumholz\textsuperscript{1,2}\textsuperscript{*} and Christopher F. McKee\textsuperscript{3}
\textsuperscript{1}Research School of Astronomy and Astrophysics, Australian National University, Canberra, ACT 2611 Australia
\textsuperscript{2}ARC Centre of Excellence for Astronomy in Three Dimensions (ASTRO-3D), Canberra, ACT 2611 Australia
\textsuperscript{3}Departments of Physics and Astronomy, University of California, Berkeley, CA 94720 USA

Accepted XXX. Received YYY; in original form ZZZ

ABSTRACT
Gravitationally-bound clusters that survive gas removal represent an unusual mode of star formation in the Milky Way and similar spiral galaxies. While forming, they can be distinguished observationally from unbound star formation by their high densities, virialised velocity structures, and star formation histories that accelerate toward the present, but extend multiple free-fall times into the past. In this paper we examine several proposed scenarios for how such structures might form and evolve, and carry out a Bayesian analysis to test these models against observed distributions of protostellar age, counts of young stellar objects relative to gas, and the overall star formation rate of the Milky Way. We show that models in which the acceleration of star formation is due either to a large-scale collapse or a time-dependent increase in star formation efficiency are unable to satisfy the combined set of observational constraints. In contrast, models in which clusters form in a “conveyor belt” mode where gas accretion and star formation occur simultaneously, but the star formation rate per free-fall time is low, can match the observations.

Key words: galaxies: star clusters: general – ISM: kinematics and dynamics – open clusters and associations: general – stars: formation

1 INTRODUCTION
The typical outcome of star formation in spiral galaxies is not a gravitationally-bound star cluster. In the Milky Way, Lada & Lada (2003) were among the first to point out that the number of observed star clusters at ages from $10 - 100$ Myr is a factor of $\sim 10$ smaller than one would expect if every observed gas-embedded star-forming clump were to go on to become a cluster of comparable mass. The natural explanation for this discrepancy is that most of the young stars that we observe in star-forming regions are in fact unbound, or will become so once the gas is removed, and that we count them as cluster members at young ages simply because they have not yet had time to drift apart. Extensive surveys of external galaxies echo this conclusion, with counts of star clusters as a function of age implying that no more than $5 - 10\%$ of stars that form will remain part of a gravitationally-bound structure several tens of Myr after formation (e.g., Adamo et al. 2015; Johnson et al. 2016; Chandar et al. 2017; Messa et al. 2018; see the recent review by Krumholz et al. 2019 for additional references).

Thus regions of star-formation that do go on to become gravitationally-bound clusters must be special in some way. Recent observational advances offer significant hints about how such regions might be special. Regions that go on to become bound clusters do not appear to represent a distinct class of gas cloud, such that most clouds unbind entirely and a small minority remain mostly bound. Instead, many star-forming regions appear to consist of a dense inner part that contains a minority of the mass, which is likely to go on to become bound, and an extended outer part whose stars will drift apart. The inner regions that go on to become bound are distinguishable in several ways.

First, they appear to feature extended star-formation histories. Low-density star-forming regions that are $\sim 10$ pc in size or larger tend to have stellar populations whose ages are comparable to their crossing times (Elmegreen 2000; Kruĳssen et al. 2019), suggesting a relatively rapid forma-
tion process. By contrast, the densest regions of star formation, with sizes $\sim 1$ pc, have star formation histories that are significantly more extended compared to their dynamical times. The best-studied example is the Orion Nebula Cluster (ONC), where the free-fall time in the central 1 pc is $\approx 0.6$ Myr (Da Rio et al. 2014), but there is extensive evidence that star formation has been ongoing for a significantly longer period (e.g., Reggiani et al. 2011; Jaehnig et al. 2015; Da Rio et al. 2016; Beccari et al. 2017). Star formation in this region appears to be accelerating (Palla & Stahler 2000; Huff & Stahler 2006; Caldwell & Chang 2018), but even accounting for this effect most stars are significantly older than a free-fall time — using the kinematically-selected sample and estimated ages of Kounkel et al. (2018), Krumholz et al. (2019) find that 50% of the stars in the ONC are older than 3 free-fall times, and 10% are older than 10 free-fall times. However, the ONC appears to be typical in this regard: similarly extended but accelerating star formation histories have been observed in NGC 6530 (Prisinzano et al. 2019), Perseus (Azimlu et al. 2015), Taurus, and $\rho$ Ophiuchus (Caldwell & Chang 2018).

Second, the regions with extended star formation histories are also distinct kinematically. While most young stars still embedded in their parent molecular clouds are characterised by unrelaxed density and velocity distributions (e.g., Fürész et al. 2008; Tobin et al. 2009), the density distribution in the central 1 pc of the ONC can be fit reasonably well by an isothermal, spherically-symmetric King (1962) model (Hillenbrand & Hartmann 1998), and the compact regions are virialised (Kim et al. 2019). This region is neither expanding nor contracting, and there is no evidence for a population of stars on primarily-radial orbits that are plausibly falling toward or escaping from it (Ward & Kruijssen 2018; Kuhn et al. 2019).

While regions like the ONC appear to be distinct in some respects, they also share one very significant commonality with the more extended envelopes around them. The density of young stellar objects (YSOs) increases smoothly with gas surface density, with no clear breaks at the densities or radii that correspond to the shift from unrelaxed, fractal stellar distributions to relaxed, virialised ones (Gutermuth et al. 2011). Once one normalises the gas surface density by the free-fall time, it correlates remarkably with the gas mass and free-fall time. If there are $N_{\mathrm{YSO}}$ YSOs asssociated with this gas that have a mean mass $M_{\mathrm{YSO}}$ and free-fall time $t_{\mathrm{ff}}$, then the sole modification to equation 1 is that $\rho_{\mathrm{ff}}$ is increased by a factor of $N_{\mathrm{YSO}}$.

We now sketch out some simple, general models for how star clusters might form. These are necessarily cartoons of much more complex processes, but they have the virtue of being analytically-computable, and capturing the essential features of a number of proposals that have appeared in the literature. All the software used to produce all the plots and analysis found in this paper are publicly available at https://bitbucket.org/krumholz/km19/.

### 2 FRAMEWORK FOR CLUSTER FORMATION

We begin by characterising a gas cloud that is in the process of forming a star cluster in terms of its instantaneous gas mass $M_\star$ and mean density $\rho_\star$; it is convenient to characterise the latter in terms of the free-fall time $t_{\mathrm{ff}} = \sqrt{3\pi/2G\rho_\star}$. Both $M_\star$ and $\rho_\star$ can in general be functions of time. At any instant, the cloud forms stars at a rate

$$M_\star = \epsilon_{\mathrm{ff}} M_{\mathrm{gas}} / t_{\mathrm{ff}}. \quad (1)$$

For simplicity we will generally only worry about mean quantities, but we note that, if instead of a uniform cloud one considers a cloud where the density profile is a power-law $\rho \propto r^{-k} \nu$, and one assumes that equation 1 holds locally (i.e., at every point the star formation density obeys $\dot{\rho}_\star \propto \rho_\star$), then the sole modification to equation 1 is that $\epsilon_{\mathrm{ff}}$ is increased by a factor of $[2/(2-k)]([3-k]/3)^{1/2}$, which is of order unity unless $k$ is very close to 2.

In addition to star formation, the cloud can gain mass by accretion and lose it by ejection of mass by stellar feedback. We take the mass removal rate by feedback to be proportional to the star formation rate $M_{\mathrm{ff}} = \eta M_\star$, while the accretion rate $M_{\mathrm{acc}}$ is an input parameter; here $\eta$ is the usual mass loading factor. The total mass of gas and stars

\[1\] Our choice to parameterise mass loss in terms of a mass-loading
therefore evolve following
\[ \dot{M}_g = M_{\text{acc}} - (1 + \eta) \frac{\dot{M}_g}{t_f}, \quad \dot{M}_* = \eta \frac{\dot{M}_g}{t_f}. \] (2)

In principle both \( M_{\text{acc}} \) and \( \eta \) can, like \( t_f \), be a function of time.

### 2.2 Scenarios of star formation

Having established this general framework, we now consider a range of scenarios for how a star cluster might be assembled. We plot example histories for each model in Figure 1, and summarise the models and their key free parameters in Table 1.

#### 2.2.1 Static cloud (ST)

Our first scenario is the simplest possible, a non-accreting cloud with constant \( t_f \) that begins with an initial gas mass \( M_g = M_{g,0} \) and starts forming stars at time \( t = 0 \). We refer to this as the static, or ST model, hereafter. Physically, this would correspond to a situation where cloud assembly is rapid compared to the process of star formation, or where a cloud is assembled in a state where it cannot form stars immediately. As first pointed out by Ginsburg et al. (2012) and Longmore et al. (2014, also see Walker et al. 2016 and Urquhart et al. 2018), such a scenario can be ruled out for at least the most massive young clusters observed in the Milky Way, on the grounds that there are no observed gas clouds dense and massive enough to be the progenitors of the most massive clusters. On the other hand, Krumholz et al. (2019) point out that there is no such difficulty for clusters near the Galactic Centre, and in this region there do indeed appear to be very massive and dense molecular clouds with little or no star formation activity such as “the Brick” (Longmore et al. 2013; Rathborne et al. 2014). These have been hypothesised to remain quiescent until star formation within them is triggered by a pericentre passage around Sgr A* (Kruĳssen et al. 2015), and thus they represent potential exemplars of the static cloud scenario.

Since there is no mass accretion in this model, \( M_{\text{acc}} = 0 \), and we will also assume \( \eta \) is constant, the solution to equation 2 is trivial:

\[ \dot{M}_* = M_{g,0} \frac{1 - e^{-\tau}}{1 + \eta}, \quad \dot{M}_g = M_{g,0} e^{-\tau}, \] (3)

factor \( \eta \), so that the mass removal rate is proportional to the star formation rate, differs from some other simple models (e.g., Lee et al. 2016) in which the mass removal rate is taken to be proportional to the total stellar mass. As discussed in Dekel & Krumholz (2013), which of these approximations is preferable depends on the feedback mechanism and the timescales under consideration. In the formation of star clusters, the dominant feedback mechanisms are likely to be protostellar outflows, photoionisation, and radiation pressure (Krumholz et al. 2019), which occur over timescales of \( \sim 0.1 \) to a few Myr after star formation. Since the star formation histories with which we are concerned are longer than this, \( \sim 10 \) Myr, we prefer to model the mass removal rate as proportional to the instantaneous star formation rate. However, arguments can be made for both choices.

### Figure 1

Example evolutionary histories of stellar mass (top), gas mass (middle), and star formation rate (bottom) for each of the models discussed in the paper (as indicated in the legend). For the purposes of this plot, we use \( \eta = 1 \) in all models. For CB and CBD we use \( \tau_{\text{acc}} = 1.5 \) and \( p = 3 \), for CBD we use \( \eta_d = 5 \), for GHC we use \( \xi = 1 \) and \( \tau_{\text{off}} = 0.75 \), and for IE we use \( \chi = 0.5 \) and \( \delta = 1 \). See main text for definitions of the various parameters.
the model result for general parameters. Note that not all of these parameters are independent, and in cases where parameters are related, we list the relationship in the table.

| Model name                  | Abbreviation | Parameter | Meaning                                      |
|-----------------------------|--------------|-----------|----------------------------------------------|
| Parameters common to all models |              | $\epsilon$| Star formation efficiency per free-fall time |
|                            |              | $\eta$    | Mass loading factor                          |
|                            |              | $t_g$     | Free-fall time                               |
|                            |              | $t_{ff}$  | Star formation timescale, $t_{ff} = t_g/[(1 + \eta)\epsilon]$ |

| Static cloud | ST | – | Accretion rate versus time $M_{acc} \propto t^p$ |
| Conveyor belt | CB | $p$ | Duration of accretion flow; dimensionless time $\tau_{acc} \equiv t_{acc}/t_{ff}$ |
| Conveyor belt + dispersal | CBD | $p$ | Duration of accretion flow; dimensionless time $\tau_{acc} \equiv t_{acc}/t_{ff}$ |
| Global hierarchical collapse | GHC | $t_{coll}$ | Collapse time; dimensionless time $\tau_{coll} \equiv t_{coll}/t_{ff}$ |
| Increasing efficiency | IE | $\delta$, $\epsilon$, $\chi$ | Efficiency per free-fall time varies as $\epsilon = \epsilon_{g,0}(t/t_{ff})^p$ Value of $\epsilon$ at $t = t_{ff}$; for this model, $t_{ff} \equiv t_g/[(1 + \eta)\epsilon_{g,0}]$ Ratio of star formation timescale to free-fall timescale, $\chi = t_{ff}/t_g$ |

Table 1. Summary of models and their parameters. Note that not all of these parameters are independent, and in cases where parameters are related, we list the relationship in the table.

mass of gas available for star formation, is

$$\epsilon_s \equiv \frac{M_s}{M_{g,0}} = \frac{1}{1 + \eta}.$$  \hspace{1cm} (5)

### 2.2.2 Conveyor belt (CB)

The absence of gas clouds as massive and dense as the densest star clusters, as noted in Section 2.2.1, led Longmore et al. (2014) to propose a “conveyor belt” model where gas accretion occurs simultaneously with cluster formation, so that the full mass of the gas cloud is never assembled at a single time; observations that regions such as the ONC frequently sit at the intersections of filaments supports this picture (Motte et al. 2018). In principle the region fed by a conveyor belt could be static, expanding, or contracting. Numerical simulations and analytic calculations by a number of authors (e.g., Klessen & Burkert 2000; Goldbaum et al. 2011; Matzner & Jumper 2015; Lee & Hennebelle 2016a,b) suggest that, as long as accretion is rapid, the input of energy supplied by the accretion flow is likely to yield a quasi-static structure, so here we will assume constant $t_g$ and $\epsilon_g$. We refer to this model as conveyor belt, or CB, hereafter.

We abstract this model as having an initial gas mass of zero, and an accretion rate that varies in time as a powerlaw $t^p$. We generically expect $p > 0$, since gravity-driven accretion rates generally rise with time until the reservoir of mass is exhausted; Goldbaum et al. (2011) show that pressureless collapse of a reservoir of constant surface density that becomes gravitationally unstable naturally produces $p = 3$; Lee & Hennebelle (2016b) find a similar value of $p$ while protoclusters are small compared to their parent reservoirs. Since that tapers to $p \approx 0$ once $\gtrsim 10\%$ of the parent reservoir has been accreted. We will adopt the Goldbaum et al. value of $p = 3$ as our fiducial choice, but for completeness we give the model result for general $p$, by taking the accretion rate to be

$$M_{acc} = H(t_{acc} - t)(p + 1)\frac{M_{g,0}}{t_{acc}} \left( \frac{t}{t_{acc}} \right)^p,$$  \hspace{1cm} (6)

where $M_{g,0}$ is the total mass that will eventually reach the protocluster, $t_{acc}$ is the time over which accretion happens, and $H(x)$ is the Heaviside step function. The initial conditions are $M_s = M_{g,0} = 0$. With this accretion rate, equation 2 has the following analytic solutions for any non-negative integer $p$:

$$M_s = \begin{cases} M_{g,0} & \text{if } \tau \leq \tau_{acc} \\
\frac{(1 + \eta)(p + 2)\rho^+_{acc}}{1 + \eta} g(\tau, p + 2), & \text{if } \tau > \tau_{acc} \end{cases}$$  \hspace{1cm} (7)

$$M_g = \begin{cases} \frac{M_{g,0}}{\tau_{acc}} + \frac{M_{g,0}}{\tau_{acc}} \delta_i \left( 1 - e^{-(p + 1)i} \right), & \text{if } \tau > \tau_{acc} \\
\frac{M_{g,0}}{\tau_{acc}^p} t \delta_i e^{-(p + 1)i} \delta_i, & \text{if } \tau > \tau_{acc} \end{cases}$$  \hspace{1cm} (8)

where for $p \geq 1$

$$g(\tau, p) = p_{\tau} e^{-\tau} \int_{\tau}^{\tau} \tau^{(p - 1)} e^{\tau^2} e^{-\tau^2} d\tau^2.$$  \hspace{1cm} (9)

$$= p! \left[ (-1)^p \frac{1}{p!} \sum_{i=1}^{p} (-1)^i \frac{1}{(p - i)!} \tau^{p-i} \right].$$  \hspace{1cm} (10)

Here $t = t_{ff}$ as in Section 2.2.1, $\tau_{acc} = t_{acc}/t_{ff}$, and we made use of the relations

$$\int_0^{\tau} g(\tau', p) d\tau^2 = \frac{g(\tau, p + 1)}{p + 1}.$$  \hspace{1cm} (11)

$$= \tau^p - g(\tau, p).$$  \hspace{1cm} (12)

To get a feeling for the magnitude of $g(\tau, p)$, we note that $g(\tau, 1) = 1 - e^{-\tau}$ and that $g(\tau, 2) = 2(\tau - 1 + e^{-\tau})$. For $\tau \leq 1$, $g(\tau, p) \approx \tau^p$ to within a factor 1.2 for $p \geq 4$. The accuracy improves as $\tau$ shrinks. Next, observe that equation 11 implies
\[
d\rho(t,p+1)/dt = (p+1)\rho(t,p) 
\]
In turn, this relation implies that \( g(t,p) \) is a monotonically increasing function of \( \tau \) since equation 9 implies that \( g(t,p) \) is positive. It follows from equation 8 that the gas mass increases monotonically until the accretion stops.

At times \( \tau \gg \tau_{\text{acc}} \), the star formation efficiency in the conveyor belt model approaches \( \epsilon_\tau = 1/(1+\eta) \), exactly as in the static cloud case, but the star formation history is different. This model satisfies the observational constraint that originally motivated it, in that the gas mass need never be large compared to the final stellar mass. Indeed, the final stellar mass (achieved in the limit \( \tau \to \infty \)) is \( M_{*,f} = M_{g,0}/(1+\eta) \) regardless of the accretion history, while the maximum gas mass (achieved when \( \tau = \tau_{\text{acc}} \)) is \( M_{g,\text{max}} \approx M_{g,0}/\tau_{\text{acc}} \), with the exact coefficient depending on the value of \( p \). Thus as long as \( \tau_{\text{acc}} \gtrsim \eta \), the maximum gas mass will be comparable to or smaller than the final stellar mass.

An important feature of this conveyor belt model is that star formation always accelerates while gas is accreting, provided \( p \geq 0 \). With the aid of equation 13, we find that the acceleration in the stellar mass is
\[
\dot{M}_* = \frac{p + 1}{(1+\eta)\tau_{\text{acc}}^{p+1}} \left( \frac{M_{g,0}}{\tau_{\text{ff}}^2} \right) g(t,p+2) \quad \tau \leq \tau_{\text{acc}},
\]
which is always positive, as noted above. Such acceleration appears to be demanded by the observations (Palla & Stahler 2000).

2.2.3 Conveyor belt model plus rapid dispersal (CBD)

A slight variation on the standard conveyor belt model is to note that, as pointed out by Goldbaum et al. (2011), mass loss rates are likely sensitive to the strength of the confining ram pressure from accretion. Consequently, it makes sense to adopt a mass loading factor that increases significantly once accretion ceases, leading to more rapid dispersal. We refer to conveyor belt models in which dispersal after the end of accretion is rapid as conveyor belt plus dispersal (CBD) models hereafter. From the standpoint of our simple analytic models, we can model this by setting \( \eta \) to one value during the accretion phase, \( t < t_{\text{acc}} \), and to another value \( \eta_d > \eta \) during the dispersal phase, \( t > t_{\text{acc}} \). In all other respects this model is identical to the sample conveyor belt model of Section 2.2.2. The solution to equation 2 in this case is modified only slightly from that given by equation 7 and equation 8:
\[
M_* = \begin{cases} 
\frac{M_{g,0}}{(1+\eta)(p+2)\tau_{\text{acc}}^{p+1}} g(t,p+2), & \tau \leq \tau_{\text{acc}} \\
M_{*,f} + \frac{M_{g,0}}{1+\eta_d} \left( 1 - e^{-\phi_d(t-\tau_{\text{acc}})} \right), & \tau > \tau_{\text{acc}} 
\end{cases}
\]
\[
M_g = \begin{cases} 
\frac{M_{g,0}}{\tau_{\text{acc}}^{p+1}} g(t,p+1), & \tau \leq \tau_{\text{acc}} \\
\frac{M_{g,0}}{\tau_{\text{acc}}^{p+1}} e^{-\phi_d(t-\tau_{\text{acc}})}, & \tau > \tau_{\text{acc}}
\end{cases}
\]
where \( \phi_d \equiv (1+\eta_d)/(1+\eta) \) can be thought of as representing the ratio of star formation efficiencies during and after the accretion phase. This model shares the key feature of the conveyor belt model: there is no need to assemble a cloud as massive as the final star cluster all at once, since the histories are identical up to the end of the accretion phase, but then a smaller fraction of the remaining gas mass is converted to stars than in the standard conveyor belt case. To be precise, the final star formation efficiency is
\[
\epsilon_\tau = \frac{1}{1+\eta} \left[ 1 - \frac{\phi_d - 1}{\phi_d} g(t_{\text{acc}},p+1) \right]
\]
Equations (11) and (12) imply that the ratio \( g(t,p+1)/\tau_{\text{acc}}^{p+1} \) is strictly smaller than unity for any \( \tau > 0 \) since \( g(t,p+2) > 0 \), so the final star formation efficiency is between \( 1/(1+\eta) \) and \( 1/(1+\eta_d) \).

2.2.4 Global hierarchical collapse (GHC)

The observation that star formation accelerates could be a reflection of gas accumulation, as in the CB or CBD models, but it could also be a result of the star formation process itself. An example of such a model is the global hierarchical collapse (GHC) scenario proposed by a number of authors (e.g., Zamora-Avilés & Vázquez-Semadeni 2014; Kuznetsova et al. 2015, 2018; Vázquez-Semadeni et al. 2017, 2019). The central idea of GHC models is that clouds are assembled in a low density state but then undergo a global collapse. Consequently, the mean free-fall time, rather than remaining constant, systematically decreases on a free-fall timescale as the mean density rises. The combination of an apparently-extended star formation history and an accelerating star formation rate is then taken to be due to the decreasing free-fall time: stars that form at early times may have ages comparable to the free-fall time of the system when they were formed, but this can be significantly longer than the free-fall time of the system at the time when it is observed. Moreover, as the system gets denser, the free-fall time decreases and thus star formation accelerates.

Mathematically we can represent this model by assuming that the mean density obeys
\[
\frac{dp}{dt} = \xi \frac{p}{t_{\text{ff}}(p)}
\]
where \( t_{\text{ff}}(p) = \sqrt{3\pi/2G\rho_0} \) is the free-fall time at the current density. The constant \( \xi \) specifies how fast the cloud contracts compared to the free-fall timescale, with higher \( \xi \) corresponding to more rapid contraction. The value of \( \xi \) will depend at least partly on geometry – \( \xi \approx 1 \) is expected for 3D structures, but values as small as \( \sim 0.1 \) are possible for highly-flattened geometries (Toalá et al. 2012). For a cloud that starts at density \( \rho_0 \) at time \( t = 0 \), the density and free-fall time evolve as
\[
\rho = \frac{\rho_0}{1 - \left( \frac{t}{t_{\text{ff},0}} \right)}^{2/\xi}, \quad t_{\text{ff}} = t_{\text{ff},0} \left[ 1 - \left( \frac{t}{t_{\text{ff},0}} \right)^{1/\xi} \right],
\]
where \( t_{\text{ff},0} = t_{\text{ff}}(\rho_0) \). The cloud reaches infinite density at time \( t = t_{\text{coll}} = 2t_{\text{ff},0}/\xi \).

Inserting this non-constant free-fall time into equation 2, holding \( \eta \) and \( \eta_d \) constant, and solving subject to the initial condition that \( M_g = M_{g,0} \) and \( M_* = 0 \) at \( t = 0 \), we
obtain
\[ M_{\ast} = \frac{M_{g,0}}{1 + \eta} \left[ 1 - (1 - \tau / \tau_{\text{coll}})^{\eta} \right], \quad \tau < \tau_{\text{coll}} \]
\[ M_{g} = \frac{M_{g,0}}{1 + \eta} \left( 1 - (1 - \tau / \tau_{\text{coll}})^{\eta} \right), \quad \tau \geq \tau_{\text{coll}} \]
(20)
(21)
where \( \tau = t/t_{\text{ff}} \) as before, but we now define \( t_{\text{ff}} = t_{\text{ff},0}/(1 + \eta) \) (c.f. equation 4), i.e., we define \( t_{\text{ff}} \) using the initial free-fall time since \( t_{\text{ff}} \) is non-constant. The quantity
\[ \tau_{\text{coll}} = \frac{2(1 + \eta) t_{\text{ff}}}{\xi} \]
is the dimensionless time at which the cloud collapses to infinite density and \( t_{\text{ff}} \to 0 \). Half the stars have formed and half the gas has been consumed at a time
\[ t_{1/2} = \left( 1 - \frac{1}{2} \right) t_{\text{coll}}. \]
(23)
and correspondingly the free-fall time then is
\[ t_{\text{ff},1/2} = \frac{t_{\text{ff},0}}{1 + \eta}. \]
(24)
For \( \tau_{\text{coll}} \geq 1 \), half the stars form at a rate not that different from the initial rate. More generally, the rate at which the star formation rate changes is
\[ \dot{M}_{\ast} = \frac{M_{g,0}}{(1 + \eta) t_{\text{ff},0}^2} \left[ 1 - \tau_{\text{coll}} \right] \left( 1 - \tau \right) \tau_{\text{coll}}^{-2}, \]
(25)
so star formation accelerates with time (\( \dot{M}_{\ast} > 0 \)) only if \( \tau_{\text{coll}} < 1 \). The final star formation efficiency is \( \epsilon_{f} = 1/(1 + \eta) \), exactly as in the ST or CB models.

2.2.5 Increasing star formation efficiency (IE)

A final potential mechanism to explain why star formation accelerates in protoclusters like the ONC is to posit that this is an intrinsic part of the star formation process itself. Lee et al. (2015) and Murray & Chang (2015) argue that, rather than being constant, \( \epsilon_{f} \) increases with time in star-forming regions as \( \epsilon_{f} \propto t^{\delta} \), with \( \delta \approx 1 \); we refer to this as the increasing efficiency (IE) model. Although somewhat similar to the GHC model, the two are conceptually distinct in that star formation accelerates in the GHC model because the mean density rises with time, while in the IE model it accelerates even though the mean density remains constant because the star formation process itself becomes more efficient. Mathematically, the two models differ in their predicted rate of acceleration. Caldwell & Chang (2018) argue that the IE model provides a good fit to observed star formation histories in resolved clusters, and Lee et al. (2016) and Ochsendorf et al. (2017) argue it provides a good fit to the observed ratio of ionising luminosity to CO luminosity, though, as we note above, the quality of the agreement is extremely sensitive to the choice of procedure for matching up non-co-spatial molecular gas and H II regions.

For the purposes of comparing this model to ours, we adopt the same parameterisation as Lee et al. (2016): \( \epsilon_{f} = \epsilon_{f,0}(t/t_{\text{ff}})^{\delta} \). Thus \( \epsilon_{f,0} \) represents the value of \( \epsilon_{f} \) one free-fall time after the onset of star formation. While the theoretical models of Lee et al. (2015) and Murray & Chang (2015) give \( \delta = 1 \), we will allow \( \delta \) to be a free parameter from 0 – 3 when we fit to observations below. The solution to equation 2 for arbitrary \( \delta \geq 0 \), holding \( \eta \) and \( t_{\text{ff}} \) constant, subject to the initial conditions \( M_{g} = M_{g,0} \) and \( M_{\ast} = 0 \) at \( t = 0 \), is
\[ M_{\ast} = \frac{M_{g,0}}{1 + \eta} \left[ 1 - \exp \left( -\frac{\chi^{\delta + 1}}{1 + \delta} \right) \right] \]
\[ M_{g} = \frac{M_{g,0}}{1 + \eta} \exp \left( -\frac{\chi^{\delta + 1}}{1 + \delta} \right) \]
(26)
(27)
where \( \tau = t/t_{\text{ff}} \), \( t_{\text{ff}} = t_{\text{ff},0}/(1 + \eta) \) (i.e., we define \( t_{\text{ff}} \) using the value of \( \epsilon_{f} \) at 1 free-fall time; c.f. equation 4), and \( \chi = t_{\text{ff}}/t_{g} = 1/(1 + \eta) \). The final star formation efficiency is \( \epsilon_{f} = 1/(1 + \eta) \), exactly as in the static model. The average efficiency with which stars form is
\[ \bar{\epsilon}_{f} = \frac{t_{\text{ff}}}{\tau_{\text{ff}}} = \frac{1 + \eta}{M_{g,0}} \int_{0}^{\infty} \epsilon_{f}(\tau) M_{\ast}(\tau) d\tau \]
\[ = \left[ \chi(1 + \delta) \right]^{\delta/(1 + \delta)} \left[ 1 + \frac{\delta}{1 + \delta} \right] \int_{0}^{\infty} \epsilon_{f,0}(\tau) d\tau. \]
(28)
For typical parameters in this model, \( \delta = 1 \) and \( \chi = 50 \), this gives \( \bar{\epsilon}_{f} \approx 8.9 \epsilon_{f,0} \), so most stars form at an efficiency substantially higher than that which prevails for the first free-fall time. Intuitively, this makes sense: in this model there are a relatively long period of near-quiescence when \( \epsilon_{f} \) is small and few stars form, but this is followed by a burst of activity after \( \epsilon_{f} \) becomes large; most stars form during this final burst. Quantitatively, the second derivative of the star formation rate is
\[ \ddot{M}_{\ast} = \frac{M_{g,0}}{1 + \eta} \left[ \chi^{\delta - 1} (\delta - \chi^{\delta} \tau^{\delta+1}) \exp \left( -\frac{\chi^{\delta + 1}}{1 + \delta} \right) \right]. \]
(29)
The sign of \( \ddot{M}_{\ast} \) therefore depends on \( \delta - \chi^{\delta} \tau^{\delta+1} \); for sufficiently small \( \tau \) this term is positive, and star formation accelerates. Later on, as gas is depleted, this term becomes negative and star formation decelerates.

3 CONFRONTATION WITH OBSERVATIONS

Having outlined the various models, we now compare them to observations.

3.1 Star formation histories

3.1.1 Data set

The first observation to which we are interested in comparing is the observed distribution of stellar ages in young clusters; as discussed in Section 1, working through the implications of the observed extended but accelerating star formation histories in such regions is one of our primary motivations in this work. For our observational data set, we select two young open clusters: the Orion Nebula Cluster and NGC 6530. We focus on these two because they both offer very clean, high-quality data: membership lists determined from Gaia 6D phase space data plus other ancillary indicators, and ages determined from spectroscopy, with star-by-star extinction corrections. The free-fall time in the ONC is \( t_{\text{ff},\text{ONC}} \approx 0.6 \) Myr as determined from dynamical modelling by Da Rio et al. (2014). For NGC 6530, Prisinzano et al. (2019) measure a stellar velocity dispersion of \( \sigma_{\text{NGC6530}} = 2.42 \) km s\(^{-1}\), and the effective radius of the cluster is 0.1°.
(Kharchenko et al. 2013), which translates to 2.3 pc for the best-fit distance of 1.32 kpc obtained by Prisinzano et al.. Thus the crossing time is $t_{\text{ff}} = r_{\text{NGC6530}} / v_{\text{NGC6530}} = 0.93$ Myr. For a virialised object, the free-fall time is approximately half the crossing time (Tan et al. 2006), so we adopt $t_{\text{ff,NGC6530}} = 0.5$ Myr.

For our stellar ages in NGC 6530, we use the fits provided by Prisinzano et al. (2019). For the ONC, we must select down from the full catalog of Kounkel et al. (2018), since their study covers the entire Orion star-forming region and includes multiple populations across a large volume. For this study, we select stars from their catalog that are within 1 pc in projection of $\theta^1$ C (the same radius within which we have estimated the free-fall time), and that are kinematically identified as part of the Orion A population. We take the ages of these stars from Kounkel et al., using only the ages based on spectroscopic determinations, since those based on colour are unreliable in the ONC due to high extinction. After applying these cuts, our sample consists of 185 stars in the ONC and 395 stars in NGC 6530.

In addition to the age estimates themselves, in order to carry out a meaningful statistical analysis we must have some understanding of the uncertainties in these measurements. Uncertainties in the ages of young stars has been a topic of considerable debate in the literature in recent years, and we refer to the readers to the reviews by Soderblom et al. (2014), Jeffries (2017), and Krumholz et al. (2019) for a detailed discussion. Young stellar ages are always subject to a systematic uncertainty of $\approx 0.1 - 0.3$ dex in the absolute age scale coming from the choice of pre-main sequence tracks. However, there is significantly less uncertainty in the relative ages of stars (e.g. Reggiani et al. 2011), which is the quantity of concern for us, since we are interested in the star formation history – a shift in absolute age just amounts to a rescaling of the timescales.

Relative age uncertainties come from a variety of factors, depending on the age-dating method. Uncertainties larger than $0.2 - 0.3$ dex can be ruled out by independent methods of constraining dispersions of stellar age (e.g., radii derived from rotation or gravity-sensitive spectral features – Jeffries 2007; Da Rio et al. 2016; Prisinzano et al. 2019), but a range of estimates below this limit have been published (e.g., Preibisch 2012; Da Rio et al. 2016; Prisinzano et al. 2019). For this work we adopt the results of Prisinzano et al. (2019): we take the error in log age to be a Gaussian with a width $\sigma = 0.13$ dex and a bias $b = -0.05$ dex (i.e., true stellar ages are on average 0.05 dex older than estimated ones). The systematic bias is due to unresolved binarity, which increases luminosity at fixed effective temperature, and thus tends to bias age estimates low. We have experimented with other choices of these parameters, subject to the overall constraint that the total error cannot exceed $\approx 0.2 - 0.3$ dex, and we find that the posterior PDFs for some parameters can be sensitive to the exact choice of $\sigma$ and $b$, as are quantitative measures of relative good-ness-of-fit such as the Akaike information criterion. Since we do not understand the true error distribution in detail, we will for this reason limit our analysis to general features that are robust against plausible changes in $\sigma$ or $b$.

### Table 2. Definitions of parameters used in computing the stellar age distribution likelihood function.

| Parameter | Meaning |
|-----------|---------|
| $t_{\text{clust}}$ | Age of cluster (time since onset of star formation) |
| $t_*$ | True age of a star |
| $t_{\text{obs}}$ | Observationally-estimated stellar age (including errors) |
| $\sigma$ | Dispersion of stellar age error distribution |
| $b$ | Bias in the stellar age error distribution |
| $t_{\text{ff,clust}}$ | Present-day free-fall time in cluster |
| $\epsilon_{\text{clust}}$ | Present-day star formation efficiency, $M_*(t_{\text{clust}}) / M_{L,0}$ |
| $g_{\text{clust}}$ | Present-day gas fraction, $M_g / (M_g + M_*)$ at $t = t_{\text{clust}}$ |

#### 3.1.2 Likelihood function

We wish to compare the observed age distribution to that predicted by our various candidate models. To this end, we now compute a likelihood function, which gives the probability density of the data given the model. For convenience we summarise the meanings of various parameters that we introduce in this calculation in Table 2.

For a cluster formation model with stellar mass as a dimensionless time, $M_*(\tau)$, the distribution of log stellar ages that will be seen a time when the cluster age is $t_{\text{clust}}$ (i.e., a time $t_{\text{clust}}$ after the onset of star formation) is

$$
\frac{dp}{d \log t_*} = \left( \ln 10 \right) \frac{t_* M_*(t_{\text{clust}} - t_*)}{t_{\text{clust}} M_*(t_{\text{clust}})}
$$

$$
\times \frac{\left( \frac{t_{\text{clust}} - t_*}{t_{\text{ff,clust}}} \right)}{M_*(t_{\text{clust}})}.
$$

(30)

where $t_*$ is the stellar age, $t_{\text{clust}} = t_{\text{clust}} / t_{\text{ff,clust}}$ and $t_* = t_{\text{ff,clust}} / t_{\text{ff,clust}}$ are the dimensionless cluster and stellar ages, respectively, and $M'_* = dM_* / dt$. The factor of $\ln 10$ is to ensure that the PDF is properly normalised to have unit integral over all log $t_*$. The stellar mass versus dimensionless time, $M_*(\tau)$, is given by equation 3, equation 7, equation 15, equation 20, and equation 26, for the ST, CB, CBD, GHC, and IE models, respectively; the corresponding gas masses, $M_g(\tau)$, are given by equation 3, equation 8, equation 16, equation 21, and equation 27.

Note that, in the GHC model, $t_{\text{ff}}$ is also a function of $\tau_{\text{clust}} - \tau$, (equation 19). The GHC model is also unique among the ones we consider in that it produces a double-peaked profile in the distribution $dp/d \log t_*$. Equation 30 shows that the age distribution is proportional to $M'_*(\tau_{\text{clust}} - \tau)$, or, in terms of the parameter $\tau$ in Figure 1, $(\tau_{\text{clust}} - \tau)M'_*(\tau)$. Reference to Figure 1 shows that this leads to a double peak in the GHC model, with one peak at $\tau \sim \tau_{\text{clust}}$ and a second at $\tau \sim \tau_{\text{clust}} - \tau_{\text{coll}}$.

To incorporate the effects of errors, we convolve the true age distribution with the error distribution. Following our discussion in Section 3.1.1, we parameterise the uncertainty in log age as a biased Gaussian, i.e., for a star whose true log age is $t_{\text{obs}}$, the distribution of measured log ages $t_{\text{obs}}$ is

$$
\frac{f(\log t_{\text{obs}} | \log t_*)}{\sqrt{2\pi}\sigma^2} \exp \left[ - \frac{(\log t_* - \log t_{\text{obs}} + b)^2}{2\sigma^2} \right],
$$

(31)

where $b$ is the bias and $\sigma$ is the dispersion, and both $b$ and
$\sigma$ are in units of dex. The full distribution of observed ages is therefore given by
\[
\frac{dp}{d \log t_{\text{obs}}} = \int_{-\infty}^{\infty} \left( \frac{dp}{d \log t} \right) f(\log t_{\text{obs}} | \log t_{\ast}) d \log t_{\ast}.
\]
We evaluate this integral numerically via Fourier transform, since it is equivalent to the convolution of the true stellar age distribution $dp/d\log t_{\ast}$ with a Gaussian. The log likelihood function $L$ is simply the probability density of the data given the model:
\[
\log L = \sum_{i=1}^{N} \log \left( \frac{dp_{i}}{d \log t_{\text{obs}}} \right)_{t_{\text{obs}}=t_{i}}
\]
where $t_{i}$ is the age estimated for the $i$th star in the observed sample.

Our stellar age distributions as written depend on two dimensional quantities: the cluster age $t_{\text{clust}}$ and the star formation timescale $t_{f}$ that scales between physical times $t$ and dimensionless times $\tau = t/t_{f}$. We treat these as free parameters to be fit. In addition, we fit free parameters for each of the models: $t_{\text{acc}}$ for model CB, $t_{\text{facc}}$ and $\phi_{g}$ for model CBD, $t_{\text{coll}}$ for model GHC, and $\delta$ for model IE. Note that we do not have to fit to $\eta$ or $\xi$ (for the GHC model), because $\eta$ is absorbed into the definition of $t_{f}$, and $\xi$ into the definition of $t_{\text{coll}}$. We adopt priors that are flat in the logarithm of all the positive-definite quantities (all timescales) or that are strictly greater than unity ($\phi_{g}$), and flat linear priors in all other parameters. We impose almost no constraint on the time of observation $t_{\text{obs}}$, allowing any value in the range 0.01 – 100 Myr, but we limit the allowed ranges of the remaining parameters based on physical considerations, which we now proceed to describe.

First, for all models we set the prior probability to zero for $t_{f}$ outside the range $10^{-4}$ to 1, on the grounds that $t_{f}$ values outside this range correspond to unphysically-inefficient or efficient star formation; to estimate $t_{f}$ from $t_{f}$, we use the observed free-fall time in NGC 6530 or the ONC, as appropriate, and $\eta = 1$. This serves to define the allowed range of $t_{f}$. Second, we apply priors based on the physical picture that motivates each model. For the CB and CBD models, the physical picture is that accretion is due to the collapse of a larger-scale, lower-density reservoir with a longer dynamical time than the cluster-forming region, a picture that requires $t_{\text{acc}} > t_{f}$; we also require $t_{\text{facc}} \leq t_{\text{clust}}$, not for any physical reason, but simply because all models with $t_{\text{facc}} > t_{\text{clust}}$ have identical age distributions for the stars that exist today, and thus cannot be distinguished in our analysis. For the GHC model, the central idea is that regions collapse on a free-fall timescale, forming stars while doing so. We therefore impose as a prior $0.1 < \xi < 10$; lower values of $\xi$ correspond to collapses so slow as to be nearly indistinguishable from the ST model, while higher values require regions to collapse much faster than a free-fall time, which is unphysical. This serves to limit the range of $t_{\text{coll}}$ (see footnote 2). Finally, for IE, theoretical models of how the density structure changes as star formation proceeds predict $\delta = 1$. We allow some range around this, by setting our prior to zero outside the range $\delta = 0 – 3$.

Our third and final prior is on the present-day star formation efficiency, $\epsilon_{\ast, \text{clust}} \equiv M_{\ast}(t_{\text{clust}})/M_{\ast,0}$, i.e., the fraction of all the gas available that has been converted to stars; note that $\epsilon_{\ast, \text{clust}}$ may be smaller than the final star formation efficiency $\epsilon_{\ast}$ that would be reached as $t_{\text{clust}} \to \infty$. For the ONC, Kim et al. (2019) find that the cluster is virialised and not expanding, which suggests that its star formation efficiency could not be too low. We have no direct dynamical evidence that the same is true for NGC 6530, but given its overall similarity with the ONC, this seems likely to be the case for it as well. The requirement that the star formation efficiency not be “too low” is somewhat difficult to quantify: when gas is removed from a protocluster rapidly compared to its dynamical time, loss of more than $\approx 70\%$ of the mass always leads to complete unbinding (Kroupa et al. 2001). However, the age distributions in the ONC and NGC 6530 imply that star formation, and presumably mass removal, have been ongoing for significantly longer than a free-fall time, and for sufficiently adiabatic gas removal, stars can remain bound down to arbitrarily small star formation efficiencies. Moreover, in order to match the observation that most stars do not form as part of bound clusters, we require that only a small fraction of the stars remain bound, and thus we do not want the efficiency to be too high. Given our uncertainties, we adopt a relatively mild prior, which disfavours efficiencies below 5%. Formally, we apply a prior $p_{\text{prior}}(\epsilon_{\ast, \text{clust}}) \propto \exp(-0.05/\min(\epsilon_{\ast, \text{clsut}}, 0.05)^2)$. For the purpose of calculating $\epsilon_{\ast}$, we adopt $\eta = 1$, corresponding to 50% instantaneous star formation efficiency, for all models, and a 50% final star formation efficiency for all but the CBD models. By allowing $\epsilon_{\ast, \text{clust}}$ to be small compared to $\epsilon_{\ast}$, we are allowing for the possibility that the clusters are observed early in the formation process, when $M_{\ast} \gg M_{\ast}$.

Finally, we note that the ONC is also observed to have a small gas fraction at the present day (Da Rio et al. 2014), $f_{g, \text{clust}} \equiv M_{g}(t_{\text{clust}})/([M_{g}(t_{\text{clust}}) + M_{\ast}(t_{\text{clust}})]) < 1$. While in principle this could serve as an additional prior, we lack quantitave constraints on the gas fraction in NGC 6530, and, with the exception of CB, none of our models contains an explicit treatment of gas clearing. For this reason, we will report $f_{g, \text{clust}}$ for our fits, but we will not impose any restrictions on it as a prior.

\[2\] Applying this prior to the GHC case requires some care, because a particular combination of $t_{\text{clust}}$, $t_{f}$, and $t_{\text{coll}}$, the parameters to which we are fitting, does not by itself determine a unique value of $\epsilon_{g}$; instead, one can change $\epsilon_{g}$ arbitrarily while leaving all these timescales unchanged by simultaneously changing $\xi$ and $t_{\text{coll}}$. To determine $\epsilon_{g}$, we must therefore choose a value of $\xi$. We can do so by considering two possible scenarios. One is that the cluster in question has not yet reached collapse ($t_{\text{clust}} < t_{\text{coll}}$), in which case we can fix $\xi$ by demanding that the free-fall time in the model match the observed present-day free-fall time $t_{\text{f, coll}}$ (0.6 Myr for the ONC, 0.5 Myr for NGC 6530, respectively). Rearranging equation 19, we find that the value of $\xi$ that satisfies this condition is $\xi = 2\pi t_{\text{clust}}/(t_{\text{coll}} - t_{\text{clust}})$. This in turn breaks the degeneracy and allows us to determine a unique value of $\epsilon_{g}$. The other possibility is that the cluster as we see it today is after the collapse to singularity ($t_{\text{clust}} > t_{\text{coll}}$), in which case the free-fall time we measure is a result of the stars rebounding to their current positions post-collapse, and has nothing to do with the free-fall time prior to collapse. In this case $\xi$ is unconstrained by the fit, and we must therefore adopt a value of $\xi$. For this case we choose a fiducial value $\xi = 1$. Our calculation of the best-fitting model is able to consider both scenarios, since we do not impose any prior on whether $t_{\text{clust}} < t_{\text{coll}}$ or $t_{\text{clust}} > t_{\text{coll}}$.\]
Table 3. Best fit parameters obtained by comparing each model to the observed distribution of stellar ages in the ONC and NGC 6530. The quantities listed as “Fit parameters” are those directly constrained in the fit, while “Derived parameters” are calculated from the fit parameters. Values are specified in the form $p(\theta|y) \propto p(y|\theta) p(\theta)$, where $p(q)$ is the qth percentile of the marginalised posterior PDF for that quantity. Times expressed as logarithms are in Myr.

| Model | log $t_{\text{obs}}$ [Myr] | log $t_{\text{fit}}$ [Myr] | Fit parameters | Derived parameters |
|-------|-----------------------------|-----------------------------|----------------|-------------------|
|       |                             |                             |                | log $\eta$ | $f_{\text{c,clust}}$ | log $\epsilon_{\text{c,clust}}$ |
| ONC   | 0.78$^{+0.02}_{-0.02}$     | 1.71$^{+0.15}_{-0.14}$     | $-2.23^{+0.14}_{-0.15}$ | 0.94$^{+0.02}_{-0.02}$ | 1.26$^{+0.13}_{-0.14}$ |
| CB, $p = 0$ | 0.92$^{+0.03}_{-0.03}$ | 1.45$^{+0.17}_{-0.16}$ | $-2.27^{+0.12}_{-0.10}$ | 0.95$^{+0.02}_{-0.02}$ | 1.29$^{+0.11}_{-0.10}$ |
| CB, $p = 3$ | 1.14$^{+0.04}_{-0.04}$ | 1.25$^{+0.32}_{-0.32}$ | $-2.07^{+0.03}_{-0.03}$ | 0.92$^{+0.03}_{-0.03}$ | 1.11$^{+0.03}_{-0.01}$ |
| CBD, $p = 3$ | 1.11$^{+0.04}_{-0.04}$ | 0.83$^{+0.23}_{-0.24}$ | $-2.15^{+0.04}_{-0.03}$ | 0.93$^{+0.04}_{-0.03}$ | 1.24$^{+0.04}_{-0.03}$ |
| GHC    | 0.94$^{+0.04}_{-0.04}$ | 1.39$^{+0.14}_{-0.14}$ | $-2.43^{+0.04}_{-0.04}$ | 0.94$^{+0.04}_{-0.04}$ | 1.32$^{+0.03}_{-0.03}$ |
| IE     | 1.08$^{+0.04}_{-0.04}$ | 4.70$^{+0.48}_{-0.47}$ | $-2.53^{+0.04}_{-0.04}$ | 0.92$^{+0.04}_{-0.04}$ | 1.39$^{+0.03}_{-0.03}$ |
| NGC6530| 0.52$^{+0.02}_{-0.02}$ | 1.65$^{+0.13}_{-0.13}$ | $-2.25^{+0.16}_{-0.16}$ | 0.96$^{+0.01}_{-0.01}$ | 1.45$^{+0.15}_{-0.15}$ |
| CB, $p = 0$ | 0.66$^{+0.02}_{-0.02}$ | 1.35$^{+0.17}_{-0.17}$ | $-2.17^{+0.12}_{-0.12}$ | 0.95$^{+0.02}_{-0.02}$ | 1.29$^{+0.11}_{-0.10}$ |
| CB, $p = 3$ | 0.92$^{+0.02}_{-0.02}$ | 1.04$^{+0.22}_{-0.22}$ | $-2.22^{+0.03}_{-0.03}$ | 0.92$^{+0.03}_{-0.03}$ | 1.11$^{+0.03}_{-0.02}$ |
| CBD, $p = 3$ | 0.89$^{+0.02}_{-0.02}$ | 0.51$^{+0.34}_{-0.34}$ | $-2.12^{+0.03}_{-0.03}$ | 0.92$^{+0.03}_{-0.03}$ | 1.24$^{+0.03}_{-0.03}$ |
| GHC    | 0.84$^{+0.04}_{-0.04}$ | 2.77$^{+0.18}_{-0.18}$ | $-2.43^{+0.04}_{-0.04}$ | 0.93$^{+0.04}_{-0.04}$ | 1.33$^{+0.03}_{-0.03}$ |
| IE     | 0.85$^{+0.02}_{-0.02}$ | 4.59$^{+0.20}_{-0.20}$ | $-2.53^{+0.04}_{-0.04}$ | 0.95$^{+0.02}_{-0.02}$ | 1.39$^{+0.03}_{-0.03}$ |

(a) We derive $\epsilon_{\eta}$ as follows: for models ST, CB, and CBD, we use equation 4 with $t_{\text{eff}}$ set equal to the observed value in NGC 6530 or the ONC. For model IE, we report the time-averaged value $f_{\text{c,clust}}$ given by equation 28. Finally, for model GHC we use the procedure described in footnote 2. In all cases our numerical value is for $\eta = 1$, and $\epsilon_{\eta}$ obeys the scaling $\epsilon_{\eta} \propto 1/(1 + \eta)$.

3.1.3 Results

Having defined the likelihood function and priors, we use the package EMCEE (Foreman-Mackey et al. 2013) to perform a Markov Chain Monte Carlo (MCMC) calculation to determine the posterior probability distribution for all the free parameters in each model as compared to the data; for the CB model we consider both a case with our fiducial value, $p = 3$, and one with $p = 0$, as predicted for late stages of collapse by Lee & Hennebelle (2016a,b). For this calculation we use 100 walkers and perform 500 MCMC steps; visual inspection of the chains indicates that this is more than adequate for convergence. We report the marginalised posterior PDFs, which we derive from the final 400 steps (i.e., we use 100 steps as a burn in period), in Table 3, and show the fits in Figure 2. We provide full posterior PDF distributions of all variables as Supplementary material (online). In Table 3 we also report three additional, derived quantities for each model, which are helpful in interpreting the results: the star formation rate per free-fall time $\epsilon_{\eta}$, the present-day gas fraction $f_{\text{c,clust}}$, and the present-day star formation efficiency $\epsilon_{\text{c,clust}}$.

Our analysis allows a few immediate conclusions. First, examining Figure 2, it is clear that the ST and CB ($p = 0$) models provide a poor description of the data in both NGC 6530 and the ONC. The underlying reason is that ST always produces a star formation rate that is highest at the start of star formation and then tapers; CB with $p = 0$ has a star formation rate that accelerates with time only weakly. Both models therefore predict a stellar age distribution that is peaked toward the oldest ages, contrary to what we observe. The MCMC attempts to compensate for this effect by favouring large star formation timescales $t_{\text{ff}}$, so that as little gas is converted to stars as possible and the star formation rate falls off due to gas depletion as little as possible; this is also why both models have very high present-day gas fraction $f_{\text{c,clust}}$ and very low present-day star formation efficiency $\epsilon_{\text{c,clust}}$.

The IE model provides a better fit to the data, but in order to do so the fit is driven to values of $\delta$, the acceleration parameter, far from the theoretically-preferred value $\delta = 1$. Indeed, the only reason $\delta$ does not go even higher is that our priors do not allow $\delta > 3$. Physically, this is because the model has difficulty producing a star formation history that extends for many free-fall times but also accelerates strongly at late times, unless $\delta$ is very large. The existence of a reasonably population of stars with ages approaching $\sim 10 \text{Myr}$ requires that the $t_{\text{obs}}$ not be too small, but then if $\delta$ is close to unity, too much gas is consumed at early times to allow the star formation rate to accelerate at later times. Thus in order to fit the data, the model requires a much larger value of $\delta$, which more strongly suppresses star formation at early times.

The most successful models are CB, CBD, and GHC. All models produce accelerating star formation that is in reasonable agreement with the observed age distribution, with an accretion time (for CB or CBD) or a collapse time (for GHC) that is nearly equal to the age of the oldest stars present, and to our best estimate for the age of the system as a whole. The posterior distribution of dimensionless star formation efficiency $\epsilon_{\eta}$ in these models is extremely broad, mainly because the star formation history is relative insensitive to gas consumption, and instead reflects the accumu-
Figure 2. Distribution of observed stellar ages $dp/d\log t_{\star,\text{obs}}$ in the ONC (top) and NGC 6530 (bottom). In all panels, the coloured lines represent 20 random samples from the final iteration of the MCMC for each of the models, as indicated in the legend. Grey histograms show the observed distribution, and are the same in every panel. The dashed vertical lines indicate 1, 3, 2, 10$\times$ the observed free-fall time, as indicated.

lation of additional mass (at a rate in good agreement with that predicted by Goldbaum et al. 2011) in CB or CBD, or to the overall increase in the density and thus decrease in the free-fall time in GHC. Interestingly, in the ONC all three models either admit or require that the present-day gas fraction $f_{g,\text{clus}}$ be small, consistent with the observations of Da Rio et al. (2014), though we did not explicitly impose this as a prior.

3.2 $\epsilon_{ff}$ from YSO counts

3.2.1 Data set

The next observational test to which we subject our models is the relationship between gas and YSOs in the gaseous objects that are the likely progenitors of star clusters. As discussed in the Section 1, estimates of $\epsilon_{ff}$ based on YSO counts cluster around $\approx 0.01$ in all observed star-forming regions, with small scatter. While this would seem to straightforwardly and directly constrain $\epsilon_{ff}$, a number of authors have suggested that this is not the case due to biases introduced by the methodologies of the measurement. For example, Lee et al. (2016) argue that some measurements preferentially select clouds early in their evolution, when, according to Lee et al.’s favoured IE model, $\epsilon_{ff}$ is smaller than its time-averaged value. Similarly, Vázquez-Semadeni et al. (2019) favour a GHC model and argue that estimates of $\epsilon_{ff}$ may be erroneous in clouds because a count of the number of YSOs present implicitly integrates the star formation rate over some period of time into the past, when the free-fall time was longer than the value we measure at the present day. We are in a position to test both these hypotheses, by directly modelling the observed distribution of $\epsilon_{ff}$ values produced by our cluster formation models.

We take our measured distribution of $\epsilon_{ff}$ values from Heyer et al. (2016), who identify class 0/1 YSOs within and measure $\epsilon_{ff}$ for gas clumps identified in the ATLASGAL survey (Schuller et al. 2009; Csengeri et al. 2014). This is the largest ($N = 517$)$^3$ and most complete sample of $\epsilon_{ff}$ measurements in the literature, and the ATLASGAL clumps that for some of this sample Heyer et al. (2016) do not detect any YSOs, and thus only obtain an upper limit on $\epsilon_{ff}$. For the purposes
it targets are very similar to the ONC and NGC 6530 in terms of mass, density, and free-fall time, making the data well-suited to the task of using both the cluster star formation history and the $\epsilon_f$ distribution together, as we do below. Specifically, the mean free-fall time of the ATLASGAL clumps is 0.3 Myr, very similar to the observed free-fall times of 0.5 Myr and 0.6 Myr in NGC 6530 and the ONC. Thus the ATLASGAL sample very likely represents a survey of YSOs in objects that will become clusters like NGC 6530 or the ONC, just at a slightly earlier evolutionary phase. However, we do note that the distribution of $\epsilon_f$ values obtained by Heyer et al. (2016) is qualitatively quite similar to those obtained from other samples that also use YSO counts for objects at a range of size and density scales (e.g., Evans et al. 2014; Ochsendorf et al. 2017).

### 3.2.2 Likelihood function

As in Section 3.1, to compare to the models to the observations we require a likelihood function that gives the probability density of the data given the model, which must properly account for averaging of the star formation rate over a finite time interval, potential biases in the sample, and observational errors. First consider the issue of averaging over a finite time. The Heyer et al. (2016) data set on which we focus estimates the star formation rate (SFR) based on number counts of class 0/I YSOs, a phase that lasts for a time $t_{YSO} \approx 0.5$ Myr (Evans et al. 2009; Gutermuth et al. 2009). We can therefore define an appropriately time-averaged $\epsilon_f$ for our models as

$$\epsilon_{f,\text{avg}}(t, \Delta t) = \frac{[M_f(t) - M_f(t - \Delta t)]}{\Delta t},$$

(34)

where $t$ is the time of observation, and $\Delta t = 0.5$ Myr is the window over which the SFR is averaged.

As with our treatment of stellar ages, we must consider not only biases (in this case introduced by averaging over a finite time), but observational errors. Errors in $\epsilon_f$ measurements are significantly more poorly modelled than errors in stellar age distributions, and involve subtleties such as making an IMF-based correction for the presence of protostars too dim to be detected. Given our ignorance, we will adopt a simple lognormal functional form, i.e., in a cloud with a true (time-averaged) logarithmic star formation efficiency $\log \epsilon_{f,\text{avg}}$, the distribution of observationally-inferred values $\log \epsilon_{f,\text{obs}}$ will be distributed as a Gaussian of width $\sigma_{\log \epsilon_f}$ centred on $\log \epsilon_{f,\text{avg}}$. That is, given a true (time-averaged) efficiency per free-fall time $\epsilon_{f,\text{avg}}$, the distribution of observationally-estimated star formation efficiency per free-fall time is

$$f(\log \epsilon_{f,\text{obs}} | \log \epsilon_{f,\text{avg}}) = \frac{1}{\sqrt{2\pi}\sigma_{\log \epsilon_f}} \exp \left[ -\frac{(\log \epsilon_{f,\text{obs}} - \log \epsilon_{f,\text{obs}})^2}{2\sigma_{\log \epsilon_f}^2} \right].$$

(35)

The value of the dispersion $\sigma_{\log \epsilon_f}$ is not well known, but we will see below that it is not necessary to adopt a model for of our analysis we take the value of $\epsilon_f$ in these clumps to be equal to the stated 2$\sigma$ upper limit.

Bound star cluster formation

Now consider a cloud observed at some time $t$ during its evolution, with an instantaneous time-averaged star formation efficiency $\epsilon_{f,\text{avg}}(t, \Delta t)$. The distribution of observed efficiencies for this cloud is $f(\log \epsilon_{f,\text{obs}} | \log \epsilon_{f,\text{avg}}(t, \Delta t))$. If we have a population of such clouds, each observed at random times $t$ between the onset of star formation at $t = 0$ and some maximum time $t_{\text{max}}$, then the distribution of observed $\epsilon_{f,\text{obs}}$ values for the population is simply the average of $f(\log \epsilon_{f,\text{obs}} | \log \epsilon_{f,\text{avg}}(t, \Delta t))$ over all possible times $t$ at which the clouds could be observed, i.e.,

$$\frac{dp}{d \log \epsilon_{f,\text{obs}}} = \frac{1}{t_{\text{max}}} \int_0^{t_{\text{max}}} f(\log \epsilon_{f,\text{obs}} | \log \epsilon_{f,\text{avg}}(t, \Delta t)) \, dt.$$  

(36)

The choice of maximum time $t_{\text{max}}$ is somewhat subtle. In simple models where $M_f$ reaches 0 in finite time, such as the GHC model, one can simply take $t_{\text{max}}$ to be the time for which $M_f(t_{\text{max}}) = 0$. However, we are interested in comparing to a more general class of models where $M_f$ may not go to exactly 0 at finite time. To choose a reasonable $t_{\text{max}}$, we note that studies of $\epsilon_f$ based on YSO counts always select YSOs and gas clouds within the same area on the sky, which limits the phase of evolution to which they are sensitive: as clusters evolve and begin to clear their gas, stars inevitably cease to be surrounded by molecular gas, so clouds that have cleared most of their gas are not included in YSO counting surveys. Our simple zero-dimensional models cannot capture this effect directly, but we crudely mimic it by choosing our time interval to correspond to that over which $M_f/M_\star > 1$, i.e., when the stellar mass has not yet exceeded the gas mass. We therefore take $t_{\text{max}}$ to be defined implicitly by the condition $M_f(t_{\text{max}})/M_\star(t_{\text{max}}) = 1$. We have verified that varying the value of $M_f/M_\star$ we use to define our time interval by a factor of ten in either direction not change the results substantially.

Given the preceding discussion, we have now write down the log likelihood function for a set of observed $\epsilon_f$ values is

$$\log L = \sum_{i=1}^{N} \log \left( \frac{dp}{d \log \epsilon_{f,\text{obs}}} \right)_{\epsilon_{f,\text{obs}}=\epsilon_{f,i}},$$

(37)

where $\epsilon_{f,i}$ is the $i$th observed value of $\epsilon_f$, and there are $N$ measurements in total. We use this likelihood function with emcee to obtain posterior PDFs for the parameters for the same models as in Section 3.1. As in our analysis of the stellar age distribution, we use priors that are flat in the log of positive-definite quantities, and flat in value for other quantities; the allowed parameter range is identical to that used in Section 3.1. In addition to the parameters included there, we must also fit for $\eta$, $\sigma_{\log \epsilon_f}$ and $\xi$ (for model GHC), since, while these do not affect the distribution of stellar ages, they do affect the distribution of observed $\epsilon_f$ values. For $\eta$ our prior is flat in log from 0.01 – 10, and for $\sigma_{\log \epsilon_f}$ it is flat in log from 0.01 – 10. We must also choose a value for the free-fall time, since this sets the ratio $\Delta t/t_f$, which determines how much the observed $\epsilon_f$ distribution is biased by averaging the star formation rate over a finite time. As noted above, the mean value of $t_f$ in the ATLASGAL sample is 0.3 Myr, and the dispersion around this is small (0.26 dex), so we use $t_f = 0.3$ Myr for our analysis of all models except GHC; GHC models sweep through all values of $t_f$ from $t_{f,0}$ to 0, so for this case we impose as a prior the...
requirement that $\delta t_{\text{ff},0} > 0.3$ Myr, i.e., the collapse must start from a state that is no denser than the observed ATLASGAL clumps.

### 3.2.3 Results

We show models evaluated using samples drawn from the MCMC chains in Figure 3, and report the posterior PDFs of all parameters in Table 4. The results show that all the models we consider can fit the observed $\epsilon_{\text{ff}}$ distribution quite well, but that both $\epsilon_{\text{ff}}$ and the level of observational error are very tightly constrained by the observations; $\epsilon_{\text{ff}}$ is required to be of order a few percent, and $\sigma_{\log \epsilon_{\text{ff}}}$ to be approximately 0.15 dex. Indeed, the models even constrain $\eta$ not to be too large, since otherwise rapid mass removal means that the gas mass is able to change significantly over the time-averaging interval $\Delta \tau$, which in turn would broaden the observed $\epsilon_{\text{ff}}$ distribution more than the data allow. Thus, despite the hypothesis in the literature that measured $\epsilon_{\text{ff}}$ distributions are biased because they average over a finite time interval and thus miss changes in the free-fall time (e.g., Vázquez-Semadeni et al. 2019), or that they miss periods of efficient star formation (e.g., Lee et al. 2016), we do not obtain significantly looser constraints on the value of $\epsilon_{\text{ff}}$ when we explicitly put those possibilities into our model.

### 3.3 Combined constraints

Having examined the constraints we can deduce from the distribution of stellar ages and the YSO-gas correlation individually, we now ask whether these constraints are compatible. That is, do there exist a set of parameters for a given model such it can simultaneously reproduce the observed stellar age distribution in young clusters and the YSO count in protoclusters? To answer this question, we use our MCMC samples to compute the dimensionless parameters $\epsilon_{\text{ff}}, \eta$, etc. that characterise each proposed model, using the constraints from both the stellar age distribution and YSO counts. We plot the posterior PDFs of these dimensionless parameters for models CBD, GHC, and IE in Figure 4, Figure 5, and Figure 6, respectively.\(^4\) These plots use the posterior PDFs derived from the stellar age distribution is NGC 6530, since it is a somewhat larger data set, but the results for the ONC are qualitatively similar. We omit ST and CB ($p = 0$) from this comparison because we have already determined that these models provide poor fits to the stellar age distribution alone, and we omit CB ($p = 3$) because it is qualitatively similar to CBD on the parameters they share. However, the corresponding plots for these clusters are provided in the Supplementary material (online).

Turning first to Figure 6, we immediately see that the IE model has a major difficulty: as discussed in Section 3.1 and shown in Figure 6, the stellar ages distributions in NGC 6530 and the ONC are best fit in the context of this model by a star formation efficiency that increases as roughly $\epsilon_{\text{ff}} \propto t^3$ (or faster, since $\delta = 3$ is the largest allowed by our priors). This is completely at odds with the constraint provided by the ATLASGAL clumps, whose tight relationship between YSOs and gas properties requires that $\epsilon_{\text{ff}}$ be nearly constant, and thus that $\delta \approx 0$. The physical explanation for this tension is simple: star formation is observed to accelerate based on stellar age distributions, and the IE model interprets this acceleration as a systematic increase in star formation efficiency with time. However, when one observes the gas clumps that are in the process of forming clusters, one finds that the number of YSOs per unit gas mass, normalised by the free-fall time, is nearly constant, completely inconsistent with large variations in star formation efficiency. There is no way to reconcile these two constraints in the context of the GHC model, or indeed in any model that assumes the acceleration of star formation is due to an increase in star formation efficiency with time. Instead, the acceleration of

\(^4\) For parameters that cannot be constrained by the stellar age distribution, we take the posterior PDF derived from stellar ages to be equal to the flat prior we use for these variables when analysing the ATLASGAL data.

Figure 3. Distribution of observed star formation efficiencies $\log \epsilon_{\text{ff}, \text{obs}}$. Grey histograms show the distribution observed by Heyer et al. (2016) for the ATLASGAL sample, and are the same in every panel. Coloured lines represent 20 random samples from the final iteration of the MCMC fit for each model, as indicated in the legend.
Table 4. Best fit parameters obtained by comparing each model to the observed distribution of measured \( \epsilon_f \) values in ATLASGAL clumps (Heyer et al. 2016).

| Model | \( \log \sigma_{\log \epsilon_f} \) [dex] | \( \log \eta^{(a)} \) | \( \log \tau_{\text{ef}} \) [Myr] | Other | Derived parameters \( \log \epsilon_f^{(b)} \) |
|-------|----------------|----------------|----------------|-------|----------------|
| ST | \(-0.79^{+0.01}_{-0.01}\) | \(-0.97^{+0.02}_{-0.01}\) | \(1.21^{+0.19}_{-0.04}\) | — | \(-1.78^{+0.02}_{-0.02}\) |
| CB, \( p = 0 \) | \(-0.79^{+0.01}_{-0.02}\) | \(-0.94^{+0.06}_{-0.01}\) | \(1.19^{+0.05}_{-0.28}\) | \( \log \tau_{\text{acc}} = 1.35^{+0.40}_{-0.00} \) | \(-1.76^{+0.02}_{-0.02}\) |
| CB, \( p = 3 \) | \(-0.79^{+0.01}_{-0.02}\) | \(-1.07^{+0.07}_{-0.61}\) | \(1.18^{+0.04}_{-0.11}\) | \( \log \tau_{\text{acc}} = 1.75^{+0.18}_{-0.30} \) | \(-1.74^{+0.02}_{-0.02}\) |
| CBD, \( p = 3 \) | \(-0.79^{+0.01}_{-0.02}\) | \(-0.70^{+0.36}_{-0.80}\) | \(1.13^{+0.31}_{-0.31}\) | \( \log \tau_{\text{acc}} = 1.69^{+0.21}_{-0.31} \) | \(-1.74^{+0.02}_{-0.02}\) |
| GHC | \(-0.79^{+0.02}_{-0.02}\) | \(-1.30^{+0.62}_{-0.48}\) | \(2.82^{+0.40}_{-0.62}\) | \( \log \tau_{\text{coll}} = 1.54^{+0.32}_{-0.31} \) | \(-1.76^{+0.03}_{-0.02}\) |
| IE | \(-0.79^{+0.02}_{-0.02}\) | \(-1.04^{+0.70}_{-0.64}\) | \(1.29^{+0.12}_{-0.12}\) | \( \delta = 0.07^{+0.10}_{-0.05} \) | \(-1.76^{+0.04}_{-0.03}\) |

Formatting is identical to that used in Table 3.

(a) The median and percentile values we report for \( \log \eta \) are strongly affected by our prior \( \log \eta > -2 \). All models with \( \eta \ll 1 \) are essentially identical, so our analysis cannot distinguish them; thus the values we report should be read as providing an upper limit at the reported 84th percentile, rather than a meaningful central estimate.

(b) The value of \( \epsilon_f \) we report here is the true value defined by the instantaneous star formation rate, not the time-averaged value \( \epsilon_{f, \text{av}} \) defined by equation 34. For model IE, we report the time-averaged value \( \tau_{\text{eff}} \) given by equation 28. We compute \( \epsilon_f \) as described in the notes to Table 3.

Figure 4. Corner plot showing the posterior PDF for the dimensionless parameters of the CBD model (\( \epsilon_f, \eta, \tau_{\text{acc}}, \) and \( \phi_d \)), derived using the distribution of stellar ages in NGC 6530 (red colours) and the counts of YSOs in ATLASGAL clumps (blue colours). In the panels on the bottom left corner, contours show 2D marginal posterior PDFs for each combination of variables, as indicated on the axes. Histograms along the central diagonal show 1D marginal posterior PDFs for each variable. PDFs in all panels are scaled so that the maximum is unity.

Figure 5. Same as Figure 4, but showing the GHC model and its dimensionless parameters.

star formation must be due either to an increase in the star-forming mass with time (as in CB or CBD) or a decrease in the free-fall time (as in GHC). We may therefore rule out the IE model.

The CBD and GHC models illustrated respectively in Figure 4 and Figure 5, on the other hand, show no contradiction between the parameter values demanded by the stellar age distributions and the ATLASGAL clumps. In both sets of models the ATLASGAL data very tightly constrain \( \epsilon_f \), while setting little constraint on \( \eta, \phi_d, \) or \( \chi \). Conversely, the stellar age distribution tightly constrains \( \tau_{\text{acc}}, \tau_{\text{coll}}, \) and \( \xi \), but provides little restriction on \( \epsilon_f \). As a result, there is a reasonable parameter space of overlap.

Thus we find that the joint set of data favours one of two scenarios. In the first, gas accretes as roughly \( M \propto t^3 \) (consistent with the theoretical models of Goldbaum et al. 2011) and forms stars inefficiently (\( \epsilon_f = 0.01 \)). Accretion continues for \( \sim 1 \) (10 star formation timescales (\( \tau_{\text{acc}} \sim 1 \) 10), and once it ends, mass is rapidly dispersed by feedback (\( \phi_d > 1 \)). In the second scenario, an initially low-density cloud undergoes
a global collapse that is fairly rapidly compared to the instantaneous free-fall time ($\xi \sim 3$), as might be expected for example in a colliding flow where the collapse is due to external pressure plus gravity rather than gravity alone, but during this collapse it forms stars quite inefficiently ($\epsilon_{ff} \approx 0.01$). As a result, the total collapse time is quite small compared to the star formation timescale ($\tau_{coll} \sim 0.03$), so that most stars form only during the final plunge when the density is running way to infinity – a value $\tau_{coll} < 1$ is required to yield an accelerating star formation history.

### 3.4 Global SFR

We now add an additional constraint to our modelling: the star formation rate of the Milky Way as a whole is $\approx 2 \, M_\odot \, \text{yr}^{-1}$ (Chomiuk & Povich 2011), so this star formation rate implied by a successful model must not exceed this value. To see what this implies, we again return to the ATLASGAL sample. As noted above, the mean free-fall time of these objects is $\tau_{ff} = 0.3$ Myr (Heyer et al. 2016), and the total mass of ATLASGAL clouds in the Galaxy is $M_{tot} \approx 1.0 \times 10^7 \, M_\odot$ (Urquhart et al. 2018). The rate at which these clumps form stars is straightforward to calculate in the ST, CB, and CBD models:

$$\text{SFR} = \epsilon_{ff} \frac{M_{tot}}{\tau_{ff}} = 0.33 \left( \frac{\epsilon_{ff}}{0.01} \right) \left( \frac{M_{tot}}{10^7 M_\odot} \right) \left( \frac{\tau_{ff}}{0.3 \text{ Myr}} \right)^{-1} \, M_\odot \, \text{yr}^{-1},$$

where we have normalised to the mean free-fall time for the ATLASGAL clumps. Thus if $\epsilon_{ff} \approx 0.01$, for these models, as suggested by our analysis so far, the total contribution of the ATLASGAL clumps to the total star formation budget of the Milky Way is $\approx 0.3 \, M_\odot \, \text{yr}^{-1}$, which is $\approx 10\%$ of the total. This is consistent with the upper limit stated above, and in fact suggests a nice consistency: the ATLASGAL clumps are much denser than the mean star-forming region or star cluster (for example, compare to Fig. 9 of Krumholz et al. 2019), and thus the stars that form within them are much more likely to remain part of a bound cluster than the typical star formed in the Galaxy. If we hypothesise that the ATLASGAL clumps correspond roughly to the bound portion of the star formation in the Galaxy, so our estimate implies that $\approx 10\%$ of all stars formed in bound clusters, which is entirely consistent with the observationally-measured fraction of stars formed in bound clusters in typical spiral galaxies (e.g., Ryon et al. 2014; Adamo et al. 2015; Johnson et al. 2016; Chandar et al. 2017). We caution, however, not to put too much weight on this agreement, since we do not in fact know if the density range that is selected by ATLASGAL corresponds well to the conditions that delineate between bound and unbound star formation.

The GHC and IE models require a more refined treatment because $\tau_{ff}$ and $\epsilon_{ff}$ can vary. Since these models do not depend on the magnitude of the mass, we can assume that the entire population of ATLASGAL clouds is born with the same mass and then evolves according to one of these models. Let $N_M = dN / dM_g$ be the number of clouds per unit mass, and let $N$ be the rate at which clouds are born with a mass $M_{g,0}$. The equation of continuity for the cloud mass distribution is then

$$\frac{\partial N_M}{\partial t} + \frac{\partial}{\partial M_g} (N_M \dot{M}_g) = N \delta (M_g - M_{g,0}),$$

so that in a steady state we have

$$-N_M \dot{M}_g = \dot{N}.$$

The SFR is then

$$\text{SFR} = \int_0^{M_{g,0}} M_g N_M dM_g = \frac{\dot{N} M_{g,0}}{1 + \eta}$$

from equation 2, since there is no accretion in these models ($M_{acc} = 0$). Now the total mass in clouds is

$$M_{tot} = \int_0^{M_{g,0}} M_g N_M dM_g = \frac{\dot{N} M_{g,0}}{1 + \eta},$$

where

$$\text{SFR} = \frac{M_{tot}}{\tau_{ff}} = \epsilon_{ff} \frac{M_{tot}}{\tau_{ff}}$$

First consider the IE model. Evaluating the integral in equation 45 with the aid of equation 27 we find

$$\text{SFR} = \frac{M_{tot}}{\tau_{ff}} = \epsilon_{ff} \frac{M_{tot}}{\tau_{ff}}$$

$$= \frac{\Gamma (1 + \frac{\delta}{1 + \delta}) (1 + \frac{1}{1 + \delta})^{\epsilon_{ff} \frac{M_{tot}}{\tau_{ff}}}}{(1 + \delta)^{1/1 + \delta}} \Gamma (1 + \frac{1}{1 + \delta})^{\epsilon_{ff} \frac{M_{tot}}{\tau_{ff}}}.$$
where in the second step we have made use of equation 28 to rewrite the star formation rate in terms of the mass-averaged star formation efficiency $e$. We have already noted that the constraints on $\delta$ arising from stellar age distributions are inconsistent with those derived from YSO counts, but the total star formation rates in both cases are similar. Consulting Table 3 and Table 4, we see that YSO counts give $e_{\text{Y}} \approx 0.01$ and $\delta \approx 0$, so overall we obtain $SFR \approx 0.01 M_{\text{tot}}/t_{\text{ff}}$.

Stellar ages give $e_{\text{St}} \approx 0.05$ and $\delta \approx 3$, which again gives $SFR \approx 0.01 M_{\text{tot}}/t_{\text{ff}}$. Thus the global star formation rate predicted by our best-fitting values of the IE model are roughly the same as those obtained in the ST, CB, or CBD models, and is consistent with the global star formation budget of the Milky Way.

Next consider the GHC model. In this case, evaluation of the integral in equation 45 gives

$$SFR = \frac{1 + \tau_{\text{coll}}}{\tau_{\text{coll}}} \left( \frac{M_{\text{tot}}}{t_{\text{ff},0}} \right).$$

(49)

The free-fall time at the onset of collapse $t_{\text{ff},0}$ is unknown, as is the total mass $M_{\text{tot}}$ at the start of collapse – ATLASGAL tells us only the instantaneous mass of clumps whose density is high enough for them to be included in the catalogue, i.e., those for which $t_{\text{ff}} \lesssim 0.3$ Myr. To make use of the ATLASGAL constraint, we can generalise equation 44 to describe not the total mass in all clouds, but only the mass in clouds denser enough that their free-fall time is at or below some threshold value $t_{\text{ff}}$:

$$M_{\text{tot}}(< t_{\text{ff}}) = N_{\text{tot}} \int_{0}^{\infty} M_{x} H(t_{\text{ff}} - t_{\text{ff}}(\tau)) d\tau,$$

(50)

where $t_{\text{ff}}(\tau)$ is the free-fall time some dimensionless time $\tau$ after the onset of collapse, and $H(\tau)$ is the usual Heaviside function, which is unity for positive argument and zero for negative arguments. In the case of the GHC model, for which $t_{\text{ff}}(\tau)$ is a monotonically-decreasing function, this reduces to

$$M_{\text{tot}}(< t_{\text{ff}}) = N_{\text{tot}} \int_{t_{\text{ff},0}}^{\infty} M_{x} d\tau = \frac{\tau_{\text{coll}}}{1 + \tau_{\text{coll}}} (1 - x)\tau_{\text{coll}}^{-1},$$

(51)

where $x = t_{\text{ff}}/t_{\text{ff},0}$. The corresponding star formation rate, analogous to equation 45, is then

$$SFR = \frac{1 + \tau_{\text{coll}}}{\tau_{\text{coll}}(1 - x)\tau_{\text{coll}}^{-1}} \left( \frac{M_{\text{tot}}(< t_{\text{ff}})}{t_{\text{ff}}} \right).$$

(52)

The term $(1 - x)\tau_{\text{coll}}^{-1}$ in the denominator is strictly less than unity, and so we can obtain a lower limit on the star formation rate by setting it to unity exactly. Thus the observed mass $M_{\text{tot}}$ and free-fall time $t_{\text{ff}}$ of the ATLASGAL clumps implies a lower limit on the star formation rate

$$SFR \gtrsim 11 \left( \frac{t_{\text{ff}}}{0.01} \right) \left( \frac{\tau_{\text{coll}}}{0.05} \right)^{-1} \left( \frac{M_{\text{tot}}}{10^{5} M_{\odot}} \right) \left( \frac{t_{\text{ff}}}{0.3} \text{ Myr} \right)^{-1} \left( \frac{M_{\odot}}{M_{\odot}} \right) \text{ yr}^{-1}.$$
the rate at which the gas produces YSOs. We carry out a Bayesian forward-modelling treatment of the observational uncertainties and possible biases in these data set, including the effects of selecting only gas-dominated systems and of changes in the gas properties on timescales shorter than the YSO lifetime, and we find that these factors do not significantly alter the overall constraint on how efficiently gas produces YSOs. The tight correlation of gas properties with YSO counts rules out the possibility that the star formation efficiency per free-fall time is time-dependent, ruling out models where the observed acceleration of star formation is due to a time-dependent increase in star formation efficiency per unit mass per unit free-fall time.

We finally consider the global star formation budget of the Milky Way, and show that the scenarios of global hierarchical collapse and conveyor belt star formation predict that the observed ATLASGAL clump population will yield very different total rates of star formation in the Galaxy. The collapse scenario is only able to recover the observed acceleration of star formation if clumps collapse globally on a timescale shorter than that on which they initially form stars locally, since otherwise depletion of the gas by star formation yields a star formation history that decelerates rather than accelerating. However, the requirement for global collapse to occur before a significant fraction of the mass can form stars in turn requires that the ATLASGAL clumps produce stars at a rate that exceeds the entire star formation rate of the Milky Way, let alone the substantially lower rate at which bound star clusters form.

By contrast, the conveyor belt model, first proposed by Longmore et al. (2014), encounters no such difficulties, because it attributes the acceleration of star formation to the fact that gas clumps form stars and accrete simultaneously, so that the gas mass available for star formation tends to increase with time until the gas is dispersed by feedback. We further find that accretion at a rate that varies with time as $M_{\text{acc}} \propto t^3$, as generically predicted for the gravitational collapse of mass reservoirs with fixed bounding pressure (Goldbaum et al. 2011), produces a distribution of stellar ages consistent with that observed in young clusters. We therefore conclude that the best available explanation for all of the available observational constraints is that bound star clusters form in a conveyor belt mode, where gas accretes at an increasing rate, but the central cluster-forming region is not in a state of global collapse, and has a star formation efficiency per unit mass that is both low and roughly constant in time.

ACKNOWLEDGEMENTS

MRK acknowledges funding from the Australian Research Council through the Future Fellowship (FT180100375) and Discovery Projects (DP190101285) funding schemes. CFM acknowledges support by NASA through NASA ATP grant NNX13AB84G. This research made use of Astropy,5 a community-developed core Python package for Astronomy (Astropy Collaboration et al. 2013, 2018).

REFERENCES

Adamo A., Kruisjes J. M. D., Bastian N., Silva-Villa E., Ryon J., 2015, MNRAS, 452, 246
Astropy Collaboration et al., 2013, A&A, 558, A33
Astropy Collaboration et al., 2018, AJ, 156, 123
Azimlu M., Martínez-Galarza J. R., Muench A. A., 2015, AJ, 150, 95
Beccari G., et al., 2017, A&A, 604, A22
Caldwell S., Chang P., 2018, MNRAS, 474, 4818
Chandar R., Fall S. M., Whitmore B. C., Mulia A. J., 2017, ApJ, 849, 128
Chomiuk L., Povich M. S., 2011, AJ, 142, 197
Csengeri T., et al., 2014, A&A, 565, A75
Da Rio N., Tan J. C., Jaehnig K., 2014, ApJ, 795, 55
Da Rio N., et al., 2016, ApJ, 818, 59
Dekel A., Krumholz M. R., 2013, MNRAS, 432, 455
Elmegreen B. G., 2000, ApJ, 530, 277
Evans N. J., et al., 2009, ApJS, 181, 321
Evans H. N. J., Heiderman A., Vutisalchavakul N., 2014, ApJ, 782, 114
Fűrész G., Hartmann L. W., Megeath S. T., Szentgyorgyi A. H., Hamden E. T., 2008, ApJ, 676, 1109
Foreman-Mackey D., Hogg D. W., Lang D., Goodman J., 2013, PASP, 125, 306
Ginsburg A., Bressert E., Bally J., Battersby C., 2012, ApJ, 758, 129
Goldbaum N. J., Krumholz M. R., Matzner C. D., McKee C. F., 2011, ApJ, 738, 101
Gutermuth R. A., Megeath S. T., Myers P. C., Allen L. E., Piper J. L., Fazio G. G., 2009, ApJS, 184, 18
Gutermuth R. A., Piper J. L., Megeath S. T., Myers P. C., Allen L. E., Allen T. S., 2011, ApJ, 739, 84
Heyer M., Gutermuth R., Urquhart J. S., Csengeri T., Wienen M., Leurini S., Menten K., Wyrowski F., 2016, A&A, 588, A29
Hillenbrand L. A., Hartmann L. W., 1998, ApJ, 492, 540
Huff E. M., Stahler S. W., 2006, ApJ, 644, 355
Jaehnig K. O., Da Rio N., Tan J. C., 2015, ApJ, 798, 126
Jeffries R. D., 2007, MNRAS, 381, 1169
Jeffries R. D., 2017, Mem. Soc. Astron. Italiana, 88, 637
Johnson L. C., et al., 2016, ApJ, 827, 33
Kharchenko N. V., Piskunov A. E., Schilbach E., Röser S., Scholz R.-D., 2013, A&A, 558, A33
Kim D., Lu J. R., Konopacky Q., Chu L., Toller E., Anderson J., Thiessen C. A., Morris M. R., 2019, AJ, 157, 109
King I. R., 1962, AJ, 67, 471
Klessen R. S., Burkert A., 2000, ApJS, 128, 287
Koukel M., et al., 2018, AJ, 156, 84
Kroupa P., Aarseth S., Hurley J., 2001, MNRAS, 321, 699
Kruisjes J. M. D., Dale J. E., Longmore S. N., 2015, MNRAS, 447, 1069
Kruisjes J. M. D., et al., 2019, Nature, 569, 519
Krumholz M. R., Tan J. C., 2007, ApJ, 654, 304
Krumholz M. R., Dekel A., McKee C. F., 2012, ApJ, 745, 69
Krumholz M. R., McKee C. F., Bland-Hawthorn J., 2019, MNRAS, 482, L29
Lee Y.-N., Hennebelle P., 2016b, A&A, 591, A31
Lee Y.-N., Hennebelle P., 2016a, A&A, 591, A30
Lada C. J., Lombardi M., Roman-Zuniga C., Forbrich J., Alves J. F., 2013, ApJ, 778, 133
Lee Y.-N., Hennebelle P., 2016a, A&A, 591, A30

5 http://www.astropy.org

MNRS 000, 1–18 (2019)
APPENDIX A: SUPPLEMENTARY MATERIAL
(ONLINE-ONLY)

In Figures A1 - A11, we provide full posterior PDFs resulting from our MCMC fits of all models to the stellar age distributions in the ONC and NGC 6530. These plots are all for our fiducial errors \((\sigma, b) = (0.12, -0.05)\) dex, and the PDFs shown are derived using samples from the final 400 iterations of the MCMC. In Figures A13 - A15 we provide the full combined posterior PDFs derived from stellar age distributions and YSO counts in ATLASGAL clumps for the ST, CB \((p = 0)\), and CB models.

This paper has been typeset from a TeX/LaTeX file prepared by the author.
Figure A1. Corner plot showing the distributions of both directly fit parameters and derived parameters from our MCMC application of model ST to the stellar age distribution in the ONC. The central panels show heat maps of the probability density in two dimensional cuts through the indicated axes, each scaled to have a maximum of unity; the outer contour marks a scale probability density of 0.01, and individual points outside this contour correspond to individual MCMC samples. Panels containing histograms show the marginal posterior probability distributions for each parameter. The parameters shown are the same as those listed in Table 3, and are computed in the same way, i.e., only $t_{\text{clust}}$ and $t_{\text{sf}}$ are fit as part of the MCMC, while all other quantities are derived from them; this is why, for example, $\epsilon_{\text{ff}}$ and $t_{\text{sf}}$ are perfectly correlated.
Figure A2. Same as Figure A1, but for model CB with $p = 0$. 
Figure A3. Same as Figure A1, but for model CB with $p = 3$. 

Bound star cluster formation 21
Figure A4. Same as Figure A1, but for model CBD with $p = 3$. 
Figure A5. Same as Figure A1, but for model IE.
Figure A6. Same as Figure A1, but for model GHC.
Figure A7. Same as Figure A1, but for NGC 6530 rather than the ONC.
Figure A8. Same as Figure A7, but for model CB with $\rho = 0$. 
Figure A9. Same as Figure A7, but for model CB with $p = 3$. 
Figure A10. Same as Figure A7, but for model CBD with $p = 3$. 

$\log t_{sf}$ [Myr] $0.0$ $0.5$ $1.0$ $1.5$ $2.0$

$\log t_{acc}$ [Myr] $0.0$ $0.5$ $1.0$ $1.5$ $2.0$

$\log \phi_d$ $0.0$ $0.5$ $1.0$

$\log \epsilon_{ff}$ $0.25$ $0.50$ $0.75$

$\log f_{g,\text{clust}}$ $0.50$ $0.75$ $1.0$

$\log \epsilon_{*,\text{clust}}$ $-0.5$ $-1.0$ $-1.5$ $-2.0$ $-2.5$

$\log t_{\text{clust}}$ [Myr] $0.75$ $1.0$ $1.25$ $1.5$ $1.75$

$\log t_{\text{sf}}$ [Myr] $0.75$ $1.0$ $1.25$ $1.5$ $1.75$

$\log \phi_d$ $0.25$ $0.50$ $0.75$

$\log \epsilon_{*,\text{clust}}$ $-0.5$ $-1.0$ $-1.5$ $-2.0$ $-2.5$

Figure A10. Same as Figure A7, but for model CBD with $p = 3$. 

$\log t_{\text{clust}}$ [Myr] $0.75$ $1.0$ $1.25$ $1.5$ $1.75$

$\log t_{\text{sf}}$ [Myr] $0.75$ $1.0$ $1.25$ $1.5$ $1.75$

$\log \phi_d$ $0.25$ $0.50$ $0.75$

$\log \epsilon_{*,\text{clust}}$ $-0.5$ $-1.0$ $-1.5$ $-2.0$ $-2.5$
Figure A11. Same as Figure A7, but for model IE.
Figure A12. Same as Figure A7, but for model GHC. Note that $\xi$ is not directly fit, but is derived from the fit parameters. However, we provide it for convenience.
Figure A13. Same as Figure 4 of the main text, but for the ST model.

Figure A14. Same as Figure 4 of the main text, but for the CB model.

Figure A15. Same as Figure 4 of the main text, but for the CB model.