Monthly Notices of the Royal Astronomical Society

MNRAS **551**, 1–16 (2026) https://doi.org/10.1093/mnras/stag1547
Advance Access publication 2026 August 14

Revisiting gravitational instability in protostellar discs with improved radiative cooling models

Alison K. Young[1,2,3,4,5]★, Ken Rice[2,3] Richard Booth[1] and Farzana Meru[4,5]

[1]School of Physics and Astronomy, University of Leeds, Sir William Henry Bragg Building, Woodhouse Ln, Leeds LS2 9JT, UK
[2]SUPA, Institute for Astronomy, University of Edinburgh, The Royal Observatory, Blackford Hill, Edinburgh EH9 3HJ, UK
[3]Centre for Exoplanet Science, University of Edinburgh, Edinburgh EH9 3HJ, UK
[4]Centre for Exoplanets and Habitability, University of Warwick, Coventry CV4 7AL, UK
[5]Department of Physics, University of Warwick, Coventry CV4 7AL, UK

Accepted 2026 August 10. Received 2026 July 7; in original form 2026 April 27

ABSTRACT

Young discs are expected to be significantly more massive than those observed at > 1 Myr, and it is at this earliest stage that planet formation likely begins. Such massive discs may be susceptible to the gravitational instability (GI); therefore, we need to determine the disc and stellar properties for which the GI is active to understand its role in early disc evolution and planet formation. Prior work has been limited by model assumptions and inaccuracies due to the complex nature of the thermodynamics of protostellar discs, so we now revisit this question using an improved method to approximate radiative cooling within hydrodynamics simulations. We have explored a wide parameter space, representative of young protostellar discs of 0.1 to 1 $M_{\odot}$ and include irradiation from the host star. The parameters for which discs form spirals and fragment were found to differ to those obtained from earlier simulations. The outer regions of discs with radii of 50 au may be susceptible to fragmentation, meaning that GI-driven planet formation is not restricted to only the most extended discs. The additional thermal support due to stellar irradiation increases the disc mass that remains stable against GI: discs may reach up to $\gtrsim 0.4\ M_*$ without fragmenting, providing a considerable quantity of material for building planets. Large-scale spiral arms only developed for $M_* \lesssim 0.3\ M_{\odot}$, except in the most compact discs. Furthermore, the long-lived spiral structures that form tend to be flocculent and compact, indicating that large-scale spiral arms should not be considered a typical outcome of GI.

Key words: hydrodynamics – radiative transfer – protoplanetary discs – stars: formation.

1 INTRODUCTION

Protostellar discs form with relatively high masses that can be a significant fraction of the mass of the host protostar, while the protostar itself has yet to reach its final mass (J. K. Jørgensen et al. 2009; J. P. Williams & L. A. Cieza 2011). With such a high mass fraction ($0.1 \lesssim m_{\rm disc}/m_* \lesssim 0.3$), young protostellar discs may be susceptible to the growth of the gravitational instability (GI; G. Laughlin & P. Bodenheimer 1994; K. Rice 2016; M. J. Maureira et al. 2026), which could accelerate planet formation via the direct collapse of fragments and/or through assisting the growth of dust grains (A. P. Boss 1998; L. Mayer et al. 2004; W. K. M. Rice et al. 2006; C. Longarini et al. 2023; S. Rowther et al. 2024a; K. Rice et al. 2025). It is at this early stage during the Class 0/I embedded phase (< 1 Myr) that planet formation is expected to begin (C. J. Nixon, A. R. King & J. E. Pringle 2018; D. M. Segura-Cox et al. 2020). Therefore, we need to understand the evolution of discs subject to the GI to make progress in explaining how planet formation begins.

The stability parameter known as 'Toomre's Q' indicates whether a disc's self-gravity could lead to the development of the gravitational instability and is given by

$$Q = \frac{c_s \kappa}{\pi G \Sigma}, \quad (1)$$

(A. Toomre 1964), where $c_{\rm s}$ is the sound speed, κ is the epicyclic frequency, and Σ is the disc surface density. In the razor-thin disc limit, the GI becomes active when $Q = 1$ (V. S. Safronov 1960; A. Toomre 1964) but for realistic discs observable effects of GI may manifest when $Q \lesssim 1.5$ (P. Goldreich & D. Lynden-Bell 1965; P. O. Vandervoort 1970; G. Bertin & A. B. Romeo 1988). A disc that is susceptible to the growth of GI will tend to a quasi-steady state in which shock heating due to the GI quenches the instability (B. Paczynski 1978). This self-regulation mechanism acts as a thermostat that maintains $Q \sim 1$ in the disc and allows spiral density waves to emerge that can act to transport angular momentum (e.g. G. Lodato & W. K. M. Rice 2004). If, however, cooling is sufficiently rapid, or mass accretion sufficiently fast, Q may decrease well below 1 and the disc may undergo fragmentation. The cooling rate is typically evaluated by comparing the cooling time to the Keplerian frequency $\Omega_{\rm K}$ at a given radius in

★ E-mail: a.k.young@leeds.ac.uk

the disc (C. F. Gammie 2001):

$$\beta_{\rm cool} = \Omega_{\rm K} t_{\rm cool}. \tag{2}$$

Sufficiently rapid cooling drives fragmentation of the disc since GI cannot—in a quasi-steady state—provide enough heating to stabilize the disc. A cooling parameter $\beta_{\rm cool} \lesssim 8$ will typically result in fragmentation (e.g. C. F. Gammie 2001; W. K. M. Rice et al. 2011; C. S. Leedham, R. A. Booth & C. J. Clarke 2025). A longer cooling time-scale of $\beta_{\rm cool} \gtrsim 10$–20 maintains the quasi-steady state (K. Kratter & G. Lodato 2016).

Of course, discs are irradiated by their host star(s), which significantly affects the thermal balance and raises the mid-plane temperature. The degree to which stellar irradiation dominates the thermal balance has been studied with various analytical models and numerical simulations. With stellar irradiation, discs can reach higher masses before becoming unstable (C. D. Matzner & Y. Levin 2005; K. Cai et al. 2008; F. Meru & M. R. Bate 2010; D. Forgan & K. Rice 2013; T. J. Haworth et al. 2020; J. Cadman et al. 2020a). However, once the GI emerges, the maximum value of gravitational stress ($\alpha_{\rm grav}$) that can be sustained is lower for an irradiated disc and the critical value of $\beta_{\rm cool}$ necessary for fragmentation is reduced by around a factor of 2 with irradiation (W. K. M. Rice et al. 2011). The result is that a long-lived self-regulated state is unlikely to be reached in irradiated discs since GI shock heating and viscous heating comprise a smaller fraction of the thermal balance. Hence, when the system is susceptible to the GI, they are probably insufficient to prevent rapid growth of the instability (K. M. Kratter & R. A. Murray-Clay 2011; S. Rowther et al. 2024b; C. S. Leedham et al. 2025).

The disc temperature and the cooling rate are functions of radius, with lower temperatures and lower values of $\beta_{\rm cool}$ reached further from the host star. Consequently, it was expected that the warm inner disc is stable and that fragmentation is only possible in the outer regions beyond ~ 70 au (R. R. Rafikov 2005, 2009; C. J. Clarke 2009; M.-K. Lin & K. M. Kratter 2016). Under these conditions, disc fragmentation due to GI is unlikely to form many planetary-mass objects (K. M. Kratter, R. A. Murray-Clay & A. N. Youdin 2010; K. Rice et al. 2015; D. H. Forgan et al. 2018; O. Schib et al. 2025). However, once magnetic fields are included, the formation of sub-Jovian mass planets may be possible (H. Deng, L. Mayer & R. Helled 2021). There are indications, too, that fragmentation may be possible a little closer to the host star than 70 au (A. Mercer & D. Stamatellos 2020; C. S. Leedham et al. 2025), and as close as ~ 30 au when dust grain growth is considered (H. Lee, S. Nayakshin & R. A. Booth 2025). Moreover, F. Meru (2015) demonstrated that the gas motions in the disc triggered by the local collapse of a fragment may in turn cause fragmentation within 50 au.

The evolution of gravitationally unstable discs sensitively depends on the balance of heating and cooling (C. F. Gammie 2001). In most of the early work, the cooling rate was imposed on the model disc by means of a fixed value of $\beta_{\rm cool}$, the ratio of cooling time to orbital period (see equation (2)). A constant $\beta_{\rm cool}$ is a poor approximation to the cooling in protostellar discs (A. Mercer, D. Stamatellos & A. Dunhill 2018) and so some simulations have employed a radial-dependent $\beta_{\rm cool} = \beta_{\rm cool}(R)$ (S. Rowther & F. Meru 2020) or temperature-dependent $\beta_{\rm cool} = \beta_{\rm cool}(T)$ (P. Cossins, G. Lodato & C. Clarke 2010). These approaches capture better the variation in cooling times throughout the disc and tend to produce discs that are more susceptible to fragmentation.

The 'β-cooling' method has proved a useful approximation for studying the evolution of massive discs but further progress with β-cooling simulations is limited because the cooling rate cannot evolve with the changing surface density and the development of structures in the disc. The cooling approximation introduced by D. Stamatellos et al. (2007) links the cooling rate directly to the local structure by assuming that each parcel of gas is embedded within a polytropic spherical pseudo-cloud, providing a physically motivated approach for a modest additional computational cost. This method works well for spherical geometries, offering significant improvement on both β-cooling and barotropic models (D. Stamatellos & A. P. Whitworth 2009a), and was developed further by J. C. Lombardi, W. G. McInally & J. A. Faber (2015) with a different approach to estimating the depth of a gas parcel within the pseudo-cloud. This effectively improved the accuracy of the method for disc geometries (A. Mercer et al. 2018).

The same approach was recently employed in the 'modified Lombardi' method developed by A. K. Young et al. (2024) but now specifically tailored for self-gravitating discs. The modified Lombardi method involves estimating the column density above a parcel of gas from the pressure gradient and then adjusting for the reduction in scale height of a self-gravitating disc compared to a passive disc in which the scale height is set exclusively by stellar heating. This method produces excellent estimates of the column density throughout the disc and therefore can give accurate estimates of the optical depth in thick discs, from which the radiative cooling rate (or heating rate) can be calculated. This cooling model can be coupled with radiation transport within the disc using the flux-limited diffusion (FLD) approximation in a 'hybrid' method (D. Forgan et al. 2009). The transfer of energy between neighbouring fluid elements in optically thick regions is therefore treated in addition to the radiative cooling that dominates where $\tau \sim 1$. Of course, a full radiative transfer calculation is necessary for complex structures that produce shadows, for example. The advantage of the hybrid radiative cooling approximation and FLD is the greatly reduced computational expense compared to full radiative transfer while maintaining accuracy for the disc structures we wish to study.

The properties of the disc and star determine its stability to GI. Of particular interest is the range of disc-to-star mass ratios and radial extents for which a disc is susceptible to fragmentation or the development of spiral waves when realistic stellar heating is considered. By knowing this, we can infer whether direct fragmentation is a likely planet formation pathway, the influence of stellar heating on the disc's behaviour, and at what stages spiral waves might drive dust trapping, for example. In this paper, we re-examine the parameter space to investigate the properties of discs that develop GI using the more accurate approximate cooling treatment of A. K. Young et al. (2024).

The method, including the implementation of the radiative cooling approximation with four different approaches, is described in detail in Section 2. We then test the polytropic cooling approximation methods introduced by A. K. Young et al. (2024) and compare the outcome of simulations conducted with each of the methods. These are presented in Section 3.1. Next, we conduct a parameter study with the most accurate of those methods ('modified Lombardi') to explore whether and how the range of disc and stellar parameters for which the GI is active changes compared to prior work (Section 3.2). The key results are discussed in detail with reference to observations and future modelling in Section 4 and are summarized in Section 5.

2 NUMERICAL METHOD

The simulations were performed with a version of PHANTOM (D. J. Price et al. 2018) that has been modified to combine a radiative cooling approximation with radiative transfer using the flux-limited diffusion method.[1] This hybrid approach was introduced by D. Forgan et al. (2009), and here we implement FLD in tandem with improved versions of the radiative cooling approximation that were described in A. K. Young et al. (2024).

2.1 Radiative cooling approximation

The radiative cooling rate is approximated by estimating the optical depth throughout the disc. This approach was introduced by D. Stamatellos et al. (2007), extended to include flux-limited diffusion (FLD) radiative transfer by D. Forgan et al. (2009), and further refined by J. C. Lombardi et al. (2015). A key limitation of this approach is the accuracy of the optical depth estimate, derived from either the local gravitational potential or the local pressure gradient. A. K. Young et al. (2024) introduced two new approaches which give estimates of the local column density far closer to that calculated directly from the particle distribution. A more accurate value of column density gives a much better estimate of the optical depth, and therefore of the radiative cooling (or heating) rate. The four methods for estimating the pseudo-mean column density $\bar{\Sigma}_i$ are described below.

(i) Stamatellos method. The column density is estimated from the local gravitational potential ψ_i and density ρ_i of particle i (D. Stamatellos et al. 2007):

$$\bar{\Sigma}_i = \zeta_n \left[\frac{-\psi_i \rho_i}{4\pi G}\right]^{1/2}. \tag{3}$$

ζ_n is a factor that depends on the polytropic index n (see D. Stamatellos et al. 2007). Here, we use $\zeta_2 = 0.368$.

(ii) Lombardi method. The local pressure P_i and hydrodynamic acceleration, $\boldsymbol{a}_{\mathrm{h},i} = -\nabla P_i/\rho_i$, which contains the pressure gradient, can be shown to give (J. C. Lombardi et al. 2015):

$$\bar{\Sigma}_i = \frac{\zeta' P_i}{|\boldsymbol{a}_{\mathrm{h},i}|}, \tag{4}$$

where the factor $\zeta' = 1.014$

(iii) Combined method. The pressure scale height is obtained for the Stamatellos method of estimating $\bar{\Sigma}_i$ (equation 3),

$$H_{\mathrm{S},i} = \bar{\Sigma}_i / \zeta' \rho_i. \tag{5}$$

Then similarly, we obtain the scale height via the Lombardi method,

$$H_{\mathrm{L},i} = \frac{P_i}{\rho_i |\boldsymbol{a}_{\mathrm{h},i}|}. \tag{6}$$

These two values are then averaged in inverse quadrature:

$$H_{\mathrm{C}} = \left(H_{\mathrm{P}}^{-2} + H_{\mathrm{S}}^{-2}\right)^{-1/2}. \tag{7}$$

This improves the estimates of scale height and the derived column density since the result tends to the lower of the two values. This means that for most of the disc, the Lombardi estimate dominates but where the pressure gradient approaches zero (near the mid-plane), the Stamatellos estimate takes over (A. K. Young et al. 2024).

(iv) Modified Lombardi: The value obtained with the Lombardi method is averaged in quadrature with the analytical value of scale height in a self-gravitating disc. In a self-gravitating disc, the scale height H_0 is reduced compared to a non-self-gravitating disc in which $H_* = c_s/\Omega_{\mathrm{K}}$, where c_s is the local sound speed. The ratio can be found to be (A. K. Young et al. 2024)

$$\frac{H_0}{H_*} = \frac{\sqrt{\pi/2}}{\sqrt{1 + 1/\left(Q_{\mathrm{3D}}\sqrt{\pi/2}\right)}}. \tag{8}$$

Here,

$$Q_{3D} = \Omega_{\mathrm{K}}^2 / 4\pi G \rho(0), \tag{9}$$

is the 3D Toomre Q parameter (G. R. Mamatsashvili & W. K. M. Rice 2010), with the mid-plane density $\rho(0)$.

The modified Lombardi scale height is calculated similarly to equation (7) but replacing H_{S} with H_0.

In Section 3.1, we compare the evolution of a disc under the four methods. In the rest of the paper, we apply the modified Lombardi method since this provides the closest estimate of local column density to the actual value, both in the disc and within clumps.

Once we have estimated the column density, we can find the optical depth. The pseudo-mean optical depth is given by

$$\bar{\tau} = \bar{\Sigma}_i \bar{\kappa}_i(\rho_i, T_i). \tag{10}$$

A look-up table of the pseudo-mean mass opacity $\bar{\kappa}_i(\rho_i, T_i)$ is pre-calculated by averaging the Rosseland mean opacity over the pseudo-cloud; see D. Stamatellos et al. (2007) and J. C. Lombardi et al. (2015) for the details.

The radiative cooling rate is then:

$$\left.\frac{\mathrm{d}u_i}{\mathrm{d}t}\right|_{\mathrm{rad}} = \frac{4\sigma_{\mathrm{B}}\left(T_0^4(\boldsymbol{r}_i) - T_i^4\right)}{\bar{\Sigma}_i^{\,2}\bar{\kappa}_i(\rho_i, T_i) + \kappa_i^{-1}(\rho_i, T_i)}, \tag{11}$$

where σ_{B} is the Stefan–Boltzmann constant, T_i is the temperature of the ith particle, and $\kappa_i^{-1}(\rho_i, T_i)$ is the Planck mean opacity. T_0 is the background temperature towards which the parcel of gas is cooling or heating. Here, this term is set by considering the stellar luminosity L_* so as to incorporate the effect of stellar heating:

$$T_{0,i}^4 = T_{\mathrm{floor}}^4 + \exp(-\Sigma_i \bar{\kappa}_i)\frac{L_*}{16\pi\sigma_{\mathrm{B}} r_i^2}. \tag{12}$$

The background heating due to the interstellar radiation field is assumed to give $T_{\mathrm{floor}} = 5$ K and the additional local stellar heating at radius r is found using the stellar luminosity L_*, the local column density Σ_i and the mean opacity $\bar{\kappa}_i$ (see D. Stamatellos et al. 2007). While the attenuation factor $\exp(-\Sigma_i\bar{\kappa}_i)$ significantly reduces the heating effect of stellar irradiation near the mid-plane, heat is transported from higher up by the diffusion term.

The equilibrium temperature is determined by assuming

$$\left.\frac{\mathrm{d}u_i}{\mathrm{d}t}\right|_{\mathrm{hydro}} + \left.\frac{\mathrm{d}u_i}{\mathrm{d}t}\right|_{\mathrm{rad}} + \left.\frac{\mathrm{d}u_i}{\mathrm{d}t}\right|_{\mathrm{FLD}} = 0, \tag{13}$$

where 'hydro' refers to the $pdV+$ viscous contribution.

The equilibrium temperature is thus

$$T_{\mathrm{eq,i}}^4 = T_0^4(\boldsymbol{r}_i) + \frac{1}{4\sigma}\left(\bar{\Sigma}^2\bar{\kappa}_i(\rho_i, T_i) + \kappa_i^{-1}(\rho_i, T_i)\right) \left[\left.\frac{\mathrm{d}u_i}{\mathrm{d}t}\right|_{\mathrm{hydro}} + \left.\frac{\mathrm{d}u_i}{\mathrm{d}t}\right|_{\mathrm{FLD}}\right]. \tag{14}$$

[1]The code is available from https://github.com/alisonkyoung1/phantom.

Table 1. Stellar luminosity used for the simulations assuming an age of 0.5 Myr.

M_* ($\mathrm{M}_\odot$)	L_* ($\mathrm{L}_\odot$)
0.1	0.074
0.2	0.275
0.3	0.575
0.5	1.143
0.7	1.900
1.0	3.310

We can then set the equilibrium internal energy $u_{\mathrm{eq,i}} = u(T_{\mathrm{eq},i}, \rho_i)$. A thermal time-scale is defined:

$$t_{\mathrm{therm},i} = \left(u_{\mathrm{eq,i}} - u_i\right)\left[\left.\frac{\mathrm{d}u_i}{\mathrm{d}t}\right|_{\mathrm{hydro}} + \left.\frac{\mathrm{d}u_i}{\mathrm{d}t}\right|_{\mathrm{rad}} + \left.\frac{\mathrm{d}u_i}{\mathrm{d}t}\right|_{\mathrm{FLD}}\right]^{-1}. \tag{15}$$

The equilibrium internal energy and thermal time-scale are used to evolve the energy equation by considering the progress of the energy of gas particle i towards the equilibrium value.

We implement the energy update slightly differently to A. K. Young et al. (2024), solving the update directly in the leapfrog integrator, for the active hydrodynamical sub-time-step δt. For the forward sub-time-steps the energy is evolved with:

$$u_{i+\frac{1}{2}} = u_i \exp(-\delta t/t_{\mathrm{therm,i}}) + u_{\mathrm{eq,i}}\left[1 - \exp(-\delta t/t_{\mathrm{therm,i}})\right]. \tag{16}$$

For the reverse sub-time-steps:

$$u_{i-\frac{1}{2}} = \frac{u_i - u_{\mathrm{eq,i}}\left[1 - \exp(-\delta t/t_{\mathrm{therm,i}})\right]}{\exp(-\delta t/t_{\mathrm{therm,i}})}. \tag{17}$$

2.2 Simulations

Simulations are performed with the smoothed particle hydrodynamics (SPH) code PHANTOM (D. J. Price et al. 2018) using the radiative cooling method implemented by A. K. Young et al. (2024), and detailed in Section 2.1. PHANTOM is an established astrophysical code that has been used extensively to model gravitationally unstable discs with the β-cooling method (e.g. S. Rowther, R. Nealon & F. Meru 2023), the polytropic cooling approximation (e.g. J. Cadman et al. 2020a), and more recently with coupled ray-tracing radiative transfer (S. Rowther et al. 2024b).

The central star is modelled as a sink particle (M. R. Bate, I. A. Bonnell & N. M. Price 1995) with the accretion radius set to 1 au, independently of the stellar mass. The luminosity of the host star is set according to the MIST stellar evolution models at 0.5 Myr (J. Choi et al. 2016; A. Dotter 2016) following T. J. Haworth et al. (2020; see Table 1). The heating due to irradiation from this central star is implemented via equation (12).

The artificial viscosity parameter α_{AV} is varied between 0.01 and 1.0 using the switch of L. Cullen & W. Dehnen (2010) and we set $\beta_{\mathrm{AV}} = 2.0$ to keep numerical dissipation low (G. Lodato & C. J. Clarke 2011) and to prevent over-reducing viscosity at shocks (F. Meru & M. R. Bate 2012).

2.3 Initial conditions

We aim for an initial structure that is consistent with that of a self-gravitating and radiatively cooled/heated disc. This is to avoid an initial disc structure that is overdense and artificially unstable to fragmentation. To this end, we implement an altered setup routine in PHANTOM to initialize the sound speed c_{s} using the luminosity of the host star. As usual, the sound speed profile is initialized with

$$c_{\mathrm{s}}(r) = c_{\mathrm{s},0}(r/r_0)^{-q}, \tag{18}$$

where $c_{\mathrm{s},0}$ is the sound speed at $r = r_0$. The exponent $q = 1/4$, following the analytical temperature profile for a passively irradiated disc. The sound speed at r_0 is

$$c_{\mathrm{s},0} = \frac{1}{\sqrt{\pi/2}}\sqrt{\frac{k_{\mathrm{B}}}{\mu m_p}}\left[\left(\frac{L_*}{4\pi\sigma_{\mathrm{B}} r_0^2}\right) + T_{\mathrm{floor}}^4\right]^{\frac{1}{8}}. \tag{19}$$

The prefactor $1/\sqrt{\pi/2}$ is from $H/H_* = \sqrt{\pi/2}$ where H is the scale height of a self-gravitating disc and $H_* = c_{\mathrm{s}}/\Omega_{\mathrm{K}}$. This is found to produce sensible initial conditions (see Appendix A).

The initial surface density profile of the disc is set via

$$\Sigma(r) = \Sigma_0 \left(\frac{r}{r_0}\right)^{-p} \exp\left[\left(\frac{r}{r_c}\right)^{2-p}\right]\left(1 - \sqrt{1/r}\right), \tag{20}$$

where Σ_0 is the surface density at $r = r_0$ and r_c is the characteristic radius chosen to taper the density towards the outer edge. Σ_0 is set according to the chosen disc mass. We set $p = 1.0$ and taper the density profile with $r_c = 0.9 r_{\mathrm{out}}$ such that the radial density profile is stable as the disc begins to evolve. Unless otherwise stated, simulations are initialized with 5×10^5 SPH particles. This resolution is sufficient to determine the large-scale evolution of the disc, while also keeping run times short enough to undertake a parameter study of over 60 simulations (see Appendix B).

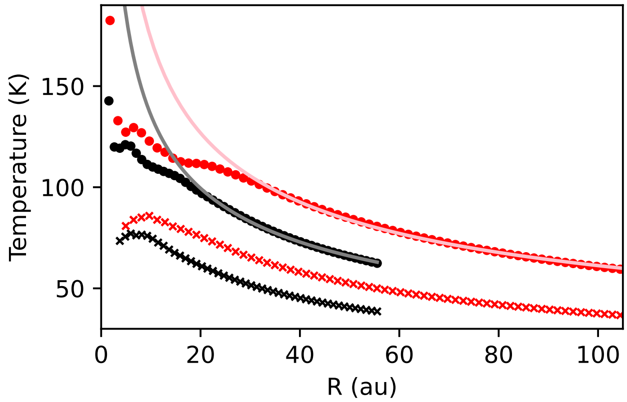

Figure 1. The radial temperature profiles for two simulated low-mass (i.e. largely passively irradiated) discs. The figure shows the mass-weighted vertically averaged temperature (circles), fitted profiles (solid lines), and mid-plane temperature (crosses). Black: star of 0.5 $\mathrm{M}_\odot$, disc of 0.005 $\mathrm{M}_\odot$, $L = 1.143$ $\mathrm{L}_\odot$, $T_{\mathrm{floor}} = 5$ K, and initial radius 50 au. Red: star of 1 $\mathrm{M}_\odot$, disc of 0.01 $\mathrm{M}_\odot$, $L = 3.31$ $\mathrm{L}_\odot$, $T_{\mathrm{floor}} = 5$ K, and initial radius 100 au. The fitted lines are $337r^{-0.41}$ (black) and $491r^{-0.45}$.

2.4 Passive disc test

First, we verify that temperatures are accurately estimated in the model by testing the method on passive discs, i.e. discs in which the temperature is set by the incident stellar radiation and the internal heating is negligible. Fig. 1 shows the radial profiles of the vertically averaged temperature for two simulated discs with $M_{\mathrm{disc}}/M_* = 0.01$. The simulation data were fitted with the profile $T = T_0 r^{-b}$, where T_0 is the temperature at $r = 1$ au, the disc inner edge. The temperature profiles are well described by power laws with $b = 0.41$ and $b = 0.45$. These values are close to the analytical value expected for passively irradiated discs ($b \approx 0.43$; E.

I. Chiang & P. Goldreich 1997). The crosses show the mid-plane temperature for each disc, found by averaging only particles with $|z| < 0.5H$. The temperature at the mid-plane is substantially reduced by the attenuation term in equation (12). Temperatures are underestimated in the inner regions of the disc but we do not expect this to affect the results. In Section 3.1, we compare the mid-plane temperatures with those derived from observations.

3 RESULTS

First, we present the results of comparing simulations with the four different approaches to estimating the optical depth in Section 3.1. Following this, the outcomes of the parameter study using the modified Lombardi are presented in Section 3.2. There, we describe the differences in disc evolution when simulated with the improved method.

In this analysis, we estimate $\beta_{\rm cool} = \Omega_{\rm K} t_{\rm cool}$ by using an estimate of the cooling time in annuli of the disc. In these simulations, heat is transferred by both flux-limited diffusion and radiative cooling via the polytropic cooling approximation. The cooling time-scale of a disc annulus at r is therefore given by

$$t_{\rm cool}(r) = \sum_i \frac{-(u_i - u_{0,i})}{\dot{u}_{{\rm rad},i} + \dot{u}_{{\rm FLD},i}}. \tag{21}$$

where u_i is the internal energy density of gas particle i, $u_0 = u(T_{0,i}, \rho_i)$ and $\dot{u}_{{\rm rad},i}$ and $\dot{u}_{{\rm FLD},i}$ are the heating rates due to radiative heating/cooling and FLD, respectively. Only particles with $|z| < 0.5H$ are included in the summation so we can probe cooling times near to the mid-plane. To allow comparison with the results of simulations that employed a variable $\beta_{\rm cool}$ of the form

$$\beta_{\rm cool}(r) = a(r/r_{\rm in})^{-b}, \tag{22}$$

we fit the values calculated from our simulations, $\log(|\beta_{\rm cool}|)$, with the function

$$f(r, a, b) = \log(a) - b\log(r/r_{\rm in}). \tag{23}$$

The Shakura–Sunyaev viscosity $\alpha_{\rm ss}$ provides a parametrized form of viscosity resulting from unresolved turbulence within a disc (I. Shakura & R. A. Sunyaev 1973). Gravitational stresses contribute to angular momentum transport in self-gravitating discs (D. Lynden-Bell & A. J. Kalnajs 1972), and it is useful to compare the resulting equivalent values of $\alpha_{\rm ss}$:

$$\alpha_{\rm grav}(r) = \left(\frac{d\ln\Omega_{\rm K}}{d\ln r}\right)^{-1} \frac{T^{\rm grav}_{r,\phi}}{\Sigma c_s^2}. \tag{24}$$

For each annulus in the disc, the gravitational stress $\alpha_{\rm grav}$ is estimated by summing the radial and azimuthal components of the gravitational acceleration g_r and g_ϕ over N particles in the annulus, which gives:

$$\alpha_{\rm grav}(r) = \left(\frac{d\ln\Omega_{\rm K}}{d\ln r}\right)^{-1} \frac{1}{4\pi G\langle c_s\rangle^2}\frac{1}{N}\sum_{i=1}^{N}\frac{g_{r,i}g_{\phi,i}}{\rho_i}. \tag{25}$$

3.1 Comparison of radiative cooling approximation methods

In Section 2.1, we described four approaches for a radiative cooling approximation. We now present the outcomes of simulations employing each method for comparison. Fig. 2 shows column density renderings of discs evolved using the four methods from four sets of initial conditions. In the first and third rows, the Stamatellos method leads to the gravitational instability, evident as spiral waves, while the other methods lead to a stable, axisymmetric disc. The simulations in the second and fourth rows display other differences in evolution. For the discs initialized with $M_* = 1\,{\rm M}_\odot$, $R_{\rm out} = 200$ au, and $M_{\rm d}/M_* = 0.6$ (Fig. 2, fourth row), the disc quickly fragments under the Stamatellos method but remains stable under the Lombardi and combined methods. With the modified Lombardi method, the disc maintains a stable spiral structure but forms a fragment after seven outer rotation periods. We now examine this latter model further in an effort to explain the differences in evolution.

The differences in the evolution are caused by the differences in estimates of optical depth, which affects both the minimum mid-plane temperature (c.f. equation 12) and how easily heat generated by the weak spiral shocks can be radiated away. The pseudo-mean optical depth $\bar{\tau}$ (equation 10), Q, effective $\beta_{\rm cool}$, mid-plane temperature and gravitational stress $\alpha_{\rm grav}$ for the simulations with $1\,{\rm M}_\odot$ star, $R_{\rm out} = 200$ au, and $M_{\rm d}/M_* = 0.6$ (bottom row of Fig. 2) are presented in Fig. 3. To study the radial profiles, these quantities are estimated in radial bins using equations (1), (21), and (25), respectively.

Disc mid-plane temperatures are difficult to infer from observations because of the high optical depths. Nevertheless, estimates for Class 0/I discs range from ∼20–70 K within ∼ 50 au (e.g. M. L. R. van't Hoff et al. 2018, 2020; J. Zamponi et al. 2021; S. Takakuwa et al. 2024; M. J. Maureira et al. 2026). We note that the disc mid-planes in our simulations with all methods are warm at ≈ 20–30 K, which is in agreement with observations of young discs.

The Stamatellos method gives rise to a largely optically thick disc, with the mid-plane optical depth dropping below $\bar{\tau} = 1$ at $R = 150$ au compared to within 100 au for the other methods. This is despite the discs having a similar surface density profile, and the reasons for the difference, discussed previously in A. K. Young et al. (2024) and A. Mercer et al. (2018), are to do with the assumption of spherical symmetry in estimating the column density from the gravitational potential, which results in an overestimate. The mid-plane temperature is correspondingly lower, due to the additional attenuation of the stellar irradiation. For all four simulations, $\beta_{\rm cool}$ is low enough to allow fragmentation where $Q \lesssim 1$. The cooler mid-plane obtained with the Stamatellos method allows Q to decrease such that the disc becomes unstable and fragments. In the other simulations, the temperature is just high enough for a quasi-stable spiral structure to be maintained. After several outer rotation periods, localized regions become unstable and collapse in the modified Lombardi simulation. In all four simulations, $\alpha_{\rm grav} > 0.1$ in most of the disc, showing that gravitational torques are driving the disc evolution, and $\alpha_{\rm grav}$ is notably higher for the Stamatellos simulation.

We estimate the radial exponent b of $\beta_{\rm cool}$ by fitting equation (23) between 70 and 200 au, and this is shown in the fourth panel of Fig. 3. $\beta_{\rm cool}$ decreases most steeply with the Stamatellos method, and for the other methods it flattens off towards in the outer disc.

For more stable discs, we find the evolution of the Lombardi, combined and modified Lombardi methods to be similar, but this often differs from the evolution under the Stamatellos method. In situations where the system is close to fragmentation, there may be small differences between the former three methods in whether fragments form, time to fragmentation, or number of fragments.

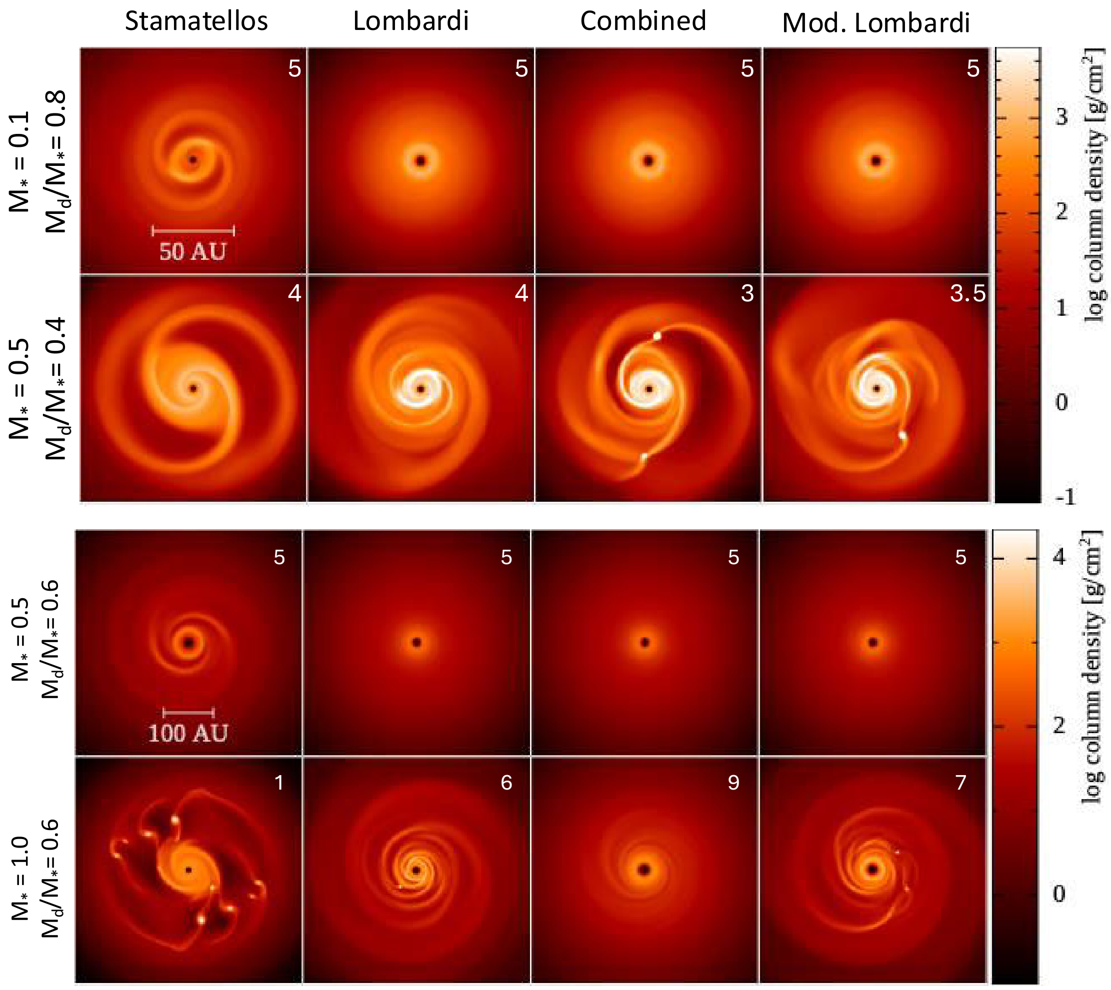

Figure 2. Snapshots from simulations performed with the four methods for four sets of initial conditions with $R_{\rm out}$ = 50 au (upper two rows) and $R_{\rm out}$ = 200 au (lower two rows). The timestamp of each snapshot is given in terms of outer rotation periods for each snapshot in white. If fragmentation occurred, the simulation was stopped shortly afterwards because the simulations run very slowly once clumps form. For the more stable disc-to-star mass ratios, the methods produce similar results but discs are generally more susceptible to the gravitational instability with the Stamatellos method. More accurate treatment of radiative cooling is more important for discs that are closer to instability.

3.2 Fragmentation and stability in discs around low mass stars

Following T. J. Haworth et al. (2020) and J. Cadman et al. (2020a), we simulated discs around stars of masses 0.25 $M_\odot$ to 1 $M_\odot$ including heating from the central star. We employ the same values of stellar luminosity that they estimated from the MIST stellar evolution tracks at 0.5 Myr (J. Choi et al. 2016; A. Dotter 2016; Table 1). Like J. Cadman et al. (2020a) and T. J. Haworth et al. (2020), we simulate discs with initial outer radii of 50, 100, and 200 au. The full initial parameters are given in Table 2 and Section 2.3. We initially ran the simulations for 5 outer rotation periods (ORP) but found that the structure evolved further after this time in some cases. For example, the 0.2 $M_\odot$, 50 au disc around a 1.0 $M_\odot$ star remained axisymmetric for nearly 9 ORPs before a spiral developed. Interestingly, when employing the cooling of D. Stamatellos et al. (2007) to compare with T. J. Haworth et al. (2020), we found this disc began to develop a spiral structure after 4 ORPs, which suggests T. J. Haworth et al. (2020) may have observed the spiral form if the simulations had run for longer than a minimum of 3 ORPs. Consequently, we evolve the discs longer to at least 10 ORPs to verify that the structure is steady. These differences could also be due to the initial temperature profiles, which were set in the optically thin limit in the simulations of T. J. Haworth et al. (2020) and J. Cadman et al. (2020a). For reference, the results of T. J. Haworth et al. (2020) and our simulations with the Stamatellos method are compared in Appendix C.

A summary of the outcomes of the simulations with the modified Lombardi and Stamatellos methods is presented in Fig. 4. We have divided the discs that formed spiral structures into two categories. Discs with ‘faint spirals’ feature low-contrast spiral arms, often with $m > 2$ and more prominent in the inner disc regions. Discs with ‘extended spirals’ have clear spiral arms throughout the entire disc. An example of a faint spiral can be seen in the first panel of Fig.5, and extended spirals can be seen in the third and fourth panels of Fig. 5.

In general, we find that with the modified Lombardi method discs remain stable for a higher $M_{\rm d}/M_*$ than that found in prior work. As shown previously, lower mass stars can support discs of higher $M_{\rm d}/M_*$ than higher mass stars without the gravitational instability developing. However, the parameter space in which discs may fragment differs from earlier findings, and we describe this next.

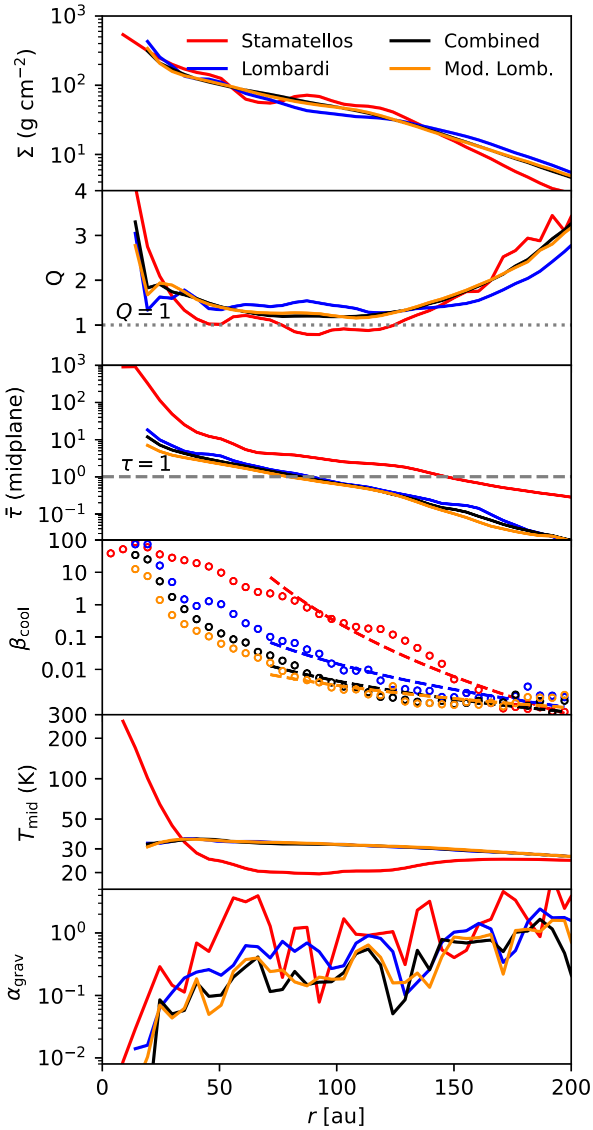

Figure 3. Azimuthally averaged values of the surface density Σ, Toomre Q parameter, mid-plane pseudo-mean optical depth ($\bar{\tau}$), and $\beta_{\rm cool}$, mid-plane temperature $T_{\rm mid}$, and gravitational stress $\alpha_{\rm grav}$ for the four methods for the simulations with 1 $\mathrm{M}_{\odot}$ star, $R_{\rm out} = 200$ au, and $M_{\rm d}/M_* = 0.6$ (bottom row of Fig. 2). The graphs were constructed by averaging simulation outputs taken before clumps formed in the Stamatellos simulation and after 5 ORPs for the others. The azimuthal averages of τ, $\beta_{\rm cool}$, and $T_{\rm mid}$ were calculated using only particles with $|z| < H/2$. For the $\beta_{\rm cool}$ panel, the dashed lines show the fitted functions of the form of equation (23) to $\beta_{\rm cool}$. The fits give $\beta_{\rm cool} \propto r^{-x}$, with $x \approx 9, 5, 3, 2$, respectively, for the Stamatellos, Lombardi, Combined, and modified Lombardi methods.

Table 2. Model parameters.

Parameter	Value(s)
$M_*(\mathrm{M}_{\odot})$	0.1, 0.2, 0.3, 0.5, 0.7, 1.0
$M_{\rm d}/M_*$	{0.1:1.4}
$r_{\rm out}$ (au)	50, 100, 200
p	1
q	1/4
$T_{\rm floor}$ (K)	5

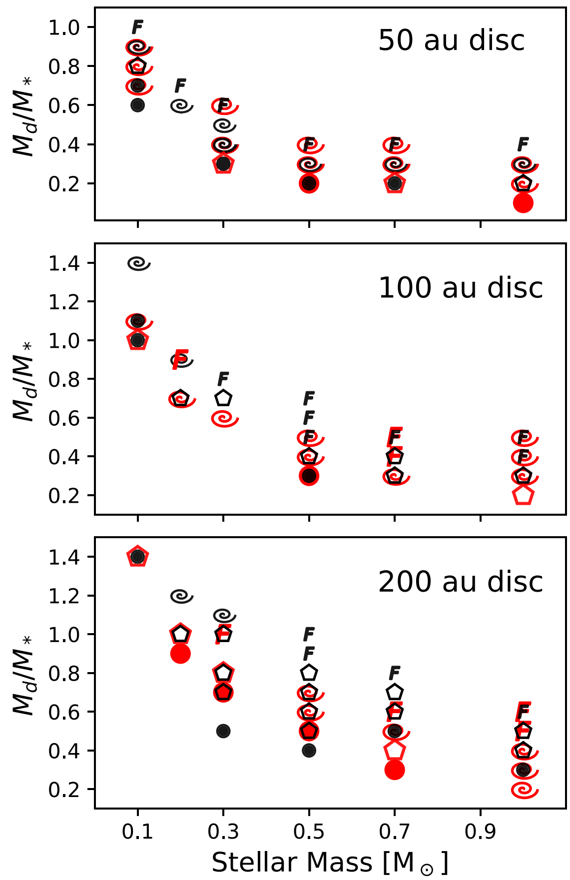

Figure 4. The outcomes of the grid of simulations, illustrating whether the disc remained axisymmetric (circles), developed a faint spiral (pentagons), developed an extended spiral (icons), or fragmented (F). The discs that fragmented developed strong spiral structures as well. The results with the modified Lombardi method are shown in black, and the results with the Stamatellos method are plotted in red.

3.2.1 *Fragmentation*

With the new modified Lombardi method, we observe fragmentation in relatively compact 50 au discs, even for a host stellar mass of 0.1 $\mathrm{M}_{\odot}$, in contrast to the Stamatellos method, in which 50 au discs were stable against fragmentation for the same parameters. We find that 50 au discs may be susceptible to fragmentation for $M_{\rm d}/M_* \gtrsim 0.4$. The more extended (and therefore less optically thick) discs also fragmented but only if $M_{\rm d}/M_* \gtrsim 0.5$. Fragments were produced between 20 and 30 au in the smallest discs, which in our Solar system is in the region of the orbits of Uranus and Neptune. If we consider that protoplanets will undergo inward migration, this allows the possibility of forming a wider variety of planets via direct fragmentation rather than just those on very wide (> 100 au) orbits.

Some simulations like $M_* = 0.5$ $\mathrm{M}_{\odot}$ $R = 50$ au $M_{\rm d}/M_* = 0.4$ form just one collapsing clump but others quickly form ten or more. GI-driven fragmentation can be expected to produce a range of outcomes: few or multiple collapsing clumps with semi-major axes of between $\sim$ 20 and $\sim$ 100 au.

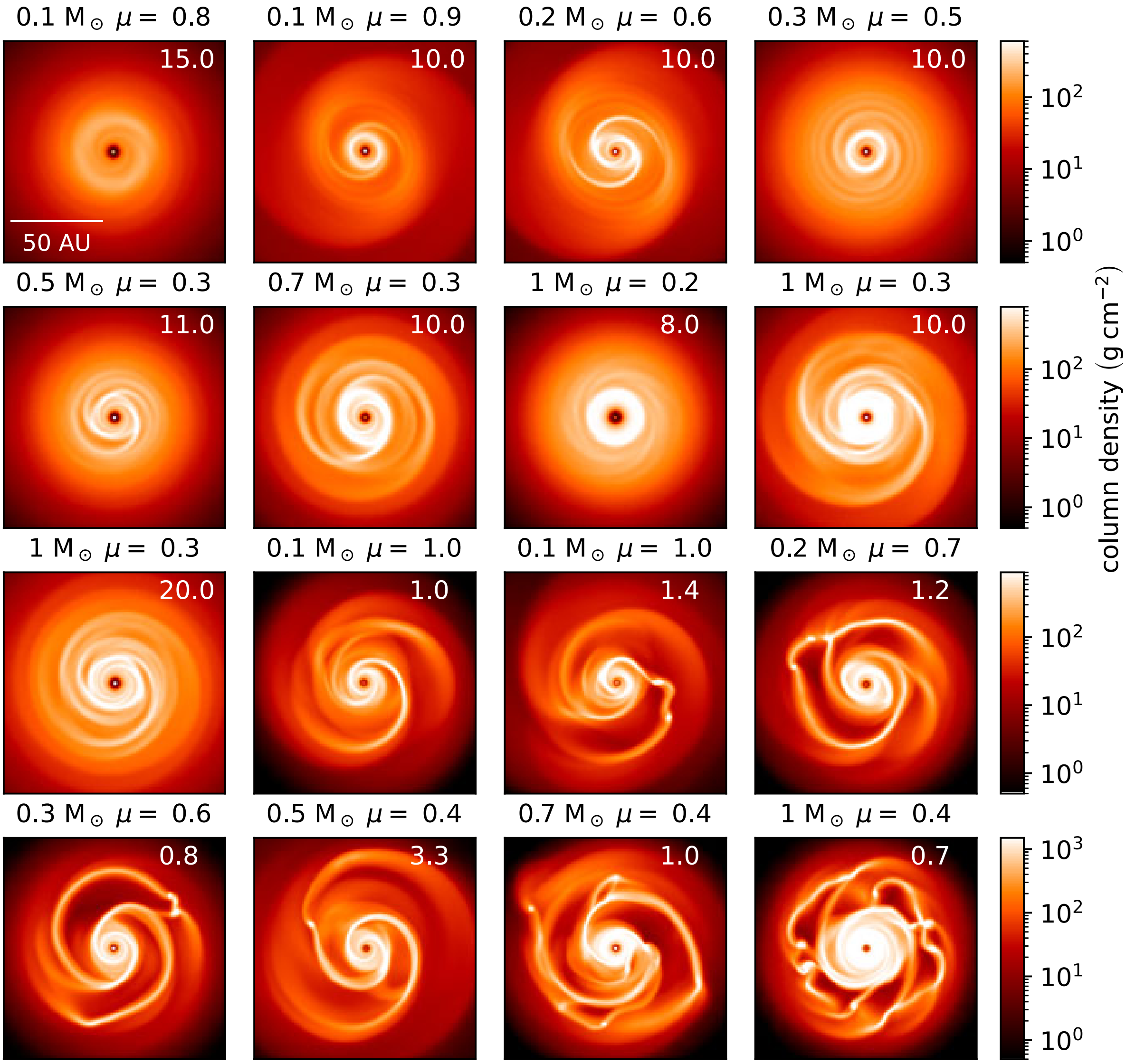

Figure 5. The morphology of spiral structures in discs with an initial radius of 50 au, from simulations employing the modified Lombardi method. The snapshots displayed have been selected to show the range of structures that develop in the discs; the timestamp is given in the upper right of each panel in units of ORP. Stellar masses are given above each panel and $\mu = M_{\rm d}/M_*$. The structures in the 0.1 M$_\odot$ $\mu = 0.8$ and 1 M$_\odot$ $\mu = 0.2$ simulations are classed as faint spirals.

3.2.2 Stability

These simulations indicate that extended $R_{\rm out} = 200$ au discs may be stable at higher masses than previously thought. The larger ($R_{\rm out} = 100$ and 200 au) discs are strongly stabilized by stellar heating such that stars of < 0.5 M$_\odot$ can support discs of the same mass as the star without significant development of the gravitational instability. T. J. Haworth et al. (2020) reported that a 0.4 M$_\odot$ disc around a 1 M$_\odot$ star will fragment, whereas we find that a disc as massive as 0.5 M$_\odot$ remains stable. For $M_* < 0.5$ M$_\odot$, 200 au discs do not fragment even if $M_{\rm d}/M_* > 1$. This indicates that low-mass stars can support relatively massive extended discs. $M_{\rm d}/M_* \sim 0.1$ is typically quoted as indicating the conditions when the gravitational instability is likely to develop. However, these simulations indicate that spiral structures do not develop unless $M_{\rm d}/M_*$ is at least 0.4, and in most cases, much higher still.

3.2.3 Spirals

The variety of structures formed in the simulation can be seen in Figs 5, 6, and 7. In this paper, we classify the spiral structures as faint or extended. Nearly all the simulation snapshots in Fig. 5 show extended spirals and the first two panels of Fig. 6 are examples of faint spirals. Extended spiral structures form in a far narrower range of discs than found by T. J. Haworth et al. (2020) or J. Cadman et al. (2020a) because higher mass discs remain axisymmetric and fragmentation occurred rapidly in many cases, leaving few discs with sustained spiral structures. Faint spirals developed in some of the 100 and 200 au discs but the large-scale spirals typically associated with GI only formed for the lowest mass stars of $M_* \lesssim 0.5$ M$_\odot$.

Long-lived spiral structures (at least 10 ORPs) form in discs of all three initial radii tested but are uncommon within the parameter space as a whole. The time evolution of the minimum

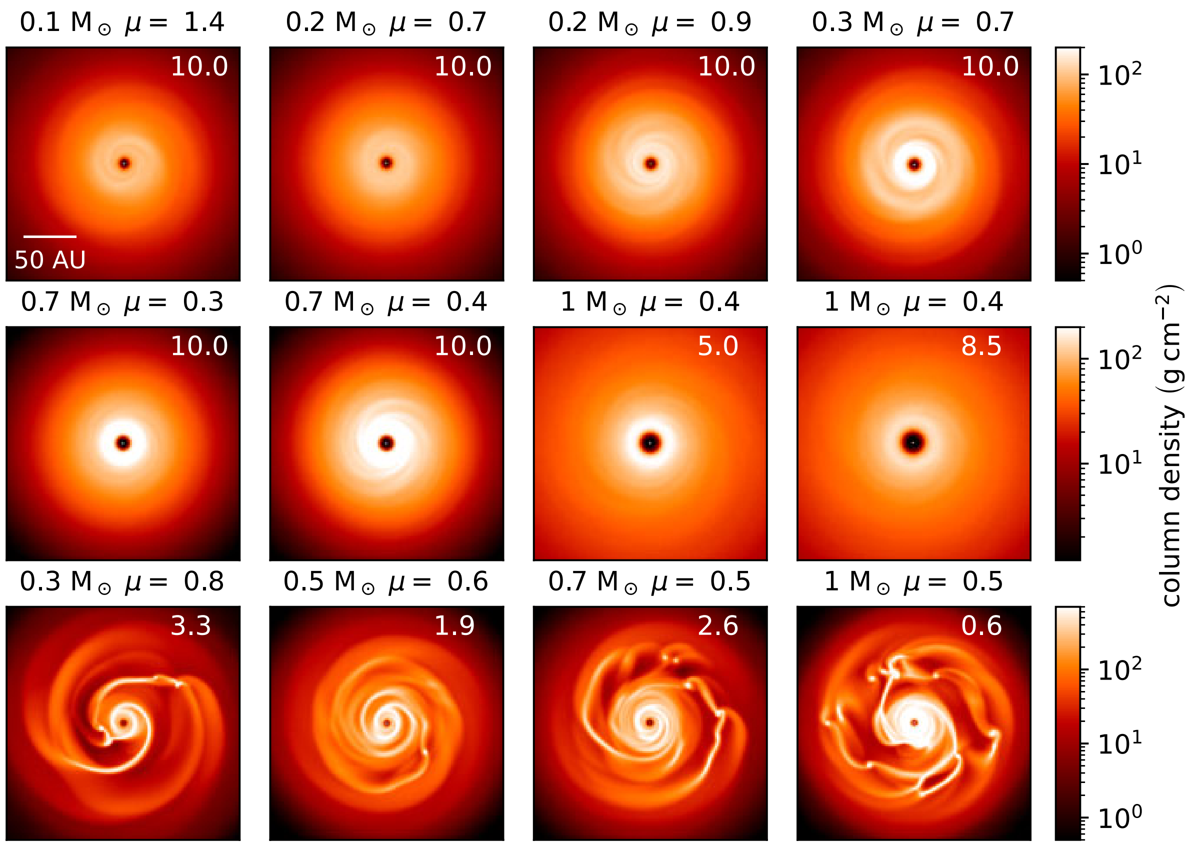

Figure 6. The morphology of spiral structures in discs with an initial radius of 100 au. The timestamp is given in the upper right of each panel in units of ORP and $\mu = M_{\rm d}/M_*$.

values of Q and the maximum values of $\alpha_{\rm grav}$ are plotted in Fig. 8 for simulations with sustained spirals that were modelled for at least 10 ORPs. The mean gravitational stress is not constant but fluctuates on a dynamical time-scale. This is evidence of self-regulation at work: Gravitational stresses heat the disc, increasing Q, weakening the instability, and reducing the stresses. The disc then cools back towards $Q = 1$, the instability strengthens, and the stresses increase (C. J. Clarke 2009). There may also be some non-local effects if these are high disc-to-star mass ratios, as discussed in G. Lodato & W. K. M. Rice (2005) and D. Forgan et al. (2009). Some discs simulated had $Q_{\rm min} \sim 1$ for many ORPs but did not show the zig-zag stress pattern. Those discs are probably not self-regulating but are cooling towards either fragmentation or self-regulation.

Examples of accretion rates over 10 ORPs are plotted in Fig. 9. The models were selected to illustrate the range of accretion rates of axisymmetric discs, discs with strong spirals and discs with faint spirals. Discs with faint and strong spirals had higher accretion rates than axisymmetric discs, as expected, because of the additional effective viscosity due to the gravitational stresses. Discs with faint spirals appear to have accretion rates within a similar range to those with strong spirals, though we note that the disc masses differ between simulations. While accretion rates tend to be overestimated in SPH simulations, we do expect quasi-steady self-gravitating discs to have high accretion rates, probably of at least 10^{-7} M$_\odot$ yr^{-1}.

3.2.4 Linking disc evolution to internal structure

We now compare the physical properties of discs that develop different morphologies. The density structures of the four simulated discs that we will examine are shown in Fig. 10 and azimuthally averaged physical properties are plotted in Fig. 11.

For the axisymmetric disc, $Q > 2$ and so there are no signs of GI-driven structure. The mid-plane optical depth is lowest in the disc with the faint spiral structure. Consequently, stellar heating is significant at the mid-plane for this disc. The disc is heated by the star to above the equilibrium temperature needed to support it against collapse and the disc has a near-isothermal temperature profile.

For discs with a higher column density, the mid-plane temperature is more sensitive to internal heating since the stellar heating is reduced there. In the discs with extended spiral arms, the disc mid-plane is optically thick within $R < 40$ au. Radiative cooling and heating are efficient outside these regions, and the equivalent $\beta_{\rm cool} \ll 1$, where the disc becomes optically thin (Fig. 11).

Like in Section 3.1, we estimate the fit $\beta_{\rm cool} \propto R^{-x}$ between 20 and 50 au to sample the outer disc, and these fits are shown as dashed lines in the fourth panel of Fig. 11. The exponents obtained are approximately 10, 2, 11, and 9 for the axisymmetric, faint spiral, spiral, and fragmenting simulations, respectively. Estimates of $\beta_{\rm cool}$ in the disc mid-plane are also presented, in Fig. 12, to show the spatial variation. $\beta_{\rm cool}$ is high enough towards the edge of the disc with the faint spiral to prevent large-scale spiral arms forming ($\beta_{\rm cool} \gtrsim 20$).

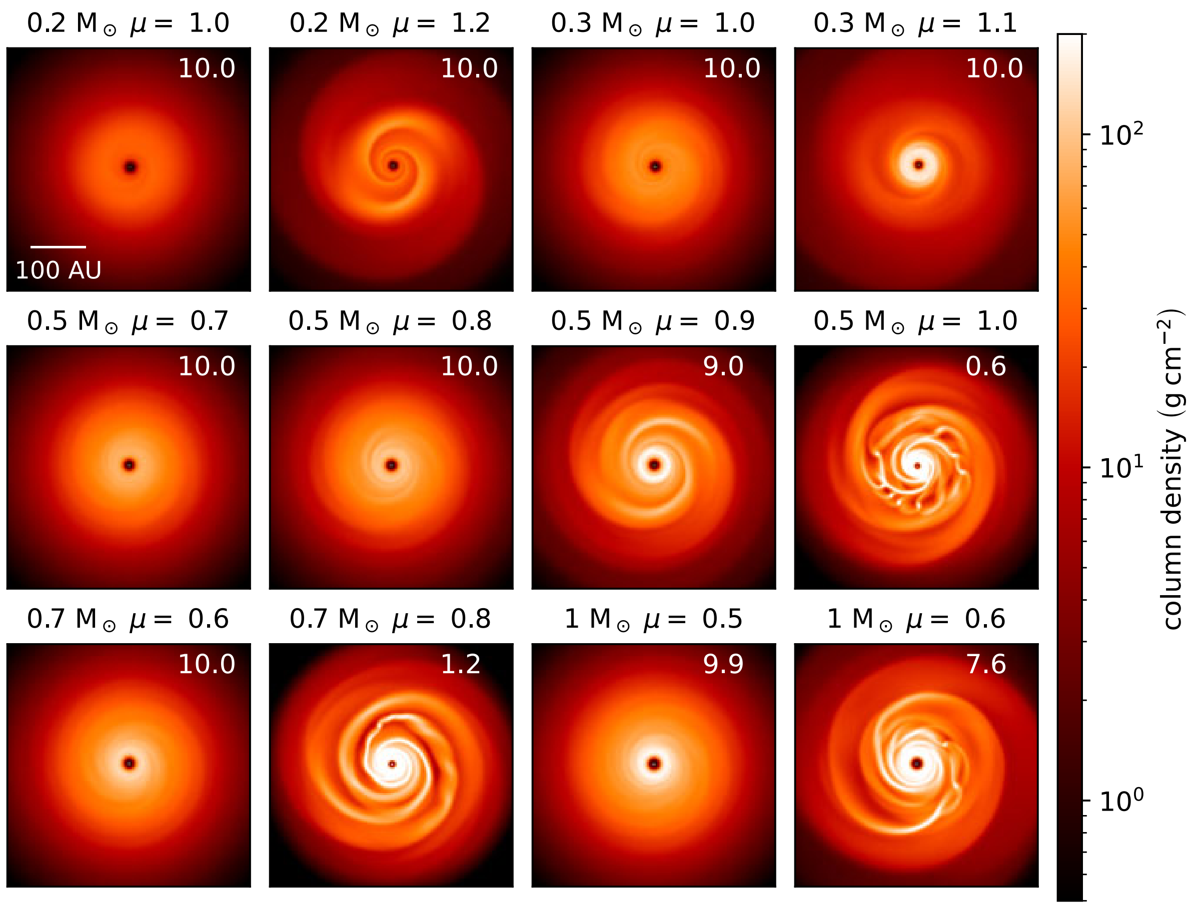

Figure 7. The morphology of spiral structures in discs with an initial outer radius of 200 au. The timestamp is given in the upper right of each panel in units of ORP and $\mu = M_{\rm d}/M_*$.

4 DISCUSSION

We began by repeating simulations of protoplanetary discs with a variety of initial conditions with four different approximate radiative cooling methods. The evolution of self-gravitating discs is very sensitive to the radiative cooling (and heating) rate, so it is important to estimate the optical depth accurately. This became evident in the outcomes of the simulations, in which discs remained axisymmetric when evolved with the more accurate methods compared to developing spiral structures in the original (Stamatellos) method because the latter method overestimates the optical depth. Not only are spiral structures less common with the more accurate methods, but the nature of those spirals appears to be different. Most of the spirals shown in the results of T. J. Haworth et al. (2020), who also employed the Stamatellos method, are also large-scale, low-m spirals but in our wider exploration of the parameter space we find that these structures are uncommon (see Figs 5, 6, and 7). This points to a broader conclusion that the more accurate cooling methods tend not to produce grand design-like spirals, consistent with S. Rowther et al. (2024b) and W. Xu et al. (2025).

4.1 Discs around low-mass stars can fragment

It has been clear for some time that stellar irradiation impacts the evolution of massive protostellar discs but we have now studied a broad parameter space with an accurate approximate method to gain a deeper understanding of the implications. Our key finding is that stellar irradiation does not prevent discs from undergoing fragmentation or from developing self-regulating spiral structures, in agreement with several prior studies (W. K. M. Rice et al. 2011; A. Mercer et al. 2018; J. Cadman et al. 2020a; T. J. Haworth et al. 2020; A. Mercer & D. Stamatellos 2020; C. S. Leedham et al. 2025).

Within the parameters studied ($0.1 < M_* < 1.0$ M$_\odot$) fragmentation occurs more readily in discs around higher mass stars, as found by J. Cadman et al. (2020a). Analytical work has indicated that fragmentation is unlikely in discs of $R_{\rm out} = 50$ au. For example, R. R. Rafikov (2005) found that disc conditions are only favourable for fragmentation beyond ~ 100 au; and C. J. Clarke (2009) found that fragmentation could not occur within ~ 70 au (their figs 5 and 6), unless the accretion rate is extremely high, at least 2 orders of magnitude higher than $\dot{m}$ measured in our simulations (Fig. 9). We find that fragmentation is not restricted to $R > 70$ au but can occur even in 50 au discs, with fragmentation occurring between ~ 30 and 40 au (see also D. Stamatellos & A. P. Whitworth 2009b; A. Mercer & D. Stamatellos 2020).

4.2 Morphology

When stellar heating strongly influences the disc temperatures, discs tend to remain axisymmetric or else fragment at higher

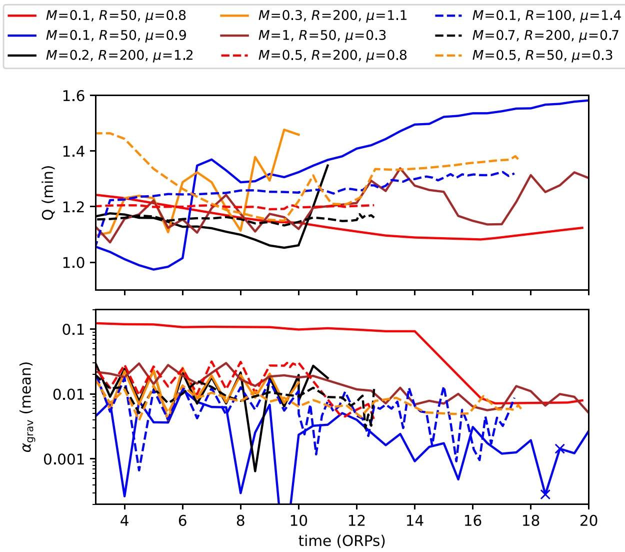

Figure 8. Long-lived spiral structures. The time evolution of the azimuthally averaged minimum value of Q and the average value of $\alpha_{\rm grav}$ within $0.5R_{\rm out} < R < R_{\rm out}$. Solid lines denote discs with large-scale, prominent spirals and dashed lines denote discs with faint spirals. Stellar masses are given in $\rm M_{\odot}$, outer radii in au, and $\mu = M_{\rm d}/M_{*}$. The points marked with '*x*' are negative values of $\alpha_{\rm grav}$, i.e. where the stresses drive outward net gas motions rather than accretion. The zig-zag pattern in $\alpha_{\rm grav}$(mean) is indicative of self-regulation.

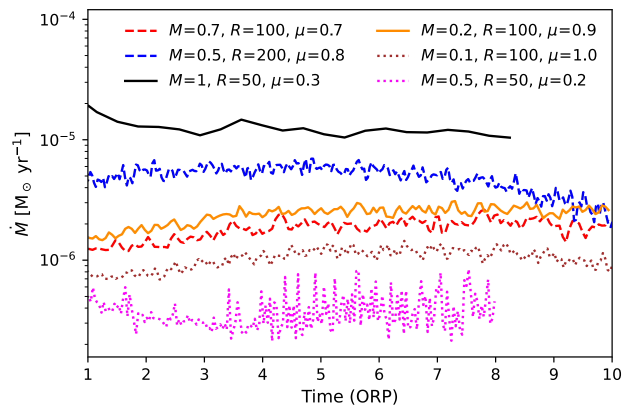

Figure 9. Accretion rates of selected discs smoothed by resampling every 20 yr. The models were chosen to represent the range in accretion rates. Solid lines indicate discs which develop strong spirals, dashed lines indicate faint spirals, and dotted lines indicate discs that remain axisymmetric.

$m_{\rm disc}/m_{\rm star}$. For lower stellar masses with lower luminosity ($m_{*} <$ $0.5\rm M_{\odot}$ here), the disc evolution is less affected by stellar irradiation. In this case, spiral structures may form, and discs remain stable to very high masses. The work of G. Laughlin & M. Rozyczka (1996) predicted that the GI would predominantly drive the $m = 2$ spiral mode; however, we find that low-m grand-design-type spirals appear to be less common. For discs in which stellar irradiation plays a greater role in setting the temperature, we see fainter, higher-order spirals. Elucidation of this effect will require higher resolution simulations of those discs, which we leave for future work.

4.3 Long-lived spirals

There has been some disagreement over whether a self-regulated thermal state exists. K. M. Kratter & R. A. Murray-Clay (2011) showed that discs around 1–3 Myr stars are irradiation dominated and therefore either remain stable or quickly fragment. S. Rowther et al. (2024b) report that the shock heating is negligible compared to the stellar irradiation for a 0.1 $\rm M_{\odot}$ disc around a 1 $\rm M_{\odot}$ star at 3 Myr, meaning gravitoturbulence does not drive disc evolution for those parameters. As a result, studies have recently looked at whether a self-regulated state driven instead by infall is possible (C. Longarini et al. 2025). For the more optically thin discs with a low surface density and $L_{*} \gtrsim 1\ \rm L_{\odot}$, we agree that the stellar irradiation is indeed dominant. However, we do find that more massive discs around young protostars can develop a self-regulated state.

The longest we were able to follow the disc evolution for was 20 orbits ($\sim$ 7000 yr at 50 au), and several discs show sustained GI-driven spiral structure over this time (see Fig. 8). Furthermore, we see evidence of the transient spiral structure as predicted for a regime of marginal stability (C. J. Clarke 2009), for example in the simulation with $M_{*} = 1\ \rm M_{\odot}$, $R = 50$ au, $\mu = 0.3$ and see Fig. 8. However, the periodic variation in gravitational stress does not appear to cause distinctive variations in the accretion rate (Fig. 9; G. Lodato & W. K. M. Rice 2005).

4.4 Implications for modelling

We have shown that the new methods (combined and modified Lombardi, A. K. Young et al. 2024) produce similar results which differ from the original Stamatellos method. Moving forward, we recommend the modified Lombardi method as a suitable radiative cooling approximation. The β cooling approximation has been extensively employed for simulations of GI, so now we examine how well imposed values of $\beta_{\rm cool}$ reflect the physical state of these discs.

The simplest simulations assume that $\beta_{\rm cool}$ is constant with radius but A. Mercer et al. (2018) demonstrated that constant $\beta_{\rm cool}$ with radius is generally not realistic. They showed that $\beta_{\rm cool}$ generally decreases with radius, but in discs with spiral arms $\beta_{\rm cool} \sim$ constant in outer discs. Our results support this finding for discs with moderate surface density (for example, the blue line in Fig. 11), though the slope is still quite steep ($\beta_{\rm cool} \propto 2$) and these discs tend to show faint, flocculent spirals. As expected, the β-cooling approximation with the exponential form is only representative of a subset of disc parameters, and we do need to treat the thermodynamics to be self-consistent.

In the last couple of years, interest has grown in simulating self-gravitating discs with radiative transfer. Numerical radiative transfer methods effectively model the transport of heat within the disc but there is still much uncertainty surrounding the best approach to incorporating cooling to the external environment and stellar irradiation because these quantities require imposed boundary conditions. Very recently Y. Ni, H. Deng & X.-N. Bai (2025) and W. Xu et al. (2025) studied the development of GI using radiation hydrodynamics. In the former model, energy loss to the environment derives from the local opacity and exposed cell surfaces, and in the latter an analytical cooling rate is applied at the disc surface to couple the disc thermodynamics to the environment. In radiation hydrodynamics simulations such as these (and also those implementing a polytropic cooling approximation), the cooling is naturally self-consistent with the development of

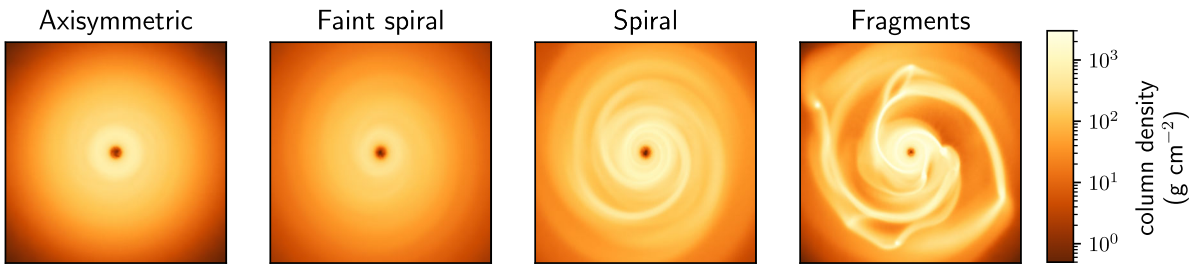

Figure 10. Examples of simulations displaying different morphology, for which various quantities are plotted in Fig. 12. These simulations all had initial outer radii of 50 au, and the width of each panel is 120 au. The simulations are from the left for: $M_* = 0.7\ \mathrm{M}_\odot$, $M_\mathrm{d}/M_* = 0.3$; $M_* = 0.1\ \mathrm{M}_\odot$, $M_\mathrm{d}/M_* = 0.8$; $M_* = 0.7\ \mathrm{M}_\odot$, $M_\mathrm{d}/M_* = 0.3$; and $M_* = 0.7\ \mathrm{M}_\odot$, $M_\mathrm{d}/M_* = 0.4$.

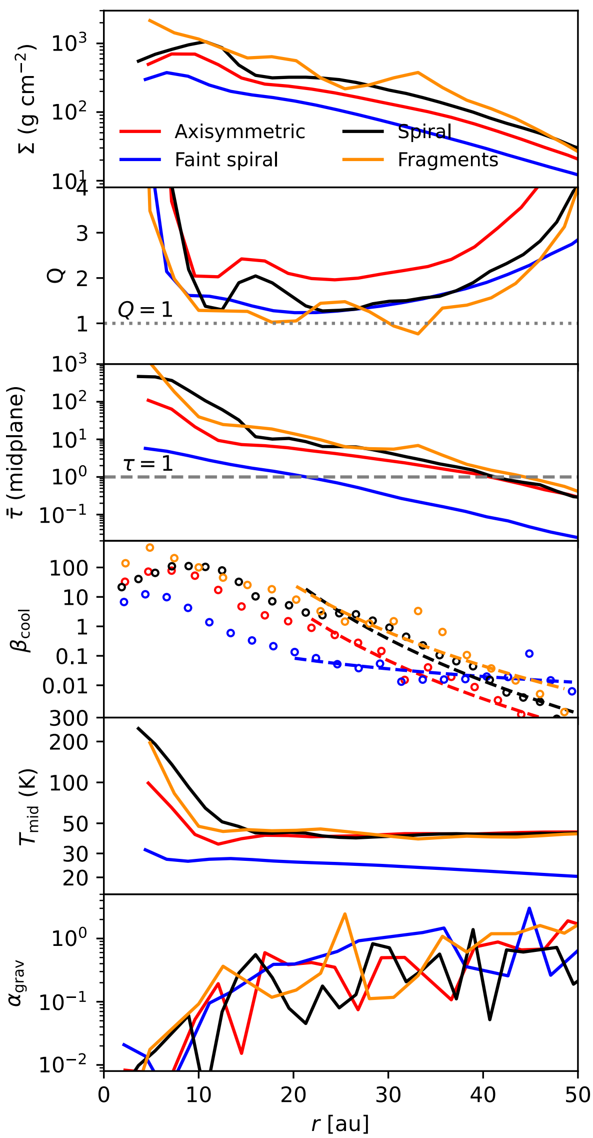

Figure 11. Azimuthally averaged values of the surface density Σ, Toomre Q parameter, mid-plane pseudo-mean optical depth ($\bar{\tau}$), β_cool, and mid-plane temperature T_mid for the $R = 50$ au discs shown in Fig. 10. The red lines (axisymmetric) are for $M_* = 0.7\ \mathrm{M}_\odot$, $M_\mathrm{d}/M_* = 0.2$; the blue lines (faint spiral) are $M_* = 0.1\ \mathrm{M}_\odot$, $M_\mathrm{d}/M_* = 0.8$; the black lines (spiral) show $M_* = 0.7\ \mathrm{M}_\odot$, $M_\mathrm{d}/M_* = 0.3$; and the yellow lines (fragments) show $M_* = 0.7\ \mathrm{M}_\odot$,$M_\mathrm{d}/M_* = 0.4$. The values were extracted for time-steps when the structure has developed.

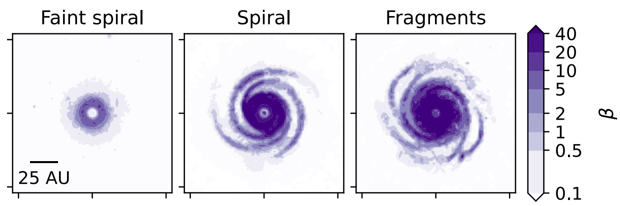

Figure 12. Estimated values of β_cool at the mid-plane for three of the simulated discs from Figs 10 and 11. The left and centre panels show snapshots after 5 ORPs and the right-hand panel shows a snapshot shortly before fragmentation occurs. β_cool is sufficiently low in the dense spiral arms for them to collapse.

GI. With such simulations, gravitoturbulent discs are generally not dominated by a single low-m spiral, and fragments seem to form with lower masses than in simulations with fixed cooling rates. A key limitation of those studies is that they do not include stellar heating. However, like us, W. Xu et al. (2025) also found that grand-design spirals may be uncommon in gravitoturbulent discs, which perhaps suggests that this is linked to a more realistic treatment of thermodynamics rather than the inclusion of stellar heating.

With the modified Lombardi radiative cooling approximation, we can simulate discs using realistic physics with much more modest computational resources than would be required for full radiative transfer. A parameter study like the one presented here would be unfeasible with full radiative transfer, which illustrates a clear benefit to the new method. The reduced computational cost also allows simulations to be run for longer, which is important since disc structures can take many orbits to develop. Based on our results, we recommend evolving simulations for at least 10 dynamical time-scales, by which time all the simulations presented here had reached their longer term state.

While the approximate radiative transfer methods compared here perform well for self-gravitating discs, there are scenarios for which full radiative transfer would be required. The approach used here is not suitable for modelling structures that cause shadowing (like all codes that employ FLD). Similarly, for complex geometries such as misaligned disc components, only ray-tracing radiative transfer would accurately capture the thermodynamics. At present, the method cannot treat local variations in dust density since it relies upon a column-averaged density and opacity. However, we note that recent published simulations that employ ray-tracing radiative transfer also use a constant dust-to-gas ratio (e.g. S. Rowther et al. 2024b; W. Xu et al. 2025).

4.5 Predictions for observations

The simulations conducted in this study indicate that discs need to be at least $0.3M_*$ for the gravitational instability to develop. Therefore, we would expect the GI to be active only very early in the lifetime of the disc unless there has been a substantial late delivery of material to the disc (C. Hall et al. 2019). The work also indicates that discs around lower mass stars may remain stable against fragmentation to $M_{\rm disc} \gtrsim M_*$. Star formation simulations predict that large (>80 au to several 100 au) discs can form (e.g. M. R. Bate 2018; J. Wurster, M. R. Bate & D. J. Price 2019; C.-C. He & M. Ricotti 2025) and that $M_{\rm disc} \gtrsim M_*$ (M. R. Bate 2011, 2018) soon after the protostar forms. However, as pointed out by A. P. Whitworth & D. Stamatellos (2006), if we do not observe massive discs it does not mean they do not form; they might just fragment quickly and have short lifetimes. Observationally, the sample of very young discs in M. J. Maureira et al. (2026) contains several discs with $M_{\rm disc} > 0.1M_*$ and in which $Q < 1$ for some radial range. This demonstrates the existence of young discs with significant masses, and more observations of similar discs may be forthcoming given the growing understanding of the problems of underestimating disc masses from observations, especially at the early, embedded stages.

Long-lived, extended, high-contrast spiral structures appear to be a less common effect of GI, and so we do not expect to find many in observations. Low-contrast, higher-m spiral structures are more common but these may be resolved out or hidden due to high optical depths (M. J. Maureira et al. 2026). It is possible that dust trapping in spiral arms might enhance the observability of these structures (G. Dipierro et al. 2015; J. Cadman et al. 2020b), and this will be an interesting avenue for future research.

5 CONCLUSION

In this paper, we have compared the evolution of massive discs around young protostars in simulations with four different implementations of a radiative cooling approximation. We find that because the earlier Stamatellos and Lombardi methods overestimate the optical depth, the parameter space for which discs develop spiral arms and/or fragment changes compared to J. Cadman et al. (2020a) and T. J. Haworth et al. (2020) for example.

Even for 50 au discs, we find that the outer regions are optically thin and fragmentation is possible. For more extended discs of $R_{\rm out} = 100$ and 200 au, the optical depth is sufficiently low for stellar irradiation to stabilize discs against fragmentation up to masses of $\gtrsim 0.4M_*$. The large-scale spiral arms thought to be characteristic of gravitational instability form only in discs around stars $\lesssim 0.3$ M$_\odot$, due to the low luminosity, and in discs with a high surface density. Our simulations indicate that discs with active gravitational instability, with $Q_{\rm min} \lesssim 1.5$, may only have a very faint spiral structure that is very unlikely to be detectable. This could explain the dearth of observed discs with these features. Nevertheless, for parameters where spiral structures do form, we find that long-lived spirals can persist in a self-regulating state in weakly irradiated discs.

Because massive discs around the youngest protostars may be stable, there is a large reservoir of planetary building material with the potential for forming massive planets quickly as required by observations (e.g. C. J. Nixon et al. 2018; R. Gratton et al. 2019; T. Currie et al. 2022). Since a self-regulated state may persist for some time, this allows the possibility of the formation of planetary cores by direct collapse of the solid component (S. Rowther et al. 2024a; K. Rice et al. 2025).

Since large disc masses and low stellar luminosities are required for the GI to develop, we expect GI to play a key role in driving the evolution of the youngest discs, of less than 1 Myr old. Now that more observations are targeting Class 0 protostars, we should soon be in a strong position to determine the extent to which GI shapes young discs and the beginning of planet formation.

ACKNOWLEDGEMENTS

We thank the anonymous reviewer for their comments that helped to improve the manuscript. AKY is grateful for a UKRI Stephen Hawking Fellowship, a University of Leeds University Academic Fellowship, and a Warwick Prize Fellowship. AKY and KR are grateful for support from the UK STFC via grant ST/V000594/1. FM and AKY are grateful for funding from the Royal Society. This work used the DiRAC Data Intensive service (DIaL2) at the University of Leicester, managed by the University of Leicester Research Computing Service on behalf of the STFC DiRAC HPC Facility (www.dirac.ac.uk). The DiRAC service at Leicester was funded by BEIS, UKRI, and STFC capital funding and STFC operations grants. DiRAC is part of the UKRI Digital Research Infrastructure. AKY thanks the DiRAC RSEs for advice on speeding up PHANTOM at runtime. This work also made use of the Avon HPC cluster operated by the Scientific Computing Research Technology Platform at the University of Warwick and the Aire HPC cluster at the University of Leeds. This work made use of SPLASH (D. J. Price 2007), NUMPY (C. R. Harris et al. 2020), and MATPLOTLIB (Hunter, J. D. 2007).

DATA AVAILABILITY

PHANTOM is a public code, and the version used in this work can be cloned from https://github.com/alisonkyoung1/phantom. The simulation set-up files and selected simulation output data files for the modified Lombardi simulations are available for download from Zenodo at DOI 10.5281/zenodo.22045439. The full set-up files and data for the remaining simulations are available on reasonable request from the corresponding author.

REFERENCES

Bate M. R., 2011, MNRAS, 417, 2036
Bate M. R., 2018, MNRAS, 475, 5618
Bate M. R., Bonnell I. A., Price N. M., 1995, MNRAS, 277, 362
Bertin G., Romeo A. B., 1988, A&A, 195, 105
Boss A. P., 1998, ApJ, 503, 923
Cadman J., Rice K., Hall C., Haworth T. J., Biller B., 2020a, MNRAS, 492, 5041
Cadman J., Hall C., Rice K., Harries T. J., Klaassen P. D., 2020b, MNRAS, 498, 4256
Cai K., Durisen R. H., Boley A. C., Pickett M. K., Mejía A. C., 2008, ApJ, 673, 1138
Chiang E. I., Goldreich P., 1997, ApJ, 490, 368
Choi J., Dotter A., Conroy C., Cantiello M., Paxton B., Johnson B. D., 2016, ApJ, 823, 102
Clarke C. J., 2009, MNRAS, 396, 1066
Cossins P., Lodato G., Clarke C., 2010, MNRAS, 401, 2587
Cullen L., Dehnen W., 2010, MNRAS, 408, 669
Currie T. et al., 2022, Nat. Astron., 6, 751
Deng H., Mayer L., Helled R., 2021, Nat. Astron., 5, 440

Dipierro G., Pinilla P., Lodato G., Testi L., 2015, MNRAS, 451, 974
Dotter A., 2016, ApJS, 222, 8
Forgan D., Rice K., 2013, MNRAS, 430, 2082
Forgan D., Rice K., Stamatellos D., Whitworth A., 2009, MNRAS, 394, 882
Forgan D. H., Hall C., Meru F., Rice W. K. M., 2018, MNRAS, 474, 5036
Gammie C. F., 2001, ApJ, 553, 174
Goldreich P., Lynden-Bell D., 1965, MNRAS, 130, 97
Gratton R. et al., 2019, A&A, 623, A140
Hall C., Dong R., Rice K., Harries T. J., Najita J., Alexander R., Brittain S., 2019, ApJ, 871, 228
Harris C. R. et al., 2020, Nature, 585, 357
Haworth T. J., Cadman J., Meru F., Hall C., Albertini E., Forgan D., Rice K., Owen J. E., 2020, MNRAS, 494, 4130
He C.-C., Ricotti M., 2025, MNRAS, 540, 175
Hunter J. D., 2007, Comput. Sci. Eng., 9, 90
Jørgensen J. K., van Dishoeck E. F., Visser R., Bourke T. L., Wilner D. J., Lommen D., Hogerheijde M. R., Myers P. C., 2009, A&A, 507, 861
Kratter K., Lodato G., 2016, ARA&A, 54, 271
Kratter K. M., Murray-Clay R. A., 2011, ApJ, 740, 1
Kratter K. M., Murray-Clay R. A., Youdin A. N., 2010, ApJ, 710, 1375
Laughlin G., Bodenheimer P., 1994, ApJ, 436, 335
Laughlin G., Rozyczka M., 1996, ApJ, 456, 279
Lee H., Nayakshin S., Booth R. A., 2025, MNRAS, 544, L18
Leedham C. S., Booth R. A., Clarke C. J., 2025, MNRAS, 539, 2780
Lin M.-K., Kratter K. M., 2016, ApJ, 824, 91
Lodato G., Clarke C. J., 2011, MNRAS, 413, 2735
Lodato G., Rice W. K. M., 2004, MNRAS, 351, 630
Lodato G., Rice W. K. M., 2005, MNRAS, 358, 1489
Lombardi J. C., McInally W. G., Faber J. A., 2015, MNRAS, 447, 25
Longarini C., Lodato G., Bertin G., Armitage P. J., 2023, MNRAS, 519, 2017
Longarini C., Price D. J., Kratter K. M., Lodato G., Clarke C. J., 2025, MNRAS, 541, 1145
Lynden-Bell D., Kalnajs A. J., 1972, MNRAS, 157, 1
Mamatsashvili G. R., Rice W. K. M., 2010, MNRAS, 406, 2050
Matzner C. D., Levin Y., 2005, ApJ, 628, 817
Maureira M. J. et al., 2026, A&A, 705, A96
Mayer L., Quinn T., Wadsley J., Stadel J., 2004, ApJ, 609, 1045
Mercer A., Stamatellos D., 2020, A&A, 633, A116
Mercer A., Stamatellos D., Dunhill A., 2018, MNRAS, 478, 3478
Meru F., 2015, MNRAS, 454, 2529
Meru F., Bate M. R., 2010, MNRAS, 406, 2279
Meru F., Bate M. R., 2012, MNRAS, 427, 2022
Nelson A. F., 2006, MNRAS, 373, 1039
Ni Y., Deng H., Bai X.-N., 2025, ApJ, 995, 96
Nixon C. J., King A. R., Pringle J. E., 2018, MNRAS, 477, 3273
Paczynski B., 1978, AcA, 28, 91
Price D. J., 2007, PASA, 24, 159
Price D. J. et al., 2018, Publ. Astron. Soc. Aust., 35, e031
Rafikov R. R., 2005, ApJ, 621, L69
Rafikov R. R., 2009, ApJ, 704, 281
Rice K., 2016, PASA, 33, e012
Rice K., Lopez E., Forgan D., Biller B., 2015, MNRAS, 454, 1940
Rice K., Baehr H., Young A. K., Booth R., Rowther S., Meru F., Hall C., Koval A., 2025, MNRAS, 539, 3421
Rice W. K. M., Lodato G., Pringle J. E., Armitage P. J., Bonnell I. A., 2006, MNRAS, 372, L9
Rice W. K. M., Armitage P. J., Mamatsashvili G. R., Lodato G., Clarke C. J., 2011, MNRAS, 418, 1356
Rowther S., Meru F., 2020, MNRAS, 496, 1598
Rowther S., Nealon R., Meru F., 2023, MNRAS, 518, 763
Rowther S., Nealon R., Meru F., Wurster J., Aly H., Alexander R., Rice K., Booth R. A., 2024a, MNRAS, 528, 2490
Rowther S., Price D. J., Pinte C., Nealon R., Meru F., Alexander R., 2024b, MNRAS, 534, 2277
Safronov V. S., 1960, Ann. Astrophys., 23, 979
Schib O., Mordasini C., Emsenhuber A., Helled R., 2025, A&A, 704, A28
Segura-Cox D. M. et al., 2020, Nature, 586, 228
Shakura N. I., Sunyaev R. A., 1973, A&A, 24, 337
Stamatellos D., Whitworth A. P., 2009a, MNRAS, 400, 1563
Stamatellos D., Whitworth A. P., 2009b, MNRAS, 392, 413
Stamatellos D., Whitworth A. P., Bisbas T., Goodwin S., 2007, A&A, 475, 37
Takakuwa S. et al., 2024, ApJ, 964, 24
Toomre A., 1964, ApJ, 139, 1217
Vandervoort P. O., 1970, ApJ, 161, 87
van't Hoff M. L. R., Tobin J. J., Harsono D., van Dishoeck E. F., 2018, A&A, 615, A83
van't Hoff M. L. R. et al., 2020, ApJ, 901, 166
Whitworth A. P., Stamatellos D., 2006, A&A, 458, 817
Williams J. P., Cieza L. A., 2011, ARA&A, 49, 67
Wurster J., Bate M. R., Price D. J., 2019, MNRAS, 489, 1719
Xu W., Jiang Y.-F., Kunz M. W., Stone J. M., 2025, ApJ, 986, 92
Young A. K., Celeste M., Booth R. A., Rice K., Koval A., Carter E., Stamatellos D., 2024, MNRAS, 531, 1746
Zamponi J., Maureira M. J., Zhao B., Liu H. B., Ilee J. D., Forgan D., Caselli P., 2021, MNRAS, 508, 2583

APPENDIX A: SETTING INITIAL CONDITIONS

We need to be careful that the choice of initial conditions does not have undue influence on the outcomes of the simulations. To verify that the results presented are robust, we repeat a sample of the simulations after taking steps to ensure the initial discs are stable and settled. The simulations chosen are given in Table A1.

The aim of altering the initial conditions was to obtain an initial disc that was as stable as possible, i.e. to minimise initial spreading inwards and outwards due to artificially abrupt pressure gradients and to reduce any initial transient overdensities. The surface density power-law exponent was reduced to $p = 0.75$ and density was tapered more gently in the outer disc from $r_c = 0.8R_{\rm out}$. The disc was initialized with 5.2×10^5 particles to allow for increased initial accretion. After 2 ORPs, there were ≈ 4 per cent more particles in the discs than at the same point in the original simulations. The discs were then allowed to evolve for 2 ORPs with fixed $\beta_{\rm cool} = 20$, so that any initial perturbations were not amplified by changes in optical depth. From there, the simulations were evolved as before with the modified Lombardi cooling.

The outcome of the new simulations is the same as the original simulations. Fig. A1 shows azimuthally averaged properties of the disc after 4 ORPs in the original simulation compared to simulation A. The density peak is much shallower at the inner edge and beyond 50 au; the sound speed and Toomre Q are similar. Simulations S and F are shown in Fig. A2 with the original simulations in the left-hand panels. We can see that the structures are largely similar, with simulation S displaying a stable spiral structure as before, and simulation F producing collapsing fragments. We can therefore conclude that the results presented in this paper are not driven by the initial conditions.

Table A1. Details of the simulations that were repeated with slight changes to initial conditions and initially evolved for 2 ORPs with β-cooling.

ID	Outcome	M_* ($M_\odot$)	$R_{\rm out}$ (au)	$M_{\rm d}/M_*$
A	Axisymmetric	1.0	200	0.4
S	Spiral	0.5	50	0.3
F	Fragments	0.7	100	0.5

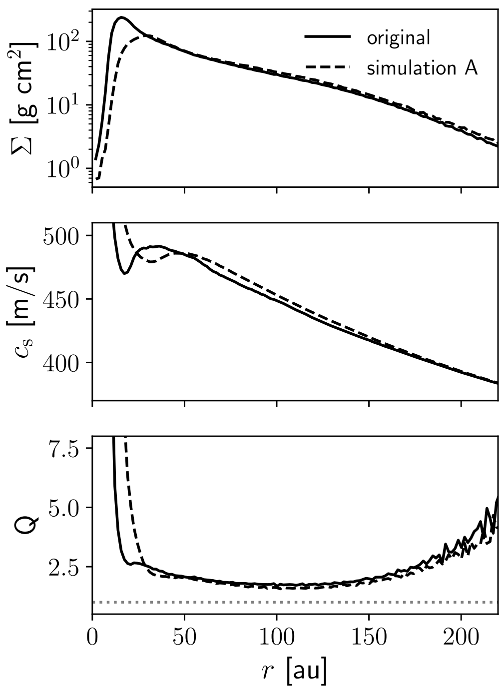

Figure A1. Azimuthally averaged surface density, sound speed and Q profiles for snapshots taken after 4 ORPs, for simulation A (axisymmetric) and the corresponding original simulation from the main paper.

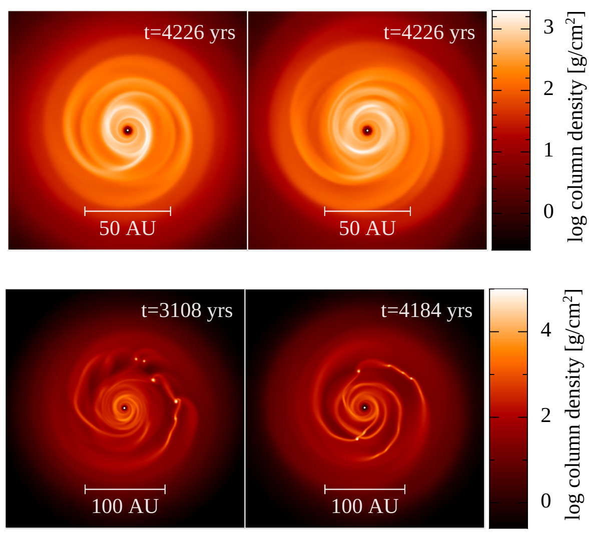

Figure A2. Top right: new simulation S; Bottom right: new simulation F. The counterpart snapshots from the original simulations are in the left-hand panels. The structures are broadly very similar in the new simulations.

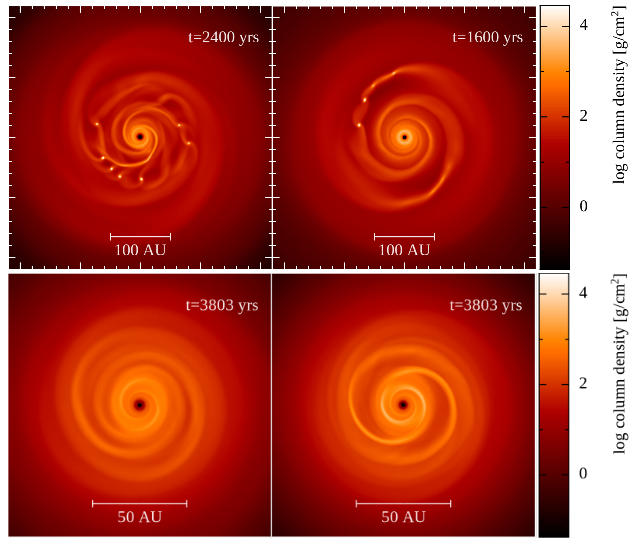

Figure B1. Comparison of the onset of fragmentation in a simulation using 5×10^5 particles (left) and 10^6 particles (right). Upper panels: model had $M_* = 0.5$ M$_\odot$, $R_{\rm out} = 200$ au and $M_{\rm d}/M_* = 1$. Lower panels: $M_* = 0.7$ M$_\odot$, $R_{\rm out} = 50$ au and $M_{\rm d}/M_* = 0.3$. Fragmentation occurs slightly sooner with the higher resolution, still within a fraction of an ORP. In the lower panels, the structure is broadly similar; however, it took around twice as long for the $m = 2$ spiral to form in the high-resolution simulation.

APPENDIX B: RESOLUTION

In this work, we chose to model the discs with 5×10^5 particles so that a parameter study was feasible with the available compute resources. This is sufficient to resolve the Toomre mass (A. F. Nelson 2006), which reaches a minimum of ~ 0.002 M$_\odot$, with > 27 SPH particles. As an additional check, we repeat some simulations at a higher resolution of 10^6 SPH particles. Fig. B1 shows that the disc evolution is very similar for a high-mass disc. Fragmentation occurs slightly sooner with the higher resolution, but the development of spiral waves and clumps proceeds in a similar manner.

In Fig. B2, we compare the surface density, Q, vertical resolution, and mid-plane optical depth for two discs modelled with increased resolution. Increasing the resolution has a small effect, causing a reduced optical depth since the mid-plane is better resolved. The vertical resolution is just sufficient with $\langle h \rangle / H \approx 0.25$ for the regions of interest. We also tested the effects of changing the lower limit of $\alpha_{\rm SPH}$ used by the viscosity switch for the $R = 50$ au disc (left panels). The long-term evolution of all four simulations was similar; only the time at which spiral structures developed differed. The optical depths and cooling rates are broadly similar, with differences due mainly to the speed of evolution.

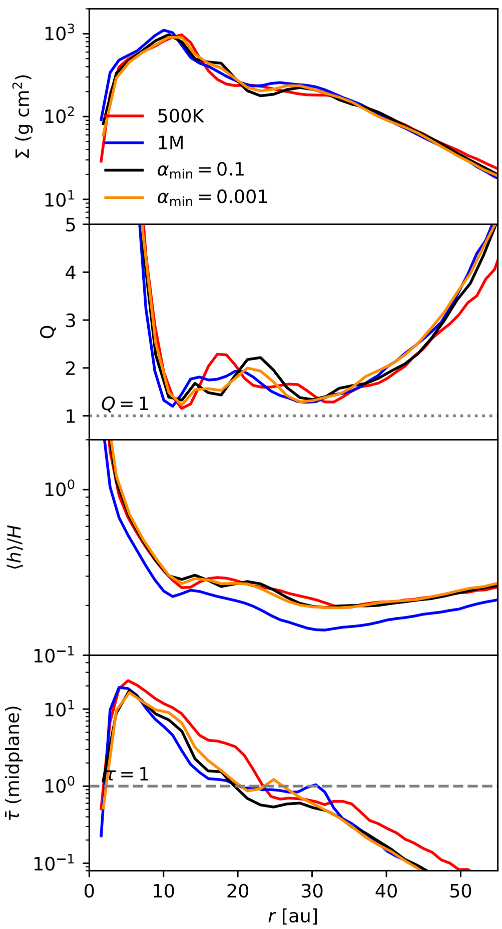

Figure B2. Comparison of simulations with standard resolution (5×10^5 particles) and high resolution. $M_* = 0.7\ \mathrm{M}_\odot$, $R_{\mathrm{out}} = 50$ au and $M_*/M_{\mathrm{d}} = 0.3$ (see also bottom panels of Fig. B1). These snapshots were taken at 10 ORPs. Also shown are the outcomes of simulations with 5×10^5 particles and a raised and reduced minimum value of α_{SPH}.

APPENDIX C: COMPARISON WITH HAWORTH ET AL. (2020)

While the method for calculating the radiative cooling rate is identical, the temperature of the disc due to stellar heating is different. In our simulations, the minimum disc temperature is determined by the stellar irradiation, accounting for some attenuation of the radiation incident on the disc surface using equation (12). T. J. Haworth et al. (2020) did not apply any attenuation, so the disc temperatures are effectively estimated in the optically thin limit, resulting in higher temperatures despite employing identical values of stellar luminosity (Table 1). We therefore now compare the evolution of discs under the Stamatellos method in this work with that of T. J. Haworth et al. (2020) to determine to what extent the difference in minimum temperature estimates affects the outcome. The results are shown in Fig. C1, with our simulations in black and those of T. J. Haworth et al. (2020) in red. We note that the latter work does not distinguish between faint and large-scale spirals.

We find evolution of the discs to be similar. In the top panel of Fig. C1 , it is apparent that the gravitational instability becomes active at slightly lower values of M_{d}/M_* for 50 au discs, driving the formation of spiral arms. This can be explained by the lower disc temperatures in the new simulations due to the attenuation of the stellar irradiation. For the 200 au discs (Fig. C1, lower panel), we find that fragmentation occurs at slightly higher disc masses. Since the mid-plane temperature is lower in our simulations, the gravitational instability is able to self-regulate effectively to a higher M_{d}/M_* by locally increasing the temperature enough to prevent fragmentation.

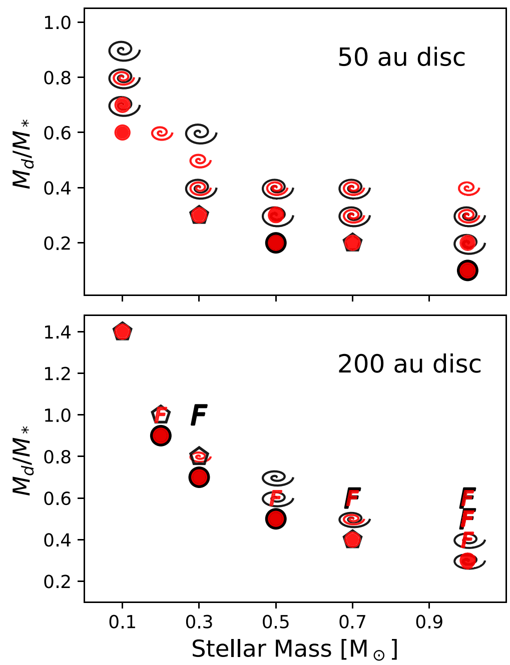

Figure C1. A comparison of the evolution of discs using the MIST stellar luminosity from T. J. Haworth et al. (2020; red icons) with our code using the Stamatellos method and the same luminosities (black icons). 'F' indicates fragmentation, spirals indicate large-scale spiral structures, pentagons indicate faint spirals (only for our simulations), and circles indicate axisymmetric discs.

Both approaches to calculating the minimum disc temperatures are approximations that represent rough upper and lower limits. Comparing with Fig. 4, it's clear that the method of estimating the radiative cooling rate has a more significant effect on the disc evolution than this difference in minimum disc temperature.

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