



## LJMU Research Online

Kruijssen, JMD, Schrubba, A, Hygate, APS, Hu, C-Y, Haydon, DT and Longmore, SN

**An uncertainty principle for star formation - II. A new method for characterising the cloud-scale physics of star formation and feedback across cosmic history**

<http://researchonline.ljmu.ac.uk/id/eprint/8641/>

### Article

**Citation** (please note it is advisable to refer to the publisher's version if you intend to cite from this work)

**Kruijssen, JMD, Schrubba, A, Hygate, APS, Hu, C-Y, Haydon, DT and Longmore, SN (2018) An uncertainty principle for star formation - II. A new method for characterising the cloud-scale physics of star formation and feedback across cosmic history. Monthly Notices of the Royal Astronomical**

LJMU has developed **LJMU Research Online** for users to access the research output of the University more effectively. Copyright © and Moral Rights for the papers on this site are retained by the individual authors and/or other copyright owners. Users may download and/or print one copy of any article(s) in LJMU Research Online to facilitate their private study or for non-commercial research. You may not engage in further distribution of the material or use it for any profit-making activities or any commercial gain.

The version presented here may differ from the published version or from the version of the record. Please see the repository URL above for details on accessing the published version and note that access may require a subscription.

For more information please contact [researchonline@ljmu.ac.uk](mailto:researchonline@ljmu.ac.uk)

<http://researchonline.ljmu.ac.uk/>



# An uncertainty principle for star formation – II. A new method for characterising the cloud-scale physics of star formation and feedback across cosmic history

J. M. Diederik Kruijssen,<sup>1,2,3\*</sup> Andreas Schruba,<sup>4</sup> Alexander P. S. Hygate,<sup>2,1</sup> Chia-Yu Hu,<sup>3,5</sup> Daniel T. Haydon<sup>1</sup> and Steven N. Longmore<sup>6</sup>

<sup>1</sup>*Astronomisches Rechen-Institut, Zentrum für Astronomie der Universität Heidelberg, Mönchhofstraße 12-14, 69120 Heidelberg, Germany*

<sup>2</sup>*Max-Planck Institut für Astronomie, Königstuhl 17, 69117 Heidelberg, Germany*

<sup>3</sup>*Max-Planck Institut für Astrophysik, Karl-Schwarzschild-Straße 1, 85748 Garching, Germany*

<sup>4</sup>*Max-Planck Institut für Extraterrestrische Physik, Giessenbachstraße 1, 85748 Garching, Germany*

<sup>5</sup>*Center for Computational Astrophysics, 160 Fifth Avenue, New York, NY 10010, USA*

<sup>6</sup>*Astrophysics Research Institute, Liverpool John Moores University, IC2, Liverpool Science Park, 146 Brownlow Hill, Liverpool L3 5RF, United Kingdom*

Accepted 2018 April 27. Received 2018 April 23; in original form 2017 October 5.

## ABSTRACT

The cloud-scale physics of star formation and feedback represent the main uncertainty in galaxy formation studies. Progress is hampered by the limited empirical constraints outside the restricted environment of the Local Group. In particular, the poorly-quantified time evolution of the molecular cloud lifecycle, star formation, and feedback obstructs robust predictions on the scales smaller than the disc scale height that are resolved in modern galaxy formation simulations. We present a new statistical method to derive the evolutionary timeline of molecular clouds and star-forming regions. By quantifying the excess or deficit of the gas-to-stellar flux ratio around peaks of gas or star formation tracer emission, we directly measure the relative rarity of these peaks, which allows us to derive their lifetimes. We present a step-by-step, quantitative description of the method and demonstrate its practical application. The method’s accuracy is tested in nearly 300 experiments using simulated galaxy maps, showing that it is capable of constraining the molecular cloud lifetime and feedback time-scale to  $< 0.1$  dex precision. Access to the evolutionary timeline provides a variety of additional physical quantities, such as the cloud-scale star formation efficiency, the feedback outflow velocity, the mass loading factor, and the feedback energy or momentum coupling efficiencies to the ambient medium. We show that the results are robust for a wide variety of gas and star formation tracers, spatial resolutions, galaxy inclinations, and galaxy sizes. Finally, we demonstrate that our method can be applied out to high redshift ( $z \lesssim 4$ ) with a feasible time investment on current large-scale observatories. This is a major shift from previous studies that constrained the physics of star formation and feedback in the immediate vicinity of the Sun.

**Key words:** stars: formation – ISM: evolution – galaxies: evolution – galaxies: formation – galaxies: ISM – galaxies: stellar content

## 1 INTRODUCTION

Current theoretical and numerical models aiming to reproduce the observed galaxy population are strongly limited by uncertainties in the baryonic physics (e.g. Hopkins et al. 2011; Haas et al. 2013; Vogelsberger et al. 2014; Schaye et al. 2015). While the boundary conditions provided by cosmology and the hierarchical growth of the dark matter haloes in which galaxies reside are relatively well-constrained (e.g. Springel et al. 2005b; Planck Collaboration et al. 2014), connecting these haloes to the visible galaxy population hinges critically on the unknown physics of star formation and feedback (McKee & Ostriker 2007; Kennicutt & Evans 2012).

Galaxy formation simulations generally describe star formation using the phenomenological, galaxy-scale relations between the gas mass and the star formation rate (SFR), observed from nearby spiral galaxies (e.g. Bigiel et al. 2008; Schruba et al. 2011; Kennicutt & Evans 2012; Leroy et al. 2013) out to high redshift (e.g. Daddi et al. 2010; Genzel et al. 2010; Tacconi et al. 2013). Broadly, these galactic ‘star formation relations’ represent variations of the form

$$\text{SFR} = \frac{\epsilon_{\text{sf}}}{\tau_{\text{sf}}} M_{\text{gas}}, \quad (1)$$

where  $\epsilon_{\text{sf}}$  represents the star formation efficiency,  $\tau_{\text{sf}}$  is the star formation time-scale, and  $M_{\text{gas}}$  indicates the gas mass. One of the main reasons that the physics underpinning the galactic star formation relation have been so challenging to constrain (see e.g.

\* kruijssen@uni-heidelberg.de

Krumholz & McKee 2005; Hennebelle & Chabrier 2011; Ostriker & Shetty 2011; Padoan & Nordlund 2011; Krumholz et al. 2012, among many others) is that the star formation efficiency and time-scale represent degenerate quantities in the above expression. It is impossible to determine from the star formation relation alone whether star formation is rapid and inefficient or slow and efficient. This problem is exacerbated by the fact that both quantities are notoriously hard to measure – star formation takes place on time-scales much longer than a human lifetime, implying that  $\tau_{\text{sf}}$  cannot be measured directly and  $\epsilon_{\text{sf}}$  cannot be obtained by simply considering the initial and final states of a single system. However, knowing either  $\epsilon_{\text{sf}}$  or  $\tau_{\text{sf}}$  would allow one to immediately constrain the other. Analogously to equation (1), the commonly-used star formation ‘recipes’ in galaxy formation simulations generally assume that gas is turned into stars on some time-scale  $\tau_{\text{sf}}$  with some efficiency  $\epsilon_{\text{sf}}$ . Both quantities are unknown on the small (10–100 pc) scales of the individual gas clouds resolved in modern galaxy formation simulations, but strongly affect the structure of simulated galaxies (e.g. Hopkins et al. 2013; Braun & Schmidt 2015; Semenov et al. 2016). It is therefore a priority to obtain empirical constraints on these quantities and their variation with the galactic environment.

Likewise, the deposition of mass, momentum and energy by feedback is often based on a simple representation of supernova (SN) feedback, i.e. by injecting an amount of energy appropriate for the expected number of SNe into the interstellar medium (ISM) surrounding each young stellar population (e.g. Springel & Hernquist 2003) or by defining an empirical mass outflow rate in units of the SFR that may depend on the local galactic environmental conditions (e.g. Oppenheimer & Davé 2006, 2008). However, it is shown by theory (e.g. Murray et al. 2010; Dale et al. 2013; Agertz et al. 2013) and observations (e.g. Pellegrini et al. 2010; Lopez et al. 2011, 2014) of nearby star-forming regions that other feedback mechanisms (e.g. photoionization, stellar winds, radiation pressure) can be at least equally effective as SNe (see e.g. the recent reviews by Kruijssen 2013; Krumholz et al. 2014; Dale 2015). Obtaining an empirically-motivated prescription for stellar feedback in galaxy formation simulations is crucial, because most discrepancies between the observed and simulated galaxy populations can be alleviated with a suitable (but possibly ad-hoc) choice of feedback model (e.g. Hopkins et al. 2014; Schaye et al. 2015).

In Kruijssen & Longmore (2014, hereafter KL14), we presented a new statistical model, named the ‘uncertainty principle for star formation’, which explains how the galaxy-scale star formation properties of a galaxy emerge by the summation over its constituent population of independent star-forming regions. In particular, the ‘KL14 principle’ predicts that the star formation relation changes from towards small ( $\lesssim$  kpc) spatial scales in a way that depends sensitively on the evolutionary timeline of star-forming regions. We showed that it can therefore be used to directly measure the efficiencies and time-, size- and velocity-scales of cloud collapse, star formation, and feedback without requiring individual clouds to be resolved. By comparing the flux ratios between tracers of gas and star formation on 0.1–1 kpc scales across a galaxy, the method effectively counts the relative occurrence of these phases without resolving them spatially, thereby directly constraining their characteristic lifetimes and spatial separations. The method is easy to apply and could potentially be used to characterize cloud-scale star formation and feedback across cosmic history.

The KL14 principle does not provide the first attempt of characterizing cloud-scale star formation and feedback. For instance, molecular cloud lifetimes have been estimated in the Local Group

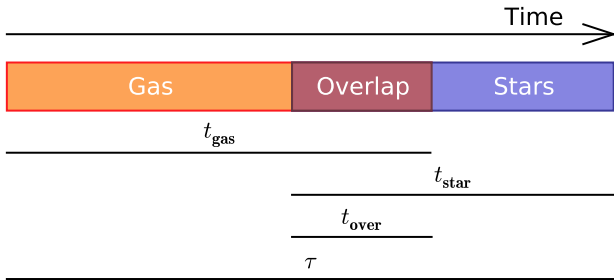
for the Milky Way, the Magellanic Clouds, and M33 (Engargiola et al. 2003; Kawamura et al. 2009; Murray 2011), as well as for M51 (Meidt et al. 2015), stellar feedback energy and momentum deposition rates have been determined for the Magellanic Clouds (Pellegrini et al. 2010; Lopez et al. 2011, 2014), and the rapidity and efficiency of star formation have been quantified for star-forming clouds in the Milky Way, albeit with a broad range of outcomes (e.g. Elmegreen 2000; Krumholz & Tan 2007; Evans et al. 2009; Kruijssen et al. 2015; Barnes et al. 2017). The common shortcoming of the methods used in these studies is that they can only be applied to the limited range of galactic environments present in Local Group galaxies or to exceptionally high-resolution observations of external galaxies, because they require individual clouds to be resolved. The main improvement of the KL14 principle is that it can be applied out to much larger distances because only the typical *separation* length between clouds must be (marginally) resolved. In addition, the robust statistical basis of the method enables its application to large galaxy samples with little computational effort or human intervention.

In this paper, a general framework for applying the KL14 principle is presented and the method is validated using numerical simulations of isolated disc galaxies. We first summarise the basic concepts behind the KL14 principle (Section 2). We then present a step-by-step description of the method’s application to integrated intensity maps of gas and star formation tracers across galaxies (Section 3). We validate the process by applying it to numerical simulations of disc galaxies and verifying the accuracy of the extracted quantities (Section 4), the results of which are summarised as a set of guidelines for the reliable application of the method (Section 4.4). The extracted quantities are then used to calculate several derived quantities that are demonstrated to accurately describe cloud-scale star formation and feedback (Section 5). We then use the several performed tests of the method to determine out to which distances reliable constraints on cloud-scale star formation and feedback can be obtained (Section 6). We include an extensive discussion of the method’s caveats and limitations, as well as of its planned future improvements (Section 7). The paper is finished with a summary of our conclusions and a discussion of the method’s potential for future applications (Section 8). Finally, Appendices A–D provide the necessary background for a number of technical considerations in this paper, as well as the complete set of test results used in Section 4.

The reader less interested in the quantitative details of the method and its validation can refer to parts of the paper for a qualitative summary. For a cursory overview of the results, we recommend consulting only Sections 1, 2, 3.1, 4.4, 6, and 8. This effectively shortens the paper to a little over ten pages.

## 2 SUMMARY OF THE FORMALISM

The fundamental concept underpinning the KL14 principle is that a galaxy consists of some number of independent star-forming regions (hereafter ‘independent regions’) separated by some length-scale  $\lambda$ , which each reside in an evolutionary phase independently of their neighbours. Figure 1 shows the evolutionary timeline for a single independent region. The physical meaning of the phases depends on the adopted tracers. While this example shows the transition from gas to stars, a similar timeline can be constructed for pairs of gas tracers or star formation tracers. By combining many such tracer pairs, a long timeline of several different, partially-overlapping phases is obtained.



**Figure 1.** Schematic representation of the timeline of an individual star-forming region, which starts out being visible in gas tracer emission (e.g. HI, CO or HCN), and ends up being visible in young stellar tracer emission (e.g. H $\alpha$  or FUV). In between, there is an overlap phase during which both tracers are visible. The duration of each corresponding time-scale is indicated below the timeline.

## 2.1 The star formation relation must break down on small spatial scales

In its most general form, the first result of KL14 is that if a macroscopic correlation (e.g. the star formation relation between the gas mass  $M_{\text{gas}}$  and the SFR) is caused by a time evolution (e.g. the conversion from gas to stars), then it *must* break down on small spatial scales because the subsequent phases of that time evolution are being resolved. When treating galaxies as a whole, this is generally not a concern, because most often they consist of a number of independent regions sufficiently large to ensure that the timeline of Figure 1 is well-sampled.<sup>1</sup> However, there must be some spatial scale below which this is no longer true. Given some typical separation length  $\lambda$  between independent regions, the KL14 principle predicts that the relative uncertainty or Poisson error on the star formation relation is only smaller than unity (i.e. the relation is well-defined and does not ‘break down’) when the condition is satisfied that

$$\Delta x \Delta t^{1/2} \geq \lambda \tau^{1/2}. \quad (2)$$

Here,  $\Delta x$  is the size scale over which the star formation relation is evaluated (the ‘aperture size’),  $\Delta t$  is the shortest<sup>2</sup> evolutionary phase from Figure 1,  $\lambda$  is the separation length of independent regions, and  $\tau$  is the total duration of the evolutionary timeline as in Figure 1. In KL14, we showed that the observed scale dependence of the scatter on the star formation relation (e.g. Bigiel et al. 2008; Schrubba et al. 2010; Onodera et al. 2010; Liu et al. 2011; Leroy et al. 2013) is quantitatively reproduced by this simple model.<sup>3</sup>

The key insight drawn from reproducing the observed, scale-dependent scatter with the simple schematic model of Figure 1, is that the observed relation on 0.1–1 kpc scales provides information on processes taking place on much smaller scales. Another important implication is that the galactic-scale star formation relation results from taking an ensemble average over the cloud-scale physics.

<sup>1</sup> Certain dwarf galaxies, high-redshift galaxies, and galaxy mergers may represent exceptions to this idea if they host a limited number of star-forming regions, or their star-forming regions are subject to a large-scale synchronisation of their evolutionary states.

<sup>2</sup> This minimization only draws from the gas and young stellar phases ( $t_{\text{gas}}$  and  $t_{\text{star}}$ ) and does not include the ‘overlap’ phase  $t_{\text{over}}$ .

<sup>3</sup> While Figure 1 shows a Lagrangian model that follows a single idealised region in time, a similar conclusion was reached by Feldmann et al. (2011, 2012) using a Eulerian model describing instantaneous snapshots of populations of independent regions.

This is particularly relevant for interpreting the recent work showing that the cloud-scale star formation relations of actively star-forming regions have gas depletion times ( $t_{\text{depl}} \equiv M_{\text{gas}}/\text{SFR}$ ) shorter by a factor of 5–50 than the galactic star formation relation (e.g. Heiderman et al. 2010; Lada et al. 2010, 2012).

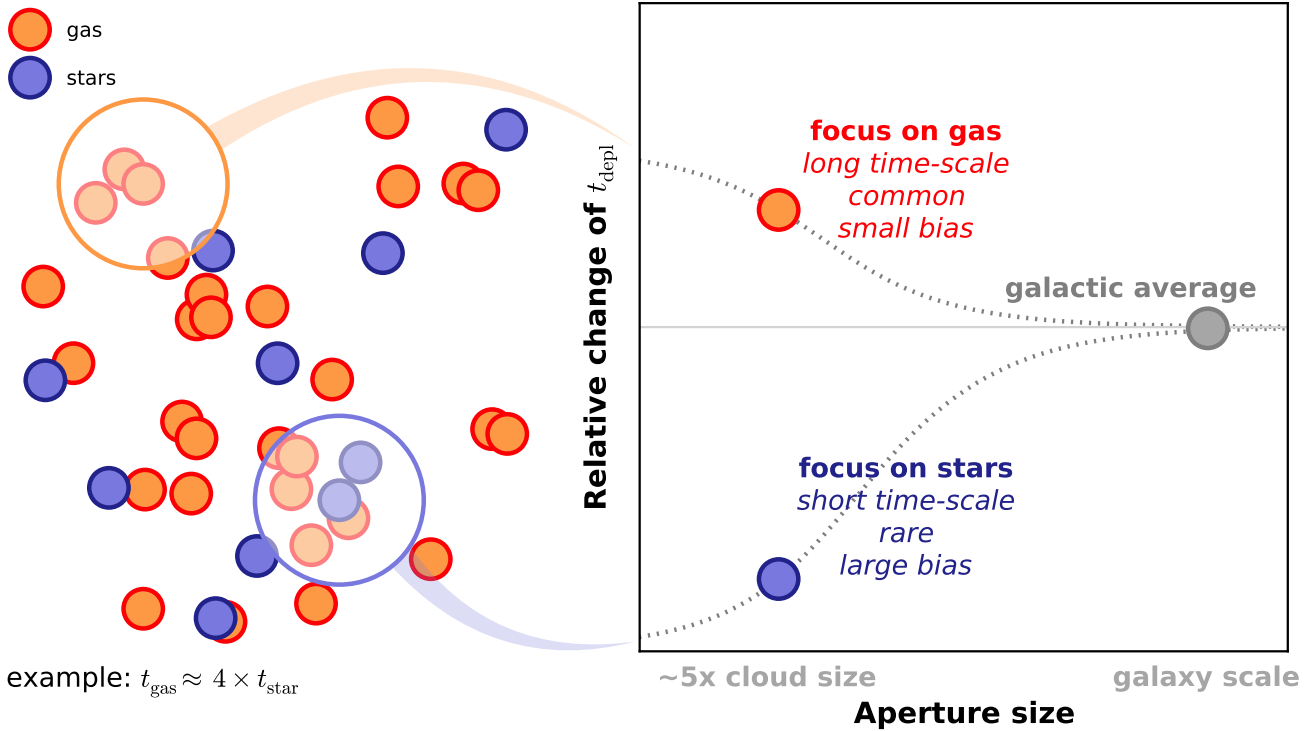
In the context of the KL14 principle, the difference between the star formation relations on cloud and galaxy scales is not surprising. In order to place a single region in the same  $M_{\text{gas}}$ –SFR diagram as galaxies, it must both contain gas and star formation tracer emission. For the timeline of Figure 1, this is equivalent to requiring the region to reside in the overlap phase. As a result, cloud-scale star formation relations cannot consider gaseous regions on the part of the timeline preceding the overlap phase and they therefore omit the gas emission from all regions outside the overlap phase. Assuming that the gas flux of a gaseous region (i.e. a region residing in the phase covered by  $t_{\text{gas}}$ ) does not strongly depend on the evolutionary phase, this means that the gas depletion times of individual regions should be a factor of  $t_{\text{over}}/t_{\text{gas}}$  shorter than on galactic scales. For fiducial values of  $t_{\text{over}} \sim 3$  Myr and  $t_{\text{gas}} \sim 30$  Myr, this implies a bias of roughly one order of magnitude, which is consistent with the large offset observed between  $t_{\text{depl}} \sim 100$  Myr on the cloud-scale (using CO, Heiderman et al. 2010; Lada et al. 2012) and  $t_{\text{depl}} \sim \text{Gyr}$  on the scales of entire galaxies (Bigiel et al. 2008; Schrubba et al. 2011).<sup>4</sup> An example of how to account for the region selection bias in targeted star formation studies is given in Schrubba et al. (2017).

Next to setting the absolute value of the gas depletion time, the part of the evolutionary timeline that is being traced should also influence the scatter of the star formation relation. Selecting higher-density gas (tracers) as in Lada et al. (2010) places even stronger limits on the part of the evolutionary timeline that is being probed than when using low-density gas (defined by using CO or a certain level of extinction), thus limiting the sample to regions at even more similar evolutionary stages. As a result, the scatter on the star formation relation should decrease towards higher gas densities, as is indeed observed by Lada et al. (2010).

## 2.2 The scale dependence of the star formation relation reveals cloud-scale physics

The second result presented in KL14 is that the *way in which* star formation relations depend on the spatial scale is a direct probe of the physics of star formation and feedback on the cloud scale. It exploits the aforementioned notion that the observed star formation

<sup>4</sup> This depletion time bias can be increased further by differences in the adopted tracers. For instance, the quoted example adopts CO-based gas masses and depletion times, but cloud-scale studies often use a dense gas tracer like HCN or a minimum extinction contour containing even less mass than traced with CO. Such a choice results in an even shorter gas depletion time. Differences in SFR tracers are less important, because these are generally calibrated to translate the observed flux to a rate (which implicitly involves division by a reference time-scale). This is why the cloud-scale SFR inferred from young stellar object counts and galaxy-scale SFR measurements using H $\alpha$  do not necessarily differ by orders of magnitude, even though they technically trace different phases in Figure 1. These differences and similarities between SFR and gas tracers are physical in nature and reflect the time evolution of the collapse and star formation process. No empirical star formation relation is therefore intrinsically incorrect. However, unless the appropriate care is taken to compare the right quantities, it may be misleading to compare different star formation relations.



**Figure 2.** Schematic example of how the **KL14** principle can be used to characterize the cloud-scale evolutionary timeline of star formation and feedback. Left panel: cartoon of a ‘galaxy’ consisting of a random distribution of independent regions (circles), which are situated on the timeline of **Figure 1** in a way that is uncorrelated to their neighbours. Orange circles indicate regions in the gas phase, whereas blue circles indicate those in the young stellar phase. In this example, the duration of the gas phase is 4 times that of the young stellar phase. The large circles represent apertures focused on a gas peak (orange) or a stellar peak (blue). Right panel: relative change of the gas depletion time (or the gas-to-stellar flux ratio) when focusing apertures on gas peaks (top branch) or young stellar peaks (bottom branch), as a function of the aperture size. On large scales, the galactic average is retrieved and the relative change is unity. However, on small scales (corresponding to the separation length  $\lambda$ , which is typically several times the cloud size), the excess or deficit of the gas depletion time in this ‘tuning fork diagram’ is a non-degenerate, direct probe of the time and size scales governing the timeline of **Figure 1**.

relation on 0.1–1 kpc scales provides information on processes taking place on much smaller scales. However, contrary to phrasing the scale-dependence in terms of the scatter of the star formation relation to access this information, this feature of the **KL14** principle uses the absolute change (or ‘bias’) of the gas depletion time when focusing a small aperture on gas or young stellar peaks to determine how rare or common the central peak is. That way, the **KL14** principle can be used to constrain the time-scales governing the evolutionary timeline of **Figure 1**.

We illustrate how the **KL14** principle is applied to characterize cloud-scale star formation with an idealised example in **Figure 2**. Imagine a two-dimensional, random distribution of points representing independent regions. Some part of these are dominated by gas, whereas the other part is dominated by young stellar emission. These regions represent two successive phases in the star formation process that do not overlap in time. Let us assume that time spent by a star-forming region in the ‘gas’ phase is 4 times longer than the time spent in the ‘stellar’ phase. This means that gas-dominated regions (‘gas peaks’) will be 4 times more numerous than stellar-dominated regions (‘stellar peaks’). If we then consider a small region around a gas peak (defining an aperture of some size smaller than the galaxy), the local gas depletion time (i.e. the gas-to-stellar flux ratio) in that region will be elevated compared to the galaxy-wide average, because focusing on a gas peak guarantees some excess gas flux to be present, whereas the rest of the aperture is randomly filled with gas or stellar peaks according to the galactic

average. However, the depletion time excess will be minor – the gas peaks are 4 times more common than the stellar peaks, hence the relative effect of guaranteeing the already-ubiquitous gas flux to be present is small. By contrast, if we focus an aperture on one of the rare stellar peaks (with a duration 4 times shorter than the gas phase), the corresponding decrease of the local gas depletion time is large compared to the galactic average, because a very rare phase is guaranteed to be present in the aperture.

This idealised example illustrates the fundamental thought behind the **KL14** principle: the relative rarity of the subsequent phases in the cloud-scale star formation process is set by their relative durations and can be constrained from variations of the gas depletion time on small spatial scales. The **KL14** principle provides a way of ‘counting’ (and assigning relative time-scales to) these phases even if the two-dimensional structure of their emission is continuous and hard to quantise. If the duration of one of the two phases is known, then the relative time-scales translate to an absolute evolutionary timeline. In practice, this known ‘reference time-scale’ will often refer to the stellar phase in the examples of **Figures 1** and **2**, because stellar population synthesis models provide the age range over which a coeval stellar population is bright in commonly-used SFR tracers such as  $H\alpha$ , FUV, and NUV (**Haydon et al. 2018**). As demonstrated in **KL14**, the small-scale excess or deficit of the depletion time when focusing apertures on gas or young stellar peaks (with a known lifetime) is then set by the three free parameters  $t_{\text{gas}}$ ,  $t_{\text{over}}$ , and  $\lambda$ . Most importantly, we showed that these free pa-



parameters describe the ‘tuning fork diagram’ of Figure 2 in a non-degenerate way. The absolute timeline of Figure 1 can therefore be uniquely constrained by fitting a model tuning fork to observations.

The sampling effects described by the KL14 principle manifest themselves on spatial scales much larger than the sizes of individual clouds or star-forming regions (Schruba et al. 2010), because they emerge on size scales comparable to the typical separation between such regions. This new method is therefore highly suitable for application to observations with physical resolutions of  $10\text{--}10^3$  pc, depending on the properties of the target galaxy. This provides a strong contrast with respect to previous methods, which required individual clouds to be resolved and could therefore only be applied to galaxies in (or around) the Local Group. Here, we present the step-by-step process for applying the KL14 principle to observations and carry out detailed numerical tests using numerical simulations of isolated disc galaxies.

### 3 STEP-BY-STEP DESCRIPTION OF APPLYING THE METHOD TO GALAXY MAPS

In this section, we first describe the method in qualitative terms, before turning to a detailed description. The reader interested in obtaining a cursory understanding of the method is referred to Section 3.1. Anyone planning practical applications of the method to observations or to simulated galaxy maps is encouraged to also give Section 3.2 a close read.

#### 3.1 Qualitative description of the method

We have named the machinery for applying the method the ‘HEISENBERG’ code. HEISENBERG has been developed in the Interactive Data Language (IDL),<sup>5</sup> which is commonly used for handling and analysing astrophysical data sets and is well-supported through several public libraries. It is compatible with IDL versions 7.0 and later. A translation of the code into PYTHON is planned for the near future. HEISENBERG currently has a couple of dependences on publicly available routines, most notably from the IDL Astronomy User’s Library,<sup>6</sup> the IDL Coyote Library,<sup>7</sup> and CLUMPFIND (Williams et al. 1994).<sup>8</sup> Several of these routines have required modifications to ensure or optimise their compatibility with HEISENBERG. The public release of HEISENBERG will therefore include these modified routines.<sup>9</sup>

In summary, the method is aimed at the systematic and quantitative application of the formalism sketched in Section 2.2 to real data sets. This process requires five basic ingredients:

- (i) selecting two maps of tracers that represent causally-related phases in a Lagrangian timeline as depicted in Figure 1;
- (ii) identifying emission peaks in this pair of maps;

<sup>5</sup> <http://www.harrisgeospatial.com/ProductsandSolutions/GeospatialProducts/IDL.aspx>

<sup>6</sup> <https://idlastro.gsfc.nasa.gov/>

<sup>7</sup> <http://idlcoyote.com/>

<sup>8</sup> <http://www.ifa.hawaii.edu/users/jpw/clumpfind.shtml>

<sup>9</sup> The HEISENBERG code for applying the described method is currently still proprietary, but it is planned to become publicly available in the near future, some time after the appearance of the present paper. It will become available at <https://github.com/mustang-project/>. The interested reader is welcome to contact the first author for further details.

(iii) measuring the flux ratio of both maps around these peaks as a function of the spatial averaging scale;

(iv) fitting a statistical model that describes the time-scale dependence of the resulting flux ratios (cf. the tuning fork diagram of Figure 2);

(v) carrying out the full error propagation and calculating any derived quantities that follow from the time-scales.

As we show in this paper, these steps enable the characterisation of the cloud-scale physics of star formation and feedback across a statistically representative galaxy sample.

In principle, the method can be applied to a pair of galaxy maps showing any tracer of interest, but physically meaningful results are only obtained if the tracer pair is related by a Lagrangian evolutionary step as in Figure 1. Throughout this paper, we will mainly assume the example case in which ‘gas’ (traced by e.g. HI, CO, HCN, HCO<sup>+</sup>, or sub-mm dust continuum) turns into (young) ‘stars’ (traced by e.g. H $\alpha$ , FUV, or NUV). However, it is also possible to mask these tracers by any physical quantity for which one expects a monotonic change with time, in order to measure the time-scale on which this change takes place. For instance, if one assumes that molecular clouds evolve towards higher densities and excitation conditions during their collapse towards star formation, it is possible to set a maximum or minimum CO(3–2)/CO(1–0) ratio for which the CO(1–0) emission is shown, resulting in two CO maps of low- and high-excitation CO. A characteristic lifetime can be obtained for each of these, providing a time-scale for the evolution towards high-excitation conditions. Another example would be to mask by velocity dispersion in order to measure turbulent energy dissipation time-scales in clouds that evolve towards star formation. Similarly, ionized emission line ratios at optical wavelengths can be used to isolate individual stellar feedback mechanisms such as photoionising radiation, stellar winds, and supernovae, each of which take place on different time-scales, or physical quantities such as electron temperatures and densities (e.g. Blair & Long 2004; Pellegrini et al. 2010; McLeod et al. 2015, 2016). Masking maps of ionized emission lines based on these mechanisms or quantities can thus provide insight in the anatomy of cloud-scale stellar feedback. These are just a few examples – there exist many more applications to a broad variety of astrophysical problems.

The practical application of the above five steps requires several additional steps to be taken. Figure 3 shows the structure of the HEISENBERG code, which has been developed for this purpose. Starting from the bottom of the schematic and proceeding in the clockwise direction, the method goes through the following steps (each also listing the subsection numbers in Section 3.2 where that step is discussed in detail).

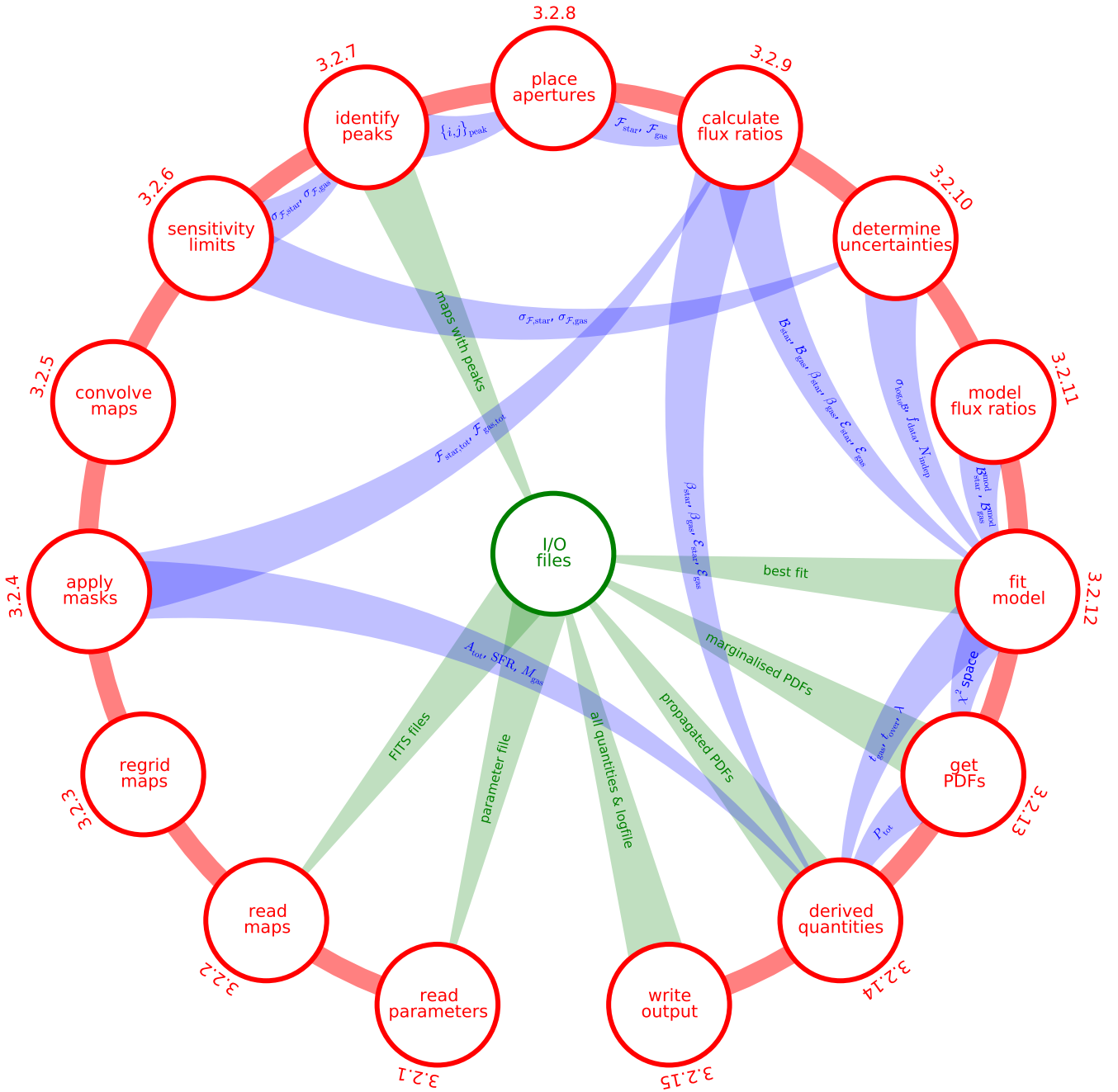
(i) Section 3.2.1: Read in the input parameter file that contains the settings for carrying out the analysis and use these to derive any additional parameters.

(ii) Section 3.2.2: Read in the FITS galaxy maps and determine the smallest possible aperture size given their spatial resolutions. One of these maps must show a phase from the timeline of Figure 1 with a known duration.

(iii) Section 3.2.3: Regrid the galaxy maps by convolving them to the best common spatial resolution and changing the pixel grid to avoid extreme oversampling of a resolution element.

(iv) Section 3.2.4: Apply any masks or cuts in galactocentric radius that are specified in the parameter file to either restrict the analysis of the galaxy maps to these areas or to omit them from the analysis.

(v) Section 3.2.5: Convolve the galaxy maps to the range of spa-



**Figure 3.** Schematic representation of the method that is presented and validated in this paper, specifically following the structure of the HEISENBERG code. The red circles indicate key steps in the process, with section numbers referring to the subsections in Section 3.2 that detail each step. The procedure runs from the bottom (‘read parameters’, Section 3.2.1) in the clockwise direction along the red connections until it is back at the bottom (‘write output’, Section 3.2.15). Several of the steps interact with files on disk, by reading files or producing output. These steps are indicated with green wedges that connect to the green central ‘I/O files’ circle. Temporary file output is not shown. The information content indicated in each wedge is contained by the circle at the wide end and flows towards the narrow end of the wedge. Likewise, the blue curved wedges indicate the flow of the highlighted variables or information between the several steps in the process. To avoid cluttering, these represent the most important (but non-exhaustive) subset of all information flows in the procedure.

tial scales (i.e. aperture sizes) specified in the parameter file using the chosen convolution kernel, with the goal of measuring the flux within a given radius around each position in the maps.

(vi) Section 3.2.6: Determine the sensitivity limit and flux zero point of each galaxy map by fitting a Gaussian to the low-flux end of the pixel flux probability distribution function (PDF).

(vii) Section 3.2.7: Identify flux peaks in the two galaxy maps

with CLUMPFIND,<sup>10</sup> if desired using a different pair of maps than

<sup>10</sup> There exist several modern, more accurate ways of defining clumps, overdensities, or structures and their properties in astrophysical data sets (e.g. Rosolowsky & Leroy 2006, Henshaw et al. 2016, and ASTRODENDRO, see <http://www.astrodendro.org>). However, for the purpose of the presented method, we only need an algorithm to identify local maxima



those used in the multi-scale flux integration, omit peaks with total fluxes below the sensitivity limit or those that should be excluded due to the area masking, and output images of the maps highlighting the positions of the included peaks.

(viii) Section 3.2.8: Calculate the fluxes within the apertures of varying size focused on each peak, as well as the effective aperture size and areas accounting for masked pixels, and the distances between all possible peak pairs.

(ix) Section 3.2.9: Calculate the flux ratio ‘bias’ (i.e. excess or deficit) relative to the galactic average for each set of peaks as in the tuning fork diagram of Figure 2. For each aperture size and within apertures focused on each of the two sets of flux peaks, first calculate the total flux enclosed by the apertures in each of the two maps. This results in four total fluxes per aperture size: the flux in the first/second map around the peaks identified in the first/second map. For each aperture size and peak type, these fluxes are obtained by drawing Monte-Carlo realisations of peak subsets that have non-overlapping apertures, to ensure no pixels are counted twice, and by then taking the mean of the resulting total fluxes across all realisations. The flux ratio excess (i.e. the top branch from Figure 2) then follows by taking the peak sample from the first map and dividing the total flux around these peaks in the first map by that in the second map. The flux ratio deficit (i.e. the bottom branch from Figure 2) then follows by taking the peak sample from the second map and dividing the total flux around these peaks in the first map by that in the second map. These flux ratio biases are then normalised by the average flux ratio between both maps across the entire field.

(x) Section 3.2.10: Determine the uncertainties on the observed flux ratio bias measurements. The sources of uncertainty that should be considered are the sensitivity limits of the maps, which leads to flux uncertainties, and the variety of region masses, which manifests itself as a non-zero variance of the fluxes in individual apertures. Because the data points are flux ratios and thus originate from two fluxes, it is necessary to subtract the covariance between the numerator and the denominator, which reflects spatial trends in region properties. Finally, the uncertainty on each individual data point differs from the effective error bar when fitting a model to the data points, because the data points are not independent between different aperture sizes. Each data point on a branch in the tuning fork (Figure 2) represents a different aperture size that is centred on the same set of peaks, implying that larger apertures (partially) contain the same flux as smaller apertures. We account for this by calculating the ‘independence fraction’ of each data point relative to all other data points.

(xi) Section 3.2.11: Derive a model that connects the flux ratio biases in the tuning fork diagram of Figure 2 to the evolutionary timeline of Figure 1. Such a model was presented in KL14, which considered a random distribution of point-like regions that are situated at random positions on the evolutionary timeline and also allowed these regions to undergo flux evolution between the overlapping and ‘isolated’ phases in Figure 1. However, practical applications of this model also require it to deal with regions of finite densities, i.e. regions characterized by extended emission rather than point particles. The present paper greatly improves on KL14 by assuming that the regions follow a certain spatially-extended profile and then using the flux density contrast between the peaks and

their immediate surroundings to determine the characteristic sizes of these regions in units of the region separation length.

(xii) Section 3.2.12: Carry out a reduced- $\chi^2$  fit of the model to the data points and their (independence-weighted) uncertainties to determine the quantities describing the time-evolution of independent regions. In the context of Figure 1, we thus constrain the best-fitting values of the lifetime of the first phase  $t_{\text{gas}}$  (assuming that the second phase provides the reference time-scale described in Section 2.2), which in applications to molecular gas maps represents the molecular cloud lifetime, the duration of the overlap phase  $t_{\text{over}}$ , which in applications to molecular gas and SFR tracer maps represents the feedback time-scale, and the mean separation length between independent regions  $\lambda$ . Because we carry out a reduced- $\chi^2$  fit, the fitting process also returns the three-dimensional PDF of the above three model parameters. Additional quantities that are also returned by the fitting process (but are not specifically fitted for) are  $\beta_{\text{star}}$  and  $\beta_{\text{gas}}$ , which refer to the mean flux ratios of regions residing in the overlap phase relative to their ‘isolated’ phases, i.e. outside of the overlap phase (see Section 3.3 of KL14). These two parameters capture the (possibly complex) flux evolution of both tracers over the evolutionary timeline. This step also generates the output tuning fork diagram including the best-fitting model as a figure and as an ASCII table.

(xiii) Section 3.2.13: Get the marginalised, one- and two-dimensional PDFs of each of the (pairs of) three free parameters in the model, i.e.  $t_{\text{gas}}$ ,  $t_{\text{over}}$ , and  $\lambda$ , and write the resulting figures and ASCII tables to disk. The error bars on the best-fitting values returned by the procedure refer to the 32nd percentile of the part of the one-dimensional PDF *below* the best-fitting value and the 68th percentile of the part of the PDF *above* the best-fitting value.

(xiv) Section 3.2.14: Calculate derived quantities using the fundamental free parameters  $t_{\text{gas}}$ ,  $t_{\text{over}}$ , and  $\lambda$ . These quantities are the total star formation tracer lifetime including the overlap phase ( $t_{\text{star}}$ ), the total duration of the evolutionary timeline ( $\tau$ ), the region radii ( $r_{\text{star}}$  and  $r_{\text{gas}}$ ), the region size-to-separation ratios or filling factors ( $\zeta_{\text{star}}$  and  $\zeta_{\text{gas}}$ ), the feedback outflow velocity ( $v_{\text{fb}}$ ), the field-wide gas depletion time ( $t_{\text{depl}}$ ), the star formation efficiency per star formation event ( $\epsilon_{\text{sf}}$ ), the star formation and mass removal rates per star formation event ( $\dot{M}_{\text{sf}}$  and  $\dot{M}_{\text{fb}}$ ), the instantaneous and time-averaged mass loading factors ( $\eta_{\text{fb}}$  and  $\bar{\eta}_{\text{fb}}$ ), and the feedback energy and momentum efficiencies ( $\chi_{\text{fb},E}$  and  $\chi_{\text{fb},p}$ ). The one-dimensional PDFs of each of these quantities are determined using Monte-Carlo error propagation, where we draw from the complete, three-dimensional PDF of  $t_{\text{gas}}$ ,  $t_{\text{over}}$ , and  $\lambda$  obtained from the fitting process as well as from the PDF of any other, independent quantities that are used to derive the above quantities. This typically leads to well-defined, nearly-noiseless cumulative PDFs. As in the previous step, each of the resulting PDFs are written to disk as figures and ASCII tables.

(xv) Section 3.2.15: Write the output files as ASCII tables that summarise all constrained quantities and their error bars, as well as the logfile containing all of the command line analysis output. These are critical for evaluating and interpreting the results.

The above steps can be carried out on modern personal computers or laptops within a reasonable computing time. On such systems, we have achieved typical runtimes of 1–20 minutes per experiment, both for the experiments carried out in this paper and for the first applications of this method (Kruijssen et al. 2018; Haydon et al. 2018; Hygate et al. in prep.; Schrubba et al. in prep.; Chevance et al. in prep.; Ward et al. in prep.). This means that the method can

in two-dimensional maps. These peak coordinates should be measured reliably, but we do not use the clump properties derived by CLUMPFIND for any important steps in the analysis. This is a relatively simple task for which the results provided by CLUMPFIND are entirely adequate.

readily be applied to observational and numerical data sets without requiring special-purpose hardware or supercomputing time.

### 3.2 Detailed description of the method

We now turn to a detailed description of the method summarised in Section 3.1 and Figure 3. In this procedure, we extract the quantities describing the time evolution of Figure 1 from two galaxy maps that each show the emission from one phase in the timeline. For simplicity and clarity, we assume in this description that one of the two maps shows molecular gas traced by CO (referred to as ‘gas’) and the other shows star formation traced by  $H\alpha$  (referred to as ‘stars’ or ‘stellar’, even though it should technically be ‘young stars’ or ‘massive stars’). Of these two maps, the stellar phase has a known duration and therefore acts as a ‘reference time-scale’ that sets the absolute scale of the entire timeline of Figure 1 – stellar population synthesis models show that the  $H\alpha$  line is emitted over a 5 Myr time-scale (Haydon et al. 2018). Even though this particular example is used here, we expressly reiterate that the method is capable of handling any pair of emission maps that represent causally-related phases in a Lagrangian timeline as shown in Figure 1. Other examples of tracers that could be used are atomic gas (i.e. HI), dense gas (e.g. traced by HCN), 24  $\mu\text{m}$  emission, ionized emission line tracers (e.g. [OIII] or [NII]), or UV emission, but these are by no means exhaustive.<sup>11</sup>

#### 3.2.1 Input parameters

In order to carry out the analysis depicted in Figure 3, we must first specify a set of flags and input parameters. These are listed in Tables 1 and 2, respectively, and are listed in an input parameter file that is read by the HEISENBERG code. Focusing first on Table 1, the flags are divided into four main categories, separated by white space in the table. The first category concerns optional code modules or steps, the second enables the use of ancillary maps for carrying out the peak identification, the third covers the masking-related options, and the fourth lists the analysis options. Table 1 highlights the default choice of each of these flags, which will be used throughout the paper unless stated otherwise.

While the descriptions of the flags in Table 1 are mostly self-explanatory, we should comment in some detail on a subset for which the choice between the available options may not be obvious.

(i) The `tophat` flag enables the use of either a two-dimensional Gaussian or `tophat` kernel to convolve the maps to larger aperture sizes. Technically, only the `tophat` kernel does this correctly – for each pixel, it adds up all flux within the specified aperture area centred on that pixel. The Gaussian kernel adds up the surrounding flux too, but weighs it by a Gaussian profile. We do not use this kernel in the present paper, because it under-represents the distant pixels within the specified aperture size, but leave it as an available option for future work.

(ii) The `loglevels` flag specifies whether the contour levels that are used to identify peaks with CLUMPFIND are equally-spaced in linear or logarithmic space. In principle, there is no correct choice – it depends on the maps to be analysed which option best identifies peaks of interest. Some trial-and-error is necessarily involved. However, given that the mass functions of molecular clouds and stellar clusters are characterized by power laws (see

e.g. the reviews by Portegies Zwart et al. 2010; Dobbs et al. 2014; Kruijssen 2014), we prefer using logarithmically-spaced contour levels. As demonstrated below, this results in a satisfactory identification of peaks in the maps used in this paper.

(iii) The `flux_weight` flag enables the peak position to correspond to the brightest pixel or the flux-weighted mean within the peak area. The latter option uses the area of the peak as defined by CLUMPFIND. Because there is no obvious physical definition of the peak edge (and hence its area), we prefer to use the brightest pixel.

(iv) The `tstar_incl` flag indicates whether, in the context of the evolutionary timeline shown in Figure 1, the specified ‘reference’ time-scale  $t_{\text{star,ref}}$  (see Section 2.2) includes the overlap phase (with duration  $t_{\text{over}}$ ) or not. This depends entirely on the tracers used. In the present paper, we use pairs of maps generated from galaxy simulations. These map pairs can consist of two stellar maps covering star particles with two different, specified ranges of ages (used to test whether the method retrieves the correct evolutionary timeline, see Section 4.2), or of one such stellar map together with a gas map (used to correct the accuracy of the method under a variety of observational conditions, see Section 4.3). In the cases where we use pairs of stellar maps, the reference time-scale includes the time overlap with the other stellar map. However, when using a stellar map and a gas map, the reference time-scale does not include this overlap phase, because new star particles may form as long as a region still contains gas. Similar considerations will hold in observational applications. When applying the presented method to measure a molecular cloud lifetime as traced by CO and using an ionized emission line like  $H\alpha$  or broadband UV to trace star formation, it is reasonable to assume that the corresponding reference time-scale does not include the overlap phase.  $H\alpha$  and UV emission trace unembedded, massive stars, which locally have blown out the residual gas while massive star formation may still be ongoing in other parts of the star-forming region (e.g. Ginsburg et al. 2016). The emission lifetimes of SFR tracers such as  $H\alpha$  and FUV or NUV are set by stellar population synthesis models, which assume a coeval stellar population (Haydon et al. 2018, also see Leroy et al. 2012), and therefore do not include the time  $t_{\text{over}}$  during which stars may still be forming – the clock starts ticking when there is no molecular gas left.

Turning to the description of the input parameters in Table 2, there are six main categories, again separated by white space in the table. The first category contains quantities describing the basic galaxy properties and manipulation of the maps, the second covers the definition of the aperture sizes to be used in the analysis, the third concerns the peak identification, the fourth describes the quantities needed to define the evolutionary timeline, the fifth sets the parameters for carrying out the fit of our model to the data, and the sixth lists conversion factors and constants needed to calculate the derived physical quantities discussed in Sections 3.2.14 and 5.

As before, many of the input parameters in Table 2 are self-explanatory, but we will discuss a subset of twelve items for which the choice of value is non-trivial.

(i) The parameters  $R_{\text{min}}$  and  $R_{\text{max}}$  can be chosen to limit the analysis to a certain galactocentric radial interval within a galaxy (if `cut_radius` = 1), which can be of great interest when studying any of the quantities constrained by our method, e.g. the molecular cloud lifetime, star formation efficiency, or outflow velocity. However, care must be taken that the radial interval does not become too narrow, i.e. when the area contains too few peaks (this statement is quantified in Section 4.3.8) or that it limits the maximum attainable effective aperture size such that the galactic average in

<sup>11</sup> Applications of the presented method to a wide variety of different tracers in the Large Magellanic Cloud will be presented by Ward et al. in prep.

**Table 1.** Flags to be set for the presented analysis

Flag	Values (default)	Description
mask_images	0/1	Mask images ( <b>off/on</b> )
generate_plot	0/1	Generate output plots and save to PostScript files ( <b>off/on</b> )
derive_phys	0/1	Calculate derived physical quantities as described in Section 3.2.14 ( <b>off/on</b> )
write_output	0/1	Write results to output files ( <b>off/on</b> )
use_star2	0/1	Use a second map for identifying stellar peaks and use the default map for performing the flux calculation ( <b>off/on</b> )
use_gas2	0/1	Use a second map for identifying gas peaks and use the default map for performing the flux calculation ( <b>off/on</b> )
mstar_ext	0/1	Mask areas in stellar map exterior to regions listed in a specified DS9 region file ( <b>off/on</b> )
mstar_int	0/1	Mask areas in stellar map interior to regions listed in a specified DS9 region file ( <b>off/on</b> )
mgas_ext	0/1	Mask areas in gas map exterior to regions listed in a specified DS9 region file ( <b>off/on</b> )
mgas_int	0/1	Mask areas in gas map interior to regions listed in a specified DS9 region file ( <b>off/on</b> )
cut_radius	0/1	Mask maps outside a specified radial interval within the galaxy ( <b>off/on</b> )
set_centre	$\begin{cases} 0 \\ 1 \end{cases}$	Pixel coordinates of the galaxy centre are set to the central pixel of the map Specify pixel coordinates of the galaxy centre ( <b>default</b> )
tophat*	$\begin{cases} 0 \\ 1 \end{cases}$	Use Gaussian kernel to convolve maps to larger aperture sizes Use tophat kernel to convolve maps to larger aperture sizes ( <b>default</b> )
loglevels*	$\begin{cases} 0 \\ 1 \end{cases}$	Contour levels for peak identification are equally-spaced in linear space Contour levels for peak identification are equally-spaced in logarithmic space ( <b>default</b> )
flux_weight*	$\begin{cases} 0 \\ 1 \end{cases}$	Peak positions correspond to brightest pixel in peak area ( <b>default</b> ) Peak positions correspond to flux-weighted mean position of peak area
calc_ap_area	0/1	Calculate the effective aperture area using the number of unmasked pixels within the target aperture area ( <b>off/on</b> )
tstar_incl*	$\begin{cases} 0 \\ 1 \end{cases}$	Reference time-scale $t_{\text{star,ref}}$ does not include the overlap phase (i.e. $t_{\text{star}} = t_{\text{star,ref}} + t_{\text{over}}$ , <b>default</b> ) Reference time-scale $t_{\text{star,ref}}$ includes the overlap phase (i.e. $t_{\text{star}} = t_{\text{star,ref}}$ )
peak_prof	$\begin{cases} 0 \\ 1 \\ 2 \end{cases}$	Model independent regions as points Model independent regions as constant-surface density discs Model independent regions as two-dimensional Gaussians ( <b>default</b> )

\*These flags are discussed specifically in Section 3.2.1. For the other flags, the descriptions in this table are considered to be largely self-explanatory. Examples of setting some of these flags are given in the pertinent subsections below.

Figure 2 cannot be reached. This requires some iteration after the first application of the analysis to a particular data set.

(ii) The parameter  $N_{\text{samp}}$  indicates the maximum number of pixels used to sample the smallest aperture size [which is typically chosen to match the full width at half maximum (FWHM) of the native point spread function (PSF), see below] after regridding the input maps to the same pixel grid. For instance, for the default value of  $N_{\text{samp}} = 10$  and a minimum aperture size of  $l_{\text{ap}} = 50$  pc, the maps will be placed on a grid with  $l_{\text{ap}}/N_{\text{samp}} = 5$  pc pixels. Classical Nyquist sampling corresponds to  $N_{\text{samp}} = 2.5$ , which yields maps with a factor of 16 fewer pixels than our default value and therefore results in improved performance on slow machines. However, we find that this sampling of the aperture size is too coarse to enable the peak identification process to identify all peaks in crowded regions. Setting  $N_{\text{samp}} = 10$  enables accurate peak identification at an acceptable increase in computing time.

(iii) The parameter  $\theta_{\text{ast}}$  defines the maximum allowed tolerance between the pixel coordinates and pixel dimensions of the maps. After the maps have been regridded to an optimal pixel scale (see Section 3.2.3), we verify in Section 3.2.4 that the pixel grids are the same to within  $\theta_{\text{ast}}$ , based on the pixel coordinates and dimensions specified in the headers of the FITS files containing the galaxy maps. We adopt a default tolerance of  $\theta_{\text{ast}} \sim 10^{-6}$  deg  $\sim 3$  mas, which corresponds to  $\sim 0.01$  pc in our simulated galaxy maps ( $D = 840$  kpc), to  $\sim 1$  pc at a distance of  $D \sim 60$  Mpc, and to  $\sim 30$  pc at  $z \sim 3$ . This default angle of 3 mas is well below the spatial resolution used in this paper or the resolution generally achieved with modern observatories. This guarantees that the pixel grid cannot be a source of positioning errors. Of course, this assumes that the pixel coordinates specified in the FITS headers are

correct. The astrometric precision of the maps should be verified manually before applying the presented method (see Section 3.2.2).

(iv) The parameter  $N_{\text{bins}}$  sets the number of bins that is used to sample the pixel flux PDF and fit a Gaussian to determine the sensitivity limit of each map. The default number is  $N_{\text{bins}} = 20$ , which is somewhat arbitrary but yields accurate results.

(v) The parameters  $l_{\text{ap,min}}$ ,  $l_{\text{ap,max}}$ , and  $N_{\text{ap}}$  together define the array of aperture sizes (i.e. diameters) to which the maps are convolved and at which the analysis is carried out. Specifically, the apertures are logarithmically spaced, such that:

$$l_{\text{ap}}(i) = l_{\text{ap,min}} \left( \frac{l_{\text{ap,max}}}{l_{\text{ap,min}}} \right)^{\frac{i}{N_{\text{ap}}-1}}, \quad (3)$$

with  $\{i \in \mathbb{N} \mid 0 \leq i \leq N_{\text{ap}} - 1\}$ . The desired values of  $l_{\text{ap,min}}$ ,  $l_{\text{ap,max}}$ , and  $N_{\text{ap}}$  strongly depend on the problem at hand. The minimum aperture size  $l_{\text{ap,min}}$  should be similar to the coarsest resolution across both maps, whereas  $l_{\text{ap,max}}$  should be large enough to enable the large-scale convergence of the curves in the tuning fork diagram of Figure 2 to the galactic average. The choice of maximum aperture size may therefore require some iteration. Likewise, there is no a priori rule for choosing the number of aperture sizes  $N_{\text{ap}}$ . Most importantly, it should be large enough to sample the shape of the curves in the tuning fork diagram. The possible oversampling of these curves with a large number of aperture sizes (and hence observational data points) is harmless, because the statistical part of our analysis corrects for correlations between the data points (see Section 3.2.10). However, using an unnecessarily large number of aperture sizes can greatly slow the analysis. Again, some iteration may be required to optimize this choice for the maps under consideration. For the tests performed in this paper, we find that the

**Table 2.** Input parameters of the presented analysis

Quantity [unit]	Default	Description
$D$ [pc]	–	Distance to galaxy
$i$ [°]	0	Inclination angle
$\phi$ [°]	0	Position angle
$i_{\text{cen}}$ [pix]	–	Index of $x$ -axis coordinate of galaxy centre (only used if <code>set_centre = 1</code> , otherwise set to central pixel in map)
$j_{\text{cen}}$ [pix]	–	Index of $y$ -axis coordinate of galaxy centre (only used if <code>set_centre = 1</code> , otherwise set to central pixel in map)
$R_{\text{min}}$ [pc]*	0	Minimum inclination-corrected radius for analysis (only used if <code>cut_radius = 1</code> )
$R_{\text{max}}$ [pc]*	10000	Maximum inclination-corrected radius for analysis (only used if <code>cut_radius = 1</code> )
$N_{\text{samp}}$ *	10	Maximum number of pixels per unit FWHM of a map resolution element after regridding the maps
$\theta_{\text{ast}}$ [°]*	$10^{-6}$	Allowed tolerance between the pixel coordinates and pixel dimensions in both maps
$N_{\text{bins}}$ *	20	Number of bins used during sensitivity limit calculation to sample the flux PDFs and fit Gaussians
$l_{\text{ap,min}}$ [pc]*	50	Minimum inclination-corrected aperture size (i.e. diameter) to convolve the input maps to
$l_{\text{ap,max}}$ [pc]*	6400	Maximum inclination-corrected aperture size (i.e. diameter) to convolve the input maps to
$N_{\text{ap}}$ *	8	Number of aperture sizes used to create logarithmically-spaced aperture size array in the range $[l_{\text{ap,min}}, l_{\text{ap,max}}]$
$N_{\text{pix,min}}$ *	20	Minimum number of pixels for a valid peak (use $N_{\text{pix,min}} = 1$ to allow single points to be identified as peaks)
$N_{\sigma}$	5	Flux multiple of the derived sensitivity limit needed for a peak to be included
$\Delta \log_{10} \mathcal{F}_{\text{star}}$ *	2	Logarithmic range below flux maximum covered by flux contour levels for stellar peak identification
$\delta \log_{10} \mathcal{F}_{\text{star}}$ *	0.5	Logarithmic interval between flux contour levels for stellar peak identification
$\Delta \log_{10} \mathcal{F}_{\text{gas}}$ *	2	Logarithmic range below flux maximum covered by flux contour levels for gas peak identification
$\delta \log_{10} \mathcal{F}_{\text{gas}}$ *	0.5	Logarithmic interval between flux contour levels for stellar peak identification
$N_{\text{lin,star}}$ *	11	Number of contours for stellar peak identification with range $\min(\mathcal{F}_{\text{star}})$ – $\max(\mathcal{F}_{\text{star}})$ (only used if <code>loglevels = 0</code> )
$N_{\text{lin,gas}}$ *	11	Number of contours for gas peak identification with range $\min(\mathcal{F}_{\text{gas}})$ – $\max(\mathcal{F}_{\text{gas}})$ (only used if <code>loglevels = 0</code> )
$t_{\text{star,ref}}$ [Myr]*	–	Reference time-scale spanned by star formation tracer
$\sigma_{-}(t_{\text{star,ref}})$ [Myr]	0	Downwards uncertainty on reference time-scale
$\sigma_{+}(t_{\text{star,ref}})$ [Myr]	0	Upwards uncertainty on reference time-scale
$t_{\text{gas,min}}$ [Myr]	0.1	Minimum value of $t_{\text{gas}}$ considered during fitting process
$t_{\text{gas,max}}$ [Myr]	3000	Maximum value of $t_{\text{gas}}$ considered during fitting process
$t_{\text{over,min}}$ [Myr]	0.01	Minimum value of $t_{\text{over}}$ considered during fitting process
$N_{\text{mc,peak}}$ *	1000	Number of Monte-Carlo realisations of independent peak samples to be generated and averaged over
$N_{\text{depth}}$ *	4	Maximum number of free parameter array refinement loops for obtaining best-fitting value
$N_{\text{try}}$ *	101	Size of each free parameter array to obtain the best-fitting value
$N_{\text{mc,phys}}$ *	$10^6$	Number of Monte-Carlo draws used for error propagation of derived physical quantities
$\log_{10} X_{\text{star}}$ *	–	Logarithm of conversion factor from map pixel value to an absolute SFR in $M_{\odot} \text{ yr}^{-1}$ (only used if <code>derive_phys = 1</code> )
$\sigma_{\text{rel}}(X_{\text{star}})$	0	Relative uncertainty (i.e. $\sigma_x/x$ ) of $X_{\text{star}}$ (only used if <code>derive_phys = 1</code> )
$\log_{10} X_{\text{gas}}$ *	–	Logarithm of conversion factor from map pixel value to an absolute gas mass in $M_{\odot}$ (only used if <code>derive_phys = 1</code> )
$\sigma_{\text{rel}}(X_{\text{gas}})$	0	Relative uncertainty (i.e. $\sigma_x/x$ ) of $X_{\text{gas}}$ (only used if <code>derive_phys = 1</code> )
$\Psi_E$ [ $\text{m}^2 \text{ s}^{-3}$ ]*	–	Light-to-mass ratio of a desired feedback mechanism (only used if <code>derive_phys = 1</code> )
$\psi_p$ [ $\text{m s}^{-2}$ ]*	–	Momentum output rate per unit mass of a desired feedback mechanism (only used if <code>derive_phys = 1</code> )

\*These parameters are discussed specifically in Section 3.2.1. For the other parameters, the descriptions in this table are considered to be largely self-explanatory. Examples of setting some of these parameters are given in the pertinent subsections below.

default values of  $l_{\text{ap,min}}$ ,  $l_{\text{ap,max}}$ , and  $N_{\text{ap}}$  provide accurate results. In Section 4.3.6, we vary  $l_{\text{ap,min}}$  and  $N_{\text{ap}}$  to carry out a resolution test and quantify the requirements for observational applications of the method.

(vi) The parameter  $N_{\text{pix,min}}$  represents the minimum area of a peak identified by CLUMPFIND (see Figure 2 of Williams et al. 1994) to be considered in the remainder of the analysis. The default value of  $N_{\text{pix,min}} = 20$  avoids unreliable detections without obstructing the identification of unresolved point sources. For illustration, the default pixel sampling rate  $N_{\text{samp}} = 10$  means that the typical area within an FWHM is 75 pixels, implying that a minimum area of 20 pixels enables the identification of unresolved point sources. To apply the analysis to maps of point-like regions (see Section 4), it is necessary to set  $N_{\text{pix,min}} = 1$ .

(vii) The parameters  $\Delta \log_{10} \mathcal{F}$  and  $\delta \log_{10} \mathcal{F}$  (with subscripts ‘star’ and ‘gas’ referring to stars and gas, respectively) set the logarithmic range and separation of the contour levels used by CLUMPFIND to identify peaks (see Section 3.2.7 for a summary of how this identification is performed, or Williams et al. 1994 for

a detailed discussion). Choosing the best values of these parameters requires the visual inspection of the maps produced by the analysis (see Figure 3). If faint but relevant peaks are not identified,  $\Delta \log_{10} \mathcal{F}$  should be increased, whereas it should be decreased if spurious peaks are identified. If adjacent but independent peaks are not distinguished, one should decrease  $\delta \log_{10} \mathcal{F}$ , whereas it should be increased if multiple peaks within a single independent region are each identified individually. Lacking a quantitative definition of ‘independent regions’ other than the idealised description given in Section 2.1, these choices are necessarily somewhat arbitrary. However, the details of the peak identification do not strongly affect the constrained quantities as long as an ‘obvious’ set of physically relevant peaks is identified, because we carry out Monte-Carlo sampling of peaks to discard close pairs (see below and Section 3.2.8). The default values of  $\Delta \log_{10} \mathcal{F}$  and  $\delta \log_{10} \mathcal{F}$  provide a good starting point for applications of the method and are used for most of the experiments discussed in this work. The same approach applies to the choice of  $N_{\text{lin,star}}$  and  $N_{\text{lin,gas}}$ . These quantities define the contour level spacing if linearly-spaced contours



are chosen (i.e.  $\text{loglevels} = 0$ ) and should be chosen by visual inspection of the maps produced by the analysis. In general, it is important to realise that the peak identification should have access to most of the tracer emission in the maps. If the peak positions of either tracer are biased to areas of the map that do not host most of the tracer emission, e.g. due to differences in spatial structure and clumpiness, this negatively affects the accuracy of the method.

(viii) The parameter  $t_{\text{star,ref}}$  represents the ‘reference time-scale’ that is used to translate the derived *relative* evolutionary timeline to an *absolute* one (cf. Figure 1). Throughout this paper, we assume that  $t_{\text{star,ref}}$  refers to the SFR tracer map, because stellar population synthesis models like STARBURST99 (Leitherer et al. 1999) can provide an absolute ‘lifetime’ for the emission from massive stars. Although the characteristic lifetimes of SFR tracers such as  $\text{H}\alpha$ , FUV, and NUV can vary by more than an order of magnitude depending on the definition used (Leroy et al. 2012, Table 3), they have been calibrated specifically for use in the presented method by Haydon et al. (2018), who also considered the effect of sampling from the initial mass function (IMF) in low-SFR environments. These lifetimes are the best choice for observational applications of the method. Of course, once the lifetime of any other tracer (e.g. CO) in the galaxy under consideration has been measured using our method, it becomes possible to use that lifetime as a reference time-scale for other measurements. In the specific context of our tests of the method using galaxy simulations in Section 4, we use maps of the star particles in specific age bins, which grants us full control over the value of  $t_{\text{star,ref}}$ .

(ix) The parameter  $N_{\text{mc,peak}}$  denotes the number of Monte-Carlo experiments used to draw independent peak samples from the parent sample at each aperture size. In other words, we generate  $N_{\text{mc,peak}}$  different (but inevitably partially overlapping) subsets of peaks such that none of these peaks has any neighbours within  $l_{\text{ap}}$ . As described in Section 3.1, this is a necessary step to make sure that flux around peaks in crowded regions is not counted more than once. The total gas and stellar fluxes around each peak type (i.e. gas peaks in the top branch of Figure 2, as well as stellar peaks in the bottom branch) are then averaged over the  $N_{\text{mc,peak}}$  experiments. Setting  $N_{\text{mc,peak}}$  too low can result in run-to-run variations of the data points and thus the best-fitting quantities. It is therefore important to choose a sufficient number of Monte-Carlo samples, even if this goes at the expense of computing time (which scales as  $N_{\text{mc,peak}}^2$  due to looping over all peak pairs). We find that at the default  $N_{\text{mc,peak}} = 1000$ , the run-to-run variation is of the order 1–2 per cent, which is appreciably smaller than the uncertainties on the best-fitting quantities. This value requires a few minutes of runtime on modern laptops and is therefore used throughout the paper.

(x) The parameters  $N_{\text{depth}}$  and  $N_{\text{try}}$  govern the fitting process. This part of the analysis takes up a significant fraction of the computing time and can even dominate altogether. As discussed in Section 3.1, we carry out the fit in a three-dimensional parameter space of  $t_{\text{gas}}$ ,  $t_{\text{over}}$ , and  $\lambda$ . To achieve a reasonable computing time without sacrificing precision, we therefore iteratively refine the fitting grid by zooming in on the best-fitting point in parameter space. The parameter  $N_{\text{depth}}$  sets the maximum number of refinement steps (see Section 3.2.12 for details) and  $N_{\text{try}}$  sets the number of array elements in each of the three free parameters that is evaluated during each refinement step, resulting in a total number of  $N_{\text{try}}^3$  elements. This third-power scaling of the total number of elements means that it is very important to optimize the choice of  $N_{\text{depth}}$  and  $N_{\text{try}}$ . For the default values of  $N_{\text{depth}} = 4$  and  $N_{\text{try}} = 101$ , we obtain well-converged results within a reasonable runtime, to the extent

that  $N_{\text{depth}}$  is typically not even reached because the condition for convergence is already satisfied during an earlier refinement step.

(xi) The parameter  $N_{\text{mc,phys}}$  sets the number of Monte-Carlo draws used to perform the error propagation for the derived quantities in Section 3.2.14. Again, this process can be quite time-intensive, implying that there is a trade-off between runtime and the smoothness of the PDFs of the derived physical quantities. For the default value of  $N_{\text{mc,phys}} = 10^6$ , the noise on the PDFs of the derived quantities is effectively unnoticeable, whereas it is entirely absent in the cumulative distribution functions (CDFs). To minimize the computing time,  $N_{\text{mc,phys}}$  can be lowered by up to two orders of magnitude, resulting in PDFs that have noticeable noise, but CDFs that are still smooth and well-defined. Throughout the paper, we adopt the default value listed in Table 2.

(xii) The parameters  $\log_{10} X_{\text{star}}$  and  $\log_{10} X_{\text{gas}}$  represent the logarithm of the conversion factors from a pixel value to an absolute SFR (in  $\text{M}_{\odot} \text{yr}^{-1}$ ) and an absolute gas mass (in  $\text{M}_{\odot}$ ), respectively. Their values account for the conversion of the input maps’ flux units to the physical units adopted in HEISENBERG. As such,  $\log_{10} X_{\text{star}}$  and  $\log_{10} X_{\text{gas}}$  depend on the units of the maps under consideration as well as the distance of the galaxy, implying that these parameters need to be calculated prior to applying the method to a certain data set. The native input maps considered in the present paper show gas and stellar surface densities on a pixel scale of  $\Delta x = 14.25 \text{ pc}$  (see Section 4), which means that the conversion factor to an SFR is given by

$$\log_{10} X_{\text{star}} = \log_{10} \left[ \left( \frac{\Delta x}{\text{pc}} \right)^2 \left( \frac{t_{\text{star}}}{\text{yr}} \right)^{-1} \right], \quad (4)$$

where the division by the stellar lifetime is necessary to convert a stellar surface density in a known age bin to a corresponding time-average SFR. For the experiments in this paper, assuming a value of  $t_{\text{star}} = 1 \text{ Myr}$  would imply  $\log_{10} X_{\text{star}} = -3.692$ . Likewise, the conversion factor to a gas mass is given by

$$\log_{10} X_{\text{gas}} = \log_{10} \left[ \left( \frac{\Delta x}{\text{pc}} \right)^2 \right], \quad (5)$$

which for the experiments in this paper becomes  $\log_{10} X_{\text{gas}} = 2.308$ . These conversion factors are updated whenever the maps are regridded to a different pixel scale. Note that other data sets will require other expressions and values. The above examples are intended to illustrate the definition of these conversion factors. Also, we emphasize that the conversion factors are only used when calculating the derived physical quantities in Section 5. The KL14 principle itself and the associated method to constrain the evolutionary timeline of Figure 1 both rely on flux *ratios* relative to the galactic average, implying that the conversion factors cancel.<sup>12</sup>

(xiii) Finally, the parameters  $\Psi_E$  and  $\psi_p$  are the light-to-mass ratio (or energy injection rate per unit mass) and momentum output rate per unit mass, respectively. They are only used when calculating the feedback energy and momentum efficiencies in Section 5, which is done by comparing the effective output energy and momentum injection rates constrained by the method to the total rates implied by the young stellar populations in the star-forming regions. These parameters are needed to derive the total injection rates – their values depend on the feedback mechanism under consideration and the time-scale over which the rates are averaged. For

<sup>12</sup> This only holds if the conversion factors do not vary across the map. See Section 3.2.2 for a discussion of ways to apply the method to galaxies with non-uniform conversion factors.

most feedback mechanisms, suitable numbers for these quantities can be obtained from stellar population synthesis models (e.g. [Leitherer et al. 1999](#); [Agertz et al. 2013](#)).

All flags and input parameters are specified in an input file, which is read in by HEISENBERG. Next to the elements listed in Tables 1 and 2, it also contains the full path of the maps and any DS9 region files used in the analysis to mask the maps.<sup>13</sup> In addition, the input file includes a small number of quantities that are not listed in Tables 1 and 2 because they are not relevant to the present paper, but are required to run the analysis. They are included in the HEISENBERG documentation when it is made publicly available.

### 3.2.2 Selecting and reading in galaxy maps

The next step in the analysis process is to read in the galaxy maps in FITS format to which the method is applied. In Section 3.1, we explained that these maps can show any tracer pair of interest that follows a Lagrangian evolutionary connection as in Figure 1. As stated previously, here we follow an example case in which ‘gas’ turns into (young) ‘stars’, but we reiterate that the presented method applies much more generally.

Irrespective of their physical nature, the pair of maps should satisfy a number of conditions.

(i) The maps should be two-dimensional and share the same dimensions and pixel grid. In the case of a spectral cube containing line emission data, this means one should use a moment-0 map. The pixel grid is defined using pixel coordinates  $\{i, j\}$ , with  $\{i \in \mathbb{N} \mid 0 \leq i \leq N_{\text{pix},x} - 1\}$  and  $\{j \in \mathbb{N} \mid 0 \leq j \leq N_{\text{pix},y} - 1\}$ , where  $\{N_{\text{pix},x}, N_{\text{pix},y}\}$  represent the number of pixels in the  $\{x, y\}$  directions. The pixel values represent the stellar and gas flux densities  $\mathcal{F}_{\text{star},ij}$  and  $\mathcal{F}_{\text{gas},ij}$ , respectively.

(ii) The astrometric precision of the pair of maps should be sufficient, such that positional uncertainties are considerably smaller than the resolution FWHM. The presented method correlates the spatial structure in pairs of tracer maps to derive the underlying evolutionary timeline. It therefore relies strongly on accurate position information. We have quantified this by carrying out experiments with a positional offset in one of the maps, which show that an acceptable astrometric precision is  $\sim 1/3$  of the FWHM. This is commonly achieved with modern observatories, which routinely achieve astrometric precision of the order of 10 per cent of the FWHM or better. For instance, ALMA achieves a precision of 0.05 times the FWHM, with a minimum of 3 mas.<sup>14</sup> Maps for which the astrometric uncertainty exceeds  $1/3$  of the FWHM cannot be used for measuring  $t_{\text{over}}$ , because such large positional offsets prohibit the statistical identification of physically co-spatial regions in both maps. If the astrometric uncertainties are not much smaller than the FWHM, the duration of the overlap phase  $t_{\text{over}}$  (see Figure 1) will be underestimated.

(iii) One has to be relatively confident that each cloud or star-forming region that is visible in one tracer is also visible in the other tracer at some point in its lifecycle. The constrained lifetimes

are a population average over all regions, meaning that if this condition is not satisfied, the lifetime includes a zero-duration contribution from those regions that never appear in one of the tracers. However, it does not necessarily pose a problem if regions never appear in either tracer, because then they are simply omitted from the population-average evolutionary timeline altogether. In such a case, the method returns the ‘visibility time-scale’ and ‘visible separation length’ of the tracers. The retrieved time-scales still match the ‘true’ underlying lifetimes if the number of invisible regions is small or these regions are consistent with being a randomly drawn subset of the parent population. There are no general guidelines for dealing with region visibility, because its implications depends strongly on the question at hand and the tracers used. Specific examples are discussed in Sections 4.3.3 and 7.2.2.

(iv) Contamination should be minimized such that the emission from each tracer can be considered to be almost exclusively associated with the physical objects of interest. In the case of H $\alpha$  emission, one would want to avoid emission from shock-heated galactic accretion or outflows that are not driven by stellar feedback. Likewise, 24 $\mu\text{m}$  emission may trace both young and evolved stars, which complicates its interpretation as a star formation tracer. In general, diffuse emission due to unresolved, low-mass objects or photon scattering at large distances from the emission sources should be removed from the maps. Ideally, this filtered emission is added back into the image by rescaling the pixel values. The main reason for filtering diffuse emission is that it does not belong to the population of peaks that is being studied. In extreme cases, the presence of a large flux reservoir without peaks in one of the tracers can result in a tracer *deficit* relative to the galactic average when focusing on the corresponding tracer peaks, because the diffuse flux contributes to the galactic average without contributing to the flux around the identified peaks (see Section 4.3.9 for a detailed discussion). This should be avoided. A new module in HEISENBERG has the purpose of automatically filtering out diffuse emission ([Hygate et al. 2018](#)), but if this module is not used, the maps should be visually inspected and pre-processed if necessary.

(v) The emission maps should be as homogeneous as possible, avoiding major variations of extinction, excitation conditions (e.g. temperature), and chemical abundances across the considered area. In principle, it is one of the method’s main strengths that it is insensitive to galaxy-to-galaxy variations of the uncertain conversion factors between gas tracer flux and the gas mass that are well-known to hamper studies of extragalactic star formation (e.g. [Daddi et al. 2010](#)). However, if these quantities do vary strongly *within* a map (e.g. in the case of radial gradients in galaxies), this may pose a problem. In that case, one can still apply the method, but it should be acknowledged that a tracer lifetime will be obtained under the population-average conditions of the varying quantity. If this is undesirable (e.g. due to a spatially-varying CO-to-H $_2$  conversion factor), it is recommended to either apply this conversion factor to the map beforehand, or to apply the method separately to radial bins in which any variation is within acceptable limits.

(vi) The sensitivities or detection limits of the maps should enable a representative fraction of the tracer emission of interest to be recovered.

(vii) The spatial resolution of the maps should enable the characteristic separation length between independent regions  $\lambda$  to be resolved. This can only be verified after applying the method and is discussed in detail in Section 4.3.6.

(viii) The largest effective aperture size (after subtracting any masked pixels from the aperture) should enable the convergence towards the galactic average in Figure 2 to be reached. If this is

<sup>13</sup> To create the DS9 region files used for masking (see below), DS9 is available at <http://ds9.si.edu>. In addition, DS9 region files may be created manually, using the region file description that can be found at <http://ds9.si.edu/doc/ref/region.html>.

<sup>14</sup> See the ALMA Technical Handbook at <https://almascience.eso.org/documents-and-tools/>.



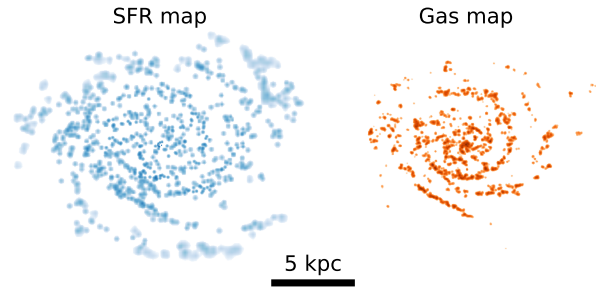
not achieved, any masks or limits on the range of galactocentric radii should be relaxed to increase the largest effective aperture size. Should the galactic average still not be retrieved in any circular aperture (this is possible for irregular fields), the individual maps may be unsuitable for applying the method. The method may still be applied to such galaxies by combining the images of multiple galaxies at the same physical scale side-by-side in a single map, or combining the peak samples obtained for the individual galaxies, as long as these combinations are carried out prior to fitting the model. In this case, the method constrains the population-average evolutionary timeline of Figure 1 for the combined galaxy sample.

(ix) If one wishes to recover absolute time-scales, then the lifetime of the tracer shown in one of the maps must be known, either directly from stellar population synthesis modelling (e.g. when using H $\alpha$ , FUV, or NUV, see Haydon et al. 2018), or by having been calibrated to such a reference time-scale in a separate application of the method. In other words, if the CO lifetime has been measured by applying the method to an H $\alpha$  map and a CO map, then either of these can be used as a reference time-scale when measuring the lifetimes of HI clouds across the same area.

While it may seem that the above points set stringent limits on the applicability of the method, there currently exists a broad range of observational data sets that satisfy these conditions. These conditions are merely intended as guidelines to carefully consider when selecting the maps. In general, the KL14 principle should be thought of as a purely empirical method that as such will not necessarily yield a ‘wrong’ answer if any of the above conditions are violated. However, the choice of maps and their properties together *define* the question that is answered by applying the method. Therefore, the interpretation of the results depends on to what extent these guidelines for selecting the maps are satisfied. In practice, some of the above guidelines cannot be evaluated in advance (e.g. points vii and viii) and must be addressed a posteriori, after the analysis with HEISENBERG is completed. The above list should be evaluated in addition to the guidelines for observational applications from Section 4.4. We note that all of the above conditions are satisfied by the simulated maps that are used in this paper.

It is possible to select different maps for the peak identification and for the flux integration. An example where this may be of interest is when using CO(1–0) observations to trace molecular gas. These position-position-velocity data cubes can be quite noisy, meaning that a signal-masked moment-0 map showing only the emission above a minimum signal-to-noise level is often needed to select the real peaks in the map and avoid selecting noise peaks. However, this omits real flux from the map, because part of the emission below the signal-to-noise threshold is real. In this case, it may be desirable to use a different map that does include this emission when calculating the fluxes in the apertures focused on each peak. For instance, a moment-0 map that is generated without a signal-to-noise threshold, but instead uses a certain velocity mask around the velocity field of atomic gas (if available) may help to include and isolate all real CO emission even in the presence of noise. A flux calculation based on such a map would be accurate on scales larger than a few times the PSF, because the noise maxima and minima should then statistically cancel.

Given a selected set of maps, HEISENBERG reads in the FITS files. Throughout the description of the method, we will use a pair of example maps to illustrate the application of the method. These are taken from the high-resolution, extended emission simulated maps of experiment ID 37 in Section 4.3. We describe later how these particular maps were generated – at this point of the paper,



**Figure 4.** Example of two simulated galaxy maps (the high-resolution, extended-emission versions of experiment ID 37 in Section 4.3) that are used for illustrating the method throughout this section. Shown are the SFR (left) and gas (right) maps on a logarithmic stretch over three orders of magnitude. The spatial scale is indicated by the scale bar at the bottom.

they exclusively serve an illustrative purpose. The maps are presented in Figure 4 and represent the input for later illustrations of the analysis. When reading in the maps, HEISENBERG immediately applies individual masks  $\xi_{\text{star},ij}$  and  $\xi_{\text{gas},ij}$  according to the DS9 region files that are specified for each of them. These region files may be composites of any number of sub-regions and can be used to either include or exclude any pixels of which the pixel centres reside within the area enclosed by the region files. The values of excluded pixels are set to NaN in each individually-masked map. The synchronisation of each of these masks across all maps is deferred to a later step, in which any cuts in galactocentric radius are also applied (see Section 3.2.4).

After reading in the input file as described in Section 3.2.1 and subsequently reading in the maps, their FITS headers, and the DS9 region files, the analysis is self-contained and does not require reading any further files. Before continuing the analysis, we compare the PSF sizes and determine the largest PSF size  $l_{\text{PSF,max}}$  across the set of maps. Assuming a Gaussian PSF, this size corresponds to an FWHM. Given the grid of aperture sizes  $l_{\text{ap}}$  defined by the input file and provided in equation (3), we determine the aperture size that is closest to the maximum PSF size in logarithmic space, i.e. we minimize the quantity  $|\log_{10}(l_{\text{ap}}/l_{\text{PSF,max}})|$ , and set  $l_{\text{ap,min}}$  to the aperture size where this minimum is reached. This ensures that we only consider aperture sizes that are at least marginally resolved in both of the maps.

### 3.2.3 Regridding the galaxy maps

The runtime of the analysis increases with the area of the maps in units of square pixels. For this reason, we regrid the galaxy maps to a smaller number of pixels if the original pixel size  $l_{\text{pix}}$  allows doing so. As discussed in Section 3.2.1, the target pixel size is  $l_{\text{ap,min}}/N_{\text{samp}}$ , which ensures that resolution elements are resolved by a sufficient number of pixels to reliably identify emission peaks. We therefore regrid the maps if the target pixel size  $l_{\text{ap,min}}/N_{\text{samp}} > 1.2l_{\text{pix}}$ , where the factor of 1.2 is added to skip the regridding if this yields less than a 30 per cent decrease of the map area in units of square pixels. In addition, the maps are convolved with a two-dimensional Gaussian PSF to a common resolution of  $l_{\text{ap,min}}$ , provided that  $l_{\text{ap,min}} > 1.05l_{\text{PSF}}$ . Again, the factor of 1.05 is added to skip the convolution if only a 5 per cent change of resolution would be achieved. If the beam is non-circular, we use the effective beam width  $l_{\text{PSF}} = (l_{\text{PSF},a}l_{\text{PSF},b})^{1/2}$ , where  $l_{\text{PSF},a}$  and  $l_{\text{PSF},b}$  indicate the beam width along the semi-minor

and semi-major axes, respectively. The convolution of the maps then proceeds with a kernel FWHM of  $(l_{\text{ap},\text{min}}^2 - l_{\text{PSF}}^2)^{1/2}$ , which ensures that the final resolution equals  $l_{\text{ap},\text{min}}$  irrespective of the original circular PSF size. The FITS header is updated with the new map dimensions  $\{N_{\text{pix},x}, N_{\text{pix},y}\}$ , coordinates of the galactic centre  $\{i_{\text{cen}}, j_{\text{cen}}\}$  (which are taken to correspond to the central pixel of the image if `set_centre = 0`), and pixel size  $\theta_{\text{pix}}$ . Finally, the updated FITS file is written to disk for later use. The above procedure is applied to each of the maps under consideration.

After the regridding is completed, we update several of the quantities needed for the remainder of the analysis.

(i) The original position of the galactic centre pixel  $\{i_{\text{cen}}, j_{\text{cen}}\}$  is converted to a fraction of the dimensions of the original maps in units of pixels  $\{N_{\text{pix},x}, N_{\text{pix},y}\}$ . Interpolation of the pixel dimensions of the regridded maps then provides the position of the central pixel in the new maps and thus updates the original values of  $\{i_{\text{cen}}, j_{\text{cen}}\}$ .

(ii) The pixel size in pc is determined from the angular size  $\theta_{\text{pix}}$  in the FITS file header of the regridded map by writing

$$l_{\text{pix}} = \frac{D \tan \theta_{\text{pix}}}{\sqrt{\cos i}}, \quad (6)$$

where the square root in the denominator accounts for the fact that a non-zero inclination only affects one of the pixel dimensions and thus does not enter linearly in the effective pixel size. If the distance or inclination angle are poorly constrained, we recommend carrying out the analysis for a reasonable range of distances or inclination angles to probe the impact of their uncertainties on the inferred quantities.

(iii) We check that there are at least two elements in the aperture size array  $l_{\text{ap}}$  that exceed the pixel diagonal  $\sqrt{2}l_{\text{pix}}$  to enable a numerically (but not necessarily physically) meaningful application of the method. If this is not satisfied, the analysis is stopped to modify the grid of aperture sizes to increase  $l_{\text{ap}}$ . This can happen if  $N_{\text{samp}} < \sqrt{2}$  and the dynamic range of  $l_{\text{ap}}$  is too small, or if the original pixel size was too large to enable regridding.

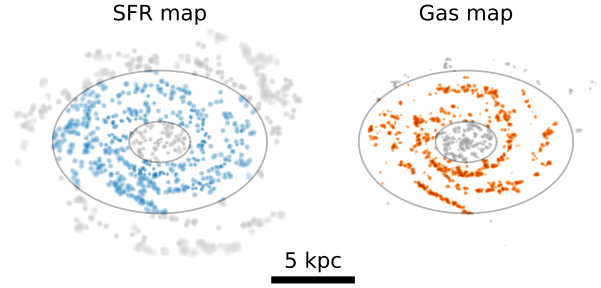
(iv) If the smallest aperture size  $l_{\text{ap},\text{min}}$  is smaller than the pixel diagonal,  $l_{\text{ap},\text{min}}$  is changed to the first element of  $l_{\text{ap}}$  that does exceed the pixel diagonal. This can only happen if  $N_{\text{samp}} < \sqrt{2}$ , or if the original pixel size was too large to enable regridding. Increasing  $l_{\text{ap},\text{min}}$  ensures that all aperture sizes can be assigned accurate enclosed fluxes.

(v) Finally, we multiply the conversion factors  $X_{\text{star}}$  and  $X_{\text{gas}}$  that translate the pixel value to an absolute SFR in  $M_{\odot} \text{ yr}^{-1}$  and an absolute gas mass in  $M_{\odot}$  by a factor of  $(l_{\text{pix}}/l_{\text{pix},\text{orig}})^2$  (with  $l_{\text{pix},\text{orig}}$  the original pixel size) to account for the fact that the pixel size was changed by regridding the maps. This step is required because the regridding process treats the pixel values as surface densities and thus does not conserve the sum of all pixel values.

### 3.2.4 Applying the masks

Having regridded the maps on to an optimal pixel grid and after redefining several of the related quantities, the masks applied to each of the individual maps  $\xi_{\text{star},ij}$  and  $\xi_{\text{gas},ij}$  in Section 3.2.2 must be synchronised and applied uniformly to all maps. In addition, any desired cuts in galactocentric radius should be made here. Before proceeding, it should first be verified that the pixel grid of the maps is the same to within the specified tolerance angle  $\theta_{\text{ast}}$ . We carry out the following checks.

(i) All maps must have identical dimensions  $\{N_{\text{pix},x}, N_{\text{pix},y}\}$ .



**Figure 5.** Example of applying a galactocentric radius mask to the simulated galaxy maps from Figure 4. Only radii  $2 < R/\text{kpc} < 7$  are included and the masked pixels are shown in greyscale. For the remainder of Section 3.2, any further illustrations of the method will be restricted to this radial interval.

(ii) All maps must have identical equinoxes (e.g. J2000).

(iii) Any differences in pixel size  $\theta_{\text{pix}}$  between the maps must be within the specified tolerance, i.e.  $|\theta_{\text{pix},\text{star}} - \theta_{\text{pix},\text{gas}}| \leq \theta_{\text{ast}}$ .

(iv) To definitively exclude any discrepancies between the maps, we carry out the brute-force test of checking whether all pixels in the map have the same right ascension  $\alpha$  and declination  $\delta$  within the specified tolerance, i.e.  $|\alpha_{\text{star}} - \alpha_{\text{gas}}| \leq \theta_{\text{ast}}$  and  $|\delta_{\text{star}} - \delta_{\text{gas}}| \leq \theta_{\text{ast}}$ .

If any of the above conditions is not satisfied, we abort the analysis. In such a case, the regridding should be redone and verified.

The individually-masked maps  $\xi_{\text{star},ij}$  and  $\xi_{\text{gas},ij}$  are then used to create a joint mask  $\Xi_{ij}$  that is applied to all maps. If any pixel has a NaN value in any of the maps, that pixel value is set to NaN across all maps. We also create a mask array  $\Xi_{ij}^0$  in which the masked pixels have a value of zero and the unmasked pixels have a value of unity. This array enables the straightforward calculation of effective aperture sizes in the later steps of the analysis, or the calculation of quantities that are integrated over the entire map. In addition, we perform any radial cuts specified in the input file, accounting for the position angle and inclination of the host galaxy. For each pixel, we calculate the  $X_{\text{pix}}$  and  $Y_{\text{pix}}$  distances to the galactic centre (which is taken to correspond to the central pixel of the image if `set_centre = 0`) as

$$X_{\text{pix}} = D \tan(\theta_{\text{pix}}) [\cos(\phi)(i_{\text{pix}} - i_{\text{cen}}) - \sin(-\phi)(j_{\text{pix}} - j_{\text{cen}})], \quad (7)$$

and

$$Y_{\text{pix}} = D \frac{\tan(\theta_{\text{pix}})}{\cos(i)} [\sin(-\phi)(i_{\text{pix}} - i_{\text{cen}}) + \cos(\phi)(j_{\text{pix}} - j_{\text{cen}})], \quad (8)$$

for a position angle  $\phi$ . The galactocentric radius of each pixel is

$$R_{\text{pix}} = \sqrt{X_{\text{pix}}^2 + Y_{\text{pix}}^2}. \quad (9)$$

We remind the reader that the units of  $i_{\text{pix}}$ ,  $i_{\text{cen}}$ ,  $j_{\text{pix}}$ , and  $j_{\text{cen}}$  are in units of pixels, whereas  $X_{\text{pix}}$  and  $Y_{\text{pix}}$  share the units of  $D$  and represent distances in pc. With the galactocentric radii of the pixels known, we set the values of the pixels across all maps to NaN if they satisfy  $R_{\text{pix}} < R_{\text{min}}$  or  $R_{\text{pix}} > R_{\text{max}}$ . Figure 5 illustrates the application of a mask that selects galactocentric radii  $2 < R/\text{kpc} < 7$  for the example maps considered in this section.

After including any cuts in galactocentric radius in the total mask arrays  $\Xi_{ij}$  and  $\Xi_{ij}^0$ , the mask is applied to all maps and the

updated FITS files are written to disk for later use. A number of important quantities describing the maps are calculated as follows. The total area  $A_{\text{tot}}$  of the unmasked pixels in the map is

$$A_{\text{tot}} = l_{\text{pix}}^2 \sum_{i,j} \Xi_{ij}^0. \quad (10)$$

The total stellar and gas flux in the map ( $\mathcal{F}_{\text{star,tot}}$  and  $\mathcal{F}_{\text{gas,tot}}$ , respectively) are given by summation over all unmasked pixels:

$$\mathcal{F}_{\text{star,tot}} = l_{\text{pix}}^2 \sum_{i,j} \mathcal{F}_{\text{star},ij} \Xi_{ij}^0, \quad (11)$$

and

$$\mathcal{F}_{\text{gas,tot}} = l_{\text{pix}}^2 \sum_{i,j} \mathcal{F}_{\text{gas},ij} \Xi_{ij}^0. \quad (12)$$

Likewise, we determine the total SFR and gas mass  $M_{\text{gas}}$  in the unmasked part of the map by summation over all unmasked pixels:

$$\text{SFR} = X_{\text{star}} \sum_{i,j} \mathcal{F}_{\text{star},ij} \Xi_{ij}^0, \quad (13)$$

and

$$M_{\text{gas}} = X_{\text{gas}} \sum_{i,j} \mathcal{F}_{\text{gas},ij} \Xi_{ij}^0. \quad (14)$$

The uncertainties on SFR and  $M_{\text{gas}}$  (including the statistical uncertainty) is captured in those of the conversion factors, i.e.

$$\sigma(\text{SFR}) = \sigma_{\text{rel}}(X_{\text{star}}) \text{SFR}, \quad (15)$$

and

$$\sigma(M_{\text{gas}}) = \sigma_{\text{rel}}(X_{\text{gas}}) M_{\text{gas}}. \quad (16)$$

These quantities represent the most important macroscopic properties of the maps. Given that we will not use the original maps at any point of the analysis below, the term ‘map’ from hereon refers to the unmasked part of a map, i.e. where  $\Xi_{ij} = 1$ . This excludes any masked pixels, unless specified otherwise.

### 3.2.5 Convolution of the maps

The next step is to create the maps in which the emission peaks will be identified and to generate the convolved maps for determining the stellar and gas flux around the peak positions. For consistency with the minimum aperture size  $l_{\text{ap,min}}$  considered in the analysis, we do not identify the peaks in the original maps, but in maps that have been convolved to a resolution with an FWHM equal to  $l_{\text{ap,min}}$  (as in Figure 5). If no separate maps are specified for this purpose, the peaks will simply be identified in the regridded maps and no convolution is needed, because the maps used as input for the regridding were already convolved to a PSF size equal to  $l_{\text{ap,min}}$ . However, if the peak identification does use separate maps, we convolve them with a two-dimensional Gaussian PSF with FWHM  $(l_{\text{ap,min}}^2 - l_{\text{PSF}}^2)^{1/2}$ , analogous to convolution carried out prior to the regridding in Section 3.2.2.

As described in Section 3.1, the calculation of the flux around emission peaks is performed by convolution of the maps with a kernel of the desired size and shape. Various papers in the literature have used Gaussian kernels for this purpose, with a final FWHM equal to the size scale of interest (e.g. Bigiel et al. 2008; Liu et al. 2011; Leroy et al. 2013, 2016). While we have included the Gaussian kernel as an option in HEISENBERG, we stress that this is not the optimal choice for the problem at hand, because it includes flux

beyond the target radius and emphasizes emission close to the aperture centre. When aiming to obtain an unbiased measurement of the total flux within some radius of a given position in the map, this is achieved by convolution with a two-dimensional tophat kernel:

$$W(r, h) = \begin{cases} 1 & \text{if } r \leq h \\ 0 & \text{if } r > h. \end{cases} \quad (17)$$

To define an aperture of size (i.e. diameter)  $l_{\text{ap}}$ , we set  $h = l_{\text{ap}}/2$ . However, the numerical application of this kernel is carried out on the pixel grid, which means that the effective aperture radius in units of pixels becomes

$$\frac{h}{\text{pix}} = \frac{l_{\text{ap}}}{2l_{\text{pix}}} = \frac{l_{\text{ap}} \sqrt{\cos i}}{D \tan \theta_{\text{pix}}}. \quad (18)$$

This expression accounts for the fact that the effective pixel size increases with the inclination, which results in an aperture size in units of the number of pixels that decreases with the inclination.

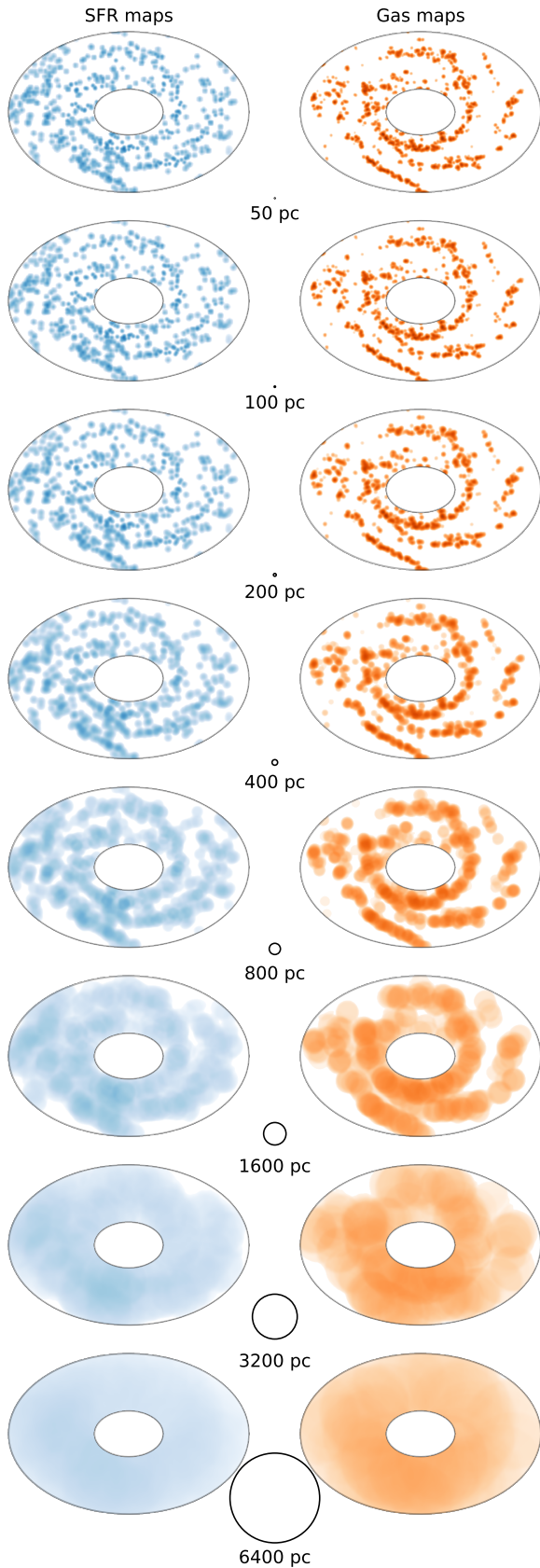
While the functional form of equation (17) is very simple, its numerical application on a discretised pixel grid is more involved. If we define  $W_{ij}$  as the kernel value of a pixel with coordinates  $\{i, j\}$ , then we trivially obtain  $W_{ij} = 1$  and  $W_{ij} = 0$  for any pixels that fall entirely within or outside the circular aperture boundary defined by  $r = h$ , respectively. However, the aperture boundary also intersects with a significant number of pixels, for which the kernel value should become a real number between zero and unity, i.e.  $\{W_{ij} \in \mathbb{R} \mid 0 \leq W_{ij} \leq 1\}$ . For each intersected pixel, we decompose the geometry of the intersection with the aperture into a combination of squares, rectangles, triangles, and/or circular segments, which can then be added or subtracted to analytically calculate the exact fraction of the pixel area  $f_{\text{encl}}$  that is enclosed by the aperture. The details of this calculation are provided in Appendix A. By setting  $W_{ij} = f_{\text{encl}}$  for any pixels that are intersected by the aperture, we obtain a tophat kernel that achieves flux conservation at machine precision when convolved with the maps.

Having defined the convolution kernel, we obtain convolved stellar and gas maps for each aperture size, i.e.  $\mathcal{F}_{\text{star},ij}(l_{\text{ap}})$  and  $\mathcal{F}_{\text{gas},ij}(l_{\text{ap}})$ , as well as convolved mask arrays  $\Xi_{ij}^0(l_{\text{ap}})$ . Note that we do not subtract the PSF size from the kernel size as we did previously when regridding the maps and convolving them to a common Gaussian FWHM (see Section 3.2.2), because the goal of the current convolution is to obtain the exact flux within the specified aperture radius. If the target aperture size  $l_{\text{ap}} \leq 1.05 l_{\text{PSF}}$ , HEISENBERG returns a warning to notify the user that flux is being measured in an aperture smaller than the PSF, but proceeds none the less. The result of convolving the maps with tophat kernels of sizes  $l_{\text{ap}}(i)$  as defined in equation (3) is shown in Figure 6 for the example pair of maps that is used throughout this section to illustrate the method. The sequence of maps nicely visualises how each pixel is assigned the mean flux surface density in a circular aperture centred on that position. While this does not noticeably affect the maps for aperture sizes  $l_{\text{ap}} \leq 100$  pc, it leads to the appearance of disc-like features on intermediate ( $l_{\text{ap}} = \{800, 1600\}$  pc) scales. Towards the largest aperture sizes, the maps become homogeneous and the ‘galactic average’ introduced in Figure 2 is retrieved.

### 3.2.6 Sensitivity limits

In order to enable a reliable peak identification and to quantify the uncertainties on the observed flux ratios, the sensitivity limits of the maps are determined next. Ideally, this should be done by fitting a one-dimensional Gaussian to the noise-dominated range of the pixel value PDF. However, it is non-trivial to estimate which range





**Figure 6.** Application of tophat convolution with different aperture sizes  $l_{\text{ap}}(i)$  to the pair of maps from Figure 5. The aperture sizes are shown in the middle of each row. The sequence of maps illustrates how each pixel value represents the mean flux surface density in a surrounding aperture.

of pixel values are affected by noise without having a measurement of the sensitivity limit. For radio or sub-mm interferometric data, noise manifests itself as a normally-distributed, random component in the pixel value PDF that is symmetric around the background flux (which we here assume to be zero). For optical data, noise results from low photon counts and manifests itself as a Poisson distribution at pixel values above the background flux. These two cases require a different range of pixel values over which the sensitivity limit is characterized.

The range over which a Gaussian is fitted to the pixel value PDF is taken to be

$$\min(\mathcal{F}_{ij}) \leq \mathcal{F}_{ij} \leq \mathcal{F}_{\text{ref}}, \quad (19)$$

where  $\mathcal{F}_{ij}$  represents the array of all pixel values, following the notation of Section 3.2.4, and we define

$$\mathcal{F}_{\text{ref}} = \begin{cases} |\min(\mathcal{F}_{ij})| & \text{if } \min(\mathcal{F}_{ij}) < 0 \\ \text{median}(\mathcal{F}_{ij}) & \text{if } \min(\mathcal{F}_{ij}) \geq 0. \end{cases} \quad (20)$$

The first of these two cases would typically apply to an interferometric image, whereas the second would typically apply to an optical image or simulated map, provided that the noise follows the pattern described above.

To carry out the fits and measure the sensitivity limits, the pixel value PDF of each image is binned over the above range, with a bin width chosen to enable an accurate Gaussian fit. We first determine the standard deviation of  $\mathcal{F}_{ij}$  in the to-be-fitted range as

$$\sigma_{\mathcal{F}} = \sqrt{\langle \mathcal{F}_{ij}^2 \rangle - \langle \mathcal{F}_{ij} \rangle^2} \quad \text{for } \min(\mathcal{F}_{ij}) \leq \mathcal{F}_{ij} \leq \mathcal{F}_{\text{ref}}, \quad (21)$$

and define the bin width as

$$\delta_{\mathcal{F}} = \frac{\sigma_{\mathcal{F}}}{N_{\text{bins}}}. \quad (22)$$

After defining the bins, we generate the corresponding pixel value PDF and carry out a least-square fit of a one-dimensional Gaussian  $p(\mathcal{F})$  to the distribution. The sensitivity limits  $\sigma_{\mathcal{F},\text{star}}$  and  $\sigma_{\mathcal{F},\text{gas}}$  of the stellar and gas map, respectively, are then defined as the dispersion of the corresponding Gaussian fit. Likewise, the background flux levels  $\mathcal{F}_{\text{back,star}}$  and  $\mathcal{F}_{\text{back,gas}}$  of the stellar and gas map, respectively, are defined as the peak flux of each corresponding Gaussian fit.

While it is appealing to handle the sensitivity calculation automatically through the Gaussian fitting process described here, this represents a somewhat idealised situation that is not always achievable. For instance, the maps at hand may be signal masked or filtered to remove the diffuse emission (see Section 4.3.9). In such cases, the noise properties have been modified and the sensitivity limit cannot be obtained by fitting a Gaussian to the pixel value PDF. The sensitivity limit should then be measured by hand using an unprocessed map and be provided to HEISENBERG.

### 3.2.7 Peak identification

To identify emission peaks in each of the maps, we use the two-dimensional version of the publicly available CLUMPFIND algorithm (Williams et al. 1994). As mentioned in Section 3.1, there exist more sophisticated identification methods, but the main goal in this step of the analysis is to obtain accurate peak coordinates, for which CLUMPFIND is entirely adequate. The peak identification is performed using the high-resolution, masked maps (see the top panels of Figure 6), unless separate maps have been provided for this purpose (see Section 3.2.1 and Table 2), in which case these separate maps will also have been masked (see Section 3.2.4).

In brief, the CLUMPFIND algorithm loops over a predefined set of flux levels, starting at the highest level, and identifies closed contours for each of them. For the default choice of `flux.weight = 0` (see Table 1), the pixel with the highest flux value within the closed contour is taken to represent the peak position  $\{i, j\}_{\text{peak}}$ <sup>15</sup>. If this position has not previously been identified (i.e. it is absent at higher flux levels), it is saved as a new peak. The lowest flux level in the set is only used to grow previously-defined peaks to their final sizes and no new peaks are assigned. The total flux of each peak  $\mathcal{F}_{\text{peak}}$  is taken to be the sum of all pixel values within its lowest contour. If the lowest contour encloses more than one peak, the pixels are assigned to individual peaks within the contour using a friend-of-friends algorithm. For further details, we refer the reader to Williams et al. (1994) and the CLUMPFIND website (see Section 3.1).

To apply CLUMPFIND within HEISENBERG to identify the emission peaks in the maps, we define the parameters listed in the third block of Table 2. First, we define the contours that are used by CLUMPFIND to perform the peak identification. For the default choice of using logarithmic spacing, we define the levels as

$$\log_{10} \mathcal{F}_{\text{lev}}(i) = \log_{10} \mathcal{F}_{\text{lev,max}} - \frac{N_{\text{lev}} - 1 - i}{N_{\text{lev}} - 1} \Delta \log_{10} \mathcal{F}, \quad (23)$$

where  $\{i \in \mathbb{N} \mid 0 \leq i \leq N_{\text{lev}} - 1\}$ . In this expression, the maximum level is defined as

$$\log_{10} \mathcal{F}_{\text{lev,max}} = \left\lfloor \left[ \frac{\log_{10} \max(\mathcal{F}_{ij})}{\delta \log_{10} \mathcal{F}} \right] - 1 \right\rfloor \delta \log_{10} \mathcal{F}, \quad (24)$$

which rounds down the maximum pixel value in the map to the nearest multiple of the level interval and subtracts another level interval. This latter step is made to ensure that the top level in which peaks are identified is sufficiently wide to encompass a significant fraction of the flux envelope around the brightest peaks and avoid fragmentation of these peaks. The number of levels follows as

$$N_{\text{lev}} = \frac{\Delta \log_{10} \mathcal{F}}{\delta \log_{10} \mathcal{F}} + 1. \quad (25)$$

These definitions imply that equation (23) covers a total logarithmic range equal to  $\Delta \log_{10} \mathcal{F}$ , with the highest level given by equation (24) and using intervals between the levels of  $\delta \log_{10} \mathcal{F}$ . Finally, if linearly-spaced contours are preferred, the levels are defined as

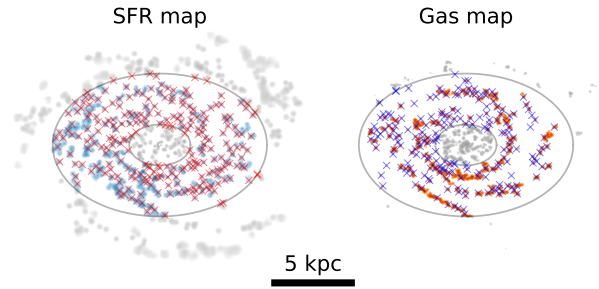
$$\mathcal{F}_{\text{lev}}(i) = \min(\mathcal{F}_{ij}) + \frac{i}{N_{\text{lin}} - 1} [\max(\mathcal{F}_{ij}) - \min(\mathcal{F}_{ij})], \quad (26)$$

where  $\{i \in \mathbb{N} \mid 0 \leq i \leq N_{\text{lin}} - 1\}$ . Throughout the present paper, we will use the logarithmic spacing of equation (23), but the linear spacing of equation (26) is available as an option for future work.

For both maps, we run CLUMPFIND at the highest working resolution ( $l_{\text{ap,min}}$ ) with the specified flux levels. Any identified peaks are included if they satisfy the following two conditions.

- (i) The total number of pixels assigned to the peak (i.e. the peak area in pixels) should be at least  $N_{\text{pix,min}}$  (defined in Table 2).
- (ii) The total flux of the peak should exceed  $N_{\sigma} \sigma_{\mathcal{F}} + \mathcal{F}_{\text{back}}$ , i.e. it must be detected against the background flux with a total signal-to-noise ratio of at least  $N_{\sigma}$  (defined in Table 2).

<sup>15</sup> If `flux.weight = 1`, then the peak position is taken to be the flux-weighted mean of all pixels associated with the peak. This option is available in HEISENBERG (see Table 1), but is not used in the present paper.



**Figure 7.** Outcome of the peak identification process for the example maps used throughout this section. The peak positions are indicated by red (left) and blue (right) crosses. Most of the peaks that one would visually identify are also identified by CLUMPFIND.

Applying the above procedure to each of the maps results in two final samples of reliable stellar peaks and gas peaks, from which the peak positions  $\{i, j\}_{\text{peak}}$  will be used in the rest of the analysis. This provides a total of  $N_{\text{peak}}$  peaks, with  $N_{\text{peak,star}}$  stellar ones and  $N_{\text{peak,gas}}$  gas peaks. Finally, HEISENBERG outputs images of the two maps on which the positions of the peaks are highlighted.

An outcome of the peak identification process is shown in Figure 7 for the standard example maps. The figure nicely illustrates that the default input parameters result in an accurate sample of peaks identified by CLUMPFIND. Applications of the method to other maps may require other values of these parameters, depending on the dynamic range of the pixel values, the relative degree of blending between peaks, the sensitivity limits of the maps, and the typical peak area in pixels. Because there is no quantitative rule for setting the input parameters that govern the peak identification, visual inspection of the output maps is required to verify if the outcome is satisfactory. It is important to reiterate that, if the identified peak sample is incomplete or biased to a certain part of the sample, the evolutionary timeline of Figure 1 that is constrained by the method is not necessarily incorrect, because the method is empirical in nature – it applies to whatever subset of the peak sample is selected, as long as it is representative for the flux-emitting regions in the maps. The only way in which the outcome of the method can be truly incorrect, is if a significant fraction of the total flux in the map is inaccessible by the peak identification process. This occurs if a map contains a large reservoir of diffuse emission (Hygate et al. 2018), or if the depth of the peak identification ( $\Delta \log_{10} \mathcal{F}$ ) is chosen to be too small to reach the peaks that contain most of the flux. As long as these situations are avoided, we find that the details of the peak identification process do not strongly affect the outcome of the method, implying that the derived timeline is robust.

### 3.2.8 Aperture placement on peaks

With the peak positions in hand, we calculate the total fluxes in circular apertures centred on these peaks, as well as the effective aperture areas and peak-to-peak distances. To quantify the flux within an aperture, we prefer to use the flux density in that aperture, which is equivalent to the pixel value at the peak position in the map convolved to the desired aperture size (see Section 3.2.5), rather than the absolute flux within the aperture. To obtain an absolute flux rather than a flux density, the latter quantity should be multiplied by  $\pi l_{\text{ap}}^2/4$ . However, the remainder of the presented analysis deals with flux ratios at each aperture size, implying that this geometric factor cancels. We therefore use  $\mathcal{F}_{\text{star},ij}(l_{\text{ap}}, \{i, j\}_{\text{peak}})$  and

$\mathcal{F}_{\text{gas},ij}(l_{\text{ap}}, \{i, j\}_{\text{peak}})$  to describe the flux within each aperture, where the terms in parentheses indicate that the flux is evaluated at each peak position  $\{i, j\}_{\text{peak}}$  in a map convolved to an aperture size  $l_{\text{ap}}$ .

Even though the aperture size is formally defined by equation (3), each aperture focused on a particular peak has an effective aperture size  $l_{\text{ap,eff}} \leq l_{\text{ap}}$  because it may contain masked pixels. To account for this effect throughout the analysis, we calculate the included fraction of the intended aperture area in each case as

$$f_{A,ij}(l_{\text{ap}}, \{i, j\}_{\text{peak}}) = \Xi_{ij}^0(l_{\text{ap}}, \{i, j\}_{\text{peak}}). \quad (27)$$

This expression is so simple, because the pixel values in the convolved mask array  $\Xi_{ij}^0(l_{\text{ap}})$  (in which the masked pixels have been set to zero) are equivalent to the unmasked fraction of the aperture area at each position. The effective aperture size then follows as

$$l_{\text{ap,eff}} = \sqrt{\langle f_{A,ij} \rangle} l_{\text{ap}}, \quad (28)$$

where  $\langle f_{A,ij} \rangle$  takes the ensemble average over all peaks and is placed within the square-root, because the flux is proportional to the aperture area rather than its size. This quantity will be used in Section 3.2.9 to place each measurement of the flux ratio excess or deficit at the correct effective aperture size in the tuning fork diagram of Figure 2.

Finally, we calculate the distances between all peak pairs as

$$d_{kl} = l_{\text{pix}} |\{i, j\}_{\text{peak},k} - \{i, j\}_{\text{peak},l}|, \quad (29)$$

where  $\{k, l \in \mathbb{N} \mid 0 \leq k, l \leq N_{\text{peak}} - 1\}$  represent the indices used to select elements from the list of peaks. This distance is used in Section 3.2.9 to draw Monte-Carlo realisations of peak samples with non-overlapping apertures and ensure that no pixels are counted twice in the flux ratio calculation.

### 3.2.9 Gas-to-stellar flux ratio bias around peaks

At this point in the analysis, all necessary flux levels have been determined. It is now possible to turn these fluxes into a gas-to-stellar flux ratio ‘bias’ (i.e. excess or deficit) relative to the galactic average as a function of the aperture size (see the cartoon of the tuning fork diagram in Figure 2). This bias is defined as the ratio of two gas-to-stellar flux ratios, i.e.

$$\mathcal{B}_{\text{star}}(l_{\text{ap}}) = \frac{\mathcal{R}_{\text{star}}(l_{\text{ap}})}{\mathcal{R}_{\text{tot}}}, \quad (30)$$

and

$$\mathcal{B}_{\text{gas}}(l_{\text{ap}}) = \frac{\mathcal{R}_{\text{gas}}(l_{\text{ap}})}{\mathcal{R}_{\text{tot}}}, \quad (31)$$

for the bias towards stellar and gas peaks, which represent the bottom and top branches of the tuning fork diagram, respectively. The denominators represent the ‘galactic average’ flux ratio, i.e. the total gas-to-stellar flux ratio across the entire unmasked area of the maps, for which the fluxes have been calculated in Section 3.2.4:

$$\mathcal{R}_{\text{tot}} = \frac{\mathcal{F}_{\text{gas,tot}}}{\mathcal{F}_{\text{star,tot}}}. \quad (32)$$

The numerators in equations (30) and (31) represent the gas-to-stellar flux ratio of apertures focused on stellar or gas peaks, respectively. In principle, such a ratio can be determined in two ways that have a different meaning. Firstly, one can obtain the gas-to-stellar flux ratio for each individual peak of a given type, after which the average ratio is taken across all peaks of that type. Secondly, one can calculate the total stellar flux and gas flux across all apertures

focused on a given peak type and then take the ratio between these total aperture fluxes. The model of the KL14 principle follows the latter approach, because it is straightforward to predict analytically and is less sensitive to peak-to-peak variations or correlations between the local stellar and gas flux.

Before being able to provide the expressions for  $\mathcal{R}_{\text{star}}(l_{\text{ap}})$  and  $\mathcal{R}_{\text{gas}}(l_{\text{ap}})$  that are used in the method to represent the observed gas-to-stellar flux ratios, there is an important practical point to consider. Across each of the two peak samples, there will be a minimum distance between all peak pairs  $d_{kl}$ , as defined in equation (29). For aperture sizes  $l_{\text{ap}} \geq \min(d_{kl})$ , this means that at least some apertures will overlap. If the total aperture flux is obtained by summation over all apertures irrespective of their (in)dependence, then pixels could be counted twice. In the best case [ $l_{\text{ap,max}} \sim \min(d_{kl})$ ], this would not significantly affect the observed gas-to-stellar flux ratio bias. However, for  $l_{\text{ap,max}} \gg \min(d_{kl})$ , the observed flux ratio bias would count large numbers of pixels multiple times, particularly in regions of high peak density. This would lead to an inaccurate measure of the true flux ratio bias.

To address this problem, we draw Monte-Carlo realisations of independent peak samples for each peak type and aperture size, which will be averaged over to obtain the gas-to-stellar flux ratio. This is done by drawing candidate peaks in a random order. For each candidate peak, it is verified whether its distance to all previously drawn peaks is larger than the aperture size, i.e.:

$$l_{\text{ap}} < \min(\widehat{d}_{kl}), \quad (33)$$

where the hat indicates that this expression only considers the subset of previously drawn and accepted peaks. If equation (33) is satisfied, then the peak candidate is accepted and added to the Monte-Carlo sample. If the condition is not satisfied, then the peak is erased from the candidate list. This process is repeated until all candidate peaks have been considered. The resulting realisations of the peak samples  $\widehat{\{i, j\}}_{\text{peak,star}}$  and  $\widehat{\{i, j\}}_{\text{peak,gas}}$  with their correspondingly-positioned apertures each only include independent apertures – no pixel is included more than once in each of the samples. For each aperture size  $l_{\text{ap}}$ , we follow the above procedure  $N_{\text{mc,peak}}$  times to generate  $N_{\text{mc,peak}}$  Monte-Carlo realisations of independent stellar and gas peak samples. At small apertures, these samples will hold close to the total number of peaks, because most peaks will have no neighbours closer than the minimum aperture size  $l_{\text{ap,min}}$ . However, at large apertures (e.g.  $l_{\text{ap}} = l_{\text{ap,max}}$ ), the samples may contain only a handful of peaks, because the aperture sizes approach the diameter of the entire unmasked area in the map. The mean number of peaks considered at each aperture size is determined as

$$N_{\text{star,mc}}(l_{\text{ap}}) = \left\langle \widehat{N}_{\text{star}}(l_{\text{ap}}) \right\rangle_{\text{mc}}, \quad (34)$$

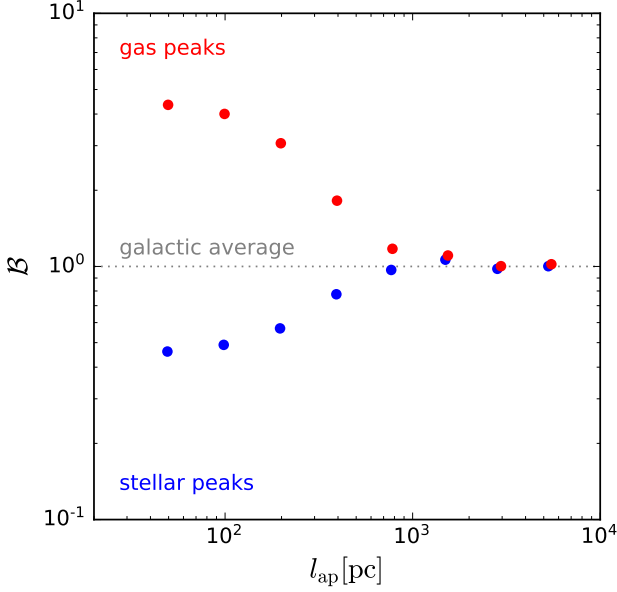
and

$$N_{\text{gas,mc}}(l_{\text{ap}}) = \left\langle \widehat{N}_{\text{gas}}(l_{\text{ap}}) \right\rangle_{\text{mc}}, \quad (35)$$

where the average  $\langle \dots \rangle_{\text{mc}}$  takes place over all Monte-Carlo realisations and the hat indicates that these expressions only consider the subset of accepted peaks.

With the Monte-Carlo realisation of the peak samples in hand, it is straightforward to calculate the gas-to-stellar flux ratios for apertures focused on stellar ( $\mathcal{R}_{\text{star}}$ ) and gas ( $\mathcal{R}_{\text{gas}}$ ) peaks as a func-





**Figure 8.** Example of the gas-to-stellar flux ratio bias as a function of the aperture size  $l_{\text{ap}}$  when focusing apertures on stellar peaks ( $\mathcal{B}_{\text{star}}$ , blue dots) or gas peaks ( $\mathcal{B}_{\text{gas}}$ , red dots) in the maps shown in Figure 7. The dotted line indicates the ‘galactic average’ gas-to-stellar flux ratio. Note the small differences in aperture size between the blue and red symbols towards the right. This reflects slightly different effective aperture areas due to masking, which affects both peak samples differently, because stellar and gas peaks are not co-spatial and thus apertures centred on the two different populations do not necessarily contain the same number of masked pixels (see the text for a detailed explanation). The characteristic ‘tuning fork’ shape of this quantitative application of the method resembles the qualitative behaviour described in Section 2.2 and Figure 2.

tion of the aperture size. We write

$$\mathcal{R}_{\text{star}} = \frac{\mathcal{F}_{\text{gas,star}}}{\mathcal{F}_{\text{star,star}}} = \frac{\left\langle \frac{\sum_{\widehat{\{i,j\}}_{\text{peak,star}}} \mathcal{F}_{\text{gas},ij}(l_{\text{ap}})}{\sum_{\widehat{\{i,j\}}_{\text{peak,star}}} \mathcal{F}_{\text{star},ij}(l_{\text{ap}})} \right\rangle_{\text{mc}}}{\left\langle \frac{\sum_{\widehat{\{i,j\}}_{\text{peak,star}}} \mathcal{F}_{\text{gas},ij}(l_{\text{ap}})}{\sum_{\widehat{\{i,j\}}_{\text{peak,star}}} \mathcal{F}_{\text{star},ij}(l_{\text{ap}})} \right\rangle_{\text{mc}}}, \quad (36)$$

and

$$\mathcal{R}_{\text{gas}} = \frac{\mathcal{F}_{\text{gas,gas}}}{\mathcal{F}_{\text{star,gas}}} = \frac{\left\langle \frac{\sum_{\widehat{\{i,j\}}_{\text{peak,gas}}} \mathcal{F}_{\text{gas},ij}(l_{\text{ap}})}{\sum_{\widehat{\{i,j\}}_{\text{peak,gas}}} \mathcal{F}_{\text{star},ij}(l_{\text{ap}})} \right\rangle_{\text{mc}}}{\left\langle \frac{\sum_{\widehat{\{i,j\}}_{\text{peak,gas}}} \mathcal{F}_{\text{gas},ij}(l_{\text{ap}})}{\sum_{\widehat{\{i,j\}}_{\text{peak,gas}}} \mathcal{F}_{\text{star},ij}(l_{\text{ap}})} \right\rangle_{\text{mc}}}, \quad (37)$$

where the average  $\langle \dots \rangle_{\text{mc}}$  again takes place over all Monte-Carlo realisations and the summation is performed only over the accepted, independent peaks in each realisation, as indicated by the hat in  $\widehat{\{i,j\}}_{\text{peak}}$ . Note that these equations contain four different fluxes due to the fact that both flux types are measured around both peak types. This results in subscripts to the flux that contain two instances of ‘star’ and ‘gas’ (appearing after the first equality in these expressions), where we follow the convention that first word indicates the type of flux and the second word indicates the type of peak that the apertures are focused on. The gas-to-stellar flux ratios  $\mathcal{R}_{\text{star}}$  and  $\mathcal{R}_{\text{gas}}$  are substituted into equations (30) and (31) to obtain the flux ratio bias when focusing apertures on stellar and gas peaks, respectively. Figure 8 shows an example of the resulting gas-to-stellar flux ratio bias for the pair of maps used throughout this section, clearly highlighting how the qualitative behaviour of the tuning fork diagram in Figure 2 is reproduced in the quantitative application of the method.

Before continuing, we emphasize that these flux ratio biases

are not affected by any uncertainties on the conversion factors from a stellar or gas flux to an SFR or gas mass, as long as these factors do not vary significantly across the maps. In that case, the conversion factors on scales of  $l_{\text{ap}} \lesssim \lambda$  are similar to the map average ( $l_{\text{ap}} \sim l_{\text{ap,max}}$ ), implying that the conversion factors cancel when taking the ratio of the two flux ratios in equations (30) and (31).

We also need to re-evaluate the effective aperture size of equation (28) after having generated the Monte-Carlo realisations, because the differing coverages of the masked pixels in each Monte-Carlo realisation of the peak sample lead to different effective aperture sizes around stellar and gas peaks. We calculate these as

$$l_{\text{ap,star}} = \left\langle \left[ \frac{\sum_{\widehat{\{i,j\}}_{\text{peak,star}}} f_{A,ij}(l_{\text{ap}})}{\widehat{N}_{\text{peak,star}}} \right]^{1/2} l_{\text{ap}} \right\rangle_{\text{mc}}. \quad (38)$$

and

$$l_{\text{ap,gas}} = \left\langle \left[ \frac{\sum_{\widehat{\{i,j\}}_{\text{peak,gas}}} f_{A,ij}(l_{\text{ap}})}{\widehat{N}_{\text{peak,gas}}} \right]^{1/2} l_{\text{ap}} \right\rangle_{\text{mc}}. \quad (39)$$

for the aperture sizes for stellar and gas peaks, respectively. In these expressions, the term in square brackets averages the unmasked fraction of the aperture area over all peaks in a Monte-Carlo realisation, after which the square-root and multiplication with  $l_{\text{ap}}$  result in an effective aperture size for that realisation. The average  $\langle \dots \rangle_{\text{mc}}$  then provides the final effective aperture sizes by averaging over all realisations. Note that these revised aperture sizes do not affect the gas-to-stellar flux ratios of equations (36) and (37), because the aperture sizes in the numerator and denominator of each of these expressions are identical. The change only affects the horizontal positions of the data points in Figure 8.

In addition to determining the gas-to-stellar flux ratio bias as a function of size scale around emission peaks, we also use the Monte-Carlo sampling to obtain the necessary constraints on the flux evolution of independent regions. As explained in KL14, regions may undergo any form of flux evolution during the evolutionary timeline of Figure 1. The KL14 principle discretises that timeline into three phases and allows one to measure their relative durations. Because this method is based on flux ratio measurements, it effectively uses the mean flux level integrated over each of these phases. If one is not interested in the overlap phase, but only in determining the ratio between the gas and stellar lifetimes, then the measurement is therefore entirely independent of the flux evolution of each tracer. Only when concerned with the duration of the overlap phase it is necessary to account for time-averaged relative flux level changes of a tracer between the different phases. To capture this evolution, we determine the quantities  $\beta_{\text{star}}$  and  $\beta_{\text{gas}}$ , which refer to the mean flux ratios of regions residing in the overlap phase relative to their ‘isolated’ phases, i.e. outside of the overlap phase. For the purpose of our method, these two parameters fully capture the (possibly complex) flux evolution of both tracers over the evolutionary timeline. Assuming that there exists a way of determining which peaks reside in the overlap and isolated phases,  $\beta_{\text{star}}$  and  $\beta_{\text{gas}}$  can be defined as

$$\beta_{\text{star}} = \left\langle \frac{\widehat{N}_{\text{peak,star,iso}} \sum_{\widehat{\{i,j\}}_{\text{peak,star,over}}} \mathcal{F}_{\text{star},ij}(l_{\text{ap,min}})}{\widehat{N}_{\text{peak,star,over}} \sum_{\widehat{\{i,j\}}_{\text{peak,star,iso}}} \mathcal{F}_{\text{star},ij}(l_{\text{ap,min}})} \right\rangle_{\text{mc}}, \quad (40)$$

and

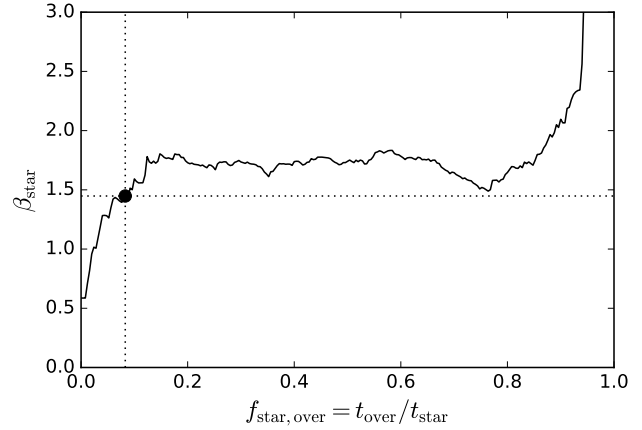
$$\beta_{\text{gas}} = \left\langle \frac{\widehat{N}_{\text{peak,gas,iso}} \sum_{\widehat{\{i,j\}}_{\text{peak,gas,over}}} \mathcal{F}_{\text{gas},ij}(l_{\text{ap,min}})}{\widehat{N}_{\text{peak,gas,over}} \sum_{\widehat{\{i,j\}}_{\text{peak,gas,iso}}} \mathcal{F}_{\text{gas},ij}(l_{\text{ap,min}})} \right\rangle_{\text{mc}}, \quad (41)$$

where  $\widehat{N}_{\text{peak,star,iso}}$ ,  $\widehat{N}_{\text{peak,gas,iso}}$ ,  $\widehat{N}_{\text{peak,star,over}}$ , and  $\widehat{N}_{\text{peak,gas,over}}$  represent the number of peaks in a Monte-Carlo realisation that reside in the isolated stellar, isolated gas, stellar overlap, and gas overlap phases, respectively, and the subscripts to  $\widehat{\{i,j\}}$  refer to the peak coordinates of these same subsets of peaks. Equations (40) and (41) determine the mean flux level within an aperture of size  $l_{\text{ap,min}}$  for isolated peaks and for peaks residing in the overlap phase, take the ratio between both flux levels, and average the resulting ratio over all Monte-Carlo realisations. They thus represent the ‘overlap-to-isolated flux ratios’ of independent regions.

The key question is now how one determines whether a peak resides in the isolated or overlap phase. Without further knowledge regarding the evolutionary timeline of Figure 1, there is no objective way of making that decision. One solution could be to define a critical flux contrast between stellar and gas flux to consider a peak in the isolated or overlap phase. However, this would not only be arbitrary, but most importantly it would also be degenerate with the idea that the duration of the overlap phase can be constrained by measuring flux ratios. Fortunately, there exists an objective way of distinguishing between isolated and overlap-phase peaks if one assumes that the identified peaks homogeneously sample the complete evolutionary timeline and are thus representative for the underlying flux evolution. This is a strong assumption – so strong in fact, that in its most extreme interpretation it would enable one to derive the evolutionary time-scales of Figure 1 by simply counting the identified peaks and using the relative counts as a proxy for the relative time-scales. However, it would be a grave mistake to do so, because it results in a timeline that is entirely dependent on the peak identification process, which (as discussed in Section 3.2.7) relies on by-eye verification and is therefore much too subjective to draw such an immediate quantitative conclusion.

Here, we propose a much weaker version of the above assumption. If the peaks reasonably homogeneously sample the evolutionary timeline of converting one tracer into the next, which happens in the discussed example by converting gas into stars and expelling the residual gas by feedback, then a peak sample that is sorted by decreasing gas-to-stellar flux ratio reflects a trend of increasing evolutionary age. The sorted peak sample can then be roughly divided into peaks residing in the overlap and isolated phases as a function of the evolutionary time-scale ratios  $f_{\text{star,over}} = t_{\text{over}}/t_{\text{star}}$  and  $f_{\text{gas,over}} = t_{\text{over}}/t_{\text{gas}}$ , by requiring that the fraction of peaks residing in a certain phase must be the same as the fraction of the tracer lifetime covered by that phase. Because these time-scales are constrained during the fitting process (see Section 3.2.12), such a dependence would imply functional dependences of  $\beta_{\text{star}}(f_{\text{star,over}})$  and  $\beta_{\text{gas}}(f_{\text{gas,over}})$ , allowing both quantities to be derived as a byproduct of the fitting process. This version of assuming that the identified peaks are representative for the evolutionary timeline is much weaker than in the earlier example of directly using peak counts to constrain the evolutionary timeline, because (as we show in the example below)  $\beta_{\text{star}}$  and  $\beta_{\text{gas}}$  only vary weakly with  $f_{\text{star,over}}$  and  $f_{\text{gas,over}}$ , respectively – typically by a few tens of per cent over factors of several in the underlying time-scales.

Quantitatively, the above method for deriving  $\beta_{\text{star}}(f_{\text{star,over}})$



**Figure 9.** Dependence of the overlap-to-isolated stellar flux ratio  $\beta_{\text{star}}$  on the relative duration of the overlap phase  $f_{\text{star,over}} = t_{\text{over}}/t_{\text{star}}$  for the example maps that are used throughout this section. The flux ratio  $\beta_{\text{star}}$  captures the (possibly complex) flux evolution of the SFR tracer in the presented method. The best fitting value of  $t_{\text{over}}$  and  $t_{\text{star}}$  (see Section 3.2.12) and the corresponding best-fitting value of  $\beta_{\text{star}}$  are shown by the black dot and the dotted lines. In this example,  $\beta_{\text{star}}$  is well-constrained, because it varies only weakly over most of the range of  $f_{\text{star,over}}$ .

and  $\beta_{\text{gas}}(f_{\text{gas,over}})$  is applied as follows. First, the arrays for  $f_{\text{star,over}}$  and  $f_{\text{gas,over}}$  are discretised by writing

$$f_{\text{star,over}}(k) = \frac{k}{\widehat{N}_{\text{peak,star}} - 1}, \quad (42)$$

with  $\{k \in \mathbb{N} \mid 0 \leq k \leq \widehat{N}_{\text{peak,star}} - 1\}$ , and

$$f_{\text{gas,over}}(l) = \frac{l}{\widehat{N}_{\text{peak,gas}} - 1}, \quad (43)$$

with  $\{l \in \mathbb{N} \mid 0 \leq l \leq \widehat{N}_{\text{peak,gas}} - 1\}$ . For each Monte-Carlo realisation and for each  $f_{\text{star,over}}(k)$ :

(i) we assign the  $\widehat{N}_{\text{peak,star,iso}} = k + 1$  stellar peaks that have the lowest  $\mathcal{F}_{\text{gas},ij}(l_{\text{ap,min}})/\mathcal{F}_{\text{star},ij}(l_{\text{ap,min}})$  ratio to the isolated stellar phase, which thus defines  $\widehat{\{i,j\}}_{\text{peak,star,iso}}$ ;

(ii) we assign the  $\widehat{N}_{\text{peak,star,over}} = \widehat{N}_{\text{peak,star}} - k - 1$  stellar peaks that have the highest  $\mathcal{F}_{\text{gas},ij}(l_{\text{ap,min}})/\mathcal{F}_{\text{star},ij}(l_{\text{ap,min}})$  ratio to the overlap phase, which thus defines  $\widehat{\{i,j\}}_{\text{peak,star,over}}$ .

Vice versa, for each Monte-Carlo realisation and for each  $f_{\text{gas,over}}(l)$ :

(i) we assign the  $\widehat{N}_{\text{peak,gas,iso}} = l + 1$  gas peaks that have the lowest  $\mathcal{F}_{\text{star},ij}(l_{\text{ap,min}})/\mathcal{F}_{\text{gas},ij}(l_{\text{ap,min}})$  ratio to the isolated gas phase, which thus defines  $\widehat{\{i,j\}}_{\text{peak,gas,iso}}$ ;

(ii) we assign the  $\widehat{N}_{\text{peak,gas,over}} = \widehat{N}_{\text{peak,gas}} - l - 1$  gas peaks that have the highest  $\mathcal{F}_{\text{star},ij}(l_{\text{ap,min}})/\mathcal{F}_{\text{gas},ij}(l_{\text{ap,min}})$  ratio to the overlap phase, which thus defines  $\widehat{\{i,j\}}_{\text{peak,gas,over}}$ .

This way, we can evaluate equations (40) and (41) and obtain  $\beta_{\text{star}}(f_{\text{star,over}})$  and  $\beta_{\text{gas}}(f_{\text{gas,over}})$  for use in the fitting process of Section 3.2.12. An example of  $\beta_{\text{star}}(f_{\text{star,over}})$  is given in Figure 9, which illustrates the earlier statement that  $\beta$  typically varies weakly with  $f_{\text{over}}$ . In addition to  $\beta_{\text{star}}(f_{\text{star,over}})$ , Figure 9 also shows the best-fitting value to  $\beta_{\text{star}}$  that results from the fitting process in Section 3.2.11 below. When fitting the observed gas-to-stellar flux ratio bias as a function of the aperture size, we obtain the time-scales  $t_{\text{gas}}$  and  $t_{\text{over}}$  for a given  $t_{\text{star}}$ . This immediately

provides  $f_{\text{star,over}}$  and  $f_{\text{gas,over}}$ , implying that  $\beta_{\text{star}}$  and  $\beta_{\text{gas}}$  are byproducts of the fitting process. We demonstrate in Section 4.2.5 that these overlap-to-isolated flux ratios give an accurate representation of the underlying flux evolution.

Finally, we determine a set of quantities making use of the flux distribution in the maps that are used at later points in the analysis. The first of these is the flux density contrast of emission peaks relative to the average flux density on a size scale  $l_{\text{ap}}$  for each tracer. This contrast will be used to determine the region sizes in units of the mean separation length  $\lambda$  (i.e. their filling factor) when fitting the model in Sections 3.2.11 and 3.2.12. Because the region separation length is unknown at this stage, we calculate the flux density contrast relative to all size scales  $l_{\text{ap}}$  and interpolate this later to the desired size scale for measuring the background flux density. Analogously to equations (36) and (37), this contrast can be expressed as

$$\mathcal{E}_{\text{star}}(l_{\text{ap}}) = \frac{\left\langle \sum_{\{i,j\}_{\text{peak,star}}} \mathcal{F}_{\text{star},ij}(l_{\text{ap,min}}) \right\rangle_{\text{mc}}}{\left\langle \sum_{\{i,j\}_{\text{peak,star}}} \mathcal{F}_{\text{star},ij}(l_{\text{ap}}) \right\rangle_{\text{mc}}} \times \frac{N_{\text{star,mc}}(l_{\text{ap}})}{N_{\text{star,mc}}(l_{\text{ap,min}})}, \quad (44)$$

and

$$\mathcal{E}_{\text{gas}}(l_{\text{ap}}) = \frac{\left\langle \sum_{\{i,j\}_{\text{peak,gas}}} \mathcal{F}_{\text{gas},ij}(l_{\text{ap,min}}) \right\rangle_{\text{mc}}}{\left\langle \sum_{\{i,j\}_{\text{peak,gas}}} \mathcal{F}_{\text{gas},ij}(l_{\text{ap}}) \right\rangle_{\text{mc}}} \times \frac{N_{\text{gas,mc}}(l_{\text{ap}})}{N_{\text{gas,mc}}(l_{\text{ap,min}})}. \quad (45)$$

These expressions effectively yield the mean flux density (averaged over all Monte-Carlo realisations and all peaks) at the smallest aperture size in units of the mean flux density at larger aperture sizes. As stated above, the resulting flux density contrasts of single peaks relative to the background population will be used when fitting the model in Sections 3.2.11 and 3.2.12.

Secondly, we calculate a quantity that is similar to the flux density contrasts of equations (44) and (45) above, but instead measures the flux density contrast between the flux on some size scale  $l_{\text{ap}}$  and the entire map. This ‘global’ flux density contrast can be used in observational applications of the method to account for large-scale morphological features of the target system (or the spatial clustering of independent regions) when converting map-averaged flux densities to the mean flux density per independent region, which in turn can be useful for deriving a variety of physical quantities (see e.g. Section 4.3.3). As before, this requires us to calculate the flux density contrast as a function of size scale  $l_{\text{ap}}$  and interpolate this to the mean separation length  $\lambda$  once it has been obtained from the fitting process in Section 3.2.12. The contrast is expressed as

$$\mathcal{E}_{\text{star,glob}}(l_{\text{ap}}) = \frac{\left\langle \sum_{\{i,j\}_{\text{peak,star}}} \mathcal{F}_{\text{star},ij}(l_{\text{ap}}) \right\rangle_{\text{mc}}}{N_{\text{star,mc}}(l_{\text{ap}})} \frac{A_{\text{tot}}}{\mathcal{F}_{\text{star,tot}}}, \quad (46)$$

and

$$\mathcal{E}_{\text{gas,glob}}(l_{\text{ap}}) = \frac{\left\langle \sum_{\{i,j\}_{\text{peak,gas}}} \mathcal{F}_{\text{gas},ij}(l_{\text{ap}}) \right\rangle_{\text{mc}}}{N_{\text{gas,mc}}(l_{\text{ap}})} \frac{A_{\text{tot}}}{\mathcal{F}_{\text{gas,tot}}}, \quad (47)$$

These expressions effectively yield the mean flux density (averaged over all Monte-Carlo realisations and all peaks) at an aperture size

$l_{\text{ap}}$  in units of the mean flux density of the entire map. As stated above, an example of using the resulting flux density contrasts is provided in Section 4.3.3.

### 3.2.10 Determination of uncertainties

Section 3.2.9 presents a derivation of the gas-to-stellar flux ratio bias when focusing on stellar and gas peaks, but these are meaningless without an estimate of the uncertainties (cf. Figure 8). The data points also require error bars in order to constrain our best-fitting statistical model (see Section 3.2.11) to the data in Section 3.2.12. We now describe how the uncertainties are quantified. This includes a discussion of how we account for the covariance between the data points in the tuning fork diagram – after all, apertures of different sizes are focused on the same peaks, implying that they are at least partially sharing the same information.

We start by identifying the main sources of uncertainty in the data points of equations (36) and (37), which are shown in Figure 8.

(i) The stellar and gas fluxes used in each of the data points are uncertain due to the finite signal-to-noise level of the data. To include this uncertainty, we use the measurements of the sensitivity limits  $\sigma_{\mathcal{F},\text{star}}$  and  $\sigma_{\mathcal{F},\text{gas}}$  in Section 3.2.6.

(ii) The peak flux levels exhibit an intrinsic variance because they do not all represent the same SFR or mass. This variance arises because the region mass function is not a delta function, but also because the individual regions may undergo some flux evolution, as discussed in Section 3.2.9. It leads to an uncertainty on the derived gas-to-stellar flux ratio biases, because we have used a finite number of peaks to determine these biases. For an infinite number of regions, this uncertainty would approach zero.

(iii) Finally, we should account for the covariance between the numerator and denominator of the gas-to-stellar flux ratios in equations (36) and (37). It is possible that bright peaks from both populations are in close proximity to one another, either because of the evolutionary connection between both peak populations (e.g. a bright stellar region emerging from a bright gas region), or because there may exist a systematic environmental variation of peak brightnesses in the considered maps. These effects are captured by subtracting the covariance between the stellar and gas flux from the total uncertainty.

Mathematically, we account for the above sources of uncertainty as follows. We first assume that the uncertainties on the aperture flux ratios are much larger than those on the galactic average flux ratios. This allows us to set the relative uncertainties of the flux ratio biases equal to the relative uncertainties of the gas-to-stellar flux ratios around stellar and gas peaks, i.e.

$$\frac{\sigma_{\mathcal{B},\text{star}}(l_{\text{ap,star}})}{\mathcal{B}_{\text{star}}(l_{\text{ap,star}})} = \frac{\sigma_{\mathcal{R},\text{star}}(l_{\text{ap,star}})}{\mathcal{R}_{\text{star}}(l_{\text{ap,star}})}, \quad (48)$$

and

$$\frac{\sigma_{\mathcal{B},\text{gas}}(l_{\text{ap,gas}})}{\mathcal{B}_{\text{gas}}(l_{\text{ap,gas}})} = \frac{\sigma_{\mathcal{R},\text{gas}}(l_{\text{ap,gas}})}{\mathcal{R}_{\text{gas}}(l_{\text{ap,gas}})}. \quad (49)$$

The relative uncertainties of the flux ratios are then defined as

$$\frac{\sigma_{\mathcal{R},\text{star}}(l_{\text{ap,star}})}{\mathcal{R}_{\text{star}}(l_{\text{ap,star}})} = \left( \sigma_{\text{sens,star}}^2 + \sigma_{\text{star,star}}^2 + \sigma_{\text{gas,star}}^2 - 2\sigma_{\text{star,star}}\sigma_{\text{gas,star}}\rho_{\text{star}} \right)^{1/2}, \quad (50)$$

and

$$\frac{\sigma_{\mathcal{R},\text{gas}}(l_{\text{ap},\text{gas}})}{\mathcal{R}_{\text{gas}}(l_{\text{ap},\text{gas}})} = \left( \sigma_{\text{sens,gas}}^2 + \sigma_{\text{star,gas}}^2 + \sigma_{\text{gas,gas}}^2 - 2\sigma_{\text{star,gas}}\sigma_{\text{gas,gas}}\rho_{\text{gas}} \right)^{1/2}, \quad (51)$$

respectively, where each of the uncertainties in parentheses represent relative uncertainties and the negative terms represent the covariance between the stellar and gas flux in the apertures. Each of these terms is defined below. As before, in the subscripts containing a comma, the word after the comma indicates the peak type that the apertures are focused on. The first term represents the uncertainty due to the finite signal-to-noise level (or non-zero sensitivity), the second term represents the uncertainty on the stellar flux, the third term represents the uncertainty on the gas flux, and the fourth term represents the covariance of the second and third terms. Finally, all terms have a dependence on the aperture size, but these dependences are omitted in the notation for brevity.

For the uncertainties due to the finite signal-to-noise level, we write

$$\sigma_{\text{sens,star}}^2 = \left[ \left( \frac{\sigma_{\mathcal{F},\text{star}}}{\mathcal{F}_{\text{star,star}}} \right)^2 + \left( \frac{\sigma_{\mathcal{F},\text{gas}}}{\mathcal{F}_{\text{gas,star}}} \right)^2 \right] \times \left( \frac{l_{\text{ap,star}}}{l_{\text{ap,min}}} \right)^{-2} N_{\text{star,mc}}, \quad (52)$$

and

$$\sigma_{\text{sens,gas}}^2 = \left[ \left( \frac{\sigma_{\mathcal{F},\text{star}}}{\mathcal{F}_{\text{star,gas}}} \right)^2 + \left( \frac{\sigma_{\mathcal{F},\text{gas}}}{\mathcal{F}_{\text{gas,gas}}} \right)^2 \right] \times \left( \frac{l_{\text{ap,gas}}}{l_{\text{ap,min}}} \right)^{-2} N_{\text{gas,mc}}, \quad (53)$$

for apertures focused on stellar and gas peaks, respectively. As in equations (36) and (37), the double subscripts of the total fluxes follow the order of indicating the flux type before the peak type. The relative uncertainty on the total flux due to noise in these expressions follows normal statistics. Firstly, as mentioned in Sections 3.2.3 and 3.2.5, the pixel flux values represent surface densities, implying that their uncertainty due to noise should decrease with the square-root of the number of pixels in the aperture  $N_{\text{pix,ap}}$  and thus  $\sigma^2 \propto N_{\text{pix,ap}}^{-1} \propto l_{\text{ap}}^{-2}$ . Secondly, the flux ratios of equations (36) and (37) include the summation over all peaks in the Monte-Carlo realisation, which means that the uncertainty  $\sigma$  on the numerator and denominator should scale with  $N_{\text{gas,mc}}^{1/2}$  and  $N_{\text{star,mc}}^{1/2}$ , respectively. Both of these dependences are included in equations (52) and (53). The division by the total flux in these equations converts the absolute uncertainties to relative ones.

The uncertainties due to the intrinsic variance of the peak flux (e.g. by flux evolution or a region mass spectrum) are also straightforward to define. Specifically, we take the standard deviation of the aperture fluxes for each of the four flux measurements (i.e. each of the two flux types around each of the two peak types) and convert this to a relative uncertainty on the sum of all flux measurements by writing

$$\sigma_{\text{star,star}}^2 = \frac{\langle \mathcal{F}_{\text{star},ij}^2 \rangle_{\text{peak,star}} - \langle \mathcal{F}_{\text{star},ij} \rangle_{\text{peak,star}}^2}{\mathcal{F}_{\text{star,star}}^2} N_{\text{star,mc}}, \quad (54)$$

and

$$\sigma_{\text{gas,star}}^2 = \frac{\langle \mathcal{F}_{\text{gas},ij}^2 \rangle_{\text{peak,star}} - \langle \mathcal{F}_{\text{gas},ij} \rangle_{\text{peak,star}}^2}{\mathcal{F}_{\text{gas,star}}^2} N_{\text{star,mc}}, \quad (55)$$

for apertures focused on stellar peaks (i.e.  $\mathcal{R}_{\text{star}}$ ). Likewise, for apertures focused on gas peaks (i.e.  $\mathcal{R}_{\text{gas}}$ ) we define

$$\sigma_{\text{star,gas}}^2 = \frac{\langle \mathcal{F}_{\text{star},ij}^2 \rangle_{\text{peak,gas}} - \langle \mathcal{F}_{\text{star},ij} \rangle_{\text{peak,gas}}^2}{\mathcal{F}_{\text{star,gas}}^2} N_{\text{gas,mc}}, \quad (56)$$

and

$$\sigma_{\text{gas,gas}}^2 = \frac{\langle \mathcal{F}_{\text{gas},ij}^2 \rangle_{\text{peak,gas}} - \langle \mathcal{F}_{\text{gas},ij} \rangle_{\text{peak,gas}}^2}{\mathcal{F}_{\text{gas,gas}}^2} N_{\text{gas,mc}}. \quad (57)$$

In these expressions, the numerator represents the variance (i.e. the standard deviation squared) of the flux in a single aperture across the population of stellar (subscript ‘peak,star’) or gas (subscript ‘peak,gas’) peaks, the denominator represents the total flux that is used in the flux ratios of equations (36) and (37), and the factor  $N$  at the end converts the variance of the flux in a single aperture to that of the total flux.

Having defined all relative uncertainty terms that contribute to equations (50) and (51), we should subtract the relative covariance between the stellar and gas fluxes in the apertures, because  $\mathcal{R}_{\text{star}}$  and  $\mathcal{R}_{\text{gas}}$  represent the ratios of these fluxes. For apertures focused on stellar and gas peaks, the relative covariances are

$$\text{COV}(\mathcal{F})_{\text{peak,star}} = \langle \mathcal{F}_{\text{star},ij} \mathcal{F}_{\text{gas},ij} \rangle_{\text{peak,star}} - \langle \mathcal{F}_{\text{star},ij} \rangle_{\text{peak,star}} \langle \mathcal{F}_{\text{gas},ij} \rangle_{\text{peak,star}}, \quad (58)$$

and

$$\text{COV}(\mathcal{F})_{\text{peak,gas}} = \langle \mathcal{F}_{\text{star},ij} \mathcal{F}_{\text{gas},ij} \rangle_{\text{peak,gas}} - \langle \mathcal{F}_{\text{star},ij} \rangle_{\text{peak,gas}} \langle \mathcal{F}_{\text{gas},ij} \rangle_{\text{peak,gas}}, \quad (59)$$

respectively. These covariances define the correlation coefficients between the stellar and gas fluxes in the apertures in equations (50) and (51) as

$$\rho_{\text{star}} = \frac{\text{COV}(\mathcal{F})_{\text{peak,star}}}{\sigma_{\text{star,star}} \sigma_{\text{gas,star}}}, \quad (60)$$

for apertures focused on stellar peaks and

$$\rho_{\text{gas}} = \frac{\text{COV}(\mathcal{F})_{\text{peak,gas}}}{\sigma_{\text{star,gas}} \sigma_{\text{gas,gas}}}, \quad (61)$$

for apertures focused on gas peaks, respectively. Together, equations (50)–(61) define the uncertainty on each individual data point in equations (30) and (31). These are converted to uncertainties on the logarithmic flux ratio bias by writing

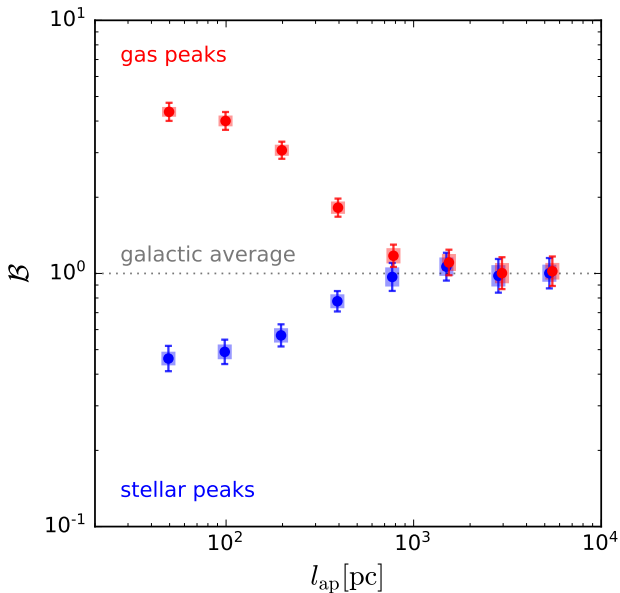
$$\sigma_{\log_{10} \mathcal{B}} = \frac{1}{\ln 10} \frac{\sigma_{\mathcal{B}}}{\mathcal{B}}. \quad (62)$$

The result of the presented uncertainty calculation is illustrated by the error bars in Figure 10 for the example maps used throughout this section.

To determine the variances and covariances in equations (54)–(59), we calculate the flux averages across all peaks, instead of evaluating the variance for each Monte-Carlo realisation of non-overlapping aperture samples and then taking the mean variance, or calculating the variance of the total flux across all Monte-Carlo realisations. As a result, equations (54)–(59) may seem lacking in two regards.

(i) Firstly, these expressions do not account for the fact that overlapping apertures are omitted in each Monte-Carlo realisation. However, this is not necessarily incorrect, because pixels contained by multiple apertures are more likely to be drawn. Determining the variance for each realisation would overestimate the variance due to the range of environments covered by the maps. In the extreme example of large aperture sizes, the flux variance in each Monte-Carlo





**Figure 10.** Illustration of the uncertainties on the example tuning fork diagram of Figure 8. The error bars with hats indicate the uncertainty on each individual data point, without accounting for the covariance between the data points. By contrast, the red (top) and blue (bottom) shaded areas behind each of the data points indicates the effective uncertainty range that accounts for the correlation between the data points (see the text). These ‘effective’ uncertainties are used when obtaining the best-fitting model in Section 3.2.12 and provide the correct point of reference when visually assessing the quality of a fit.

realisation may be based on a small number of peaks in widely different parts of a galaxy, leading to a large aperture-to-aperture flux variance, even if the total flux may be similar in each Monte-Carlo realisation. Such a scenario arises naturally when the same environments are included in each realisation. Overestimating the variance in this way would be highly undesirable.

(ii) Alternatively, one could take the variance of the total fluxes across all Monte-Carlo realisations. However, this would underestimate the variance – if none of the peaks is overlapping, then the total flux variance between the Monte-Carlo realisations will be zero, because all peaks are included in each realisation. This would give the erroneous impression that the variance is constrained to infinite precision, which gravely underestimates the variance at small aperture sizes. However, the non-zero variance only reflects the point that the total number of peaks in the maps is finite. This finite peak sample itself should be seen as one possible realisation of an underlying physical distribution of peak fluxes, which implies a non-zero variance of the total flux even if all peaks are included.

We find that taking the variance of all aperture fluxes, irrespective of whether an aperture is included or not, provides an accurate middle ground between the above two extremes. It accounts for the fact that pixels contained by multiple apertures are more likely to be drawn in each Monte-Carlo realisation, while also properly accounting for the fact that the peak population itself represents one possible realisation of the underlying physical flux distribution function. The uncertainties due to the intrinsic variance of the peak flux are therefore defined by averaging over all peaks in equations (54)–(59).

The error bars derived above are corrected for the correlation between the numerator and denominator in the gas-to-stellar flux

ratios of equations (36) and (37), but they do not account for the fact that the data points in the tuning fork diagram are not independent. This dependence arises because the apertures of different sizes are focused on the same peaks, implying that some fraction of the flux contained by apertures of a certain size is also present in apertures of other sizes. We correct for this dependence between the data points during the fitting process in Section 3.2.12 by calculating their normalised ‘independence fraction’, which represents the fraction of the flux in a data point that is independent from the flux in all other data points, and using this quantity to weigh the contribution of each data point to the goodness-of-fit statistic. The sum of these independence fractions corresponds to the total number of independent data points that the model from Section 3.2.11 is fitted to. The net effect of accounting for the correlation between data points is to decrease the number of degrees of freedom, leading to a larger difference between the model and the data. This is equivalent to decreasing the error bars on the data points, which is illustrated by the red and blue shaded areas behind each data point in Figure 10. These shaded areas represent ‘effective’ uncertainties that should be used for visually evaluating the quality of a fit. In the text below, we derive these effective uncertainties.

First, we determine for each aperture size which fraction of the flux is independent from the flux in the other apertures. For the stellar and gas flux around stellar peaks, we define the fraction of the flux at an aperture size  $l_{\text{ap}}(j)$  that is also contained by an aperture of (smaller) size  $l_{\text{ap}}(i)$ , with  $\{i, j \in \mathbb{N} | i \leq j\}$ , to be

$$N_{\text{star,star}} = \frac{\mathcal{F}_{\text{star,star}}[l_{\text{ap}}(i)]}{\mathcal{F}_{\text{star,star}}[l_{\text{ap}}(j)]} \frac{N_{\text{star,mc}}[l_{\text{ap}}(j)]}{N_{\text{star,mc}}[l_{\text{ap}}(i)]} \left[ \frac{l_{\text{ap}}(i)}{l_{\text{ap}}(j)} \right]^2, \quad (63)$$

and

$$N_{\text{gas,star}} = \frac{\mathcal{F}_{\text{gas,star}}[l_{\text{ap}}(i)]}{\mathcal{F}_{\text{gas,star}}[l_{\text{ap}}(j)]} \frac{N_{\text{star,mc}}[l_{\text{ap}}(j)]}{N_{\text{star,mc}}[l_{\text{ap}}(i)]} \left[ \frac{l_{\text{ap}}(i)}{l_{\text{ap}}(j)} \right]^2, \quad (64)$$

respectively. Similarly, for the stellar and gas flux around gas peaks, we define the fraction of the flux at an aperture size  $l_{\text{ap}}(j)$  that is also contained by an aperture of (smaller) size  $l_{\text{ap}}(i)$ , with  $\{i, j \in \mathbb{N} | i \leq j\}$ , to be

$$N_{\text{star,gas}} = \frac{\mathcal{F}_{\text{star,gas}}[l_{\text{ap}}(i)]}{\mathcal{F}_{\text{star,gas}}[l_{\text{ap}}(j)]} \frac{N_{\text{gas,mc}}[l_{\text{ap}}(j)]}{N_{\text{gas,mc}}[l_{\text{ap}}(i)]} \left[ \frac{l_{\text{ap}}(i)}{l_{\text{ap}}(j)} \right]^2, \quad (65)$$

and

$$N_{\text{gas,gas}} = \frac{\mathcal{F}_{\text{gas,gas}}[l_{\text{ap}}(i)]}{\mathcal{F}_{\text{gas,gas}}[l_{\text{ap}}(j)]} \frac{N_{\text{gas,mc}}[l_{\text{ap}}(j)]}{N_{\text{gas,mc}}[l_{\text{ap}}(i)]} \left[ \frac{l_{\text{ap}}(i)}{l_{\text{ap}}(j)} \right]^2, \quad (66)$$

respectively. As before, we have followed the convention that the first word of a subscript indicates the type of flux and the second word indicates the peak type that the apertures are focused on. We symmetrise these four arrays by setting

$$N(j, i | i \leq j) = N(i, j | i \leq j), \quad (67)$$

for each of them. As a result, array elements with  $i > j$  indicate the fraction of the flux at an aperture size  $l_{\text{ap}}(i)$  that is also contained by an aperture of (smaller) size  $l_{\text{ap}}(j)$ . For all four arrays, setting  $i = j$  gives  $N = 1$ , indicating the trivial result that an aperture contains all of the flux within that aperture.

Equations (63)–(66) define the fraction of the flux in a larger aperture that is ‘reused’ from a smaller aperture, for each aperture size pair. The above definitions imply that by performing a summation over one of the dimensions in these ‘reuse’ fraction arrays, one obtains the non-integer number of data points that are constituted by the flux contained within an aperture of size  $l_{\text{ap}}(i)$ , i.e.

$$N_{\text{data}}(i) = \sum_j N(i, j), \quad (68)$$

which returns a number  $1 < N_{\text{data}}[l_{\text{ap}}(i)] \leq N_{\text{ap}}$ . We then define the ‘independence fraction’ of the flux at an aperture size  $l_{\text{ap}}(i)$  as the inverse of this number

$$f_{\text{data}}(i) = \frac{1}{N_{\text{data}}} = \frac{1}{\sum_j N(i, j)}, \quad (69)$$

which takes values of  $0 < f_{\text{data}} < 1$ . Because the flux ratio biases of equations (30) and (31) that are shown in Figure 10 represent flux ratios, we define the independence fraction of each data point as the mean of the fluxes in the numerator and the denominator, resulting in

$$f_{\text{data,star}}(i) = \frac{1}{2 \sum_j N_{\text{star,star}}(i, j)} + \frac{1}{2 \sum_j N_{\text{gas,star}}(i, j)}, \quad (70)$$

and

$$f_{\text{data,gas}}(i) = \frac{1}{2 \sum_j N_{\text{star,gas}}(i, j)} + \frac{1}{2 \sum_j N_{\text{gas,gas}}(i, j)}, \quad (71)$$

respectively. The total number of independent data points in each of the arms of the tuning fork diagram then follows as

$$N_{\text{indep,star}} = \sum_i f_{\text{data,star}}(i), \quad (72)$$

and

$$N_{\text{indep,gas}} = \sum_i f_{\text{data,gas}}(i), \quad (73)$$

for the bottom arm and the top arm of the diagram, focusing on stellar peaks and gas peaks, respectively. The total number of independent data points is therefore smaller than twice the number of aperture sizes, i.e.

$$N_{\text{indep}} = N_{\text{indep,star}} + N_{\text{indep,gas}} < 2N_{\text{ap}}. \quad (74)$$

As stated above, the correlation between data points is accounted for during the fitting process of Section 3.2.12 by weighing the contribution of each data point to the goodness-of-fit statistic by its normalised independence fraction. We also decrease the number of degrees of freedom by using the total number of independent data points  $N_{\text{indep}}$  rather than the total number of data points  $2N_{\text{ap}}$ . The net effect of these changes is equivalent to decreasing the uncertainties  $\sigma_B$  to becoming ‘effective uncertainties’  $\sigma'_B$ . The weights and degrees of freedom are discussed in detail in Section 3.2.12, but here we provide these effective uncertainties. Relative to the error bars defined in equations (48) and (49), we obtain correction factors of

$$\frac{\sigma'_{\log_{10} \mathcal{B}_{\text{star}}}(i)}{\sigma_{\log_{10} \mathcal{B}_{\text{star}}}(i)} = \sqrt{\frac{N_{\text{indep}}}{2f_{\text{data,star}}(i)N_{\text{ap}}} \frac{N_{\text{indep}} - N_{\text{fit}}}{2N_{\text{ap}} - N_{\text{fit}}}}, \quad (75)$$

and

$$\frac{\sigma'_{\log_{10} \mathcal{B}_{\text{gas}}}(i)}{\sigma_{\log_{10} \mathcal{B}_{\text{gas}}}(i)} = \sqrt{\frac{N_{\text{indep}}}{2f_{\text{data,gas}}(i)N_{\text{ap}}} \frac{N_{\text{indep}} - N_{\text{fit}}}{2N_{\text{ap}} - N_{\text{fit}}}}, \quad (76)$$

for the gas-to-stellar flux ratio bias when focusing on stellar peaks and gas peaks, respectively. The square-roots in these expressions contain two ratios. The first term reflects the inverse of the weight that is assigned to the data points when evaluating the goodness-of-fit statistic. The second term represents the decrease of the number of degrees of freedom by only counting the effective number of independent data points, where  $N_{\text{fit}}$  refers to the number of free parameters (this is  $N_{\text{fit}} = 3$  for the problem considered here, see Section 3.2.12 for details). The above two expressions provide the effective uncertainties shown by the shaded areas in Figure 10.

### 3.2.11 Model for interpreting the gas-to-stellar flux ratio bias

We now turn to the derivation of the theoretical model for the gas-to-stellar flux ratio bias that will be fitted to the observed tuning fork diagram of Figure 10 in Section 3.2.12 below. The presentation in the original KL14 principle paper already included a theoretical model for the flux ratio bias (see their Appendix C), but the model presented here eliminates an important assumption of the previous version. Previously, the model assumed that all independent regions in the map resemble point sources, implying that the surface densities approach infinity in the limit of  $l_{\text{ap}} \downarrow 0$ . This then leads to a flux ratio bias that diverges at small aperture sizes if  $t_{\text{over}} = 0$  (see Figure 4 of KL14). In the present paper, we describe the regions by spatial profiles with finite central densities, which is obviously appropriate for real-Universe regions. This implies that a non-diverging flux ratio bias at small aperture sizes does not necessarily require  $t_{\text{over}} > 0$ . This change is critical for the accuracy of observational applications of the method. After presenting the derivation of the new model, we will briefly discuss the effect on the predicted flux ratio bias of changing the model parameters. These include the to-be-fitted, free parameters  $t_{\text{gas}}$ ,  $t_{\text{over}}$ , and  $\lambda$ , as well as input parameters of the method ( $t_{\text{star}}$  and `peak_profile`) and parameters that have been extracted from the maps in the previous sections ( $\beta_{\text{star}}$ ,  $\beta_{\text{gas}}$ ,  $\mathcal{E}_{\text{star}}$ , and  $\mathcal{E}_{\text{gas}}$ ).

In deriving the new model, we will start by taking the result from KL14 and then modify it to introduce the peak surface density profiles. In the original model, the gas-to-stellar flux ratio bias is obtained by counting regions within the aperture. The flux from the peak type (i.e. stars or gas) that the apertures are focused on is constituted by the sum of the central peak and the statistically expected background flux from other peaks. By contrast, the flux from the other tracer is constituted by the sum of the statistically expected background flux from other peaks and the central peak if it spends any time in the overlap phase. This latter contribution depends on the time-scales involved – for stellar peaks, the probability that the central peak also contains gas flux is  $f_{\text{star,over}} = t_{\text{over}}/t_{\text{star}}$ , whereas the probability that gas peaks also contain stellar flux is  $f_{\text{gas,over}} = t_{\text{over}}/t_{\text{gas}}$ . This contribution should also be corrected for any evolution of the peak flux, which may result in fluxes during the overlap phase that differ from those in the isolated phase (as captured by the parameters  $\beta_{\text{star}}$  and  $\beta_{\text{gas}}$ , see Section 3.2.9).

The expressions for the gas-to-stellar flux ratio bias are derived in Appendix C of KL14 and we refer the interested reader to that earlier work for the details of the derivation. The gas-to-stellar flux ratio bias when focusing apertures on stellar peaks (corresponding to the bottom branch in the tuning fork diagram) is

$$\mathcal{B}_{\text{star}}^{\text{KL14}} = \frac{\left[1 + \beta_{\text{gas}}^{-1} \left(\frac{t_{\text{gas}}}{t_{\text{over}}} - 1\right)\right]^{-1} + \frac{t_{\text{star}}}{\tau} \left(\frac{l_{\text{ap}}}{\lambda}\right)^2}{1 + \frac{t_{\text{star}}}{\tau} \left(\frac{l_{\text{ap}}}{\lambda}\right)^2}. \quad (77)$$

When focusing on gas peaks (top branch in the tuning fork diagram), the bias is given by

$$\mathcal{B}_{\text{gas}}^{\text{KL14}} = \frac{1 + \frac{t_{\text{gas}}}{\tau} \left(\frac{l_{\text{ap}}}{\lambda}\right)^2}{\left[1 + \beta_{\text{star}}^{-1} \left(\frac{t_{\text{star}}}{t_{\text{over}}} - 1\right)\right]^{-1} + \frac{t_{\text{gas}}}{\tau} \left(\frac{l_{\text{ap}}}{\lambda}\right)^2}, \quad (78)$$

Each of the variables in these expressions have already been introduced, but for clarity we briefly repeat them. The time-scale  $\tau = t_{\text{star}} + t_{\text{gas}} - t_{\text{over}}$  is the total duration of the evolutionary timeline in Figure 1,  $t_{\text{star}}$  is the duration of the stellar phase,  $t_{\text{gas}}$  is the duration of the gas phase,  $t_{\text{over}}$  is the duration of the overlap phase,



$\lambda$  is the mean separation length between independent regions,  $l_{\text{ap}}$  is the aperture size, and  $\beta_{\text{star}}$  and  $\beta_{\text{gas}}$  represent the flux ratios between the overlap phase and the isolated phase of stellar peaks and gas peaks, respectively. In both of the above two expressions, the terms containing the square of the aperture size represent the background flux from other peaks, the ‘1’ represents the peak that the aperture is focused on, and the terms containing  $t_{\text{over}}$  represent the contribution from that peak to the other tracer if it resides in the overlap phase. A change relative to KL14 that is not immediately visible in these expressions is that  $\beta_{\text{star}}$  and  $\beta_{\text{gas}}$  are now treated as functions of  $f_{\text{star,over}} = t_{\text{over}}/t_{\text{star}}$  and  $f_{\text{gas,over}} = t_{\text{over}}/t_{\text{gas}}$ . In the exploration of the model behaviour below, we will treat these as independent parameters, but when fitting the model to the observed tuning fork diagram in Section 3.2.12,  $\beta_{\text{star}}$  and  $\beta_{\text{gas}}$  are not independent variables.

As a consequence of fully counting the peak that is being focused on (as well as fully counting any contribution from this peak to the other tracer if it resides in the overlap phase), irrespective of the aperture size, these formulations implicitly assume that the peak profiles are delta functions, i.e. that their radii are infinitesimally small. As a result, gas-to-stellar flux ratios of equations (77) and (78) diverge when  $t_{\text{over}} \downarrow 0$ . This can be seen directly by taking the limits of  $l_{\text{ap}} \downarrow 0$ , which yields

$$\lim_{l_{\text{ap}} \downarrow 0} \mathcal{B}_{\text{star}}^{\text{KL14}} = \left[ 1 + \beta_{\text{gas}}^{-1} \left( \frac{t_{\text{gas}}}{t_{\text{over}}} - 1 \right) \right]^{-1}, \quad (79)$$

and

$$\lim_{l_{\text{ap}} \downarrow 0} \mathcal{B}_{\text{gas}}^{\text{KL14}} = 1 + \beta_{\text{star}}^{-1} \left( \frac{t_{\text{star}}}{t_{\text{over}}} - 1 \right). \quad (80)$$

If  $t_{\text{over}} \downarrow 0$ , these expressions approach zero and infinity, respectively. This result led us to conclude in KL14 that the flattening of the tuning fork diagram at small aperture sizes provides a way of measuring  $t_{\text{over}}$ . While this is still true, we pointed out above that the real-Universe situation is more complex, because the flattening of the tuning fork diagram is not unique to a non-zero duration of the overlap phase. Real molecular clouds or star-forming regions are not infinitesimally small and their finite central densities provide another way of avoiding divergence.

We account for the peak surface density profiles by introducing an extra factor in equations (77) and (78) that takes values between 0 and 1 and represents the fraction of the flux from the central peak that is contained within the aperture. This factor is only added to the terms that represent the stellar and gas flux contribution from the central peak. For the randomly-distributed, background population of spatially-extended peaks in and around the aperture, the flux contributions from peaks with centres outside of the aperture cancel on geometric grounds with the flux loss from peaks with centres within the aperture. As a result, the background flux in the aperture is exclusively set by the number surface density of the peaks and is insensitive to their individual surface density profiles.<sup>16</sup> This allows us to formulate the updated version of the KL14 principle model as

$$\mathcal{B}_{\text{star}}^{\text{mod}} = \frac{f_{\text{gas}} \left[ 1 + \beta_{\text{gas}}^{-1} \left( \frac{t_{\text{gas}}}{t_{\text{over}}} - 1 \right) \right]^{-1} + \frac{t_{\text{star}}}{\tau} \left( \frac{l_{\text{ap}}}{\lambda} \right)^2}{f_{\text{star}} + \frac{t_{\text{star}}}{\tau} \left( \frac{l_{\text{ap}}}{\lambda} \right)^2}, \quad (81)$$

<sup>16</sup> We thank Sophie Kruijssen for pointing out this logic, which has led to an important simplification of the model.

and

$$\mathcal{B}_{\text{gas}}^{\text{mod}} = \frac{f_{\text{gas}} + \frac{t_{\text{gas}}}{\tau} \left( \frac{l_{\text{ap}}}{\lambda} \right)^2}{f_{\text{star}} \left[ 1 + \beta_{\text{star}}^{-1} \left( \frac{t_{\text{star}}}{t_{\text{over}}} - 1 \right) \right]^{-1} + \frac{t_{\text{gas}}}{\tau} \left( \frac{l_{\text{ap}}}{\lambda} \right)^2}, \quad (82)$$

where  $f_{\text{star}}$  and  $f_{\text{gas}}$  represent the fraction of the central peak flux that is contained by the aperture, with the subscripts referring to stellar peaks and gas peaks respectively. These expressions are used throughout the rest of this paper.

Equations (81) and (82) do not yet show how  $f_{\text{star}}$  and  $f_{\text{gas}}$  depend on the aperture size. We will specify this below for a small number of useful peak surface density profiles. However, in defining the above expressions, we have already made an important assumption. The enclosed flux fraction of the central peak is assumed to differ between both tracers, because a factor of  $f_{\text{gas}}$  appears in the numerator and a factor of  $f_{\text{star}}$  appears in the denominator of equations (81) and (82). This means that we allow the flux from both tracers to follow different profiles, which provides some useful flexibility and is consistent with observations (e.g. Walker et al. 2015, 2016). However, we also assume that there is no difference in profile shape between peaks of a given tracer that reside in the isolated and overlap phases, because the same factors of  $f_{\text{star}}$  and  $f_{\text{gas}}$  are used in equations (81) and (82). This may introduce a minor inconsistency in cases where there is considerable time-evolution of the peak size. In practical applications, however, this typically happens on size scales below the resolution limit, because we convolve the maps to a common resolution  $l_{\text{ap,min}}$  (see Section 3.2.3). It is demonstrated in Section 4 that the method yields accurate measurements of  $t_{\text{over}}$ , to such an extent that any effect of this inconsistency must fall within acceptable limits. Indeed, we have verified that the results do not change (i.e. they stay well within the obtained uncertainties) when dividing the peak sample in peaks residing in the isolated and overlap phases before determining the flux density profiles. It is therefore undesirable to add another layer of complexity that could address this, in particular because we have no a priori constraints on how the peak profiles change with evolutionary stage.

We consider three different types of surface density profiles to model the peak emission, all of which are available in HEISENBERG. However, only the third of these profiles will be used throughout the rest of the paper, for reasons outlined below. Each of these profiles is normalised to have a total flux fraction of unity (i.e.  $\lim_{l_{\text{ap}} \rightarrow \infty} f(l_{\text{ap}}) = 1$ ), so that at infinitely large aperture sizes the entire peak resides within the aperture. The profiles are as follows.

(i) A two-dimensional delta function that is centred on the peak position, resulting in an enclosed flux fraction as a function of aperture size that is given by

$$f_{\delta}(l_{\text{ap}}) = 1. \quad (83)$$

This corresponds to the case derived in KL14, where we assumed that the peaks correspond to points.

(ii) A two-dimensional, constant surface density disc that is centred on the peak position, resulting in an enclosed flux fraction as a function of aperture size that is given by

$$f_{\text{disc}}(l_{\text{ap}}) = \begin{cases} (l_{\text{ap}}/2r_{\text{peak}})^2 & \text{if } l_{\text{ap}} < 2r_{\text{peak}} \\ 1 & \text{if } l_{\text{ap}} \geq 2r_{\text{peak}}. \end{cases} \quad (84)$$

In this expression,  $r_{\text{peak}}$  represents the radius of the constant surface density disc. We discuss below how this radius is obtained from the galaxy maps.

(iii) A two-dimensional Gaussian centred on the peak position, resulting in an enclosed flux fraction as a function of aperture size that is given by

$$f_{\text{Gauss}}(l_{\text{ap}}) = 1 - \exp \left[ -\frac{1}{2} \left( \frac{l_{\text{ap}}}{2r_{\text{peak}}} \right)^2 \right]. \quad (85)$$

In this expression,  $r_{\text{peak}}$  represents the dispersion of the two-dimensional Gaussian. We discuss below how this scale radius is obtained from the galaxy maps.

The second and third of these profiles require the characteristic peak radius  $r_{\text{peak}}$  to be known. However, it is straightforward to obtain this from the galaxy maps once a peak profile has been assumed to approximate the flux density profiles of the peaks in the maps. In Section 3.2.9, we measured the flux density contrast of emission peaks to the average flux density on a size scale  $l_{\text{ap}}$  in equations (44) and (45). This flux ratio contrast can be related to the ratio between the central surface density of peaks and the background surface density of the peak population. In turn, this allows us to express the peak radius  $r_{\text{peak}}$  in units of the region separation length  $\lambda$ . For the stellar and gas flux, the background number densities of independent regions are

$$\Sigma_{\text{back,star}} = \frac{t_{\text{star}}}{\tau} \left( \frac{\pi\lambda^2}{4} \right)^{-1}, \quad (86)$$

and

$$\Sigma_{\text{back,gas}} = \frac{t_{\text{gas}}}{\tau} \left( \frac{\pi\lambda^2}{4} \right)^{-1}, \quad (87)$$

respectively. In these expressions, the term in parentheses uses the mean separation length between independent regions to represent the area per region, whereas the multiplications by  $t_{\text{star}}/\tau$  and  $t_{\text{gas}}/\tau$  reduce the resulting total number densities to only refer to the number densities of stellar peaks and gas peaks, respectively.

For a constant surface density disc normalised to unity, the central surface density is

$$\Sigma_{\text{disc}} = \frac{1}{\pi r_{\text{peak}}^2}, \quad (88)$$

whereas for a two-dimensional Gaussian normalised to unity, the central surface density is

$$\Sigma_{\text{Gauss}} = \frac{1}{2\pi r_{\text{peak}}^2}. \quad (89)$$

With the expected background and central flux densities in hand, it is now possible to relate these to the flux density contrasts  $\mathcal{E}_{\text{star}}$  and  $\mathcal{E}_{\text{gas}}$  as a function of aperture size obtained in equations (44) and (45) above (see Section 3.2.9). This will provide the peak radius  $r_{\text{peak}}$  in units of the region separation length  $\lambda$ .

Because  $\mathcal{E}_{\text{star}}$  and  $\mathcal{E}_{\text{gas}}$  represent flux density contrasts measured relative to a variable aperture size, there is some freedom in choosing the size scale on which the flux density contrast is evaluated. We find that it is desirable to define  $\mathcal{E}_{\text{star}}$  and  $\mathcal{E}_{\text{gas}}$  on a size scale  $\lambda$ , because on larger scales the considered maps may contain voids or other morphological features that change the surface number density of regions from the density in the direct vicinity of the peaks implied by the best-fitting  $\lambda$ . By setting the size scale to  $\lambda$ , the background surface number density of neighbouring peaks best corresponds to the measured region separation length. In practice, this means that when fitting the observed tuning fork diagram in Section 3.2.12 below,  $\mathcal{E}_{\text{star}}$  and  $\mathcal{E}_{\text{gas}}$  are both varied with  $\lambda$  and are constrained as a byproduct of the fitting process.

On a size scale  $\lambda$ , the measured flux ratio contrasts defined in equations (44) and (45) can be expressed in terms of the above central and background number densities as

$$\mathcal{E}_{\text{star}} \equiv \mathcal{E}_{\text{star}}(\lambda) = \frac{\Sigma_{\text{peak}} + \Sigma_{\text{back,star}}}{(\pi\lambda^2/4)^{-1} + \Sigma_{\text{back,star}}}, \quad (90)$$

and

$$\mathcal{E}_{\text{gas}} \equiv \mathcal{E}_{\text{gas}}(\lambda) = \frac{\Sigma_{\text{peak}} + \Sigma_{\text{back,gas}}}{(\pi\lambda^2/4)^{-1} + \Sigma_{\text{back,gas}}}, \quad (91)$$

where  $\Sigma_{\text{peak}}$  refers to  $\Sigma_{\text{disc}}$  or  $\Sigma_{\text{Gauss}}$  depending on the adopted profile. In these two expressions, the numerator represents the flux density on a size scale  $l_{\text{ap,min}}$ , consisting of the central surface density of the peak (first term) and the background population (second term), and the denominator represents the flux density on a size scale  $\lambda$ , consisting of the entire central peak (first term) and the background population (second term).

Rearranging the above expressions allows the characteristic peak radii to be expressed in terms of quantities that are either measured directly from the maps ( $\mathcal{E}_{\text{star}}$  and  $\mathcal{E}_{\text{gas}}$ ), or represent free parameters during the fitting process in Section 3.2.12 ( $\tau$ ,  $t_{\text{gas}}$ , and  $\lambda$ ). If we represent the peaks with constant surface density discs, we obtain for the radii of stellar peaks

$$r_{\text{star}} = \frac{\lambda}{2} \sqrt{\frac{\tau/t_{\text{star}}}{\mathcal{E}_{\text{star}}(1 + \tau/t_{\text{star}}) - 1}}, \quad (92)$$

and for the radii of gas peaks

$$r_{\text{gas}} = \frac{\lambda}{2} \sqrt{\frac{\tau/t_{\text{gas}}}{\mathcal{E}_{\text{gas}}(1 + \tau/t_{\text{gas}}) - 1}}. \quad (93)$$

Likewise, if we represent the peaks with two-dimensional Gaussian profiles, we obtain for the radii of stellar peaks

$$r_{\text{star}} = \frac{\lambda}{2} \sqrt{\frac{\tau/t_{\text{star}}}{2\mathcal{E}_{\text{star}}(1 + \tau/t_{\text{star}}) - 2}}, \quad (94)$$

and for the radii of gas peaks

$$r_{\text{gas}} = \frac{\lambda}{2} \sqrt{\frac{\tau/t_{\text{gas}}}{2\mathcal{E}_{\text{gas}}(1 + \tau/t_{\text{gas}}) - 2}}. \quad (95)$$

Equations (92)–(95) are substituted into the enclosed flux fractions as a function of aperture size from equations (84) and (85). Substitution of the result into equations (81) and (82) yields the predicted gas-to-stellar flux ratio bias when focusing apertures on stellar peaks and gas peaks, respectively, in terms of measurable quantities and free parameters of the fitting process in Section 3.2.12.

When applying the method to real-Universe observations or to the simulated galaxy maps in Section 4, the Gaussian peak surface density profile  $f_{\text{Gauss}}$  is generally the best choice to make. Above all, it is versatile – in the limit of a very large flux ratio excess  $\mathcal{E}$ , the peak radii approach zero and hence the possible range of profiles includes a reasonable approximation of the delta function profile  $f_{\delta}$  to represent point particles. This feature is also shared by the constant surface density disc profile  $f_{\text{disc}}$ , but the Gaussian profile is more suitable for two other reasons. Firstly, the regridding of the maps in Section 3.2.3 is accompanied by a convolution of the maps with a Gaussian PSF to a common resolution. This means that unresolved peaks have Gaussian profiles by definition and resolved peaks at least have a Gaussian component. Secondly, real-Universe molecular clouds and star-forming regions generally follow centrally concentrated morphologies, rather than having constant surface densities (e.g. Walker et al. 2015, 2016). Gaussian profiles

provide a better representation of this behaviour. Of course, it is straightforward to select any other desired analytical profile and derive expressions equivalent to equations (92)–(95) for that particular choice. This may yield differences in the best-fitting quantities, but for the above reasons we expect these to be minor. In the remainder of this paper, we use the Gaussian profile to represent the peaks in the map.

Substitution of the stellar and gas peak radii into the Gaussian profile of equation (85) provides the enclosed flux fraction of stellar and gas peaks as a function of the aperture size, i.e.

$$f_{\text{star}}(l_{\text{ap}}) = 1 - \exp \left[ \frac{1 - \mathcal{E}_{\text{star}} (1 + \tau/t_{\text{star}})}{\tau/t_{\text{star}}} \left( \frac{l_{\text{ap}}}{\lambda} \right)^2 \right], \quad (96)$$

and

$$f_{\text{gas}}(l_{\text{ap}}) = 1 - \exp \left[ \frac{1 - \mathcal{E}_{\text{gas}} (1 + \tau/t_{\text{gas}})}{\tau/t_{\text{gas}}} \left( \frac{l_{\text{ap}}}{\lambda} \right)^2 \right], \quad (97)$$

respectively. We will use these two expressions in combination with the gas-to-stellar flux ratios of equations (81) and (82) to fit the observed tuning fork diagram that was obtained in Section 3.2.10.

To avoid overly long expressions, we do not explicitly provide the (trivial) substitution of the peak density profiles into equations (81) and (82), but it is useful to consider the limit of  $l_{\text{ap}} \downarrow 0$  to highlight the difference in behaviour relative to the original model of KL14. For small aperture sizes, we obtain

$$\lim_{l_{\text{ap}} \downarrow 0} \mathcal{B}_{\text{star}}^{\text{mod}} = \frac{1}{\mathcal{E}_{\text{star}} \left( 1 + \frac{\tau}{t_{\text{star}}} \right)} + \frac{1 - \mathcal{E}_{\text{star}}^{-1} \left( 1 + \frac{\tau}{t_{\text{star}}} \right)^{-1}}{1 + \beta_{\text{gas}}^{-1} \left( \frac{t_{\text{gas}}}{t_{\text{over}}} - 1 \right)}, \quad (98)$$

for the gas-to-stellar flux ratio bias when focusing on stellar peaks and

$$\lim_{l_{\text{ap}} \downarrow 0} \mathcal{B}_{\text{gas}}^{\text{mod}} = \left[ \frac{1}{\mathcal{E}_{\text{gas}} \left( 1 + \frac{\tau}{t_{\text{gas}}} \right)} + \frac{1 - \mathcal{E}_{\text{gas}}^{-1} \left( 1 + \frac{\tau}{t_{\text{gas}}} \right)^{-1}}{1 + \beta_{\text{star}}^{-1} \left( \frac{t_{\text{star}}}{t_{\text{over}}} - 1 \right)} \right]^{-1}, \quad (99)$$

when focusing on gas peaks. It is straightforward to verify that we obtain the old limits of equations (79) and (80) when we consider point particles and thus let  $\mathcal{E}_{\text{star}} \rightarrow \infty$  and  $\mathcal{E}_{\text{gas}} \rightarrow \infty$ . However, the change relative to the old limits is best illustrated by considering the case of  $t_{\text{over}} \downarrow 0$ , which yields

$$\lim_{t_{\text{over}} \downarrow 0} \left( \lim_{l_{\text{ap}} \downarrow 0} \mathcal{B}_{\text{star}}^{\text{mod}} \right) = \frac{1}{\mathcal{E}_{\text{star}} \left( 1 + \frac{\tau}{t_{\text{star}}} \right)}, \quad (100)$$

and

$$\lim_{t_{\text{over}} \downarrow 0} \left( \lim_{l_{\text{ap}} \downarrow 0} \mathcal{B}_{\text{gas}}^{\text{mod}} \right) = \mathcal{E}_{\text{gas}} \left( 1 + \frac{\tau}{t_{\text{gas}}} \right), \quad (101)$$

for the bottom and top branches of the tuning fork diagram, respectively. These limits show immediately that setting  $\mathcal{E}_{\text{star}} \rightarrow \infty$  and  $\mathcal{E}_{\text{gas}} \rightarrow \infty$  to represent point-like regions results in a gas-to-stellar flux ratio bias that diverges to 0 or  $\infty$  for stellar and gas peaks, respectively, as in KL14. However, they also quantify that for peak surface density profiles with a finite central density, the gas-to-stellar flux ratio bias does not diverge to 0 or  $\infty$  towards small aperture sizes as in KL14, but instead takes values proportional to the (inverse) flux density contrast, i.e.  $\mathcal{E}^{-1}$  or  $\mathcal{E}$ . The two

**Table 3.** Model parameter values explored in Figure 11

Quantity	low	middle	high
free parameter			
$t_{\text{gas}}$ [Myr]	5	<b>10</b>	20
$t_{\text{over}}$ [Myr]	0	<b>2</b>	4
$\lambda$ [pc]	100	<b>300</b>	1000
function of free parameters			
$t_{\text{star}}$ [Myr]	2.5	<b>5</b>	10
$\beta_{\text{star}}$	0.5	<b>1</b>	2
$\beta_{\text{gas}}$	0.5	<b>1</b>	2
$\mathcal{E}_{\text{star}}$	2.5	<b>5</b>	10
$\mathcal{E}_{\text{gas}}$	2.5	<b>5</b>	10
fixed parameter			
peak_prof	points	discs	<b>Gaussians</b>

The boldface numbers indicate the fiducial values used in each of the panels in Figure 11.

factors in these limits represent the contribution from the peak flux relative to a size scale  $\lambda$  (the factor  $\mathcal{E}$ ) and from the background peak population within  $\lambda$  relative to the galactic average (the factor  $1 + \tau/t$ ). The first of these factors follows from the definition of  $\mathcal{E}$  and the second is easily verified by evaluating equations (81) and (82) at  $l_{\text{ap}} = \lambda$ . By combining these two factors, we obtain the central flux ratio bias of a peak relative to the galactic average. The qualitative behaviour of these limits is also as expected when setting  $t_{\text{over}} = 0$ . Firstly, a low filling factor (high  $\mathcal{E}$ ) implies a large difference between the central peak and the background population within  $\lambda$ , leading to a strong bias. Secondly, a rarity of tracer peaks within an aperture of size  $\lambda$  (i.e. when  $t_{\text{star}}$  or  $t_{\text{gas}}$  is much smaller than  $\tau$ ) implies a large difference between the flux density on a size scale  $\lambda$  and the galactic average, also leading to a strong bias.

The immediate implication of equations (100) and (101) is that the flattening of the tuning fork diagram towards small aperture sizes is still a measure of the duration of the overlap phase  $t_{\text{over}}$ , but there is a degeneracy with the central flux ratio excess of peaks relative to the background. However, this degeneracy is lifted by directly obtaining the peak flux ratio excess from the maps, as discussed in Section 3.2.9. In practice, a finite flux ratio excess provides a minimum amount of flattening. Any additional flattening indicates a non-zero duration of the overlap phase and thus still provides a way of directly measuring  $t_{\text{over}}$ .

We now turn to a brief discussion of how the gas-to-stellar flux ratio biases as a function of aperture size vary with each of the quantities involved in our model. Table 3 lists the nine quantities that together determine the flux ratio biases of equations (81) and (82), as well as three typical values for each of these quantities that we will consider during a quantitative illustration of the model below. The first three of these ( $t_{\text{gas}}$ ,  $t_{\text{over}}$ , and  $\lambda$ ) represent the three free parameters in the model and are obtained by fitting the model in Section 3.2.12 to the observed gas-to-stellar flux ratio bias of Section 3.2.10 and Figure 10. The second set of five quantities ( $t_{\text{star}}$ ,  $\beta_{\text{star}}$ ,  $\beta_{\text{gas}}$ ,  $\mathcal{E}_{\text{star}}$ , and  $\mathcal{E}_{\text{gas}}$ ) depend on the three free parameters and are therefore constrained during the fitting process, but are not free parameters themselves. The default definition of  $t_{\text{star}} = t_{\text{star,ref}} + t_{\text{over}}$  implies a dependence on the duration of the overlap phase, but if the flag `tstar_incl = 1` (see Table 1), then  $t_{\text{star}} = t_{\text{star,ref}}$  is fixed in the input file and becomes independent of the free parameters. In this example, we specify  $t_{\text{star}}$ , implying that  $t_{\text{star,ref}}$  changes when  $t_{\text{over}}$  changes, which effectively corresponds to `tstar_incl = 1`. By contrast,  $\beta_{\text{star}}(f_{\text{star,over}})$

and  $\beta_{\text{gas}}(f_{\text{gas,over}})$  always depend on the free parameters, because they depend on the ratios  $t_{\text{over}}/t_{\text{star}}$  and  $t_{\text{over}}/t_{\text{gas}}$ , respectively, as described in Section 3.2.9 and illustrated in Figure 9. Likewise,  $\mathcal{E}_{\text{star}}(\lambda)$  and  $\mathcal{E}_{\text{gas}}(\lambda)$  both depend on the mean separation length  $\lambda$  (see Section 3.2.9). While these five quantities will vary as a function of  $t_{\text{gas}}$ ,  $t_{\text{over}}$ , and  $\lambda$  in practical applications of the method, we consider them as free parameters in the example given here, with the goal of isolating their effects. The final quantity in Table 3 (`peak_prof`) represents a fixed parameter that is set in the input parameter file. This quantity therefore influences the shape of the modelled tuning fork diagram, but does not vary during the fitting process of Section 3.2.12.

Figure 11 systematically shows how each of the nine quantities from Table 3 affect the shape of the modelled gas-to-stellar flux ratio bias as a function of the aperture size. The main result of the parameter study in the figure is that each of the nine quantities in general, but especially the three free parameter in particular, have a unique effect on the shapes of the curves. This means that the gas-to-stellar flux ratio bias is a non-degenerate function of these quantities, allowing the free parameters to be measured by fitting the model to observed tuning fork diagrams. The basic behaviour is the same as in the schematic of Figure 2. When focusing apertures on stellar (gas) peaks, the gas-to-stellar flux ratio is smaller (larger) than the galactic average. This gas-to-stellar flux ratio bias increases towards decreasing aperture sizes, because the central peak contributes a larger fraction to the total flux ratio in the aperture when a smaller population of other peaks is included. However, the bias does not grow indefinitely towards small apertures. As discussed in the above derivation, this arises for two reasons. Firstly, if the stellar and gas phases overlap in time, then there is a non-zero probability that a peak resides in the overlap phase and emission from both tracers is present. This limits the gas-to-stellar flux ratio bias when preferentially selecting one type of tracer. Secondly, peaks have finite central surface densities, implying that for the smallest aperture sizes, the gas-to-stellar flux ratio bias reflects the overdensity of the peaks relative to the background population of other peaks, as demonstrated in equations (100) and (101).

Starting with the top row of Figure 11, we see in panel (a) that increasing  $t_{\text{gas}}$  decreases the bottom branch of the tuning fork diagram, i.e. it leads to a larger gas-to-stellar flux deficit when focusing on stellar peaks. This happens because for larger  $t_{\text{gas}}$ , the overlap phase (which sets the bias at small aperture sizes) takes up a smaller fraction of the total duration of the gas phase, implying that stellar peaks residing in the overlap phase contribute a smaller fraction of the galactic gas flux. Due to the relatively smaller ‘damping’ effect of the gas flux from stellar peaks in the overlap phase the bottom branch flattens less if  $t_{\text{gas}}$  is larger. By contrast, changing the duration of the gas phase does not affect the relative contribution of the stellar flux from gas peaks in the overlap phase to the galactic stellar flux, because this is mainly set by the ratio  $t_{\text{over}}/t_{\text{star}}$  in equation (82). The flattening of the top branch is therefore only weakly affected by changing  $t_{\text{gas}}$ . There is a small decrease of the aperture size below which the gas-to-stellar flux ratio bias sets in when increasing  $t_{\text{gas}}$ . This is easily understood – gas peaks become more common as  $t_{\text{gas}}$  increases, causing their number density to increase too. Therefore, a smaller aperture size is required to isolate the contribution from the gas peak that the aperture is focused on.

The second free parameter is the duration of the overlap phase  $t_{\text{over}}$ . Panel (b) shows that the effect of increasing  $t_{\text{over}}$  is that the probability that the central peak in the aperture also contains emission from the other tracer increases. This limits the gas-to-stellar flux ratio bias relative to the galactic average that can be achieved

at the smallest aperture sizes and therefore leads to a stronger flattening of the model tuning fork diagram towards larger  $t_{\text{over}}$ . The figure again illustrates the key difference relative to KL14, namely that the branches of the tuning fork do not diverge towards small aperture sizes when  $t_{\text{over}} = 0$ , due to the finite central surface densities of the two-dimensional Gaussians used here. Most importantly, the behaviour due to  $t_{\text{over}}$  fundamentally differs from the influence of  $t_{\text{gas}}$  on the gas-to-stellar flux ratio bias. While the latter affects the vertical position of one of the branches and only weakly affects the horizontal position of the other branch, the former symmetrically sets the vertical positions of both branches.

The third free parameter is the mean separation length between independent regions  $\lambda$ . Its effect on the shape of the tuning fork is easy to understand. Panel (c) shows that increasing the separation length means that the entire model tuning fork diagram shifts to larger aperture sizes, because recovering the galactic average by encompassing a sufficient number of regions in the aperture requires a larger aperture area. Crucially, the effect of the separation length is orthogonal to that of  $t_{\text{gas}}$  and  $t_{\text{over}}$ , as it controls the overall horizontal position of the diagram.

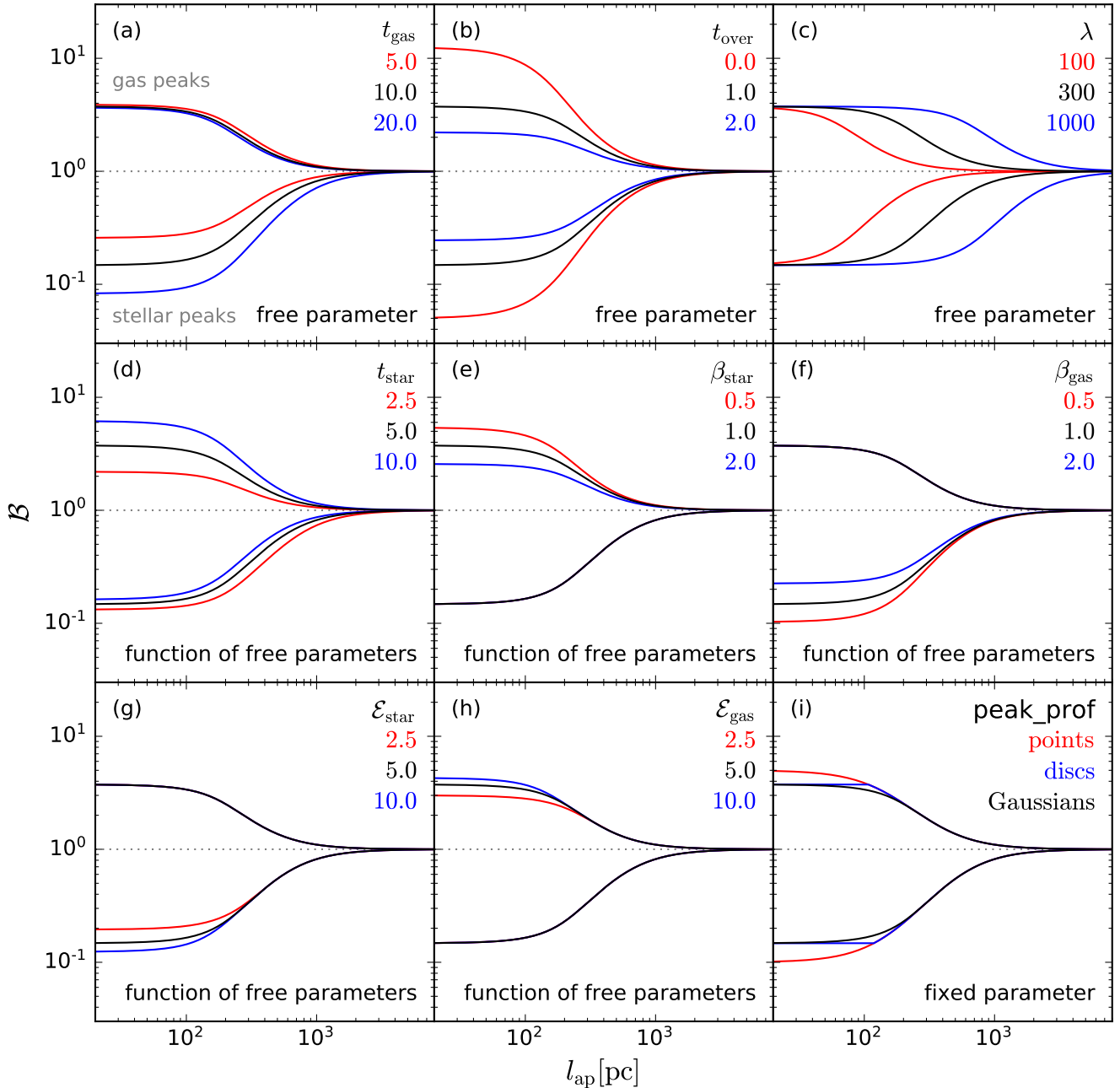
We briefly summarise the effects of the three free parameters during the fitting process of our model.

- (i) The ratio between  $t_{\text{gas}}$  and  $t_{\text{star}}$  sets the vertical asymmetry of the model tuning fork diagram.
- (ii) The duration of the overlap phase  $t_{\text{over}}$  sets the flattening of the tuning fork towards small aperture sizes.
- (iii) The mean separation length of independent regions  $\lambda$  sets the horizontal position of the model tuning fork.

The quantities in panels (d)–(h) of Figure 11 depend on the free parameters  $t_{\text{gas}}$ ,  $t_{\text{over}}$ , and  $\lambda$  from the first row. When changing either of these free parameters during the fitting process, their effect on the tuning fork diagram will therefore be accompanied by the change seen in one or more of panels (d)–(h). The duration of the stellar phase  $t_{\text{star}} = t_{\text{star,ref}} + t_{\text{over}}$  is varied in panel (d) and mirrors the behaviour of changing  $t_{\text{gas}}$  in panel (a). In addition, because  $t_{\text{star}}$  increases with  $t_{\text{over}}$ , the increased flattening of the curves towards large values of  $t_{\text{over}}$  is accompanied by a slight increase of the gas-to-stellar flux ratio excess in the top branch and an even smaller inwards shift of the bottom branch. We note that the effect of changing  $t_{\text{star}}$  on the bottom branch in panel (d) is larger than that of changing  $t_{\text{gas}}$  on the top branch in panel (a). This is caused by the fact that we have set  $t_{\text{gas}} > t_{\text{star}}$  in this example. As a result of this dissimilarity, the flattening of the top branch in panel (a) is affected more strongly by the high probability that young stellar peaks reside in the overlap phase and are therefore included when focusing on gas peaks, which does not depend on  $t_{\text{gas}}$ , whereas the flattening of the bottom branch in panel (d) is affected more strongly by the flux density contrast of the central peak relative to its immediate surroundings, which does depend on  $t_{\text{star}}$ . This difference can be quantitatively verified by inspection of equations (98) and (99).

Panel (e) shows the influence of changing the overlap-to-isolated flux ratio of stellar emission. Unsurprisingly, this only affects the top branch, where the incidence of stellar emission from gas peaks in the overlap phase flattens the tuning fork. This flattening becomes stronger if the stellar emission from peaks in the overlap phase is brighter than from those in the isolated phase (i.e.  $\beta_{\text{star}} > 1$ ) and weakens if the stellar emission from peaks in the overlap phase is fainter (i.e.  $\beta_{\text{gas}} < 1$ ). The same behaviour is seen for the bottom branch in Panel (f), which shows the influence of changing the overlap-to-isolated flux ratio of gas emission. Be-





**Figure 11.** Predicted gas-to-stellar flux ratio bias as a function of the aperture size when focusing apertures on stellar peaks (bottom branches) or gas peaks (top branches), as indicated by the grey labels in panel (a). In each panel, the effect of varying one model quantity is shown relative to the fiducial parameter set (see Table 3). The values in the top right corner of each panel show the values used, with the fiducial choice shown in black. In the bottom right corner of each panel, we list the nature of each quantity. The quantities varied in panels (a)–(c) are all *free parameters* in the model, meaning that they are fitted for in Section 3.2.12. The quantities varied in panels (d)–(h) are all a function of the three free parameters in the top row, implying that while they will be constrained during the fitting process, they are not free parameters as such. Finally, the quantity `peak_prof` that is varied in panel (i) is fixed prior to starting the fitting process. The figure shows that each of the free parameters has a unique effect on the shapes of the curves, implying that they are non-degenerate and can therefore be measured by fitting our model to observed tuning fork diagrams.

cause  $\beta_{\text{star}}$  and  $\beta_{\text{gas}}$  depend on  $t_{\text{over}}/t_{\text{star}}$  and  $t_{\text{over}}/t_{\text{gas}}$ , respectively, a change of the model tuning fork diagram due to any of the three time-scales can be accompanied by a change as seen in panels (e) or (f). This could hamper the accuracy of the method if the corresponding changes in  $\beta_{\text{star}}$  and  $\beta_{\text{gas}}$  would systematically cancel the impact of changing the above time-scales. However, there is no obvious reason why the relations between  $\beta_{\text{star}}$  and  $\beta_{\text{gas}}$  should

scale with these time-scale ratios in a particular direction, or even monotonically. The applications to simulated maps in Section 4, as well as early applications of the method to observational data (e.g. Kruijssen et al. 2018; Hygate et al. in prep.; Chevance et al. in prep.) show no strong systematic patterns in these relations.

If the relation between  $\beta_{\text{star}}(f_{\text{star,over}})$  and  $\beta_{\text{gas}}(f_{\text{gas,over}})$  would be monotonic, we would still not expect any significant

change of the tuning fork diagram due to the propagation of time-scale changes through  $\beta_{\text{star}}$  and  $\beta_{\text{gas}}$ , because  $\beta_{\text{star}}$  and  $\beta_{\text{gas}}$  typically only change by a small amount across the range of  $t_{\text{over}}/t_{\text{star}}$  and  $t_{\text{over}}/t_{\text{gas}}$  (see e.g. Figure 9, where  $\beta_{\text{star}} = 1.3\text{--}1.7$  for  $t_{\text{over}}/t_{\text{star}} = 0.05\text{--}0.85$  and compare this dynamic range to Figure 11). Even if this would occur in individual cases, it is important to note that the effects of  $\beta_{\text{star}}$  and  $\beta_{\text{gas}}$  differ fundamentally from the free parameters in panels (a)–(c). The greatest similarity is between  $t_{\text{gas}}$  and  $\beta_{\text{gas}}$ , both of which affect the vertical position of the bottom branch. In that case, a systematic decrease of  $\beta_{\text{gas}}$  with  $t_{\text{over}}/t_{\text{gas}}$  would mean that the widening of the tuning fork by increasing  $t_{\text{gas}}$  would be compensated by the accompanying decrease of  $t_{\text{over}}/t_{\text{gas}}$ , the resulting increase of  $\beta_{\text{gas}}$ , and the corresponding vertical compression of the tuning fork. In such a situation,  $t_{\text{gas}}$  would be ill-constrained, but this is unlikely to occur for two reasons. Firstly, in the vast majority of the experiments considered in this paper, we find that  $\beta_{\text{star}}$  and  $\beta_{\text{gas}}$  weakly increase with  $t_{\text{over}}/t_{\text{star}}$  and  $t_{\text{over}}/t_{\text{gas}}$ , respectively (see e.g. Figure 9). This trend is opposite to what would be needed for changes in the tuning fork diagram due to the evolutionary timeline to be compensated by changes in  $\beta_{\text{star}}$  and  $\beta_{\text{gas}}$ . Secondly, even in this rather specific case, the detailed change of the tuning fork due to varying  $t_{\text{gas}}$  and  $\beta_{\text{gas}}$  differs. The former affects the vertical position of the bottom branch over the entire range of aperture sizes, whereas the latter only affects the part where the overlap phase dominates the flux ratio bias in the aperture (i.e. where  $l_{\text{ap}} < \lambda$ ). In summary, we therefore do not expect any systematic variation of  $\beta_{\text{star}}$  or  $\beta_{\text{gas}}$  with  $t_{\text{over}}/t_{\text{star}}$  or  $t_{\text{over}}/t_{\text{gas}}$  to negatively influence the accuracy with which the free parameters can be constrained.

Panels (g) and (h) demonstrate the influence of changing the flux ratio contrast  $\mathcal{E}_{\text{star}}$  and  $\mathcal{E}_{\text{gas}}$  between the central peak and the flux measured on a size scale  $\lambda$ . The effect of changing these quantities on the model tuning fork diagram is similar to that of  $\beta_{\text{star}}$  and  $\beta_{\text{gas}}$ , with the key difference that  $\mathcal{E}_{\text{star}}$  and  $\mathcal{E}_{\text{gas}}$  are functions of  $\lambda$  rather than  $t_{\text{over}}/t_{\text{star}}$  or  $t_{\text{over}}/t_{\text{gas}}$ . We see that increasing the flux ratio contrast of single peaks relative to the background leads to a larger gas-to-stellar flux ratio excess (when focusing on gas peaks) or deficit (when focusing on stellar peaks) at the smallest aperture sizes. This is easily understood, because on these size scales, the central surface density of the peak sets the flux ratio bias. Because  $\mathcal{E}_{\text{star}}$  and  $\mathcal{E}_{\text{gas}}$  depend on  $\lambda$ , a change of  $\lambda$  can be accompanied by a change as seen in panels (g) or (h). As before, this could hamper the accuracy of the method if the corresponding changes in  $\mathcal{E}_{\text{star}}$  and  $\mathcal{E}_{\text{gas}}$  would systematically cancel the impact of changing the separation length  $\lambda$ . Fortunately, Figure 11 clearly shows that the effects of changing  $\lambda$  and  $\mathcal{E}$  are orthogonal – the former causes a horizontal shift of the model tuning fork, while the latter vertically shifts a single branch. Therefore, we do not expect any systematic variation of  $\mathcal{E}_{\text{star}}$  or  $\mathcal{E}_{\text{gas}}$  with  $\lambda$  to negatively influence the accuracy with which the free parameters can be constrained.

The final quantity in panel (i) of Figure 11 is fixed and therefore does not change during the fitting process. However, it is still useful to understand its impact of the gas-to-stellar flux ratio bias as a function of aperture size. Panel (i) shows how the shape of the surface density profile chosen to represent the peaks affects the resulting model tuning fork diagram. Because this only concerns the central peak in each aperture, we expect this choice to affect the shape only for apertures  $l_{\text{ap}} < \lambda$ . The delta function profile, reflecting point-like regions, corresponds to the model from KL14. If we instead represent the regions as constant surface density discs with radius  $r_{\text{peak}}$ , the tuning fork abruptly flattens once  $l_{\text{ap}} < 2r_{\text{peak}}$ , because there the enclosed surface density no longer increases with

decreasing aperture size. Adopting a two-dimensional Gaussian profile results in the same central gas-to-stellar flux ratio bias as for the disc profile, but the additional flattening sets in at larger aperture sizes, because the extended nature of the Gaussian profile leads to flux loss from the aperture already for  $l_{\text{ap}} > 2r_{\text{peak}}$ . While the effective change from the point particle representation to the two-dimensional Gaussian may seem minor, it is critical for accurately retrieving the duration of the overlap phase. As we will discuss in Section 3.2.14, this change is therefore instrumental for deriving any feedback-related physical quantities.

Thanks to its purely analytical form, the derived model for the gas-to-stellar flux ratio bias is easy to interpret and readily applicable to the type of observational measurement made in Sections 3.2.1–3.2.10. By fitting the model to observed ‘tuning fork diagrams’, the six quantities in the top two rows of Figure 11 can be obtained, which provide a direct measurement of the evolutionary timeline of independent star-forming regions. With this timeline in hand, it is possible to derive a broad range of other quantities describing the cloud-scale physics of star formation and feedback.

### 3.2.12 Fitting the model to the data

Having obtained the observed gas-to-stellar flux ratio bias and its uncertainties as a function of the aperture size in Section 3.2.10 and having derived a model to describe its behaviour in Section 3.2.11, we now describe how the model is fitted to the data. The best fit is obtained by minimising the goodness-of-fit statistic  $\chi^2_{\text{red}}$  over  $\{t_{\text{gas}}, t_{\text{over}}, \lambda\}$  space. We then marginalise the resulting three-dimensional PDF to obtain the one-dimensional PDF of each free parameter (see Section 3.2.13). At the end of the fitting process, a figure of the best-fitting tuning fork diagram is written to disk, together with ASCII tables containing the data points, their error bars, and the best-fitting model.

Before performing the fit, we need to define the arrays describing the values of the free parameters that are evaluated. Because the parameter space under consideration is three-dimensional, increasing the number of elements  $N_{\text{try}}$  in these arrays significantly slows down the fitting procedure. For that reason, we attain the desired precision of the fit not by using a high value of  $N_{\text{try}}$ , but by iteratively ‘zooming in’ on the best-fitting part of parameter space and refining the free parameter arrays. We define  $N_{\text{depth}}$  as the maximum number of zooms (including the first, unrefined fit), with a default value of  $N_{\text{depth}} = 4$  (see Table 2). Note that most fits do not reach this number of iterations, but finish earlier to avoid omitting too much of parameter space (see below). When using the default number of  $N_{\text{try}} = 101$  array elements per free parameter (and thus  $N_{\text{try}}^3$  elements in total), this approach allows the fit to reach a precision that falls comfortably within the error bars on each free parameter.

The free parameter arrays are defined as follows. For the duration of the gas phase we use

$$t_{\text{gas}}(i) = t_{\text{gas},\text{min}} \left( \frac{t_{\text{gas},\text{max}}}{t_{\text{gas},\text{min}}} \right)^{\frac{i+1/2}{N_{\text{try}}}}, \quad (102)$$

where  $\{i \in \mathbb{N} \mid 0 \leq i \leq N_{\text{try}} - 1\}$  and the range of  $t_{\text{gas}}$  is specified in the input parameter file (see Table 2), with default limits of  $t_{\text{gas},\text{min}} = 0.1$  Myr and  $t_{\text{gas},\text{max}} = 5000$  Myr. Note that we do not allow  $t_{\text{gas},\text{max}}$  to exceed the gas depletion time  $t_{\text{depl}} = M_{\text{gas}}/\text{SFR}$ , which would be unphysical. If  $t_{\text{depl}} < t_{\text{gas},\text{max}}$ , we set  $t_{\text{gas},\text{max}} = t_{\text{depl}}$ . Likewise, for the duration of the overlap phase



we define

$$t_{\text{over}}(j) = t_{\text{over,min}} \left( \frac{t_{\text{over,max}}}{t_{\text{over,min}}} \right)^{\frac{j+1/2}{N_{\text{try}}}}, \quad (103)$$

where  $\{j \in \mathbb{N} \mid 0 \leq j \leq N_{\text{try}} - 1\}$  and the default value of  $t_{\text{over,min}} = 0.01$  Myr (see Table 2) effectively corresponds to  $t_{\text{over}} = 0$ , but is chosen to avoid divergence in the model expressions for the gas-to-stellar flux ratio bias (see Section 3.2.11). The upper limit of the range of  $t_{\text{over}}$  follows from the fact that the duration of the overlap phase cannot exceed the lifetime of the stellar or gas tracer. We therefore set  $t_{\text{over,max}} = t_{\text{gas,max}}$  in the default case of `tstar_incl = 0` (i.e.  $t_{\text{star}}$  does not include the duration of the overlap phase) and  $t_{\text{over,max}} = \min(t_{\text{star}}, t_{\text{gas,max}})$  if `tstar_incl = 1`. Finally, for the mean separation length we use

$$\lambda(k) = \lambda_{\text{min}} \left( \frac{\lambda_{\text{max}}}{\lambda_{\text{min}}} \right)^{\frac{k+1/2}{N_{\text{try}}}}, \quad (104)$$

where  $\{k \in \mathbb{N} \mid 0 \leq k \leq N_{\text{try}} - 1\}$  and we define the limits as  $\lambda_{\text{min}} = 0.3 \min(l_{\text{ap,star}}, l_{\text{ap,gas}})$  and  $\lambda_{\text{max}} = 3 \max(l_{\text{ap,star}}, l_{\text{ap,gas}})$ . For all three free parameters, the defined limits effectively enclose (even more than) the physically reasonable part of parameter space.

By defining the three arrays in equations (102)–(104) as we have, the lowest and highest values in the array do not correspond to the specified limits. This is caused by the addition of  $1/2$  in each of the exponents. However, this offset of half a logarithmic separation step enables the straightforward integration of the PDF using these same arrays. In logarithmic space, the integration steps are constant and the array elements are situated in the centre of each step. The integration steps are defined as

$$\frac{dt_{\text{gas}}(i)}{t_{\text{gas}}(i)} = \left[ \left( \frac{t_{\text{gas,max}}}{t_{\text{gas,min}}} \right)^{\frac{1}{2N_{\text{try}}}} - \left( \frac{t_{\text{gas,max}}}{t_{\text{gas,min}}} \right)^{-\frac{1}{2N_{\text{try}}}} \right], \quad (105)$$

for the duration of the gas phase,

$$\frac{dt_{\text{over}}(j)}{t_{\text{over}}(j)} = \left[ \left( \frac{t_{\text{over,max}}}{t_{\text{over,min}}} \right)^{\frac{1}{2N_{\text{try}}}} - \left( \frac{t_{\text{over,max}}}{t_{\text{over,min}}} \right)^{-\frac{1}{2N_{\text{try}}}} \right], \quad (106)$$

for the duration of the overlap phase, and

$$\frac{d\lambda(k)}{\lambda(k)} = \left[ \left( \frac{\lambda_{\text{max}}}{\lambda_{\text{min}}} \right)^{\frac{1}{2N_{\text{try}}}} - \left( \frac{\lambda_{\text{max}}}{\lambda_{\text{min}}} \right)^{-\frac{1}{2N_{\text{try}}}} \right], \quad (107)$$

for the mean separation length.

Across all  $N_{\text{try}}^3$  elements covered by  $\{t_{\text{gas}}(i), t_{\text{over}}(j), \lambda(k)\}$ , we use equations (81) and (82) to calculate the predicted gas-to-stellar flux ratio bias when focusing on stellar peaks ( $\mathcal{B}_{\text{star}}^{\text{mod}}$ ) or gas peaks ( $\mathcal{B}_{\text{gas}}^{\text{mod}}$ ). If `tstar_incl = 0`, the duration of the stellar phase is defined as  $t_{\text{star}} = t_{\text{star,ref}} + t_{\text{over}}(j)$  and we thus update its value for each  $j$ . Likewise, the values of  $\beta_{\text{star}}$  and  $\beta_{\text{gas}}$  depend on  $f_{\text{star,over}} = t_{\text{over}}(j)/t_{\text{star}}$  and  $f_{\text{gas,over}} = t_{\text{over}}(j)/t_{\text{gas}}(i)$ , respectively. The dependence of both quantities on the time-scales that the fitting process explores is quantified in Section 3.2.9. For each combination of  $t_{\text{gas}}(i)$  and  $t_{\text{over}}(j)$ , we linearly interpolate the grids of  $f_{\text{star,over}}$  and  $f_{\text{gas,over}}$  to obtain the appropriate values of  $\beta_{\text{star}}$  and  $\beta_{\text{gas}}$ . In the same way, the values of  $\mathcal{E}_{\text{star}}$  and  $\mathcal{E}_{\text{gas}}$  depend on  $\lambda(k)$ . These dependences are quantified in equations (44) and (45). We interpolate  $\mathcal{E}(\lambda)$  in double-logarithmic space to obtain the values of  $\mathcal{E}_{\text{star}}$  and  $\mathcal{E}_{\text{gas}}$  for each explored value of  $\lambda(k)$ .

The above process provides us with the predicted tuning fork diagram for each point in parameter space, i.e.  $\mathcal{B}_{\text{star}}^{\text{mod}}(i, j, k, l_{\text{ap}})$  and  $\mathcal{B}_{\text{gas}}^{\text{mod}}(i, j, k, l_{\text{ap}})$ .

For each point in parameter space, the goodness-of-fit statistic  $\chi_{\text{red}}^2$  is calculated by evaluating the difference between the observations and the model prediction in logarithmic space. We define this difference as

$$\Delta_{\text{star}}[l_{\text{ap,star}}(l)] = \left\{ \frac{\log_{10}[\mathcal{B}_{\text{star}}(l_{\text{ap,star}})/\mathcal{B}_{\text{star}}^{\text{mod}}(l_{\text{ap,star}})]}{\sigma_{\log_{10} \mathcal{B}_{\text{star}}(l_{\text{ap,star}})}} \right\}^2, \quad (108)$$

for the gas-to-stellar flux ratio bias when focusing on stellar peaks (bottom branch in the tuning fork diagram) and

$$\Delta_{\text{gas}}[l_{\text{ap,gas}}(l)] = \left\{ \frac{\log_{10}[\mathcal{B}_{\text{gas}}(l_{\text{ap,gas}})/\mathcal{B}_{\text{gas}}^{\text{mod}}(l_{\text{ap,gas}})]}{\sigma_{\log_{10} \mathcal{B}_{\text{gas}}(l_{\text{ap,gas}})}} \right\}^2, \quad (109)$$

for the gas-to-stellar flux ratio bias when focusing on gas peaks (top branch in the tuning fork diagram). In both expressions, we have  $\{l \in \mathbb{N} \mid 0 \leq l \leq N_{\text{ap}} - 1\}$ , analogously to equation (3). The  $\chi^2$  statistic of the point in parameter space then follows as

$$\chi^2 = \sum_{l=0}^{N_{\text{ap}}-1} \Delta_{\text{star}}[l_{\text{ap,star}}(l)] w_{\text{star}}(l) + \Delta_{\text{gas}}[l_{\text{ap,gas}}(l)] w_{\text{gas}}(l), \quad (110)$$

where  $w_{\text{star}}(l)$  and  $w_{\text{gas}}(l)$  represent the weights of the data points on the bottom and top branches of the tuning fork diagram, respectively. These weights were briefly mentioned in Section 3.2.10 and here we define them formally. As discussed in Section 3.2.10, none of the data points in the tuning fork diagram is truly independent, because they reflect the flux ratios in apertures of different sizes that are focused on the same emission peaks. We defined the ‘independence fractions’  $f_{\text{data,star}}(l)$  and  $f_{\text{data,gas}}(l)$  of the data points in equations (70) and (71). These quantities represent the fraction of each data point that is truly independent from all other data points. When calculating the goodness-of-fit statistic, we should weigh the difference between model and observation for each data point by its independence fraction, so that highly independent data points contribute more strongly than correlated data points. This ensures that each independent measurement has the same total contribution to the goodness-of-fit statistic. The weights are defined by normalising the independence fractions to their mean value, i.e.

$$w_{\text{star}}(l) = f_{\text{data,star}} \left( \frac{N_{\text{indep}}}{2N_{\text{ap}}} \right)^{-1}, \quad (111)$$

and

$$w_{\text{gas}}(l) = f_{\text{data,gas}} \left( \frac{N_{\text{indep}}}{2N_{\text{ap}}} \right)^{-1}. \quad (112)$$

In these weights,  $N_{\text{indep}}$  represents the total number of independent data points as defined in equation (74), which is the sum of the independence fractions across both branches of the tuning fork diagram. The quantity  $N_{\text{indep}}/2N_{\text{ap}}$  thus represents the mean independence fraction across all data points. By normalising the independence fractions to their mean value, we guarantee that the mean weight is unity. That way, any change of  $\chi^2$  due to weighing the data points in equation (110) only reflects a change of their relative contributions, rather than a change of the absolute number of data points. This absolute number is decreased by independence

fractions smaller than unity, but that effect is accounted for by decreasing the number of degrees of freedom below, not by modifying the weights of the data points.

The  $\chi^2$  statistic is converted to the goodness-of-fit statistic  $\chi_{\text{red}}^2$  by division by the number of degrees of freedom  $N_{\text{deg}}$ , i.e.

$$\chi_{\text{red}}^2(i, j, k) = \frac{\chi^2(i, j, k)}{N_{\text{deg}}}, \quad (113)$$

where the indices  $(i, j, k)$  have been added to emphasize that the described process is repeated for each point in parameter space. The number of degrees of freedom is defined as

$$N_{\text{deg}} = N_{\text{indep}} - N_{\text{fit}}. \quad (114)$$

In this expression, the number of free parameters is  $N_{\text{fit}} = 3$ . If none of the data points in the tuning fork diagram would have been correlated, then  $N_{\text{indep}} = 2N_{\text{ap}}$  by definition. However, as discussed at length, the fact that the data points are not independent implies that  $N_{\text{indep}} < 2N_{\text{ap}}$ . This leads to a decrease of the number of degrees of freedom, which in turn causes the goodness-of-fit statistic  $\chi_{\text{red}}^2$  to increase.

The best-fitting solution across the covered parameter space corresponds to the element where  $\chi_{\text{red}}^2$  is minimal, i.e.

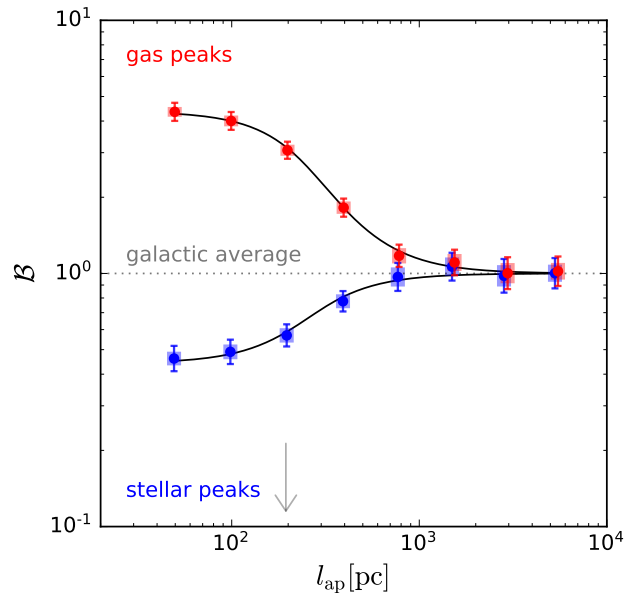
$$\chi_{\text{red}, \text{min}}^2 \equiv \min[\chi_{\text{red}}^2(i, j, k)], \quad (115)$$

with best-fitting values of  $\{t_{\text{gas}}(i_{\text{best}}), t_{\text{over}}(j_{\text{best}}), \lambda(k_{\text{best}})\}$ . However, this best fit may be improved by refining the free parameter arrays in the space around the best fit. A refinement step consists of first updating the minimum and maximum parameter values in equations (102)–(104), defining the refined free parameter arrays, and then repeating the fitting process. This way, we carry out a maximum of  $N_{\text{depth}}$  fitting loops. The criterion for refinement makes use of the one-dimensional PDFs of the free parameters, which are described in the error analysis of Section 3.2.13 below. We therefore defer the quantitative description of the refinement process to that discussion.

After the last refinement step is completed, the final best-fitting solution has been obtained. HEISENBERG outputs a figure of the tuning fork diagram, figures of  $\beta_{\text{star}}$  and  $\beta_{\text{gas}}$  as a function of the time fraction spent in the overlap phase (i.e. Figure 9), as well as ASCII tables that contain the data points, their error bars, and the best-fitting model. Figure 12 shows the best-fitting solution for the example maps used throughout this section. The figure demonstrates that the method provides an accurate representation of the ‘observed’ gas-to-stellar flux ratio bias, which in this case is reflected in the goodness-of-fit statistic of  $\chi_{\text{red}, \text{min}}^2 = 0.20$ . The best-fitting quantities are  $t_{\text{gas}} = 3.06$  Myr,  $t_{\text{over}} = 0.90$  Myr, and  $\lambda = 194$  pc, whereas the additional quantities that are also constrained during the fitting process are  $t_{\text{star}} = 10.90$  Myr,  $\beta_{\text{star}} = 1.45$  (as was already shown in Figure 9),  $\beta_{\text{gas}} = 0.44$ ,  $\mathcal{E}_{\text{star}} = 1.34$ , and  $\mathcal{E}_{\text{gas}} = 1.49$ . We emphasize that these numbers have little-to-no physical meaning, because we applied the method to a simulated data set in this example. In Section 3.2.13, we use these numbers to show that the relative uncertainties on the best-fitting quantities are small, implying that the fitting process provides precise measurements of these quantities.

### 3.2.13 Error analysis and parameter space refinement

After having performed a first fit using the unrefined free parameter arrays of equations (102)–(104), equation (113) provides the goodness-of-fit statistic for each element in parameter space. It is



**Figure 12.** Best-fitting model to the example tuning fork diagram of Figure 10. When visually assessing the quality of the fit, the reader should use the shaded areas as the effective error bars, because they account for the correlation between the data points and the resulting decrease of the number of degrees of freedom (see the text in Sections 3.2.10 and 3.2.12). By fitting the model of Section 3.2.11, we obtain the best-fitting values and their uncertainties of the lifetimes of gas peaks  $t_{\text{gas}}$ , the coexistence time-scale of gas and stellar peaks  $t_{\text{over}}$ , and the characteristic peak separation length  $\lambda$ . The aperture size where  $l_{\text{ap}} = \lambda$  is highlighted by the grey arrow.

straightforward to convert this statistic to a three-dimensional PDF of the free parameters by writing

$$P_{\text{tot}}(t_{\text{gas}}, t_{\text{over}}, \lambda) \equiv \frac{d^3p}{dt_{\text{gas}}dt_{\text{over}}d\lambda} = I_{\chi}^{-1} e^{-\chi_{\text{red}}^2/2}, \quad (116)$$

with a normalisation given by

$$I_{\chi} = \iiint e^{-\chi_{\text{red}}^2/2} dt_{\text{gas}} dt_{\text{over}} d\lambda \\ = \sum_{i,j,k} e^{-\chi_{\text{red}}^2(i,j,k)/2} dt_{\text{gas}}(i) dt_{\text{over}}(j) d\lambda(k), \quad (117)$$

where the second equality accounts for the discretisation of parameter space into the free parameter arrays of equations (102)–(104). The one-dimensional, differential PDFs of the individual free parameters then follow trivially, for the duration of the gas phase:

$$P_{1D}(t_{\text{gas}}) \equiv \frac{dp}{dt_{\text{gas}}} = \iint P_{\text{tot}}(t_{\text{gas}}, t_{\text{over}}, \lambda) dt_{\text{over}} d\lambda, \quad (118)$$

for the duration of the overlap phase:

$$P_{1D}(t_{\text{over}}) \equiv \frac{dp}{dt_{\text{over}}} = \iint P_{\text{tot}}(t_{\text{gas}}, t_{\text{over}}, \lambda) dt_{\text{gas}} d\lambda, \quad (119)$$

and for the region separation length:

$$P_{1D}(\lambda) \equiv \frac{dp}{d\lambda} = \iint P_{\text{tot}}(t_{\text{gas}}, t_{\text{over}}, \lambda) dt_{\text{gas}} dt_{\text{over}}. \quad (120)$$

These are converted to their cumulative forms by defining

$$P_{1D}(X \leq x) \equiv \int_{x_{\text{min}}}^x P_{1D}(X) dX, \quad (121)$$

for each of the PDFs in equations (118)–(120). The numerical implementations in HEISENBERG of the six (differential and cumulative) PDFs described by equations (118)–(121) are discretised

analogously to the second equality of equation (117). Finally, we also calculate the three different two-dimensional PDFs by separately integrating out each single variable in equation (116). These two-dimensional distributions can be used to visualise the degree of correlation and thus degeneracy between the three free parameters.

To compute the uncertainties on the best-fitting values of the free parameters, we use the one-dimensional cumulative PDFs. For most of the applications discussed in Section 4 and the early observational applications of the presented method, the differential PDFs are asymmetric and non-Gaussian. One way of appropriately representing this asymmetry would be to define the uncertainties by using the  $1\sigma$  percentiles in a Gaussian, i.e. 16 and 84 per cent. However, this can result in situations where the best-fitting value is found outside of the 16–84 percentile range, leading to negative uncertainties. To avoid such a (rare) situation, we choose to define the uncertainties used in this work differently. Given a best-fitting value  $x_{\text{best}}$ , we divide the variable range in two around  $x_{\text{best}}$  and define the uncertainties analogously to the Gaussian case as the 32<sup>nd</sup> percentile of the part of the PDF below  $x_{\text{best}}$  and the 68<sup>th</sup> percentile of the part of the PDF above  $x_{\text{best}}$ . If the PDF is Gaussian, this corresponds to the formal definition of the  $1\sigma$  uncertainties, but for asymmetric PDFs it deviates.

Mathematically, the above procedure results in the following steps. First, we define the percentiles of the  $1\sigma$  uncertainties in a Gaussian PDF as usual, i.e.

$$p_{\text{min,Gauss}} = 1 - \frac{1}{2} \left[ 1 + \operatorname{erf} \left( \frac{1}{\sqrt{2}} \right) \right] \approx 0.16, \quad (122)$$

for the downward uncertainty, and

$$p_{\text{max,Gauss}} = \frac{1}{2} \left[ 1 + \operatorname{erf} \left( \frac{1}{\sqrt{2}} \right) \right] \approx 0.84, \quad (123)$$

for the upward uncertainty. We can normalise these to the range on either side of the best-fitting value in a Gaussian PDF (i.e.  $p_{\text{best,Gauss}} = 0.5$ ) and thus determine an equivalent fraction of that range corresponding to the above percentiles by writing

$$f_{\text{min,Gauss}} = \frac{p_{\text{min,Gauss}}}{p_{\text{best,Gauss}}} \approx 0.32, \quad (124)$$

for the downward uncertainty, and

$$f_{\text{max,Gauss}} = \frac{p_{\text{max,Gauss}} - p_{\text{best,Gauss}}}{1 - p_{\text{best,Gauss}}} \approx 0.68, \quad (125)$$

for the upward uncertainty. The corresponding percentiles around the best-fitting value  $x_{\text{best}}$  then follow by multiplication with the PDF ranges on either side of  $x_{\text{best}}$ , i.e.

$$p_{\text{min}} = f_{\text{min,Gauss}} P_{1\text{D}}(x \leq x_{\text{best}}) \approx 0.32 P_{1\text{D}}(x \leq x_{\text{best}}), \quad (126)$$

for the downward uncertainty, and

$$p_{\text{max}} = P_{1\text{D}}(x \leq x_{\text{best}}) + f_{\text{max,Gauss}} [1 - P_{1\text{D}}(x \leq x_{\text{best}})] \approx 0.32 P_{1\text{D}}(x \leq x_{\text{best}}) + 0.68, \quad (127)$$

for the upward uncertainty. For the Gaussian value of  $P_{1\text{D}}(x \leq x_{\text{best}}) = 0.5$ , these reduce to equations (122) and (123). The parameter values  $x_-$  and  $x_+$  at the percentiles of equations (126) and (127) are obtained by numerically solving

$$\int_{x_{\text{min}}}^{x_-} P_{1\text{D}}(x) dx = p_{\text{min}}, \quad (128)$$

for  $x_-$  and

$$\int_{x_{\text{min}}}^{x_+} P_{1\text{D}}(x) dx = p_{\text{max}}, \quad (129)$$

for  $x_+$ . This results in downward and upward error bars on the best-fitting values of

$$\sigma_-(x) = x_{\text{best}} - x_-, \quad (130)$$

and

$$\sigma_+(x) = x_+ - x_{\text{best}}, \quad (131)$$

respectively.

The uncertainties obtained through the above procedure are used after each iteration of the fitting process to determine whether it requires another iteration and, if so, within which value range the free parameter arrays of equations (102)–(104) should be re(de)efined. At the end of a complete fitting step with these arrays, we determine the part of the three-dimensional parameter space that corresponds to the  $3\sigma$  confidence ellipsoid and encloses 99.73 per cent of the total probability, with the intention of restricting the fitting range to this volume during the next fitting step. The surface corresponding to this ellipsoid has a goodness-of-fit parameter value of

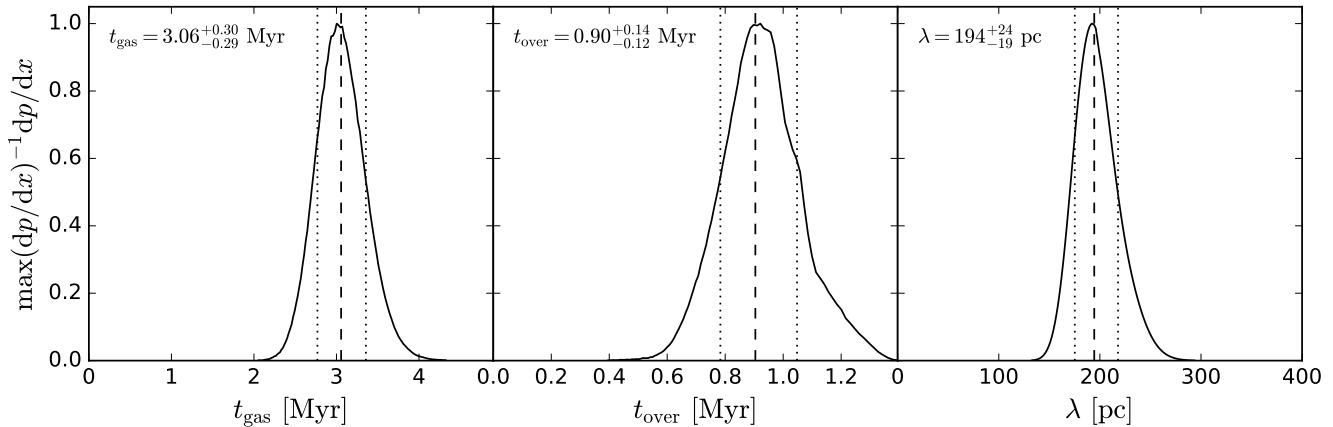
$$\chi_{\text{red}}^2 = \chi_{\text{red,min}}^2 + \frac{\Delta\chi^2}{N_{\text{deg}}}, \quad (132)$$

where  $\Delta\chi^2$  depends on the number of free parameters. The problem under consideration has  $N_{\text{fit}} = 3$ , which for the  $3\sigma$  ellipsoid results in  $\Delta\chi^2 = 14.2$  (Press et al. 1992, chapter 15.6). For each of the free parameter arrays of equations (102)–(104), we then determine at each element what the minimum value of  $\chi_{\text{red}}^2$  is across the full range of both other free parameters. The refined array for the next fitting step is restricted to the range of elements where  $\chi_{\text{red}}^2$  is smaller than the reference value of equation (132), plus one element on either side of this range. This removes any free parameter values for which no elements in the complete, three-dimensional  $\chi_{\text{red}}^2$  array fall within the  $3\sigma$  ellipsoid.

The above steps update the values of  $t_{\text{gas,min}}$ ,  $t_{\text{gas,max}}$ ,  $t_{\text{over,min}}$ ,  $t_{\text{over,max}}$ ,  $\lambda_{\text{min}}$ , and  $\lambda_{\text{max}}$  that are used in equations (102)–(104) for the next fitting step, with the added condition that if  $t_{\text{over,min}} > t_{\text{gas,min}}$ , we set  $t_{\text{over,min}} = t_{\text{gas,min}}$  to avoid the unphysical situation that the minimum duration of the overlap phase exceeds the duration of the gas phase for some part of the range of  $t_{\text{gas}}$ . Even if this enables a next fitting step in principle, we carry out four additional checks of the new free parameter limits, three of which can prevent entering the next fitting step. These checks are as follows.

(i) If the best-fitting value resides within four array elements of the edge of a new parameter space range, which can happen in the rare situation that  $\chi_{\text{red}}^2$  increases very steeply from  $\chi_{\text{red,min}}^2$  to  $\chi_{\text{red,min}}^2 + \Delta\chi^2/N_{\text{deg}}$ , the free parameter range at that edge is expanded. For lower limits, we expand the updated minimum through multiplication by  $x_{\text{min}}/x_{\text{max}}$ , whereas for upper limits, the updated maximum is expanded through multiplication by  $x_{\text{max}}/x_{\text{min}}$ . In either case, the updated minimum (maximum) is not allowed to fall below (exceed) the value set at the beginning of the fitting process, as defined in Section 3.2.12 for equations (102)–(104). This check does not prevent the next fitting step, but it ensures that the best-fitting solution does not end up at the edge of parameter space.

(ii) If both the downward and upward  $1\sigma$  ranges of all free parameters reach within four array elements of the edge of the current, unrefined parameter space, the refinement is aborted and the fitting



**Figure 13.** One-dimensional PDFs of  $t_{\text{gas}}$  (left),  $t_{\text{over}}$  (middle), and  $\lambda$  (right), which are the free parameters in the fit of the tuning fork diagram in Figure 12. The PDFs are normalised to their maximum probability density. Vertical dashed lines indicate the best-fitting values (also indicated in the top left of each panel), whereas the vertical dotted lines represent the  $1\sigma$  uncertainty range as defined in the text. These three PDFs are representative of the PDFs that are typically obtained when applying the method and illustrate that they are close to symmetric and Gaussian.

process is concluded. This check is added as a safety feature against removing relevant parts of parameter space.

(iii) If the absolute logarithmic difference between all six of the old and new parameter space limits is smaller than a certain tolerance value  $\delta_{\text{lim}}$ , another refinement step is deemed unnecessary, because the added precision is negligible. We use a value of  $\delta_{\text{lim}} = 0.01$ , which implies that if the change of all parameter space limits is smaller than 2.3 per cent, the refinement is aborted and the fitting process is concluded.

(iv) Finally, we cap the number of fitting steps at  $N_{\text{depth}}$ , corresponding to a maximum of  $N_{\text{depth}} - 1$  refinements. For the default value of  $N_{\text{depth}} = 4$  (see Table 2, this cap is often not reached. Instead, the refinement is aborted because the absolute logarithmic difference between the old and new limits falls below the specified tolerance (as discussed in the previous check).

The fitting process continues through the cycle of obtaining a fit and refining the free parameter arrays until any of the three exit conditions specified above are met. Once it is completed, we obtain the final best-fitting values and their uncertainties. For the example galaxy maps used throughout this section, these are  $t_{\text{gas}} = 3.06^{+0.30}_{-0.29}$  Myr,  $t_{\text{over}} = 0.90^{+0.14}_{-0.12}$  Myr, and  $\lambda = 194^{+24}_{-19}$  pc. As stated before in Section 3.2.12, due to the nature of the experiment these values have no particular physical meaning, but they highlight that the quantities are very well-constrained, with relative errors of 16 per cent or smaller in all three cases. Finally, HEISENBERG writes the differential one-dimensional PDFs to disk in ASCII format and also outputs figures of all nine free parameter PDFs calculated here, i.e. the three differential one-dimensional PDFs, the three cumulative one-dimensional PDFs, and the three two-dimensional PDFs of each free parameter pair. The best-fitting values and their uncertainties are also indicated in these plots.

For the familiar example maps, the one-dimensional differential PDFs of  $t_{\text{gas}}$ ,  $t_{\text{over}}$ , and  $\lambda$  are shown in Figure 13. These PDFs are well-behaved in that they are quite symmetric and even fairly close to Gaussian, as evidenced by the fact that the  $1\sigma$  lines intersect with the PDFs near the expected  $\exp(-1/2) \approx 0.61$  times the peak value. The kinks in the  $t_{\text{over}}$  PDF result from the discrete nature of the relation between  $\beta_{\text{star}}$ ,  $\beta_{\text{gas}}$ , and  $t_{\text{over}}$  (see Figure 9), which is unavoidable given that any galaxy map hosts a finite number of emission peaks. We note that these PDFs are representative

for the applications of our method in Section 4, as well as the ongoing observational applications. Having said that, it is possible to retrieve asymmetric PDFs if the tuning fork diagram provides less strong constraints than in this example (e.g. when one of the branches remains nearly flat over the full range of aperture sizes, which is indicative of a long duration of the overlap phase). For this reason, we emphasize that future applications of the method should generally present the PDFs rather than the best-fitting values and their uncertainties, especially when the downward and upward uncertainties on one of the best-fitting values are dissimilar.

### 3.2.14 Calculation of derived quantities and error propagation

As the final quantitative step of the method, we use the constraints on the evolutionary timeline of independent star-forming regions and their characteristic separation lengths to derive additional physical quantities describing the star formation and feedback process. Unsurprisingly, having access to the evolutionary time-scales provides a powerful avenue for constraining these physics. By carrying out Monte-Carlo sampling from the complete, three-dimensional PDF of  $t_{\text{gas}}$ ,  $t_{\text{over}}$ , and  $\lambda$  from Section 3.2.13, we self-consistently obtain the one-dimensional PDFs of all derived quantities and fully account for the covariance between the three free parameters. This error propagation process also provides us with the complete covariance matrix between all constrained quantities. In what follows, we first discuss the expressions used to obtain each of the derived quantities, before treating the Monte-Carlo error propagation.

We first describe a set of quantities that represent *byproducts* of the fitting process, in that they play a role in the equations describing the model in Section 3.2.11, but do not act as free parameters in the fitting process, because they are functions of  $t_{\text{gas}}$  and  $t_{\text{over}}$ . In Section 3.2.1, we described that the input file contains the stellar ‘reference time-scale’ used to convert the relative timeline constrained by the method to an absolute timeline. However, we also introduced the flag `tstar.incl = {0,1}`, which encodes whether the reference time-scale excludes or includes the overlap time-scale, respectively. A physical example of the first case would be H $\alpha$  emission, because star-forming regions with a non-zero age spread may continue forming stars while they are already visible in H $\alpha$ , implying that only when star formation has ceased and any residual star-forming gas is expelled, the final massive star has



formed and the stellar evolutionary ‘clock’ of the reference time-scale starts ticking. By contrast, in the numerical experiments carried out in Section 4.2, we use star particles within a certain age range to provide the stellar map and reference time-scale, in which case the reference time-scale is equivalent to this age range and includes the duration of the overlap phase. Mathematically, these cases correspond to

$$t_{\text{star}} = t_{\text{star,ref}} + t_{\text{over}}, \quad (133)$$

if `tstar_incl = 0` and  $t_{\text{star,ref}}$  does not include  $t_{\text{over}}$ , and

$$t_{\text{star}} = t_{\text{star,ref}}, \quad (134)$$

if `tstar_incl = 1` and  $t_{\text{star,ref}}$  does include  $t_{\text{over}}$ . Throughout this paper, we will indicate which definition of  $t_{\text{star,ref}}$  we are using for each experiment. Similarly to  $t_{\text{star}}$ , the total duration of the evolutionary timeline shown in Figure 1 follows by combining the time-scales

$$\tau = t_{\text{star}} + t_{\text{gas}} - t_{\text{over}}, \quad (135)$$

where the subtraction of  $t_{\text{over}}$  is needed to avoid counting it twice, because the duration of the overlap phase is included both in  $t_{\text{star}}$  and  $t_{\text{gas}}$ .

The next quantities that are obtained as byproducts of the fitting process are  $\beta_{\text{star}}$  and  $\beta_{\text{gas}}$ . They are defined by equations (40) and (41), where they are given as a function of  $f_{\text{star,over}} = t_{\text{over}}/t_{\text{star}}$  and  $f_{\text{gas,over}} = t_{\text{over}}/t_{\text{gas}}$ , respectively (also see Figure 9). HEISENBERG stores the numerical relations  $\beta_{\text{star}}(f_{\text{star,over}})$  and  $\beta_{\text{gas}}(f_{\text{gas,over}})$  from Section 3.2.9 and uses these after constraining the best fit to the tuning fork diagram to determine the values of  $\beta_{\text{star}}$  and  $\beta_{\text{gas}}$ . The PDFs of these quantities are determined below using the same relations in the Monte-Carlo error propagation process.

The third set of quantities that are constrained as byproducts of the fitting process are  $\mathcal{E}_{\text{star}}$ ,  $\mathcal{E}_{\text{gas}}$ ,  $\mathcal{E}_{\text{star,glob}}$ , and  $\mathcal{E}_{\text{gas,glob}}$ , which are defined in equations (44)–(47) as a function of a size scale  $l_{\text{ap}}$  that is set to  $l_{\text{ap}} = \lambda$  in Section 3.2.11. As with  $\beta_{\text{star}}$  and  $\beta_{\text{gas}}$ , HEISENBERG stores the relations between these quantities and  $\lambda$  and uses these after obtaining the best-fitting value of  $\lambda$  to determine the corresponding values of  $\mathcal{E}_{\text{star}}$ ,  $\mathcal{E}_{\text{gas}}$ ,  $\mathcal{E}_{\text{star,glob}}$ , and  $\mathcal{E}_{\text{gas,glob}}$ . These relations are also used for determining the PDFs of these quantities by Monte-Carlo error propagation.

We now turn to a number of *derived* quantities that are themselves not used directly in the fitting process, but follow immediately from the quantities determined so far. Firstly, the radii of the stellar and gas emission peaks,  $r_{\text{star}}$  and  $r_{\text{gas}}$ , are defined by equations (92)–(95), depending on whether it is assumed that the peak profiles follow constant surface density discs or two-dimensional Gaussians. Throughout this paper, we adopt the latter case, as this is more representative of typical emission peak profiles, both in observations and in our simulated maps (see Sections 3.2.11 and 4). From these radii, we immediately derive the relative filling factors of the peaks, i.e. their diameters in units of the mean separation length of independent regions. For the stellar peaks, we write this quantity as

$$\zeta_{\text{star}} = \frac{2r_{\text{star}}}{\lambda}, \quad (136)$$

while for the gas peaks this becomes

$$\zeta_{\text{gas}} = \frac{2r_{\text{gas}}}{\lambda}. \quad (137)$$

In Section 4, we find that a reliable measurement of the duration of the overlap phase  $t_{\text{over}}$  and, hence, the characterisation of

feedback-related quantities requires  $\zeta_{\text{star}}$  and  $\zeta_{\text{gas}}$  to be lower than a certain maximum value (also see Appendix B). These quantities therefore represent an important benchmark in assessing the accuracy of the method’s results.

Finally, by combining the mean separation length and the duration of the overlap phase, we obtain the characteristic velocity  $v_{\text{fb}}$  by which the residual gas is removed from independent star-forming regions by feedback:

$$v_{\text{fb}} = \frac{\lambda}{2t_{\text{over}}}. \quad (138)$$

The physical meaning of this velocity depends on the feedback mechanism responsible for removing the gas. If the gas is kinetically removed, the feedback velocity simply represents the outflow velocity averaged over the distance from the feedback source to half the region separation length  $\lambda/2$ . However, if the gas tracer is rendered invisible by a change in phase or emissivity (e.g. through photoevaporation, ionization, excitation conditions, or a chemical reaction), then the feedback velocity corresponds to the velocity of the phase transition or emissivity front. In that case, the relevant length scale is not the region separation length, but the gas region radius  $r_{\text{gas}}$ . The same applies if the gas becomes undetectable by dilution once the shell escapes the cloud boundary. These situations may arise in regions of low gas surface density, where molecular clouds are embedded in an atomic-dominated medium (e.g. Blitz & Rosolowsky 2006; Krumholz et al. 2009). The feedback velocity is then defined as

$$v_{\text{fb},r} = \frac{r_{\text{gas}}}{t_{\text{over}}}. \quad (139)$$

Both feedback velocities are calculated by HEISENBERG. Most generally,  $v_{\text{fb}}$  and  $v_{\text{fb},r}$  bracket the range of possible feedback velocities, but distinguishing between these is possible through a variety of metrics. In the context of the presented method, the duration of the overlap phase  $t_{\text{over}}$  may rule out the importance of certain feedback mechanisms, e.g. when the gas is removed on a time-scale shorter than 3 Myr, it is certain that core-collapse SNe are not responsible (e.g. Ekström et al. 2012). In addition, we specify the feedback energy and momentum coupling efficiencies in equations (148), (149), (152) and (153) below, which indicate whether a given feedback mechanism provides the energy or momentum required for the removal of the gas over length scales  $\lambda/2$  or  $r_{\text{gas}}$ . Independent constraints from ancillary observations can also help. For instance, it is possible to directly measure the expansion velocities of shell-like features around individual star-forming regions. Comparing these to  $v_{\text{fb}}$  as defined above provides a conclusive way of identifying the relevant length scale – if the typical expansion velocity is similar to  $v_{\text{fb},r}$  rather than  $v_{\text{fb}}$ , this means that the gas tracer vanishes at a radius  $r_{\text{gas}}$  rather than  $\lambda/2$ , due to dilution or a change in phase or emissivity.

The final set of derived physical quantities is referred to as *composite* quantities, because they depend on additional quantities that are not directly constrained by the method itself. In particular, by combining the obtained time-scales and size scales with the absolute SFR and gas mass (or their surface densities), we obtain several of the key quantities describing cloud-scale star formation and feedback. This is an important implication of the presented method, albeit not necessarily surprising. Recall the example given in equation (1), which showed that the degeneracy between the star formation time-scale and the star formation efficiency can be lifted by directly measuring one of these. This is exactly what we have done in the method presented above.

Firstly, we define the SFR surface density as

$$\Sigma_{\text{SFR}} = \frac{\text{SFR}}{A_{\text{tot}}}, \quad (140)$$

which uses the total area and the SFR defined in equations (10) and (13). The uncertainty on  $\Sigma_{\text{SFR}}$  is assumed to be dominated by the uncertainty of the conversion factor  $X_{\text{star}}$  and we therefore set the relative uncertainty  $\sigma_{\text{rel}}(\Sigma_{\text{SFR}}) = \sigma_{\text{rel}}(X_{\text{star}})$ . Likewise, the gas surface density is defined as

$$\Sigma_{\text{gas}} = \frac{M_{\text{gas}}}{A_{\text{tot}}}, \quad (141)$$

which uses the total area and the gas mass defined in equations (10) and (14). Again, we assume that the uncertainty on the conversion factor  $X_{\text{gas}}$  dominates and therefore set the relative uncertainty  $\sigma_{\text{rel}}(M_{\text{gas}}) = \sigma_{\text{rel}}(X_{\text{gas}})$ . The ratio of the gas mass and the SFR (surface densities) corresponds to the gas depletion time

$$t_{\text{depl}} = \frac{\Sigma_{\text{gas}}}{\Sigma_{\text{SFR}}}, \quad (142)$$

which represents the time it takes to exhaust the presently available gas reservoir at the present SFR.

As shown by equation (1), the ratio between the gas mass and the SFR (surface densities) provides the ratio between the star formation efficiency and the star formation time-scale. Given that our method provides the duration of the gas phase (over which star formation can proceed), we obtain the mean star formation efficiency per unit cloud lifetime by dividing the surface density of the formed stars  $t_{\text{gas}}\Sigma_{\text{SFR}}$  by the gas surface density:

$$\epsilon_{\text{sf}} = \frac{t_{\text{gas}}\Sigma_{\text{SFR}}}{\Sigma_{\text{gas}}} = \frac{t_{\text{gas}}}{t_{\text{depl}}}, \quad (143)$$

where the second equality shows that this is equivalent to the ratio between the duration of the gas phase (or cloud lifetime) and the gas depletion time. By constraining the evolutionary timeline of cloud-scale star formation, we can now determine whether star formation is rapid and inefficient or slow and efficient.

The expressions derived so far allow us to formulate a set of important dimensionless quantities describing cloud-scale feedback. These quantities are similar to those used in galaxy formation simulations, but here they describe feedback on a size scale  $\lambda$ , whereas in simulations they are often considered on a larger spatial scale. Firstly, a comparison of the instantaneous SFR and outflow rates per region of equations (154) and (155) provide the instantaneous mass loading factor, i.e. the outflow rate during the overlap phase in units of the SFR during the gas phase:

$$\eta_{\text{fb}} = \frac{\dot{M}_{\text{fb}}}{\dot{M}_{\text{sf}}} = \frac{(1 - \epsilon_{\text{sf}})t_{\text{gas}}}{\epsilon_{\text{sf}}t_{\text{over}}} = \frac{t_{\text{depl}} - t_{\text{gas}}}{t_{\text{over}}}. \quad (144)$$

This expression does not reflect the total mass budget, because it considers the instantaneous *rates* of star formation and mass outflow, both of which take place on different time-scales ( $t_{\text{gas}}$  and  $t_{\text{over}}$ ). By contrast, the time-averaged mass loading factor does reflect the total mass budget, i.e. it balances the total mass ejected by feedback  $(1 - \epsilon_{\text{sf}})\Sigma_{\text{gas}}$  by the total stellar mass formed  $t_{\text{gas}}\Sigma_{\text{SFR}}$  and thus becomes a simple ratio of either star formation efficiencies or time-scales:

$$\bar{\eta}_{\text{fb}} = \frac{(1 - \epsilon_{\text{sf}})\Sigma_{\text{gas}}}{t_{\text{gas}}\Sigma_{\text{SFR}}} = \frac{1 - \epsilon_{\text{sf}}}{\epsilon_{\text{sf}}} = \frac{t_{\text{depl}}}{t_{\text{gas}}} - 1. \quad (145)$$

While the mass loading factor of equation (144) may be appropriate for comparison to observational estimates of the mass loading factor based on observed mass outflow rates and SFRs, the definition in equation (145) is appropriate for comparison to theoretical

models, which often consider the total ejected mass in terms of the total stellar mass (e.g. Krumholz et al. 2017).

Likewise, it is possible to compare the total energy and momentum imparted by feedback on the ISM to the total injected energy and momentum, resulting in the feedback energy and momentum coupling efficiencies  $\chi_{\text{fb},E}$  and  $\chi_{\text{fb},p}$ . The total imparted kinetic energy per unit area is given by

$$E_{\text{fb}} = \frac{1}{2}(1 - \epsilon_{\text{sf}})\Sigma_{\text{gas}}v^2, \quad (146)$$

where  $v$  is  $v_{\text{fb}}$  or  $v_{\text{fb},r}$ . The total injected energy per unit area is

$$E_{\text{tot}} = \epsilon_{\text{sf}}\Sigma_{\text{gas}}\Psi_E t_{\text{over}}, \quad (147)$$

with  $\Psi_E$  the energy output rate per unit mass (i.e. the light-to-mass ratio). The first half of this expression ( $\epsilon_{\text{sf}}\Sigma_{\text{gas}}$ ) represents the total mass surface density of new-born stars, whereas the second half ( $\Psi_E t_{\text{over}}$ ) represents the total energy output per unit mass. With these two energy budgets in hand, it is straightforward to define the feedback energy efficiency as the ratio between  $E_{\text{fb}}$  and  $E_{\text{tot}}$ :

$$\chi_{\text{fb},E} = \frac{E_{\text{fb}}}{E_{\text{tot}}} = \frac{(1 - \epsilon_{\text{sf}})v_{\text{fb}}^2}{2\epsilon_{\text{sf}}\Psi_E t_{\text{over}}}, \quad (148)$$

for expansion out to  $\lambda/2$  and

$$\chi_{\text{fb},E,r} = \frac{E_{\text{fb}}}{E_{\text{tot}}} = \frac{(1 - \epsilon_{\text{sf}})v_{\text{fb},r}^2}{2\epsilon_{\text{sf}}\Psi_E t_{\text{over}}}, \quad (149)$$

if the gas vanishes already at  $r_{\text{gas}}$  by dilution or a change in phase or emissivity. Clearly, the feedback energy efficiency depends on the light-to-mass ratio  $\Psi_E$ , which should be specified. However, the appropriate value of  $\Psi_E$  depends on the problem under consideration. With the output from stellar population synthesis models, it is possible to define this parameter for individual feedback mechanisms or any combination thereof. Typical values range from  $\Psi_E = 10^{-4} - 10^0 \text{ m}^2 \text{ s}^{-3}$ , which covers core-collapse SNe, stellar winds, and radiative feedback (e.g. Agertz et al. 2013). A feedback energy efficiency in excess of unity ( $\chi_{\text{fb},E} > 1$ ) indicates that the feedback mechanisms accounted for by the adopted value of  $\Psi_E$  are insufficient to drive the observed feedback energy imparted on the ISM, whereas  $\chi_{\text{fb},E} < 1$  provides constraints on the inefficiency of feedback energy coupling to the ISM due to e.g. the porosity of the ISM or radiative losses.

Analogously to the total imparted kinetic energy per unit area, we define the total imparted momentum per unit area as

$$p_{\text{fb}} = (1 - \epsilon_{\text{sf}})\Sigma_{\text{gas}}v, \quad (150)$$

where  $v$  again refers to  $v_{\text{fb}}$  or  $v_{\text{fb},r}$ . The total injected momentum per unit area is

$$p_{\text{tot}} = \epsilon_{\text{sf}}\Sigma_{\text{gas}}\psi_p t_{\text{over}}, \quad (151)$$

with  $\psi_p$  the momentum output rate per unit mass, which has units of acceleration. As in equation (147), the first half of this expression ( $\epsilon_{\text{sf}}\Sigma_{\text{gas}}$ ) represents the total mass surface density of new-born stars, whereas the second half ( $\psi_p t_{\text{over}}$ ) represents the total momentum output per unit mass. With these two momentum budgets in hand, it is straightforward to define the feedback momentum efficiency as the ratio between  $p_{\text{fb}}$  and  $p_{\text{tot}}$ :

$$\chi_{\text{fb},p} = \frac{p_{\text{fb}}}{p_{\text{tot}}} = \frac{(1 - \epsilon_{\text{sf}})v_{\text{fb}}}{\epsilon_{\text{sf}}\psi_p t_{\text{over}}}. \quad (152)$$

for expansion out to  $\lambda/2$  and

$$\chi_{\text{fb},p,r} = \frac{p_{\text{fb}}}{p_{\text{tot}}} = \frac{(1 - \epsilon_{\text{sf}})v_{\text{fb},r}}{\epsilon_{\text{sf}}\psi_p t_{\text{over}}}. \quad (153)$$

if the gas vanishes already at  $r_{\text{gas}}$  by dilution or a change in phase or emissivity. Evaluating these expressions requires the momentum output rate per unit mass  $\psi_p$  to be specified. As before, the appropriate value can be chosen to represent individual feedback mechanisms or to reflect their combined effect. For stellar winds, typical values are  $\psi_p = 10^{-11}$ – $10^{-10}$  m s $^{-2}$  (e.g. Agertz et al. 2013). Analogously to the feedback energy efficiency, a feedback momentum efficiency in excess of unity ( $\chi_{\text{fb},p} > 1$ ) indicates that the feedback mechanisms accounted for by the adopted value of  $\psi_p$  are insufficient to drive the observed feedback momentum imparted on the ISM, whereas  $\chi_{\text{fb},p} < 1$  provides constraints on the inefficiency of the feedback momentum coupling to the ISM due to e.g. the porosity of the ISM or opposing forces (such as the ambient gas pressure).

In general, a comparison of the measured  $\chi_{\text{fb},E}$  and  $\chi_{\text{fb},p}$  helps determine whether outflow is energy-driven or momentum-driven. If  $\Psi_E$  and  $\psi_p$  are chosen such that they contain all possible contributions to the injected energy and momentum and either  $\chi_{\text{fb},E}$  or  $\chi_{\text{fb},p}$  exceeds unity, this indicates that the *other* driving process dominates, because the observed outflow cannot be generated when  $\chi_{\text{fb}} > 1$ . When both  $\chi_{\text{fb},E} \gtrsim 1$  and  $\chi_{\text{fb},p} \gtrsim 1$ , this suggests that the gas is rendered invisible at a radius  $r_{\text{gas}}$  rather than  $\lambda/2$ , necessitating the use of  $\chi_{\text{fb},E,r}$  and  $\chi_{\text{fb},p,r}$  instead. If all four coupling efficiencies exceed unity, this either points towards a problem in the input parameters chosen for the analysis, leading to inaccurate measurements of  $t_{\text{over}}$ ,  $v_{\text{fb}}$ ,  $v_{\text{fb},r}$ , or  $\epsilon_{\text{sf}}$ , or indicates that the (models motivating the) choice of  $\Psi_E$  and  $\psi_p$  are incorrect or incomplete. The former situation would require the user to revisit the input and correct any mistakes, but the latter case would count as a falsification of the tested feedback description and would hint at the importance of additional feedback mechanisms.

Finally, the obtained characteristic time-scales and size scales governing cloud-scale star formation and feedback can be used to derive absolute properties of the average cloud or star-forming region in the considered maps. This can be done by multiplying the mean surface densities across the maps with the region area  $\pi(\lambda/2)^2$ . However, this requires the strong assumption that the number densities of regions on a size scale  $\lambda$  are similar to that across the entire map, i.e. that the regions are randomly distributed in space. While a similar assumption is made in deriving the model of Section 3.2.11, it is much stronger here. In the general context of the model, we assume a random distribution of independent regions (see Section 7.2.1) on size scales up to a few times the mean separation length  $\lambda$ . On larger size scales, the flux ratio biases of the tuning fork diagram converge to the galactic average, irrespectively of (possibly strong) deviations from a random spatial distribution. However, it is a much stronger assumption to obtain a region mass by multiplying the map-averaged surface density with the region separation length, because this conversion will be affected by large-scale morphological features such as spiral arms, rings, and bars. Due to the associated uncertainty of calculating the absolute properties of independent regions, we intend to consider these quantities in more detail in future work and only provide two examples here.

We determine the typical absolute SFRs and feedback outflow rates of the independent star-forming regions described by our model. The mean separation length  $\lambda$  provides the characteristic size scale within which star formation proceeds, implying a region-averaged SFR during the gas phase of

$$\dot{M}_{\text{sf}} = \pi \left( \frac{\lambda}{2} \right)^2 \Sigma_{\text{SFR}} = \pi \left( \frac{\lambda}{2} \right)^2 \frac{\epsilon_{\text{sf}} \Sigma_{\text{gas}}}{t_{\text{over}}}. \quad (154)$$

The first expression given here is trivial – it simply multiplies the

area-averaged SFR surface density by the area per region to get the SFR per region. The second expression obtains the SFR surface density in terms of gas tracer properties to enable a straightforward comparison to equation (155) below. Analogously to equation (154), we can derive the region-averaged mass outflow (or removal) rate due to feedback during the overlap phase:

$$\dot{M}_{\text{fb}} = \pi \left( \frac{\lambda}{2} \right)^2 \frac{(1 - \epsilon_{\text{sf}}) \Sigma_{\text{gas}}}{t_{\text{over}}}, \quad (155)$$

where the numerator denotes the gas surface density of the residual gas and the denominator represents the feedback time-scale (i.e. the duration of the overlap phase). We reiterate that both equations (154) and (155) assume a homogeneous distribution of independent regions, without large-scale morphological features or a strong spatial clustering of the regions, such that the average surface densities  $\Sigma_{\text{SFR}}$  and  $\Sigma_{\text{gas}}$  yield meaningful SFRs or masses when multiplying by  $\pi(\lambda/2)^2$ . When applying the method to galaxies with prominent morphological features, the risk exists that the global surface density underestimates the surface density in the direct vicinity of star-forming regions or gas clouds. In such a situation, the above expressions for the absolute SFR and mass outflow rate are accurate when a part of the galaxy with relatively homogeneous structure is selected by masking. Alternatively, it is possible to correct for morphology using the flux density contrasts  $\mathcal{E}_{\text{star, glob}}$  and  $\mathcal{E}_{\text{gas, glob}}$  from equations (46) and (47). We do not provide a detailed derivation here, but defer it to a future application of the method for which a morphological correction is needed.

After having determined all of the quantities described in this section for the best-fitting model of the tuning fork diagram, we obtain their PDFs through Monte-Carlo error propagation. To fully account for any covariance between the three free parameters  $t_{\text{gas}}$ ,  $t_{\text{over}}$ , and  $\lambda$ , it is necessary to draw from their complete three-dimensional PDF of equation (116). Drawing numbers from a discretised PDF is trivial in one dimension, because one simply interpolates the cumulative one-dimensional PDF, but this is not possible when dealing with a three-dimensional PDF. We therefore draw each Monte-Carlo realisation of the three free parameters through three correlated, one-dimensional representations of the relevant parts of the three-dimensional PDF.

First, we draw a value  $t_{\text{gas, mc}}$  from its marginalised, one-dimensional cumulative PDF by numerically solving

$$q_l = P_{1\text{D}}(t_{\text{gas}} \leq t_{\text{gas, mc}}), \quad (156)$$

for  $t_{\text{gas, mc}}$ , where  $q_l$  represents a random number  $\{q_l \in \mathbb{R} | 0 \leq q_l \leq 1\}$ . In practice, this requires interpolating the discretised form of  $P_{1\text{D}}(t_{\text{gas}} < t_{\text{gas, mc}})$ .<sup>17</sup> To obtain  $t_{\text{over, mc}}$ , we evaluate the marginalised, two-dimensional PDF of  $t_{\text{gas}}$  and  $t_{\text{over}}$  at  $t_{\text{gas, mc}}$ , resulting in the one-dimensional PDF of  $t_{\text{over}}$  given the value of  $t_{\text{gas, mc}}$ . This one-dimensional PDF is obtained by interpolation between the two discrete one-dimensional PDFs:

$$\begin{aligned} P_{2\text{D}}[t_{\text{over}} \leq t_{\text{over, mc}} | t_{\text{gas}}(i)], \\ P_{2\text{D}}[t_{\text{over}} \leq t_{\text{over, mc}} | t_{\text{gas}}(i+1)], \end{aligned} \quad (157)$$

where  $i$  denotes the index such that  $t_{\text{gas}}(i) \leq t_{\text{gas, mc}} \leq t_{\text{gas}}(i+1)$ . We then draw  $t_{\text{over, mc}}$  by numerically solving:

$$q_{l+1} = P_{2\text{D}}(t_{\text{over}} \leq t_{\text{over, mc}} | t_{\text{gas, mc}}), \quad (158)$$

<sup>17</sup> When drawing the Monte-Carlo realisations from the three-dimensional PDF, the interpolations are carried out on the logarithmic grids of equations (102)–(104), but they are linear in the random number and the cumulative probability.

for  $t_{\text{over},\text{mc}}$ , where the increment of the subscript in  $q_{l+1}$  indicates that we use a new random number. As before, the drawn value  $t_{\text{over},\text{mc}}$  is obtained by interpolating the discretised form of  $P_{2\text{D}}(t_{\text{over}} < t_{\text{over},\text{mc}} | t_{\text{gas},\text{mc}})$ . To obtain  $\lambda_{\text{mc}}$ , we evaluate the complete three-dimensional PDF of  $t_{\text{gas}}$ ,  $t_{\text{over}}$ , and  $\lambda$  at  $\{t_{\text{gas},\text{mc}}, t_{\text{over},\text{mc}}\}$ , resulting in the one-dimensional PDF of  $\lambda$  given the values of  $t_{\text{gas},\text{mc}}$  and  $t_{\text{over},\text{mc}}$ . This one-dimensional PDF is obtained by interpolation between the four discrete one-dimensional PDFs:

$$\begin{aligned} P_{3\text{D}}[\lambda \leq \lambda_{\text{mc}} | t_{\text{gas}}(i), t_{\text{over}}(j)], \\ P_{3\text{D}}[\lambda \leq \lambda_{\text{mc}} | t_{\text{gas}}(i), t_{\text{over}}(j+1)], \\ P_{3\text{D}}[\lambda \leq \lambda_{\text{mc}} | t_{\text{gas}}(i+1), t_{\text{over}}(j)], \\ P_{3\text{D}}[\lambda \leq \lambda_{\text{mc}} | t_{\text{gas}}(i+1), t_{\text{over}}(j+1)], \end{aligned} \quad (159)$$

where  $i$  is defined as before and  $j$  denotes the index such that  $t_{\text{over}}(j) \leq t_{\text{over},\text{mc}} \leq t_{\text{over}}(j+1)$ . We then draw  $\lambda_{\text{mc}}$  by numerically solving:

$$q_{l+2} = P_{3\text{D}}(\lambda \leq \lambda_{\text{mc}} | t_{\text{gas},\text{mc}}, t_{\text{over},\text{mc}}), \quad (160)$$

for  $\lambda_{\text{mc}}$ , where we have again used a new random number. As before, the drawn value  $\lambda_{\text{mc}}$  is obtained by interpolating the discretised form of  $P_{3\text{D}}(\lambda \leq \lambda_{\text{mc}} | t_{\text{gas},\text{mc}}, t_{\text{over},\text{mc}})$ .

Through the above procedure, we obtain a Monte-Carlo realisation of a fit to the observed tuning fork diagram, i.e.  $\{t_{\text{gas},\text{mc}}, t_{\text{over},\text{mc}}, \lambda_{\text{mc}}\}$ , by drawing from the three-dimensional PDF. In addition, we generate Monte-Carlo realisations of the duration of the reference time-scale  $t_{\text{star},\text{ref}}$  and the SFR and gas surface densities  $\Sigma_{\text{SFR}}$  and  $\Sigma_{\text{gas}}$ . The uncertainties on each of these quantities are defined either directly in the input file (see Table 2) or by analytical error propagation, as detailed in the discussion of equations (140) and (141). We assume that the underlying PDFs are Gaussian,<sup>18</sup> implying that the Monte-Carlo realisations of these three quantities are obtained by drawing Gaussian random numbers. This is repeated a total of  $N_{\text{mc,phys}}$  times and thus requires a total of  $N_{\text{rnd}} = 3N_{\text{mc,phys}}$  random numbers  $\{q_l \in \mathbb{R} | 0 \leq q_l \leq 1\}$  and the same number of Gaussian random numbers  $\{q'_l \in \mathbb{R}\}$ , where we adopt a default value of  $N_{\text{mc,phys}} = 10^6$  for a total of  $N_{\text{rnd}} = 6 \times 10^6$  random numbers. We find that  $10^6$  Monte-Carlo realisations yield PDFs of the derived quantities (obtained below) that are essentially noise-free and well-defined.

For each Monte-Carlo realisation, we calculate the quantities described in this section. The total ensemble of Monte-Carlo realisations then provides the marginalised, one-dimensional PDF for each of these quantities. We obtain the upward and downward uncertainties from these PDFs in the same way as for the three free parameters  $t_{\text{gas}}$ ,  $t_{\text{over}}$ , and  $\lambda$ , as described in equations (122)–(131) of Section 3.2.13. In addition, we obtain the covariance and correlation matrices for all quantity pairs  $\{X_i, X_j\}$  as usual, i.e.

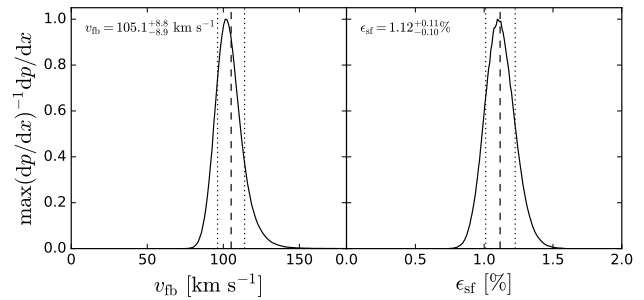
$$\text{cov}(X_i, X_j) = \langle X_i X_j \rangle - \langle X_i \rangle \langle X_j \rangle, \quad (161)$$

and

$$\rho_{ij} = \frac{\text{cov}(X_i, X_j)}{\sigma_i \sigma_j}, \quad (162)$$

for the covariance and correlation matrix, respectively, where  $\langle \dots \rangle$  takes the mean over all Monte-Carlo realisations. In equation (162),

<sup>18</sup> We allow asymmetric error bars on  $t_{\text{star},\text{ref}}$  in Table 2. If the upward and downward uncertainties differ, we use the mean uncertainty and still generate the uncertainties from a symmetric Gaussian.



**Figure 14.** One-dimensional PDFs of the feedback outflow velocity  $v_{\text{fb}}$  (left) and the star formation efficiency per star formation event  $\epsilon_{\text{sf}}$  (right), providing two examples of the quantities that can be derived from the fit in Figure 12. The PDFs are normalised to their maximum probability density. Vertical dashed lines indicate the best-fitting values (also indicated in the top left of each panel), whereas the vertical dotted lines represent the  $1\sigma$  uncertainty range as defined in the text. These two PDFs are representative for the PDFs of the derived quantities and illustrate that they are well-defined by the Monte-Carlo error propagation procedure described in the text.

the quantities  $\sigma_i$  and  $\sigma_j$  represent the standard deviations of the Monte-Carlo samples of  $X_i$  and  $X_j$ . The covariance and correlation matrices are written to disk in ASCII format. HEISENBERG writes each differential one-dimensional PDF to disk in ASCII format too and also outputs figures of all PDFs calculated here, i.e. the differential and cumulative one-dimensional PDFs for each derived quantity. The best-fitting values and their uncertainties are indicated in these plots. For illustration, Figure 14 shows the differential PDFs of the feedback velocity  $v_{\text{fb}}$  and the star formation efficiency per star formation event  $\epsilon_{\text{sf}}$  for the example maps used throughout Section 3. The PDFs are well-defined and show almost no statistical noise. In addition, the uncertainty ranges on the quantities are small, with relative errors less than 10 per cent, showing that even the derived quantities can be determined to high precision.

### 3.2.15 Model output

At this point in the application of the method, all quantities of interest have been calculated and the information needed for further analysis should be written to disk. Table 4 lists all constrained quantities for which PDFs and uncertainties are calculated and for which ASCII tables and figures have been produced. In the final step of the process, HEISENBERG generates an ASCII table that contains the best-fitting quantities from Table 4 and their uncertainties, as well as the absolute SFR and gas mass  $M_{\text{gas}}$ , the number of stellar and gas peaks  $N_{\text{peak,star}}$  and  $N_{\text{peak,gas}}$ , and the minimum aperture size used  $l_{\text{ap,min}}$ . In addition, HEISENBERG outputs a single table row containing all of these quantities, which is straightforward to copy into a master table of multiple runs. Finally, a log file is written to disk, which contains all terminal output produced during steps described in this section.

The text files generated during this final step add to the output files written to disk at earlier steps, as summarised in Figure 3. Upon completing the application of the method to a pair of galaxy maps, we have access to:

- (i) The masked maps showing the identified peak positions.
- (ii) Figures and ASCII tables of the observed and best-fitting tuning fork diagram.
- (iii) ASCII tables of  $\beta_{\text{star}}(f_{\text{star,over}})$ ,  $\beta_{\text{gas}}(f_{\text{gas,over}})$ ,  $\mathcal{E}_{\text{star}}(\lambda)$ , and  $\mathcal{E}_{\text{gas}}(\lambda)$ .



**Table 4.** Quantities constrained by the presented analysis

Quantity [unit]	Class <sup>a</sup>	Equation	Description
$\chi_{\text{red,min}}^2$	<i>fundamental</i>	113	Goodness-of-fit statistic
$t_{\text{gas}}$ [Myr]	<i>fundamental</i>	116	Best-fitting gas tracer lifetime (e.g. the cloud lifetime)
$t_{\text{over}}$ [Myr]	<i>fundamental</i>	116	Best-fitting overlap lifetime (e.g. the feedback time-scale)
$\lambda$ [pc]	<i>fundamental</i>	116	Best-fitting mean separation length of independent regions (e.g. the fragmentation length)
$t_{\text{star}}$ [Myr]	<i>byproduct</i>	133, 134	Star formation tracer lifetime
$\tau$ [Myr]	<i>byproduct</i>	135	Total region lifetime
$\beta_{\text{star}}$	<i>byproduct</i>	40	Ratio between the mean flux of overlapping and isolated star formation tracer peaks
$\beta_{\text{gas}}$	<i>byproduct</i>	41	Ratio between the mean flux of overlapping and isolated gas tracer peaks
$\mathcal{E}_{\text{star}}$	<i>byproduct</i>	44	Ratio between the central peak flux density and that on a size scale $\lambda$ for star formation tracer peaks
$\mathcal{E}_{\text{gas}}$	<i>byproduct</i>	45	Ratio between the central peak flux density and that on a size scale $\lambda$ for gas tracer peaks
$\mathcal{E}_{\text{star, glob}}$	<i>byproduct</i>	46	Ratio between the flux density on a size scale $\lambda$ and the map average for star formation tracer peaks
$\mathcal{E}_{\text{gas, glob}}$	<i>byproduct</i>	47	Ratio between the flux density on a size scale $\lambda$ and the map average for gas tracer peaks
$r_{\text{star}}$ [pc]	<i>derived</i>	92, 94	Disc radius or Gaussian dispersion radius of star formation tracer peaks
$r_{\text{gas}}$ [pc]	<i>derived</i>	93, 95	Disc radius or Gaussian dispersion radius of gas tracer peaks
$\zeta_{\text{star}}$	<i>derived</i>	136	Star formation tracer peak concentration parameter
$\zeta_{\text{gas}}$	<i>derived</i>	137	Gas tracer peak concentration parameter
$v_{\text{fb}}$ [km s <sup>-1</sup> ]	<i>derived</i>	138	Feedback-driven expansion velocity of ejecta if gas vanishes at $\lambda/2$
$v_{\text{fb},r}$ [km s <sup>-1</sup> ]	<i>derived</i>	139	Feedback-driven expansion velocity of ejecta if gas vanishes at $r_{\text{gas}}$
$\Sigma_{\text{SFR}}$ [M <sub>⊙</sub> yr <sup>-1</sup> pc <sup>-2</sup> ]	<i>composite</i>	140	SFR surface density
$\Sigma_{\text{gas}}$ [M <sub>⊙</sub> pc <sup>-2</sup> ]	<i>composite</i>	141	Gas surface density
$t_{\text{depl}}$ [Gyr]	<i>composite</i>	142	Gas depletion time
$\epsilon_{\text{sf}}$	<i>composite</i>	143	Star formation efficiency per star formation event
$\eta_{\text{fb}}$	<i>composite</i>	144	Instantaneous mass loading factor
$\bar{\eta}_{\text{fb}}$	<i>composite</i>	145	Time-integrated mass loading factor
$\chi_{\text{fb},E}$	<i>composite</i>	148	Feedback energy efficiency using $v_{\text{fb}}$
$\chi_{\text{fb},E,r}$	<i>composite</i>	149	Feedback energy efficiency using $v_{\text{fb},r}$
$\chi_{\text{fb},p}$	<i>composite</i>	152	Feedback momentum efficiency using $v_{\text{fb}}$
$\chi_{\text{fb},p,r}$	<i>composite</i>	153	Feedback momentum efficiency using $v_{\text{fb},r}$
$\dot{M}_{\text{sf}}$ [M <sub>⊙</sub> yr <sup>-1</sup> ]	<i>composite</i>	154	SFR per star formation event during the gas phase
$\dot{M}_{\text{fb}}$ [M <sub>⊙</sub> yr <sup>-1</sup> ]	<i>composite</i>	155	Feedback mass removal rate per star formation event during the overlap phase

<sup>a</sup> The listed classes of quantities are *fundamental* (obtained directly from the fitting process), *byproducts* (of the fitting process), *derived* (from fundamental quantities), or *composite* (obtained using additional constants such as conversion factors).

(iv) Figures of the marginalised, two-dimensional PDFs of the three free parameters.

(v) Figures and ASCII tables of the marginalised, one-dimensional PDFs of all constrained quantities (see Table 4).

(vi) The correlation and covariance matrices of all constrained quantities.

(vii) The output files listing the best-fitting values and uncertainties of all constrained quantities.

(viii) The log file of the terminal window.

In future work, we aim to extend the method further. In particular, we have included the option of closing the loop in Figure 3 and iteratively applying the method, such that the output of one analysis run can be used to modify the input maps and rerun the analysis with the updated maps (Hygate et al. 2018). This approach can be used to filter out diffuse emission from the maps that does not belong to independent regions as defined by the application of the method, or plausibly to account for distinct morphological features in the galaxy maps such as strong spiral arms or rings.

#### 4 VALIDATION USING DISC GALAXY SIMULATIONS

To test the method described in Section 3, we require pairs of galaxy maps to which the method can be applied. For this purpose, we carry out hydrodynamical simulations of isolated disc galaxies and create a large number of different maps from these simulations,

showing their gas and stellar content in a variety of ways. Because these are simulated rather than observed systems, we can perform a variety of controlled experiments to assess the accuracy of the method. These simulations are carried out using commonly-used, sub-grid models for star formation (e.g. Katz 1992) and feedback (see Hu et al. 2014 for a detailed description of the code used). Even though such models result in galaxies with macroscopic properties that provide a reasonable match to the observed galaxy population (e.g. Vogelsberger et al. 2014; Schaye et al. 2015), the star formation and feedback prescriptions are certainly inadequate when concerned with the cloud-scale processes discussed here (e.g. Hopkins et al. 2014). However, this is not a concern in the context of the problem at hand. The goal of this work is not to accurately model cloud-scale star formation and feedback, but to determine whether the presented method accurately retrieves the quantities describing what happens in the simulated maps, irrespective of whether the underlying physics are correct. If the method successfully extracts these quantities, then it can be applied to observations of real galaxies to motivate improved models for star formation and feedback in galaxy simulations. As we show in this section, the method indeed provides accurate measurements of these quantities and is suitable for application to a large variety of observed galaxies.

#### 4.1 Disc galaxy models

We briefly describe the setup and properties of the numerical disc galaxy models used. In order, we summarise the hydrodynamical method used, the cooling, star formation, and feedback model, the initial conditions, and the resulting properties of the isolated disc galaxy models.

##### 4.1.1 Hydrodynamical method

The galaxy simulations presented here have been performed with the smoothed particle hydrodynamics (SPH, Monaghan 1992) code P-GADGET-3 (Springel 2005), using the SPHGAL implementation of the hydrodynamics solver by Hu et al. (2014). SPH is a Galilean invariant, Lagrangian method for hydrodynamical simulations that adopts a particle representation of the fluid. The implementation by Hu et al. (2014) contains a number of important improvements relative to classical SPH flavours and alleviates many of the numerical issues that affected SPH models in recent years (see Agertz et al. 2007 and Hu et al. 2014 for discussions). In brief, these improvements are as follows.

(i) We use a Wendland  $C^4$  smoothing kernel (Dehnen & Aly 2012) to enable the use of  $N_{\text{ngb}} = 200$  neighbouring particles without suffering from the pairing instability (Price 2012). Using this high number of neighbours reduces the noise originating from the ‘ $E_0$  error’ in the SPH momentum equation (Read et al. 2010) and improves the convergence rate.

(ii) The classical density-entropy formulation of SPH is unable to accurately model contact discontinuities or fluid mixing. To remedy this, we use the pressure-entropy formulation of the SPH equations of motion (e.g. Ritchie & Thomas 2001; Saitoh & Makino 2013; Hopkins 2013). This formulation smoothes over the pressure rather than the density and therefore avoids spurious pressure jumps, which fundamentally improves the numerical approximation of the force field.

(iii) We include a strong artificial viscosity limiter as defined in Hu et al. (2014) and progressing from the weak limiter from Cullen & Dehnen (2010). This updated limiter keeps the fluid from being too viscous in a differentially rotating disc, while retaining the ability to capture shocks properly.

(iv) We adopt an artificial thermal energy conduction term to dampen the entropy jumps at hydrodynamic shocks, which smoothes the resulting pressure and force fields (see e.g. Read & Hayfield 2012). This term only operates across shocks (not shearing flows) and is highly pertinent in pressure-entropy SPH, because pressure jumps can exceed density jumps by orders of magnitude.

(v) Finally, the time steps of particles in the vicinity of strong shocks are limited to ensure that they are similar to within a factor of a few (Saitoh & Makino 2009). In addition, particles become active whenever they are subject to a feedback energy injection (Durier & Dalla Vecchia 2012).

Together, these improvements are chosen to optimise the accuracy of the hydrodynamical modelling in differentially-rotating disc galaxies (Hu et al. 2014), as considered in this paper. The simulations include the evolutionary cycling between gas and star particles, as well as a live dark matter halo.

##### 4.1.2 Cooling, star formation, and feedback

The thermal evolution of the gas particles proceeds according to the metal enrichment and cooling scheme from Aumer et al. (2013),

**Table 5.** Particle resolution of the simulated galaxies

Component	$N_{\text{LR}}$	$N_{\text{HR}}$	$m_{\text{LR}} [M_{\odot}]$	$m_{\text{HR}} [M_{\odot}]$
Halo	400000	1000000	$2.3 \times 10^6$	$9.0 \times 10^5$
Stellar disc	462000	2310000	$1.4 \times 10^4$	$2.7 \times 10^3$
Gas disc	308000	1540000	$1.4 \times 10^4$	$2.7 \times 10^3$
Bulge	20000	100000	$1.4 \times 10^4$	$2.7 \times 10^3$

which traces the 11 elements H, He, C, N, O, Ne, Mg, Si, S, Ca, and Fe. Once a star particle forms (see below), it inherits the abundances from its parent gas particle. The chemical enrichment includes ejecta from SNe types Ia and II, as well as AGB stars. After metals are deposited into the ISM, they are advected with the flow and undergo metal diffusion as in Aumer et al. (2013) to account for turbulent mixing. The resulting cooling rate of each particle is obtained for each element by assuming the gas is optically thin and subject to the UV/X-ray background (Wiersma et al. 2009). We adopt a cooling floor of  $T = 10$  K, which is not reached in practice (see Section 4.1.4).

Gas particles are eligible for star formation once they have temperatures  $T < 1.2 \times 10^4$  K and hydrogen particle densities  $n_{\text{H}} > 0.5 \text{ cm}^{-3}$  (assuming atomic gas). Star-forming gas particles are converted into stars stochastically, according to the prescription from Katz (1992):

$$\rho_{\text{SFR}} = \epsilon \frac{\rho_{\text{g}}}{t_{\text{dyn}}}, \quad (163)$$

where  $\rho_{\text{SFR}}$  is the SFR volume density,  $\rho_{\text{g}}$  is the gas particle volume density,  $\epsilon = 0.02$  is the star formation efficiency per dynamical time, and

$$t_{\text{dyn}} = \frac{1}{\sqrt{4\pi G \rho_{\text{g}}}}, \quad (164)$$

is the dynamical time. In practice, this expression implies that, during a timestep  $dt$ , the probability that a gas particle is converted into a star particle is given by

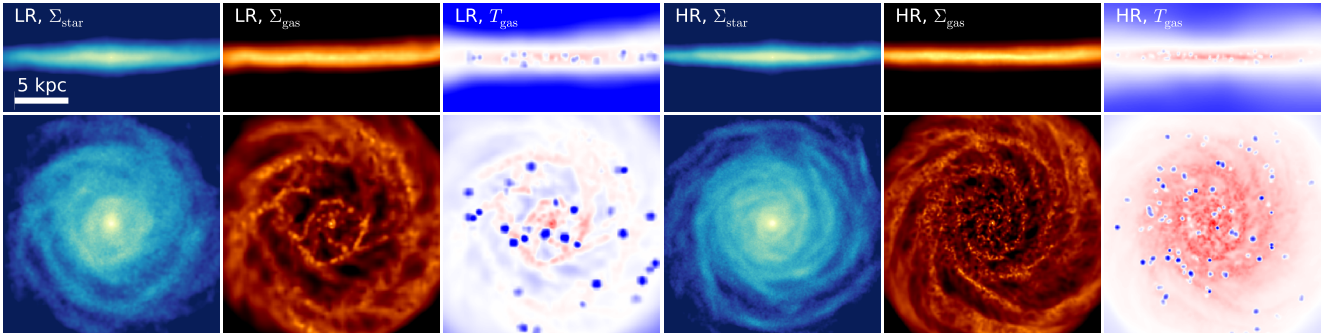
$$p_{\text{sf}} = 1 - \exp(-\epsilon dt/t_{\text{dyn}}). \quad (165)$$

The effective timescale of the exponential is thus  $t_{\text{dyn}}/\epsilon$ , which at the minimum density for star formation corresponds to 1.6 Gyr. In practice, clouds consist of a large number of particles (recall that  $N_{\text{ngb}} = 200$ ), implying that they begin forming stars on a correspondingly shorter time-scale.

The deposition of mass, (radial) momentum, and (thermal) energy into the ISM is driven by SN explosions, which adds these to the 10 gas particles closest to the star particle according to the kernel weighting. Type II SNe take place after 3 Myr, whereas Type Ia SNe occur continuously between 0.1–1 Gyr. SN ejecta correspond to  $\sim 19$  per cent of the particle mass and carry  $10^{51}$  erg of energy, with a velocity of  $3000 \text{ km s}^{-1}$ . The momentum is added to the gas particles according to an inelastic collision and the remaining energy added to the thermal energy budget of these particles. See Scannapieco et al. (2006) and Hu et al. (2014) for further details.

##### 4.1.3 Initial conditions

We generate the initial conditions of two isolated disc galaxies according to the method by Springel et al. (2005a). The galaxies consist of a live dark matter halo, a mixed gas-stellar disc, and a stellar bulge. The halo is taken to follow a Hernquist (1990) profile with an equivalent Navarro et al. (1997) concentration parameter of  $c = 12$ , a virial mass of  $M_{200} = 9.0 \times 10^{11} M_{\odot}$ , and a virial



**Figure 15.** Maps of the simulated galaxies, showing (from left to right) the stellar surface density, gas surface density, and density-weighted temperature along the line of sight, for the low-resolution simulation (LR, left three columns) and the high-resolution simulation (HR, right three columns). The top row shows the edge-on views of the galaxies, whereas the bottom row shows the face-on views. The colour scale is logarithmic in all panels, with stellar surface densities covering  $\log_{10}(\Sigma/M_{\odot} \text{ pc}^{-2}) = 0.5\text{--}3.5$  in the edge-on maps and  $\log_{10}(\Sigma/M_{\odot} \text{ pc}^{-2}) = 0.5\text{--}2.5$  in the face-on maps. The upper limits on the gas surface densities are 0.5 dex lower. For both the edge-on and face-on temperature maps, {red, white, blue} corresponds to  $\log_{10}(T/K) = \{3, 4, 5\}$ . All panels are 20 kpc on a side and the top-left panel includes a scale bar for reference.

radius of  $r_{200} = 2.2 \times 10^2$  kpc. We set the halo spin parameter to  $\lambda = |J|E^{1/2}/GM_{200}^{5/2} = 0.030$ , where  $J$  is the halo angular momentum and  $E$  is the halo energy. The disc has a total mass of  $M_d = 1.05 \times 10^{10} M_{\odot}$  and an angular momentum fraction set to be equal to the disc mass fraction. The stellar disc constitutes 60 per cent of the total disc mass ( $M_{d,s} = 6.3 \times 10^9 M_{\odot}$ ) and follows an exponential profile with a scale radius of  $r_{d,s} = 3.1$  kpc and a scale height of  $h_d = 0.14$  kpc. The gas disc constitutes 40 per cent of the total disc mass ( $M_{d,g} = 4.2 \times 10^9 M_{\odot}$ ) and follows a composite radial profile, consisting of an exponential profile with a scale radius of  $r_{d,g} = 6.3$  kpc (holding 5 per cent of the gas mass) and a constant surface density HI disc that extends out to  $r_{\text{HI}} = 12.5$  kpc (holding 95 per cent of the gas mass). Finally, the galaxies host a [Hernquist \(1990\)](#) bulge with mass  $M_b = 2.7 \times 10^8 M_{\odot}$  and a scale length of  $r_b = 0.23$  kpc. Over the course of the simulation,  $h_d$  and  $r_b$  increase somewhat, by up to 40 per cent. The above parameters have been chosen to be somewhat representative of near- $L^*$ , flocculent spiral galaxies in the local Universe.

To sample the three-dimensional mass distribution of the galaxy models, we use four families of particles with numbers and masses listed in [Table 5](#). The dependence on numerical resolution is assessed by considering two galaxy models with different resolutions. The low-resolution (subscript ‘LR’) model consists of a total of  $1.19 \times 10^6$  particles, whereas the high-resolution (subscript ‘HR’) model consists of a total of  $4.95 \times 10^6$  particles. Gravitational forces between the particles are calculated using a softening length for the halo particles of  $h_{\text{halo}} = 100$  pc and for the baryonic particles in the {low, high}-resolution simulations of  $h_{\text{LR,bary}} = 20$  pc and  $h_{\text{HR,bary}} = 10$  pc, respectively. Note that the smoothing lengths enclosing  $N_{\text{ngb}} = 200$  neighbours used for determining the local fluid properties are considerably larger than the above softening lengths, with typical map-averaged values of  $h_{\text{smooth,LR}} \sim 250$  pc and  $h_{\text{smooth,HR}} \sim 100$  pc (see [Section 5.1](#)). Finally, the particle masses listed in [Table 5](#) are not sufficient to resolve the cold gas, because the mass contained in an SPH kernel is  $N_{\text{ngb}} m_{\text{HR,g}} \sim 5 \times 10^5 M_{\odot}$ . However, they are sufficient to marginally resolve the Jeans mass at the onset of cloud formation, which suffices for the main goals of this paper (see [Section 4.1.4](#)).

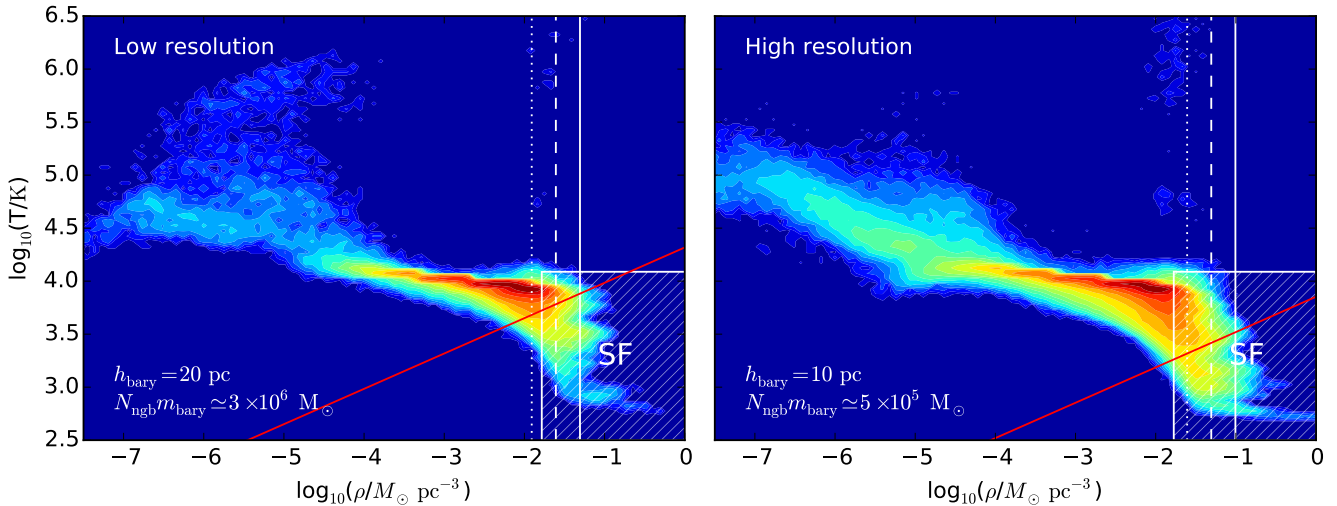
#### 4.1.4 Simulated disc galaxy properties

We run the disc galaxy simulations for a total duration of  $t_{\text{run}} = 2.2$  Gyr before generating the maps on which the method of [Section 3](#) is to be tested. For reference, we show the stellar and gas particle maps at  $t = t_{\text{run}}$  in [Figure 15](#). The time covered by the simulations exceeds the dynamical time at the virial radius and corresponds to over a dozen dynamical times within the galaxy itself (at  $R < 10$  kpc). The galaxies achieve equilibrium on the order of a few disc dynamical times, well before the snapshots at  $t_{\text{run}}$  that we will consider throughout the rest of this paper. Indeed, [Figure 15](#) shows that the simulated galaxies are stable and do not exhibit any obvious transient morphological features. This is confirmed by visual inspection of all preceding snapshots.

To assess the behaviour of the ISM, [Figure 16](#) shows the phase diagram of the gas discs at  $t = t_{\text{run}}$ , demonstrating that the ISM in the simulations is stable too. Star formation takes place in the regime where cooling from the atomic phase has set in and the Jeans mass is sufficiently well-resolved to identify where the cooling track sets in. Some hot and dense feedback ejecta can be seen along the vertical white lines in the top right of each panel. The increased resolution of the high-resolution simulation enables more efficient gas cooling than in the low-resolution simulation due to the higher densities. In [Section 4.3](#), we will be testing the method of this paper by using gas maps of these galaxies above minimum volume densities indicated by the vertical white lines. [Figure 16](#) shows that these volume density thresholds trace star-forming gas in all but one (LR, dotted line) case. The total SFR in the simulations averaged over the 10–300 Myr preceding these snapshots is  $0.2\text{--}0.3 M_{\odot} \text{ yr}^{-1}$ .

## 4.2 Age-binned stellar maps

As a first test of the method presented in [Section 3](#), we exclusively use maps of the star particles in the simulations. The major advantage of using star particles is that their ages are known. By only displaying the star particles in specified age ranges, the duration of the evolutionary phase displayed in each map is known, allowing the accuracy of the method to be directly quantified. In this subsection, we use these ‘age-binned’ stellar maps to assess how well the method retrieves the ‘cloud lifetime’  $t_{\text{gas}}$ , the ‘feedback time-scale’  $t_{\text{over}}$ , and the overlap-to-isolated flux ratios  $\beta_{\text{star}}$  and  $\beta_{\text{gas}}$ . Even



**Figure 16.** Phase diagram of the gas particles in the low-resolution (left) and high-resolution (right) simulations. The part of phase space where gas particles are eligible for star formation is represented by the hatched region, indicated with ‘SF’. The vertical dotted, dashed, and solid lines indicate the critical volume densities above which the simulated gas maps are generated in Section 4.3. Converting densities from  $M_{\odot} \text{pc}^{-3}$  to  $\text{cm}^{-3}$  or  $m_{\text{H}} \text{cm}^{-3}$  requires the addition of 1.25 dex (assuming a mean molecular weight  $\mu = 2.3$ ) or 1.61 dex, respectively. These phase diagrams show that the simulations resolve cooling down to temperatures of several 100 K and densities of  $\sim 10 \text{cm}^{-3}$ , covering the onset of molecular cloud condensation. The Jeans mass is resolved with at least 200 particles for fluid elements above the red diagonal line, which applies to all gas at the onset of the collapse towards star formation, in the top left corner of the hatched region. At both resolutions, some feedback-heated and shocked gas in star-forming regions is present at high densities and temperatures.

though we only use stellar maps in this subsection and all phases of the evolutionary timeline therefore represent stellar phases, we keep referring to the first phase as ‘gas’ (this includes its associated quantities such as  $t_{\text{gas}}$  and  $\beta_{\text{gas}}$ ) to remain consistent with the rest of the paper.

#### 4.2.1 Procedure for creating the maps

To generate age-binned maps of the star particles in the galaxy models, we first need to define the age bins for which the particles are displayed. The duration of the stellar phase  $t_{\text{star}}$  represents the width of the age bin for which the star particles are shown in the stellar map. This time-scale is taken to be known and is used during the fitting process to determine the other time-scales. The duration of the ‘gas’ phase  $t_{\text{gas}}$  represents the width of the age bin for which the star particles are shown in the ‘gas’ map. The duration of the ‘overlap’ phase  $t_{\text{over}}$  represents the age range for which star particles are included in both maps. We add a fourth time-scale  $t_{\text{off}}$  that represents the offset of the age bins from an age of zero ( $t = 0$ ). For each experiment, these four time-scales define the adopted age bins. The stellar phase includes the star particles in the age bin spanning  $t = \langle t_{\text{off}} + t_{\text{gas}} - t_{\text{over}}, t_{\text{off}} + t_{\text{gas}} + t_{\text{star}} - t_{\text{over}} \rangle$ , whereas the ‘gas’ phase includes the star particles in the age bin spanning  $t = \langle t_{\text{off}}, t_{\text{off}} + t_{\text{gas}} \rangle$ . The above definitions are consistent with the evolutionary timeline in Figure 1 and imply that the youngest star particles are shown in the ‘gas’ map, whereas the oldest star particles are shown in the stellar map. As will be discussed in Section 4.2.2, we have set  $t_{\text{off}} = 0$  for all but two experiments.

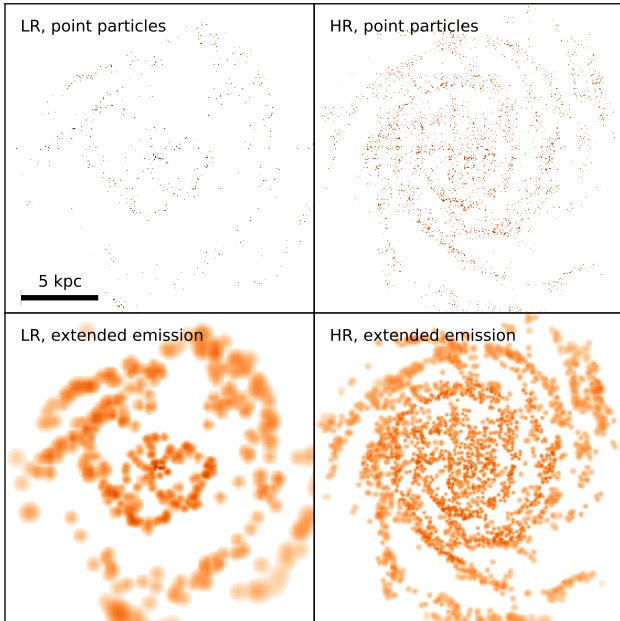
In addition to the above time-scales, we specify the overlap-to-isolated flux ratios  $\beta_{\text{star}}$  and  $\beta_{\text{gas}}$ . To do so, we first identify the star particles in the age range  $t = \langle t_{\text{off}} + t_{\text{gas}} - t_{\text{over}}, t_{\text{off}} + t_{\text{gas}} \rangle$ , which corresponds to the overlap phase. When creating the stellar map, we then multiply the masses of these star particles by  $\beta_{\text{star}}$ , while for the ‘gas’ map, we multiply their masses by  $\beta_{\text{gas}}$ . Note that out of the six quantities discussed here, only  $t_{\text{star}}$  is used as

an input parameter in the method of Section 3. The method is not informed of the values of the other five quantities. Instead, we are using the experiments carried out here to assess how well the best-fitting values of these quantities match their input values.

Having defined the age bins for which the star particles are shown in each map, we generate the surface density maps of the particles in these bins. These maps are square with a diameter of 20 kpc and 1403 pixels, corresponding to a pixel scale of 14.25 pc. Two types of maps are made for each galaxy model. The ‘point particle’ surface density maps show the distribution of particles as-is, for each pixel adding up the enclosed mass of the particles (modulo a factor  $\beta$  if it contains particles in the overlap phase) and dividing by the pixel area. At the adopted numerical resolution of the simulations and the pixel scale of 14.25 pc, the typical number of particles per pixel with ages in the selected age bins is zero or unity. By contrast, the ‘extended emission’ surface density maps account for the extended morphology expected for real star-forming regions. In the case of gas particles, this extended emission is set by the smoothing kernel and smoothing length, but normally there is no need to smoothen the star particles (only their gravitational force is softened). For consistency with the gas maps discussed in Section 4.3.1 below, we generate the extended emission surface density maps of the star particles by calculating their smoothing length following the same procedure as for the gas particles, i.e. by using the 200 nearest neighbours to define the smoothing length and adopting a Wendland  $C^4$  smoothing kernel (Dehnen & Aly 2012) to distribute the mass around each particle position. This choice implies coarser resolution in low-density regions and is made for consistency with P-Gadget, in which the same kernel is used to smoothen the gas particles in the simulations (as discussed in Section 4.1).

The maps resulting from the above procedure are written to disk in the FITS file format and include a FITS header that contains their key properties, consistent with common conventions for observational data delivery. For the particular case of  $t_{\text{gas}} = 30 \text{ Myr}$  and  $t_{\text{off}} = 0 \text{ Myr}$ , Figure 17 shows four examples of a typical ‘gas’





**Figure 17.** The four different types of ‘gas’ maps used in Section 4.2. From left to right, top to bottom, these are maps of the low-resolution (LR) point particle distribution, the high-resolution (HR) point particle distribution, the LR extended emission (generated by using a smoothing kernel, see the text), and the HR extended emission. In these examples, the maps show the distribution of star particles in the age range  $t = (0, 30)$  Myr. In Section 4.2.2, we present the 22 different sets of age ranges (and thus maps) for which our method is tested, for a total of  $4 \times 22 = 88$  experiments.

map that these experiments are carried out with, illustrating the similarities and differences between the two types of maps (‘point particles’ and ‘extended emission’) for both disc galaxy models (‘low resolution’ and ‘high resolution’). Because we are using age-binned stellar maps both to generate the stellar maps and the ‘gas’ maps shown in Figure 17. In contrast to the pairs of age-binned stellar maps discussed here, Section 4.3 below concerns a second set of experiments that each combine a stellar map with an actual gas map of the gas particles in the disc galaxy models.

#### 4.2.2 Description of the experiments

Table 6 summarises the experiments carried out using the age-binned stellar maps, specifying the values of  $t_{\text{star}}$ ,  $t_{\text{gas}}$ ,  $t_{\text{over}}$ ,  $\beta_{\text{star}}$ , and  $\beta_{\text{gas}}$  for each experiment. It also lists the symbols used to indicate each experiment in the figures presented in Section 4.2.3 and 4.2.4. The experiments are divided into six main categories, reflected by the differently coloured symbols in the final column.

(i) The first set of experiments (black symbols) is aimed at investigating the role of galactic dynamics. This is the simplest possible experiment, with equal stellar and ‘gas’ time-scales ( $t_{\text{star}} = t_{\text{gas}}$ ), no overlap ( $t_{\text{over}} = 0$  Myr), no age offset ( $t_{\text{off}} = 0$  Myr), and no flux difference between particles in the isolated and overlap phases ( $\beta_{\text{star}} = 1$  and  $\beta_{\text{gas}} = 1$ ). By spanning 1.5 dex in total duration (ranging from just a few Myr in the first experiment to more than an orbital time in the fourth experiment), we can assess whether the gradual dispersion of groups of young star particles affects the accuracy of our method.

(ii) The second set of experiments (blue symbols) adds a first

**Table 6.** Experiments carried out using age-binned stellar maps

ID	$t_{\text{star}}$	$t_{\text{gas}}$	$t_{\text{over}}$	$t_{\text{off}}$	$\beta_{\text{star}}$	$\beta_{\text{gas}}$	Symbol
1	3	3	0	0	1.0	1.0	■
2	10	10	0	0	1.0	1.0	■
3	30	30	0	0	1.0	1.0	■
4	100	100	0	0	1.0	1.0	■
5	1	3	0	0	1.0	1.0	◀
6	3	1	0	0	1.0	1.0	▶
7	10	30	0	0	1.0	1.0	◀
8	30	10	0	0	1.0	1.0	▶
9	1	10	0	0	1.0	1.0	▼
10	10	1	0	0	1.0	1.0	▲
11	3	30	0	0	1.0	1.0	▼
12	30	3	0	0	1.0	1.0	▲
13	10	10	2	0	1.0	1.0	■
14	10	10	4	0	1.0	1.0	■
15	10	10	8	0	1.0	1.0	■
16	10	30	4	0	1.0	1.0	◀
17	30	10	4	0	1.0	1.0	▶
18	10	10	4	0	0.5	1.0	■
19	10	10	4	0	1.0	0.5	■
20	10	10	4	0	0.5	0.5	■
21	10	30	4	500	1.0	1.0	◀
22	30	3	0	500	1.0	1.0	▶

All time-scales are in units of Myr.

level of additional complexity by considering various degrees of asymmetry between  $t_{\text{star}}$  and  $t_{\text{gas}}$ . In the first four experiments, both time-scales differ only by a factor of three, whereas the final four experiments have time-scales differing by an order of magnitude. These experiments are intended to test how well the method retrieves the  $t_{\text{gas}}$  if it is very dissimilar from the reference time-scale  $t_{\text{star}}$ .

(iii) The third set of experiments (red symbols) reverts to equal stellar and ‘gas’ time-scales, but adds a non-zero overlap ( $t_{\text{over}} > 0$  Myr). These experiments allow us to determine how well our method retrieves  $t_{\text{over}}$  for various durations of the overlap phase.

(iv) The fourth set of experiments (green symbols) combines the second and third sets, but considering dissimilar  $t_{\text{star}}$  and  $t_{\text{gas}}$  and including an overlap phase. This setup reminisces the situation one would expect to occur in nature. We consider two cases, one where  $t_{\text{gas}} > t_{\text{star}}$  and one where  $t_{\text{gas}} < t_{\text{star}}$ .

(v) The fifth set of experiments (cyan symbols) is an extension of the third set, this time varying the overlap-to-isolated flux ratios  $\beta_{\text{star}}$  and  $\beta_{\text{gas}}$ . We consider two cases where only one of these quantities deviates from unity (meaning that either  $\beta_{\text{star}} = 0.5$  or  $\beta_{\text{gas}} = 0.5$ , which corresponds to the particles being half as bright when they are in the overlap phase than when they are in the isolated phase) and also include an experiment where both  $\beta_{\text{star}} = 0.5$  and  $\beta_{\text{gas}} = 0.5$ . These experiments are included to evaluate how well our method retrieves  $t_{\text{over}}$  when there is any flux evolution of independent regions during the covered timeline, as well as to see how well we retrieve the parameters  $\beta_{\text{star}}$  and  $\beta_{\text{gas}}$  that capture this flux evolution.

(vi) The sixth set of experiments (magenta symbols) considers a large age offset of  $t_{\text{off}} = 500$  Myr from the star particle formation time. One experiment represents a typical set of time-scales (cf. the fourth set) and one reflects the extreme contrast between the stellar and ‘gas’ time-scale (cf. the second set). These experiments

consider cases where any initial structure has entirely been erased by galactic dynamics and is aimed at testing how well the method works for smoothed distributions of particles or regions.

The above sets of experiments are carried out using the ‘point particle’ or the ‘extended emission’ surface density maps (see Section 4.2.1), both for the low and high-resolution simulations. We thus have a total of 88 experiments using age-binned stellar maps. Some of the parameters used in these experiments deviate from the default values in Section 3. Firstly, we are dealing with maps showing star particles with specified age ranges, for which the reference time-scale includes the time for which these age ranges overlap. We therefore set  $t_{\text{star.incl}} = 1$ , indicating that  $t_{\text{star.ref}}$  includes  $t_{\text{over}}$  and thus  $t_{\text{star}} = t_{\text{star.ref}}$ . Secondly, we set  $N_{\text{pix,min}} = 1$  for the point particle experiments, to enable single pixels to be identified as peaks. Thirdly, for the low-resolution runs we set  $l_{\text{ap,min}} = 100$  pc to remove sub-resolution apertures, as well as  $\Delta \log_{10} \mathcal{F} = 2.5$  and  $\delta \log_{10} \mathcal{F} = 0.25$  for the identification of both the stellar peaks and gas peaks, because the low-resolution maps used in this work require a finer and deeper contour level spacing to identify the relevant emission peaks.

#### 4.2.3 Accuracy of the ‘cloud lifetime’ $t_{\text{gas}}$

The first test carried out using the  $4 \times 22 = 88$  experiments presented in Section 4.2.2 is to determine how well the duration of the ‘gas’ phase ( $t_{\text{gas}}$ ) is retrieved by our new method. The quantity  $t_{\text{gas}}$  can have a variety of physical meanings depending on the (gas) tracers used. For the specific, real-Universe example of tracing (molecular) gas with CO, this quantity refers to the time spent by a region in a CO-bright state preceding a star formation event, commonly referred to as the molecular cloud lifetime. As discussed in Section 3, our method provides the complete PDF for each of the quantities constrained in the fitting process, including  $t_{\text{gas}}$ . Figure 18 shows the best-fitting value of  $t_{\text{gas}}$  (referred to as  $t_{\text{gas,out}}$ ) as a function of the input value of  $t_{\text{gas}}$  from Table 6 (referred to as  $t_{\text{gas,in}}$ ) for all 88 experiments.

**4.2.3.1 Point particles** Because the method assumes that independent regions can be easily identified and the evolutionary timeline is well-sampled, the best agreement between input and output values is expected for the high-resolution point particle experiments (top-right panel). A comparison between the scatter of the data points around the one-to-one relation ( $\sigma_{\text{true}}$ ) in the four panels of Figure 18 shows that this is indeed the case. The logarithmic scatter in the top-right panel is just 0.06 dex, corresponding to just  $\sim 14$  per cent. The fact that this scatter is similar in magnitude to the mean error bar size of 0.07 dex means that the observed scatter is entirely accounted for by the uncertainties on the best-fitting values  $t_{\text{gas,out}}$ . In other words, the method does not have any fundamental systematic biases in determining  $t_{\text{gas}}$  for the high-resolution point particle maps.

Even though this subset of experiments is expected to return the best results, the close agreement is non-trivial. For instance, the method assumes that independent regions are randomly distributed in space up to a spatial scale of a few times the mean separation length  $\lambda$ , which corresponds to  $\sim 1$  kpc in these experiments. As is evident from Figure 17, the distribution of particles in the simulation is substructured. It is encouraging that the deviation from a random distribution seen here does not strongly influence the accuracy of the method. In Section 7.2, we will discuss more extreme situations in which it may be important to account for substructure

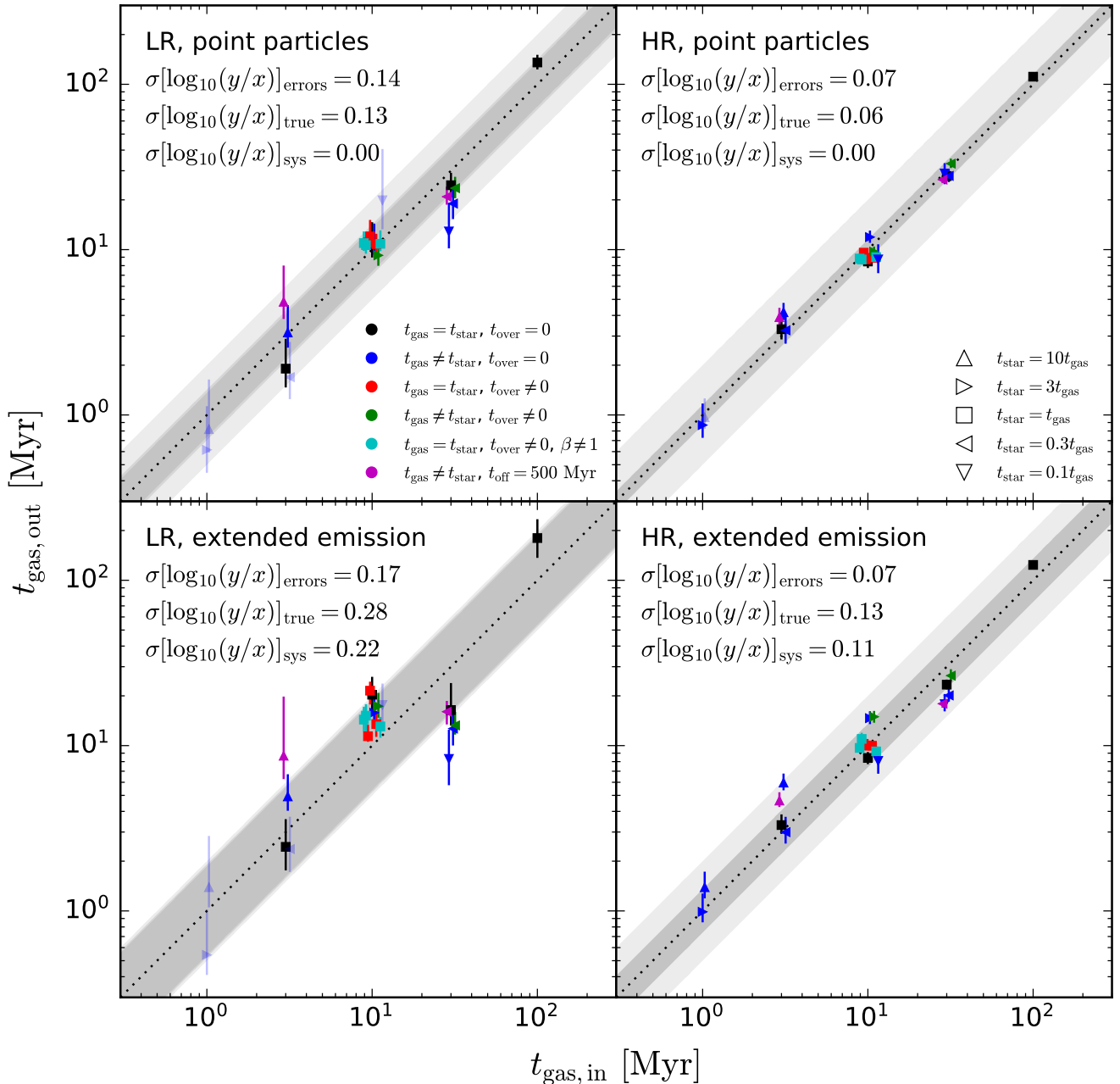
in the spatial distribution of independent regions. The method also assumes that the SFR has been constant to within the error bars on the best-fitting value of  $t_{\text{gas}}$  – any excess star formation during one phase relative to the other phase will proportionally bias the measured time-scales. Again, the good agreement observed here shows that variations of the SFR in the simulated isolated disc galaxy with an M33-like mass are sufficiently small in principle to avoid biases in the measured value of the ‘gas’ phase lifetime.

The agreement between input and output is less pronounced in the low-resolution point particle experiments (top-left panel). With a smaller number of independent regions, the evolutionary timeline of Figure 1 is more sparsely sampled, implying that the Poisson noise of the star formation history increases and the measured time-scales are more prone to being biased. The number of independent regions scales inversely with the particle mass, which is a factor of five larger than in the high-resolution simulation. We thus expect the Poisson noise to be a factor of  $\sqrt{5} \approx 2.2$  larger than in the high-resolution simulation, which is indeed the difference between the scatter observed in the top-left and top-right panels ( $0.13/0.06 \approx 2.2$ ). This ‘true’ scatter around the one-to-one relation is consistent with the scatter expected for the size of the error bars ( $\sigma_{\text{errors}}$ ), indicating that the error bars on the measurements are accurate.

The scatter around the one-to-one relation does not need to be random. Because these experiments all use the same simulation snapshot, deviations from a constant SFR may affect all experiments covering a similar age range in the same way. Especially for the low numbers of independent regions considered in the top-left panel of Figure 18, this can lead to a systematic deviation of the measured time-scales, which explains at least part of the discrepant data points at  $t_{\text{gas,in}} = 30$  Myr. If we consider the example of experiment 11, the number ratio between the particles in the stellar and ‘gas’ phases is 0.08 dex (20 per cent) higher than expected for the relative durations of these phases. This causes the retrieved value of  $t_{\text{gas}}$  to be 0.08 dex lower than it should be and represents yet another example of Poisson noise, illustrating that the sampling statistics of having smaller numbers of regions contribute to many of the differences between the top-left and top-right panels. In addition, the spatial structure of these maps deviates more strongly from a random distribution than for the high-resolution experiments (see Figure 17). As we will discuss in Section 7.2, this substructure may also contribute somewhat to the deviation from the one-to-one relation.

**4.2.3.2 Extended emission** The bottom two panels of Figure 18 show the results of the experiments based on maps of the extended emission, i.e. for a non-zero smoothing length (see Figure 17). Using extended emission maps rather than point particle maps again increases the scatter around the one-to-one relation, as would be expected. After all, compared to the point particle maps it is less trivial to identify independent regions as peaks when their emission is smoothed over a certain area. This smoothing also adds an emission background around each region that may overlap in space with neighbouring regions and may be mistaken for a time overlap (see Section 4.2.4).

In the method of Section 3, we deal with the above issues by representing the independent regions as two-dimensional Gaussians. In our formalism, each Gaussian stands for one region, implying that its two-dimensional integral must be equal to unity. This property is used to relate the surface density contrast between peaks and the average across the map to the width (i.e. the Gaussian dispersion) of each peak. Mathematically, this is the correct approach and it works very well in practice when peaks are marginally re-



**Figure 18.** Accuracy of the ‘gas’ phase lifetimes ( $t_{\text{gas}}$ ) retrieved by our new method, for the low-resolution (LR, left panels) and high-resolution (HR, right panels) simulations, using point particles (top panels) and extended emission (bottom panels). Shown is the best-fitting value  $t_{\text{gas, out}}$  as a function of the input value  $t_{\text{gas, in}}$  specified for each experiment (see Table 6). The data points indicate the results for each experiment (see Table 6 and the legends in the top panels for the meanings of the symbols), with error bars indicating the 16th and 84th percentiles of the  $t_{\text{gas, out}}$  PDF. Note that only five values of  $t_{\text{gas, in}} = \{1, 3, 10, 30, 100\}$  Myr are considered in these experiments. The small horizontal offsets of the data points relative to these exact values have been added to improve the legibility of the figure. Transparent symbols do not pass the guidelines for the reliable application of our method presented in Section 4.4 below and should therefore be omitted from further analysis. The dotted line indicates the one-to-one agreement and the grey areas illustrate different amounts of scatter. The lightest shade represents a factor of two, the middle shade represents the standard deviation of  $\log_{10}(t_{\text{gas, out}}/t_{\text{gas, in}})$  (i.e. the scatter around the diagonal) for all data points (including the unreliable, transparent ones), and the darkest shade indicates the standard deviation of  $\log_{10}(t_{\text{gas, out}}/t_{\text{gas, in}})$  exclusively for the opaque (reliable) data points. Three different measures of the scatter around the one-to-one relation are given in the top left corner of each panel. The first (subscript ‘errors’) is the scatter around the one-to-one relation expected for the error bars of the opaque data points (this is effectively the mean error bar). The second (subscript ‘true’) is the actual scatter around the one-to-one relation for the opaque data points (corresponding to the darkest shade of grey). The third (subscript ‘sys’) is the difference between the first and second values (defined as  $\sigma_{\text{sys}}^2 = \sigma_{\text{true}}^2 - \sigma_{\text{errors}}^2$ , with the additional requirement that  $\sigma_{\text{sys}}^2 \geq 0$ ), which is the part of the scatter that cannot be accounted for by the error bars. As such, this represents the ‘systematic’ inaccuracy of the method. In all cases, the intrinsic accuracy of the method is  $\sigma_{\text{sys}} \lesssim 0.2$  dex, whereas at the high resolutions reached for local-Universe galaxies with telescopes like ALMA, the accuracy is even higher with  $\sigma_{\text{sys}} \lesssim 0.1$  dex.

solved in observed galaxy maps, because the observational point spread function typically resembles a Gaussian (see Section 5). However, it is possible (e.g. due to substructure, blending, highly-resolved independent regions, or a different point spread or smoothing function<sup>19</sup>) that the assumption of Gaussian surface density profiles is invalid. In this case, the measured time-scales may be a bad representation of the true evolutionary timeline.

The bottom panels of Figure 18 demonstrate that any deviations resulting from the above assumptions is modest. The excess of the scatter beyond the mean error bar size (which is a measure for the inaccuracy of the method) indeed increases, but it remains within reasonable limits. The experiments using high-resolution extended emission maps have a spatial resolution that is comfortably obtained with ALMA. For these experiments, the excess scatter around the one-to-one relation is 0.11 dex (which is less than 30 per cent). While this residual scatter is non-zero, it is an encouraging accuracy given the discussed assumptions. It is also sufficient for astrophysical purposes. Current discussions in the literature regarding the molecular cloud lifetime discuss order-of-magnitude differences (e.g. Dobbs et al. 2014; Jeffreson & Kruijssen 2018), which our method can plausibly distinguish at the  $> 8\sigma$  level.

**4.2.3.3 General conclusions** We see that across all panels of Figure 18, the largest outliers (ignoring the transparent symbols) are upward or downward-pointing triangles. These symbols represent the experiments in which  $t_{\text{star}}$  and  $t_{\text{gas}}$  differ by a factor of 10. Throughout the experiments carried out for this study (also see Section 4.3.4), we have noticed that the most accurate results are obtained when  $t_{\text{star}}$  and  $t_{\text{gas}}$  are similar. This occurs because the ‘tuning fork’ diagram of Figure 2 becomes asymmetric for dissimilar values of  $t_{\text{star}}$  and  $t_{\text{gas}}$ , which moves one of the two branches in the diagram closer to the  $\mathcal{B} = 1$  line, where the shape of the diagram becomes less sensitive to the underlying time-scales. This can limit the accuracy of the method.

To optimize the reliability of the results, the tracer for which the time-scale is known (i.e. the ‘reference’ time-scale, which in practice will often refer to the star formation tracer lifetime  $t_{\text{star}}$ ) should be chosen to be as similar as possible to the measured time-scale (which in practice will often refer to the cloud lifetime  $t_{\text{gas}}$ ). This may require iteration or running multiple experiments with different reference maps (e.g. using H $\alpha$ , FUV, and NUV, see Haydon et al. 2018) to converge on the best tracer to use. Figure 18 shows that the inaccuracy of the upward and downward-pointing triangles is not unacceptable – the deviations from the one-to-one relation are generally less than a factor of two and there are several such symbols that lie right on top of the one-to-one relation. However, larger deviations would begin to impede astrophysical applications of the method. For practical applications, we therefore recommend an a posteriori evaluation to discard results for which the difference between  $t_{\text{star}}$  and  $t_{\text{gas}}$  exceeds an order of magnitude. By contrast, time-scale measurements can be considered reliable when  $|\log_{10}(t_{\text{star}}/t_{\text{gas}})| \leq 1$ .

#### 4.2.4 Accuracy of the ‘feedback time-scale’ $t_{\text{over}}$

The second test carried out using the 88 experiments from Section 4.2.2 is to determine how well the duration of the ‘overlap’ phase ( $t_{\text{over}}$ ) is retrieved by our method. Like  $t_{\text{gas}}$  before,  $t_{\text{over}}$  can

have a variety of physical meanings depending on the gas and star formation tracers used. For the specific, real-Universe example of tracing (molecular) gas with CO and star formation with an ionization tracer like H $\alpha$ , this quantity refers to the time it takes the CO emission from a molecular cloud to disappear once the first ionising photons have started to emerge from the embedded star-forming region. We refer to this duration as the feedback time-scale. Analogously to Section 4.2.3, we obtain the complete PDF of  $t_{\text{over}}$  and show the best-fitting value of  $t_{\text{over}}$  (referred to as  $t_{\text{over,out}}$ ) as a function of the input value of  $t_{\text{over}}$  from Table 6 (referred to as  $t_{\text{over,in}}$ ) for all 88 experiments in Figure 19. On both axes, we normalise  $t_{\text{over}}$  to the total duration of the timeline  $\tau$ , thus comparing the input fraction of a region’s lifetime during which it is visible in both maps to the retrieved fraction.

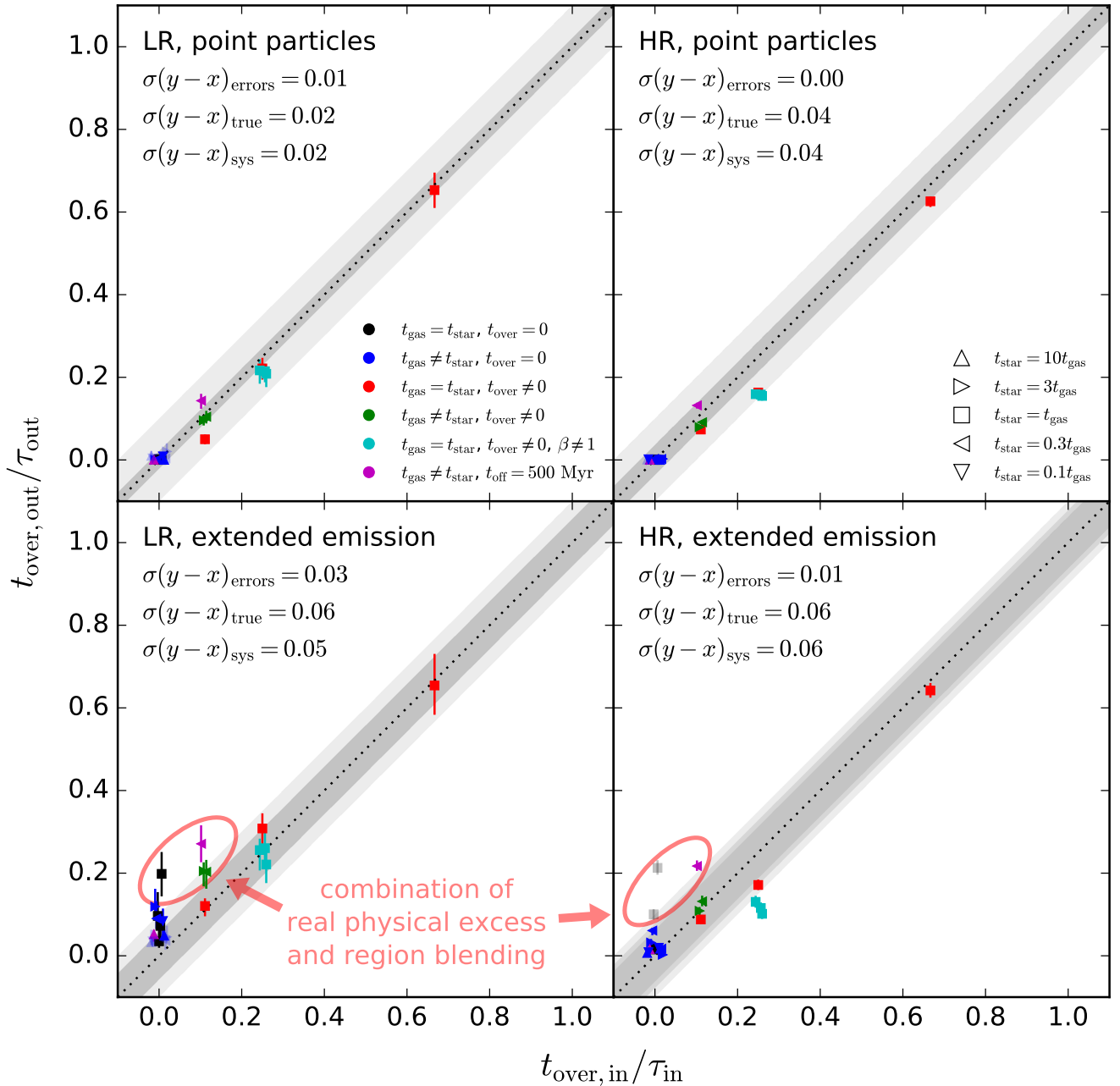
**4.2.4.1 Point particles** As in Section 4.2.3, we expect the point particle experiments to yield the best results, owing to the straightforward identification of independent regions. The top panels of Figure 19 indeed demonstrate good agreement, with standard deviations around the one-to-one relation ( $\sigma_{\text{true}}$ ) of 2–4 per cent in  $t_{\text{over}}/\tau$ . In both cases, this scatter is larger than expected based on the errors, indicating that there is a systematic uncertainty associated with the method. This uncertainty does not take the form of a systematic bias, but represents an increase of the statistical uncertainty. As previously, there are two possible sources of the added scatter. Firstly, it is possible that the substructure in the spatial distribution of regions somehow influences  $t_{\text{over}}$  more strongly than  $t_{\text{gas}}$  (on which the influence for the amount of substructure in these simulations is negligible, see Section 4.2.3). Secondly, deviations from a constant SFR could lead to biases in the measurements of  $t_{\text{over}}$ . This should affect  $t_{\text{over}}$  more strongly than  $t_{\text{gas}}$ , because it covers a shorter duration and thus is more susceptible to Poisson noise.

We test the hypothesis that variations in the SFR are responsible for the main outliers in the top panels of Figure 19. In the left-hand panel, the main outlier is experiment 13 (the red square at  $\{t_{\text{over,in}}/\tau_{\text{in}}, t_{\text{over,out}}/\tau_{\text{out}}\} = \{0.11, 0.05\}$ ), whereas in the right-hand panel, the main outliers are experiments 14, 18, 19, and 20 (the cluster of squares at  $\{t_{\text{over,in}}/\tau_{\text{in}}, t_{\text{over,out}}/\tau_{\text{out}}\} = \{0.25, 0.16\}$ ). The latter four experiments cover the same age bins and only differ in terms of the overlap-to-isolated flux ratios  $\beta_{\text{star}}$  and  $\beta_{\text{gas}}$  (see Section 4.2.5). The fact that these four points show an almost identical discrepancy compared to the input values while sharing the same timeline coverage strongly suggests that SFR variations are the cause of their deviation. We verify this by comparing the fraction  $f_M(t_{\text{over}})$  of the total stellar mass in each map formed during  $t_{\text{over}}$  to the corresponding time fractions  $t_{\text{over,in}}/\tau_{\text{in}}$  and  $t_{\text{over,out}}/\tau_{\text{out}}$ . If SFR variations cause the observed offsets, then  $f_M(t_{\text{over}})$  should match the measured value of  $t_{\text{over,out}}/\tau_{\text{out}}$  rather than the stellar mass fraction expected for a constant SFR  $t_{\text{over,in}}/\tau_{\text{in}}$ .

For experiment 13 in the top-left panel of Figure 19, we find that  $f_M(t_{\text{over}}) = 0.05$ , whereas for the cluster of points around experiment 14 in the top-right panel, we find that  $f_M(t_{\text{over}}) = 0.16$ . Within the displayed precision, both of these numbers are identical to the retrieved values of  $t_{\text{over,out}}/\tau_{\text{out}}$  (0.05 and 0.16, respectively). Therefore, we conclude that the systematic uncertainty of  $\sigma_{\text{sys}} = 2\text{--}4$  per cent identified in the top panels of Figure 19 is caused by deviations from a constant SFR during the covered part of the timeline. This is a real effect that will also affect observational applications of our method. Assuming that the simulations considered in this paper provide a reasonable representation of SFR

<sup>19</sup> We note that the shape of the Wendland  $C^4$  smoothing kernel used here to generate the maps is very similar to a Gaussian.





**Figure 19.** Accuracy of the ‘overlap’ phase lifetimes ( $t_{\text{over}}$ ) retrieved by our new method, normalised to the total duration of the timeline ( $\tau$ ), for the low-resolution (LR, left panels) and high-resolution (HR, right panels) simulations, using point particles (top panels) and extended emission (bottom panels). Shown is the best-fitting value  $t_{\text{over,out}}/\tau_{\text{out}}$  as a function of the input value  $t_{\text{over,in}}/\tau_{\text{in}}$  specified for each experiment (see Table 6). The data points indicate the results for each experiment (see Table 6 and the legends in the top panels for the meanings of the symbols), with error bars indicating the 16th and 84th percentiles of the  $t_{\text{over,out}}/\tau_{\text{out}}$  PDF. Note that only four values of  $t_{\text{over,in}}/\tau_{\text{in}} = \{0, 0.11, 0.25, 0.67\}$  are considered in these experiments. The small horizontal offsets of the data points relative to these exact values have been added to improve the legibility of the figure. Transparent symbols do not pass the guidelines for the reliable application of our method presented in Section 4.4 below and should therefore be omitted from further analysis. The red ellipses and annotation in the bottom panels indicate real outliers that are not due to any inaccuracy of the method (part of this  $t_{\text{over}}/\tau$  excess is caused by the fact that the extended emission maps contain peaks consisting of multiple particles that have a non-zero age spread, see the text). The dotted line indicates the one-to-one agreement and the grey areas illustrate different amounts of scatter. The lightest shade represents a scatter of 0.1 (i.e. 10 per cent of the timeline  $\tau$ ), the middle shade represents the standard deviation of  $t_{\text{over,out}}/\tau_{\text{out}} - t_{\text{gas,in}}/\tau_{\text{in}}$  (i.e. the scatter around the diagonal) for all data points (including the unreliable, transparent ones), and the darkest shade indicates the standard deviation of  $t_{\text{over,out}}/\tau_{\text{out}} - t_{\text{gas,in}}/\tau_{\text{in}}$  exclusively for the opaque (reliable) data points. Three different measures of the scatter around the one-to-one relation are given in the top left of each panel. The first (subscript ‘errors’) is the scatter around the one-to-one relation expected for the error bars of the opaque data points (this is effectively the mean error bar). The second (subscript ‘true’) is the actual scatter around the one-to-one relation for the opaque data points (corresponding to the darkest shade of grey). The third (subscript ‘sys’) is the difference between the first and second values (defined as  $\sigma_{\text{sys}}^2 = \sigma_{\text{true}}^2 - \sigma_{\text{errors}}^2$ , with the additional requirement that  $\sigma_{\text{sys}}^2 \geq 0$ ), which is the part of the scatter that cannot be accounted for by the error bars. As such, this represents the ‘systematic’ inaccuracy of the method. The intrinsic accuracy of the method is  $\sigma_{\text{sys}} \lesssim 0.05$  (see the text), implying that  $t_{\text{over}}$  can be determined to about 5 per cent of  $\tau$ . Any  $t_{\text{over}} < 0.05\tau$  is therefore consistent with  $t_{\text{over}} = 0$ .

variations in disc galaxies, we recommend a minimum uncertainty of 5 per cent in  $t_{\text{over}}/\tau$ , even if the formal error is smaller. In practice, this means that  $t_{\text{over}}/\tau < 0.05$  is consistent with no overlap and  $t_{\text{over}}/\tau > 0.95$  is consistent with completely overlapping phases.

**4.2.4.2 Extended emission** Contrary to the good performance of the method in retrieving  $t_{\text{over}}/\tau$  when applying it to the point particle maps in the top panels of Figure 19, the bottom panels of Figure 19 show a group of experiments with measured values of  $t_{\text{over}}$  that are significantly higher than the input values (highlighted by the light red ellipses). This excess of  $t_{\text{over}}$  only occurs in some (but not all) of the experiments with low  $t_{\text{over,in}}/\tau_{\text{in}}$ . While this deviation from the one-to-one relation may look like a major inaccuracy of the method at first sight, it is largely physical in nature.

When determining how well  $t_{\text{over}}$  is retrieved from the extended emission maps discussed in Figure 17, it is important to realise that only the point particle maps may exhibit one-to-one agreement between input and output. In these maps, it is possible to control the value of  $t_{\text{over}}$  by specifying the age ranges for which the particles are included in the map, because individual star particles have a single formation time. This does not hold for the extended emission maps, because their independent regions consist of groups of star particles that have a non-zero, intrinsic age spread – the star formation history of each progenitor cloud is not a delta function. As a result, the descendant groups of star particles may be present in both maps even in the experiments for which we specify  $t_{\text{over,in}} = 0$ . Therefore, any value of  $t_{\text{over}}$  retrieved from the extended emission maps should correspond to the sum of the number specified in Table 6 and the age spread of each region. This explains why mostly experiments with low  $t_{\text{over,in}}/\tau_{\text{in}}$  show the  $t_{\text{over}}$  excess. In these experiments, the intrinsic age spread of the regions in the simulation is more likely to dominate over the imposed age spread  $t_{\text{over,in}}$ .

It is important to realise that  $t_{\text{over}}$  can only contain an intrinsic age spread for regions that consist of a sufficient number of star particles to sample this age spread. In practice, the mean number of particles per region depends on the total duration of each phase. A very rough estimate of this number can be made by dividing the number of regions in the point particle experiments by the number of regions in the extended emission experiments (cf. the tables in Appendix D1). A minimum of two particles per region is required for both tracers to sample any age spread at all. This is only consistently achieved in experiments 3 and 4 (see Table 6), where the duration of both phases exceeds 30 Myr and the mean number of star particles per region is 4–15 for each phase. The other experiments have less than two particles per region on average, implying that the age spread is only resolved in a subset of the regions in the map. This dependence on the numerical sampling implies that the excess of  $t_{\text{over}}$  depends on the total duration of the timeline, and is only fully included in experiments 3 and 4. For phase durations shorter than  $\sim 30$  Myr, the  $t_{\text{over}}$  excess is smaller than the true age spread, because the likelihood of sampling the timeline with a sufficient number of particles to cover the full overlap phase is smaller than unity.

We reiterate that the above increase of the intrinsic age spread with the total duration of the timeline is entirely numerical in nature and results from the quantisation of the mass reservoir into particles of  $1.4 \times 10^4 M_{\odot}$  (low resolution) and  $2.7 \times 10^3 M_{\odot}$  (high resolution). In real star-forming regions, this type of quantisation could potentially take place in ionization-based star formation tracers such as H $\alpha$ , because such emission is generated by individual mas-

sive stars. However, such a quantisation will have a much weaker effect than in our simulations, because roughly one O-star is formed per  $100 M_{\odot}$ , assuming a normal IMF (e.g. Chabrier 2003). A single star particle in our high-resolution simulation would host 27 O-stars, i.e. the sampling rate of a real star-forming region is a factor of  $\sim 30$  higher, meaning that an age spread like the one we see in the simulations would be well-sampled already for  $t_{\text{star}} \sim 1$  Myr rather than the limit of  $t_{\text{star}} \sim 30$  Myr that we see in the experiments. The commonly-used star formation tracers H $\alpha$ , FUV, and NUV all cover time-scales well in excess of 1 Myr (Haydon et al. 2018). Therefore, we conclude that the method is capable of accurately retrieving the overlap (or feedback) time-scale from observed (molecular) gas and star formation tracer maps.

The bottom panels of Figure 19 contain two additional transparent data points (indicative of not satisfying the guidelines for the reliable application of our method presented in Section 4.4 below) relative to those in Figure 18. These are the largest outliers in the bottom-right panel and correspond to experiments 3 and 4, which have large-to-extreme  $t_{\text{over}}$  excesses of  $t_{\text{over,out}} = 4.87^{+0.58}_{-0.62}$  Myr and  $t_{\text{over,out}} = 39.27^{+3.13}_{-2.74}$  Myr, respectively. Visual inspection of the maps reveals extreme blending of the regions, caused by the long timelines covered by these experiments (see Table 6) and the resulting high spatial density of the regions. The large discrepancy between the input and output value of  $t_{\text{over}}$  in these experiments highlights that blending can negatively affect the accuracy of measuring  $t_{\text{over}}$  with our method.

Interestingly, blending does not influence the accuracy of the  $t_{\text{gas}}$  measurements in Section 4.2.3, because it affects both phases of the timeline in Figure 1 in proportion to their number densities in the maps and therefore in proportion to their durations, resulting in the same *relative* change of either time-scale. Fundamentally, our method measures the relative time-scale difference between both phases, implying that the measurement of  $t_{\text{gas}}$  is not affected by blending. However, this does not hold when measuring  $t_{\text{over}}$ , because that quantity is obtained by measuring the correlation between both phases in excess of the statistical correlation expected for a random distribution in the plane. It is easy to picture the influence of blending on this measurement with an example. If one would convolve both the stellar and ‘gas’ maps with a point spread function that matches the size of the simulated galaxy, both phases would be fully correlated in the convolved maps and the ratio between both maps would equal the galactic average everywhere. Naturally, this should result in a duration of the overlap phase equal to the total duration of the timeline, irrespective of what the true duration of the overlap phase would have been in the unconvolved maps. This extreme example shows that blending adds to the true time overlap between the phases.

In order to minimize the effects of blending on the measurement of  $t_{\text{over}}$  with our method, two potential effects should be accounted for. Firstly, blending can inhibit the identification of independent regions, because the blending decreases the contrast between the flux density minimum in between adjacent peaks and the peaks themselves. For a logarithmic spacing  $\delta \log_{10} \mathcal{F}$  between the flux levels used to identify peaks (see Section 3), there exists a minimum distance between the peaks (in units of the peak radius) such that they can still be identified. The mathematical derivation of this distance is provided in Appendix B, but here we give the ratios between the separation lengths and the peak radius needed to identify the peaks in our experiments. For the {low, high} resolution simulations, we use a logarithmic spacing of  $\delta \mathcal{F} = \{0.25, 0.5\}$ , which requires that  $\lambda/r > \{3.14, 3.84\}$  or, when expressed in

terms of the region concentration parameter  $\zeta \equiv 2r/\lambda$ , requires that  $\zeta < \{0.64, 0.52\}$ .

Secondly, the blending of adjacent peaks also causes the emission of each individual region to be increased by some factor. This leads to an overestimation of the duration of the overlap phase, because blending combines the emission from regions with independent (and therefore most likely different) ages into a single region. We quantify this effect in Appendix B, where we consider two adjacent, two-dimensional Gaussian peaks and calculate as a function of  $\lambda/r$  which fraction of the emission on one side of the equidistance line is constituted by contamination from the neighbour. This fraction steeply decreases as a function of distance and should be lower than the desired accuracy of the  $t_{\text{over}}$  measurement. The top panels of Figure 19 demonstrate that the intrinsic accuracy of the method is  $\sim 0.05$  in  $t_{\text{over}}/\tau$  due to deviations from a constant SFR. Therefore, it is desirable that much less than 5 per cent of the emission in any region comes from neighbouring emission peaks. The condition that peaks are separated by  $\sim 1.7$  times their FWHM (corresponding to  $\lambda/r > 4$  or  $\zeta \equiv 2r/\lambda < 0.5$ ) results in a contamination percentage of  $\sim 5$  per cent. This satisfies the above requirements and we will use this condition to determine if blending may have influenced the measurements of  $t_{\text{over}}$ . If regions are marginally resolved, this requirement translates to having a resolution (FWHM) of  $\lambda/1.7$  or better. This is discussed further in Section 4.3.6, where we carry out resolution tests designed to explore these issues further.

The above upper limits on  $\zeta$  (and hence blending) successfully identify experiments 3 and 4 in the bottom-right panel of Figure 19 as an inaccurate measurement of the duration of the overlap phase. It also successfully identifies several of the experiments using the real gas maps of the simulations in Section 4.3 below, which upon visual inspection are all clearly affected by blending as well. However, it fails to identify experiments 4 and 21 in the bottom-left panel of Figure 19, which also have large  $t_{\text{over}}$  excesses due to blending. We expect that this is caused by a large radial variation of  $\zeta$  within the maps, which implies that the mean  $\zeta$  is not representative for a non-negligible fraction of the regions, especially towards the inner galaxy. Therefore, we recommend to always carry out a visual inspection of the maps to verify if they show any region blending. This is particularly useful because the measured value of  $\zeta$  is an average over the entire map. If there is substantial spatial variation of  $\zeta$  (e.g. as a function of galactocentric radius), then it is possible that blending can only be identified by visual inspection, after which parts of the maps may be masked to obtain a more homogeneous and representative measurement of  $\zeta$ . We have expanded the method with a new way of filtering out emission on size scales exceeding the region separation length  $\lambda$  (Hygate et al. 2018), which will help to avoid blending without human intervention.

**4.2.4.3 General conclusions** In summary, we see that the method is capable of measuring the fraction of the timeline in which both phases overlap ( $t_{\text{over}}/\tau$ ) with an accuracy of about 0.05 or 5 per cent. A higher accuracy could be obtained in principle, but fluctuations of the SFR imply that our assumption of a constant SFR is violated, thus limiting the attainable accuracy. We propose that the above accuracy is appropriate for disc galaxies, but will naturally be worse in systems with strongly varying SFRs, i.e. starburst systems like dwarf galaxies or mergers. Should the star formation history of the observed field be known over the entire duration of the timeline  $\tau$  (e.g. by stellar population synthesis fitting), then it is possible to correct the time-scales retrieved by our method for

variations in the SFR over the fraction of the timeline covered by the stellar phase.

The method’s accuracy may be decreased if no measures are taken to avoid the blending of regions in the tracer maps used. For this reason, we recommend that  $\max(\zeta_{\text{star}}, \zeta_{\text{gas}}) < \zeta_{\text{crit}}$  (to allow the regions to be identified by the clump finding algorithm; see Appendix B for the derivation of  $\zeta_{\text{crit}}$  as a function of  $\delta \log_{10} \mathcal{F}$ ) and  $\max(\zeta_{\text{star}}, \zeta_{\text{gas}}) < 0.5$  (to avoid the contamination of peaks by nearby regions). In case of a strong spatial variation of  $\zeta$  across either of the maps, we recommend to carry out a visual inspection to verify the degree of blending and to mask the maps if necessary, or to filter out diffuse emission on size scales larger than  $\lambda$  (Hygate et al. 2018). If any of these conditions is not satisfied and region blending does affect the analysis, then the measured value of  $t_{\text{over}}$  is not necessarily incorrect, but it reflects the duration for which both tracers coexist at the blending scale. This composite value of  $t_{\text{over}}$  may then exceed the ‘Lagrangian’ duration of the overlap phase when following an individual region through its evolution.

As discussed at the beginning of this section, we expect that real-Universe applications of the method’s ability to constrain the duration of the overlap phase will help provide insight in (molecular cloud destruction by) stellar feedback. Of course, the physical meaning of  $t_{\text{over}}$  depends on the adopted tracers. In the example of tracing (molecular) gas with CO and star formation with H $\alpha$ ,  $t_{\text{over}}$  corresponds to the feedback time-scale, i.e. the time over which excess CO emission disappears once ionising photons start to escape from a region.

In order to distinguish between various feedback mechanisms, it is essential that the accuracy of the  $t_{\text{over}}$  measurement allows resolving the time-scale differences expected between these mechanisms. For instance, ruling out SNe in favour of early (e.g. photoionization) feedback would require an accuracy of the order 3 Myr. In view of the expected  $\sim 5$  per cent accuracy in  $t_{\text{over}}/\tau$ , this example would necessitate the use of gas and star formation tracers that together cover  $\tau < 60$  Myr. This can be ensured by adopting tracers with short lifetimes, e.g. H $\alpha$  (which contributes  $t_{\text{star}} \sim 5$  Myr, Haydon et al. 2018) to trace star formation. Contrary to some star formation tracers, the lifetime of the molecular gas tracer is not known a priori. If the molecular gas tracer [e.g.  $^{12}\text{CO}(1-0)$  or  $^{12}\text{CO}(2-1)$ ] turns out to be longer-lived than  $t_{\text{gas}} \sim 50$  Myr, one can probe feedback-driven gas dispersal by switching to a higher-density gas tracer such as  $^{13}\text{CO}(1-0)$  or HCN (e.g. Leroy et al. 2017), which are expected to have much shorter lifetimes. In addition to using the time-scale associated with ionising photons to determine which feedback mechanism is responsible for molecular cloud destruction, it should be possible to distinguish between feedback mechanisms in even more detail if additional tracers can be identified that are associated with single mechanisms (see Section 8.2). Combining the above considerations, our method should then allow measurements of the feedback time-scale in galaxies with an accuracy better than 1 Myr, which is sufficient to distinguish between several of the possible feedback mechanisms. The above distinctions between feedback mechanisms share the benefit of not depending on feedback models. However, it will also be possible to directly compare the cloud destruction time-scales predicted for individual feedback mechanisms (or combinations thereof) to measurements of  $t_{\text{over}}$  obtained with our method.

**Table 7.** Mean best-fitting values of  $\beta$  using age-binned stellar maps

Runs	$\langle\beta_{\text{star}}\rangle_{1.0}$	$\langle\beta_{\text{star}}\rangle_{0.5}$	$\langle\beta_{\text{gas}}\rangle_{1.0}$	$\langle\beta_{\text{gas}}\rangle_{0.5}$
LR points	$0.99^{+0.03}_{-0.02}$	$0.52^{+0.04}_{-0.02}$	$0.94^{+0.04}_{-0.03}$	$0.48^{+0.02}_{-0.00}$
HR points	$0.98^{+0.01}_{-0.00}$	$0.50^{+0.01}_{-0.00}$	$0.93^{+0.01}_{-0.01}$	$0.47^{+0.01}_{-0.00}$
LR extended	$0.95^{+0.26}_{-0.11}$	$0.55^{+0.09}_{-0.05}$	$1.12^{+0.39}_{-0.10}$	$0.87^{+0.03}_{-0.02}$
HR extended	$1.02^{+0.03}_{-0.03}$	$0.74^{+0.03}_{-0.04}$	$0.91^{+0.03}_{-0.02}$	$0.62^{+0.05}_{-0.04}$

#### 4.2.5 Accuracy of the overlap-to-isolated flux ratio $\beta$

Next to the gas phase lifetime  $t_{\text{gas}}$  and the duration of the overlap phase  $t_{\text{over}}$ , our method also measures the mean flux ratio between regions in the overlap phase to those not in the overlap phase (which we refer to as being ‘isolated’ in time rather than spatially). This overlap-to-isolated flux ratio  $\beta$  is calculated for both tracers, implying that we have measurements of  $\beta_{\text{star}}$  and  $\beta_{\text{gas}}$ . As discussed in Section 3 and KL14, knowledge of this flux ratio is sufficient to account for the effect of the flux evolution history of either tracer on the retrieved time-scales. In Table 6, all experiments with a non-zero value of  $t_{\text{over}}$  result in a meaningful measurement of  $\beta_{\text{star}}$  and  $\beta_{\text{gas}}$ . For most experiments, we have set both quantities to unity, but experiments 18–20 are chosen to explore the effects of varying  $\beta$  on the best-fitting quantities.

Table 7 shows the mean obtained values of  $\beta_{\text{star}}$  and  $\beta_{\text{gas}}$  for different subsets of experiments. We divide up the experiments in the four types of Figure 17, i.e. low-resolution point particles, high-resolution point particles, low-resolution extended emission, and high-resolution extended emission. For each of these, the mean output values of  $\beta_{\text{star}}$  and  $\beta_{\text{gas}}$  are calculated for the experiments where the input values are set to unity (columns indicated by  $\langle\beta\rangle_{1.0}$ ) and to 0.5 (columns indicated by  $\langle\beta\rangle_{0.5}$ ). We only consider experiments for which  $t_{\text{over, in}} > 0$ . As expected, the values of  $\beta_{\text{star}}$  and  $\beta_{\text{gas}}$  in the point particle experiments are in excellent agreement with the input values. This shows that the method is capable of capturing the mean flux evolution of independent regions by correctly retrieving the overlap-to-isolated flux ratio, even if the error bars may be underestimated by up to a factor of 2.<sup>20</sup> For the extended emission experiments with input value  $\beta = 0.5$ , the agreement is generally poor. However, as in Section 4.2.4, this is a real effect rather than a flaw in the method, because the regions in the extended emission maps consist of collections of particles. This means that it is no longer possible to control which particles reside in the overlap phase, implying that the overlap-to-isolated flux ratio of these composite regions should be in the range  $\beta = 0.5$ – $1.0$ . Table 7 shows that this is indeed the case to within the error bars. While this broad range clearly shows that the extended emission experiments provide no particularly strong test of the method’s accuracy in retrieving the overlap-to-isolated flux ratios  $\beta_{\text{star}}$  and  $\beta_{\text{gas}}$ , it is encouraging that the output values are consistent with the expected range.

### 4.3 Gas and stellar maps

Our second set of tests is based on maps of both the gas and the star particles in the simulations. Having established the accuracy of the method with the controlled experiments of Section 4.2, we can

<sup>20</sup> We obtain this factor by considering that across eight numbers, only two or three should deviate by more than  $1\sigma$ , whereas currently this applies to half of the measurements. Increasing the error bars by a factor of 2 would result in two numbers deviating by more than  $1\sigma$ .

now turn to a more realistic set of experiments. By using the actual distributions of gas particles in the simulations and combining these with the ‘age-binned’ stellar maps as before, we can investigate dependences that may affect practical applications of the method to observations. Specifically, we consider the effects of

- (i) diffuse emission in the maps, i.e. emission emerging from scales  $> \lambda$ ;
- (ii) the (stellar) reference time-scale, reflecting the star formation tracer used (e.g. Leroy et al. 2012; Haydon et al. 2018);
- (iii) the volume density threshold above which the gas is visible, reflecting the gas tracer used (e.g. Shirley 2015; Leroy et al. 2017);
- (iv) the spatial resolution, reflecting the observational setup and the target distance or redshift;
- (v) the galaxy inclination angle;
- (vi) the number of independent regions in the maps, represented by the identified number of peaks.

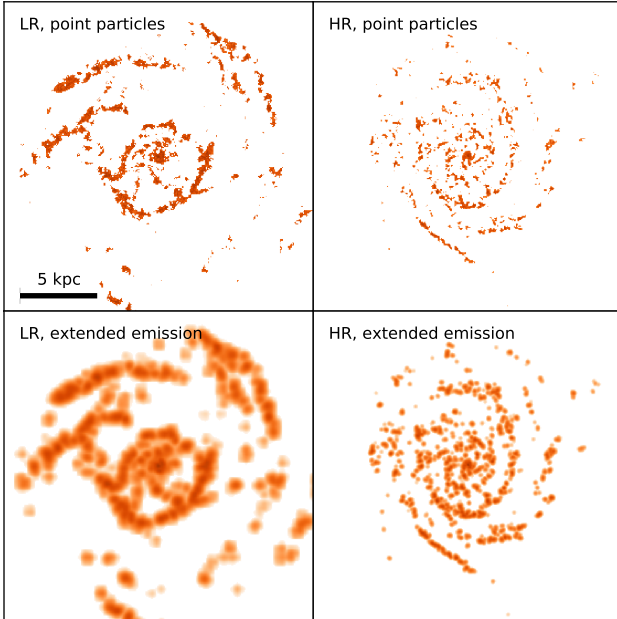
#### 4.3.1 Procedure for creating the maps

For the experiments discussed here, the age-binned maps of the star particles are generated for different age ranges in exactly the same way as described in Section 4.2.1. To generate maps of the gas particles in the galaxy models, we first need to define the minimum volume densities above which the particles are displayed. In the phase diagrams of Figure 16, vertical white lines indicate volume densities of  $\rho = \{0.5, 1, 2\} \text{ cm}^{-3}$  (low resolution) and  $\rho = \{1, 2, 4\} \text{ cm}^{-3}$  (high resolution). As shown by the hatched region, these critical densities provide a reasonable match to the minimum density for particles to become star-forming ( $\rho = 0.7 \text{ cm}^{-3}$ ) and together span roughly an order of magnitude, across which different fractions of the star-forming ISM are traced. Therefore, we use the three volume densities indicated in each panel of Figure 16 to generate the gas particle maps, where the central values of  $\rho_{\text{min, LR}} = 1 \text{ cm}^{-3}$  (low resolution) and  $\rho_{\text{min, HR}} = 2 \text{ cm}^{-3}$  (high resolution) act as the default choices.

In addition to the above reference time-scales  $t_{\text{star}}$  and minimum gas volume densities  $\rho_{\text{min}}$ , we specify two other quantities that are investigated here. Firstly, the spatial resolution of the maps is specified by changing  $l_{\text{ap, min}}$  in Table 2 and setting  $N_{\text{ap}}$  such that the other aperture sizes remain unchanged. For instance, to change the spatial resolution to 200 pc instead of the default 50 pc, we set  $l_{\text{ap, min}} = 200 \text{ pc}$  and  $N_{\text{ap}} = 6$  (instead of the default  $N_{\text{ap}} = 8$ ) to retain the same aperture size spacing as before. Secondly, we vary the inclination angle of the galaxy maps. This is achieved by rotating the simulation output around the  $x$ -axis by an angle in the range  $i = [0, 90]^\circ$  prior to generating the maps. The remaining quantities of interest (diffuse emission fraction and number of independent regions) are not specified, but vary sufficiently across the set of experiments to draw meaningful conclusions about their influence on the results. We note that out of the quantities used to generate the maps, only  $t_{\text{star}}$  and  $i$  are used as input parameters in the method of Section 3. The method is not informed of the values of the other quantities. Having said that, the choice of stellar and gas tracer (each associated with their own time-scale or minimum density, respectively) needs to be considered when setting the input parameters  $X_{\text{star}}$  and  $X_{\text{gas}}$  for converting between pixel values and SFR and gas mass.

Having defined the subsets of particles that are shown in each map, we generate the surface density maps of these particles as described in Section 4.2.1. The maps are square with a projected diameter of 20 kpc and 1403 pixels, corresponding to a projected





**Figure 20.** The four different types of ‘gas’ maps used in Section 4.3. From left to right, top to bottom, these are maps of the low-resolution (LR) point particle distribution, the high-resolution (HR) point particle distribution, the LR extended emission (generated by using a smoothing kernel, see the text), and the HR extended emission. In these examples, the {LR, HR} maps show the distribution of gas particles with densities in excess of  $\rho_{\min} = \{1, 2\} \text{ cm}^{-3}$ , respectively. In Section 4.3.2, we present 50 different sets of map pairs for which our method is tested, for a total of  $4 \times 50 = 200$  experiments.

pixel scale of 14.25 pc before correcting for inclination.<sup>21</sup> As before, two types of maps are made for each galaxy model. The ‘point particle’ surface density maps show the distribution of particles as is, for each pixel adding up the enclosed mass of the particles and dividing by the pixel area. At the adopted numerical resolution of the simulations and the pixel scale of 14.25 pc, the typical number of gas particles per pixel is zero or exceeds unity – contrary to the stellar particles, which each generally occupy their own pixel (see Figure 17). This difference arises because high-density gas particles cluster more strongly than the young star particles, which are stochastically spawned from the high-density gas particles. As a result, the point particle maps display a morphology that is broadly similar to the ‘extended emission’ surface density maps, which account for the extended morphology expected for real star-forming regions. This extended emission is obtained by using a smoothing kernel with a smoothing length chosen to be consistent with the treatment of the gas particles in the simulations (see Section 4.1), i.e. by using the 200 nearest neighbours to define the smoothing length and adopting a Wendland  $C^4$  smoothing kernel (Dehnen & Aly 2012) to distribute the mass around each particle position. As before, this results in coarser resolution in low-density regions.

The maps resulting from the above procedure are written to disk in the FITS file format and include a FITS header that contains their key properties, consistent with common conventions for observational data delivery. The header contains the resolution of

the map to determine the minimum aperture size used (see Section 3.2.3), which is taken to be 50 per cent of the minimum smoothing length across the displayed subset of particles. This translation from the smoothing length to the resolution (or FWHM) is a reasonable approximation, because the Wendland  $C^4$  kernel adopted here has an FWHM of 0.56 times the smoothing length. Figure 20 shows four examples of a typical gas map that these experiments are carried out with, illustrating the similarities and differences between the two types of maps (‘point particles’ and ‘extended emission’) for both disc galaxy models (‘low resolution’ and ‘high resolution’), adopting  $\rho_{\min} = \{1, 2\} \text{ cm}^{-3}$  (for {low, high} resolution) and  $i = 0^\circ$ . As expected due to the increased clustering of high-density gas particles relative to young stellar particles, the point particle and extended emission maps differ considerably less than seen in the stellar maps from Section 4.2. In the following experiments, the stellar maps have the same general properties as before (see Figure 17).

### 4.3.2 Description of the experiments

Table 8 summarises the experiments carried out using the combination of age-binned stellar maps and gas maps above a certain critical volume density, specifying the values of  $t_{\text{star}}$ ,  $i$ ,  $\rho_{\min, \text{ref}}$ ,  $l_{\text{ap}, \text{min}, \text{LR}}$ , and  $l_{\text{ap}, \text{min}, \text{HR}}$  for each experiment. Because the high-resolution simulation reaches higher gas volume densities than the low-resolution simulation, we set  $\{\rho_{\min, \text{LR}}, \rho_{\min, \text{HR}}\} = \{1, 2\} \rho_{\min, \text{ref}}$  for the {low, high}-resolution experiments. The experiments are divided into four main categories.

(i) The first set of experiments is aimed at investigating the role of the adopted SFR tracer through varying the reference time-scale. This is an important experiment, because the duration of the young stellar phase cannot be freely chosen in practical applications of the method – it depends on the availability of the observational data. It is therefore necessary to identify any biases of the derived timeline as a function of  $t_{\text{star}}$  or  $t_{\text{star}}/t_{\text{gas}}$ . By spanning 2.5 dex in the reference time-scale (ranging from just 1 Myr in the first experiment to more than an orbital revolution in the eleventh experiment), we can assess whether the choice of reference time-scale affects the accuracy of our method.

(ii) The second set of experiments investigates the choice of gas tracer by considering various minimum volume densities for the gas to be visible. For both the low and high-resolution simulations, the densities cover a factor of 4. As stated above, the minimum densities for the high-resolution simulations are taken to be twice the reference values listed here. As a result, the low and high-resolution experiments together span an order of magnitude in minimum density. For each minimum density, we consider a set of three different reference time-scales  $t_{\text{star}}$ . These experiments are intended to show if applying the method to different gas tracers results in different time-scales, so that a multi-tracer chronology of cloud condensation and collapse can be obtained from observations.

(iii) The third set of experiments probes the dependence on the observational setup and target distance or redshift by varying the spatial resolution through the minimum aperture size  $l_{\text{ap}, \text{min}}$ . We consider spatial resolutions in the range  $l_{\text{ap}, \text{min}} = 25\text{--}800$  pc, each at three different reference time-scales  $t_{\text{star}}$ , with the goal of determining the critical value of the size scale ratio  $l_{\text{ap}, \text{min}}/\lambda$  above which our method fails. This will limit the cosmic distances out to which the method can be applied with current and upcoming observational facilities and, hence, how representatively the galaxy population can be covered.

<sup>21</sup> A non-zero inclination results in non-circular resolution elements that are elongated in the direction of the inclination by a factor of  $(\cos i)^{-1}$ , implying that the effective resolution is larger by a factor of  $(\cos i)^{-1/2}$  [see equation (6)].

(iv) The fourth set of experiments varies the galaxy inclination angle  $i$  to identify any corresponding limits on the applicability of our method. As before, we consider three different reference time-scales for each inclination angle. This set of experiments is expected to show some degeneracy with the third set (spatial resolution), because increasing the inclination results in an increase of the effective resolution scale by a factor of  $(\cos i)^{-1/2}$ .

As outlined in Section 4.3.1, we additionally investigate how the accuracy of the results changes as a function of the number of independent regions hosted in each of the maps and make a first estimate of how diffuse emission affects the results (this latter dependence will be quantified in more detail by Hygate et al. 2018). Both of these are relevant tests for applications of our method to observations. The diffuse fraction cannot always be controlled a priori (e.g. Pety et al. 2013; Caldú-Primo & Schruba 2016) and obtaining an understanding of its effects is thus important. By contrast, if we can identify a minimum number of independent regions needed for reliable applications of the method, this can be an important driver for choosing the spatial coverage in observational campaigns.

The above sets of experiments are carried out using the ‘point particle’ or the ‘extended emission’ surface density maps (see Section 4.3.1), both for the low and high-resolution simulations. We thus have a total of 200 experiments using pairs of age-binned stellar maps and gas maps. Some of the parameters used in these experiments deviate from the default values in Section 3. Firstly, the inclinations and spatial resolutions (i.e. minimum aperture sizes) of the maps vary on a case-by-case basis, as shown by Table 8. We therefore adjust  $i$ ,  $l_{\text{ap},\text{min}}$ , and  $N_{\text{ap}}$  accordingly for each experiment. Secondly, we adjust the range of galactocentric radii to avoid large empty areas in the maps and set  $R_{\text{max}} = 10$  kpc (7 kpc) for the low-resolution (high-resolution) simulations. Thirdly, we set  $N_{\text{pix},\text{min}} = 1$  for the point particle experiments, to enable single pixels to be identified as peaks. Finally, for the low-resolution runs we choose a default resolution of  $l_{\text{ap},\text{min},\text{LR}} = 100$  pc to remove sub-resolution apertures, as well as  $\Delta \log_{10} \mathcal{F} = 2.5$  and  $\delta \log_{10} \mathcal{F} = 0.25$  for the identification of both the stellar peaks and gas peaks, because the low-resolution maps used in this work require a finer and deeper contour level spacing to identify the relevant emission peaks. Contrary to Section 4.2.2, these experiments use the default value of  $t_{\text{star,incl}} = 0$  (see Table 1), implying that  $t_{\text{star,ref}}$  represents the duration of the isolated stellar phase. This choice is motivated by the fact that a gas concentration may form star particles with a certain age spread, implying that the total lifetime of the group of star particles spawned can exceed the duration of the stellar age bin used for generating the map.

In the remainder of this section, we will evaluate how the free parameters constrained by the method (i.e.  $t_{\text{gas}}$ ,  $t_{\text{over}}$ , and  $\lambda$ ) are affected by changing the parameters listed in Table 8, as well as by the number of independent regions. Each of these effects is discussed in a dedicated subsection and makes use of  $2 \times 2 \times 50$  experiments (cf. Table 8).

### 4.3.3 Expected values of the free parameters

Before discussing the measurements of the ‘cloud lifetime’  $t_{\text{gas}}$ , the ‘feedback time-scale’  $t_{\text{over}}$ , and the region separation length  $\lambda$  in each of the experiments discussed above, we derive which values we expect to obtain based on the baryonic physics captured in the simulations. This is helpful for assessing the accuracy of the method, even if this is not quite straightforward, because the experiments discussed in this section are less controlled than those in

**Table 8.** Experiments carried out using gas maps

ID	$t_{\text{star}}$ [Myr]	$i$ [ $^{\circ}$ ]	$\rho_{\text{min,ref}}$ [ $\text{m}_{\text{H}} \text{cm}^{-3}$ ]	$l_{\text{ap},\text{min},\text{LR}}$ [pc]	$l_{\text{ap},\text{min},\text{HR}}$ [pc]
1	1	0.0	1	100	50
2	1.75	0.0	1	100	50
3	3	0.0	1	100	50
4	6	0.0	1	100	50
5	10	0.0	1	100	50
6	17.5	0.0	1	100	50
7	30	0.0	1	100	50
8	60	0.0	1	100	50
9	100	0.0	1	100	50
10	175	0.0	1	100	50
11	300	0.0	1	100	50
12	3	0.0	0.5	100	50
13	10	0.0	0.5	100	50
14	30	0.0	0.5	100	50
15	3	0.0	2	100	50
16	10	0.0	2	100	50
17	30	0.0	2	100	50
18	3	0.0	1	25	25
19	10	0.0	1	25	25
20	30	0.0	1	25	25
21	3	0.0	1	50	100
22	10	0.0	1	50	100
23	30	0.0	1	50	100
24	3	0.0	1	200	200
25	10	0.0	1	200	200
26	30	0.0	1	200	200
27	3	0.0	1	400	400
28	10	0.0	1	400	400
29	30	0.0	1	400	400
30	3	0.0	1	800	800
31	10	0.0	1	800	800
32	30	0.0	1	800	800
33	3	33.6	1	100	50
34	10	33.6	1	100	50
35	30	33.6	1	100	50
36	3	48.2	1	100	50
37	10	48.2	1	100	50
38	30	48.2	1	100	50
39	3	60.0	1	100	50
40	10	60.0	1	100	50
41	30	60.0	1	100	50
42	3	70.5	1	100	50
43	10	70.5	1	100	50
44	30	70.5	1	100	50
45	3	80.4	1	100	50
46	10	80.4	1	100	50
47	30	80.4	1	100	50
48	3	87.1	1	100	50
49	10	87.1	1	100	50
50	30	87.1	1	100	50

Section 4.2. There, the duration of the entire evolutionary timeline was predefined, whereas the current set of experiments only specifies the duration of the (isolated) stellar phase  $t_{\text{star,ref}}$ . The values of  $t_{\text{gas}}$  and  $t_{\text{over}}$  are emergent properties of the simulations. We remind the reader that the goal is not to reproduce the star formation and feedback properties of galaxies in the real Universe, but to retrieve what happens in the simulations. In terms of the sub-grid recipes for star formation and feedback, the simulations differ markedly from observational results. As such, the expected values

derived here are not necessarily relevant in the context of the observational literature, but only serve as a point of reference when interpreting the experiments summarised in Section 4.3.2.

The expected duration of the gas phase  $t_{\text{gas,exp}}$  is obtained by assuming that feedback from star formation destroys the dense gas that is displayed in the gas maps. Fundamentally, star formation in the simulations is possible for all gas particles above the critical density for star formation<sup>22</sup>  $\rho_{\text{crit}}$  and is represented as a Poisson process that takes place on the dynamical time defined in equation (164), with a star formation efficiency  $\epsilon = 0.02$ . These numbers result in an effective star formation time-scale of  $t_{\text{dyn}}/\epsilon$ , which is about 1.6 Gyr at  $\rho_{\text{g}} = \rho_{\text{crit}}$ . This would be the expected gas lifetime if the gas maps would show all gas above the critical density, i.e.  $\rho_{\text{min}} = \rho_{\text{crit}}$ . However, Table 8 shows that most maps have higher minimum gas densities, implying not only shorter dynamical times, but also that some star formation is taking place below the gas visibility threshold. To phrase this in terms of the Lagrangian timeline of Figure 1, the gaseous progenitors of some of the young star particles never reached the gas densities high enough to have been visible in the gas map. The effect of this invisibility in the context of the formalism used here is that the expected gas lifetime is decreased in proportion to the fraction of young star particles for which the progenitors reached densities  $\rho_{\text{g}} \geq \rho_{\text{min}}$ . Statistically, this fraction can be calculated as the probability of a gas particle to have been unaffected by star formation prior to reaching these densities.

It is important to emphasise that the conversion of gas into stars below the density threshold of the simulated gas map is unique to these simulations (as well as many other simulations in the literature) and does not apply to observed gas maps. In the real Universe, the collapse of gas clouds continues to extremely high densities, before the densest gas pockets reach their eventual end state as stars. Gas tracers with increasing critical densities therefore remain the progenitors of these stars and do not represent a phase that can simply be skipped. Indeed, the correlation between gas and star formation becomes tighter with increasing critical density (e.g. Usero et al. 2015), up to values ( $n > 10^5 \text{ cm}^{-3}$ , see e.g. Leroy et al. 2017) that are three orders of magnitude higher than the critical density of a classical molecular gas tracer like CO(1-0). Of course, it may be that the gas tracer is not detected (e.g. at low metallicity), but that is a matter of sensitivity and can be alleviated with a longer observing time. It does not correspond to a fundamental absence of the gas tracer under consideration, contrary to the stochastic star formation model used in the simulations.

In Appendix C, we derive an expression for the probability of gas particles to be unaffected by star formation until reaching a density  $\rho_{\text{g}}$ , under the assumption that the gas density increases with time according to the gravitational free-fall of a homogeneous sphere. This assumption is appropriate for simulations that resolve the disc scale height, which is achieved for our high-resolution simulation, but not for the low-resolution simulation. We account for this by introducing a correction factor below. Note that in the statement ‘unaffected by star formation’, we include any of the  $N_{\text{ngb}}$  particles within a particle’s smoothing kernel. The final expression

obtained this way is

$$p_{\text{gas}}(\rho_{\text{g}}) = \frac{1 + \sqrt{1 - (\rho_{\text{crit}}/\rho_{\text{g}})^{1/3}}}{1 - \sqrt{1 - (\rho_{\text{crit}}/\rho_{\text{g}})^{1/3}}}^{-\sqrt{3/2}\epsilon N_{\text{ngb}}}, \quad (166)$$

which is defined for  $\rho_{\text{g}} \geq \rho_{\text{crit}}$  and behaves as expected: for  $\rho_{\text{g}} = \rho_{\text{crit}}$ , we obtain  $p_{\text{gas}} = 1$ , whereas in the limit  $\rho_{\text{g}} \rightarrow \infty$ , we find  $p_{\text{gas}} = 0$ .

We can now combine the gas particle ‘survival fraction’ with the time-scale for star formation to obtain an expression for the expected lifetime of gas with densities  $\rho_{\text{g}} \geq \rho_{\text{min}}$

$$t_{\text{gas,exp}}(\rho_{\text{min}}) = \left[ \frac{t_{\text{dyn}}(\rho_{\text{min}})}{\epsilon} + t_{\text{fb}}(\rho_{\text{min}}) \right] \frac{p_{\text{gas}}(\rho_{\text{min}})}{f_t}, \quad (167)$$

where we include the time expected for feedback to remove the gas once star particles have formed ( $t_{\text{fb}}$ , as defined below). The expression also includes a correction factor  $f_t \geq 1$ , which applies to simulations that do not resolve the disc scale height and thus have insufficient resolution to estimate the mid-plane free-fall time using a kernel average. For the high-resolution simulation, the scale height of the gas disc is marginally resolved and we set  $f_t = 1$ , but for the low-resolution simulation this is not the case, implying that the mass distribution within the smoothing kernel is anisotropic and the kernel-averaged free-fall time overestimates the true free-fall time in the disc mid-plane. In addition, the intrinsic feedback energy of a single star particle in the low-resolution simulation is higher by a factor of 5 than that in the high-resolution simulation due to the difference in particle mass, causing feedback events from single particles (i.e. 0.5 per cent of a smoothing kernel) to be much more disruptive at fixed ambient gas pressure. This may lead to the destruction of the gas condensation by the first star particle that forms, whereas this need not be the case at higher resolution. Both of these effects mean that the expected gas lifetime is overestimated for the low-resolution simulation. We can correct for this problem by setting  $f_t > 1$ . For the experiments based on the low-resolution simulation discussed in this section, a value  $f_t = 3$  provides an accurate match to the obtained results.

Using the maps at the fiducial densities (e.g. experiment IDs 1–11) as an example, equation (167) predicts that we should expect to find gas phase lifetimes of  $t_{\text{gas,exp}} = 15$  Myr for the low-resolution simulations and  $t_{\text{gas,exp}} = 2.6$  Myr for the high-resolution simulations. As stated previously, these values have no physical meaning, but act as a point of reference when interpreting the results of the experiments.

The expected duration of the overlap phase  $t_{\text{over,exp}}$  can be obtained analogously to the duration of the gas phase. Following Kruijssen (2012, equation 20), Reina-Campos & Kruijssen (2017, equation 4), and Pfeffer et al. (2018, equation 5), we define the expected feedback time-scale as the time required for the feedback-driven energy density to overcome the ambient turbulent gas pressure, i.e.

$$t_{\text{fb}} = \frac{t_{\text{SN}}}{2} \left( 1 + \sqrt{1 + \frac{4t_{\text{dyn}}(\rho_{\text{g}})\sigma^2}{\chi_{\text{fb,E}}\phi_{\text{fb}}\epsilon t_{\text{SN}}^2}} \right), \quad (168)$$

where  $t_{\text{SN}} = 3$  Myr is the time after which the SN feedback is applied,  $\sigma \approx 15 \text{ km s}^{-1}$  is the local turbulent velocity dispersion, and  $\phi_{\text{fb}} \approx 3.2 \times 10^{-4} \text{ m}^2 \text{ s}^{-3} \approx 10^4 (\text{km s}^{-1})^2 \text{ Myr}^{-1}$  is a constant indicating the energy deposition rate per unit mass for a normal (e.g. Chabrier 2003) stellar IMF, and  $\chi_{\text{fb,E}} \approx 1 \times 10^{-3}$  is the feedback-to-ISM coupling efficiency that we typically find for the experiments with minimum densities closest to the density

<sup>22</sup> This critical density corresponds to hydrogen particle density of  $n_{\text{H}} = 0.5 \text{ cm}^{-3}$  and a total mass density of  $\rho_{\text{crit}} = 0.68 \text{ m}_{\text{H}} \text{ cm}^{-3}$  for an assumed hydrogen mass fraction of 0.76.

threshold for star formation. As before, we account for star formation below the density threshold used to generate the gas maps, i.e. stellar regions of which the gaseous progenitor would have been invisible at the adopted minimum volume density, by writing

$$t_{\text{over,exp}} = t_{\text{fb}}(\rho_{\text{min}}) \frac{p_{\text{gas}}(\rho_{\text{min}})}{f_t}, \quad (169)$$

with  $p_{\text{gas}}(\rho_g)$  as defined in equation (166). This final expression provides the expected measured duration of the overlap phase. For the maps at the fiducial densities (e.g. experiment IDs 1–11), equation (169) predicts that we should expect to find overlap phase lifetimes of  $t_{\text{over,exp}} = 1.7$  Myr for the low-resolution simulations and  $t_{\text{over,exp}} = 0.35$  Myr for the high-resolution simulations.

Finally, the expected region separation length  $\lambda_{\text{exp}}$  can be estimated by geometric considerations. For this, we require the SFR, the total duration of the evolutionary timeline, and the stellar mass formed per region to calculate the total number of regions, which in combination with the total area of the map gives an indication of how ‘crowded’ the maps are and what the expected separation length is. First, we calculate the expected total number of regions across both the gas and the young stellar maps as

$$N_{\text{reg}} = \frac{\tau_{\text{exp}} \text{SFR}}{N_{\text{part,sf}} m_{\text{part}}}, \quad (170)$$

where we use  $\text{SFR} = 0.3 M_{\odot} \text{ yr}^{-1}$  (see Section 4.1.4) and set  $m_{\text{part}}$  according to Table 5 for the low-resolution and high-resolution simulations. This expression assumes conservation of numbers, such that a single stellar region is formed from a single gaseous region. The expected total duration of the evolutionary timeline  $\tau_{\text{exp}}$  is

$$\tau_{\text{exp}} = t_{\text{gas,exp}} + t_{\text{star,ref}}, \quad (171)$$

which does not subtract  $t_{\text{over,exp}}$  (see below), because in these experiments we have set  $t_{\text{star,incl}} = 0$  (see Table 1) and hence  $t_{\text{star,ref}}$  represents the duration of the isolated stellar phase (see Section 4.3.2). Finally, the number of star particles formed per region ( $N_{\text{part,sf}}$ ) is obtained by combining the typical star formation efficiency per star formation event that we find in Section 5 ( $\epsilon_{\text{sf}} \approx 0.01$ ) with the number of neighbours, i.e.

$$N_{\text{part,sf}} = \epsilon_{\text{sf}} N_{\text{ngb}}, \quad (172)$$

which results in  $N_{\text{part,sf}} = 2$  and is consistent with the idea that the first 1–2 star particles formed in a region cause its destruction by feedback. Together, equations (170)–(172) provide the total number of regions expected to be present in both maps.

With the number of regions in hand, their geometrically expected mean separation length follows as

$$\lambda_{\text{geo}} = 2 f_{\text{struc}} \sqrt{\frac{R_{\text{max}}^2 - R_{\text{min}}^2}{N_{\text{reg}}}}, \quad (173)$$

where  $R_{\text{min}}$  and  $R_{\text{max}}$  reflect the minimum and maximum radii of the area considered (see Table 2), which for the experiments discussed here are set to  $R_{\text{min}} = 0$  and  $R_{\text{max}} = \{10, 7\}$  kpc for the {low, high}-resolution simulations. The quantity  $f_{\text{struc}}$  in equation (173) is a correction factor that is necessary to account for the substructure in the map. For a random, unclustered distribution of points, we have  $f_{\text{struc}} = 1$ . However, the simulated galaxy maps contain substructure and voids (see Figure 20), which means that the particles cluster in space and the geometric average implied by setting  $f_{\text{struc}} = 1$  overestimates the region separation length in the direct vicinity of emission peaks. Fortunately, the value of  $f_{\text{struc}}$  can be constrained using the quantities  $\mathcal{E}_{\text{star,glob}}$  and  $\mathcal{E}_{\text{gas,glob}}$  (see

Section 3.2.9), which represent the flux surface density contrast between a size-scale  $\lambda$  and the entire map, thus providing a measure of the degree of flux clustering. Assuming that changes in the flux density with spatial scale are driven by changes in filling factor, we expect that  $f_{\text{struc}}$  scales with  $\mathcal{E}_{\text{glob}}$  as

$$f_{\text{struc}} \approx \mathcal{E}_{\text{glob}}^{-1/2}. \quad (174)$$

In the experiments considered here, we measure mean values of  $\mathcal{E}_{\text{glob}} = 6$ –12 for the low-resolution simulation and  $\mathcal{E}_{\text{glob}} = 5$ –10 for the high-resolution simulation, resulting in  $f_{\text{struc}} = 0.29$ –0.41 and  $f_{\text{struc}} = 0.32$ –0.46, respectively. Therefore, we adopt  $f_{\text{struc}} = 0.35$  for the low-resolution simulation and  $f_{\text{struc}} = 0.40$  for the high-resolution simulation.<sup>23</sup>

For an infinitely large number of regions,  $\lambda_{\text{geo}}$  approaches zero. However, infinitesimally small separation lengths cannot be attained in practice due to limits imposed by the spatial resolution of the maps. As will be discussed in Section 4.3.6, the retrieved separation length never drops below two resolution elements. We therefore expand the expression from equation (173) to write the expected region separation length as

$$\lambda_{\text{exp}} = \sqrt{\lambda_{\text{geo}}^2 + (2l_{\text{ap,min}})^2}, \quad (175)$$

in which the first term represents the geometrically expected separation length and the second term reflects the minimum separation imposed by the spatial resolution of the maps. For the experiments of Table 8, the range of expected region separation lengths is  $\lambda_{\text{exp}} = 200$ –1000 pc for the low-resolution simulation and  $\lambda_{\text{exp}} = 100$ –400 pc for the high-resolution simulation.

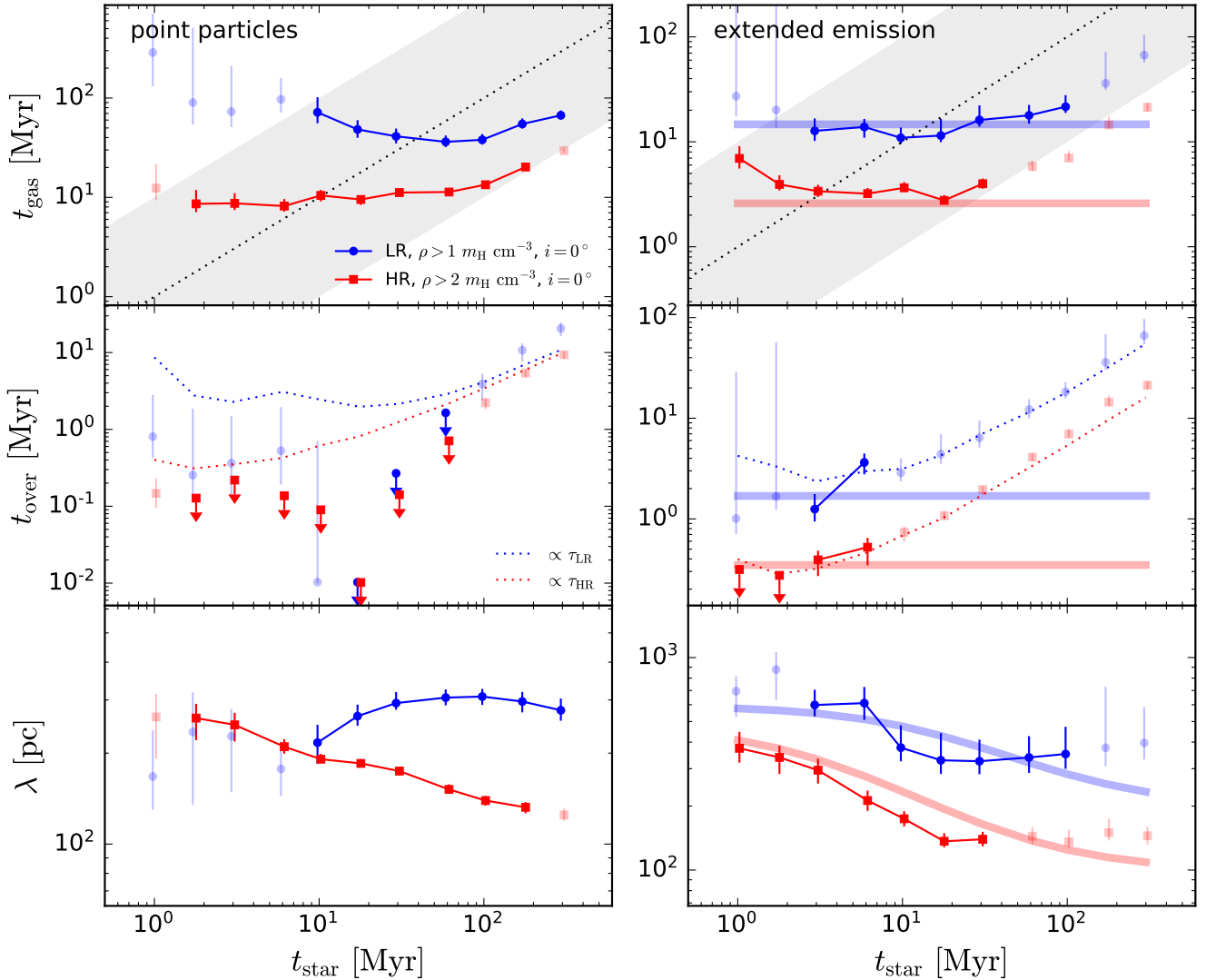
The expected values of  $t_{\text{gas,exp}}$ ,  $t_{\text{over,exp}}$ , and  $\lambda_{\text{exp}}$  from equations (167), (169), and (175) will be included in the figures discussed in Sections 4.3.4–4.3.7. Note that they should only be compared to measurements using the extended emission maps (cf. Figure 20). Any comparison to the results for the point particle maps would not be insightful, because each of the expressions depends on kernel-averaged densities and the number of neighbours used to construct the kernel. As will be shown below, the method accurately retrieves the numbers expected from the above expressions when applied to the extended emission maps.

#### 4.3.4 Dependence on the reference time-scale

We first carry out a test using the top  $4 \times 11 = 44$  experiments presented in Section 4.3.2 (i.e. ID 1–11) to determine how the choice of the reference map (which sets the reference time-scale  $t_{\text{star,ref}}$ ) affects the constrained quantities  $t_{\text{gas}}$ ,  $t_{\text{over}}$ , and  $\lambda$ . In Section 4.2.3, we showed that the reference time-scale is important, because the method provides accurate constraints on cloud-scale star formation and feedback when  $t_{\text{star}}$  and  $t_{\text{gas}}$  differ by at most an order of magnitude. Fortunately, real-Universe applications of our method have some freedom in making the choice of reference map. When using an SFR tracer to set the reference time-scale,  $\text{H}\alpha$ , FUV, and

<sup>23</sup> We stress that these numbers are specific to the simulations used. It is unclear whether they would also provide a reasonable description of observed galaxies and it is undoubtedly a function of the galaxy morphology. In observational applications, the quantities  $\mathcal{E}_{\text{star,glob}}$  and  $\mathcal{E}_{\text{gas,glob}}$  can be used to infer the degree of spatial clustering and, through equation (174), the difference between the geometrically-expected value of the separation length for a random distribution of regions (corresponding to  $f_{\text{struc}} = 1$ ) and the measured, typical separation length around emission peaks in a possibly substructured medium ( $f_{\text{struc}} < 1$ ).





**Figure 21.** Influence of the choice of SFR tracer (captured by varying the reference time-scale  $t_{\text{star}}$ ) on the best-fitting values of the duration of the gas phase ( $t_{\text{gas}}$ , top), the duration of the overlap phase ( $t_{\text{over}}$ , middle), and the mean separation length between independent regions ( $\lambda$ , bottom), for experiments ID 1–11 in Table 8 using point particle maps (left panels) and extended emission maps (right panels). Transparent symbols represent experiments that do not satisfy each of the conditions for the reliable application of the method summarised in Section 4.4. Non-transparent symbols are connected by lines as shown by the legend. In the right-hand panels, the thick blue (low resolution) and red (high resolution) lines show the expected values from Section 4.3.3. In the top panel, the black dotted line indicates  $t_{\text{gas}} = t_{\text{star}}$  with the grey-shaded area indicating the area where both time-scales differ by less than a factor of 10 (i.e.  $|\log_{10}(t_{\text{star}}/t_{\text{gas}})| \leq 1$ ). In the middle panel, downward arrows indicate upper limits (see Section 4.4) and the blue (low resolution) and red (high resolution) dotted lines illustrate a proportionality to the total duration of the timeline  $\tau$ , as indicated by the legends. This figure shows that the (non-transparent) values of  $t_{\text{gas}}$ ,  $t_{\text{over}}$ , and  $\lambda$  behave as expected, in that they either show little variation with  $t_{\text{star}}$  or follow the predicted dependence, implying that the method can be applied reliably to constrain the cloud-scale physics of star formation and feedback using a variety of star formation tracers.

NUV together cover characteristic time-scales from  $t_{\text{star,ref}} = 5\text{--}50$  Myr (Haydon et al. 2018). This provides a large dynamic range, implying that in many observational applications it is feasible to minimise the difference between  $t_{\text{star}}$  and  $t_{\text{gas}}$ . To quantify the stability of the constrained quantities across a wide dynamic range, we consider values of  $t_{\text{star}} = 1\text{--}300$  Myr, which extends well beyond the limits of observational applications.

Figure 21 shows the best-fitting values of  $t_{\text{gas}}$ ,  $t_{\text{over}}$ , and  $\lambda$  for all 44 experiments considered here. As discussed in Section 4.3.3, when calculating the expected values of the constrained quantities we make use of the particle smoothing kernels, which cover  $N_{\text{ngb}} = 200$  neighbours. Therefore, the expected values are only shown in comparison to the extended emission experiments in the

right-hand panels. Across the diagram, we see that the agreement with the expected values is good. The non-transparent symbols either show no dependence on  $t_{\text{star}}$  (for  $t_{\text{gas}}$  and  $t_{\text{over}}$ ) or the expected dependence (for  $\lambda$ ). The first-order result of these experiments is therefore that the method can be applied reliably, irrespective of the choice of reference map, as long as the requirements from Section 4.4 below are satisfied.

Specifically, the top panels of Figure 21 confirm the finding of Section 4.2.3 that the duration of the gas phase is accurately constrained provided that  $t_{\text{star}}$  and  $t_{\text{gas}}$  differ by at most an order of magnitude, which is indicated by the grey-shaded region in the figure. Experiments outside of this area have highly dissimilar durations of the (young) stellar and gas phases, which results

in highly asymmetric tuning fork diagrams. In such cases, one of the two branches in the diagram closely approaches the  $\mathcal{B} = 1$  line (corresponding to the galactic average gas-to-stellar flux ratio), where the shape of the diagram is much less sensitive to the underlying time-scales. As a result, the method becomes less accurate at constraining the time-scales and the measurements of  $t_{\text{gas}}$  begin to deviate considerably from the expected values, in extreme cases even to the extent that  $t_{\text{gas}} \propto t_{\text{star}}$ . However, as long as  $|\log_{10}(t_{\text{star}}/t_{\text{gas}})| \leq 1$ , we find that  $t_{\text{gas}}$  varies as a function of  $t_{\text{star}}$  by less than a factor of 2, and typically by less than 50 per cent. The most stable and reliable solutions are found when  $t_{\text{star}}$  and  $t_{\text{gas}}$  differ by less than a factor of 4, or  $|\log_{10}(t_{\text{star}}/t_{\text{gas}})| \leq 0.6$ . This means that, when combining the range of reference time-scales covered by H $\alpha$  ( $t_{\text{star,ref}} \sim 5$  Myr, Haydon et al. 2018), FUV, and NUV ( $t_{\text{star,ref}} = 15\text{--}35$  Myr, Haydon et al. 2018) emission, the method can be applied at the highest accuracy for  $t_{\text{gas}} = 1\text{--}140$  Myr, which comfortably encompasses the (hotly debated) range of molecular cloud lifetimes considered in the literature (cf. Dobbs et al. 2014).

Moving to the duration of the overlap phase ( $t_{\text{over}}$ ) in the middle panels of Figure 21, we see that the point particle experiments invariably provide upper limits on  $t_{\text{over}}$ . This occurs in cases where the retrieved duration occupies less than 5 per cent of the entire timeline (i.e.  $t_{\text{over}}/\tau < 0.05$ ), because that percentage corresponds to the precision at which the method can constrain  $t_{\text{over}}$  at the adopted critical region filling factor  $\zeta$  (see Section 4.2.4 and Appendix B). Any value  $t_{\text{over}}/\tau < 0.05$  is therefore consistent with  $t_{\text{over}} = 0$ . Indeed, the values shown in the left middle panel of the figure are very low ( $t_{\text{over}} \sim 0.1$  Myr), which is expected given that single gas and star particles can never coexist in the simulations.

The above behaviour changes when considering the extended emission experiments in the middle right panel of Figure 21. In these cases, the retrieved durations of the overlap phase are longer (of the order  $t_{\text{over}} \sim 1$  Myr) and in reasonable agreement with the expected values from Section 4.3.3. This change relative to the point particle experiments occurs because the extended emission maps are generated by smoothing the particle population over  $N_{\text{ngb}} = 200$  neighbours, implying that we should start to see a physical time overlap between the gas and stellar phases. Interestingly, many of the symbols in this panel are transparent, particularly at  $t_{\text{star}} \geq 10$  Myr. For these experiments, we find that the retrieved region filling factors  $\zeta_{\text{star}}$  and  $\zeta_{\text{gas}}$  both exceed the critical value  $\zeta_{\text{crit}} = 0.5$  above which the spatial blending of regions affects the accuracy of the  $t_{\text{over}}$  measurement (see Section 4.2.4). When  $\zeta > \zeta_{\text{crit}}$ , it becomes difficult (and sometimes impossible) to statistically distinguish between the time overlap of the stellar and gas phases in a single region and the spatial overlap between physically unrelated regions. The magnitude of this degeneracy is determined by the level of region crowding in the maps, which in turn is set by the total duration of the timeline  $\tau$ . For this reason, experiments that are strongly affected by blending result in retrieved values of  $t_{\text{over}}$  that increase linearly with  $t_{\text{star}}$  and, given that these are found where  $t_{\text{star}} \gg t_{\text{gas}}$ , also with  $\tau$ . This shows the importance of applying our recommended upper limit of  $\zeta_{\text{crit}} = 0.5$ . If observational applications of the method yield higher values of  $\zeta$ , it should be kept in mind that the uncertainty on  $t_{\text{over}}$  will exceed  $0.05\tau$ , as shown in Appendix B.

Finally, the constrained values of the region separation length  $\lambda$  shown in the bottom panels of Figure 21 are in good agreement with the dependence on  $t_{\text{star}}$  expected from Section 4.3.3. Overall, the separation length should decrease with increasing  $t_{\text{star}}$  for geometric reasons. A longer duration of the evolutionary timeline

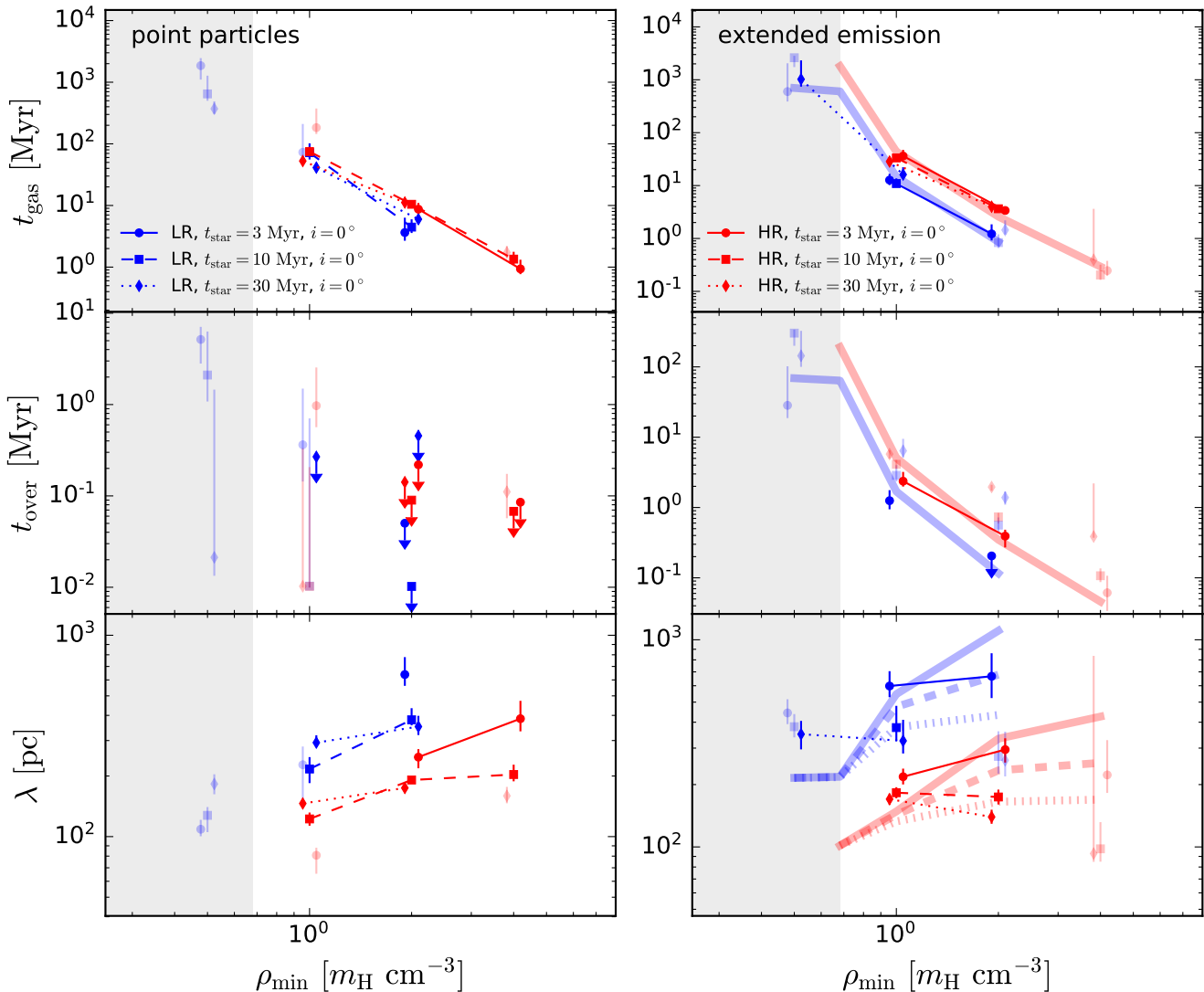
implies a larger number of regions, which in turn results in smaller separation lengths. While this behaviour is mostly followed by the experiments, there are two exceptions. The first occurs in the low-resolution point particle experiments, for which  $\lambda$  increases with  $t_{\text{star}}$  at  $t_{\text{star}} < 100$  Myr. This is caused by the fact that the young star particles in the low-resolution simulation are closely packed in spiral structures (see Figure 17) mirroring the gas structures from which they were born (see Figure 20). Over time, the older star particles disperse and occupy the large voids outside the dense birth sites, increasing the distance to their neighbours. This results in an increasing trend of  $\lambda$  with  $t_{\text{star}}$  that is only reversed when the voids get crowded and  $\lambda$  decreases for geometric reasons, as expected. The high-resolution and extended emission experiments do not follow this same trend reversal, because either the star formation is much more distributed and the maps are lacking major voids (in the high-resolution simulation), or the closely-packed young star particles are merged to constitute single regions by the use of a smoothing kernel when generating the extended emission maps.

The second case of  $\lambda$  (gently) increasing with  $t_{\text{star}}$  occurs in the extended emission experiments at large values of  $t_{\text{star}}$ . This reflects a crowding of the regions. By increasing  $t_{\text{star}}$ , the space between the regions has vanished and the overlapping envelopes of the regions give rise to a diffuse emission reservoir, within which single regions become increasingly hard to distinguish. As a result, the retrieved value of  $\lambda$  increases and the accuracy of all constrained quantities is negatively affected. In Section 4.3.9, we provide a quantitative criterion to identify applications of the method that are affected by diffuse emission reservoirs. Illustrative examples in Figure 21 are the two low-resolution, extended emission experiments at  $t_{\text{star}} > 100$  Myr, which are therefore shown as transparent symbols.

In summary, the experiments summarised in Figure 21 show that the method can be reliably applied across a broad range of reference time-scales. Primarily,  $t_{\text{star}}$  and  $t_{\text{gas}}$  should not differ by more than an order of magnitude due to the decreasing sensitivity of the method to the underlying timeline towards larger differences. It can be verified after applying the method whether this requirement is satisfied. For typical SFR tracers such as H $\alpha$  ( $t_{\text{star,ref}} \sim 5$  Myr, Haydon et al. 2018), FUV, and NUV ( $t_{\text{star,ref}} = 15\text{--}35$  Myr, Haydon et al. 2018) emission, this requirement should not be problematic, as together the tracers cover a wide range of gas phase durations that would be deemed reliable, i.e.  $t_{\text{gas}} = 0.5\text{--}50$  Myr and  $t_{\text{gas}} = 1.5\text{--}350$  Myr, respectively. Next to this immediate requirement of the method that (somewhat) limits the range of  $t_{\text{star}}$ , changing the duration of  $t_{\text{star}}$  can have secondary implications that negatively affect the results, mostly due to region crowding and blending. These effects are captured not in intrinsic limits on  $t_{\text{star}}$ , but on other quantities, such as the region filling factor  $\zeta$  (see Section 4.2.4) and the shape of the tuning fork diagram  $\mathcal{B}$  (see Section 4.3.9). All requirements for reliable applications of the method are summarised in Section 4.4 below.

#### 4.3.5 Dependence on the gas density threshold

The next test we carry out uses  $4 \times 9 = 36$  experiments from Table 8, with IDs 3, 5, 7, and 12–17, to determine how the choice of gas tracer (represented by the minimum gas volume density used to construct the gas map) affects the constrained quantities  $t_{\text{gas}}$ ,  $t_{\text{over}}$ , and  $\lambda$ . The choice of gas tracer or its excitation density is expected to strongly influence the retrieved duration of the gas phase, because higher-density gas represents a more advanced stage of collapse towards star formation. More specifically, high-density gas



**Figure 22.** Influence of the choice of gas tracer (captured by varying the minimum gas volume density  $\rho_{\text{min}}$ ) on the best-fitting values of the duration of the gas phase ( $t_{\text{gas}}$ , top), the duration of the overlap phase ( $t_{\text{over}}$ , middle), and the mean separation length between independent regions ( $\lambda$ , bottom), for experiments ID 3, 5, 7, and 12–17 in Table 8 using point particle maps (left panels) and extended emission maps (right panels). Transparent symbols represent experiments that do not satisfy each of the conditions for the reliable application of the method summarised in Section 4.4. Non-transparent symbols are connected by lines as shown by the legend. The grey-shaded area indicates the density range where gas particles are ineligible for star formation (see Section 4.1). In the right-hand panels, the thick blue (low resolution) and red (high resolution) lines show the expected values from Section 4.3.3, with line styles chosen to match the legend. In the middle panel, downward arrows indicate upper limits (see Section 4.4). This figure illustrates how a combination of gas tracers covering a variety of critical densities can be used to probe the evolutionary time-scales at different stages of the cloud lifecycle.

is short-lived in any star formation theory in which star formation proceeds on the gas free-fall time (e.g. Elmegreen 2002; Krumholz & McKee 2005; Padoan & Nordlund 2011; Hennebelle & Chabrier 2011; Federrath & Klessen 2012). In the simulations considered here, star formation does indeed proceed on a free-fall time, which will lead to a decrease of  $t_{\text{gas}}$  towards higher values of  $\rho_{\text{min}}$ . Furthermore, we showed in Section 4.3.3 that the use of minimum gas volume densities above the star formation threshold enables star particles to form from gas that could not have been visible in the gas map. The absence of visible progenitors to young stellar emission peaks leads to an additional (and even stronger) decrease of  $t_{\text{gas}}$  and  $t_{\text{over}}$  with  $\rho_{\text{min}}$ , through the factor  $p_{\text{gas}}$  in equation (167) and (169), which is propagated into the expected region separation length  $\lambda$  through equations (170)–(175). This strong variation of

the constrained quantities as a function of the minimum gas density is physical. By varying the gas tracer such that it covers a range of different excitation densities, the method can thus be used to assign lifetimes to gas at these densities and, by inversion, to probe the gas density evolution of the gas towards star formation as a function of absolute time. In order to test this behaviour, we consider an order of magnitude range in  $\rho_{\text{min}}$ , by using  $\rho_{\text{min,LR}} = 0.5\text{--}2 m_{\text{H}} \text{ cm}^{-3}$  and  $\rho_{\text{min,HR}} = 1\text{--}4 m_{\text{H}} \text{ cm}^{-3}$ .

In Figure 22, we test if the dependence of the constrained quantities  $t_{\text{gas}}$ ,  $t_{\text{over}}$ , and  $\lambda$  on the minimum gas density of the gas map follows the relations predicted in Section 4.3.3. As before in Section 4.3.4, we are only including the expected values in the right-hand panels showing the extended emission experiments. Across the three panels, we find remarkably good agreement be-

tween the values retrieved using the presented methods and the expected values. To first order, we see that the method performs well in tracing the gas density evolution as a function of absolute time. This is encouraging in the light of the wide variety densities probed by gas tracers that are readily detectable with ALMA in nearby galaxies (e.g. Shirley 2015; Leroy et al. 2017).

Specifically, the top panels of Figure 22 show that the expected, steep dependence of  $t_{\text{gas}}$  on  $\rho_{\text{min}}$  is quantitatively reproduced by the method. There is remarkably little difference between the point particle and extended emission experiments, which highlights that at the high densities probed here, the individual gas particles within a single region are so crowded (see Figure 20) that they are not separated, but together constitute a region in the point particle maps. The small vertical offset between the low-resolution and high-resolution measurements is attributed to a numerical resolution effect in Section 4.3.3.

The middle panels of Figure 22 exhibit contrasting trends. In the point particle experiments (middle left panel), the upper limits on the data points show that the duration of the overlap phase  $t_{\text{over}}$  is very short ( $t_{\text{over}} < 1$  Myr), which may be expected given that a single star particle never coexists with its parent gas particle. However, the point-like nature of the emission peaks in these maps only fundamentally applies to the stellar maps (Figure 17), because the gas particles are so strongly spatially correlated that it is hard to discriminate individual gas particles in the gas maps (Figure 20). It turns out that the point-like nature of the stellar maps is sufficient to maintain a short duration of the overlap phase.

The measurements of  $t_{\text{over}}$  obtained in the extended emission experiments (middle right panel) behave quite differently to the point particle experiments, yet follow a relation with  $\rho_{\text{min}}$  that is very similar to the gas phase lifetime in the top panels. As explained in the derivation of Section 4.3.3, this is not surprising. Because the regions consist of large numbers of neighbours with long star formation time-scales, the expected duration of the overlap phase is long ( $\sim 100$  Myr) at the density threshold for star formation  $\rho_{\text{crit}}$ . At higher densities, the gaseous progenitors to young stellar regions may be lost to star formation before they reach sufficiently high densities to be visible in the gas maps. This decreases the inferred duration of the entire timeline and thus affects  $t_{\text{gas}}$  and  $t_{\text{over}}$  to an equal extent.

We remind the reader that, even though the steep dependence of  $t_{\text{gas}}$  and  $t_{\text{over}}$  on  $\rho_{\text{min}}$  behaves as predicted in Section 4.3.3, this is strictly a numerical effect due to the stochastic star formation algorithm used in the simulations. In observational applications of the method, any dependence of the evolutionary timeline on the gas tracer density should be considerably shallower. In real-Universe systems, gaseous regions attain higher densities as they evolve towards star formation. As long as the observational sensitivity is sufficiently high, the high-density gaseous progenitor to young stellar populations should always be detectable, because stars represent the evolutionary end state of the evolution towards increasing densities. This presents a strong contrast to the simulations, in which stars may form out of low-density gas, provided that it has a density in excess of the star formation threshold ( $\rho_{\text{g}} \geq \rho_{\text{crit}}$ ).

Despite the fundamentally different physics at play, the steep dependence of the retrieved time-scales on  $\rho_{\text{min}}$  serves as both a warning and an exciting prospect for observational applications of the method. The warning is that the observations must be sufficiently sensitive to have been able to detect the gaseous progenitors to the young stellar emission peaks considered, as well as to have been able to detect the young stellar populations that may eventually emerge from the gas emission peaks hosted by the maps (also

see the discussion in KL14). The exciting prospect of the density dependences shown in the top and middle panels of Figure 22 is that by probing gas of various densities, it is not only possible to reconstruct the density evolution as a function of time towards star formation, but also the dissociation of the dense gas during the stellar feedback phase.

Finally, the bottom two panels in Figure 22 show the region separation lengths retrieved by the method, which (in the bottom right panel) again show good agreement with the expected values. The naive expectation shown by the thick lines is that these should increase with  $\rho_{\text{min}}$ . At higher minimum gas densities, the emission peaks are shorter lived, implying that the gas maps should host fewer of them, in turn leading to larger separation lengths. This behaviour is indeed followed by the point particle experiments, but for the extended emission maps this expectation only applies to the experiments with short stellar reference time-scales ( $t_{\text{star,ref}}$ ). For  $t_{\text{star,ref}} \geq 10$  Myr, the bottom right panel shows that  $\lambda$  decreases with  $\rho_{\text{min}}$ . We find that this trend occurs due to the fact that higher-density gas more closely traces the spiral structure in the simulations than the low-density gas. As a result, the high-density gas is more spatially correlated, which may lead to smaller region separation lengths if the stellar regions follow the spiral structure too. Indeed they do, but the clustering of the stellar regions near spiral arms only becomes statistically noticeable if they are sufficiently numerous to sample the spiral structure well, which happens for  $t_{\text{star,ref}} \geq 10$  Myr.

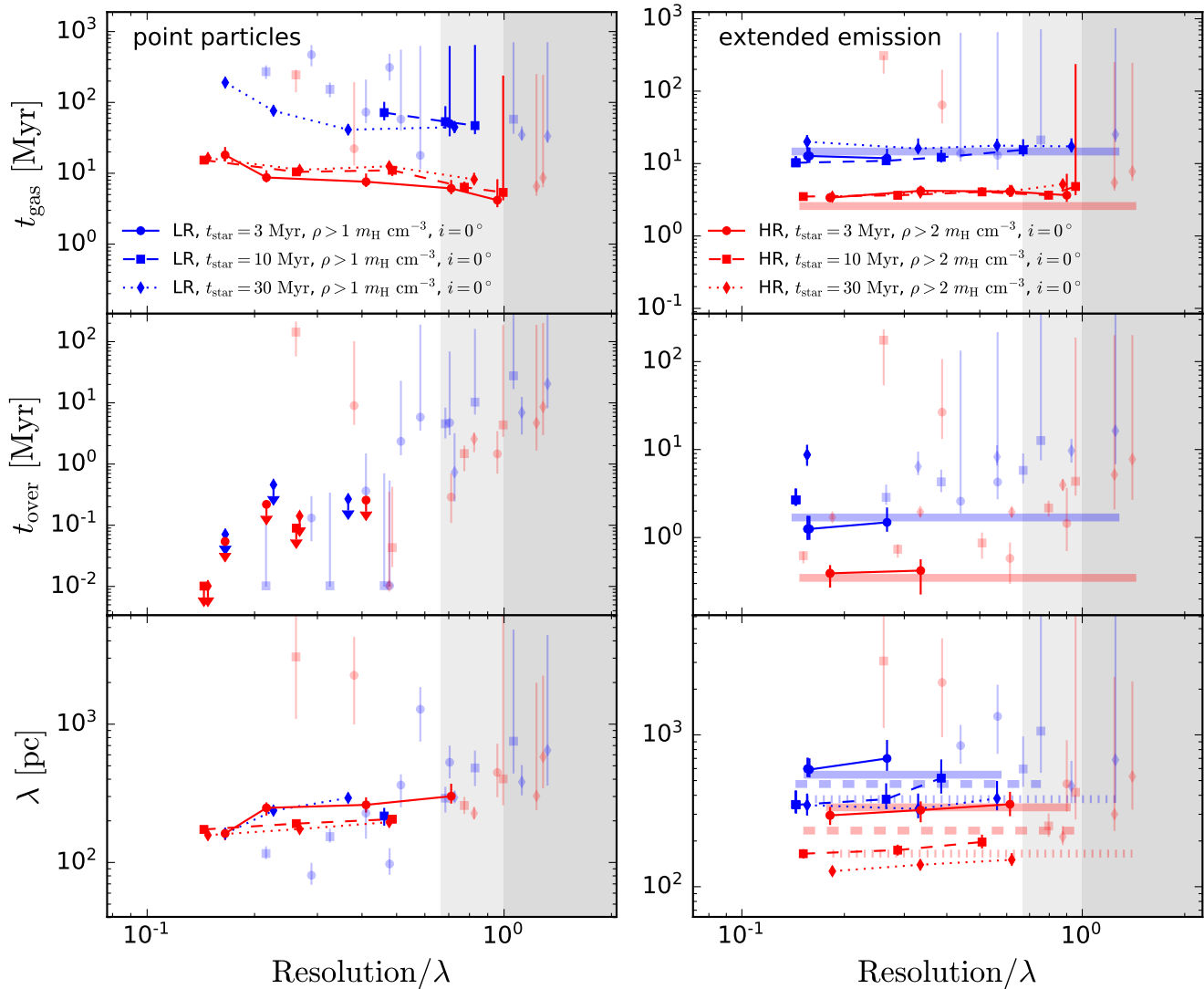
In summary, the experiments summarised in Figure 22 show that by varying the tracer (and specifically such that different characteristic excitation densities are probed), the physical dependence of the constrained quantities on the gas density can be probed. The presented examples cover an order of magnitude in gas density, which is considerably smaller than the  $\sim 3$  orders of magnitude in density that are plausibly spanned by observations of nearby galaxies with ALMA (e.g. Leroy et al. 2017). Therefore, this test illustrates how the application of the method to well-chosen sets of gas tracers can provide even more insight in the cloud-scale star formation process.

#### 4.3.6 Dependence on the spatial resolution or distance

We now use  $4 \times 18 = 72$  experiments from Table 8, with IDs 3, 5, 7, and 18–32, to test how the spatial resolution of the maps (in units of the mean separation length) affects the constrained quantities  $t_{\text{gas}}$ ,  $t_{\text{over}}$ , and  $\lambda$ . This is a critical test for future applications of the method, because any limitations placed on the spatial resolution will directly set the maximum distance up to which the method can be applied (also see Section 6 below). With high-sensitivity interferometers like ALMA being operational and 30-metre class optical telescopes with adaptive optics coming online in the near future, the application of the method out to large distances using high-resolution imaging presents a logical research avenue. To define a point of reference for these applications, we convolve the simulated maps to a wide range of spatial resolutions that are defined by the minimum aperture size, covering  $l_{\text{ap,min}} = 25\text{--}800$  pc or 1.5 dex in PSF size.

In Figure 23, we test the accuracy of the method across the above range of spatial resolutions, which for the purpose of the comparison here are normalised to the region separation length  $\lambda$ . This is a necessary normalisation to enable a quantitative comparison between all experiments and reflects the fact that the characteristic size scale of the gas-to-stellar flux ratio bias modelled in Section 3.2.11 is set entirely by the region separation length (see





**Figure 23.** Influence of the physical spatial resolution (FWHM or  $l_{\text{ap,min}}$ ) in units of the mean separation length on the best-fitting values of the duration of the gas phase ( $t_{\text{gas}}$ , top), the duration of the overlap phase ( $t_{\text{over}}$ , middle), and the mean separation length between independent regions ( $\lambda$ , bottom), for experiments ID 3, 5, 7, and 18–32 in Table 8 using point particle maps (left panels) and extended emission maps (right panels). Transparent symbols represent experiments that do not satisfy each of the conditions for the reliable application of the method summarised in Section 4.4. Non-transparent symbols are connected by lines as shown by the legend. The light grey shaded region indicates resolutions worse than  $0.67\lambda$  (or  $\zeta > 0.56$  for unresolved point sources), above which the measurements of  $t_{\text{over}}$  and  $\lambda$  start to show systematic deviations. The dark grey shaded regions indicate resolutions worse than  $\lambda$  (or  $\zeta > 0.84$  for unresolved point sources), above which  $t_{\text{gas}}$  shows minor deviations. In the right-hand panels, the thick blue (low resolution) and red (high resolution) lines show the expected values from Section 4.3.3, with line styles chosen to match the legend. In the middle panel, downward arrows indicate upper limits (see Section 4.4). This figure shows that (especially) the non-transparent retrieved quantities depend only weakly on the spatial resolution, implying that the method can be reliably applied as long as the region separation length  $\lambda$  is resolved by at least one (for  $t_{\text{gas}}$ ) or 1.5 (for  $t_{\text{over}}$  and  $\lambda$ ) resolution elements.

the top right panel of Figure 11). Again, we display the expected values from Section 4.3.3 in the right-hand panels showing the results from the extended emission experiments. Across the three panels, we find good agreement between the values retrieved using the presented methods and the expected values, as well as little dependence of the constrained quantities on the spatial resolution, provided that  $l_{\text{ap,min}}/\lambda \ll 1$ . Around  $l_{\text{ap,min}}/\lambda \sim 1$ , some of the measurements start to show deviations from both the higher-resolution measurements and the expected values. We see that all of the experiments have  $l_{\text{ap,min}}/\lambda < 1.5$ , despite the fact that the largest PSF size used ( $l_{\text{ap,min}} = 800$  pc) is an order of magnitude larger than the smallest region separation length ( $\lambda \approx 80$  pc). This shows that convolving the maps with a PSF much larger than

the region separation length will simply result in a larger retrieved value of  $\lambda$ , reflecting the unsurprising result that is not possible to retrieve sub-resolution separations. Therefore, the first-order result of Figure 23 is that the constrained quantities  $t_{\text{gas}}$ ,  $t_{\text{over}}$ , and  $\lambda$  do not strongly depend on the spatial resolution, as long as the region separation length is well-resolved. Below, we quantify this statement further for each of the three quantities.

Another general observation that can be made for each of  $t_{\text{gas}}$ ,  $t_{\text{over}}$ , and  $\lambda$ , is that the results of the point particle experiments approach those of the extended emission experiments towards large  $l_{\text{ap,min}}/\lambda$ . Again, this is not necessarily surprising – the lifecycle and separation of individual particles differs from those of concentrations of particles within a smoothing kernel, but the convolution

with a large PSF to worsen the resolution of the maps obscures that difference. Indeed, for  $l_{\text{ap},\text{min}} \gg h_{\text{smooth}}$ , the point particle maps are visually nearly indistinguishable from the extended emission maps. For the {low, high} resolution simulations, we find that the differences between the point particle and extended emission experiments start to vanish at aperture sizes of  $l_{\text{ap},\text{min}} \geq \{400, 200\}$  pc. Given that typical values of  $\lambda$  are of the same order at high spatial resolutions (see Appendix D), we expect the results of the point particle and extended emission experiments to yield similar results for  $l_{\text{ap},\text{min}}/\lambda > 0.5$ . Figure 23 shows that this is indeed the case.

If we now first focus on the top panels in Figure 23, it is clear that the retrieved durations of the gas phase closely match the expected value, independently of the spatial resolution. The point particle maps return longer lifetimes than the extended emission maps, which is to be expected given that individual particles should be longer-lived than collections of particles if these regions are destroyed by stellar feedback. The lack of correlation between  $t_{\text{gas}}$  and  $l_{\text{ap},\text{min}}/\lambda$  is encouraging for future application of the method. Only at  $l_{\text{ap},\text{min}}/\lambda \gtrsim 1$  (highlighted by the dark grey area), some of the data points show large upward error bars or overestimate  $t_{\text{gas}}$  relative to the expected value. The large upward error bars are easy to understand in the context of the tuning fork diagram of Figure 2 – at low spatial resolution, the separation of the two branches is not well-resolved. If one of the branches does not deviate strongly from the galactic average (i.e. it is consistent with  $\mathcal{B} = 1$  at  $1-2\sigma$ ), then the retrieved duration of the gas phase may be arbitrarily long-lived (up to the galaxy-wide gas depletion time). Likewise, the measurement of longer lifetimes at large  $l_{\text{ap},\text{min}}/\lambda$  can be understood in terms of a smaller deviation from the galactic average gas-to-stellar flux ratio at larger aperture sizes, which results in more extended evolutionary timelines. This is an undesirable outcome and shows that measurements of  $t_{\text{gas}}$  obtained with our method should only be used if  $l_{\text{ap},\text{min}} < \lambda$ .

In the middle panels of Figure 23, the retrieved durations of the overlap phase ( $t_{\text{over}}$ ) behave roughly as expected. In the point particle experiments, we obtain very low ( $t_{\text{over}} < 1$  Myr) upper limits, which approach the higher values found for the extended emission experiments as  $l_{\text{ap},\text{min}}/\lambda$  approaches unity. The results of the extended emission experiments are consistent with the expectations from Section 4.3.3. Unfortunately, most of the experiments do not pass the requirements for reliably constraining  $t_{\text{over}}$  listed in Section 4.4 below, which inhibits drawing firm conclusions regarding any trends of  $t_{\text{over}}$  with the spatial resolution of the maps. The transparent symbols mostly do not satisfy these requirements because the retrieved region filling factors ( $\zeta \equiv 2r/\lambda$ ) exceed the critical value ( $\zeta_{\text{crit}} = 0.5$ ), implying that spatial overlap between regions obstructs constraining their temporal overlap. However, if we focus on the trustworthy experiments (shown as the opaque data points) that reach the largest  $l_{\text{ap},\text{min}}/\lambda$ , i.e. the experiments with  $t_{\text{star}} = 3$  Myr (filled circles on solid lines), then the middle right panel of Figure 23 does show that following set of unreliable (transparent) experiments remain close to the results obtained at high spatial resolution until  $\lambda$  is resolved by less than 1.5 resolution elements (highlighted by the light grey area). Therefore, we recommend to disregard measurements of  $t_{\text{over}}$  resulting from applications for which  $l_{\text{ap},\text{min}}/\lambda > 0.67$ .

Finally, the bottom panels of Figure 23 show that the region separation length  $\lambda$  is well-constrained until about  $l_{\text{ap},\text{min}}/\lambda \sim 0.67$ , similarly to  $t_{\text{over}}$ . As discussed above, this is not surprising. The retrieved value of  $\lambda$  cannot be much smaller than a resolution element, implying that applications of the method to low-resolution maps must overestimate the region separation length. In the ex-

treme case of PSF sizes much larger than the sub-resolution region separation length, we expect that the measured separation length satisfies  $\lambda \propto l_{\text{ap},\text{min}}$ . Indeed, this behaviour can be seen in the bottom right panel of Figure 23 near  $l_{\text{ap},\text{min}}/\lambda = 1$ . This shows that region separation lengths measured with our method should only be used if they are sufficiently well-resolved, which we quantify as  $l_{\text{ap},\text{min}}/\lambda < 0.67$  (highlighted by the light grey area).

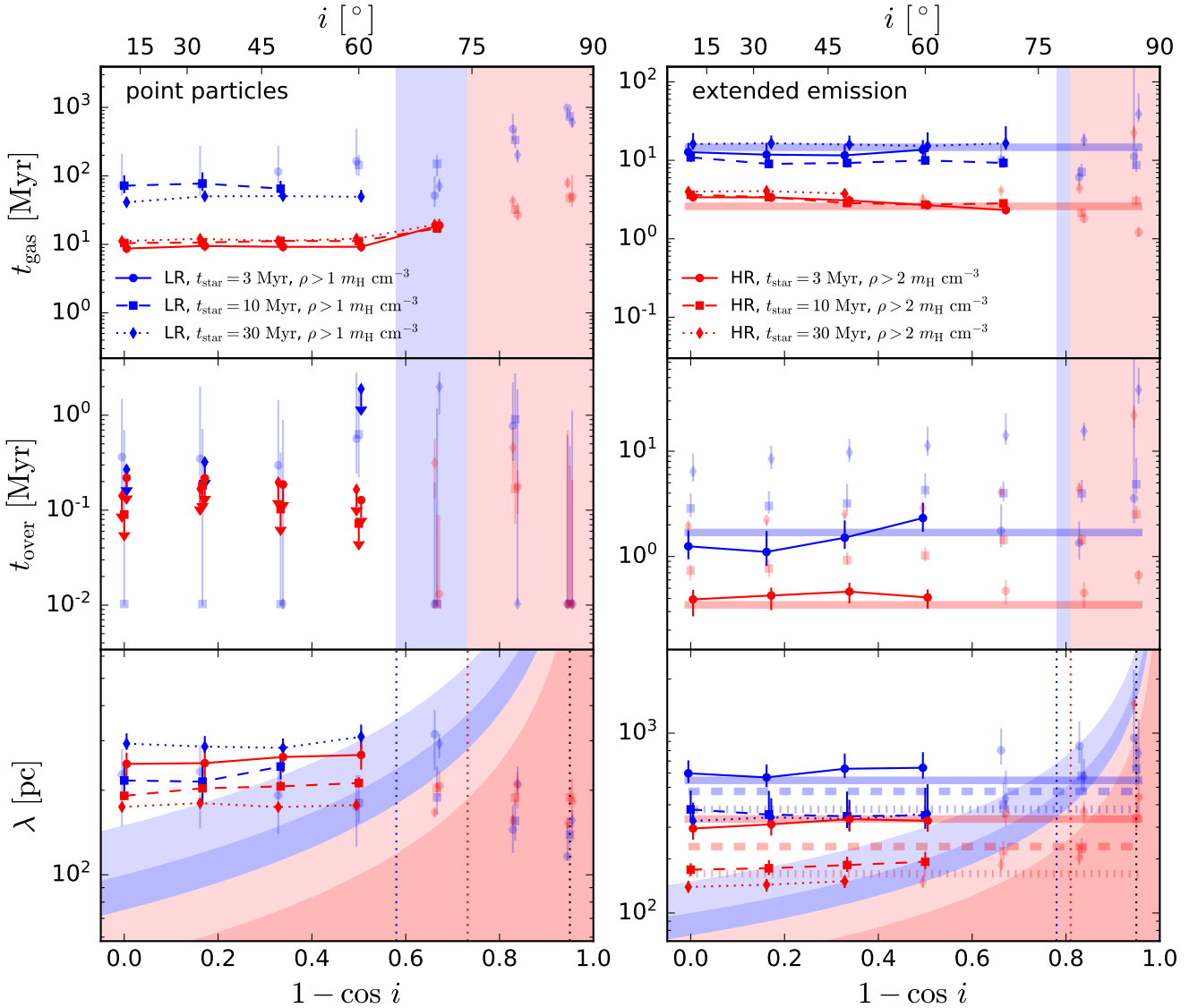
In summary, the spatial resolution of the map is a key quantity in assessing the accuracy of measurements obtained with the method. We require  $l_{\text{ap},\text{min}}/\lambda < 1$  when measuring  $t_{\text{gas}}$  and  $l_{\text{ap},\text{min}}/\lambda < 0.67$  when measuring  $t_{\text{over}}$  or  $\lambda$ . Fortunately, these are rather weak requirements. The typical separation lengths obtained in the presented experiments (see Appendix D) and in the first observational applications that we are currently carrying out (Kruijssen et al. 2018; Hygate et al. in prep.; Schruha et al. in prep.; Chevance et al. in prep.; Ward et al. in prep.) are all comfortably resolved out to tens of Mpc with modern observatories (see Section 6). This implies that our method enables the systematic characterisation of the evolutionary timeline of cloud-scale star formation across the galaxy population.

#### 4.3.7 Dependence on the galaxy inclination

We now use  $4 \times 21 = 84$  experiments from Table 8, with IDs 3, 5, 7, and 33–50, to test how the galaxy inclination affects the constrained quantities  $t_{\text{gas}}$ ,  $t_{\text{over}}$ , and  $\lambda$ . While the inclination angles of galaxies are weakly correlated with the orientation of the cosmic web by which they have been fed and torqued (e.g. White 1984; Dekel et al. 2009a; Forero-Romero et al. 2014), their distribution of inclination angles can be assumed to be flat for the purposes in this paper. As a result, by determining the inclination angle above which the method no longer returns accurate results, we can effectively determine which fraction of the galaxy population should be avoided in observational applications. Because galaxies are found with all inclination angles, our experiments are chosen to cover a (nearly) complete range of  $i = 0.0-87.1^\circ$ . At the maximum inclination angle, the projected diameter and thickness of the simulated galaxy discs are equal.

In Figure 24, we test the accuracy of the method across the above range of galaxy inclination angles, which for the purpose of the comparison here are expressed as  $1 - \cos i$ . This is a useful representation, because the distances in the disc plane are scaled by a factor of  $\cos i$  in projection. The term  $\cos i$  is subtracted from unity to have the inclination increase towards the right (see the top axis). As before, we display the expected values from Section 4.3.3 in the right-hand panels showing the results from the extended emission experiments. Across the three panels, we find good agreement between the values retrieved using the presented methods and the expected values from Section 4.3.3, as well as little dependence of the constrained quantities on the inclination.

This figure exhibits great similarity to Figure 23, in that the retrieved quantities are accurate until the galaxy is inclined so far that the region separation length is no longer resolved. In the figure, this regime is shown by the blue (low-resolution) and red (high-resolution) shaded areas – in the bottom panels, the light shades apply to the requirement for  $t_{\text{over}}$  and  $\lambda$  to resolve the region separation length with 1.5 resolution elements, whereas the dark shades reflect the requirement for  $t_{\text{gas}}$  that  $\lambda$  is resolved with a single resolution element (see Section 4.3.6). The critical inclination angles above which the projected region separation lengths are no longer resolved are  $i = 65-75^\circ$  for the point particle experiments and  $i = 75-80^\circ$  for the extended emission experiments.



**Figure 24.** Influence of the galaxy inclination angle  $i$  on the best-fitting values of the duration of the gas phase ( $t_{\text{gas}}$ , top), the duration of the overlap phase ( $t_{\text{over}}$ , middle), and the mean separation length between independent regions ( $\lambda$ , bottom), for experiments ID 3, 5, 7, and 33–50 in Table 8 using point particle maps (left panels) and extended emission maps (right panels). Transparent symbols represent experiments that do not satisfy each of the conditions for the reliable application of the method summarised in Section 4.4. Non-transparent symbols are connected by lines as shown by the legend. In the right-hand panels, the thick blue (low resolution) and red (high resolution) lines show the expected values from Section 4.3.3, with line styles chosen to match the legend. In the bottom panel, the light blue (low-resolution) and red (high-resolution) shaded regions indicate resolutions worse than  $0.67\lambda$  (or  $\zeta > 0.56$ ), above which the measurements of  $t_{\text{over}}$  and  $\lambda$  start to show systematic deviations. Likewise, the dark shaded regions indicate resolutions worse than  $\lambda$  (or  $\zeta > 0.84$ ), above which  $t_{\text{gas}}$  shows minor deviations. The coloured dotted vertical lines indicate the inclination angles above which the retrieved region separation length is no longer well-resolved, whereas the black dash-dotted vertical line marks the inclination angle at which the galaxy diameter and vertical thickness have the same projected dimension. In the top and middle panel, the shaded regions indicate the inclination angles for which the retrieved region separation lengths are ill-resolved. In the middle panel, downward arrows indicate upper limits. This figure shows that all of the retrieved quantities show only a weak dependence on the inclination angle, enabling reliable applications of the method up to  $i \sim 75^\circ$ .

In the top panels of Figure 24, we see that the duration of the gas phase does not depend on the inclination angle until  $i > \cos^{-1}(l_{\text{ap,min}}/\lambda)$  (indicated by the blue and red shaded areas). Above these inclinations, the scatter of the measurements increases and the retrieved values of  $t_{\text{gas}}$  are slightly elevated relative to low inclinations. These effects are relatively minor, implying that measurements of the gas phase lifetime are robust against variations of the inclination angle. If real-Universe systems have similar region separation lengths as the experiments used here ( $\lambda = 100\text{--}400$  pc)

and are observed at a similar spatial resolution ( $l_{\text{ap,min}} = 50$  pc), then  $t_{\text{gas}}$  can be reliably constrained up to  $i = 75^\circ$ . More generally, we require  $\lambda \geq l_{\text{ap,min}}/\cos i$  when measuring  $t_{\text{gas}}$ , which generalises the result from Section 4.3.6 to arbitrary inclinations.

The middle panels of Figure 24 show similar results as before in Section 4.3.4–4.3.6. The point particle experiments return durations of the overlap phase that are short, as indicated by upper limits  $t_{\text{over}} \lesssim 1$  Myr. Similarly to Figure 23 showing the dependence of the constrained quantities on the spatial resolution, the opaque

symbols (indicative of satisfying all requirements for the reliable application of the method listed in Section 4.4) do not extend to high inclination angles in the measurements of  $t_{\text{over}}$ . Perhaps unsurprisingly, this occurs for the same reason as in Figure 23. Already at intermediate inclinations ( $i \sim 60^\circ$ ), the region filling factors exceed the critical value above which region blending prohibits placing accurate constraints on the duration of the overlap phase (i.e.  $2r/\lambda > 0.5$ ). For the measurements that can be considered to be accurate, the middle right panel shows that the retrieved values of  $t_{\text{over}}$  reproduce the expectation from Section 4.3.3. At large inclinations, we observe a slight increase of the retrieved  $t_{\text{over}}$ , which mirrors the trend from Figure 23, as well as the behaviour of  $t_{\text{gas}}$  in the top panels of Figure 24. Therefore, we maintain the requirement from Section 4.3.6 that the region separation length should be resolved by at least 1.5 resolution elements to place accurate constraints on  $t_{\text{over}}$  and generalise it to  $\lambda \geq 1.5l_{\text{ap,min}}/\cos i$ . For  $\lambda = 200$  pc and  $l_{\text{ap,min}} = 50$  pc, this corresponds to  $i \leq 68^\circ$ .

Finally, the bottom panels of Figure 24 confirm the above results by demonstrating that the region separation length is insensitive to the inclination angle, provided that  $\lambda$  is resolved by at least 1.5 resolution elements in projection (indicated with the light shaded blue and red areas). When this requirement is satisfied, the retrieved values of  $\lambda$  are consistent with the expected values from Section 4.3.3. In line with Section 4.3.6, we thus again require  $\lambda \geq 1.5l_{\text{ap,min}}/\cos i$ , or  $i \leq 68^\circ$  for  $\lambda = 200$  pc and  $l_{\text{ap,min}} = 50$  pc. We note that this is a very conservative limit, because the bottom panels in Figure 24 both do not show significant deviations from the low-inclination measurements for  $i < 75^\circ$ .

In summary, the galaxy inclination angle only weakly affects the applicability of the method. For the dimensions and properties of the simulated galaxies considered here, we require  $i \leq 75^\circ$  (or generally  $\lambda \geq l_{\text{ap,min}}/\cos i$ ) when measuring  $t_{\text{gas}}$  and  $i \leq 68^\circ$  (or generally  $\lambda \geq 1.5l_{\text{ap,min}}/\cos i$ ) when measuring  $t_{\text{over}}$  or  $\lambda$ . Assuming a flat distribution of inclination angles, this implies that future applications of our method are not noticeably affected by the galaxy inclination for  $\sim 80$  per cent of the galaxy population. In conjunction with the resolution requirements from Section 4.3.6, this means that the method is readily applicable to a statistically significant galaxy sample. We quantify this statement in Section 6.

#### 4.3.8 Dependence on the number of independent regions

Finally, we use all  $4 \times 50 = 200$  experiments from Table 8 to quantify how the number of identified emission peaks (related to the size of the maps in units of the region separation length) affects the error bars on the constrained quantities  $t_{\text{gas}}$ ,  $t_{\text{over}}$ , and  $\lambda$ . At low numbers of emission peaks, the uncertainties are expected to increase due to Poisson noise. If we make the assumption that the peak identification algorithm roughly selects independent regions (see Section 3.2.7), then this test places a lower limit on the galaxy size for a given value of the region separation length. Specifically, if we define a circular, face-on disc galaxy with outer radius  $R_{\text{max}}$  that hosts a population of randomly distributed independent regions and require at least a number of emission peaks  $N_{\text{min}}$ , this places a lower limit on the galaxy radius of

$$R_{\text{max}} \geq \frac{\lambda}{2} \frac{\sqrt{N_{\text{min}}}}{f_{\text{struc}}}, \quad (176)$$

where the factor  $0 < f_{\text{struc}} \leq 1$  accounts for deviations from a random distribution ( $f_{\text{struc}} = 1$ ), as in equation (173) of Section 4.3.3.

Of course, the above assumptions only enable an order-of-magnitude estimate of  $R_{\text{max}}$ . We have already seen in the pre-

ceding discussion that independent regions are not randomly distributed – for the simulations considered here, we find  $f_{\text{struc}} \approx 0.4$  (see Section 4.3.3). In addition, one of the reasons for using the presented method rather than just counting emission peaks to derive their relative time-scales is that it allows us to be agnostic about what constitutes an independent region. After all, it is unknown a priori under which conditions the structures selected by peak identification algorithms correspond to independent regions. In spite of these caveats, equation (176) serves as a rough indication of how identifying a minimum number of emission peaks may help select suitable target galaxies given an expected value of  $\lambda$ . More generally, it is of immediate empirical relevance to define a value of  $N_{\text{min}}$  above which the error bars on the constrained quantities are acceptably small, because a visual check of a possible target galaxy can already give an indication of the rough number emission peaks that may plausibly be identified. Across all 200 experiments in Table 8, we define  $N_{\text{min}}$  for each experiment as

$$N_{\text{min}} = \min(N_{\text{peak,star}}, N_{\text{peak,gas}}), \quad (177)$$

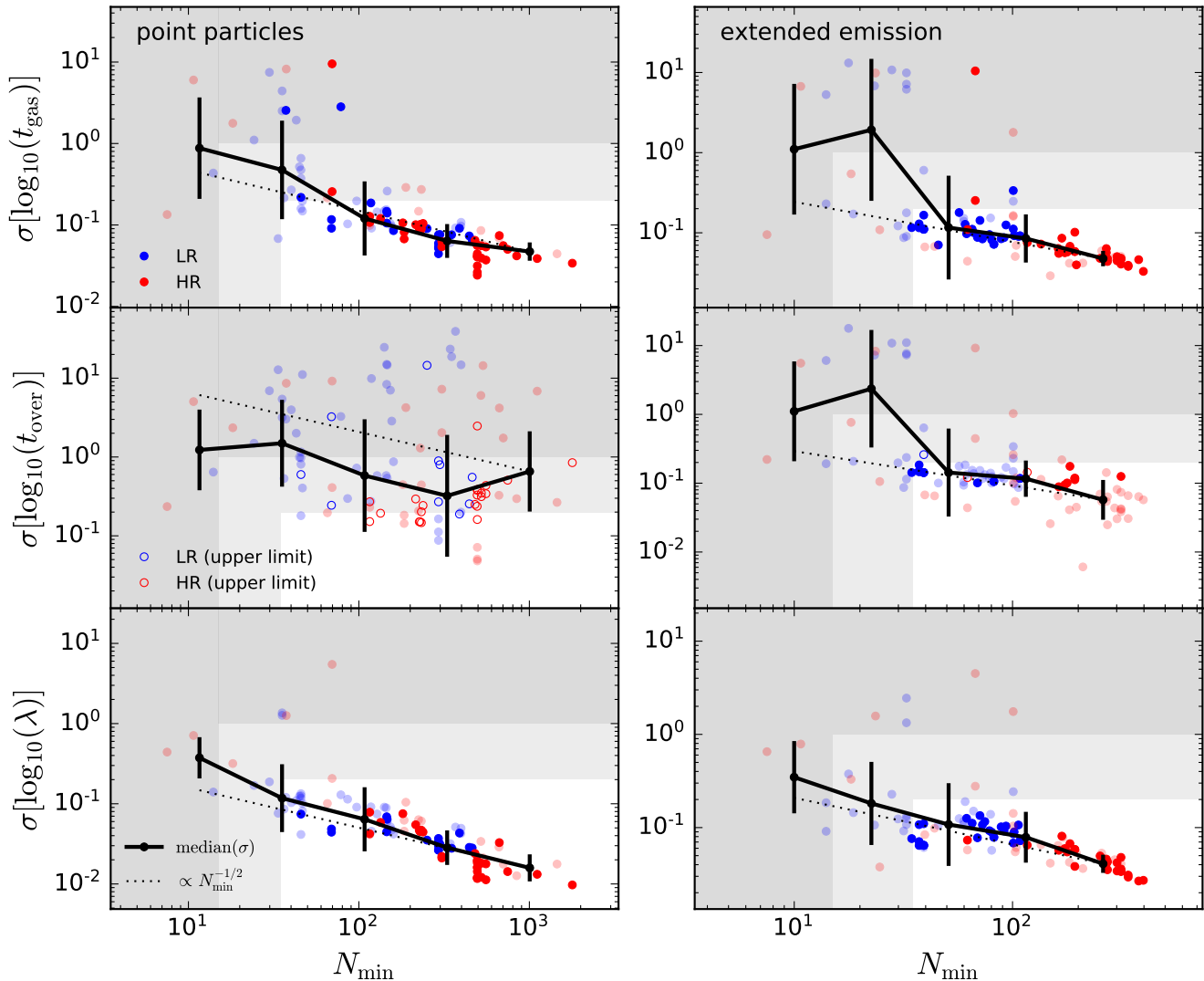
so that it reflects the smallest number of peaks identified between both maps. By using all experiments, we cover the largest possible range in  $N_{\text{min}}$ , spanning nearly 2.5 orders of magnitude from  $N_{\text{min}} = 7$ –1700.

In Figure 25, we show the logarithmic (i.e. relative) error bars on the retrieved quantities, taking the logarithmic average of the upward and downward error for each experiment, as a function of  $N_{\text{min}}$ . Even if the figure can be used to determine the method’s precision continuously as a function of  $N_{\text{min}}$ , we consider two characteristic uncertainties of 50 per cent (i.e. 0.2 dex, corresponding to the boundary between the white and light grey areas) and one order of magnitude (i.e. 1 dex, corresponding to the boundary between the dark and light grey areas) to determine a minimum number of emission peaks. The first of these defines the number of peaks above which the evolutionary timeline of cloud-scale star formation and feedback is characterised at high precision, whereas the second limit represents a firm minimum – while order-of-magnitude precision may be sufficient at high redshift to place first constraints on clump lifetimes, the diagnostic power of the method vanishes at even lower precision.

Across all panels in Figure 25 except the middle left (which is discussed below), the error bars roughly scale as the inverse square-root of the number of emission peaks  $N_{\text{min}}$ , as expected for Poisson-dominated errors. Another similarity between the panels is the normalisation of this trend – in nearly all panels, the median relation (thick black line) reaches the critical 0.2 dex uncertainty at just below 50 emission peaks and the critical 1 dex uncertainty at a little over 10 emission peaks. However, note that the median is calculated for all experiments. If we restrict ourselves to the experiments that satisfy all of the requirements listed in Section 4.4 (opaque symbols), we find that the number of emission peaks at which a 0.2 dex uncertainty is reached can be constrained quite well, at  $N_{\text{min}} \approx 35$ . This is not possible for the 1 dex uncertainty level, because none of the opaque data points have such large error bars. Using the transparent data points, we find that a 1 dex uncertainty is typically reached at  $N_{\text{min}} \approx 15$ .

Exceptions exist to both of these rules of thumb. Most notably, there are four opaque data points with error bars well in excess of 1 dex in the top panels. These experiments are easily identified as the data points with large error bars in the top panels of Figure 23 and have in common that they barely resolve the region separation length. While this boosts their upward error bars, both their absolute values and downward error bars are not negatively affected by





**Figure 25.** Influence of the smallest number of emission peaks in either of the two maps on the logarithmic uncertainties of the duration of the gas phase ( $t_{\text{gas}}$ , top), the duration of the overlap phase ( $t_{\text{over}}$ , middle), and the mean separation length between independent regions ( $\lambda$ , bottom), for all experiments in Table 8 using point particle maps (left panels) and extended emission maps (right panels). The uncertainties correspond to the logarithmic average of the upward and downward error for each experiment. Transparent symbols represent experiments that do not satisfy each of the conditions for the reliable application of the method summarised in Section 4.4. Non-transparent symbols represent the low- and high-resolution simulations as indicated by the legends, with open symbols in the middle panels denoting upper limits. The thick black lines show the median uncertainties across five bins that are equally spaced in the logarithm of the number of peaks  $N_{\text{min}}$ . For reference, the dotted line shows the expected trend for the standard deviation of the average, i.e.  $\propto N_{\text{min}}^{-1/2}$ . In the white region, the typical logarithmic uncertainties are  $< 0.2$  (i.e.  $< 50$  per cent), whereas the dark grey shaded area indicates where the typical uncertainty exceeds unity. The figure shows that meaningful quantitative measurements are obtained when each of the maps contains  $N_{\text{min}} > 35$  emission peaks, whereas order-of-magnitude estimates are possible if  $N_{\text{min}} > 15$ .

the poor resolution. The other main exception to the above minimum numbers of emission peaks is found in the middle left panel, showing the uncertainty on the duration of the overlap phase in the point particle experiments. Nearly all data points in this panel have large error bars, with a median error of about 1 dex. Again, this is not surprising. The retrieved values of  $t_{\text{over}}$  in the point particle experiments are all close to zero and often represent upper limits. Physically, such an outcome is understandable, because individual gas particles in the simulation are instantaneously converted into stellar particles, without any time overlap. However, if we compare the retrieved small values of  $t_{\text{over}} \sim 0.1$  Myr (see Figure 21–24) to its typical precision of  $0.05\tau \gtrsim 1$  Myr, it is instantly clear that the typical uncertainty on  $t_{\text{over}}$  must span an order of magnitude, even

if the absolute uncertainty is still small. Therefore, the elevated error bars in the middle left panel are no fundamental shortcoming of the method, but correspond to an acceptable uncertainty.

In summary, Figure 25 shows that the method constrains the quantities  $t_{\text{gas}}$ ,  $t_{\text{over}}$ , and  $\lambda$  to a precision of  $< 0.2$  dex (or 50 per cent) if both maps contain at least  $N_{\text{min}} \geq 35$  identified emission peaks. For order-of-magnitude estimates (1 dex precision), we require  $N_{\text{min}} > 15$ . Turning to equation (176), these numbers of regions can be translated to minimum galaxy radii. We assume a region separation length of  $\lambda = 200$  pc, which is reasonable in view of both the experiments carried out here and the first observational applications of the method (Krujissen et al. 2018; Hygate et al. in prep.; Schruha et al. in prep.; Chevance et al. in prep.; Ward et al. in

prep.), and adopt a typical correction factor for galactic structure of  $f_{\text{struc}} = 0.4$  to find minimum galaxy radii of  $R_{\text{max}} \geq 1.5$  kpc for 0.2 dex precision and  $R_{\text{max}} \geq 1$  kpc for order-of-magnitude estimates. If we characterise galaxies by their half-light radii, then the relation between stellar mass and half-light radius for local-Universe, late-type star-forming galaxies (Shen et al. 2003) can be used to convert the above limits on galaxy radii to a limit on the galaxy stellar mass of  $M_{\star} \gtrsim 10^9 M_{\odot}$ . Practically, the above fields of view can also comfortably be covered, even with high-resolution (e.g.  $l_{\text{ap,min}} = 50$  pc) observations on facilities like ALMA.

For a region separation length more appropriate for a high-redshift environment of  $\lambda = 1.5$  kpc (see Section 6) and a galactic structure correction factor appropriate for galaxies without dominant spiral structure ( $f_{\text{struc}} = 1$ ), we instead find minimum galaxy radii of  $R_{\text{max}} \geq 4.4$  kpc for 0.2 dex precision and  $R_{\text{max}} \geq 2.9$  kpc for order-of-magnitude estimates. Again, such fields of view are easily achieved with current facilities. However, these numbers do show that some care should be taken in targeting sufficiently large galaxies (or galaxy samples) at high redshift. In view of the above estimates, we conclude that most galaxies of interest should host a sufficient number of emission peaks for the method to provide meaningful constraints on the cloud-scale physics of star formation and feedback.

#### 4.3.9 The influence of diffuse emission

Finally, we briefly comment on how the presence of a diffuse emission reservoir affects the derived evolutionary timeline. In this context, ‘diffuse’ refers to emission on spatial scales larger than the region separation length  $\lambda$ . If such a reservoir is present in a given tracer, it decreases the rarity of the emission found in the identified peaks, resulting in different lifetimes for that tracer. Physically, this means that the diffuse phase is included in the timeline and therefore affects its duration. This is not necessarily incorrect, but does require that the diffuse and clumpy emission follow the same large-scale distribution throughout the maps. If diffuse emission is more prominent in one region and emission peaks can only be identified in another, the diffuse emission only contributes to the galactic average emission (obtained for large aperture sizes) and not to the peaks (which dominate at small aperture sizes). As a result, the presence of diffuse emission that is systematically displaced from emission peaks can lead to the counterintuitive situation in which focusing a small aperture on emission peaks of a certain tracer causes a flux deficit of that tracer relative to the galactic average gas-to-stellar flux ratio (Hygate et al. 2018). In the tuning fork diagram of Figure 2, this would manifest itself as a branch crossing the galactic average ( $\mathcal{B} = 1$ ) into the opposite half of the diagram. Therefore, quantitative applications of our method can be fundamentally obstructed by considerable amounts of diffuse emission.

To minimise the impact of diffuse emission on the constrained quantities, any tuning fork diagram showing signs of a substantial contribution from diffuse emission should be discarded. In the present work, we therefore rule out the extreme case in which focusing on an emission peak of a certain tracer results in a deficit of that tracer by more than  $1\sigma$ . Formally, we thus require

$$\max(\mathcal{B}_{\text{star}} - \sigma_{\log_{10} \mathcal{B}_{\text{star}}}) \leq 1, \quad (178)$$

and

$$\min(\mathcal{B}_{\text{gas}} + \sigma_{\log_{10} \mathcal{B}_{\text{gas}}}) \geq 1. \quad (179)$$

These conditions can be satisfied in two ways. Firstly, there may not exist any substantial diffuse emission reservoirs in either map,

which is the simplest situation and the one most conducive to the application of the method. Alternatively, diffuse emission reservoirs may exist, but they have similar contributions to both the young stellar and gas maps while having a similar spatial distribution to the clumpy emission. Such a configuration does not change the galactic average gas-to-stellar flux ratio relative to a strictly clumpy distribution and implies that the flux ratios around the identified peaks are representative for most of the emission in the map. As a result, the gas phase lifetime obtained from the tuning fork diagram is accurate, even if the presence of diffuse emission may increase the effective region filling factors  $\zeta_{\text{star}}$  and  $\zeta_{\text{gas}}$  and thus lead to overestimated durations of the overlap phase. However, such cases are easily dealt with by following the guidelines for the permitted range of region filling factors presented in Section 4.2.4.

While the above conditions already provide a good way of avoiding any major uncertainties due to diffuse emission, we will quantify this further in Hygate et al. (2018) by testing how the accuracy of the quantities constrained by the method depends on the ‘diffuse fraction’ of the maps, i.e. the total fraction of the emission that arises on size scales larger than  $\lambda$ , obtained by analysing the two-dimensional Fourier transforms of the maps. We will use this to define a critical diffuse fraction above which the obtained results are no longer reliable. While the definition of such a critical value is useful for evaluating the accuracy of a measurement, it does not allow one to improve on a rejected experiment, because a map’s suitability for the method is unchanged. However, a more active approach is to filter out the diffuse emission in a map and obtain the best-fitting quantities for the resulting map. This requires an iterative approach, in which the original map is used to derive  $\lambda$ , after which emission on scales  $> \lambda$  is filtered out, and the analysis is repeated using the updated, filtered map, until convergence is reached. This process is expected to lead to convergence, because filtering a map should increase  $\lambda$  and thus require less filtering during the next iteration. We have developed a module of HEISENBERG that extends the method described in this paper by filtering out diffuse emission in Fourier space, enabling the application of the method to observed maps that contain substantial diffuse emission reservoirs. This optional module is described and validated in Hygate et al. (2018).

#### 4.4 Summary: guidelines for observational applications

Sections 4.2 and 4.3 report on how we have systematically tested the method presented in Section 3 with 288 experiments, in which it has been applied to simulated galaxy maps. The goal of these experiments has been to push the method to its extremes and quantitatively assess under which circumstances the retrieved characterisation of the cloud lifecycle are still accurate. Across these tests, we have found that the resulting limits on the applicability of the method are encouraging, enabling systematic applications from the local Universe out to high redshift (see Section 6 below). We now summarise the requirements for obtaining accurate constraints on the three quantities that are fundamentally obtained with the method, i.e. the duration of the gas phase  $t_{\text{gas}}$ , the duration of the overlap phase  $t_{\text{over}}$ , and the region separation length  $\lambda$ . For a detailed explanation of the origin and consequences of these requirements, we refer to the above tests. The accuracy of the additional, derived quantities is discussed in Section 5.

We first list the requirements that apply to all three quantities:

(i)  $|\log_{10}(t_{\text{star}}/t_{\text{gas}})| \leq 1$  (Sections 4.2.3 and 4.3.4): This condition states that the durations of the gas and stellar phases should

not differ by more than an order of magnitude. For larger differences, the retrieved time-scales exhibit systematic biases.

(ii)  $\lambda \geq N_{\text{res}} l_{\text{ap, min}} / \cos i$  (Sections 4.3.6 and 4.3.7): This condition states that the region separation length should be resolved in projection by at least  $N_{\text{res}}$  resolution elements. We require  $N_{\text{res}} = 1$  for  $t_{\text{gas}}$  and  $N_{\text{res}} = 1.5$  for  $t_{\text{over}}$  and  $\lambda$ . At insufficient resolution (or too small a separation length given the resolution), all three quantities exhibit systematic biases.

(iii)  $N_{\text{min}} \geq \{15, 35\}$  (Section 4.3.8): This condition states that the smallest number of emission peaks identified in either of the two maps should be at least 15 or 35. The identification of at least 15 emission peaks enables order-of-magnitude estimates of the constrained quantities, whereas a minimum of 35 peaks yield a precision better than 0.2 dex (50 per cent). In Section 4.3.8, we convert the minimum number of peaks to a lower limit on the spatial extent of the target galaxy. Finally, we note that this is a soft requirement in the sense that the logarithmic precision scales continuously as  $\sigma_{\log} \propto N_{\text{min}}^{-1/2}$ . A higher precision is achieved at a larger number of identified emission peaks.

(iv)  $\max(\mathcal{B}_{\text{star}} - \sigma_{\log_{10} \mathcal{B}_{\text{star}}}) \leq 1$  when focusing on stellar peaks and  $\min(\mathcal{B}_{\text{gas}} + \sigma_{\log_{10} \mathcal{B}_{\text{gas}}}) \geq 1$  when focusing on gas peaks (Section 4.3.9): This condition states that focusing on a stellar or gaseous emission peak should never lead to a deficit of that tracer relative to the galactic average. Such a deficit indicates the presence of a diffuse emission reservoir, which negatively impacts the accuracy of the derived quantities.

(v)  $|\log_{10}[\text{SFR}(t \leq \tau)/\text{SFR}(0)]|_{dt=\{t_{\text{star}}, t_{\text{gas}}\}} \leq 0.2$  (Section 4.2.3): This condition states that, if available, the SFR as a function of age  $t \leq \tau$  should not vary by more than 0.2 dex when averaged over age intervals with a width of  $t_{\text{star}}$  or  $t_{\text{gas}}$ . This ensures that any bias of the retrieved  $t_{\text{gas}}$  due to SFR variations is less than 50 per cent.

(vi) Each independent region should be detectable in both tracer maps at some point in its lifecycle (Section 4.3.5): This condition states that tracer pairs in which some regions in one tracer would never be visible in the other tracer should be avoided. This is unlikely to occur for the transition from gas to young stars traced by e.g. CO and H $\alpha$ , as gas concentrations generally pass through a molecular phase before forming massive stars. However, other (rare) tracers may not be visible in all regions. In that case, the measured duration of that phase is decreased proportionally.

Across all tests discussed in this section, we have removed any experiments that do not satisfy any of the conditions (i)–(iv). These experiments are shown as transparent symbols in the figures. In order for an experiment to be represented by a non-transparent symbol in any of the panels, it must satisfy all four conditions (i)–(iv).

For feedback-related quantities depending on  $t_{\text{over}}$ , we formulate an additional set of requirements:

(vii)  $\max(\zeta_{\text{star}}, \zeta_{\text{gas}}) < \zeta(\delta \log_{10} \mathcal{F})$  (Section 4.2.4 and Appendix B): This condition states that the emission peak size in units of the region separation length should be small enough for the peak identification algorithm to identify adjacent peaks. If this condition is not satisfied, the value of  $t_{\text{over}}$  is overestimated.

(viii)  $\max(\zeta_{\text{star}}, \zeta_{\text{gas}}) < 0.5$  (Section 4.2.4 and Appendix B): This condition states that the emission peak size in units of the region separation length should be small enough to limit the flux contamination from adjacent peaks to 5 per cent. For unresolved point sources, the peak size is set by the PSF (i.e.  $l_{\text{ap, min}}$ ) and this requirement closely matches the condition of point (ii). If this con-

dition is not satisfied, the value of  $t_{\text{over}}$  may be overestimated by more than the corresponding uncertainty of  $0.05\tau$ .

(ix)  $0.05 < t_{\text{over}}/\tau < 0.95$  (Sections 4.2.4 and 4.3.8): This condition states that the duration of the overlap phase should not be too close to zero or the duration of the entire timeline. If this condition is not satisfied, then the retrieved value of  $t_{\text{over}}$  is either an upper limit (if  $t_{\text{over}}\tau < 0.05$ ) or a lower limit (if  $t_{\text{over}}/\tau > 0.95$ ). These numbers are based on a fixed precision at which  $t_{\text{over}}$  is measured of  $0.05\tau$ . However, this precision is set by  $\zeta_{\text{star}}$  and  $\zeta_{\text{gas}}$  as summarised in point (viii) and shown in Appendix B, so any value of  $\zeta > 0.5$  will increase the 5 per cent systematic uncertainty range due to region blending.

(x)  $|\log_{10}[\text{SFR}(t \leq \tau)/\text{SFR}(0)]|_{dt=t_{\text{over}}} \leq 0.2$  (Section 4.2.4): This condition states that, if available, the SFR as a function of age  $t \leq \tau$  should not vary by more than 0.2 dex when averaged over age intervals with a width of  $t_{\text{over}}$ . This ensures that any bias of the retrieved  $t_{\text{over}}$  due to SFR variations is less than 50 per cent.

(xi) A visual inspection of the maps does not reveal abundant region blending (Section 4.2.4): This condition accounts for the fact that, while the above quantitative requirements greatly increase the accuracy of  $t_{\text{over}}$  measurements, a visual check of the maps is desirable to rule out region blending. Visibly isolated emission peaks rule out any contamination of the retrieved temporal overlap by spatial overlap.

Across all tests discussed in this section that assess the accuracy of constraining  $t_{\text{over}}$ , we have removed any experiments not satisfying conditions (vii) and (viii), in addition to conditions (i)–(iv). These experiments are shown as transparent symbols in the figures. In order for an experiment to be represented by a non-transparent symbol in  $t_{\text{over}}$ -related panels, it must satisfy all six conditions (i)–(iv) and (vii)–(viii). If condition (ix) is not satisfied, then the quoted value of  $t_{\text{over}}$  is considered to be a lower limit (if  $t_{\text{over}}/\tau \leq 0.05$ ) or an upper limit (if  $t_{\text{over}}/\tau \geq 0.95$ ). Throughout Section 4.3, such cases are represented by symbols with arrows.

When all of the above conditions are satisfied, the *fitted* quantities ( $t_{\text{gas}}$ ,  $t_{\text{over}}$ ,  $\lambda$ ) should be considered to be fully accurate for point particle maps and accurate to within  $\sim 0.1$  dex for extended emission maps, as shown by the various tests in this section. This refers to a systematic uncertainty and adds in quadrature to the formal logarithmic uncertainty returned by the method. The accuracy of the derived quantities varies (cf. Section 5) and is obtained through the formal error propagation of the systematic error according to the equations provided in Section 3.2.14. Depending on the specific quantity, this can decrease the logarithmic accuracy by some factor  $> 1$ . The most extreme case is the feedback energy efficiency of equations (148), which in the limit  $\epsilon_{\text{sf}} \ll 1$  effectively scales as  $\chi_{\text{fb, E}} \propto \lambda^2 t_{\text{gas}}^{-1} t_{\text{over}}^{-3}$ . If we assume the extreme case of zero covariance between the systematic uncertainties on  $t_{\text{gas}}$ ,  $t_{\text{over}}$ , and  $\lambda$ , this boosts the systematic uncertainty by a factor of  $\sqrt{6} \approx 2.5$  relative to a single fitted quantity. We therefore conclude that, for extended emission maps, the systematic uncertainties of the *derived* quantities fall within a range of 0.1–0.3 dex.

The public release of HEISENBERG will automatically verify that the guidelines listed in this section are satisfied upon the completion of the analysis. Any violations of these conditions are identified in the output. This evaluation also includes those guidelines for selecting appropriate input maps from Section 3.2.2 that are suitable for an automated quantitative assessment.

## 5 DERIVED QUANTITIES DESCRIBING CLOUD-SCALE STAR FORMATION AND FEEDBACK

The previous section showed that the three free parameters  $t_{\text{gas}}$ ,  $t_{\text{over}}$ , and  $\lambda$  are accurately retrieved by the method presented in Section 3. We now turn to a brief set of examples demonstrating that also the derived quantities describing cloud-scale star formation and feedback are well-constrained. This section serves as a proof of concept and is not as exhaustive as Section 4 – rather than providing a discussion of all quantities listed in Table 4, we focus on five of these that we expect to be of immediate physical interest in future applications of the method. These are the radii of stellar regions  $r_{\text{star}}$ , the radii of gas regions  $r_{\text{gas}}$ , the feedback outflow velocity  $v_{\text{fb}}$ , the star formation efficiency per star formation event  $\epsilon_{\text{sf}}$ , and the mass loading factor  $\eta_{\text{fb}}$ . Firstly, we estimate the values that we expect to measure based on our knowledge of the simulations. Secondly, we carry out a quantitative comparison between these values and the retrieved quantities. This comparison exclusively uses the experiments combining the gas maps and young stellar maps from Section 4.3 and only considers those based on extended emission maps (cf. Figure 20). To avoid large spreads in the discussed quantities due to their likely physical dependence on the minimum gas density used to generate the maps, we also restrict ourselves to the maps with the fiducial values of  $\rho_{\text{min}}$ , i.e.  $\rho_{\text{min,LR}} = 1 \text{ m}_H \text{ cm}^{-3}$  and  $\rho_{\text{min,HR}} = 2 \text{ m}_H \text{ cm}^{-3}$ . This allows us to narrow down the range of expected and retrieved values and isolate the accuracy of the method in constraining the derived quantities. Finally, we only consider the experiments that satisfy the conditions for the reliable application of the method listed in Section 4.4, i.e. the opaque data points from Figures 21–25.

### 5.1 Expected values of the derived quantities

Here, we briefly provide expressions for the expected values of the five derived quantities considered here. Analogously to the discussion in Section 4.3.3, these cannot necessarily be obtained directly from the simulations themselves, because the minimum gas densities exceed the density threshold for star formation (i.e.  $\rho_{\text{min}} > \rho_{\text{crit}}$ ). As a result, only a fraction of the progenitors of the young stellar regions could have been visible in the gas maps, which changes the retrieved time-scales and other quantities from the true values in the simulation. This is strictly a numerical effect, caused by the stochastic model for star formation used in the simulations, in which stars can be spawned from low-density gas without ever going through a high-density phase. Real-Universe applications of the method will not suffer from this behaviour, because the collapse towards star formation eventually causes the gas to emit in any of the commonly-used gas tracers. Most importantly for the question at hand, we can easily accommodate how the above effect changes the retrieved quantities, such that their expected values can still be accurately predicted based on our knowledge of the simulations. In equations (167) and (169), this is done by including a factor  $p_{\text{gas}}(\rho_{\text{g}})$ , which represents the fraction of gaseous regions that survives to a density  $\rho_{\text{g}}$  without forming stars. The values of  $t_{\text{gas,exp}}$ ,  $t_{\text{over,exp}}$ , and  $\lambda_{\text{exp}}$  that we thus predict to be retrieved using the method are accurate, as shown in Sections 4.3.4–4.3.7.

The first two quantities that we consider are the radii of the young stellar and gaseous regions, denoted as  $r_{\text{star}}$  and  $r_{\text{gas}}$ , respectively. Out of the five quantities discussed in this section, these are the only ones that can be predicted directly from the simulations, without accounting for the numerical effects due to the stochastic star formation model. As evident from equation (85), we

define the region radius as its Gaussian dispersion, such that we can express the expected radius as

$$r_{\text{exp}} = \frac{1}{2} \sqrt{\frac{f_{\text{Wend}}^2 \bar{h}_{\text{smooth}}^2 + l_{\text{ap,min}}^2}{2 \ln 2}}, \quad (180)$$

where the first term in the numerator represents the mean FWHM of the smoothing kernels of the particles from the simulation shown in the maps, the second term in the numerator represents the minimum aperture size, which sets the FWHM of the resolution at which the maps are analysed, and the denominator converts the FWHM to a Gaussian dispersion. For a Wendland  $C^4$  kernel defined on the interval  $0 < r/h < 1$ , the FWHM spans a fraction  $f_{\text{Wend}} \equiv \Delta(r/h) = 0.56$  of the smoothing length. The mean smoothing length  $\bar{h}$  is determined by averaging over all gas particles with densities above the minimum density (i.e.  $\rho_{\text{g}} \geq \rho_{\text{min,LR}}$  for the low-resolution simulation and  $\rho_{\text{g}} \geq \rho_{\text{min,HR}}$  for the high-resolution simulation), which is  $h_{\text{smooth,LR}} = 267 \text{ pc}$  for the low-resolution simulation and  $h_{\text{smooth,HR}} = 126 \text{ pc}$  for the high-resolution simulation. The minimum aperture size ranges from  $l_{\text{ap,min,LR}} = 50\text{--}400 \text{ pc}$  for the low-resolution simulation and  $l_{\text{ap,min,HR}} = 25\text{--}400 \text{ pc}$  for the high-resolution simulation. These minimum aperture sizes cover a wide range, because in Section 4.3.6 we changed the spatial resolution to quantify the requirements for the reliable application of the method. Together, the above numbers imply expected radii of  $r_{\text{exp,LR}} = 67\text{--}181 \text{ pc}$  for the low-resolution simulation and  $r_{\text{exp,HR}} = 32\text{--}172 \text{ pc}$  for the high-resolution simulation. Because the young stellar maps and the gas maps are both constructed using a Wendland  $C^4$  smoothing kernel with 200 neighbours, we do not distinguish between the radii expected for both types of maps – the retrieved values of  $r_{\text{star}}$  and  $r_{\text{gas}}$  should be similar.

The third quantity that we consider is the feedback outflow velocity  $v_{\text{fb}}$ . As explained above, the expected value of this quantity cannot be estimated directly from the simulations, because the measured time-scales are affected by the stochastic star formation model used in the simulations. Fortunately, it is straightforward to express the value of  $v_{\text{fb}}$  that we expect to retrieve with the method in terms of the expected duration of the overlap phase and the expected region separation length, i.e.

$$v_{\text{fb,exp}} = \frac{\lambda_{\text{exp}}}{2t_{\text{over,exp}}}. \quad (181)$$

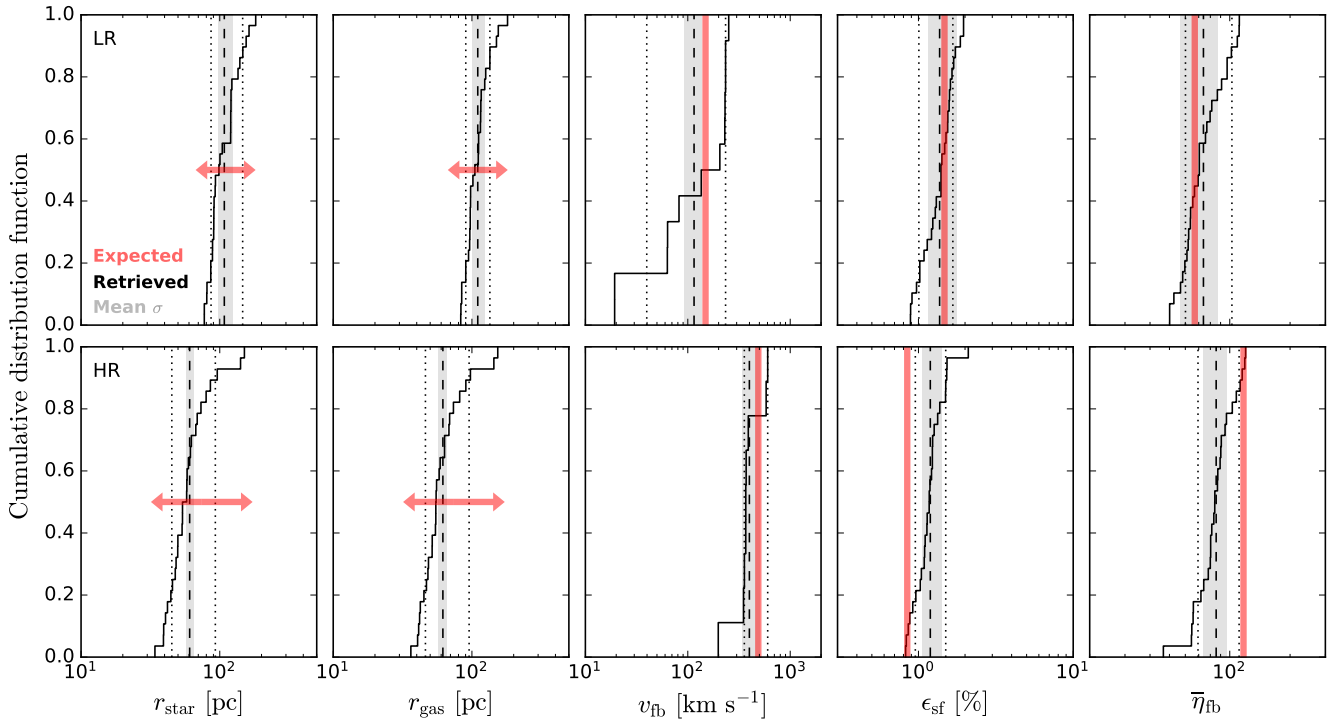
Given the mean expected values of  $t_{\text{over,exp}} = \{1.7, 0.35\} \text{ Myr}$  and  $\lambda_{\text{exp}} = \{548, 284\} \text{ pc}$  for the experiments considered here (see Section 4.3.3) based on the {low, high}-resolution simulations, we thus expect to measure typical feedback outflow velocities of  $v_{\text{fb,exp}} = \{162, 408\} \text{ km s}^{-1}$ , where we have taken the mean velocity over all experiments.

The final two quantities discussed here are closely related. The first is the star formation efficiency per star formation event, which by definition follows from the star formation relation in equation (1) as the ratio between the cloud lifetime (over which stars can form) and the galactic gas depletion time (reflecting how long it takes the galaxy to run out of gas at the current SFR), i.e.

$$\epsilon_{\text{sf,exp}} = \frac{t_{\text{gas,exp}}}{t_{\text{depl}}}. \quad (182)$$

For the {low, high}-resolution simulations, we find expected durations of the gas phase of  $t_{\text{gas,exp}} = \{15, 2.6\} \text{ Myr}$  (see Section 4.3.3) while for  $\rho_{\text{g}} \geq \{1, 2\} \text{ m}_H \text{ cm}^{-3}$ , the simulations have gas depletion times of  $t_{\text{depl}} = \{1.0, 0.3\} \text{ Gyr}$ . After again taking the mean over all experiments, we thus expect to measure typical star formation efficiencies per event of  $\epsilon_{\text{sf,exp}} =$





**Figure 26.** Comparison between five ‘derived’ quantities describing cloud-scale star formation and feedback and the expected values based on the simulations (as described in Section 5.1). Shown are the cumulative distributions (black solid lines) of the measured quantities for the subset of experiments that satisfy the guidelines from Section 4.4 (see the text for details on the sample selection), with the median values, dispersions, and logarithmic mean measurement uncertainties highlighted by vertical dashed lines, dotted lines, and grey bands, respectively. The dispersions of the distributions are derived using the same approach as for the PDFs described in Section 3.2.13. For comparison, the (range of) expected values from the simulations as derived in the text are highlighted by thick red arrows or vertical lines. These results are shown for the low and high-resolution simulations (top and bottom panels, respectively). The figure shows that not just the three free parameters discussed in Section 4 are well-constrained, but also the derived quantities are accurately retrieved using the method presented in this paper.

{1.5, 0.85} per cent. The final considered quantity is the average mass loading factor, i.e. the time-integrated mass outflow rate in units of the star formation rate, of which the expected value follows from  $\epsilon_{\text{sf,exp}}$  immediately as

$$\bar{\eta}_{\text{fb,exp}} = \frac{1 - \epsilon_{\text{sf,exp}}}{\epsilon_{\text{sf,exp}}}, \quad (183)$$

dividing the outflowing mass fraction by the star-forming mass fraction. For the above numbers, we expect to find typical mass loading factors of  $\bar{\eta}_{\text{fb,exp}} = \{67, 117\}$ , averaged over all experiments as before.

## 5.2 Comparison to the measurements using the method

Figure 26 shows the cumulative distributions of the five derived quantities considered in this section as retrieved by our method from the low and high-resolution, extended emission experiments that satisfy the conditions for the reliable application of the method (see Section 4.4). For reference, the expected values based on the lines of reasoning presented in Section 5.1 are displayed in red. Across the ten panels in Figure 26, there is good agreement between the red and black lines. In the bottom-right two panels, the expected values fall marginally outside the the  $1\sigma$ -equivalent percentiles (dotted lines), which is approximately the statistically expected number when considering ten distributions. Therefore, the main result of this comparison is that the method does not only accurately retrieve the free parameters  $t_{\text{gas}}$ ,  $t_{\text{over}}$ , and  $\lambda$ , but also puts

reliable constraints on the derived quantities describing cloud-scale star formation and feedback.

It is particularly worth commenting on the region radii  $r_{\text{star}}$  and  $r_{\text{gas}}$ , for which the horizontal arrows highlight ranges rather than typical values. For these quantities, we expect a range of values, because the spatial resolution of the experiments used to construct the distribution varies through the minimum aperture size ( $l_{\text{ap,min}}$ ). This is highlighted by the fact that the dispersion of the retrieved values (dotted lines in Figure 26) significantly exceeds the mean measurement uncertainty (indicated with the grey area), especially for the high-resolution simulation. Equation (180) quantifies how we expect a change of  $l_{\text{ap,min}}$  to affect the retrieved region radii. While this is not directly shown by the cumulative distribution shown here, there is indeed a strong correlation between  $r_{\text{star}}$  or  $r_{\text{gas}}$  and  $l_{\text{ap,min}}$ , which quantitatively traces the prediction from equation (180) with a mean difference of just 30 per cent. In fact, all experiments fall within a factor of two of the radius expected for their minimum aperture size.

The range of minimum aperture sizes is large enough to cover both the regimes where the smoothing kernels are resolved into several (1–10) resolution elements, which occurs at  $l_{\text{ap,min}} < \{150, 70\}$  pc for the {low, high}-resolution simulations, and where they are smaller than one resolution element, which applies to larger values of  $l_{\text{ap,min}}$ . Thanks to this dynamic range, the good agreement between the measured and expected region radii has two important implications. Firstly, it shows the obvious result that the radii (and hence densities) of clouds and star-forming regions can

be constrained to high accuracy with the presented method. Secondly, the good performance at  $l_{\text{ap},\text{min}} > \{150, 70\}$  pc (for the {low, high}-resolution simulations) also illustrates that even sub-resolution sizes can be measured by subtracting the PSF dispersion in quadrature from the measured region radii, somewhat analogously to the early work on stellar clusters by [Larsen \(1999\)](#).

Across the remaining three variables, i.e. the feedback velocity  $v_{\text{fb}}$ , the star formation efficiency  $\epsilon_{\text{sf}}$ , and the mass loading factor  $\bar{\eta}_{\text{fb}}$ , the dispersions of the retrieved values (dotted lines) largely match the mean measurement uncertainties (grey areas). This implies that the method obtains well-converged measurements, as is evidenced by the good agreement between the medians and the expected values shown by the red vertical lines in [Figure 26](#). We also note that the dispersions are small in an absolute sense, at about 30 per cent (or 0.1 dex), which is similar to the mean difference between the measured region radii and their expected values. Only the feedback velocity of the low-resolution simulation seems to show a systematic variation, with a large (0.4 dex) dispersion that clearly exceeds the mean uncertainty. This discrepancy is largely driven by experiment IDs 20 and 23, which are run with a longer reference time-scale ( $t_{\text{star}} = 30$  Myr) than the other experiments considered here ( $t_{\text{star}} = 3\text{--}10$  Myr). Even though the stellar maps of these experiments do not have unacceptably large values of the region filling factor  $\zeta_{\text{star}}$ , the relatively long reference time-scale does lead to visible region blending, because the emission peaks have relatively extended envelopes in the low-resolution simulation. As a result, the two experiments pass the quantitative criteria for reliable measurements of the duration of the overlap phase in [Section 4.4](#), but they do not pass the qualitative criterion that the maps should not exhibit any visual signs of region blending. In practical applications of the method, they would therefore have been considered unreliable. Quantitatively, the blending leads to overestimated durations of the overlap phase and, hence, underestimated feedback velocities. This behaviour does not occur for the high-resolution simulation, which better resolves the regions and does not show as high a degree of blending as the low-resolution simulation. Therefore, the dispersion of the retrieved feedback velocities in the high-resolution simulation matches the mean measurement uncertainty.

In summary, [Figure 26](#) demonstrates that the derived quantities describing cloud-scale star formation and feedback that are obtained using the method are accurate. While we have only considered five examples in this section to act as a proof of concept, the other quantities are derived using similarly simple and straightforward steps (see [Section 3.2.14](#)). We thus conclude that the diagnostic power of the method is very promising, provided that it can be systematically applied to large samples of galaxies. In the next section, we quantitatively show that this can readily be achieved with current observational facilities.

## 6 APPLICABILITY ACROSS COSMIC HISTORY

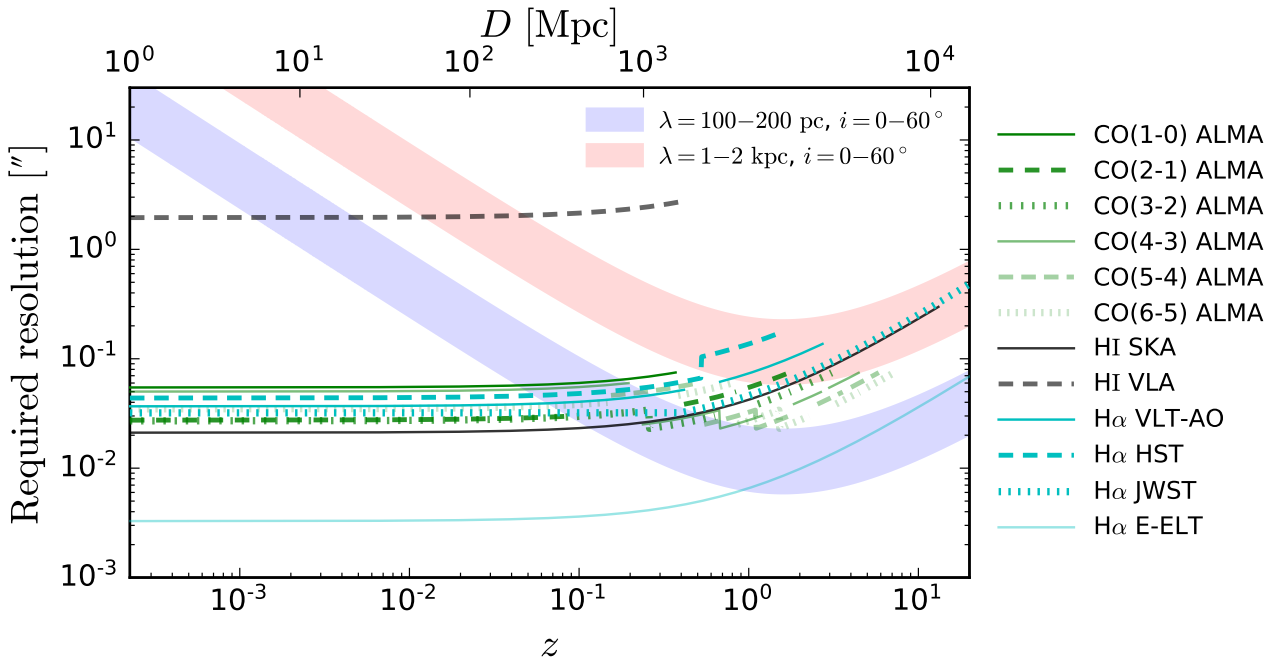
In [Section 4.4](#), we have summarised the requirements for the successful application of the method detailed in [Section 3](#) to observed galaxy maps. Broadly speaking, these requirements can be divided into a ‘soft’ and a ‘hard’ category. By soft requirements, we refer to necessary target properties, which can be satisfied by picking the right galaxy. Examples are a low inclination, a sufficient number of independent regions (usually corresponding to a sufficient spatial extent of the galaxy), a relatively constant SFR over a time-scale  $\tau$ , and best-fitting time-scales  $t_{\text{gas}}$  and  $t_{\text{over}}$  in the range where they are considered reliable. Not all of these are necessarily known a pri-

ori, but when applying the method to a sample of sensibly-selected galaxies with a variety of gas and SFR tracers to choose from, these requirements should generally be satisfied for at least some (and usually most) of the targets.

By hard requirements we mean those that are technically restrictive when formulating observing strategies for applying the method. The two main examples of such requirements are the desired sensitivity [see requirement (vi) in [Section 4.4](#)] and spatial resolution [see requirement (ii) in [Section 4.4](#)]. In principle, any sensitivity can be obtained with infinite observing time, but the spatial resolution is fundamentally limited by the specifications of the observational facility. Therefore, resolution requirements present the most stringent condition for the observational application of our method. The method’s validation in [Section 4](#) demonstrates that accurate results are obtained for a minimum aperture size (i.e. resolution) of at most  $\lambda \cos i$  (for  $t_{\text{gas}}$ ) or  $\lambda \cos i/1.5$  (for  $t_{\text{over}}$  and  $\lambda$ ). In other words, a suitable observation must resolve the typical separation between independent regions along the minor axis of an inclined galaxy. This requirement can be evaluated for current and upcoming observational facilities with a high resolving power.

In [Figure 27](#), we show the spatial resolution required for the application of our method to a typical low-redshift star-forming galaxy (blue band) and a typical high-redshift star-forming galaxy (red band). In this context, a ‘galaxy’ is characterised by a choice of the region separation length  $\lambda$  and the galaxy inclination  $i$ . For local-Universe galaxies, it is possible to carry out a physically-motivated decomposition of position-velocity data cubes into coherent structures that are commonly referred to as clouds or HII regions (e.g. [Rosolowsky & Leroy 2006](#); [Henshaw et al. 2016](#)), which have typical separation lengths of the order  $\lambda = 100\text{--}200$  pc (e.g. [Colombo et al. 2014](#); [Freeman et al. 2017](#)). Early applications of our method yield broadly similar values of  $\lambda$  ([Kruijssen et al. 2018](#); [Hygate et al. in prep.](#); [Schruba et al. in prep.](#); [Chevance et al. in prep.](#); [Ward et al. in prep.](#)). We therefore characterise low-redshift star-forming galaxies with  $\lambda = 100\text{--}200$  pc in [Figure 27](#). By contrast, high-redshift galaxies exhibit clumpy morphologies on much larger scales, of order 1–2 kpc (e.g. [Genzel et al. 2011, 2014](#); [Swinbank et al. 2011, 2012](#); [Hodge et al. 2012](#)). While observations using gravitational lensing ([Dessauges-Zavadsky et al. 2017](#)) and numerical simulations (e.g. [Behrendt et al. 2016](#); [Oklopčić et al. 2017](#)) suggest that these clumps may fragment into smaller structures when observed at higher spatial resolution, it is important to consider that  $\lambda$  represents the separation length between *independent* regions. The clumps seen in high-redshift galaxies are thought to represent the largest self-gravitating scale ([Dekel et al. 2009b](#); [Reina-Campos & Kruijssen 2017](#)), which means that the evolutionary phases of the expected substructure within these clumps will be correlated rather than independent. We therefore adopt  $\lambda = 1\text{--}2$  kpc to characterise the separation length in high-redshift star-forming galaxies. Finally, we adopt a range of inclinations  $i = 0\text{--}60^\circ$ , although we note that this is mainly for illustrative purposes. For these choices, the spreads of  $\lambda$  and  $i$  contribute equally to the width of each band in [Figure 27](#).

For comparison, the lines in [Figure 27](#) show the spatial resolutions that can be attained as a function of redshift for a number of key spectral lines with current and upcoming observational facilities. The legend indicates which spectral lines are observed at the maximum resolution of which facility. We consider several CO transitions for tracing molecular gas, HI for tracing atomic gas, and H $\alpha$  for tracing recent star formation. For the CO lines, we show the maximum attainable resolution with ALMA, the resolution in HI is shown for the Karl G. Jansky Very Large Array (JVLA)



**Figure 27.** Comparison between the spatial resolution required for applying the method to galaxies with certain properties (shaded bands) and the resolution attained by current and upcoming observational facilities (lines) as a function of redshift (bottom  $x$ -axis) or distance (top  $x$ -axis). The blue-shaded band assumes mean separation lengths 100–200 pc, which is characteristic for local-Universe galaxies, whereas the red-shaded band assumes mean separation lengths 1–2 kpc, which is seen in high-redshift systems. Both shaded bands cover a range of inclination angles  $i = 0\text{--}60^\circ$ . The spreads in  $\lambda$  and  $i$  contribute equally to the widths of the bands. As long as the lines are below or within the bands, the attainable resolution is sufficient for the application of our method. The figure shows that, in terms of spatial resolution and wavelength coverage, the method is comfortably applicable out to  $z \sim 4$ . However, the rapid decrease of the sensitivity with redshift implies that observations at  $z > 1$  require a major (but feasible) time investment (see the text).

and the Square Kilometer Array (SKA), and the resolution in  $H\alpha$  is shown for the Very Large Telescope (VLT; using the MUSE or SINFONI instruments with adaptive optics enabled), the Hubble Space Telescope (HST; using WFC3), the James Webb Space Telescope (JWST; using NIRCcam or MIRI), and the European Extremely Large Telescope (E-ELT; using HARMONI, MICADO, or METIS). Some of the lines exhibit resolution jumps at certain redshifts, which is caused by the lines redshifting into a different band or instrument, leading to a change in the maximum attainable resolution. All resolutions are taken from the instruments’ technical specifications prior to the submission of this paper. Whenever the documentation provides a single resolution per band, it is assumed that it applies to the centre of the central wavelength of that band  $\lambda_{\text{band}}$ .<sup>24</sup> The resolution at the wavelength of a line at a given redshift  $\lambda_{\text{line}}(z)$  is obtained by scaling the band’s central resolution by a factor  $\lambda_{\text{line}}(z)/\lambda_{\text{band}} = (1+z)\lambda_{\text{line}}(0)/\lambda_{\text{band}}$ .

Figure 27 shows that currently available facilities (JVLA, ALMA, VLT, HST) allow the application of our method to typical low-redshift galaxies up to distances of  $D \sim 500$  Mpc (or redshift  $z \sim 0.1$  for CO and  $H\alpha$ , whereas for atomic hydrogen this can be achieved within  $D < 10$  Mpc (with prohibitively long integration times already beyond a few Mpc). This implies that the lifecycle of molecular clouds and star-forming regions can be probed systematically using our method across hundreds of low-redshift galaxies covering a wide variety of properties. First efforts in this direction are currently ongoing (Krujissen et al. 2018; Hygate et al. in prep.;

Schruba et al. in prep.; Chevance et al. in prep.; Ward et al. in prep.). Upon the arrival of the SKA, we will be able to include atomic hydrogen across a distance range similar to the CO and  $H\alpha$ , enabling the evolutionary analysis of molecular cloud condensation from the atomic phase.

Only a modest time investment is required for achieving the required spatial resolution and sensitivity across a representative sample of nearby galaxies. Observations of the molecular gas are generally the most prohibitive. We are currently undertaking a 75-hour ALMA Large Programme as part of the PHANGS collaboration, which targets all 80 massive ( $10^{9.75} < M_*/M_\odot < 10^{11}$ ), not edge-on ( $i < 75^\circ$ ), actively star-forming ( $\text{SFR}/M_* > 10^{-2} \text{ Gyr}^{-1}$ ) galaxies at distances  $D < 17$  Mpc (Leroy et al., in preparation), which is sensitive to individual molecular clouds of masses  $M > 10^5 M_\odot$  at a spatial resolution  $1''$  (corresponding to 50–80 pc, depending on distance). Optical and UV observations of SFR tracers at matched spatial resolution can either be obtained with small, ground-based telescopes or have already been taken with HST. This shows that the observations required for the systematic application of our method can be attained with modern facilities.

In addition, Figure 27 demonstrates that the method can be applied to typical high-redshift galaxies out to  $z \sim 2.6$  using currently available facilities. This redshift range is limited by the accessibility of  $H\alpha$  with the VLT and HST until the first light of the JWST and the E-ELT. Once these facilities are online, the redshift range will instead be limited by access to the peak of the CO spectral line energy distribution (SLED) with ALMA, which is expected to be around CO(5-4) in high-redshift galaxies (Carilli & Walter

<sup>24</sup> For the JVLA, we include an additional factor of 1.5 to represent natural weighting rather than uniform weighting.

2013). For CO line observations up to the peak of the CO SLED, the method’s range of application thus reaches out to  $z = 4\text{--}5$ .

We do note that these kinds of high-redshift observations are highly time-consuming. At  $z \sim 1$ , ALMA observations of galaxies with the highest CO surface brightnesses at a sensitivity of 0.1 mJy and a required resolution of  $< 0.2''$  take  $> 20$  hours per galaxy. By  $z \sim 4$ , reaching the required sensitivity at a similar resolution requires  $> 100$  hours (also see Hodge et al. 2012), which is not fundamentally prohibitive, but does imply that a large programme is required to apply the presented method at  $z \gtrsim 2$ .<sup>25</sup> Eventually, it will be possible to systematically characterise the lifecycle of molecular and ionised clumps in high-redshift galaxies for 10–20 galaxies at  $z \sim 1$  and a handful of targets at  $z > 2$ . These numbers increase by a factor of several if the dust continuum is used to trace the molecular ISM. Such maps lack kinematic information, but require less integration time because the dust continuum is brighter than CO line emission. While this may not seem like a large number, we stress that the questions under consideration here previously represented Local Group or solar neighbourhood science. Being able to systematically probe the cloud-scale physics of star formation and feedback for any statistical sample of galaxies (let alone out to  $z \sim 4$ ) is an unprecedented and exciting prospect that is now within reach.

## 7 DISCUSSION

In Sections 3–6, we have presented and validated a new method for probing the cloud-scale physics of star formation and feedback across cosmic history. The described elements of the method represent an optimised set of tools – they have been carefully designed to yield the most accurate measurements of the constrained quantities. We now turn to a broader discussion of the results, paying particular attention to other approaches that we have tested and did not yield satisfactory results, as well as the limitations of the method as described in Section 3.

Specifically, the presented method is based on several choices of parameters and algorithmic approaches, most of which are either straightforward or carry a clear advantage relative to the alternatives. However, there are a number of these choices where (at face value) reasonable alternatives would have been possible. These warrant some further discussion. Below, we discuss the methods used for determining the overlap-to-isolated flux ratios  $\beta_{\text{star}}$  and  $\beta_{\text{gas}}$ , the way in which we account for the spatially extended nature of emission peaks, the influence of galactic morphology or substructure, and the general reliance of our method on the central limit theorem. Readers interested in specific, quantitative guidelines for observational applications are referred to Sections 4.4 and 8.

<sup>25</sup> Another way of gaining access to a larger galaxy sample may be to include high-resolution observations of gravitationally lensed galaxies (e.g. Swinbank et al. 2011; Livermore et al. 2015). However, our method relies strongly on the spatial structure of the galaxy maps, implying that lensed galaxies will require an extremely high-precision lens model to obtain accurate images in the source plane.

## 7.1 Unsuccessful alternative approaches

### 7.1.1 Methods for determining the overlap-to-isolated flux ratios $\beta_{\text{star}}$ and $\beta_{\text{gas}}$

The (possibly complex) time evolution of the young stellar or gas flux in a region can influence the best-fitting evolutionary timeline (see Section 3.2.11). In the context of our method, this time evolution is captured by the parameters  $\beta_{\text{star}}$  and  $\beta_{\text{gas}}$ , which represent the ratio between the flux regions residing in the overlap phase of Figure 1 relative to that of regions outside of that phase (i.e. in the ‘isolated’ phase). This simplistic representation of flux evolution is enabled by the fact that the method decomposes that evolution in just two phases per tracer. As discussed in Section 3.2.9 and Figure 9,  $\beta_{\text{star}}$  and  $\beta_{\text{gas}}$  are determined by sorting the gas-to-stellar (stellar-to-gas) flux ratios of all stellar (gas) peaks and assigning the fraction  $t_{\text{over}}/t_{\text{star}}$  ( $t_{\text{over}}/t_{\text{gas}}$ ) of the highest flux ratios to the overlap phase, whereas the remainder is taken to reside in the isolated phase. However, the best-fitting time-scales  $t_{\text{star}}$ ,  $t_{\text{gas}}$ , and  $t_{\text{over}}$  are only obtained after fitting the model to the data in Section 3.2.12. It is therefore necessary to calculate  $\beta_{\text{star}}$  and  $\beta_{\text{gas}}$  as a function of  $t_{\text{over}}/t_{\text{star}}$  and  $t_{\text{over}}/t_{\text{gas}}$ , respectively, and let these two parameters vary with the time-scales during the fitting process.

As discussed in Section 4.2.5, the above procedure results in accurate measurements of  $\beta_{\text{star}}$  and  $\beta_{\text{gas}}$ . However, it assumes that the identified peaks reasonably sample the underlying timeline. Even if the uncertainty resulting from this assumption is small (see Section 3.2.9), it is worth investigating if other approaches yield more accurate results. Therefore, we have evaluated two alternative methods for determining  $\beta_{\text{star}}$  and  $\beta_{\text{gas}}$ .

Firstly, one could attempt to define a critical gas-to-stellar flux ratio (or its contrast relative to the galactic average) for deciding whether an emission peak is considered to reside in the overlap phase. However, for gradually evolving flux levels of independent regions, defining a critical flux ratio is entirely ad hoc. In practice, it would also depend on the specific tracer used, as different tracers will exhibit different typical evolutionary histories. As demonstrated by Figure 9, there are no obvious jumps near the best-fitting value of (in this example)  $\beta_{\text{star}}$  that would enable identifying a threshold value. The other experiments show similar behaviour. We have experimented with several different thresholds, such as requiring the gas-to-stellar flux ratio to reside within a certain factor of the galactic average to identify a region as ‘overlap’, where we let that factor (or its inverse) vary from 0.05–0.5. The resulting values of  $\beta_{\text{star}}$  and  $\beta_{\text{gas}}$  vary greatly as a function of the adopted threshold value, even if the impact on the best-fitting time-scales is modest (as expected, see Section 3.2.9). Only for the subset of stellar point particle experiments discussed in Section 4.2.5 it is possible to define a threshold, but this is enabled exclusively by the imposed step function flux evolution of the particles in these experiments, which experience a flux change by a factor of 2 upon entering or emerging from the overlap phase (see Section 4.2.2). In other words, defining a critical flux ratio for deciding whether an emission peak resides in the overlap phase may only work with prior knowledge of the evolutionary timeline. This approach is therefore undesirable.

Secondly, one could try to iteratively constrain  $\beta_{\text{star}}$  and  $\beta_{\text{gas}}$ , i.e. by applying the procedure of Section 3.2.9–3.2.13 for some initial guess of these parameters to obtain the best-fitting  $t_{\text{star}}$ ,  $t_{\text{gas}}$ , and  $t_{\text{over}}$  and use these best-fitting values to refine  $\beta_{\text{star}}$  and  $\beta_{\text{gas}}$  analogously to how we currently use the time-scales to determine the fraction of peaks that should reside in the overlap phase. This can then be repeated until convergence is obtained. To some extent, this approach is a less elegant version of how we determine  $\beta_{\text{star}}$



and  $\beta_{\text{gas}}$  in Section 3.2.9, because during a single iteration this version (incorrectly) assumes that the time-scales can be varied without changing the overlap-to-isolated flux ratios accordingly. In addition, it is computationally more expensive, because the procedure of Section 3.2.9–3.2.13 needs to be repeated multiple times. Most importantly, an iterative approach ends up being highly unreliable – tests of this approach often identify two ‘attractor’ solutions that the iteration keeps alternating between, implying that convergence is never obtained.

In summary, the adopted approach for determining the overlap-to-isolated flux ratios  $\beta_{\text{star}}$  and  $\beta_{\text{gas}}$  is the most accurate, most computationally efficient, and the most self-consistent of the options that we have considered. Due to a lack of feasible alternatives, we have not included any other (optional) approaches in the HEISENBERG code.

### 7.1.2 Accounting for the extended nature of emission peaks

Contrary to the original model (KL14) for the ‘tuning fork diagram’ of Figure 2, the model described in Section 3.2.11 does not assume that emission peaks are point sources, but accounts for their extended nature. This is a critical enhancement of the model – in the original form, the flattening of the tuning fork diagram at small aperture sizes is unambiguously interpreted as being due to a non-zero overlap time-scale  $t_{\text{over}} > 0$ , whereas such a flattening is also caused by extended emission peaks with finite central surface densities. As a result, distinguishing the influences of extended emission and the duration of the overlap phase on the flattening of the gas-to-stellar flux ratio at aperture sizes smaller than the region separation length (i.e.  $l_{\text{ap}} < \lambda$ ) is critical for obtaining a non-degenerate characterisation of the evolutionary timeline. As evident by comparing panel (b) with panels (g) and (h) in Figure 11 (in which  $\mathcal{E} \rightarrow \infty$  corresponds to delta function-shaped peak profiles), avoiding this degeneracy may be difficult, but can be achieved using our method. However, this relies quite strongly on which functional form is assumed for the emission peak profiles. This represents the main uncertainty when using our method to constrain feedback-related physics, because it predominantly affects the measurement of the duration of the overlap phase  $t_{\text{over}}$ . We have therefore evaluated various profiles and their possible variation with tracer or evolutionary phase.

Practical applications of the method will often need to deal with objects near the resolution limit. Therefore, we make a strong case in Section 3.2.11 that two-dimensional Gaussians are the preferred profile, because these represent both the observational PSF and the convolution kernel used when convolving the input maps to a common spatial resolution. However, HEISENBERG also includes the option of using point (delta function) or constant surface density disc profiles. When applying these profiles to the extended emission experiments from Section 4, the results become less accurate than when using two-dimensional Gaussians, as expected. Adopting point particle profiles invariably leads to overestimated overlap time-scales, because any flattening of the tuning fork caused by the peak profile is erroneously attributed to the duration of the overlap phase. A constant surface density disc generally does better, because its finite surface density at small aperture sizes appropriately decreases  $t_{\text{over}}$  and could be preferred over a two-dimensional Gaussian, because it is ‘simple’. However, it also provides less accurate results, because a constant surface density does not properly describe the actual peak emission profiles. In the context of the tuning fork in Figure 2, this profile requires the gas-to-stellar flux ratios to be constant once  $l_{\text{ap}} < 2r$ , which implies a sudden transition in

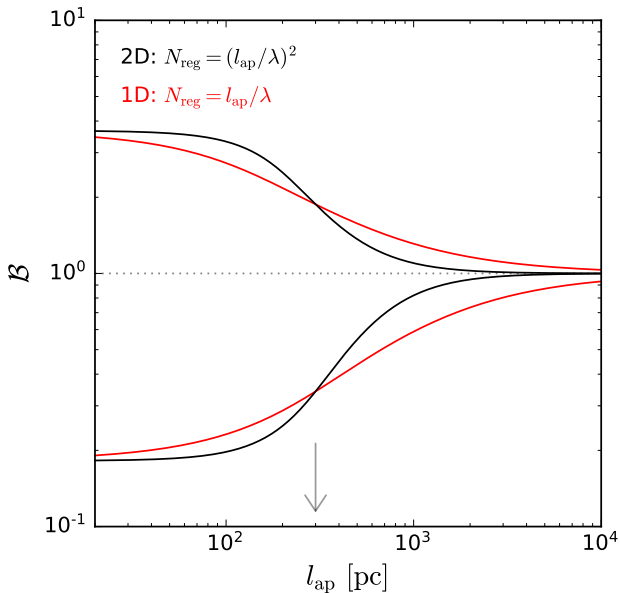
the model tuning fork (i.e. a discontinuity in its first derivative) as shown in panel (i) of Figure 11 that is not present in the observed data points. This mismatch leads to an unsatisfactory recovery of the input time-scales in the experiments of Section 4.2. Therefore, we recommend using a two-dimensional Gaussian peak profile.

Even if a Gaussian profile is preferred, it remains to be decided how to set its width, which we express in terms of a dispersion radius  $r$ . In Section 3.2.11, we use the contrast between the flux level of the emission peaks and the average across the map to determine the ratio  $\zeta \equiv 2r/\lambda$ . This is done separately for young stellar peaks and gas peaks, which is the sensible choice for peaks in the ‘isolated’ phase, i.e. that do not reside in the overlap phase. However, it is not obvious how to determine the peak radii for peaks residing in the overlap phase. Should the stellar and gas components of a single ‘overlap’ peak be allowed to have different radii? If so, should these radii be assumed to be the same across the entire peak sample, irrespective of whether they reside in the isolated or overlap phase, or should a distinction be made between the evolutionary phases? Depending on these choices, it may be necessary to obtain up to four different peak radii.

We start by addressing the first of these questions with an example. If we select a stellar or gas peak that resides in the overlap phase and allow its radius to be different for the gas and stellar flux in the numerator and denominator of the gas-to-stellar flux ratio, then the extreme of the gas-to-stellar flux ratio bias may be reached at a radius larger than the smallest aperture size if the radius difference between both tracers is large enough. In other words, the ‘tuning fork’ curve is no longer monotonically increasing (when focusing on a gas peak) or decreasing (when focusing on a young stellar peak) towards smaller aperture sizes, but exhibits a minimum or maximum at some finite aperture size. This happens when the peak that is being focused on is more extended in its primary tracer (by which it was selected) than in its secondary tracer (the presence of which makes the peak reside in the overlap phase). In that case, the primary tracer (which drives the bias of the gas-to-stellar flux ratio relative to the galactic average) reaches the central surface density at a larger size scale than the secondary tracer does, implying that at smaller apertures the secondary tracer contributes more of the flux density and the degree of flux ratio bias decreases.

The behaviour in which a secondary tracer is considerably more centrally concentrated than the primary tracer might manifest itself on the small scales of individual molecular clumps, which often exhibit a higher star formation efficiency in their central regions than globally (e.g. Kruijssen et al. 2012; Longmore et al. 2014; Ginsburg et al. 2016). However, on the scales of entire molecular clouds (or ‘independent regions’) considered here, the structure of the ISM is scale-free (e.g. Elmegreen & Falgarone 1996; Mac Low & Klessen 2004), which implies that the distribution of star formation should roughly trace the distribution of mass in the cloud even if the stars themselves form at the small-scale density peaks. As a result, the trend reversal described above has not been observed in the several observational applications of our method that are currently underway (Kruijssen et al. 2018, Hygate et al. in prep., Schrubba et al. in prep., Chevance et al. in prep., Ward et al. in prep.), which cover over a dozen different galaxies and tracer pairs. This suggests that using different radii for the stellar and gas emission of a peak in the overlap phase is not the right approach.

However, differences in peak radius are expected as a function of evolutionary stage. For instance, a molecular cloud may contract due to self-gravity or grow by accretion, whereas HII regions may expand before shrinking when the most massive stars have ended their lives. As a result, it is not obvious that the peak radii during



**Figure 28.** Influence of galactic morphology on the predicted gas-to-stellar flux ratio bias as a function of the aperture size when focusing apertures on stellar peaks (bottom branches) or gas peaks (top branches), with parameters corresponding to the fiducial model in Figure 11 (see Table 3). The black lines represent the random distribution of independent regions in two dimensions adopted in our model, whereas the red lines indicate the same model when the regions are randomly positioned along a line. These two extremes bracket the range of ‘tuning fork’ shapes expected for real galaxies with prominent morphological features. The aperture size where  $l_{\text{ap}} = \lambda$  is highlighted by the grey arrow.

the overlap phase should be the same as during the isolated stellar and gas phases. Indeed, tests using a description in which peaks in the isolated and overlap phases share the same radius (but still differ between gas and stars) show less accurate results than when the peak radius is allowed to vary with evolutionary phase for each tracer. Therefore, it is necessary to account for the evolution of the peak radii – the flux contrast between the emission peaks and the map average should be determined separately for peaks in the isolated and overlap phases, as is done in Section 3.2.11.

In summary, exploring alternative descriptions for the radius evolution shows that the model presented in Section 3.2.11 provides the best match to simulated and observed tuning fork diagrams. It thus most accurately constrains the free parameters.

## 7.2 Current limitations of the method

### 7.2.1 The influence of galactic morphology or substructure

The model that is fitted to observed ‘tuning fork’ diagrams in Section 3.2.12 and is described in Section 3.2.11 assumes that independent regions are randomly distributed in two dimensions. Specifically, this assumption implies that the number of regions enclosed by an aperture  $N_{\text{reg}}$  scales as  $N_{\text{reg}} = (l_{\text{ap}}/\lambda)^2$ , which determines how steeply the gas-to-stellar flux ratio bias (in small apertures) transitions to the galactic average (in large apertures) around a characteristic aperture size  $l_{\text{ap}} = \lambda$ . By contrast, if the regions are not randomly distributed in two dimensions, but are situated along a line, then the number of regions enclosed by an aperture scales as  $N_{\text{reg}} = l_{\text{ap}}/\lambda$ . Real galaxies that exhibit some degree of substructure

like spiral arms or a ring may have an  $N_{\text{reg}}$  scaling in between these two extremes.

Figure 28 illustrates the effect of morphology or substructure on the tuning fork diagram by considering both the two-dimensional distribution and the linear sequence of regions. The difference between both models is modest, especially when considering that the black and red lines represent morphological extremes. However, a characteristic distinction is that the linear model transitions more gradually between the single-peak and galactic-average regimes of the aperture size range. Based on these model predictions, it is clear that galactic morphology does not affect the measurement of  $\lambda$  (as both models match at  $l_{\text{ap}} = \lambda$ ), but may influence the obtained evolutionary timeline, particularly if any morphological features are more prominent in one tracer than in the other (i.e. one branch of the tuning fork follows a red line in Figure 28, whereas the other follows a black line).

We have attempted to account for galactic morphology by letting the  $(l_{\text{ap}}/\lambda)^2$  terms in equations (36) and (37) vary according to the actual substructure in the considered maps by stacking the radial emission profiles around the identified emission peaks. However, this is a highly delicate exercise. As before, much of the uncertainty in accounting for morphology comes from the treatment of regions in the overlap phase, for which the spatial structure in both maps needs to be combined. In our experiments investigating this problem, we find that a morphology correction to the default  $N_{\text{reg}} = (l_{\text{ap}}/\lambda)^2$  is very sensitive to the definition of which regions reside in the isolated or overlap phases. To remedy such a dependence, the present paper follows the conservative assumption that the regions are randomly distributed in two dimensions. The validity of this assumption is supported by the fact that the galaxy simulations on which the method is tested do exhibit some degree of substructure (see Figure 15, 17, and 20), yet the retrieved evolutionary timelines are in good agreement with the input values (see Section 4.2). This suggests that the influence of galactic morphology on the accuracy of the method is limited.

As shown by Figure 28, our model for the gas-to-stellar flux ratio as a function of aperture size can provide predictions for the two bracketing cases of substructure, as well as morphologies in between. The challenge is to characterise the substructure in a way that provides a clean measurement of  $N_{\text{reg}}$  as a function of  $l_{\text{ap}}/\lambda$  for any pair of galaxy maps without relying too strongly on the definition of a ‘region’. This line of research is currently being continued and is planned to be addressed in a future paper. Until the completion of that work, we recommend applications of the method to galaxies dominated by strong spiral arms or rings to visually inspect how well the model matches the slope of the observed tuning fork near  $l_{\text{ap}} = \lambda$ . A discrepancy between both slopes would indicate non-random morphological features. Such cases may be improved by the partial masking of the young stellar and gas maps used, or by filtering them at different emission levels to enhance or suppress their morphological features.

If the current version of the method is applied to a galaxy with dominant one-dimensional morphological features such as spiral arms, bars, or rings that propagate down to scales  $l_{\text{ap}} \approx \lambda$ , the constrained quantities may be inaccurate. Figure 28 shows that  $\lambda$  is unaffected, as the flux ratio bias at  $l_{\text{ap}} = \lambda$  is insensitive to the dimensionality of the emission structure. By contrast, the best-fitting value of  $t_{\text{gas}}$  will be incorrect if the morphological features are present down to a scale  $l_{\text{ap}} \approx \lambda$  only (or mainly) in one of the maps. In that case, the tuning fork diagram becomes asymmetric, with one branch providing a bad match to the modelled two-dimensional morphology. This affects the accuracy of  $t_{\text{gas}}$ , because

in our model it controls the asymmetry of the tuning fork diagram (see Figure 11). In other words, varying  $t_{\text{gas}}$  is somewhat degenerate with morphological differences between both maps. Fortunately, such cases are easily recognised, because the fitting procedure then often returns a best-fitting solution that reproduces one branch of the tuning fork diagram, but not the other. Finally,  $t_{\text{over}}$  is relatively unaffected, because it is set on the smallest scales, where  $l_{\text{ap}} < \lambda$  and the  $l_{\text{ap}}/\lambda$  term in the model becomes unimportant.

In closing, it is important to emphasise that while our model assumes a random distribution of regions, this only affects the quality of the fit around  $l_{\text{ap}} = \lambda$ . At aperture sizes  $l_{\text{ap}} \ll \lambda$ , the gas-to-stellar flux ratio bias is set by the properties of the central peak, specifically the central flux density and the duration of the overlap phase. At aperture sizes  $l_{\text{ap}} \gg \lambda$ , the gas-to-stellar flux ratio converges to the galactic average, meaning that large-scale morphological features do not affect the accuracy of the results. In an observational context, this means that spiral arms, bars, or rings can only affect the constrained quantities if they persist down to  $l_{\text{ap}} \approx \lambda$ . In practice, we find that a reasonable degree of two-dimensional spatial isotropy is only needed for  $l_{\text{ap}} \leq 2\lambda$ . This implies a much weaker requirement than ruling out any applications of the method to all galaxies with prominent morphological features. We plan to further investigate the applicability of the method to galaxies with pronounced morphological features in future work, as part of our ongoing observational surveys of the nearby galaxy population (see Section 6).

### 7.2.2 The method's reliance on the central limit theorem

A final point worth discussing is that our method relies quite strongly on the central limit theorem by attributing a common evolutionary timeline to all regions. This approach effectively assumes that region-to-region variations can be linearly combined to retrieve an intensity-weighted average gas-to-stellar flux ratio bias [see equations (36) and (37) in Section 3.2.9]. We demonstrated in KL14 (Figure 5) that this assumption is correct when the regions follow a mass spectrum of non-zero width. The only effect of such a mass spectrum is to increase the statistical uncertainty on the observed gas-to-stellar flux ratio bias, but it does not affect the absolute value of the bias.<sup>26</sup> We note that this increased uncertainty due to the mass spectrum is accounted for in the derivation of the uncertainties in Section 3.2.10. Even if regions of different masses experience different evolutionary timelines, our method of constraining a timeline is not incorrect as long as it is kept in mind that it represents a flux-weighted population average. In essence, it provides a characteristic evolutionary history for Lagrangian mass elements in a galaxy.

Our approach is less adequate when some subset of the independent regions never evolves through a phase in which it is bright in either of the tracers. Section 4.3.5 provides a numerical example based on the stochastic star formation model used in our simulations. In real galaxies, a subset of the molecular cloud population may be disrupted by galactic shear, without ever being associated with high-mass star formation. Alternatively, low-mass clouds may

have too low star formation efficiencies to produce massive stars that are traced by  $\text{H}\alpha$ , FUV, or NUV. In both cases, the rarity of the stellar phase is increased relative to the gas phase, implying that the cloud lifetime inferred with our method will be longer than the lifetime of the star-forming subset of clouds. Specifically, if a fraction  $f_{\text{vis}}$  of all clouds will at a later stage also be visible in the young stellar phase, then the observed lifetime will be a factor of  $f_{\text{vis}}^{-1}$  longer than the lifetime of (eventually) star-forming clouds. A good way of picturing this is that our method measures the statistical average of the total amount of time spent by a mass element in the gas phase before becoming associated with the young stellar phase. For instance, if a cloud is disrupted by shear before it forms massive stars, its lifetime is effectively added to that of a similar cloud that will become associated with star formation.

It depends on the science question at hand whether the above behaviour is desirable. For instance, the theory for cloud lifetimes by Jeffreson & Kruijssen (2018) follows a similar statistical approach by adding up cloud destruction rates due to various mechanisms and subtracting the mechanism that is not associated with star formation (i.e. shear), such that it increases the effective cloud lifetime. In that case, theory and observation are dealing with the same quantities. However, if one is interested in the absolute lifetime of a certain ionised emission line (e.g.  $\text{H}\beta$ , [OII], [OIII], [NII], or [SII]), which is linked to a stellar evolutionary time-scale, it is undesirable to obtain the population-integrated lifetime. In such a case, an effort should be made to mask or filter out emission from regions that will never be visible in the other tracer. For gas maps, this can be achieved by filtering out emission from low-mass clouds (possibly in a diffuse form), which typically have higher virial parameters (Dobbs et al. 2011) and are more prone to destruction by shear. Low-mass clouds are also statistically less likely to form high-mass stars due to the effects of IMF sampling, but this effect is accounted for in our calibrations of the lifetimes of young stellar tracers such as  $\text{H}\alpha$ , FUV, and NUV, which decrease towards low region masses due to stochastic effects (Haydon et al. 2018). In practice, the impact of low-mass clouds and star-forming regions may be limited, because the luminosity functions ( $dN/dL \propto L^\gamma$ ) of both populations have slopes  $\gamma > -2$ , implying that most of the flux emerges from massive clouds and regions (e.g. Elmegreen & Falgarone 1996; Lee et al. 2012).

The examples considered above show that there are feasible ways of dealing with tracer pairs for which the population-averaged evolutionary timeline may not provide meaningful time-scales. However, there is no universal way of doing so. The best approach depends on the tracer pair and sometimes even on the particular data set. This highlights a fundamental point – the method presented in this paper does not represent a ‘black box’ and should not be used as such. A sound astrophysical interpretation of the measurements requires the user to carefully evaluate to what extent the constrained quantities match the physical quantities of interest. When such care is taken, the presented method can provide strong and powerful constraints on important physical quantities describing cloud-scale star formation and feedback across cosmic history.

## 8 CONCLUSIONS

### 8.1 Summary

We present a new method for measuring the key quantities describing the cloud-scale physics of star formation and feedback from high-resolution imaging of the ISM and star formation across

<sup>26</sup> This also applies to the environmental variation of other region properties, such as their sizes, expansion velocities, or emissivities. If a major variation of these properties is a concern (e.g. due to a sharp radial metallicity gradient in the target galaxy), we recommend sub-dividing the galaxy into fields or radial annuli across which these properties are more homogeneous.

galaxy discs. These quantities include the (molecular) cloud lifetime, the feedback time-scale, the mean separation length between independent regions, the region size, the star formation efficiency per star formation event, the feedback outflow velocity, the mass loading factor, the feedback-to-ISM coupling efficiency, as well as several other important quantities (Table 4). This paper explains the method in detail (Section 3) and demonstrates its accuracy using nearly 300 controlled applications to hydrodynamical simulations of star-forming galaxies (Sections 4 and 5), finding that the method is suitable for systematic applications across statistically relevant samples of galaxies, from the nearby Universe out to high redshift (Section 6). This extends the ‘Local Group science’ of cloud-scale star formation and feedback into the realm of galaxy evolution across cosmic time. The main results of this work are as follows.

(i) *Section 2*: Star formation in galaxies can be described as an ensemble average over a population of ‘independent regions’, i.e. concentrations of gas or young stars that reside on a timeline of star formation in an evolutionary phase that is independent of their neighbours. The star formation relation between the gas mass (surface density) and the SFR (surface density) observed when averaging these quantities over large areas of galaxies breaks down below a few times the region separation length  $\lambda$ , where the discretisation of a galaxy into independent regions introduces significant departures from the galaxy-wide average relation due to region-to-region differences in evolutionary phase. These departures become systematic when *focusing* small ( $l_{\text{ap}} \lesssim 2\lambda$ ) apertures on emission peaks of either gas or young stars, such that the gas-to-stellar flux ratio (tracing the gas mass per unit SFR or the gas depletion time) becomes biased to elevated or suppressed values, respectively, because specific evolutionary stages are preferentially selected. The magnitude of these biases directly probes the underlying evolutionary timeline of the regions.

(ii) *Section 3*: We introduce the code HEISENBERG, developed in IDL, which applies the above formalism to obtain quantitative measurements of the duration of each evolutionary phase and the region separation length from observed galaxy maps. The evolutionary phases are defined by the emission maps used, such that a timeline consists of two phases reflected by each map and a third ‘overlap’ phase during which regions are visible in both maps. The method works by identifying emission peaks in the maps, measuring the flux ratio between both maps around the identified peaks as a function of the spatial averaging scale (or aperture size) in a ‘tuning fork diagram’ (Figure 2), and fitting a statistical model that predicts the flux ratio bias as a function of size scale and the underlying evolutionary timeline. This way, the method directly translates the observed flux ratio bias when focusing apertures on emission peaks in either map into the (relative) lifetimes of these peaks. Fundamentally, the method only measures relative lifetimes, but these can be converted into absolute time-scales by using a ‘reference time’ if one of the tracer lifetimes is known.

In the context of star formation in molecular gas clouds and feedback from young stars, obvious emission tracers are CO and  $\text{H}\alpha$ , although it is one of the method’s main strengths that it can be applied to any pair of emission maps that represent different phases of an evolutionary progression. The choice of the SFR tracer is important, because it can be used to set the aforementioned ‘reference time’. For instance,  $\text{H}\alpha$  has a lifetime of about 5 Myr (Leroy et al. 2012; Haydon et al. 2018). Because the method makes use of flux ratio biases, only relative flux differences are important. As a result, the measured time-scales and separation lengths are themselves independent of the (uncertain) conversion factors from flux

to physical quantities such as gas masses or SFRs, provided that these conversion factors are roughly constant across the analysed field. However, by combining the measured evolutionary timeline with these physical quantities, a range of important derived quantities (such as the star formation efficiency and mass loading factor) can be constrained. The method employs Monte-Carlo error propagation, providing the complete PDF for each constrained quantity.

(iii) *Section 4*: The method is systematically tested and validated using simulated galaxy maps, with the goal of identifying exactly under which conditions it reliably characterises cloud-scale star formation and feedback. To this end, we performed hydrodynamical simulations of isolated, star-forming disc galaxies, which are used to generate gaseous and young stellar emission maps. The first set of 88 controlled experiments (described in Section 4.2) uses pairs of stellar maps showing only the stars in specific age bins, granting us complete control over the duration of each phase. These tests show that the obtained time-scales are accurate to within 30 per cent or 0.11 dex. The largest source of uncertainty is the variation of the SFR across the duration of the timeline. We also find that the duration of the overlap phase can be obtained to an accuracy of  $\sim 5$  per cent of the total duration of the evolutionary timeline jointly spanned by both maps. In order to accurately measure the duration of the overlap phase, it is important that the region diameters in both maps do not exceed half the region separation length to avoid the misidentification of spatial overlap as temporal overlap.

The second set of 200 controlled experiments (described in Section 4.3) combines age-binned stellar maps with maps of the gas in the simulations, with the goal of addressing the applicability of the method as a function of observational conditions, such as the choice of SFR tracer and its characteristic lifetime, the choice of gas tracer and its characteristic gas volume density, the spatial resolution of the maps, the inclination angle of the galaxy, the number of independent regions per map, and the diffuse emission fraction. The main results of these tests are as follows.

(a) The lifetimes of the gas and SFR tracers should not differ by more than an order of magnitude. This means that the choice of SFR tracer (which sets the reference time-scale as described by Haydon et al. 2018) may need to be modified after applying the method. For an SFR tracer lifetime of 5 Myr (typical of e.g.  $\text{H}\alpha$ ), accurate cloud lifetimes are obtained for  $t_{\text{gas}} = 0.5\text{--}50$  Myr. By including high-resolution FUV or NUV coverage, the upper limit of this range is increased to several 100 Myr, implying that the method is not limited by the availability of suitable reference time-scales.

(b) The measured cloud lifetimes depend on the gas tracer used and the characteristic volume densities it traces. This dependence is physical in nature and opens up the exciting possibility that the density evolution of gas clouds towards star formation can be probed as a function of absolute time by combining different gas tracers.

(c) After correcting for galaxy inclination, the maps need to have a spatial resolution sufficient to resolve the region separation length  $\lambda$  in order to accurately measure the tracer lifetimes. To obtain the duration of the overlap phase and  $\lambda$  itself, we find that it is necessary to resolve  $\lambda/1.5$ . For cloud separation lengths that are characteristic of nearby galaxies ( $\lambda \sim 200$  pc), this implies spatial resolutions of the order  $2''$ , whereas at clump separations seen in high-redshift galaxies ( $\lambda \sim 2$  kpc), this implies spatial resolutions of the order  $0.2''$ .

(d) In terms of the galaxy inclination angle, the above require-



ments on the spatial resolution typically require  $i < 75^\circ$ , even though this is somewhat dependent on galaxy properties. This implies that galaxy inclination does not significantly limit the method’s applicability.

(e) To constrain the measured time-scales to better than 50 per cent, at least 35 emission peaks per map are needed. Making order-of-magnitude estimates of the evolutionary timeline requires 15 emission peaks per map. For the aforementioned separation lengths and accounting for substructure, the requirement of 35 emission peaks translates to minimum galaxy radii of  $R \geq 1.5$  kpc in local galaxies and  $R \geq 4.4$  kpc at high redshift.

(f) The presence of a large diffuse emission reservoir can cause the selected tracer to be underemphasised around emission peaks relative to the galactic average. This manifests itself as a negative emission bias in the tuning fork diagram of [Figure 2](#) and leads to inaccurate results. Applications showing signs of large diffuse emission fractions should be discarded or spatially filtered (see [Hygate et al. 2018](#)).

As discussed in [Section 4](#), the above conditions are comfortably satisfied with modern observatories such as ALMA and VLT/MUSE across the nearby galaxy population. The method is even readily applicable at high redshift (see below).

(iv) [Section 5](#): The accuracy of the derived physical quantities is assessed by comparing the retrieved values to those expected from the numerical simulations. Specifically, we consider the region radii in both the gas and stellar maps, the feedback velocity, the star formation efficiency per star formation event, and the mass loading factor. We show that each of these is consistent with the values expected based on the included physics, usually to within the uncertainties, but universally to within 50 per cent. This shows that the method provides an accurate way of quantitatively constraining the cloud-scale physics of star formation and feedback.

(v) [Section 6](#): We consider the spatial resolutions attainable by current observational facilities as a function of distance and redshift and compare these to the spatial resolutions required for applications of our method, assuming the aforementioned region separation lengths that are typical of galaxies in the local Universe and at high redshift. With currently available facilities (ALMA, VLT, HST), the method can be applied to typical low-redshift galaxies up to  $D \sim 500$  Mpc or  $z \sim 0.1$  for molecular gas traced by CO and star formation traced by H $\alpha$ . The atomic gas phase can currently be included up to a few Mpc with the JVLA before it becomes prohibitively time-consuming, but the arrival of the SKA will extend the method’s applicability to HI to a similar distance range as for CO. Across the wavelength range, the method can be applied to hundreds of nearby galaxies spanning a variety of masses, morphologies, and SFRs. At high redshifts ( $z > 1$ ), the method can be applied to CO line imaging of 10-20 galaxies at  $z \sim 1$  with a reasonable time investment and a handful of galaxies at  $z > 2$ . When dust continuum is used to trace the gas phase, the number of galaxies that can be observed within a feasible time-scale increases considerably. These numbers mean that the lifecycle of molecular clouds and star-forming regions can be probed systematically for more than a hundred galaxies across cosmic time, spanning a representative range of cosmic environments. This is a major step relative to previous methods and facilities, for which studying the cloud-scale physics of star formation and feedback was restricted to the Local Group or the solar neighbourhood.

(vi) [Section 7](#): The presented method is the end result of a wide variety of renditions that we have explored over the past years. We document the most important subset of these various approaches

by discussing a number of key choices and current limitations of the method and comparing these to the alternatives. We show that our current method for quantifying the flux evolution of independent regions between the isolated and overlap phases (captured by the parameters  $\beta_{\text{star}}$  and  $\beta_{\text{gas}}$ ) is accurate and performs better than reasonable alternatives. The same holds for the way in which we account for the extended density profiles of the emission peaks. However, extremely extended emission components (i.e. diffuse emission) still negatively affect the obtained results. We present reasonable ways of dealing with this problem in [Section 4.3.9](#), whereas a new approach to systematically filter diffuse components from emission maps is presented by [Hygate et al. \(2018\)](#).

Another current limitation of the method is that it assumes a random spatial distribution of independent regions on a size scale  $l_{\text{ap}} \approx \lambda$  and does not account for galactic structure such as dominant spiral arm patterns or bars, which could persist down to these small scales. We quantitatively show that this does not negatively impact the accuracy of  $t_{\text{over}}$  and  $\lambda$ , but can prohibit accurate measurements of  $t_{\text{gas}}$ . Fortunately, such cases are easily recognised as a bad fit to the observed ‘tuning fork diagram’, reproducing only one branch while poorly matching the other. We intend to explore this further in future work and currently list points of attention for applications to galaxies with prominent morphological features to ensure accurate results. Finally, we discuss how the method’s reliance on the central limit theorem, in which regions are connected through an evolutionary timeline and the galactic average incidence ratio between their evolutionary phases emerges on scales much larger than the region separation length, can fail if some regions are never visible in one of the tracers considered, or appear multiple times in one tracer before becoming visible in the other. We present practical ways of interpreting the results in such cases.

## 8.2 Applicability and future work

To conclude this paper, we present a brief outlook on how real-Universe applications of the method are expected to provide insight into the physics of star formation and feedback. Throughout this work, we have mostly presented examples in which our method is applied to a star formation rate tracer map and a (molecular) gas tracer map. However, it has also been reiterated that the method can be applied to any pair of tracer maps that share an evolutionary connection according to the schematic timeline of [Figure 1](#). For instance, the method can be applied to pairs of different gas tracer maps, pairs of different ionised emission line maps, or any combination of these. This enables a broad range of applications, including measurements of the molecular cloud lifetime, the cloud condensation time-scale, the feedback time-scale for cloud destruction, feedback outflow velocity, the mass loading factor, the feedback-ISM coupling efficiency, the star formation efficiency, the fragmentation length of galaxy discs, and many more. The systematic applicability to a broad wavelength and redshift range is key in achieving a census of cloud-scale star formation and feedback as a function of galactic environment.

Each of the above quantities can not only be obtained across a sample of different galaxies, but also as a function of galactocentric radius or local environment within single galaxies. Different tracers can be combined to obtain maps of derived physical quantities such as the density or electron temperature, which can then be used to mask the emission maps prior to applying the method, providing a way to measure the time-scales on which these quantities change. The generality of the method provides an enormous freedom in choosing such applications. Varying the gas tracer (e.g. HI,

CO, HCN, HCO<sup>+</sup>) constrains the absolute time-evolution towards increasing densities during cloud condensation, gravitational collapse, and star formation. When considering the physics of feedback, it may even be possible to choose different tracers to highlight individual feedback mechanisms. For instance, SN remnants can be traced by an elevated [SII]/H $\alpha$  ratio (e.g. Kreckel et al. 2017), implying that their time evolution can be probed by combining a reference map with a young stellar map masked by a critical [SII]/H $\alpha$  line ratio. Similarly, the infrared photon processing that drives radiation pressure feedback may be traced with JWST in the near-infrared, whereas photoionization feedback is traced by H $\alpha$ . These examples show that a promising way forward is to simultaneously observe a variety of feedback tracers with MUSE on the VLT to measure the individual time-scales of each mechanism.

In addition to the observational applications of the presented method, it may be applied to numerical simulations of galaxy formation and evolution. Such applications can serve a variety of purposes. For instance, the method may be used to infer the lifecycle of clouds and star-forming regions in simulations with sparse output intervals, in which the evolution of individual structures cannot be followed between different snapshots. In addition, galaxy simulations may be used to test the method further by considering galaxies spanning a wider variety of physical properties than included in this work. By post-processing such simulations to generate synthetic observations of (molecular) gas or SFR tracers (see e.g. Pawlik & Schaye 2011; da Silva et al. 2012; Krumholz 2014; Haworth et al. 2017), it is possible to derive the tracer lifetimes and test how the method’s performance is affected by environmental variations in tracer emissivity. First efforts in this direction are currently in progress.

This reference paper is accompanied by two companion papers. Firstly, in Haydon et al. (2018), we apply the method to synthetic SFR tracer maps of simulated disc galaxies, to calibrate the reference time-scales of the main SFR tracers (e.g. H $\alpha$ , FUV, NUV) as a function of the metallicity and SFR surface density. Both of these quantities are expected to change the SFR tracer lifetimes due to differences in massive star lifetimes and varying degrees of IMF sampling, respectively. The results presented by Haydon et al. (2018) systematically quantify these dependences and thus provide the reference time-scales needed for the systematic application of the method to large galaxy samples. Secondly, in Hygate et al. (2018), we develop a new approach for quantifying the amount of diffuse emission in tracer maps and correcting for its presence. To this end, we extend the method with a module for filtering out the diffuse component. This enables the method’s application to galaxies with high diffuse fractions, which generally have high surface densities. The approach presented by Hygate et al. (2018) thus greatly increases the method’s range of possible applications.

With the presented method at hand, it is possible to empirically constrain the main unknowns in galaxy formation simulations, such as the star formation time-scale, the star formation efficiency, the feedback outflow rate, and its coupling efficiency. We are currently carrying out systematic applications of the method to a large sample of nearby galaxies aimed at probing and understanding these physical quantities as a function of the galactic environment. Initially, this focuses on single-tracer observations of individual galaxies, such as NGC300 (Kruijssen et al. 2018), M33 (Hygate et al. in prep.), and M31 (Schruba et al. in prep.), enabling a detailed understanding of the interplay between galactic environment and the cloud lifecycle across the face of nearby galaxies, as well as providing a point of reference for comparison to previous work. In the intermediate term, we intend to capitalise on the method’s straightfor-

ward applicability by extending these applications to large galaxy samples (Chevance et al. in prep.) and panchromatic observations of many different pairs of emission maps, initially in the Large Magellanic Cloud (Ward et al. in prep.), but eventually covering the nearby galaxy population. We anticipate that the resulting measurements will provide the most systematic and accurate constraints to date on the cloud-scale physics of star formation and feedback across the galaxy population, that way providing a critical point of reference for calibrating the sub-grid physics in the next generation of large-scale galaxy formation models.

## ACKNOWLEDGEMENTS

JMDK gratefully acknowledges funding from the German Research Foundation (DFG) in the form of an Emmy Noether Research Group (grant number KR4801/1-1, PI Kruijssen), from the European Research Council (ERC) under the European Union’s Horizon 2020 research and innovation programme via the ERC Starting Grant MUSTANG (grant agreement number 714907, PI Kruijssen), and from Sonderforschungsbereich SFB 881 “The Milky Way System” (subproject P1) of the DFG. APSH and DTH are fellows of the International Max Planck Research School for Astronomy and Cosmic Physics at the University of Heidelberg (IMPRS-HD). The Flatiron Institute is supported by the Simons Foundation. The numerical simulations were performed on the Odin cluster hosted by the Max Planck Computing & Data Facility (<http://www.mpcdf.mpg.de/>). This work has made use of Numpy (van der Walt et al. 2011), Scipy (Jones et al. 2001), Astropy (Astropy Collaboration et al. 2013), and Pynbody (Pontzen et al. 2013). All figures in this paper were produced using the Python library Matplotlib (Hunter 2007). JMDK thanks Simon White for many stimulating conversations that helped shape the results presented in this paper. We thank Frank Bigiel, Andi Burkert, Daniela Calzetti, Mélanie Chevance, Bruce Elmegreen, Robert Feldmann, Simon Glover, Jonathan Henshaw, Annie Hughes, Sarah Jeffreson, Cliff Johnson, Rob Kennicutt, Ralf Klessen, Kathryn Kreckel, Sophie Kruijssen, Mark Krumholz, Charlie Lada, Adam Leroy, Sharon Meidt, Thorsten Naab, Eve Ostriker, Paolo Padoan, Eric Pellegrini, Marta Reina-Campos, Hans-Walter Rix, Joop Schaye, Eva Schinnerer, Andreas Schmidt, Linda Tacconi, Fabian Walter, and Jacob Ward for insightful comments and discussions during the development of this work. We are greatly indebted to an anonymous referee for their careful reading of the manuscript and for their constructive feedback, which improved the presentation of this paper.

## REFERENCES

- Agertz O., et al., 2007, *MNRAS*, **380**, 963  
 Agertz O., Kravtsov A. V., Leitner S. N., Gnedin N. Y., 2013, *ApJ*, **770**, 25  
 Astropy Collaboration et al., 2013, *A&A*, **558**, A33  
 Aumer M., White S. D. M., Naab T., Scannapieco C., 2013, *MNRAS*, **434**, 3142  
 Barnes A. T., Longmore S. N., Battersby C., Bally J., Kruijssen J. M. D., Henshaw J. D., Walker D. L., 2017, *MNRAS*, **469**, 2263  
 Behrendt M., Burkert A., Schartmann M., 2016, *ApJ*, **819**, L2  
 Bigiel F., Leroy A., Walter F., Brinks E., de Blok W. J. G., Madore B., Thornley M. D., 2008, *AJ*, **136**, 2846  
 Blair W. P., Long K. S., 2004, *ApJS*, **155**, 101  
 Blitz L., Rosolowsky E., 2006, *ApJ*, **650**, 933  
 Braun H., Schmidt W., 2015, *MNRAS*, **454**, 1545  
 Caldú-Primo A., Schruba A., 2016, *AJ*, **151**, 34

- Carilli C. L., Walter F., 2013, *ARA&A*, 51, 105
- Chabrier G., 2003, *PASP*, 115, 763
- Colombo D., et al., 2014, *ApJ*, 784, 3
- Cullen L., Dehnen W., 2010, *MNRAS*, 408, 669
- Daddi E., et al., 2010, *ApJ*, 714, L118
- Dale J. E., 2015, *New Astron. Rev.*, 68, 1
- Dale J. E., Ercolano B., Bonnell I. A., 2013, *MNRAS*, 430, 234
- Dehnen W., Aly H., 2012, *MNRAS*, 425, 1068
- Dekel A., et al., 2009a, *Nature*, 457, 451
- Dekel A., Sari R., Ceverino D., 2009b, *ApJ*, 703, 785
- Dessauges-Zavadsky M., Schaerer D., Cava A., Mayer L., Tamburello V., 2017, *ApJ*, 836, L22
- Dobbs C. L., Burkert A., Pringle J. E., 2011, *MNRAS*, 413, 2935
- Dobbs C. L., et al., 2014, *Protostars and Planets VI*, pp 3–26
- Durier F., Dalla Vecchia C., 2012, *MNRAS*, 419, 465
- Ekström S., et al., 2012, *A&A*, 537, A146
- Elmegreen B. G., 2000, *ApJ*, 530, 277
- Elmegreen B. G., 2002, *ApJ*, 577, 206
- Elmegreen B. G., Falgarone E., 1996, *ApJ*, 471, 816
- Engargiola G., Plambeck R. L., Rosolowsky E., Blitz L., 2003, *ApJS*, 149, 343
- Evans II N. J., et al., 2009, *ApJS*, 181, 321
- Federrath C., Klessen R. S., 2012, *ApJ*, 761, 156
- Feldmann R., Gnedin N. Y., Kravtsov A. V., 2011, *ApJ*, 732, 115
- Feldmann R., Gnedin N. Y., Kravtsov A. V., 2012, *ApJ*, 758, 127
- Forero-Romero J. E., Contreras S., Padilla N., 2014, *MNRAS*, 443, 1090
- Freeman P., Rosolowsky E., Kruijssen J. M. D., Bastian N., Adamo A., 2017, *MNRAS*, 468, 1769
- Genzel R., et al., 2010, *MNRAS*, 407, 2091
- Genzel R., et al., 2011, *ApJ*, 733, 101
- Genzel R., et al., 2014, *ApJ*, 785, 75
- Ginsburg A., et al., 2016, *A&A*, 595, A27
- Haas M. R., Schaye J., Booth C. M., Dalla Vecchia C., Springel V., Theuns T., Wiersma R. P. C., 2013, *MNRAS*, 435, 2931
- Haworth T. J., Glover S. C. O., Koeperl C. M., Bisbas T. G., Dale J. E., 2017, *New Ast. Rev. in press*, arXiv:1711.05275,
- Haydon D. T., Kruijssen J. M. D., Hygate A. P. S., Schruha A., Krumholz M. R., Longmore S. N., 2018, *MNRAS* to be submitted
- Heiderman A., Evans II N. J., Allen L. E., Huard T., Heyer M., 2010, *ApJ*, 723, 1019
- Hennebelle P., Chabrier G., 2011, *ApJ*, 743, L29
- Henshaw J. D., et al., 2016, *MNRAS*, 457, 2675
- Hernquist L., 1990, *ApJ*, 356, 359
- Hodge J. A., Carilli C. L., Walter F., de Blok W. J. G., Riechers D., Daddi E., Lentati L., 2012, *ApJ*, 760, 11
- Hopkins P. F., 2013, *MNRAS*, 428, 2840
- Hopkins P. F., Quataert E., Murray N., 2011, *MNRAS*, 417, 950
- Hopkins P. F., Narayanan D., Murray N., 2013, *MNRAS*, 432, 2647
- Hopkins P. F., Kereš D., Oñorbe J., Faucher-Giguère C.-A., Quataert E., Murray N., Bullock J. S., 2014, *MNRAS*, 445, 581
- Hu C.-Y., Naab T., Walch S., Moster B. P., Oser L., 2014, *MNRAS*, 443, 1173
- Hunter J. D., 2007, *Computing In Science & Engineering*, 9, 90
- Hygate A. P. S., Kruijssen J. M. D., Walter F., Schruha A., Haydon D. T., Longmore S. N., Chevance M., 2018, *MNRAS* to be submitted
- Jeffreson S. M. R., Kruijssen J. M. D., 2018, *MNRAS*, 476, 3688
- Jones E., Oliphant T., Peterson P., et al., 2001, *SciPy: Open source scientific tools for Python*, <http://www.scipy.org/>
- Katz N., 1992, *ApJ*, 391, 502
- Kawamura A., et al., 2009, *ApJS*, 184, 1
- Kennicutt R. C., Evans N. J., 2012, *ARA&A*, 50, 531
- Kreckel K., Groves B., Bigiel F., Blanc G. A., Kruijssen J. M. D., Hughes A., Schruha A., Schinnerer E., 2017, *ApJ*, 834, 174
- Kruijssen J. M. D., 2012, *MNRAS*, 426, 3008
- Kruijssen J. M. D., 2013, arXiv:1304.4600,
- Kruijssen J. M. D., 2014, *Classical and Quantum Gravity*, 31, 244006
- Kruijssen J. M. D., Longmore S. N., 2014, *MNRAS*, 439, 3239
- Kruijssen J. M. D., Maschberger T., Moeckel N., Clarke C. J., Bastian N., Bonnell I. A., 2012, *MNRAS*, 419, 841
- Kruijssen J. M. D., Dale J. E., Longmore S. N., 2015, *MNRAS*, 447, 1059
- Kruijssen J. M. D., et al., 2018, to be submitted
- Krumholz M. R., 2014, *MNRAS*, 437, 1662
- Krumholz M. R., McKee C. F., 2005, *ApJ*, 630, 250
- Krumholz M. R., Tan J. C., 2007, *ApJ*, 654, 304
- Krumholz M. R., McKee C. F., Tumlinson J., 2009, *ApJ*, 693, 216
- Krumholz M. R., Dekel A., McKee C. F., 2012, *ApJ*, 745, 69
- Krumholz M. R., et al., 2014, *Protostars and Planets VI*, pp 243–266
- Krumholz M. R., Kruijssen J. M. D., Crocker R. M., 2017, *MNRAS*, 466, 1213
- Lada C. J., Lombardi M., Alves J. F., 2010, *ApJ*, 724, 687
- Lada C. J., Forbrich J., Lombardi M., Alves J. F., 2012, *ApJ*, 745, 190
- Larsen S. S., 1999, *A&AS*, 139, 393
- Lee E. J., Murray N., Rahman M., 2012, *ApJ*, 752, 146
- Leitherer C., et al., 1999, *ApJS*, 123, 3
- Leroy A. K., et al., 2012, *AJ*, 144, 3
- Leroy A. K., et al., 2013, *AJ*, 146, 19
- Leroy A. K., et al., 2016, *ApJ*, 831, 16
- Leroy A. K., et al., 2017, *ApJ*, 835, 217
- Liu G., Koda J., Calzetti D., Fukuhara M., Momose R., 2011, *ApJ*, 735, 63
- Livermore R. C., et al., 2015, *MNRAS*, 450, 1812
- Longmore S. N., et al., 2014, *Protostars and Planets VI*, pp 291–314
- Lopez L. A., Krumholz M. R., Bolatto A. D., Prochaska J. X., Ramirez-Ruiz E., 2011, *ApJ*, 731, 91
- Lopez L. A., Krumholz M. R., Bolatto A. D., Prochaska J. X., Ramirez-Ruiz E., Castro D., 2014, *ApJ*, 795, 121
- Mac Low M.-M., Klessen R. S., 2004, *Reviews of Modern Physics*, 76, 125
- McKee C. F., Ostriker E. C., 2007, *ARA&A*, 45, 565
- McLeod A. F., Dale J. E., Ginsburg A., Ercolano B., Gritschneider M., Ramsay S., Testi L., 2015, *MNRAS*, 450, 1057
- McLeod A. F., et al., 2016, *MNRAS*, 462, 3537
- Meidt S. E., et al., 2015, *ApJ*, 806, 72
- Monaghan J. J., 1992, *ARA&A*, 30, 543
- Murray N., 2011, *ApJ*, 729, 133
- Murray N., Quataert E., Thompson T. A., 2010, *ApJ*, 709, 191
- Navarro J. F., Frenk C. S., White S. D. M., 1997, *ApJ*, 490, 493
- Oklopčić A., Hopkins P. F., Feldmann R., Kereš D., Faucher-Giguère C.-A., Murray N., 2017, *MNRAS*, 465, 952
- Onodera S., et al., 2010, *ApJ*, 722, L127
- Oppenheimer B. D., Davé R., 2006, *MNRAS*, 373, 1265
- Oppenheimer B. D., Davé R., 2008, *MNRAS*, 387, 577
- Ostriker E. C., Shetty R., 2011, *ApJ*, 731, 41
- Padoan P., Nordlund Å., 2011, *ApJ*, 730, 40
- Pawlik A. H., Schaye J., 2011, *MNRAS*, 412, 1943
- Pellegrini E. W., Baldwin J. A., Ferland G. J., 2010, *ApJS*, 191, 160
- Pety J., et al., 2013, *ApJ*, 779, 43
- Pfeffer J., Kruijssen J. M. D., Crain R. A., Bastian N., 2018, *MNRAS*, 475, 4309
- Planck Collaboration et al., 2014, *A&A*, 571, A16
- Pontzen A., Roškar R., Stinson G., Woods R., 2013, pynbody: N-Body/SPH analysis for python, Astrophysics Source Code Library (ascl:1305.002)
- Portegies Zwart S. F., McMillan S. L. W., Gieles M., 2010, *ARA&A*, 48, 431
- Press W. H., Teukolsky S. A., Vetterling W. T., Flannery B. P., 1992, *Numerical recipes in FORTRAN. The art of scientific computing*. Cambridge: University Press
- Price D. J., 2012, *Journal of Computational Physics*, 231, 759
- Read J. I., Hayfield T., 2012, *MNRAS*, 422, 3037
- Read J. I., Hayfield T., Agertz O., 2010, *MNRAS*, 405, 1513
- Reina-Campos M., Kruijssen J. M. D., 2017, *MNRAS*, 469, 1282
- Ritchie B. W., Thomas P. A., 2001, *MNRAS*, 323, 743
- Rosolowsky E., Leroy A., 2006, *PASP*, 118, 590
- Saitoh T. R., Makino J., 2009, *ApJ*, 697, L99
- Saitoh T. R., Makino J., 2013, *ApJ*, 768, 44
- Scannapieco C., Tissera P. B., White S. D. M., Springel V., 2006, *MNRAS*, 371, 1125



- Schaye J., et al., 2015, *MNRAS*, **446**, 521
- Schruba A., Leroy A. K., Walter F., Sandstrom K., Rosolowsky E., 2010, *ApJ*, **722**, 1699
- Schruba A., et al., 2011, *AJ*, **142**, 37
- Schruba A., et al., 2017, *ApJ*, **835**, 278
- Semenov V. A., Kravtsov A. V., Gnedin N. Y., 2016, *ApJ*, **826**, 200
- Shen S., Mo H. J., White S. D. M., Blanton M. R., Kauffmann G., Voges W., Brinkmann J., Csabai I., 2003, *MNRAS*, **343**, 978
- Shirley Y. L., 2015, *PASP*, **127**, 299
- Springel V., 2005, *MNRAS*, **364**, 1105
- Springel V., Hernquist L., 2003, *MNRAS*, **339**, 289
- Springel V., Di Matteo T., Hernquist L., 2005a, *MNRAS*, **361**, 776
- Springel V., et al., 2005b, *Nature*, **435**, 629
- Swinbank A. M., et al., 2011, *ApJ*, **742**, 11
- Swinbank A. M., Smail I., Sobral D., Theuns T., Best P. N., Geach J. E., 2012, *ApJ*, **760**, 130
- Tacconi L. J., et al., 2013, *ApJ*, **768**, 74
- Usero A., et al., 2015, *AJ*, **150**, 115
- Vogelsberger M., et al., 2014, *Nature*, **509**, 177
- Walker D. L., Longmore S. N., Bastian N., Kruijssen J. M. D., Rathborne J. M., Jackson J. M., Foster J. B., Contreras Y., 2015, *MNRAS*, **449**, 715
- Walker D. L., Longmore S. N., Bastian N., Kruijssen J. M. D., Rathborne J. M., Galván-Madrid R., Liu H. B., 2016, *MNRAS*, **457**, 4536
- White S. D. M., 1984, *ApJ*, **286**, 38
- Wiersma R. P. C., Schaye J., Smith B. D., 2009, *MNRAS*, **393**, 99
- Williams J. P., de Geus E. J., Blitz L., 1994, *ApJ*, **428**, 693
- da Silva R. L., Fumagalli M., Krumholz M., 2012, *ApJ*, **745**, 145
- van der Walt S., Colbert S. C., Varoquaux G., 2011, *Computing in Science and Engg.*, **13**, 22

## APPENDIX A: CALCULATING THE EXACT ENCLOSED PIXEL AREA WHEN INTERSECTED BY A CIRCULAR APERTURE

As discussed in Section 3.2.5, the numerical use of a tophat kernel convolution to determine all flux within an aperture of a certain size must account for the fact that the kernel boundary intersects with pixels, for which the kernel value is a real number between zero and unity, i.e.  $\{W_{ij} \in \mathbb{R} \mid 0 \leq W_{ij} \leq 1\}$ . This number represents the fraction of the pixel area enclosed by the aperture. In this appendix, we explain how the exact fraction of each intersected pixel's surface area for which  $r \leq h$  is analytically calculated by dividing up the pixel area into square, rectangular, triangular, or circular segment shapes and adding or subtracting the appropriate parts of the pixel.

To determine the area of an intersected pixel that falls within the aperture boundary, we first determine the coordinates of the pixel corners and consider the five separate cases of  $N = \{0, 1, 2, 3, 4\}$  corners enclosed by the aperture. By considering these different cases, we can easily generalise the problem to dealing with aperture-pixel intersections having eight different topologies, which are shown in Figure A1. Intersections 0(a) and 4 represent the trivial cases of pixels residing entirely outside or within the aperture, respectively. Each of the six remaining cases can be assembled by adding and subtracting squares, rectangles, triangles, and circular segments. Intersection 0(c) shows the shape of a single circular segment, which consists of a straight line and a circular arc. If we define the thickness of the circular segment  $l_{cs}$  as the maximum distance between the straight line and the arc, measured perpendicularly to the straight line, then the area under the circular segment is expressed analytically as

$$A_{cs} = \frac{l_{ap}^2}{4} \cos^{-1} \left( \frac{R - l_{cs}}{R} \right) - (R - l_{cs}) \sqrt{2Rl_{cs} - l_{cs}^2}. \quad (A1)$$

In combination with the familiar expressions for the areas of a square, rectangle, and triangle, this means that for each of the pixels that are intersected by the aperture boundary, it is straightforward to obtain the exact fraction of the pixel area enclosed by the aperture.

## APPENDIX B: REQUIREMENTS FOR MINIMISING THE EFFECTS OF BLENDING FROM ADJACENT PEAKS

Here we quantify the effects of the blending between adjacent peaks with two-dimensional Gaussian surface density profiles (see Section 4.2.4). Specifically, we focus on the identification of those peaks using the clump finding algorithm (see Section 3) and on the mutual contamination between these peaks.

### B1 Peak identification

To identify peaks, our method uses a commonly-used clump finding algorithm (Williams et al. 1994) that selects peaks by looking for closed contours in the maps. By default, these contours are logarithmically spaced by a specified amount  $\delta \log_{10} \mathcal{F}$ . For the {low, high} resolution simulations, we use  $\delta \log_{10} \mathcal{F} = \{0.25, 0.5\}$ . In order to identify two adjacent peaks, the contrast between the logarithms of the peak flux and the minimum in between both peaks should be at least  $\delta \log_{10} \mathcal{F}$ . For a set of two Gaussian profiles with standard deviation  $r$ , centred on  $(x, y) = (0, 0)$  and  $(x, y) = (\lambda, 0)$ , the flux in the two-dimensional plane is

$$\mathcal{F}(x, y) \propto e^{-(x^2+y^2)/2r^2} + e^{-[(x-\lambda)^2+y^2]/2r^2}, \quad (B1)$$

with the partial derivative along the line connecting both Gaussians ( $y = 0$ ) given by

$$\frac{\partial \mathcal{F}(x, 0)}{\partial x} \propto (\lambda - x) e^{-(x-\lambda)^2/2r^2} - x e^{-x^2/2r^2}. \quad (B2)$$

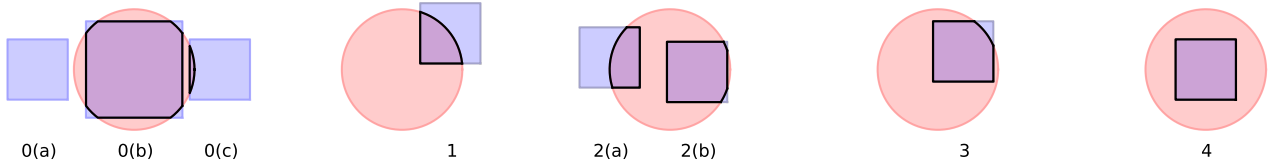
We numerically solve  $\partial \mathcal{F}(x, 0)/\partial x = 0$  to locate the peaks and the minimum between both peaks. The profile only has a minimum for separation lengths  $\lambda > 2r$  (or filling factors  $\zeta \equiv 2r/\lambda < 1$ ), because for smaller separations (or larger filling factors) the Gaussians blend together into a single peak. For  $\lambda > 2r$ , the positions of both peaks are shifted inwards relative to the centres of the isolated Gaussians, due to the asymmetry of having a neighbour contributing emission from only one direction. The observed peak separation is therefore smaller than the true separation between the underlying Gaussians and the measured filling factor  $\zeta$  exceeds the true filling factor. However, the difference is minor – for  $\zeta \leq 0.6$ , the true and observed filling factor are effectively indistinguishable.

Having located the locations of the blended minimum and maxima in the double-Gaussian profile, it is trivial to evaluate the combined flux at each of these positions and determine the flux contrast between maximum and minimum as a function of  $\zeta$ . This is shown in the top panel of Figure B1. We highlight the flux contrasts used to identify the peaks in the simulated maps as horizontal and vertical dotted lines, which define the maximum values  $\zeta_{crit}$  below which the contrast is large enough to identify peaks. We discard all experiments with  $\zeta > \zeta_{crit}$ , where for the {low, high} resolution simulations we obtain  $\zeta_{crit} = \{0.64, 0.52\}$ .

### B2 Contamination

Given two adjacent Gaussians, the integrated flux under each of these includes some contribution from the other peak. For our method to accurately determine the fraction of time for which both





**Figure A1.** Illustration of all possible intersections (black lines enclosing purple-shaded areas) between circular apertures (red circles) and pixels (blue squares). From left to right, the panels show all cases of having  $\{0, 1, 2, 3, 4\}$  corners of the pixel within the aperture. The numbers (with letters) along the bottom of the figure indicate the eight possible topologies. These illustrate that the surface areas of all intersections can be calculated through simple geometry by adding and subtracting squares, rectangles, triangles, and circular segments. For reference, the intersection of pixel 0(c) resembles a single circular segment. Note that the relative sizes and positions of the apertures and pixels are chosen to best illustrate the different topologies, but are not representative of the relative sizes (or positions) when calculating the total flux within apertures of various sizes in Section 3.2.5. Typical aperture sizes exceed the pixel size by a factor 3–500 (see Sections 3 and 4), implying that the curvature of the aperture boundary on the scale of a pixel varies as a function of aperture size from being highly significant to effectively unnoticeable.

tracers coexist (i.e. the duration of the overlap phase  $t_{\text{over}}$ ), we wish to limit the contamination by adjacent peaks. To determine this contamination as a function of the filling factor  $\zeta$ , we determine the fraction of the integrated flux on one side of the equidistance line ( $x = \lambda/2$ ) between both Gaussians that is contributed by the peak on the other side of that line. Due to symmetry, the total integrated flux on each side of the line must equal to the integrated flux of a single Gaussian. If we normalise each (two-dimensional) Gaussian to unity, the contamination fraction  $f_{\text{cont}}$  reduces to the integral of a single two-dimensional Gaussian with standard deviation  $r$  over  $x \geq \lambda/2$ . It is straightforward to show that  $f_{\text{cont}}$  then follows as

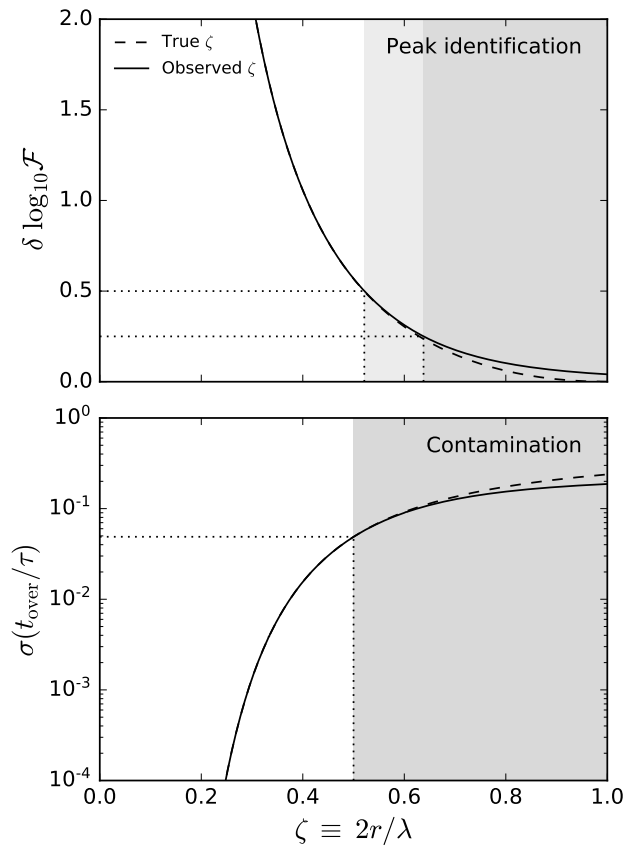
$$f_{\text{cont}} = \int_{\lambda/2}^{\infty} \frac{1}{\pi} \arccos\left(\frac{\lambda}{2x}\right) x e^{-x^2/2r^2} dx, \quad (\text{B3})$$

which is a one-dimensional function of  $\zeta \equiv 2r/\lambda$ .

Our method determines the duration of the overlap phase by measuring the correlation between both tracers in excess of the statistical correlation expected for a random distribution in the plane. This measurement assumes a linear relation between the population-integrated photon flux and the mean lifetime. Therefore, the amount by which blending affects the relative duration of the overlap phase  $t_{\text{over}}/\tau$  is equal to the contamination fraction from equation (B3). If we express this amount as a blending-induced uncertainty, we can thus define  $\sigma(t_{\text{over}}/\tau) \equiv f_{\text{cont}}$ . This uncertainty is shown as a function of  $\zeta$  in the bottom panel of Figure B1. Given that our method is able to determine  $t_{\text{over}}/\tau$  at an accuracy of  $\sim 5$  per cent (see Section 4.2.4), we require that  $\sigma(t_{\text{over}}/\tau) < 0.05$ . This condition is satisfied for  $\zeta < 0.50$ , which corresponds to the requirement that regions are separated by at least 1.69 times their FWHM. We impose this condition to guarantee the reliability of the measured values of  $t_{\text{over}}$ , resulting in an average blending-induced uncertainty of  $\sigma(t_{\text{over}}/\tau) < 0.05$ . Of course, it is possible to impose a higher upper limit on  $\zeta$ , as long as Figure B1 is used to quantify the associated blending-induced uncertainty on (or possible overestimation of) the duration of the overlap phase.

### APPENDIX C: THE PROBABILITY OF GAS REACHING HIGH DENSITIES PRIOR TO STAR FORMATION

As discussed in Section 4.3.3, the duration of the gas phase retrieved from the simulations is affected by the fact that some regions never reach the gas densities necessary for being shown in the gas map before being destroyed by feedback. Because star-forming regions in the real Universe do not form stars stochastically like our simulations do, this is not expected to influence observational applications of our method. However, for interpreting the results of



**Figure B1.** Effects of Gaussian blending on the identification of peaks and on the contamination of regions by neighbouring peaks. Top panel: logarithmic density contrast  $\delta \log_{10} \mathcal{F}$  between the peak flux and the minimum in between two adjacent Gaussian profiles with standard deviation  $r$  and separation  $\lambda$ , as a function of the filling factor  $\zeta \equiv 2r/\lambda$ . The dotted lines indicate the contrasts used to identify peaks in the simulated maps of Section 4. The grey-shaded areas indicate the range of filling factors where blending inhibits the identification of the peaks. Bottom panel: systematic uncertainty of the relative duration of the overlap phase due to blending as a function of the filling factor  $\zeta$ . This uncertainty is equal to the fraction of the integrated flux on one side of the equidistance line between two Gaussians ( $x = \lambda/2$ ) that is contributed by the peak on the other side of that line. The dotted line indicates the desired maximum level of such contamination (see the text) and the grey-shaded area indicates the range of filling factors where this maximum is exceeded. In both panels, the dashed line uses the true separation between both Gaussians ( $\lambda$ ), whereas the solid line accounts for the blending-induced shift of both peaks towards one another.

our experiments, it is important to derive the fraction of gas particles that is capable of reaching the minimum density shown in the gas map  $\rho_{\min}$  without being affected by star formation.

The formation of star particles within a given smoothing kernel in the simulation is a Poisson process with an occurrence rate  $\lambda_p$  set by

$$\lambda_p = \frac{\epsilon N_{\text{ngb}}}{t_{\text{dyn}}}, \quad (\text{C1})$$

where the factor  $N_{\text{ngb}} = 200$  represents the number of neighbours and accounts for the fact that any star formation event within the smoothing kernel will associate that kernel with star formation. For a constant dynamical time-scale, the probability of  $k$  particles being turned into stars after a time  $t$  is given by

$$p(k) = \frac{(\lambda_p t)^k}{k!} e^{-\lambda_p t}, \quad (\text{C2})$$

such that the probability of a gas concentration to be unaffected by star formation (i.e.  $k = 0$ ) for at least a time  $t$  is

$$p_{\text{gas}} = e^{-\lambda_p t}. \quad (\text{C3})$$

However, the dynamical time-scale on which the particles are spawned is time-dependent, because it is set by the kernel-averaged gas density. We therefore use a more general expression

$$p_{\text{gas}}(t) = e^{-\tilde{\lambda}_p(t)}, \quad (\text{C4})$$

where we define

$$\tilde{\lambda}_p(t) = \int_0^t \lambda_p(t') dt', \quad (\text{C5})$$

as the integral over a time-dependent occurrence rate  $\lambda[t_{\text{dyn}}(t)]$ .

While equation (C5) accounts for the time-dependence of  $\lambda_p(t)$ , we are not interested in obtaining  $p_{\text{gas}}$  at a given time, but at a given gas density. In order to obtain such an expression, we need to describe how the density changes with time. For this purpose, we define a dimensionless quantity

$$\xi_\rho \equiv \left( \frac{\rho_{\text{crit}}}{\rho_g} \right)^{1/3}, \quad (\text{C6})$$

that reflects the relative density increase beyond the star formation threshold  $\rho_{\text{crit}}$ . This will allow us to derive a relation  $\xi_\rho(t)$  and rewrite equation (C5) as

$$\tilde{\lambda}_p(\xi_\rho) = \int_1^{\xi_\rho} \lambda_p(\xi'_\rho) \frac{dt}{d\xi'_\rho} d\xi'_\rho, \quad (\text{C7})$$

so that we obtain  $p_{\text{gas}}(\xi_\rho)$  and, hence,  $p_{\text{gas}}(\rho_g)$ .

The key missing ingredient allowing us to derive an expression for  $p_{\text{gas}}(\xi_\rho)$  is the time-evolution of the gas density, or  $\xi_\rho(t)$ , so that we can specify its (inverse) time-derivative  $dt/d\xi_\rho$  in equation (C7). For this purpose, we describe the star-forming gas by the gravitational free-fall of a homogeneous sphere. Defining the initial and current radii of the sphere as  $R$  and  $r$ , respectively, and its mass as  $M$ , the textbook rate of collapse is

$$\frac{dr}{dt} = -\sqrt{2GM \left( \frac{1}{r} - \frac{1}{R} \right)}, \quad (\text{C8})$$

which can be rewritten as

$$\frac{dt}{d\xi_\rho} = -\left( \frac{3}{8\pi G \rho_{\text{crit}}} \right)^{1/2} \left( \frac{\xi_\rho}{1 - \xi_\rho} \right)^{1/2}, \quad (\text{C9})$$

where we have set the initial density of the homogeneous sphere

equal to the critical density for star formation in the simulations, i.e.  $\rho_{\text{crit}} = 3M/4\pi R^3$ .

We now combine equation (C9) with the density-dependent occurrence rate, which follows from equation (C1) as

$$\lambda_p(\xi_\rho) = \epsilon N_{\text{ngb}} (4\pi G \rho_{\text{crit}})^{1/2} \xi_\rho^{-3/2}, \quad (\text{C10})$$

upon which substitution into equation (C7) results in an integral that can be solved analytically (recall that  $0 < \xi_\rho \leq 1$ ):

$$\begin{aligned} \tilde{\lambda}_p(\xi_\rho) &= -\left( \frac{3}{2} \right)^{1/2} \epsilon N_{\text{ngb}} \int_1^{\xi_\rho} \left( \xi'_\rho{}^2 - \xi'_\rho{}^3 \right)^{-1/2} d\xi'_\rho \\ &= \left( \frac{3}{2} \right)^{1/2} \epsilon N_{\text{ngb}} \left[ 2 \tanh^{-1} \left( \sqrt{1 - \xi'_\rho} \right) \right]_{\xi'_\rho=1}^{\xi'_\rho=\xi_\rho} \\ &= \left( \frac{3}{2} \right)^{1/2} \epsilon N_{\text{ngb}} \ln \left( \frac{1 + \sqrt{1 - \xi_\rho}}{1 - \sqrt{1 - \xi_\rho}} \right). \end{aligned} \quad (\text{C11})$$

This provides the required expression of the integrated density-dependent occurrence rate. Substitution into equation (C4) then yields the probability that a gas concentration is unaffected by star formation until at least a density  $\rho_g$  has been reached as

$$p_{\text{gas}}(\xi_\rho) = \left( \frac{1 + \sqrt{1 - \xi_\rho}}{1 - \sqrt{1 - \xi_\rho}} \right)^{-\sqrt{3/2} \epsilon N_{\text{ngb}}}, \quad (\text{C12})$$

or, in terms of the volume density, as

$$p_{\text{gas}}(\rho_g) = \left[ \frac{1 + \sqrt{1 - (\rho_{\text{crit}}/\rho_g)^{1/3}}}{1 - \sqrt{1 - (\rho_{\text{crit}}/\rho_g)^{1/3}}} \right]^{-\sqrt{3/2} \epsilon N_{\text{ngb}}}, \quad (\text{C13})$$

which is valid for  $\rho_g \geq \rho_{\text{crit}}$ . For  $\rho_g < \rho_{\text{crit}}$ , we set  $p_{\text{gas}} = 1$ .

## APPENDIX D: COMPLETE RESULTS OF THE PERFORMED EXPERIMENTS

Here we tabulate the best-fitting solutions from all experiments discussed in Section 4, including some of the derived quantities.

### D1 Age-binned stellar maps

Section 4.2 discusses 88 different experiments using the star particles in the numerical simulations described in Section 4.1. These are constituted by a  $2 \times 2$  matrix of the 22 experiments listed in Table 6, using maps of {point particles, extended emission} in the {low, high} resolution galaxy simulations. The output is listed in Table D1–D4, which represent the low-resolution point particle experiments, the high-resolution point particle experiments, the low-resolution extended emission experiments, and the high-resolution extended emission experiments, respectively.

### D2 Gas and stellar maps

Section 4.3 discusses 200 different experiments using the gas and star particles in the numerical simulations described in Section 4.1. These are constituted by a  $2 \times 2$  matrix of the 50 experiments listed in Table 8, using maps of {point particles, extended emission} in the {low, high} resolution galaxy simulations. The output is listed in Table D5–D8, which represent the low-resolution point particle experiments, the high-resolution point particle experiments, the low-resolution extended emission experiments, and the high-resolution extended emission experiments, respectively.

This paper has been typeset from a  $\text{\TeX}/\text{\LaTeX}$  file prepared by the author.



**Table D3.** Best-fitting solutions for age-binned stellar maps (low resolution, extended emission)

ID	$N_{\text{star}}$	$N_{\text{gas}}$	$t_{\text{gas}}$	$t_{\text{over}}$	$\lambda$	$\beta_{\text{star}}$	$\beta_{\text{gas}}$	$\zeta_{\text{star}}$	$\zeta_{\text{gas}}$	$\chi_{\text{red}}^2$	Symbol
1	49	43	2.44 <sup>+1.15</sup> <sub>-0.68</sub>	0.18 <sup>+0.16</sup> <sub>-0.07</sub>	908 <sup>+225</sup> <sub>-129</sub>	1.35 <sup>+0.16</sup> <sub>-0.17</sub>	0.89 <sup>+0.12</sup> <sub>-0.22</sub>	0.30 <sup>+0.04</sup> <sub>-0.05</sub>	0.31 <sup>+0.04</sup> <sub>-0.04</sub>	0.36	■
2	87	87	20.31 <sup>+5.75</sup> <sub>-3.83</sub>	2.04 <sup>+0.76</sup> <sub>-0.71</sub>	533 <sup>+80</sup> <sub>-81</sub>	0.58 <sup>+0.02</sup> <sub>-0.02</sub>	1.58 <sup>+0.10</sup> <sub>-0.12</sub>	0.44 <sup>+0.04</sup> <sub>-0.03</sub>	0.44 <sup>+0.04</sup> <sub>-0.03</sub>	0.76	■
3	117	105	16.41 <sup>+7.49</sup> <sub>-3.15</sub>	4.08 <sup>+2.42</sup> <sub>-0.77</sub>	700 <sup>+246</sup> <sub>-95</sub>	1.94 <sup>+0.28</sup> <sub>-0.44</sub>	1.62 <sup>+0.13</sup> <sub>-0.04</sub>	0.44 <sup>+0.02</sup> <sub>-0.05</sub>	0.44 <sup>+0.02</sup> <sub>-0.06</sub>	0.10	■
4	145	114	180.02 <sup>+54.06</sup> <sub>-43.37</sub>	46.33 <sup>+9.52</sup> <sub>-8.42</sub>	547 <sup>+199</sup> <sub>-130</sub>	1.23 <sup>+0.15</sup> <sub>-0.03</sub>	1.95 <sup>+0.08</sup> <sub>-0.13</sub>	0.49 <sup>+0.06</sup> <sub>-0.06</sub>	0.47 <sup>+0.05</sup> <sub>-0.04</sub>	0.65	■
5	22	43	2.37 <sup>+1.35</sup> <sub>-0.65</sub>	0.12 <sup>+0.07</sup> <sub>-0.05</sub>	1273 <sup>+303</sup> <sub>-267</sub>	0.92 <sup>+0.09</sup> <sub>-0.04</sub>	0.64 <sup>+0.04</sup> <sub>-0.01</sub>	0.26 <sup>+0.07</sup> <sub>-0.05</sub>	0.22 <sup>+0.04</sup> <sub>-0.03</sub>	0.16	◀
6	46	15	0.54 <sup>+0.48</sup> <sub>-0.13</sub>	0.16 <sup>+0.09</sup> <sub>-0.04</sub>	1292 <sup>+647</sup> <sub>-253</sub>	0.90 <sup>+0.06</sup> <sub>-0.05</sub>	0.87 <sup>+0.11</sup> <sub>-0.01</sub>	0.22 <sup>+0.04</sup> <sub>-0.05</sub>	0.26 <sup>+0.06</sup> <sub>-0.08</sub>	0.28	▶
7	94	105	12.67 <sup>+3.92</sup> <sub>-2.66</sub>	1.88 <sup>+0.85</sup> <sub>-0.43</sub>	773 <sup>+268</sup> <sub>-161</sub>	1.50 <sup>+0.20</sup> <sub>-0.25</sub>	1.66 <sup>+0.62</sup> <sub>-0.91</sub>	0.39 <sup>+0.05</sup> <sub>-0.06</sub>	0.41 <sup>+0.04</sup> <sub>-0.06</sub>	0.19	◀
8	111	87	15.81 <sup>+5.46</sup> <sub>-2.06</sub>	4.88 <sup>+1.75</sup> <sub>-0.75</sub>	616 <sup>+138</sup> <sub>-81</sub>	0.92 <sup>+0.01</sup> <sub>-0.06</sub>	0.95 <sup>+0.04</sup> <sub>-0.01</sub>	0.42 <sup>+0.03</sup> <sub>-0.03</sub>	0.43 <sup>+0.03</sup> <sub>-0.04</sub>	0.43	▶
9	11	87	17.60 <sup>+5.42</sup> <sub>-6.14</sub>	0.63 <sup>+0.11</sup> <sub>-0.19</sub>	875 <sup>+180</sup> <sub>-317</sub>	0.98 <sup>+0.01</sup> <sub>-0.25</sub>	0.82 <sup>+0.07</sup> <sub>-0.06</sub>	0.39 <sup>+0.07</sup> <sub>-0.09</sub>	0.31 <sup>+0.05</sup> <sub>-0.06</sub>	0.18	▼
10	81	15	1.39 <sup>+1.45</sup> <sub>-0.34</sub>	0.36 <sup>+0.33</sup> <sub>-0.08</sub>	1078 <sup>+594</sup> <sub>-179</sub>	0.91 <sup>+0.01</sup> <sub>-0.10</sub>	1.06 <sup>+0.01</sup> <sub>-0.13</sub>	0.28 <sup>+0.03</sup> <sub>-0.06</sub>	0.31 <sup>+0.06</sup> <sub>-0.11</sub>	0.39	▲
11	50	105	8.31 <sup>+4.17</sup> <sub>-2.55</sub>	0.88 <sup>+0.25</sup> <sub>-0.35</sub>	977 <sup>+202</sup> <sub>-240</sub>	1.38 <sup>+0.65</sup> <sub>-0.08</sub>	1.16 <sup>+1.02</sup> <sub>-0.24</sub>	0.32 <sup>+0.08</sup> <sub>-0.04</sub>	0.34 <sup>+0.05</sup> <sub>-0.03</sub>	0.23	▼
12	101	43	4.93 <sup>+1.78</sup> <sub>-0.89</sub>	1.62 <sup>+0.47</sup> <sub>-0.23</sub>	574 <sup>+193</sup> <sub>-82</sub>	0.59 <sup>+0.21</sup> <sub>-0.25</sub>	0.79 <sup>+0.03</sup> <sub>-0.02</sub>	0.42 <sup>+0.03</sup> <sub>-0.04</sub>	0.45 <sup>+0.04</sup> <sub>-0.06</sub>	0.42	▲
13	85	87	21.49 <sup>+2.87</sup> <sub>-3.76</sub>	3.39 <sup>+0.30</sup> <sub>-0.61</sub>	564 <sup>+116</sup> <sub>-94</sub>	0.68 <sup>+0.05</sup> <sub>-0.04</sub>	1.28 <sup>+0.14</sup> <sub>-0.06</sub>	0.44 <sup>+0.04</sup> <sub>-0.04</sub>	0.42 <sup>+0.04</sup> <sub>-0.04</sub>	0.54	■
14	81	87	13.49 <sup>+2.70</sup> <sub>-2.20</sub>	5.53 <sup>+0.24</sup> <sub>-0.71</sub>	569 <sup>+173</sup> <sub>-146</sub>	0.68 <sup>+0.01</sup> <sub>-0.02</sub>	1.10 <sup>+0.17</sup> <sub>-0.07</sub>	0.42 <sup>+0.07</sup> <sub>-0.05</sub>	0.42 <sup>+0.08</sup> <sub>-0.05</sub>	0.38	■
15	77	87	11.39 <sup>+1.97</sup> <sub>-0.85</sub>	8.40 <sup>+0.82</sup> <sub>-0.08</sub>	931 <sup>+722</sup> <sub>-423</sub>	1.44 <sup>+0.42</sup> <sub>-0.23</sub>	1.33 <sup>+0.03</sup> <sub>-0.11</sub>	0.29 <sup>+0.12</sup> <sub>-0.09</sub>	0.31 <sup>+0.13</sup> <sub>-0.10</sub>	0.17	■
16	98	105	13.24 <sup>+2.78</sup> <sub>-1.58</sub>	3.92 <sup>+0.51</sup> <sub>-0.62</sub>	848 <sup>+213</sup> <sub>-213</sub>	1.13 <sup>+0.05</sup> <sub>-0.06</sub>	1.09 <sup>+0.99</sup> <sub>-0.17</sub>	0.38 <sup>+0.06</sup> <sub>-0.05</sub>	0.38 <sup>+0.05</sup> <sub>-0.04</sub>	0.13	◀
17	114	87	17.31 <sup>+3.52</sup> <sub>-2.57</sub>	8.05 <sup>+0.73</sup> <sub>-1.28</sub>	674 <sup>+130</sup> <sub>-137</sub>	1.02 <sup>+0.11</sup> <sub>-0.12</sub>	1.04 <sup>+0.14</sup> <sub>-0.04</sub>	0.41 <sup>+0.04</sup> <sub>-0.03</sub>	0.41 <sup>+0.05</sup> <sub>-0.04</sub>	0.25	▶
18	82	87	15.18 <sup>+2.63</sup> <sub>-3.18</sub>	5.20 <sup>+0.26</sup> <sub>-0.72</sub>	612 <sup>+227</sup> <sub>-174</sub>	0.55 <sup>+0.12</sup> <sub>-0.04</sub>	1.28 <sup>+0.06</sup> <sub>-0.10</sub>	0.40 <sup>+0.07</sup> <sub>-0.06</sub>	0.40 <sup>+0.08</sup> <sub>-0.06</sub>	0.48	■
19	81	84	13.06 <sup>+1.84</sup> <sub>-1.90</sub>	4.70 <sup>+0.34</sup> <sub>-0.78</sub>	574 <sup>+163</sup> <sub>-139</sub>	0.69 <sup>+0.52</sup> <sub>-0.04</sub>	0.83 <sup>+0.03</sup> <sub>-0.02</sub>	0.42 <sup>+0.06</sup> <sub>-0.05</sub>	0.41 <sup>+0.07</sup> <sub>-0.05</sub>	0.25	■
20	82	84	14.36 <sup>+2.83</sup> <sub>-2.21</sub>	4.41 <sup>+0.50</sup> <sub>-0.65</sub>	574 <sup>+164</sup> <sub>-136</sub>	0.54 <sup>+0.05</sup> <sub>-0.05</sub>	0.91 <sup>+0.04</sup> <sub>-0.03</sub>	0.41 <sup>+0.06</sup> <sub>-0.05</sub>	0.41 <sup>+0.07</sup> <sub>-0.05</sub>	0.38	■
21	110	141	16.03 <sup>+2.60</sup> <sub>-2.61</sub>	5.55 <sup>+0.68</sup> <sub>-0.68</sub>	776 <sup>+207</sup> <sub>-178</sub>	1.03 <sup>+0.08</sup> <sub>-0.10</sub>	0.69 <sup>+0.03</sup> <sub>-0.10</sub>	0.39 <sup>+0.06</sup> <sub>-0.06</sub>	0.34 <sup>+0.05</sup> <sub>-0.04</sub>	0.19	◀
22	145	48	8.68 <sup>+11.11</sup> <sub>-2.41</sub>	1.91 <sup>+0.37</sup> <sub>-0.32</sub>	1119 <sup>+507</sup> <sub>-233</sub>	1.83 <sup>+2.47</sup> <sub>-1.15</sub>	2.45 <sup>+2.18</sup> <sub>-0.41</sub>	0.29 <sup>+0.04</sup> <sub>-0.04</sub>	0.27 <sup>+0.05</sup> <sub>-0.06</sub>	0.18	▲

All time-scales are in units of Myr. All length-scales are in units of pc.

**Table D4.** Best-fitting solutions for age-binned stellar maps (high resolution, extended emission)

ID	$N_{\text{star}}$	$N_{\text{gas}}$	$t_{\text{gas}}$	$t_{\text{over}}$	$\lambda$	$\beta_{\text{star}}$	$\beta_{\text{gas}}$	$\zeta_{\text{star}}$	$\zeta_{\text{gas}}$	$\chi_{\text{red}}^2$	Symbol
1	248	230	3.31 <sup>+0.53</sup> <sub>-0.38</sub>	0.10 <sup>+0.07</sup> <sub>-0.04</sub>	455 <sup>+36</sup> <sub>-28</sub>	0.89 <sup>+0.00</sup> <sub>-0.08</sub>	0.95 <sup>+0.14</sup> <sub>-0.12</sub>	0.31 <sup>+0.02</sup> <sub>-0.02</sub>	0.30 <sup>+0.01</sup> <sub>-0.02</sub>	0.22	■
2	447	435	8.42 <sup>+0.73</sup> <sub>-0.70</sub>	0.32 <sup>+0.13</sup> <sub>-0.17</sub>	283 <sup>+15</sup> <sub>-14</sub>	0.64 <sup>+0.09</sup> <sub>-0.17</sub>	0.69 <sup>+0.05</sup> <sub>-0.09</sub>	0.44 <sup>+0.01</sup> <sub>-0.01</sub>	0.44 <sup>+0.01</sup> <sub>-0.01</sub>	0.47	■
3	567	514	23.37 <sup>+1.70</sup> <sub>-1.49</sub>	4.87 <sup>+0.58</sup> <sub>-0.62</sub>	253 <sup>+19</sup> <sub>-19</sub>	0.85 <sup>+0.03</sup> <sub>-0.05</sub>	0.85 <sup>+0.02</sup> <sub>-0.03</sub>	0.49 <sup>+0.02</sup> <sub>-0.02</sub>	0.51 <sup>+0.02</sup> <sub>-0.01</sub>	0.21	■
4	454	396	123.83 <sup>+7.34</sup> <sub>-7.22</sub>	39.27 <sup>+3.13</sup> <sub>-2.74</sub>	198 <sup>+19</sup> <sub>-16</sub>	0.83 <sup>+0.02</sup> <sub>-0.01</sub>	0.98 <sup>+0.03</sup> <sub>-0.03</sub>	0.59 <sup>+0.01</sup> <sub>-0.02</sub>	0.58 <sup>+0.01</sup> <sub>-0.02</sub>	0.56	■
5	114	230	3.00 <sup>+0.71</sup> <sub>-0.33</sub>	0.07 <sup>+0.03</sup> <sub>-0.03</sub>	593 <sup>+55</sup> <sub>-64</sub>	1.13 <sup>+0.17</sup> <sub>-0.15</sub>	0.79 <sup>+0.09</sup> <sub>-0.02</sub>	0.27 <sup>+0.03</sup> <sub>-0.02</sub>	0.23 <sup>+0.02</sup> <sub>-0.01</sub>	0.34	◀
6	250	83	0.99 <sup>+0.35</sup> <sub>-0.13</sub>	0.01 <sup>+0.02</sup> <sub>-0.00</sub>	588 <sup>+25</sup> <sub>-27</sub>	0.46 <sup>+0.22</sup> <sub>-0.01</sub>	0.77 <sup>+0.09</sup> <sub>-0.00</sub>	0.24 <sup>+0.02</sup> <sub>-0.02</sub>	0.25 <sup>+0.01</sup> <sub>-0.02</sub>	0.40	▶
7	449	514	20.11 <sup>+1.87</sup> <sub>-1.65</sub>	1.73 <sup>+0.16</sup> <sub>-0.30</sub>	298 <sup>+23</sup> <sub>-26</sub>	0.87 <sup>+0.01</sup> <sub>-0.02</sub>	0.84 <sup>+0.26</sup> <sub>-0.13</sub>	0.44 <sup>+0.02</sup> <sub>-0.02</sub>	0.46 <sup>+0.02</sup> <sub>-0.02</sub>	0.31	◀
8	527	435	14.65 <sup>+1.47</sup> <sub>-1.16</sub>	1.38 <sup>+0.38</sup> <sub>-0.34</sub>	267 <sup>+19</sup> <sub>-18</sub>	1.02 <sup>+0.05</sup> <sub>-0.22</sub>	0.92 <sup>+0.01</sup> <sub>-0.03</sub>	0.47 <sup>+0.02</sup> <sub>-0.01</sub>	0.47 <sup>+0.02</sup> <sub>-0.02</sub>	0.43	▶
9	99	435	8.07 <sup>+1.34</sup> <sub>-1.31</sub>	0.08 <sup>+0.03</sup> <sub>-0.03</sub>	398 <sup>+47</sup> <sub>-35</sub>	0.79 <sup>+0.12</sup> <sub>-0.14</sub>	0.82 <sup>+0.08</sup> <sub>-0.01</sub>	0.38 <sup>+0.02</sup> <sub>-0.04</sub>	0.33 <sup>+0.02</sup> <sub>-0.02</sub>	0.39	▼
10	432	83	1.39 <sup>+0.34</sup> <sub>-0.20</sub>	0.08 <sup>+0.05</sup> <sub>-0.04</sub>	426 <sup>+53</sup> <sub>-33</sub>	0.52 <sup>+0.29</sup> <sub>-0.07</sub>	0.63 <sup>+0.18</sup> <sub>-0.02</sub>	0.32 <sup>+0.02</sup> <sub>-0.02</sub>	0.34 <sup>+0.03</sup> <sub>-0.04</sub>	0.36	▲
11	201	514	17.79 <sup>+2.72</sup> <sub>-1.68</sub>	0.40 <sup>+0.09</sup> <sub>-0.09</sub>	336 <sup>+24</sup> <sub>-30</sub>	0.98 <sup>+0.04</sup> <sub>-0.15</sub>	0.71 <sup>+0.06</sup> <sub>-0.02</sub>	0.43 <sup>+0.02</sup> <sub>-0.02</sub>	0.43 <sup>+0.02</sup> <sub>-0.01</sub>	0.24	▼
12	524	230	5.97 <sup>+0.80</sup> <sub>-0.61</sub>	0.60 <sup>+0.19</sup> <sub>-0.14</sub>	290 <sup>+24</sup> <sub>-17</sub>	0.82 <sup>+0.10</sup> <sub>-0.30</sub>	0.80 <sup>+0.03</sup> <sub>-0.15</sub>	0.44 <sup>+0.01</sup> <sub>-0.01</sub>	0.44 <sup>+0.02</sup> <sub>-0.02</sub>	0.50	▲
13	429	435	10.02 <sup>+0.86</sup> <sub>-0.66</sub>	1.62 <sup>+0.19</sup> <sub>-0.21</sub>	294 <sup>+25</sup> <sub>-20</sub>	0.93 <sup>+0.02</sup> <sub>-0.00</sub>	0.84 <sup>+0.06</sup> <sub>-0.02</sub>	0.44 <sup>+0.02</sup> <sub>-0.02</sub>	0.42 <sup>+0.02</sup> <sub>-0.02</sub>	0.53	■
14	423	435	9.98 <sup>+0.68</sup> <sub>-0.58</sub>	2.92 <sup>+0.20</sup> <sub>-0.18</sub>	312 <sup>+29</sup> <sub>-25</sub>	1.11 <sup>+0.02</sup> <sub>-0.03</sub>	0.96 <sup>+0.02</sup> <sub>-0.03</sub>	0.42 <sup>+0.02</sup> <sub>-0.02</sub>	0.41 <sup>+0.02</sup> <sub>-0.02</sub>	0.44	■
15	433	435	10.25 <sup>+0.27</sup> <sub>-0.30</sub>	7.92 <sup>+0.15</sup> <sub>-0.16</sub>	323 <sup>+86</sup> <sub>-55</sub>	1.11 <sup>+0.05</sup> <sub>-0.06</sub>	0.88 <sup>+0.03</sup> <sub>-0.03</sub>	0.39 <sup>+0.02</sup> <sub>-0.05</sub>	0.38 <sup>+0.01</sup> <sub>-0.05</sub>	0.11	■
16	432	514	26.49 <sup>+2.49</sup> <sub>-2.13</sub>	4.25 <sup>+0.36</sup> <sub>-0.24</sub>	287 <sup>+29</sup> <sub>-26</sub>	1.11 <sup>+0.02</sup> <sub>-0.02</sub>	1.05 <sup>+0.02</sup> <sub>-0.03</sub>	0.46 <sup>+0.02</sup> <sub>-0.02</sub>	0.46 <sup>+0.02</sup> <sub>-0.02</sub>	0.37	◀
17	503	435	14.95 <sup>+1.27</sup> <sub>-1.19</sub>	4.41 <sup>+0.36</sup> <sub>-0.43</sub>	274 <sup>+22</sup> <sub>-19</sub>	1.01 <sup>+0.04</sup> <sub>-0.01</sub>	0.91 <sup>+0.02</sup> <sub>-0.03</sub>	0.46 <sup>+0.02</sup> <sub>-0.02</sub>	0.46 <sup>+0.02</sup> <sub>-0.02</sub>	0.39	▶
18	424	435	10.97 <sup>+0.98</sup> <sub>-0.82</sub>	2.18 <sup>+0.23</sup> <sub>-0.15</sub>	301 <sup>+26</sup> <sub>-20</sub>	0.81 <sup>+0.01</sup> <sub>-0.03</sub>	0.90 <sup>+0.01</sup> <sub>-0.01</sub>	0.43 <sup>+0.02</sup> <sub>-0.02</sub>	0.41 <sup>+0.02</sup> <sub>-0.02</sub>	0.59	■
19	423	423	9.23 <sup>+0.75</sup> <sub>-0.72</sub>	2.22 <sup>+0.24</sup> <sub>-0.23</sub>	309 <sup>+25</sup> <sub>-19</sub>	0.95 <sup>+0.04</sup> <sub>-0.01</sub>	0.68 <sup>+0.02</sup> <sub>-0.02</sub>	0.42 <sup>+0.02</sup> <sub>-0.02</sub>	0.41 <sup>+0.02</sup> <sub>-0.02</sub>	0.37	■
20	424	423	9.68 <sup>+0.71</sup> <sub>-0.59</sub>	1.82 <sup>+0.20</sup> <sub>-0.21</sub>	307 <sup>+25</sup> <sub>-25</sub>	0.66 <sup>+0.05</sup> <sub>-0.05</sub>	0.56 <sup>+0.06</sup> <sub>-0.05</sub>	0.42 <sup>+0.02</sup> <sub>-0.02</sub>	0.42 <sup>+0.02</sup> <sub>-0.02</sub>	0.42	■
21	607	675	17.92 <sup>+1.10</sup> <sub>-0.97</sub>	4.98 <sup>+0.23</sup> <sub>-0.16</sub>	281 <sup>+24</sup> <sub>-22</sub>	0.96 <sup>+0.02</sup> <sub>-0.05</sub>	0.85 <sup>+0.01</sup> <sub>-0.02</sub>	0.45 <sup>+0.02</sup> <sub>-0.02</sub>	0.46 <sup>+0.02</sup> <sub>-0.02</sub>	0.58	◀
22	651	276	4.65 <sup>+0.57</sup> <sub>-0.40</sub>	0.54 <sup>+0.13</sup> <sub>-0.11</sub>	296 <sup>+21</sup> <sub>-16</sub>	0.98 <sup>+0.03</sup> <sub>-0.11</sub>	0.77 <sup>+0.05</sup> <sub>-0.02</sub>	0.45 <sup>+0.01</sup> <sub>-0.01</sub>	0.45 <sup>+0.02</sup> <sub>-0.02</sub>	0.29	▲

All time-scales are in units of Myr. All length-scales are in units of pc.





**Table D6.** Best-fitting solutions for gas maps (high resolution, point particles)

ID	$N_{\text{star}}$	$N_{\text{gas}}$	$t_{\text{gas}}$	$t_{\text{over}}$	$\lambda$	$\beta_{\text{star}}$	$\beta_{\text{gas}}$	$\zeta_{\text{star}}$	$\zeta_{\text{gas}}$	$\chi^2_{\text{red}}$
1	61	462	$12.37^{+9.18}_{-3.03}$	$0.15^{+0.08}_{-0.05}$	$263^{+51}_{-72}$	$0.99^{+0.02}_{-0.00}$	$1.05^{+0.23}_{-0.43}$	$0.18^{+0.07}_{-0.03}$	$0.30^{+0.07}_{-0.03}$	0.60
2	125	462	$8.61^{+3.25}_{-1.52}$	$0.13^{+0.06}_{-0.05}$	$261^{+30}_{-41}$	$1.02^{+0.01}_{-0.01}$	$0.70^{+0.26}_{-0.48}$	$0.18^{+0.03}_{-0.02}$	$0.31^{+0.03}_{-0.02}$	0.16
3	217	462	$8.71^{+2.30}_{-1.27}$	$0.22^{+0.07}_{-0.08}$	$248^{+24}_{-30}$	$1.01^{+0.00}_{-0.01}$	$0.68^{+0.19}_{-0.41}$	$0.18^{+0.03}_{-0.02}$	$0.33^{+0.03}_{-0.02}$	0.32
4	448	462	$8.17^{+1.45}_{-0.99}$	$0.14^{+0.09}_{-0.07}$	$210^{+12}_{-11}$	$1.01^{+0.01}_{-0.01}$	$0.16^{+0.28}_{-0.03}$	$0.21^{+0.01}_{-0.01}$	$0.38^{+0.01}_{-0.01}$	0.42
5	622	462	$10.45^{+1.40}_{-1.34}$	$0.09^{+0.05}_{-0.05}$	$191^{+6}_{-6}$	$1.01^{+0.05}_{-0.05}$	$0.13^{+0.04}_{-0.04}$	$0.23^{+0.01}_{-0.01}$	$0.41^{+0.01}_{-0.01}$	0.44
6	1131	462	$9.51^{+0.83}_{-1.11}$	$0.01^{+0.12}_{-0.08}$	$185^{+6}_{-4}$	$1.08^{+0.01}_{-0.04}$	$0.26^{+0.09}_{-0.05}$	$0.23^{+0.00}_{-0.01}$	$0.43^{+0.01}_{-0.01}$	0.66
7	1603	462	$11.17^{+1.03}_{-1.05}$	$0.14^{+0.16}_{-0.11}$	$174^{+5}_{-4}$	$1.26^{+0.09}_{-0.11}$	$0.17^{+0.06}_{-0.00}$	$0.25^{+0.01}_{-0.01}$	$0.45^{+0.01}_{-0.01}$	0.83
8	2860	462	$11.32^{+0.90}_{-0.74}$	$0.71^{+0.29}_{-0.24}$	$152^{+6}_{-4}$	$0.98^{+0.03}_{-0.00}$	$0.30^{+0.20}_{-0.08}$	$0.29^{+0.01}_{-0.01}$	$0.49^{+0.01}_{-0.01}$	1.45
9	4070	462	$13.40^{+0.88}_{-0.74}$	$2.23^{+0.33}_{-0.40}$	$139^{+5}_{-5}$	$0.96^{+0.02}_{-0.03}$	$0.55^{+0.06}_{-0.06}$	$0.32^{+0.01}_{-0.01}$	$0.52^{+0.01}_{-0.01}$	2.79
10	5250	462	$20.21^{+1.16}_{-1.09}$	$5.42^{+0.57}_{-0.70}$	$132^{+5}_{-6}$	$1.04^{+0.01}_{-0.01}$	$0.59^{+0.02}_{-0.06}$	$0.36^{+0.01}_{-0.01}$	$0.53^{+0.01}_{-0.01}$	3.80
11	6021	462	$29.49^{+1.81}_{-1.53}$	$9.34^{+1.03}_{-1.04}$	$125^{+6}_{-6}$	$1.14^{+0.03}_{-0.03}$	$0.50^{+0.03}_{-0.05}$	$0.41^{+0.01}_{-0.01}$	$0.55^{+0.01}_{-0.01}$	4.20
12	217	1037	$183.35^{+191.23}_{-39.14}$	$0.97^{+1.58}_{-0.41}$	$81^{+7}_{-15}$	$1.01^{+0.02}_{-0.01}$	$0.52^{+0.18}_{-0.05}$	$0.54^{+0.07}_{-0.03}$	$0.61^{+0.04}_{-0.01}$	0.62
13	622	1037	$74.85^{+14.60}_{-10.95}$	$0.01^{+0.20}_{-0.00}$	$122^{+9}_{-9}$	$0.99^{+0.03}_{-0.00}$	$0.27^{+0.08}_{-0.00}$	$0.39^{+0.03}_{-0.03}$	$0.53^{+0.02}_{-0.01}$	3.65
14	1603	1037	$52.66^{+4.57}_{-4.82}$	$0.01^{+0.32}_{-0.00}$	$146^{+5}_{-4}$	$0.03^{+0.00}_{-0.00}$	$0.27^{+0.19}_{-0.00}$	$0.32^{+0.01}_{-0.01}$	$0.51^{+0.00}_{-0.01}$	6.46
15	217	108	$0.94^{+0.38}_{-0.17}$	$0.09^{+0.04}_{-0.02}$	$385^{+87}_{-52}$	$1.05^{+0.00}_{-0.02}$	$0.91^{+0.31}_{-0.11}$	$0.10^{+0.02}_{-0.01}$	$0.20^{+0.03}_{-0.03}$	0.62
16	622	108	$1.35^{+0.42}_{-0.24}$	$0.07^{+0.05}_{-0.03}$	$203^{+24}_{-15}$	$1.12^{+0.04}_{-0.00}$	$0.38^{+0.04}_{-0.04}$	$0.19^{+0.01}_{-0.01}$	$0.35^{+0.02}_{-0.03}$	0.70
17	1603	108	$1.73^{+0.47}_{-0.32}$	$0.11^{+0.06}_{-0.05}$	$159^{+17}_{-13}$	$1.50^{+0.16}_{-0.08}$	$0.30^{+0.02}_{-0.01}$	$0.25^{+0.02}_{-0.02}$	$0.42^{+0.03}_{-0.03}$	0.68
18	222	1667	$18.14^{+5.32}_{-3.39}$	$0.05^{+0.04}_{-0.03}$	$162^{+16}_{-17}$	$1.01^{+0.02}_{-0.01}$	$0.52^{+0.09}_{-0.14}$	$0.14^{+0.02}_{-0.01}$	$0.31^{+0.01}_{-0.01}$	3.62
19	696	1667	$15.34^{+3.82}_{-1.73}$	$0.01^{+0.02}_{-0.00}$	$173^{+6}_{-5}$	$1.02^{+0.02}_{-0.00}$	$0.47^{+0.10}_{-0.00}$	$0.12^{+0.00}_{-0.00}$	$0.32^{+0.00}_{-0.00}$	4.66
20	2013	1667	$16.42^{+1.15}_{-1.41}$	$0.01^{+0.04}_{-0.00}$	$158^{+4}_{-3}$	$1.12^{+0.22}_{-0.00}$	$0.54^{+0.13}_{-0.04}$	$0.13^{+0.00}_{-0.00}$	$0.35^{+0.00}_{-0.00}$	7.49
21	201	286	$7.60^{+2.32}_{-1.21}$	$0.26^{+0.19}_{-0.15}$	$261^{+35}_{-31}$	$1.09^{+0.01}_{-0.07}$	$0.32^{+0.21}_{-0.20}$	$0.35^{+0.05}_{-0.04}$	$0.44^{+0.04}_{-0.03}$	0.04
22	489	286	$11.02^{+1.29}_{-1.88}$	$0.04^{+0.38}_{-0.02}$	$206^{+15}_{-6}$	$0.47^{+0.51}_{-0.02}$	$0.08^{+0.03}_{-0.00}$	$0.44^{+0.01}_{-0.03}$	$0.54^{+0.01}_{-0.02}$	0.83
23	905	286	$12.55^{+1.37}_{-1.75}$	$0.01^{+0.34}_{-0.00}$	$196^{+11}_{-8}$	$0.82^{+0.07}_{-0.00}$	$0.02^{+0.07}_{-0.00}$	$0.46^{+0.01}_{-0.02}$	$0.57^{+0.01}_{-0.02}$	1.80
24	169	172	$6.12^{+1.92}_{-1.07}$	$0.29^{+0.41}_{-0.18}$	$301^{+70}_{-35}$	$1.01^{+0.07}_{-0.11}$	$0.14^{+0.50}_{-0.08}$	$0.54^{+0.04}_{-0.07}$	$0.58^{+0.03}_{-0.05}$	0.21
25	264	172	$6.35^{+1.26}_{-0.71}$	$1.48^{+0.53}_{-0.72}$	$258^{+40}_{-34}$	$1.18^{+0.09}_{-0.14}$	$0.31^{+0.15}_{-0.22}$	$0.59^{+0.06}_{-0.05}$	$0.63^{+0.04}_{-0.03}$	0.37
26	321	172	$8.19^{+1.99}_{-1.09}$	$2.57^{+0.70}_{-1.00}$	$226^{+26}_{-21}$	$1.69^{+0.08}_{-0.07}$	$0.22^{+0.10}_{-0.13}$	$0.66^{+0.03}_{-0.03}$	$0.67^{+0.03}_{-0.03}$	0.39
27	94	65	$4.17^{+4.07}_{-0.86}$	$1.46^{+1.99}_{-0.77}$	$447^{+275}_{-151}$	$1.17^{+0.31}_{-0.14}$	$0.38^{+0.19}_{-0.11}$	$0.65^{+0.06}_{-0.14}$	$0.66^{+0.04}_{-0.13}$	0.26
28	94	65	$5.37^{+233.79}_{-1.22}$	$4.35^{+182.55}_{-1.58}$	$402^{+9935}_{-143}$	$1.43^{+1.29}_{-0.05}$	$0.42^{+0.04}_{-0.01}$	$0.69^{+0.02}_{-0.42}$	$0.70^{+0.01}_{-0.47}$	0.28
29	35	65	$6.52^{+244.80}_{-1.70}$	$4.69^{+182.89}_{-3.02}$	$303^{+1691}_{-64}$	$3.65^{+2.99}_{-1.04}$	$0.29^{+0.03}_{-0.01}$	$0.71^{+0.00}_{-0.28}$	$0.71^{+0.00}_{-0.41}$	0.60
30	17	20	$22.22^{+171.97}_{-9.37}$	$9.01^{+92.86}_{-4.66}$	$2255^{+2037}_{-1260}$	$3.26^{+0.38}_{-2.43}$	$0.46^{+0.01}_{-0.09}$	$0.46^{+0.15}_{-0.04}$	$0.39^{+0.23}_{-0.09}$	0.08
31	7	20	$244.49^{+45.77}_{-105.33}$	$143.26^{+69.35}_{-85.76}$	$3063^{+4255}_{-1969}$	$1.54^{+1.18}_{-0.00}$	$0.35^{+0.08}_{-0.15}$	$0.40^{+0.06}_{-0.01}$	$0.33^{+0.25}_{-0.06}$	0.01
32	10	20	$8.63^{+236.94}_{-2.31}$	$8.57^{+193.38}_{-5.60}$	$580^{+1664}_{-236}$	$3.48^{+0.12}_{-1.17}$	$0.27^{+0.05}_{-0.00}$	$0.71^{+0.01}_{-0.17}$	$0.71^{+0.00}_{-0.33}$	0.62
33	211	461	$9.47^{+2.40}_{-1.53}$	$0.22^{+0.08}_{-0.07}$	$250^{+25}_{-29}$	$0.99^{+0.05}_{-0.01}$	$0.75^{+0.15}_{-0.28}$	$0.18^{+0.02}_{-0.02}$	$0.33^{+0.03}_{-0.02}$	0.24
34	613	461	$10.64^{+1.60}_{-1.15}$	$0.18^{+0.10}_{-0.10}$	$203^{+10}_{-9}$	$1.02^{+0.06}_{-0.02}$	$0.33^{+0.20}_{-0.17}$	$0.21^{+0.01}_{-0.01}$	$0.40^{+0.01}_{-0.01}$	0.58
35	1548	461	$12.03^{+1.19}_{-1.11}$	$0.17^{+0.16}_{-0.09}$	$180^{+5}_{-5}$	$1.21^{+0.07}_{-0.11}$	$0.25^{+0.10}_{-0.05}$	$0.24^{+0.01}_{-0.01}$	$0.45^{+0.01}_{-0.01}$	1.10
36	213	488	$9.20^{+2.70}_{-1.35}$	$0.19^{+0.06}_{-0.07}$	$263^{+23}_{-31}$	$0.99^{+0.00}_{-0.00}$	$0.84^{+0.16}_{-0.38}$	$0.17^{+0.02}_{-0.01}$	$0.33^{+0.02}_{-0.01}$	0.54
37	611	488	$11.24^{+1.54}_{-1.33}$	$0.10^{+0.11}_{-0.06}$	$206^{+10}_{-7}$	$0.95^{+0.09}_{-0.00}$	$0.26^{+0.20}_{-0.10}$	$0.21^{+0.01}_{-0.01}$	$0.40^{+0.01}_{-0.01}$	0.69
38	1776	488	$11.23^{+1.09}_{-0.94}$	$0.20^{+0.17}_{-0.11}$	$174^{+5}_{-4}$	$1.09^{+0.18}_{-0.04}$	$0.28^{+0.10}_{-0.01}$	$0.25^{+0.01}_{-0.01}$	$0.46^{+0.01}_{-0.01}$	0.82
39	217	519	$9.22^{+2.71}_{-1.44}$	$0.13^{+0.06}_{-0.06}$	$267^{+27}_{-30}$	$0.89^{+0.03}_{-0.08}$	$0.90^{+0.35}_{-0.40}$	$0.17^{+0.02}_{-0.02}$	$0.35^{+0.02}_{-0.02}$	0.31
40	627	519	$11.15^{+1.46}_{-1.29}$	$0.07^{+0.10}_{-0.04}$	$212^{+10}_{-7}$	$0.99^{+0.04}_{-0.02}$	$0.22^{+0.32}_{-0.02}$	$0.21^{+0.01}_{-0.01}$	$0.42^{+0.01}_{-0.01}$	0.65
41	1679	519	$12.09^{+1.06}_{-1.00}$	$0.17^{+0.16}_{-0.10}$	$177^{+5}_{-4}$	$1.01^{+0.04}_{-0.01}$	$0.34^{+0.11}_{-0.16}$	$0.25^{+0.01}_{-0.01}$	$0.48^{+0.00}_{-0.01}$	1.04
42	214	786	$18.94^{+4.61}_{-4.26}$	$0.01^{+0.08}_{-0.00}$	$208^{+26}_{-18}$	$1.02^{+0.00}_{-0.02}$	$0.27^{+0.29}_{-0.00}$	$0.23^{+0.02}_{-0.03}$	$0.42^{+0.01}_{-0.02}$	1.32
43	656	786	$17.10^{+2.10}_{-2.29}$	$0.01^{+0.08}_{-0.00}$	$203^{+10}_{-7}$	$1.00^{+0.00}_{-0.00}$	$0.27^{+0.07}_{-0.03}$	$0.22^{+0.01}_{-0.01}$	$0.44^{+0.00}_{-0.01}$	1.81
44	1865	786	$19.33^{+1.96}_{-1.76}$	$0.31^{+0.25}_{-0.18}$	$167^{+5}_{-4}$	$1.05^{+0.03}_{-0.05}$	$0.26^{+0.16}_{-0.04}$	$0.27^{+0.01}_{-0.01}$	$0.49^{+0.00}_{-0.00}$	2.24
45	209	1356	$27.22^{+13.17}_{-5.33}$	$0.18^{+0.08}_{-0.09}$	$210^{+24}_{-17}$	$1.03^{+0.00}_{-0.01}$	$0.54^{+0.10}_{-0.20}$	$0.18^{+0.04}_{-0.02}$	$0.34^{+0.02}_{-0.01}$	2.18
46	619	1356	$32.18^{+5.96}_{-4.31}$	$0.17^{+0.16}_{-0.10}$	$188^{+37}_{-12}$	$1.03^{+0.01}_{-0.01}$	$0.46^{+0.24}_{-0.15}$	$0.20^{+0.02}_{-0.01}$	$0.36^{+0.01}_{-0.01}$	4.27
47	1710	1356	$43.00^{+4.94}_{-3.83}$	$0.45^{+0.29}_{-0.27}$	$157^{+12}_{-6}$	$1.04^{+0.03}_{-0.01}$	$0.48^{+0.22}_{-0.16}$	$0.24^{+0.01}_{-0.01}$	$0.40^{+0.00}_{-0.00}$	9.27
48	176	499	$49.87^{+51.85}_{-14.52}$	$0.01^{+0.20}_{-0.00}$	$184^{+36}_{-53}$	$1.06^{+0.01}_{-0.04}$	$0.74^{+0.05}_{-0.00}$	$0.38^{+0.16}_{-0.06}$	$0.48^{+0.11}_{-0.03}$	3.54
49	485	499	$47.56^{+11.39}_{-7.34}$	$0.01^{+0.28}_{-0.00}$	$189^{+15}_{-15}$	$1.07^{+0.00}_{-0.07}$	$0.26^{+0.33}_{-0.00}$	$0.36^{+0.03}_{-0.03}$	$0.48^{+0.02}_{-0.01}$	6.19
50	1147	499	$79.12^{+11.28}_{-9.60}$	$0.01^{+0.68}_{-0.00}$	$152^{+9}_{-6}$	$0.89^{+0.15}_{-0.00}$	$0.26^{+0.21}_{-0.00}$	$0.45^{+0.02}_{-0.03}$	$0.55^{+0.01}_{-0.01}$	12.44

All time-scales are in units of Myr. All length-scales are in units of pc.

Table D7. Best-fitting solutions for gas maps (low resolution, extended emission)

ID	$N_{\text{star}}$	$N_{\text{gas}}$	$t_{\text{gas}}$	$t_{\text{over}}$	$\lambda$	$\beta_{\text{star}}$	$\beta_{\text{gas}}$	$\zeta_{\text{star}}$	$\zeta_{\text{gas}}$	$\chi^2_{\text{red}}$
1	15	137	$27.21^{+653.94}_{-9.80}$	$1.01^{+27.84}_{-0.30}$	$695^{+121}_{-171}$	$1.54^{+0.31}_{-0.22}$	$1.31^{+0.18}_{-0.11}$	$0.46^{+0.08}_{-0.04}$	$0.42^{+0.07}_{-0.03}$	0.38
2	25	137	$20.13^{+625.50}_{-6.50}$	$1.67^{+55.31}_{-0.44}$	$878^{+184}_{-247}$	$1.25^{+1.79}_{-0.03}$	$1.06^{+0.09}_{-0.17}$	$0.38^{+0.10}_{-0.05}$	$0.37^{+0.08}_{-0.04}$	0.50
3	42	169	$12.72^{+3.99}_{-2.51}$	$1.25^{+0.51}_{-0.31}$	$598^{+107}_{-70}$	$0.66^{+0.05}_{-0.02}$	$1.03^{+0.11}_{-0.09}$	$0.40^{+0.03}_{-0.03}$	$0.37^{+0.03}_{-0.03}$	0.85
4	74	169	$13.85^{+2.67}_{-2.92}$	$3.64^{+0.82}_{-0.87}$	$610^{+117}_{-101}$	$0.87^{+0.04}_{-0.01}$	$0.97^{+0.03}_{-0.07}$	$0.39^{+0.04}_{-0.03}$	$0.38^{+0.05}_{-0.04}$	1.25
5	86	169	$10.92^{+2.76}_{-1.35}$	$2.87^{+1.13}_{-0.51}$	$377^{+103}_{-52}$	$0.92^{+0.09}_{-0.02}$	$0.83^{+0.08}_{-0.12}$	$0.52^{+0.02}_{-0.06}$	$0.52^{+0.03}_{-0.07}$	0.66
6	100	169	$11.47^{+5.06}_{-1.55}$	$4.40^{+2.57}_{-0.87}$	$329^{+112}_{-45}$	$0.99^{+0.01}_{-0.05}$	$0.81^{+0.15}_{-0.10}$	$0.52^{+0.03}_{-0.05}$	$0.55^{+0.03}_{-0.07}$	0.37
7	108	169	$16.14^{+6.10}_{-2.19}$	$6.40^{+3.07}_{-1.29}$	$325^{+86}_{-43}$	$1.06^{+0.07}_{-0.07}$	$0.62^{+0.09}_{-0.05}$	$0.55^{+0.02}_{-0.03}$	$0.56^{+0.03}_{-0.04}$	0.20
8	109	169	$17.87^{+4.60}_{-2.96}$	$12.20^{+3.23}_{-2.23}$	$339^{+88}_{-52}$	$0.96^{+0.10}_{-0.03}$	$0.71^{+0.02}_{-0.02}$	$0.54^{+0.03}_{-0.03}$	$0.56^{+0.03}_{-0.05}$	0.30
9	114	169	$21.59^{+6.21}_{-2.75}$	$18.26^{+4.69}_{-2.57}$	$352^{+120}_{-52}$	$1.04^{+0.20}_{-0.08}$	$0.59^{+0.05}_{-0.02}$	$0.53^{+0.03}_{-0.05}$	$0.56^{+0.03}_{-0.07}$	0.29
10	108	169	$36.08^{+36.07}_{-5.00}$	$35.98^{+32.34}_{-5.82}$	$376^{+352}_{-68}$	$1.08^{+0.14}_{-0.09}$	$1.74^{+0.00}_{-1.11}$	$0.52^{+0.03}_{-0.11}$	$0.55^{+0.03}_{-0.17}$	0.90
11	85	169	$66.92^{+37.49}_{-9.73}$	$66.15^{+30.82}_{-11.96}$	$397^{+189}_{-65}$	$0.83^{+0.11}_{-0.04}$	$0.68^{+0.05}_{-0.13}$	$0.51^{+0.03}_{-0.07}$	$0.54^{+0.03}_{-0.11}$	0.65
12	42	206	$598.74^{+1452.38}_{-210.80}$	$28.35^{+73.57}_{-9.67}$	$444^{+73}_{-52}$	$0.22^{+0.03}_{-0.07}$	$1.59^{+0.02}_{-0.09}$	$0.50^{+0.03}_{-0.03}$	$0.53^{+0.03}_{-0.03}$	0.70
13	86	206	$2608.31^{+222.74}_{-872.57}$	$301.92^{+37.24}_{-102.52}$	$381^{+57}_{-43}$	$0.39^{+0.13}_{-0.10}$	$1.30^{+0.03}_{-0.03}$	$0.56^{+0.02}_{-0.03}$	$0.56^{+0.02}_{-0.03}$	0.94
14	108	206	$1022.09^{+1303.32}_{-280.18}$	$143.67^{+181.32}_{-43.61}$	$350^{+56}_{-54}$	$1.09^{+1.23}_{-0.52}$	$1.30^{+0.06}_{-0.02}$	$0.58^{+0.02}_{-0.02}$	$0.57^{+0.02}_{-0.02}$	0.58
15	42	64	$1.22^{+0.63}_{-0.30}$	$0.20^{+0.13}_{-0.12}$	$666^{+194}_{-143}$	$0.93^{+0.02}_{-0.01}$	$0.45^{+0.24}_{-0.26}$	$0.34^{+0.05}_{-0.05}$	$0.29^{+0.06}_{-0.05}$	0.46
16	86	64	$0.84^{+0.36}_{-0.18}$	$0.56^{+0.16}_{-0.11}$	$274^{+88}_{-48}$	$1.36^{+0.08}_{-0.10}$	$1.28^{+0.20}_{-0.42}$	$0.56^{+0.04}_{-0.05}$	$0.54^{+0.06}_{-0.07}$	0.24
17	108	64	$1.44^{+0.76}_{-0.25}$	$1.38^{+0.31}_{-0.29}$	$262^{+98}_{-43}$	$2.08^{+0.09}_{-0.01}$	$1.18^{+0.10}_{-0.42}$	$0.57^{+0.03}_{-0.05}$	$0.55^{+0.05}_{-0.08}$	0.20
18	42	169	$12.72^{+3.96}_{-2.52}$	$1.25^{+0.51}_{-0.31}$	$598^{+109}_{-70}$	$0.66^{+0.05}_{-0.02}$	$1.03^{+0.11}_{-0.09}$	$0.40^{+0.03}_{-0.03}$	$0.37^{+0.03}_{-0.03}$	0.85
19	88	198	$10.25^{+2.35}_{-1.05}$	$2.68^{+0.92}_{-0.38}$	$346^{+83}_{-43}$	$0.85^{+0.00}_{-0.03}$	$0.75^{+0.12}_{-0.13}$	$0.47^{+0.02}_{-0.03}$	$0.48^{+0.03}_{-0.04}$	0.99
20	116	198	$20.01^{+4.68}_{-3.76}$	$8.72^{+2.55}_{-2.16}$	$345^{+65}_{-49}$	$0.77^{+0.07}_{-0.02}$	$0.67^{+0.06}_{-0.06}$	$0.45^{+0.02}_{-0.02}$	$0.49^{+0.03}_{-0.03}$	0.38
21	42	169	$12.72^{+3.98}_{-2.51}$	$1.25^{+0.51}_{-0.31}$	$591^{+109}_{-66}$	$0.66^{+0.05}_{-0.02}$	$1.03^{+0.11}_{-0.09}$	$0.41^{+0.03}_{-0.03}$	$0.38^{+0.03}_{-0.03}$	0.85
22	88	198	$10.25^{+2.35}_{-1.05}$	$2.68^{+0.92}_{-0.38}$	$348^{+83}_{-43}$	$0.85^{+0.00}_{-0.03}$	$0.75^{+0.12}_{-0.13}$	$0.46^{+0.02}_{-0.03}$	$0.48^{+0.03}_{-0.04}$	0.97
23	116	198	$20.01^{+4.69}_{-3.76}$	$8.72^{+2.55}_{-2.16}$	$345^{+66}_{-49}$	$0.77^{+0.07}_{-0.02}$	$0.67^{+0.06}_{-0.05}$	$0.45^{+0.02}_{-0.02}$	$0.49^{+0.03}_{-0.03}$	0.39
24	40	137	$11.86^{+4.19}_{-2.11}$	$1.49^{+0.71}_{-0.32}$	$699^{+227}_{-120}$	$1.04^{+0.32}_{-0.11}$	$1.00^{+0.18}_{-0.17}$	$0.46^{+0.05}_{-0.08}$	$0.43^{+0.05}_{-0.06}$	0.33
25	80	137	$12.21^{+3.39}_{-1.85}$	$4.29^{+1.62}_{-1.01}$	$519^{+169}_{-108}$	$1.28^{+0.06}_{-0.10}$	$0.90^{+0.11}_{-0.13}$	$0.54^{+0.07}_{-0.07}$	$0.52^{+0.07}_{-0.07}$	0.58
26	95	137	$17.70^{+4.29}_{-2.57}$	$8.26^{+2.96}_{-1.67}$	$381^{+116}_{-63}$	$1.11^{+0.17}_{-0.08}$	$0.58^{+0.04}_{-0.06}$	$0.63^{+0.02}_{-0.06}$	$0.61^{+0.03}_{-0.07}$	0.24
27	35	84	$13.81^{+626.59}_{-3.13}$	$2.59^{+130.64}_{-0.71}$	$850^{+314}_{-206}$	$1.32^{+1.12}_{-0.14}$	$0.90^{+0.27}_{-0.12}$	$0.52^{+0.08}_{-0.09}$	$0.51^{+0.07}_{-0.09}$	0.26
28	65	84	$15.49^{+6.21}_{-2.89}$	$5.80^{+3.24}_{-1.68}$	$597^{+380}_{-142}$	$1.22^{+0.28}_{-0.14}$	$1.04^{+0.11}_{-0.06}$	$0.61^{+0.06}_{-0.11}$	$0.61^{+0.07}_{-0.12}$	0.60
29	71	84	$17.25^{+5.06}_{-2.93}$	$9.70^{+3.54}_{-2.62}$	$461^{+213}_{-106}$	$1.46^{+0.30}_{-0.17}$	$0.64^{+0.07}_{-0.06}$	$0.67^{+0.03}_{-0.06}$	$0.67^{+0.04}_{-0.07}$	0.43
30	30	35	$13.03^{+641.06}_{-4.81}$	$4.27^{+211.70}_{-1.55}$	$1324^{+814}_{-574}$	$1.18^{+6.36}_{-0.16}$	$1.04^{+0.47}_{-0.41}$	$0.54^{+0.17}_{-0.13}$	$0.56^{+0.15}_{-0.12}$	0.20
31	44	35	$21.18^{+692.71}_{-4.90}$	$12.67^{+442.59}_{-5.14}$	$1058^{+5997}_{-495}$	$1.87^{+4.25}_{-0.34}$	$0.69^{+0.42}_{-0.09}$	$0.62^{+0.09}_{-0.33}$	$0.62^{+0.09}_{-0.34}$	0.49
32	37	35	$25.26^{+713.28}_{-5.54}$	$16.33^{+539.97}_{-9.51}$	$685^{+7399}_{-328}$	$2.09^{+6.93}_{-0.93}$	$0.60^{+0.10}_{-0.04}$	$0.71^{+0.00}_{-0.43}$	$0.71^{+0.00}_{-0.47}$	0.71
33	40	172	$11.82^{+4.95}_{-2.08}$	$1.11^{+0.64}_{-0.29}$	$568^{+101}_{-67}$	$0.70^{+0.07}_{-0.01}$	$0.85^{+0.19}_{-0.07}$	$0.43^{+0.03}_{-0.03}$	$0.41^{+0.03}_{-0.04}$	0.90
34	78	172	$9.01^{+2.41}_{-1.07}$	$3.02^{+1.15}_{-0.47}$	$352^{+126}_{-48}$	$0.79^{+0.09}_{-0.02}$	$0.91^{+0.09}_{-0.04}$	$0.53^{+0.02}_{-0.06}$	$0.55^{+0.03}_{-0.07}$	0.32
35	98	172	$16.50^{+4.26}_{-2.35}$	$8.45^{+2.81}_{-1.71}$	$340^{+94}_{-50}$	$0.86^{+0.11}_{-0.02}$	$0.65^{+0.04}_{-0.02}$	$0.52^{+0.03}_{-0.04}$	$0.57^{+0.03}_{-0.05}$	0.11
36	40	153	$11.56^{+4.31}_{-2.07}$	$1.51^{+0.69}_{-0.32}$	$634^{+136}_{-70}$	$0.73^{+0.06}_{-0.04}$	$0.96^{+0.15}_{-0.05}$	$0.43^{+0.02}_{-0.03}$	$0.40^{+0.03}_{-0.04}$	1.12
37	70	153	$9.24^{+3.57}_{-1.15}$	$3.19^{+1.73}_{-0.54}$	$345^{+129}_{-48}$	$0.76^{+0.04}_{-0.02}$	$0.89^{+0.13}_{-0.09}$	$0.52^{+0.03}_{-0.05}$	$0.57^{+0.02}_{-0.07}$	0.40
38	84	153	$15.90^{+4.77}_{-2.19}$	$9.75^{+3.38}_{-1.84}$	$332^{+94}_{-48}$	$0.77^{+0.09}_{-0.03}$	$0.66^{+0.04}_{-0.01}$	$0.52^{+0.03}_{-0.04}$	$0.58^{+0.02}_{-0.05}$	0.10
39	37	139	$13.62^{+4.52}_{-2.76}$	$2.32^{+0.93}_{-0.60}$	$642^{+142}_{-83}$	$0.52^{+0.05}_{-0.02}$	$0.84^{+0.07}_{-0.03}$	$0.46^{+0.03}_{-0.03}$	$0.42^{+0.03}_{-0.04}$	0.72
40	66	139	$9.98^{+3.24}_{-1.23}$	$4.27^{+1.93}_{-0.93}$	$350^{+147}_{-55}$	$0.57^{+0.07}_{-0.03}$	$0.83^{+0.07}_{-0.13}$	$0.52^{+0.03}_{-0.06}$	$0.58^{+0.02}_{-0.08}$	0.30
41	76	139	$15.25^{+7.47}_{-2.55}$	$11.30^{+5.77}_{-2.22}$	$356^{+164}_{-57}$	$0.78^{+0.05}_{-0.01}$	$0.82^{+0.17}_{-0.11}$	$0.51^{+0.03}_{-0.05}$	$0.59^{+0.02}_{-0.09}$	0.14
42	35	90	$10.25^{+6.04}_{-2.36}$	$1.76^{+1.38}_{-0.53}$	$804^{+263}_{-136}$	$0.64^{+0.10}_{-0.01}$	$1.05^{+0.13}_{-0.08}$	$0.48^{+0.03}_{-0.06}$	$0.45^{+0.02}_{-0.08}$	0.46
43	49	120	$9.26^{+1.82}_{-1.18}$	$4.00^{+1.20}_{-0.65}$	$400^{+133}_{-61}$	$0.64^{+0.07}_{-0.01}$	$0.90^{+0.06}_{-0.12}$	$0.52^{+0.03}_{-0.06}$	$0.58^{+0.02}_{-0.07}$	0.35
44	61	120	$16.49^{+8.81}_{-2.87}$	$14.10^{+8.81}_{-2.63}$	$435^{+185}_{-66}$	$0.67^{+0.05}_{-0.07}$	$1.06^{+0.40}_{-0.27}$	$0.46^{+0.02}_{-0.03}$	$0.57^{+0.03}_{-0.08}$	0.37
45	32	78	$6.10^{+2.29}_{-1.13}$	$1.35^{+0.81}_{-0.41}$	$849^{+315}_{-169}$	$0.50^{+0.17}_{-0.07}$	$0.72^{+0.20}_{-0.10}$	$0.47^{+0.05}_{-0.06}$	$0.51^{+0.05}_{-0.08}$	0.27
46	35	78	$7.09^{+1.93}_{-1.02}$	$3.97^{+1.38}_{-0.71}$	$579^{+236}_{-96}$	$0.75^{+0.03}_{-0.01}$	$0.96^{+0.41}_{-0.06}$	$0.53^{+0.04}_{-0.06}$	$0.60^{+0.03}_{-0.06}$	0.12
47	34	82	$18.08^{+4.17}_{-3.06}$	$15.56^{+3.14}_{-3.06}$	$573^{+145}_{-98}$	$0.65^{+0.04}_{-0.01}$	$1.27^{+0.51}_{-0.32}$	$0.45^{+0.04}_{-0.04}$	$0.57^{+0.03}_{-0.03}$	0.18
48	19	33	$11.19^{+675.54}_{-3.34}$	$3.57^{+290.54}_{-1.50}$	$940^{+1310}_{-320}$	$0.37^{+9.10}_{-0.06}$	$0.82^{+0.10}_{-0.08}$	$0.58^{+0.07}_{-0.18}$	$0.60^{+0.06}_{-0.21}$	0.15
49	20	41	$8.70^{+5.29}_{-1.59}$	$4.86^{+3.75}_{-0.91}$	$632^{+294}_{-127}$	$0.84^{+0.08}_{-0.26}$	$0.92^{+0.07}_{-0.11}$	$0.57^{+0.03}_{-0.07}$	$0.64^{+0.02}_{-0.05}$	0.18
50	15	41	$39.00^{+32.87}_{-8.42}$	$38.18^{+23.76}_{-9.98}$	$776^{+411}_{-249}$	$0.37^{+0.06}_{-0.03}$	$0.73^{+0.10}_{-0.02}$	$0.44^{+0.09}_{-0.04}$	$0.61^{+0.03}_{-0.08}$	0.21

All time-scales are in units of Myr. All length-scales are in units of pc.

**Table D8.** Best-fitting solutions for gas maps (high resolution, extended emission)

ID	$N_{\text{star}}$	$N_{\text{gas}}$	$t_{\text{gas}}$	$t_{\text{over}}$	$\lambda$	$\beta_{\text{star}}$	$\beta_{\text{gas}}$	$\zeta_{\text{star}}$	$\zeta_{\text{gas}}$	$\chi^2_{\text{red}}$
1	58	293	$6.95^{+2.13}_{-1.39}$	$0.31^{+0.11}_{-0.06}$	$374^{+73}_{-54}$	$1.13^{+0.16}_{-0.14}$	$0.80^{+0.12}_{-0.10}$	$0.33^{+0.04}_{-0.05}$	$0.29^{+0.03}_{-0.04}$	0.21
2	109	293	$3.93^{+0.87}_{-0.49}$	$0.27^{+0.09}_{-0.09}$	$339^{+46}_{-55}$	$1.31^{+0.02}_{-0.16}$	$0.60^{+0.13}_{-0.23}$	$0.34^{+0.04}_{-0.03}$	$0.33^{+0.03}_{-0.03}$	0.38
3	180	293	$3.38^{+0.51}_{-0.40}$	$0.39^{+0.09}_{-0.12}$	$295^{+40}_{-40}$	$1.08^{+0.09}_{-0.08}$	$0.57^{+0.12}_{-0.16}$	$0.37^{+0.04}_{-0.03}$	$0.37^{+0.04}_{-0.03}$	0.61
4	297	293	$3.21^{+0.42}_{-0.27}$	$0.52^{+0.12}_{-0.18}$	$212^{+24}_{-23}$	$1.01^{+0.12}_{-0.12}$	$0.38^{+0.12}_{-0.16}$	$0.47^{+0.03}_{-0.03}$	$0.49^{+0.03}_{-0.03}$	0.33
5	353	293	$3.64^{+0.49}_{-0.30}$	$0.74^{+0.10}_{-0.15}$	$174^{+15}_{-14}$	$1.17^{+0.12}_{-0.12}$	$0.32^{+0.11}_{-0.04}$	$0.52^{+0.02}_{-0.01}$	$0.54^{+0.02}_{-0.01}$	0.37
6	414	293	$2.76^{+0.33}_{-0.18}$	$1.07^{+0.12}_{-0.09}$	$137^{+13}_{-9}$	$1.41^{+0.03}_{-0.19}$	$0.46^{+0.08}_{-0.05}$	$0.58^{+0.01}_{-0.02}$	$0.59^{+0.01}_{-0.02}$	0.20
7	409	293	$3.99^{+0.48}_{-0.40}$	$1.96^{+0.15}_{-0.23}$	$140^{+12}_{-10}$	$1.43^{+0.07}_{-0.01}$	$0.54^{+0.04}_{-0.05}$	$0.58^{+0.01}_{-0.01}$	$0.59^{+0.01}_{-0.01}$	0.32
8	351	293	$5.84^{+0.78}_{-0.58}$	$4.11^{+0.33}_{-0.40}$	$144^{+15}_{-12}$	$1.39^{+0.12}_{-0.08}$	$0.49^{+0.06}_{-0.03}$	$0.57^{+0.02}_{-0.02}$	$0.59^{+0.02}_{-0.02}$	0.29
9	274	293	$6.99^{+1.18}_{-0.42}$	$6.96^{+0.89}_{-0.51}$	$136^{+19}_{-9}$	$1.45^{+0.01}_{-0.03}$	$0.46^{+0.17}_{-0.00}$	$0.60^{+0.01}_{-0.02}$	$0.60^{+0.01}_{-0.02}$	1.59
10	239	293	$14.48^{+2.50}_{-1.07}$	$14.46^{+2.38}_{-1.24}$	$150^{+25}_{-12}$	$1.26^{+0.01}_{-0.04}$	$0.70^{+0.00}_{-0.22}$	$0.60^{+0.01}_{-0.02}$	$0.58^{+0.02}_{-0.03}$	0.63
11	170	293	$21.31^{+2.27}_{-1.82}$	$21.31^{+1.98}_{-2.05}$	$145^{+15}_{-13}$	$1.51^{+0.01}_{-0.00}$	$0.38^{+0.12}_{-0.00}$	$0.60^{+0.01}_{-0.01}$	$0.59^{+0.02}_{-0.02}$	2.99
12	180	371	$36.17^{+11.18}_{-5.75}$	$2.37^{+0.84}_{-0.18}$	$218^{+21}_{-18}$	$0.86^{+0.02}_{-0.01}$	$1.03^{+0.02}_{-0.01}$	$0.49^{+0.03}_{-0.03}$	$0.50^{+0.02}_{-0.02}$	0.80
13	353	371	$33.33^{+3.81}_{-3.28}$	$4.11^{+0.64}_{-0.57}$	$183^{+11}_{-11}$	$0.88^{+0.04}_{-0.04}$	$0.75^{+0.04}_{-0.08}$	$0.55^{+0.01}_{-0.01}$	$0.55^{+0.01}_{-0.01}$	0.95
14	409	371	$28.50^{+2.42}_{-0.91}$	$5.75^{+0.74}_{-0.78}$	$170^{+11}_{-11}$	$1.02^{+0.05}_{-0.01}$	$0.73^{+0.02}_{-0.01}$	$0.57^{+0.01}_{-0.01}$	$0.58^{+0.01}_{-0.01}$	1.11
15	180	94	$0.25^{+0.13}_{-0.05}$	$0.06^{+0.05}_{-0.03}$	$223^{+105}_{-40}$	$1.17^{+0.23}_{-0.12}$	$0.61^{+0.22}_{-0.29}$	$0.41^{+0.06}_{-0.10}$	$0.40^{+0.06}_{-0.12}$	0.46
16	353	94	$0.20^{+0.12}_{-0.04}$	$0.11^{+0.03}_{-0.02}$	$98^{+34}_{-13}$	$2.90^{+0.24}_{-0.05}$	$0.56^{+0.10}_{-0.22}$	$0.64^{+0.01}_{-0.07}$	$0.63^{+0.02}_{-0.08}$	0.33
17	409	94	$0.39^{+3.25}_{-0.01}$	$0.39^{+1.82}_{-0.01}$	$93^{+743}_{-8}$	$1.93^{+0.75}_{-0.38}$	$0.88^{+0.16}_{-0.34}$	$0.65^{+0.01}_{-0.36}$	$0.63^{+0.01}_{-0.49}$	1.17
18	180	293	$3.38^{+0.51}_{-0.40}$	$0.39^{+0.09}_{-0.12}$	$295^{+39}_{-40}$	$1.08^{+0.09}_{-0.08}$	$0.57^{+0.12}_{-0.16}$	$0.37^{+0.04}_{-0.03}$	$0.37^{+0.04}_{-0.03}$	0.61
19	361	316	$3.50^{+0.33}_{-0.29}$	$0.62^{+0.07}_{-0.11}$	$165^{+12}_{-12}$	$0.95^{+0.05}_{-0.05}$	$0.43^{+0.02}_{-0.08}$	$0.48^{+0.01}_{-0.01}$	$0.51^{+0.01}_{-0.01}$	0.40
20	467	316	$3.49^{+0.37}_{-0.26}$	$1.70^{+0.15}_{-0.09}$	$127^{+10}_{-7}$	$1.31^{+0.03}_{-0.05}$	$0.61^{+0.03}_{-0.02}$	$0.54^{+0.01}_{-0.01}$	$0.57^{+0.01}_{-0.02}$	0.29
21	171	251	$4.21^{+0.75}_{-0.57}$	$0.42^{+0.14}_{-0.20}$	$320^{+45}_{-54}$	$1.21^{+0.14}_{-0.25}$	$0.56^{+0.07}_{-0.38}$	$0.43^{+0.06}_{-0.03}$	$0.43^{+0.05}_{-0.03}$	0.47
22	294	251	$4.06^{+0.66}_{-0.39}$	$0.87^{+0.28}_{-0.29}$	$197^{+23}_{-18}$	$1.30^{+0.03}_{-0.11}$	$0.27^{+0.13}_{-0.14}$	$0.59^{+0.02}_{-0.04}$	$0.60^{+0.01}_{-0.03}$	0.25
23	298	251	$4.19^{+0.72}_{-0.51}$	$1.95^{+0.24}_{-0.25}$	$150^{+12}_{-12}$	$1.98^{+0.03}_{-0.06}$	$0.41^{+0.02}_{-0.12}$	$0.64^{+0.01}_{-0.01}$	$0.65^{+0.01}_{-0.01}$	0.43
24	157	159	$4.11^{+0.95}_{-0.67}$	$0.58^{+0.30}_{-0.28}$	$349^{+72}_{-58}$	$1.13^{+0.16}_{-0.05}$	$0.47^{+0.04}_{-0.35}$	$0.55^{+0.05}_{-0.06}$	$0.56^{+0.04}_{-0.05}$	0.12
25	198	159	$3.66^{+0.57}_{-0.44}$	$2.17^{+0.45}_{-0.43}$	$251^{+53}_{-38}$	$1.28^{+0.06}_{-0.03}$	$0.55^{+0.03}_{-0.09}$	$0.64^{+0.05}_{-0.05}$	$0.65^{+0.04}_{-0.04}$	0.37
26	151	159	$5.17^{+0.92}_{-0.59}$	$3.96^{+0.65}_{-0.53}$	$213^{+38}_{-25}$	$1.76^{+0.06}_{-0.08}$	$0.55^{+0.05}_{-0.16}$	$0.69^{+0.02}_{-0.03}$	$0.69^{+0.02}_{-0.04}$	0.52
27	90	63	$3.66^{+3.56}_{-0.71}$	$1.44^{+2.22}_{-0.74}$	$476^{+446}_{-163}$	$1.30^{+0.21}_{-0.15}$	$0.43^{+0.16}_{-0.24}$	$0.63^{+0.07}_{-0.17}$	$0.65^{+0.06}_{-0.16}$	0.28
28	81	63	$4.83^{+231.65}_{-1.17}$	$4.35^{+183.22}_{-1.37}$	$418^{+8529}_{-141}$	$1.46^{+1.61}_{-0.02}$	$0.80^{+0.08}_{-0.37}$	$0.67^{+0.03}_{-0.40}$	$0.69^{+0.02}_{-0.46}$	0.21
29	22	63	$5.46^{+246.66}_{-1.25}$	$5.18^{+194.68}_{-3.10}$	$300^{+2109}_{-67}$	$3.43^{+1.00}_{-0.97}$	$0.51^{+0.02}_{-0.22}$	$0.71^{+0.00}_{-0.31}$	$0.71^{+0.00}_{-0.43}$	0.55
30	17	20	$64.46^{+132.87}_{-28.57}$	$26.60^{+80.29}_{-13.42}$	$2212^{+2111}_{-1248}$	$1.63^{+1.18}_{-0.84}$	$0.47^{+0.01}_{-0.09}$	$0.47^{+0.16}_{-0.05}$	$0.39^{+0.23}_{-0.09}$	0.07
31	7	20	$309.14^{+0.00}_{-135.03}$	$175.19^{+56.84}_{-121.53}$	$3065^{+7254}_{-1956}$	$1.56^{+1.18}_{-0.00}$	$0.34^{+0.07}_{-0.16}$	$0.41^{+0.06}_{-0.01}$	$0.33^{+0.24}_{-0.08}$	0.00
32	10	20	$7.76^{+238.08}_{-2.02}$	$7.75^{+192.65}_{-5.06}$	$532^{+1719}_{-210}$	$3.65^{+0.66}_{-1.32}$	$0.27^{+0.05}_{-0.00}$	$0.71^{+0.00}_{-0.16}$	$0.71^{+0.00}_{-0.34}$	0.65
33	166	281	$3.37^{+0.47}_{-0.42}$	$0.43^{+0.08}_{-0.12}$	$311^{+36}_{-42}$	$1.04^{+0.03}_{-0.08}$	$0.65^{+0.04}_{-0.16}$	$0.37^{+0.04}_{-0.02}$	$0.37^{+0.04}_{-0.02}$	0.61
34	318	281	$3.38^{+0.40}_{-0.30}$	$0.77^{+0.16}_{-0.13}$	$177^{+19}_{-15}$	$1.30^{+0.02}_{-0.03}$	$0.38^{+0.05}_{-0.11}$	$0.54^{+0.01}_{-0.02}$	$0.54^{+0.02}_{-0.02}$	0.27
35	329	281	$4.06^{+0.50}_{-0.45}$	$2.23^{+0.03}_{-0.28}$	$144^{+10}_{-12}$	$1.44^{+0.08}_{-0.01}$	$0.52^{+0.04}_{-0.04}$	$0.58^{+0.01}_{-0.01}$	$0.59^{+0.02}_{-0.01}$	0.13
36	162	254	$3.07^{+0.47}_{-0.33}$	$0.46^{+0.10}_{-0.10}$	$333^{+46}_{-45}$	$1.06^{+0.06}_{-0.07}$	$0.62^{+0.08}_{-0.13}$	$0.38^{+0.03}_{-0.03}$	$0.38^{+0.04}_{-0.03}$	0.41
37	281	254	$2.86^{+0.35}_{-0.23}$	$0.93^{+0.17}_{-0.10}$	$185^{+21}_{-16}$	$1.34^{+0.00}_{-0.05}$	$0.49^{+0.03}_{-0.06}$	$0.54^{+0.01}_{-0.02}$	$0.56^{+0.01}_{-0.02}$	0.19
38	259	254	$3.76^{+0.43}_{-0.39}$	$2.52^{+0.09}_{-0.20}$	$150^{+12}_{-13}$	$1.52^{+0.01}_{-0.03}$	$0.64^{+0.02}_{-0.10}$	$0.59^{+0.01}_{-0.01}$	$0.60^{+0.01}_{-0.01}$	0.12
39	152	235	$2.68^{+0.36}_{-0.33}$	$0.41^{+0.08}_{-0.09}$	$326^{+43}_{-42}$	$1.19^{+0.09}_{-0.10}$	$0.64^{+0.07}_{-0.12}$	$0.41^{+0.03}_{-0.02}$	$0.42^{+0.03}_{-0.02}$	0.45
40	236	235	$2.72^{+0.36}_{-0.24}$	$1.02^{+0.20}_{-0.13}$	$193^{+26}_{-18}$	$1.39^{+0.01}_{-0.03}$	$0.52^{+0.07}_{-0.04}$	$0.56^{+0.01}_{-0.03}$	$0.58^{+0.01}_{-0.03}$	0.22
41	196	235	$3.18^{+0.44}_{-0.17}$	$2.90^{+0.03}_{-0.05}$	$148^{+14}_{-11}$	$1.76^{+0.12}_{-0.24}$	$0.56^{+0.05}_{-0.02}$	$0.60^{+0.01}_{-0.01}$	$0.62^{+0.01}_{-0.01}$	0.16
42	126	164	$2.32^{+0.50}_{-0.27}$	$0.47^{+0.13}_{-0.12}$	$357^{+53}_{-55}$	$1.35^{+0.06}_{-0.23}$	$0.52^{+0.04}_{-0.04}$	$0.48^{+0.01}_{-0.03}$	$0.51^{+0.03}_{-0.03}$	0.41
43	183	187	$2.83^{+0.26}_{-0.25}$	$1.44^{+0.14}_{-0.16}$	$221^{+31}_{-24}$	$1.07^{+0.16}_{-0.03}$	$0.56^{+0.02}_{-0.04}$	$0.54^{+0.03}_{-0.03}$	$0.57^{+0.03}_{-0.03}$	0.17
44	139	187	$4.11^{+0.36}_{-0.19}$	$4.10^{+0.23}_{-0.26}$	$186^{+22}_{-14}$	$1.31^{+0.02}_{-0.00}$	$0.58^{+0.08}_{-0.07}$	$0.58^{+0.01}_{-0.02}$	$0.61^{+0.01}_{-0.02}$	0.45
45	97	121	$1.81^{+0.35}_{-0.23}$	$0.45^{+0.12}_{-0.13}$	$365^{+52}_{-50}$	$1.10^{+0.03}_{-0.06}$	$0.42^{+0.06}_{-0.03}$	$0.52^{+0.03}_{-0.03}$	$0.56^{+0.02}_{-0.02}$	0.44
46	96	138	$2.13^{+0.29}_{-0.21}$	$1.43^{+0.20}_{-0.16}$	$227^{+34}_{-23}$	$1.39^{+0.06}_{-0.02}$	$0.48^{+0.01}_{-0.02}$	$0.56^{+0.02}_{-0.02}$	$0.61^{+0.02}_{-0.02}$	0.25
47	58	138	$4.37^{+0.71}_{-0.18}$	$4.37^{+0.55}_{-0.24}$	$203^{+36}_{-18}$	$1.36^{+0.00}_{-0.04}$	$0.81^{+0.00}_{-0.27}$	$0.54^{+0.01}_{-0.02}$	$0.63^{+0.01}_{-0.03}$	1.16
48	41	45	$1.20^{+0.22}_{-0.15}$	$0.67^{+0.08}_{-0.12}$	$440^{+106}_{-94}$	$1.24^{+0.32}_{-0.26}$	$0.50^{+0.07}_{-0.05}$	$0.57^{+0.05}_{-0.05}$	$0.63^{+0.03}_{-0.03}$	0.28
49	37	48	$3.04^{+0.61}_{-0.33}$	$2.52^{+0.53}_{-0.26}$	$336^{+79}_{-51}$	$1.42^{+0.04}_{-0.11}$	$0.55^{+0.02}_{-0.07}$	$0.58^{+0.02}_{-0.03}$	$0.62^{+0.02}_{-0.03}$	0.12
50	23	48	$22.36^{+5.34}_{-5.88}$	$22.01^{+5.15}_{-5.56}$	$1456^{+67}_{-186}$	$2.14^{+0.06}_{-0.14}$	$0.20^{+0.06}_{-0.00}$	$0.44^{+0.02}_{-0.01}$	$0.36^{+0.03}_{-0.01}$	4.31

All time-scales are in units of Myr. All length-scales are in units of pc.