Quantifying the effects of spatial resolution and noise on galaxy metallicity gradients

Ayan Acharyya [1,2][*] Mark R. Krumholz [1,2] Christoph Federrath,[1,2] Lisa J. Kewley,[1,2] Nathan J. Goldbaum[3] and Rob Sharp[1]

[1]*Research School of Astronomy and Astrophysics, Australian National University, Canberra, ACT 2611, Australia*
[2]*ARC Centre of Excellence for All Sky Astrophysics in 3 Dimensions (ASTRO 3D), Australia*
[3]*National Center for Supercomputing Applications, University of Illinois at Urbana-Champaign, 1205 W. Clark Str, Urbana, IL 61801, USA*

ABSTRACT

Metallicity gradients are important diagnostics of galaxy evolution, because they record the history of events such as mergers, gas inflow, and star formation. However, the accuracy with which gradients can be measured is limited by spatial resolution and noise, and hence, measurements need to be corrected for such effects. We use high-resolution (~ 20 pc) simulation of a face-on Milky Way mass galaxy, coupled with photoionization models, to produce a suite of synthetic high-resolution integral field spectroscopy (IFS) datacubes. We then degrade the datacubes, with a range of realistic models for spatial resolution ($2-16$ beams per galaxy scale length) and noise, to investigate and quantify how well the input metallicity gradient can be recovered as a function of resolution and signal-to-noise ratio (SNR) with the intention to compare with modern IFS surveys like MaNGA and SAMI. Given appropriate propagation of uncertainties and pruning of low SNR pixels, we show that a resolution of 3–4 telescope beams per galaxy scale length is sufficient to recover the gradient to ~ 10–20 per cent uncertainty. The uncertainty escalates to ~ 60 per cent for lower resolution. Inclusion of the low SNR pixels causes the uncertainty in the inferred gradient to deteriorate. Our results can potentially inform future IFS surveys regarding the resolution and SNR required to achieve a desired accuracy in metallicity gradient measurements.

Key words: galaxies: evolution – galaxies: ISM – ISM: abundances – H II regions.

1 INTRODUCTION

Metal enrichment of the interstellar medium (ISM) is of fundamental importance in understanding the formation and evolution of galaxies. The gas-phase oxygen abundance (henceforth referred to as metallicity) distributions of galaxies contain a wealth of information about their star formation histories (Maiolino & Mannucci 2019). For example, if the stars at smaller galactocentric radii formed earlier than those in the outskirts, as expected in the inside–out galaxy formation scenario, this would lead to a negative metallicity gradient in the disc, because the star-formation going on for a longer time in the central part of the disc would enrich the ISM more than in the outskirts. Interaction with another galaxy can cause gas inflow and mixing of metal enriched gas, thus leading to shallower metallicity gradients than isolated galaxies (e.g. Krabbe et al. 2008; Bresolin et al. 2009; Kewley et al. 2010; Rupke, Kewley & Barnes 2010; Rich et al. 2012; Miralles-Caballero et al. 2014; Vogt et al. 2015; Molina et al. 2017; Muñoz-Elgueta et al. 2018). Such flattening due to mergers has also been observed in

numerical models (Mihos & Hernquist 1994; Torrey et al. 2012; Fu et al. 2013; Zinchenko et al. 2015; Sillero et al. 2017).

It is well known that most disc galaxies exhibit a radial distribution of oxygen abundance with a negative gradient (e.g. Garnett & Shields 1987; Kennicutt & Garnett 1996; Garnett et al. 1997; Bresolin et al. 2002; Kennicutt, Bresolin & Garnett 2003; Bresolin 2007), and this has been further established by recent studies with large numbers of samples, for both nearby galaxies (e.g. Moustakas et al. 2010; Rupke, Kewley & Chien 2010; Cecil, Croom & The SAMI Galaxy Survey Team 2014; Sánchez et al. 2014; Sánchez-Menguiano et al. 2016; Molina et al. 2017; Belfiore et al. 2017; Poetrodjojo et al. 2018; Sánchez-Menguiano et al. 2018) and those at redshifts $z \sim 1$–2 (e.g. Molina et al. 2017). The steepness of the gradient appears to be correlated with a number of galaxy properties, e.g. the metallicity gradient–effective radius relation of discs (Diaz 1989), and the correlation between metallicity gradients and galaxy morphology (e.g. Zaritsky, Kennicutt & Huchra 1994; Martin & Roy 1994; Sánchez et al. 2014). Ho et al. (2015) provide a benchmark metallicity gradient of ~ -0.4 dex/R_{25}[1] for local field

---

[*] E-mail: ayan.acharyya@anu.edu.au

[1]Radius of the 25 mag arcsec^{-2} isophote in *B* band.

star-forming galaxies, suggesting co-evolution of the stellar and gas disc in local galaxies. Wuyts et al. (2016), on the other hand, report a lack of significant correlation between metallicity gradients and other global galaxy properties in a sample of ∼180 galaxies. Moreover, it is not well established if the shape of the metallicity profile in galaxies is universal (e.g. Ho et al. 2015) or depends on the stellar mass (e.g. Belfiore et al. 2017; Sánchez-Menguiano et al. 2018; Mingozzi et al. 2020). These disagreements, and the strength of the conclusions that one can potentially draw, show that it is important to measure metallicity gradients accurately.

However, measured metallicity gradients can be affected by the limitations of the instruments used and the quality of the spectroscopic observations, such as spatial resolution and signal-to-noise ratio (SNR). Resolution, noise, and similar effects are important because metallicity diagnostics are based on ratios of emission lines with intensities that have a complex, non-linear relationship with the true, underlying metallicity. Redistribution of flux from inner to outer regions of a galaxy due to beam smearing can introduce systematic biases in the line ratio profiles and therefore the metallicity gradient. The non-linearity of the metallicity diagnostics exacerbates this bias and the errors in the inferred metallicity. For example, recent studies find that metallicity gradients appear shallower at lower spatial resolution, irrespective of the metallicity indicator used (e.g. Yuan, Kewley & Rich 2013; Mast et al. 2014; Poetrodjojo et al. 2019).

Previous attempts to quantify this effect have employed either smearing of high spatial resolution observations with coarser point spread functions (PSFs; Yuan et al. 2013) or application of PSF convolution to model discs (Carton et al. 2017). The former method is ideal, in that one starts from real data and then degrades it. However, it is severely limited by data availability: in order to carry out a systematic study of the likely errors as a function of intrinsic metallicity gradient, resolution, and SNR, one would need to start from a catalogue of measurements with both very high spatial resolution and very high SNR, covering a range of intrinsic metallicity gradients. At present no such catalogue is available. The latter approach has only been employed in 'toy-models' of disc galaxies so far, which assume a smooth radial variation of the star formation rate (SFR) surface density and consequently of the nebular emission line profiles. Neither approach allows one to consider effects like beam-smearing between neighbouring regions of very different SFR densities or mean stellar population ages, or mixing together light from H II regions with very different densities or ionization parameters. A complete study of these effects using physically realistic disc models along with metallicity diagnostics commonly used in observations is still lacking. The aim of this paper is to provide quantitative results from such a study.

We produce synthetic integral field spectroscopy (IFS) observations from high-resolution simulations of disc galaxies that resolve the phase structure of the ISM and include realistic stellar feedback. We then perform spectral line fitting and metallicity measurements in order to understand how well we can reproduce the intrinsic metallicity gradient. Simulations have the unique advantage that we have complete control over the input physics and hence know the 'true' values of the physical properties. Converting these simulations to mock observations and treating the synthetic data cubes similar to an observed data cube allows us to directly compare theory with observations, and to explore the effects of observational limitations such as finite SNR and resolution. Such comparisons, by treating simulations and observations equally, are essential to understand the effect of instrument properties and observational parameters on metallicity gradients.

This paper is organized as follows. In Section 2, we briefly describe the simulations and the steps involved in translating them to synthetic IFU data cubes, including the sub-grid modelling of H II regions. Section 3 details the analysis pipeline we use to derive metallicities from the synthetic data. We present results in Section 4, followed by discussion of their implications in Section 5. Finally, we summarize our work and draw conclusions in Section 6.

We use a solar oxygen abundance $12 + \log(\text{O/H}) = 8.77$ based on the local galactic concordance scale (Dopita et al. 2016) throughout the paper unless otherwise stated. We assume a standard flat Λ cold dark matter cosmology with $H_0 = 70$ km s^{-1} Mpc^{-1} and matter density $\Omega_M = 0.27$ wherever required for our calculations.

2 METHODS FOR PRODUCTION OF SYNTHETIC OBSERVATIONS

Fig. 1 summarizes the procedure we use to produce our synthetic data cubes and then extract metallicity information from them. We explain the data generation method in this section and the analysis method in the following section. To produce our synthetic data, we start with the simulations (Section 2.1) and then model H II regions around each young star particle (Section 2.2). We then construct a MAPPINGS photoionization H II region model grid (Section 2.4) and use it to assign nebular line luminosities to each pixel of our simulated galaxy. To produce the full position-position-velocity (PPV) cube, we add the nebular fluxes to the stellar continuum derived from Starburst99 (Section 2.5). Thereafter, we convolve the datacube with a smoothing kernel to simulate finite observational resolution (Section 2.6) and spectrally bin the datacube, followed by addition of noise (Section 2.7) to simulate sky emission combined with instrumental effects.

2.1 Galaxy simulations

We use the isolated Milky Way (MW)-type disc galaxy simulations of Goldbaum, Krumholz & Forbes (2016, G16 hereafter), which are publicly available, in order to generate synthetic observables. The halo mass of $10^{12} M_\odot$ and stellar mass of $10^{10} M_\odot$ have been chosen to loosely resemble that of the MW. The bulge-to-disc ratio is 0.1 (hoon Kim et al. 2013), slightly less bulge-dominated than the MW (∼0.3 Bland-Hawthorn & Gerhard 2016), but not very different. The simulations include hydrodynamics, self-gravity, star formation, photoionization, and supernova feedback, a live stellar disc (i.e. rather than using a fixed stellar potential as is common in some galaxy simulations, in G16 the stellar potential is generated by stars that move self-consistently and thus are free to develop spiral structure, scatter off structures in the gas, etc.), and stellar plus dark matter halo. Their high spatial resolution (20 pc) means that even if we are unable to resolve individual H II regions, we have only few (∼10) H II regions within each cell. The simulations include photoionization heating using a Strömgren volume approximation, but do not include full ionizing radiative transfer. Consequently, we need to further use photoionization models to appropriately model the expansion of the H II regions (see Section 2.2) and the emission-line fluxes from them. G16 carry out simulations with three different gas fractions f_g (defined as ratio of gas to total baryonic mass within the disc): low gas fraction (LGF; $f_g = 0.1$), fiducial ($f_g = 0.2$) and high gas fraction (HGF; $f_g = 0.4$). For each case, we use the final time slice from the simulations. We show a comparison between the projected gas densities of the simulations with different gas fractions in Fig. 2 to demonstrate the higher gas content in the fiducial gas
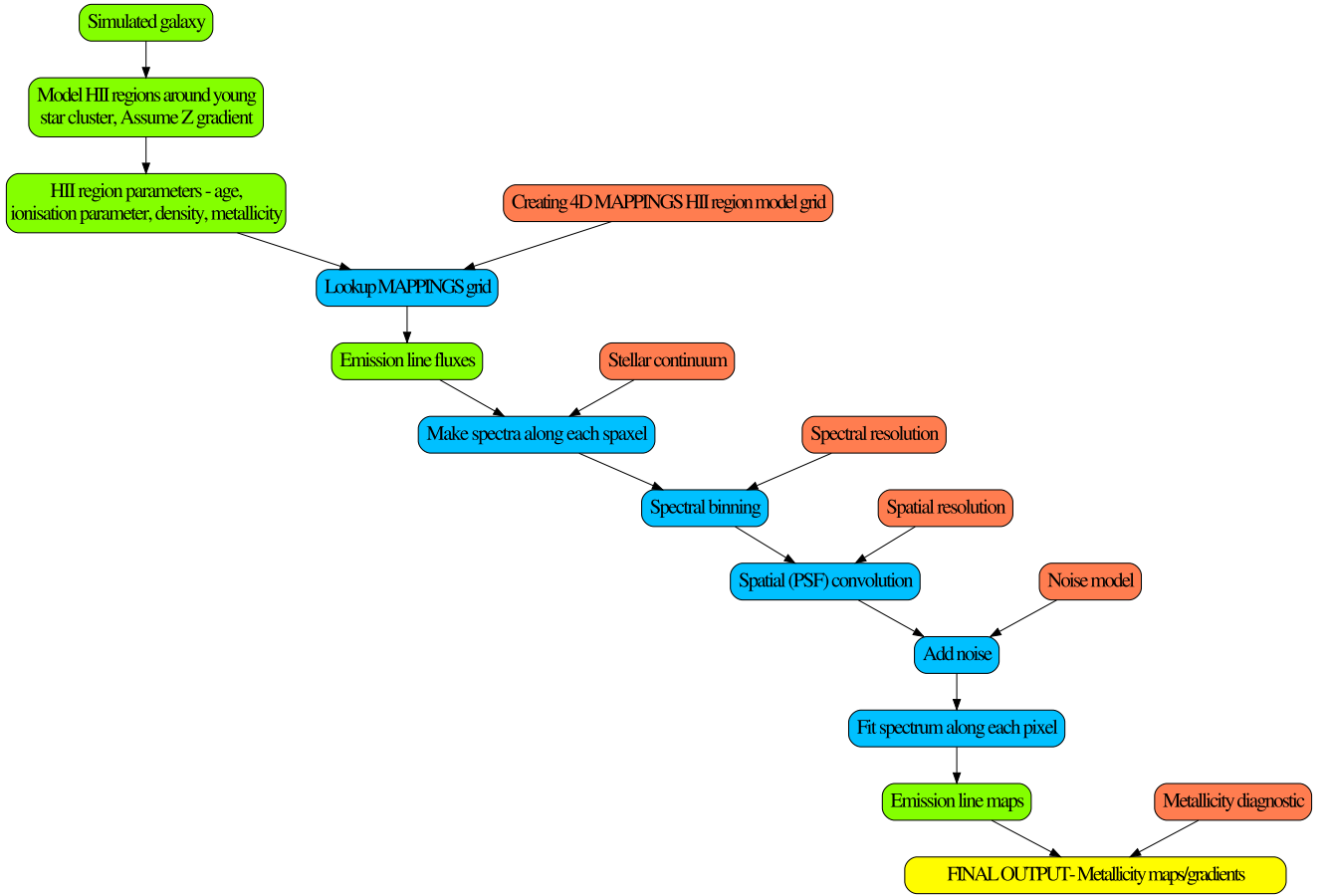
Figure 1. Flowchart for the methodology followed in this paper. The red blocks denote inputs and assumptions, blue blocks show the key steps involved, and the intermediate outputs are in the green blocks.
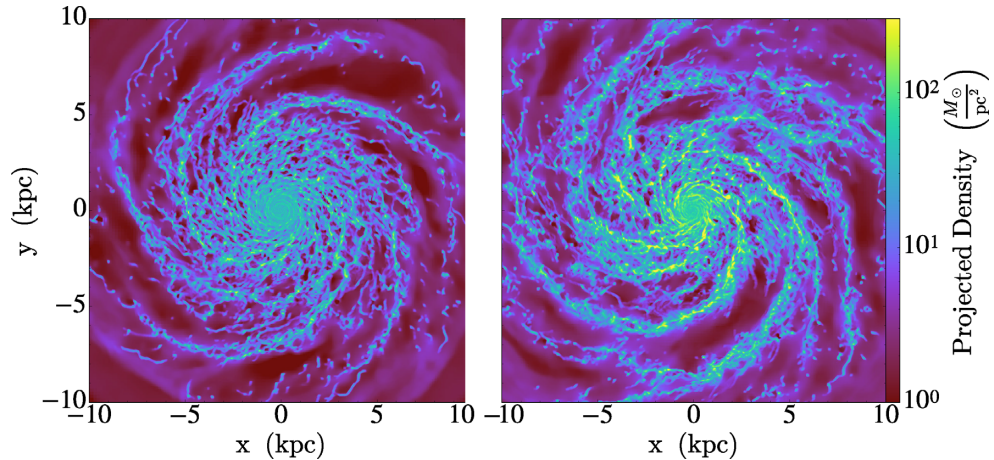

Figure 2. Comparison of the projected gas density of the LGF ($f_g = 0.1$; left) and fiducial gas fraction ($f_g = 0.2$; right) G16 simulations at the final times in the simulations.

fraction simulation (right) than the LGF scenario (left). The current MW gas fraction (∼15 per cent) is between the LGF and fiducial cases of the simulated galaxy. We compare results obtained from the fiducial and LGF versions of the G16 simulations in Appendix A; in the main text, we will focus on the fiducial case as that has an SFR (∼2 M$_\odot$ yr^{-1}) comparable to that of the MW. Our tests show that the results for the LGF case are qualitatively identical.

The G16 simulations and their successors using similar methodology (Fujimoto, Krumholz & Tachibana 2018; Fujimoto et al. 2019) are particularly well-suited for this experiment, because they have been subjected to extensive observational comparisons and show very good agreement with all diagnostics on ≳100 pc and larger scales. In particular, the simulations show excellent agreement with the observed mass spectrum and structural properties of
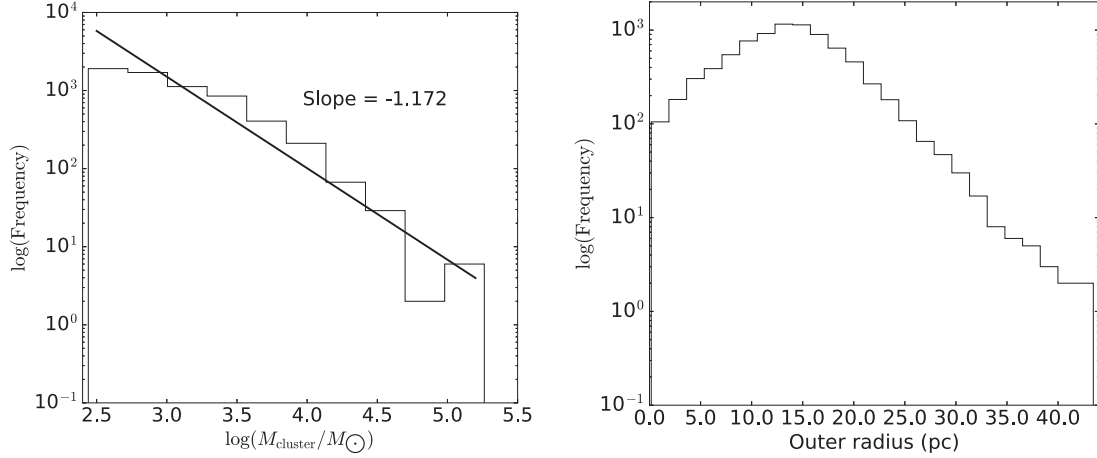
Figure 3. Distribution functions for star cluster masses (*left*) and outer radii (*right*) of H II regions for the fiducial gas fraction ($f_g = 0.2$) G16 simulations.

giant molecular clouds, and with the spatially resolved Kennicutt–Schmidt relation (see Fujimoto et al. 2019 for details); thus, the spatial distribution of star formation within them provides a good approximation to the MW.

2.2 Modelling H II regions

The first step in producing the synthetic observations is to identify the young (≤ 5 Myr) star particles in the simulations. We only include the young star particles in our calculations because the ionizing luminosities of stellar populations drop rapidly at ages above 5 Myr; for the stellar population models, we adopt in this work (see below), > 97 per cent of the total ionizing flux is emitted in the first 5 Myr.

Star particles that are in close proximity can potentially lead to merging of H II regions resulting in a superbubble being jointly driven by the host stellar associations. In order to treat the overlapping H II regions, we merge star particles in every 40 pc^{3}[2] cell by summing their luminosities and masses. We assign the luminosity-weighted mean of position, velocity, and age of the individual star particles to the merged particle. Henceforth, any reference to the star particles denotes the merged star particle in every cell.

To verify that our procedure results in a reasonable distribution of cluster sizes, in Fig. 3 we show the resulting cluster mass function (CMF), where a 'cluster' here refers to one of our merged particles. The CMF is close to a power-law distribution $dN/dlog\,M \propto M^{-1}$ from 300 M$_\odot$ (the mass of individual star particles in the G16 simulation) to $\approx 10^5$ M$_\odot$. This is in good agreement with the CMFs frequently observed in spiral galaxies – see the review by Krumholz, McKee & Bland-Hawthorn (2019) for a summary of recent measurements. However, to ensure that our results are not overly dependent on the merging procedure, in Appendix B we demonstrate that an alternative approach – not merging stars at all (so that each cluster is ≈ 300 M$_\odot$) – produces nearly identical results for the inferred metallicity gradient compared to when the stars are merged. For this reason, in the remainder of the main text we will focus only on the case of merging within 40 pc^{3} cells, yielding the CMF shown in Fig. 3. Given our list of merged clusters, we

follow the method of Verdolini et al. (2013) to compute for each star particle a series of quantities as follows.

2.2.1 Bolometric and ionizing luminosity

We run fixed mass ($M = 10^6$ M$_\odot$) Starburst99 (Leitherer et al. 1999) models with a Kroupa (2001) initial mass function (IMF) to obtain the bolometric and ionizing luminosity for star clusters aged 0–5 Myr, with a uniform linear spacing of 0.1 Myr. The IMF uses exponents 1.3 and 2.3 for mass ranges $0.1\,M_\odot \leq M \leq 0.5\,M_\odot$ and $0.5\,M_\odot \leq M \leq 120\,M_\odot$, respectively. We then assign a luminosity to each star particle in the simulation by linearly interpolating the Starburst99 outputs to the star particle age, and scaling linearly to the star particle mass. We note that this effect implicitly neglects stochastic sampling of the IMF, which can be significant in clusters with masses $\lesssim 10^{3.5}$ M$_\odot$ (e.g. da Silva, Fumagalli & Krumholz 2012). However, given our CMF (Fig. 3), such clusters contribute somewhat less than half of the total ionizing photon budget. Consequently, neglecting stochastic sampling has relatively modest effects as long as a typical observational beam contains enough total young stars that we expect it to contain a significant number of clusters with masses $\gtrsim 10^{3.5}$ M$_\odot$. This is the case for the resolution of all surveys to date with the exception of MUSE/PHANGS, and thus we will not worry about this complication further here (Arora et al. in preparation).

2.2.2 Stall radius

We define r_{stall} as the maximum radius up to which an H II region (assumed spherical, isotropic, and constant density) can expand, until its internal pressure (sum of gas pressure P_{gas} and radiation pressure P_{rad}) reaches a state of equilibrium with the ambient pressure P_{amb}, taken to be the pressure in the computational cell that contains the star particle. The stall radius is defined implicitly by the set of equations (Krumholz & Matzner 2009; Verdolini et al. 2013):

$$P_{amb} = P_{gas} + P_{rad}, \tag{1}$$

$$P_{rad} = \frac{\psi S \epsilon_0 f_{trap}}{4\pi r_{stall}^2 c}, \tag{2}$$

$$P_{gas} = \mu_H n_{II} m_H c_{II}^2, \tag{3}$$

[2]Although the resolution of the G16 simulations is 20 pc at the highest refinement level, we chose 40 pc as the base spatial resolution for this study for computational ease.

where $\psi = L/S\epsilon_0$ is the ratio of the star cluster's total bolometric luminosity to ionizing power, S is the total number of ionizing photons injected per unit time, and $\epsilon_0 = 13.6$ eV. $f_{\rm trap}$ is the factor by which radiation force is enhanced by trapping the energy within the expanding shell. The mean mass per H nucleus $\mu_{\rm H}$ is given by $1 + Y/X + Z/X \approx 1.4$, for Solar composition ($X = 0.73$, $Y = 0.25$, $Z = 0.02$), where X, Y, and Z are the mass fractions of hydrogen, helium, and all other elements, respectively. The mass of a hydrogen atom, the sound speed in the H II region, and the speed of light are denoted by $m_{\rm H}$, $c_{\rm II}$, and c, respectively. To fully specify the system, we also need to know the number density of H nuclei $n_{\rm II}$ in the H II region, which we derive by assuming ionization balance:

$$\phi S = \frac{4}{3}\pi r_{\rm stall}^3 \alpha_B n_{\rm II} n_{\rm e}, \tag{4}$$

$$n_{\rm e} = \left(1 + \frac{Y}{4X}\right) n_{\rm II}, \tag{5}$$

$$\Rightarrow n_{\rm II} = \sqrt{\frac{\phi S}{\frac{4}{3}\pi r_{\rm stall}^3 \alpha_B \left(1 + \frac{Y}{4X}\right)}}, \tag{6}$$

where ϕ is the fraction of ionizing radiation not being absorbed by the dust, and $\alpha_B = 2.59 \times 10^{-13}$ cm^{-3}s^{-1} is the recombination coefficient (assuming electron temperature of 10^4 K). Substituting the above expressions into equation (1) gives

$$P_{\rm amb} = \frac{a}{r_{\rm stall}^{\frac{3}{2}}} + \frac{b}{r_{\rm stall}^2}, \tag{7}$$

where $a = 0.65 m_{\rm H} c_{\rm II}^2 \sqrt{\phi S}$ and $b = \psi S \epsilon_0 f_{\rm trap}/4\pi c$ are constants. We solve equation (7) using a 1D Newton's method (initial guess provided by the analytically obtained solution to $P_{\rm amb} = P_{\rm gas}$) to get the values of $r_{\rm stall}$ for each H II region in the simulation. We take $\psi = 3.2$, $\phi = 0.73$, and $f_{\rm trap} = 2$, following Verdolini et al. (2013).

2.2.3 Instantaneous radius and density

To compute the instantaneous radius r of our H II regions, we use the analytic approximation provided by Krumholz & Matzner (2009), which interpolates between the similarity solutions describing the expansion of gas pressure driven ($r_{\rm gas}$) or radiation pressure driven ($r_{\rm rad}$) H II regions:

$$r = r_{\rm ch} \left(x_{\rm rad}^{(7-\kappa_\rho)/2} + x_{\rm gas}^{(7-\kappa_\rho)/2}\right)^{2/(7-\kappa_\rho)}, \tag{8}$$

$$x_{\rm rad} = \left(\frac{4 - \kappa_\rho}{2} \tau^2\right)^{1/(4-\kappa_\rho)}, \tag{9}$$

$$x_{\rm gas} = \left(\frac{(7 - 2\kappa_\rho)^2}{4(9 - 2\kappa_\rho)} \tau^2\right)^{2/(7-2\kappa_\rho)}, \tag{10}$$

where $\tau = \frac{t}{t_{\rm ch}}$, κ_ρ is the index of the power-law density distribution of the gas into which the H II region is expanding (for which we adopt $\kappa_\rho = 0$ in this work), and $r_{\rm ch}$ and $t_{\rm ch}$ are the characteristic spatial and time scales as described by equations (4) and (9) of Krumholz & Matzner (2009), given by (for the case $\kappa_\rho = 0$)

$$r_{\rm ch} = \frac{\alpha_B}{12\pi\phi} \left(\frac{\epsilon_0}{\kappa_B T_{\rm II}}\right)^2 f_{\rm trap}^2 \frac{\psi^2 S}{c^2}, \tag{11}$$

$$t_{\rm ch} = \sqrt{\frac{4\pi}{3} \frac{\rho_0 c}{f_{\rm trap} L} r_{\rm ch}^4}. \tag{12}$$

Here, ρ_0 is the ambient density of the material into which the H II region is expanding and $r_{\rm ch}$ is the characteristic radius where the radiation pressure and gas pressure are equal, with the gas pressure

being dominant at $r > r_{\rm ch}$ and radiation pressure dominating the $r < r_{\rm ch}$ regime, and $t_{\rm ch} = r_{\rm ch}/c_{\rm II}$, where $c_{\rm II}$ is the ionized gas sound speed. The *lesser* of r and $r_{\rm stall}$ is assigned as the outer radius $r_{\rm out}$ of the H II region. We provide the distribution of radii ($r_{\rm out}$) for all the H II regions in the simulation in Fig. 3. We show in Section 2.3 that the relationship between pressure and size for our simulated H II regions agrees within the uncertainties with that observed by Galactic H II regions by Tremblin et al. (2014). Once the radius is known, we can immediately compute the H II region density from ionization balance, following equation (6) with $r_{\rm stall}$ replaced by $r_{\rm out}$: $n_{\rm II} = \left[3\phi S/4\pi r_{\rm out}^3 \alpha_B (1 + Y/4X)\right]^{1/2}$. At this stage, we have the age, density, and outer radius of every young H II region in the simulation.

It should be noted that our calculation of the radius and density (and, below, the volume-averaged ionization parameter) implicitly assumes that H II regions have constant internal density. We emphasize that we make this approximation only for the purposes of computing the dynamics and that our calculation of line emission uses the more common isobaric approximation. The primary effect of assuming constant density is to somewhat lower the total recombination rate compared to that which would occur in regions with internal density variations, which in turn leads to a slight overestimate of the expansion rate. However, this is only a tens of per cent level effect (Bisbas et al. 2015).

2.3 Comparison of size distribution of H II regions with literature

It is difficult to make a direct comparison of our model H II region sizes with that of observed H II regions, particularly for extragalactic studies, due to spatial resolution limitations of the observations. Numerous extra-galactic studies report a H II region size distribution peaking between ~50 and 100 pc, which is generally the resolution limit at such distances. Tremblin et al. (2014) however study Galactic H II regions through radio recombination lines, which enables them to probe smaller length scales. Completeness remains an issue, since they have finite sensitivity, but we can minimize the problem by not comparing the absolute size distributions (which are sensitive to completeness) of our simulated H II regions to the data, but instead compare the pressure–radius relation. This should be less sensitive to completeness, but none the less captures the physics of H II region expansion. We make this comparison in Fig. 4, which shows that a majority of our models lie reasonably well within the parameter space spanned by the observations. However, we emphasize that a direct comparison of size distribution between the two is difficult to carry out without accounting for sensitivity and completeness issues in detail – which is beyond the scope of this paper.

2.4 Constructing the 4D MAPPINGS grid

Once we have the physical size and density for each H II region in the simulation, we need to compute a grid of H II region models in order to be able to read off the nebular emission-line fluxes. The strength of emission lines originating from a H II region depends on the shape of its driving radiation field, density structure, ionization structure, and chemical composition. The driving radiation field, in turn, depends on the age and mass distribution of the stars in the stellar association that produces the H II region. Hence, we need to compute a 4D grid of age, metallicity Z, volume-averaged ionization parameter $\langle U \rangle$, and density $n_{\rm II}$. We know the age of the star cluster from the simulation and the density from our modelling
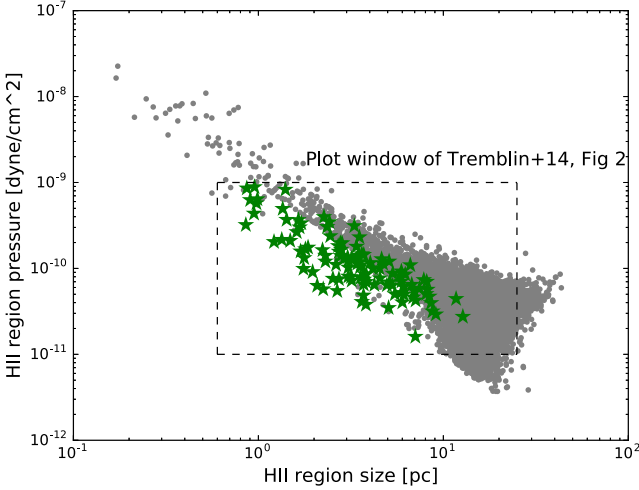
Figure 4. Pressure–radius parameter space of our model H II regions (*grey circles*) compared with observed Galactic H II regions from Tremblin et al. (2014) (*green stars*). The models shown here are based on a radially varying Ω, which, as demonstrated in Section 1.1 of the SOM, does not produce any qualitative difference from our fiducial $\Omega = 0.5$ models.

(Section 2.2). Next, we need to determine the ionization parameter and metallicity of each H II region.

2.4.1 Ionization parameter

The ionization parameter U is defined as ratio of photon to hydrogen number density. In order to determine the volume-averaged ionization parameter $\langle U \rangle$, we introduce a parameter Ω (following Yeh & Matzner 2012) as a measure of the relative volume occupied by the stellar wind cavity to the ionized bubble:

$$\Omega = \frac{r_i^3}{r_{\text{out}}^3 - r_i^3}, \tag{13}$$

where r_i is the inner radius of the ionized shell. A high value of Ω corresponds to an H II region that is wind-dominated and low Ω implies domination by either radiation or gas pressure. Throughout the main text of the paper, we assume a fiducial value of $\Omega = 0.5$, as suggested by Yeh & Matzner (2012) for radiation-confined dust-limited H II region shells (their fig. 1). Yeh & Matzner (2012) also point out that H II regions within the central 500 pc of M83, NGC 3256, NGC 253 typically have $\Omega \sim 0.5$ (their figs 8 and 9). We discuss the implications of this choice in Section 5.1.3. Assuming the gas density n_{II} to be constant throughout, the total number of ionizing photons per unit time passing through a shell at a distance r is given by (Draine 2011b)

$$Q(r) = Q_0 \left[1 - \left(\frac{r}{r_s} \right)^3 + \left(\frac{r_i}{r_s} \right)^3 \right], \tag{14}$$

where Q_0 is the rate of emission of ionizing photons from the driving star cluster and r_s is the Stromgen radius given by

$$r_s = \left(\frac{3 Q_0}{4\pi \alpha_B n_{\text{II}}^2} \right)^{1/3}. \tag{15}$$

The local ionization parameter U at a distance r from the stellar source is

$$U = \frac{Q(r)}{4\pi r^2 c n_{\text{II}}}, \tag{16}$$

which, when volume-averaged over the entire H II region, relates to n_{II}, Q_0, and Ω as

$$\langle U \rangle = \left(\frac{81 \alpha_B^2 n_{\text{II}} Q_0}{256\pi c^3} \right)^{1/3} \left((1 + \Omega)^{4/3} - \left(\frac{4}{3} + \Omega \right) \Omega^{1/3} \right). \tag{17}$$

Relaxing the assumption of constant density would lead to a slightly lower mean ionization parameter, since it would cause greater recombination losses, but given that density variations in H II regions are only at the tens of per cent level (Bisbas et al. 2015), this is a minor effect.

2.4.2 Metallicity

The G16 simulations do not have an intrinsic metallicity (Z) gradient because they do not track the chemical evolution of the gas. None of the current simulations that build up a metallicity gradient self-consistently over cosmological time-scales have high enough spatial resolution for this study. We choose G16 simulations because their structure is realistic on scales smaller than the beam of any survey we intend to imitate. Since the purpose of this paper is to study the effect of spatial resolution on metallicity gradients, we therefore 'paint' a metallicity gradient on the G16 galaxy. Having control of the metallicity gradient gives us the advantage of being able to vary it so that we can study the effect of resolution on a broad range of gradients. We assume smooth radial exponential profiles for Z and choose a value of the gradient $\nabla_r \log Z$ and the central metallicity Z_0. Thus, for every H II region at a given radius r we can compute its metallicity. In the light of recent works having reported azimuthal variations in metallicity gradients (Ho et al. 2018; Kreckel et al. 2019), we note that a smooth radial gradient may not be the most realistic representation. However, we choose this profile for simplicity and to quantify the effects of spatial resolution and noise with fewer free parameters.

Having defined the parameter space that describes our H II regions, we next use the MAPPINGS v5.1 code (Sutherland et al. 2013) to run a series of photoionization models to sample it. The parameter space sampled is as follows. The models are computed at metallicity values of 0.05, 0.1, 0.2, 0.3, 0.5, 1.0, 2.0, 3.0, and 5.0 in units of solar metallicity $Z_{\odot}$[3] and six uniformly spaced ages, ranging from 0 to 5 Myr (both boundaries included). We use four $\log (\langle U \rangle)$ values ranging from -4 to -1 and six $\log (n_{\text{II}})$ values uniformly spaced in the range -1 to 6 (in units of cm^{-3}). The MAPPINGS code requires as input the metallicity Z and total rate of ionizing photons emission rate Q_0, together with a series of quantities at the H II region inner edge: the ionization parameter U_i, pressure P, and temperature T. We determine these quantities from the parameters making up the 4D grid – Z, $\langle U \rangle$, n_{II} and age – as follows. Substituting the Ω, $\langle U \rangle$, and n_{II} values into equations (17), (15), and (14) we solve for Q_0 and r_i. We then derive the ionization parameter at the inner edge of the H II region (U_i) by substituting r_i in equation (16). The pressure in the photoionized gas $\log (P/k)$ is derived from n_{II} using $\log (P/k) = \log (n_{\text{II}}) + \log T$, where T is the electron temperature inside the H II region, which for the purposes of computing the input pressure we assume to

---

[3]We have verified that this grid is sufficiently well sampled that interpolation errors do not dominate our error budget by running the pipeline we describe below on a case with no spatial smoothing or noise. When we do so we find that we are able to recover metallicities to within a few per cent, which is the accuracy of our metallicity diagnostics, so that interpolation is not contributing significantly to the error budget.

be constant at 10^4 K.[4] The quantities Z, U_i, $\log(P/k)$, T, and Q_0 are provided to MAPPINGS to generate each H II region model. For each MAPPINGS model, we record the line flux for the [N II] $\lambda6584$, H α, [S II] $\lambda6717$, and [S II] $\lambda6730$ lines. We then assign line luminosities to each simulation H II region by linearly interpolating the line flux on our grid of Z, $\langle U \rangle$, n_{II}, age values.

2.5 Constructing the synthetic IFS cube

The next step in our synthetic data construction procedure is to produce a synthetic integral field spectrograph (IFS) data cube by adding the nebular line fluxes output from MAPPINGS on top of an underlying stellar continuum provided by Starburst99 (Leitherer et al. 1999). The Starburst99 model parameters are described in Section 2.2. We ensure that both MAPPINGS and Starburst99 models are based on the same stellar parameters. First, we run Starburst99 models for a range of ages from 0 to 5 Myr. Then, we compute the driving stellar continuum radiation field for each H II region in the simulation by interpolating the spectra output by Starburst99 linearly in age and rescaling to the cluster mass. The interpolation in age is performed at a finely sampled resolution of 0.1 Myr. We sample the continuum at a resolution of $10\,\mathrm{km\,s^{-1}}$ within a $\pm\,500\,\mathrm{km\,s^{-1}}$ window around the central wavelengths of emission lines, and at $20\,\mathrm{km\,s^{-1}}$ outside that window. The above wavelength sampling for the continuum is sufficient because the Starburst99 continuum itself has poor spectral resolution.

Next, we add the emission line fluxes for each H II region, computed in Section 2.4, as Gaussians with a velocity dispersion $\sigma_v = 15\,\mathrm{km\,s^{-1}}$ centred on the rest wavelength of each line, Doppler shifted by the velocity of the cell in which the H II region is located; for the purposes of computing the Doppler shift, we assume that the galaxy is being viewed face-on. We assume $\sigma_v = 15\,\mathrm{km\,s^{-1}}$ as a reasonable estimate of thermal and turbulent broadening (Krumholz & Burkhart 2016), although this depends on local turbulent properties of the individual galaxy observed (Federrath et al. 2017; Zhou et al. 2017). Our synthetic data cubes have a rest-frame wavelength range of 6400–6783 Å. This range is motivated by the fact that we only need the emission lines H α, [N II] $\lambda6584$, [S II] $\lambda6717$, and [S II] $\lambda6730$ for the metallicity diagnostic (Section 3.2) we employ, all of which lie within the said range.

A sufficiently high spectral resolution continuum, such as those in observed data, is marked with stellar absorption features – particularly the Balmer trough near the H α line. Observers adopt sophisticated methods to fit that trough in order to measure the H α emission-line flux. However, for our work a simplistic continuum fitting routine is sufficient because the low-resolution continuum we use does not include the aforementioned trough. In reality, a more complicated fitting algorithm may be able to measure the H α line with greater accuracy, which our method is currently incapable of doing. However, because we consider only young (<5 Myr) stellar populations, the equivalent width (W) of the H α absorption is expected to be $W \sim 2$–3 Å (Groves, Brinchmann & Walcher 2012), whereas that of the H α emission in the bright pixels of our simulation is $W \sim 40$ Å. Thus, the Balmer absorption is a small fraction of the emission-line strength and failing to account for it would only affect the flux at a few per cent level, which is smaller

than the uncertainties introduced by other steps, and hence can be ignored.

To produce the ideal IFS cube, we construct a grid spanning $30\,\mathrm{kpc} \times 30\,\mathrm{kpc} \times 380$ Å with each spaxel being $40\,\mathrm{pc} \times 40\,\mathrm{pc} \times 10\,\mathrm{km\,s^{-1}}$, and sum the nebular and stellar light from all the H II regions in each spatial pixel. Note that the G16 simulations have a spatial resolution of 20 pc at the maximum refinement level. However, we assume a base cell size of 40 pc in the interest of minimizing computation cost. The choice for the base resolution does not affect our results, as long as it is considerably smaller than our smallest PSF size ($\sim$200 pc). The end result of this step is an ideal ($10\,\mathrm{km\,s^{-1}}$ spectral resolution, 40 pc spatial resolution, noise free) 3D IFS datacube.

The next set of steps is aimed at incorporating instrument effects into the ideal cube. First, we rebin the spectral dimension of this cube down to a resolution of $30\,\mathrm{km\,s^{-1}}$, comparable to the resolutions of modern IFS systems. We then proceed to coarsen the spatial resolution of the cube, a process we describe in the next section.

2.6 Simulating finite spatial resolution

We place our simulated galaxy at a redshift $z = 0.04$, which is typical of the SAMI survey (Green et al. 2018).[5] Based on this distance, we convert the instrument PSF to a physical spatial resolution element, defined as the full width at half-maximum (FWHM) of the beam in the source plane of the galaxy. We then spatially convolve each wavelength slice with a normalized 2D Moffat kernel (Moffat 1969). The Moffat profile is defined by a width σ and power index β, which are related as follows:

$$\sigma = \frac{\mathrm{FWHM}}{2\sqrt{2^{\frac{1}{\beta}} - 1}}, \tag{18}$$

where FWHM is the full width at half-maximum for the kernel. Trujillo et al. (2001) report that the above analytical form provides the best fit to the PSF predicted from atmospheric turbulence theory for $\beta = 4.7$, and we use this value of β throughout this paper. To reduce computational cost, we truncate the Moffat kernel at 5σ, which incorporates 99.99 per cent of the total power. After the convolution, we resample each slice to 2 spaxels per beam, i.e. we choose our pixel scale to be half of the FWHM, in order to mimic typical IFS datacubes. We ensure that both the convolution and resampling procedures conserve flux to machine precision. Fig. 5 shows a comparison between H α maps when smoothed with a 0.15 arcsec (left) and 1 arcsec (right) wide PSF. No noise has been included in these cases in order to highlight the effect of spatial smoothing.

2.7 Simulating noise

Sky noise (including contribution from OH airglow and continuum) has a non-negligible impact on the otherwise surface-brightness limited ground-based IFS observations, and thus it is important that our noise model capture this properly. Read-noise is often not the dominant source of noise in IFS data but Poisson noise, however, is important because it depends on the signal. In order to account for all these different noise sources, we use realistic wavelength-dependent noise, with the wavelength dependence taken from an observed sky

---

[4]The actual temperature profile is computed self-consistently by MAPPINGS; we only adopt a temperature here for the purposes of turning our mean density into a starting pressure.

[5]The choice of the distance is immaterial because we discuss our results in the reference frame of the galaxy, i.e. our models span a wide range of PSFs expressed as number of beams per scale length of the galaxy
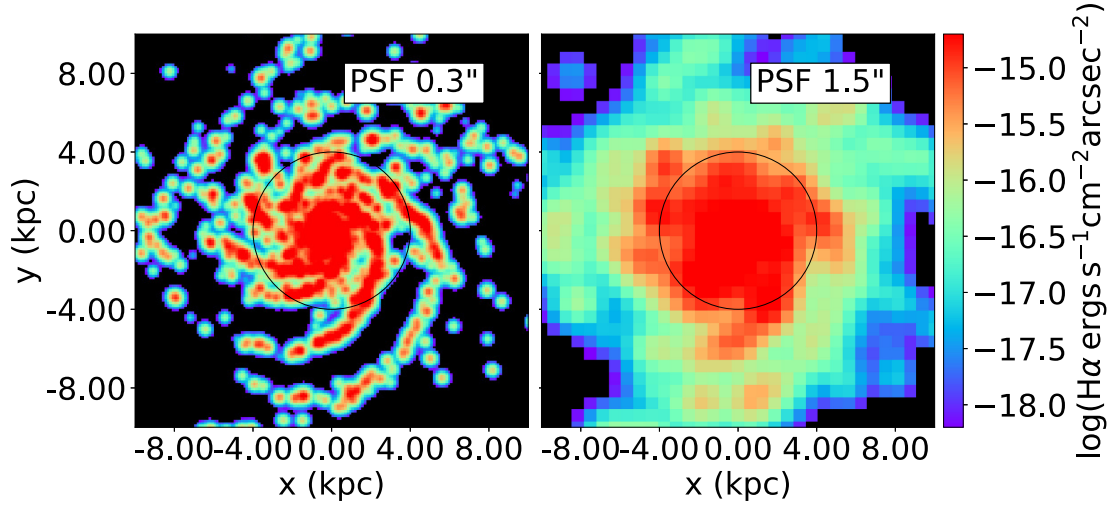
Figure 5. H α maps produced by using a small PSF (left; FWHM = 0.3 arcsec = 0.25 kpc, on source frame of galaxy) and large PSF (right; FWHM = 1.5 arcsec = 1.25 kpc, on source frame) in order to highlight the effect of convolution with a Moffat profile. The thin black circle in each map marks the scale radius of the disc, which in this case is 4 kpc. Both these cases are free of noise in order to isolate the effect of spatial smoothing.
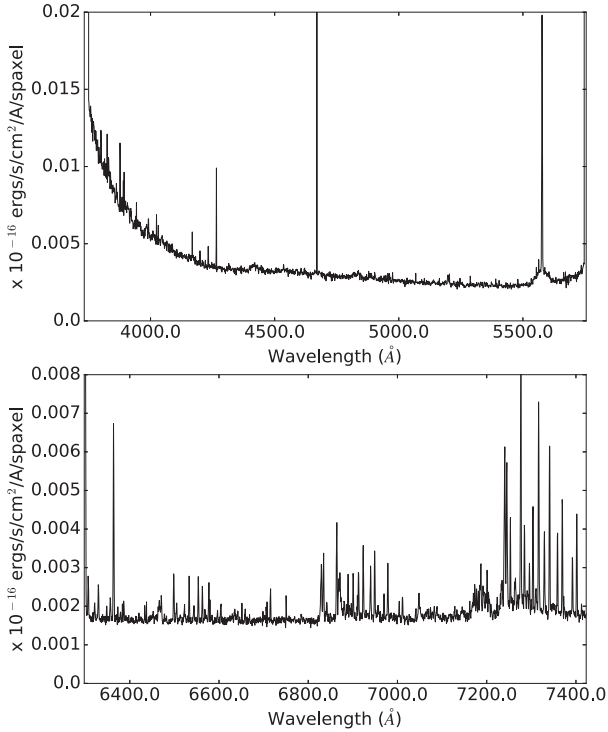

Figure 6. SAMI noise spectra for blue (top) and red (bottom) wavelengths. This includes noise contribution from the sky background as well as instrumental read-noise. We use this wavelength-dependent noise template to implement our noise model (Section 2.7).

frame of SAMI observations (Croom et al. 2012; Allen et al. 2015; Green et al. 2018; Scott et al. 2018). Fig. 6 shows the SAMI noise spectra we used for both the blue (top) and red (bottom) channel. The blue channel noise spectra is relevant only for the additional analyses we conducted that is discussed in Appendix A. Being a representative variance array from the outer regions of a SAMI observation, this noise spectrum captures a sky-noise limited data

set with some contribution from the detector read-noise (we include Poisson noise next). We apply this noise spectrum at every position in the synthetic IFS cube, with a normalization level chosen as follows. For every wavelength slice of the IFS cube, we define

$$f_{\text{sky}}(\lambda) = \frac{I_{\text{sky}}(\lambda)}{N_{\text{sky}}} \tag{19}$$

as the sky noise contribution at that wavelength, where $I_{\text{sky}}(\lambda)$ is the sky intensity as a function of wavelength, and N_{sky} is the sky intensity at the wavelength of the [N II] λ6584 line, i.e. f_{sky} is normalized to have unit intensity at the centre of the [N II] λ6584 line. We then take the noiseless [N II] λ6584 emission-line map (Section 3.1) and compute S_{gal} – the mean intensity in an annulus between 90 per cent and 110 per cent of the scale length r_{scale} = 4 kpc of the galaxy. The effective (half-light) radius r_{e} for the galaxy is 3 kpc i.e. r_{e} = $0.75 r_{\text{scale}}$. S_{gal} is used as the normalization value for the noise across the whole field of view (FoV). In order to produce an observation with a specified SNR, we take the zero noise spectrum at every spatial pixel for every wavelength slice and add a random number drawn from a Poisson distribution with standard deviation:

$$N = f_{\text{sky}}(\lambda) \times \frac{S_{\text{gal}}}{\text{SNR}}. \tag{20}$$

This allows for a signal-dependent Poisson component in our noise, which is otherwise absent in the SAMI noise spectrum we used. Thus, by definition, the mean SNR within an (0.9–1.1) × r_{scale} annulus of the [N II] λ6584 emission-line map should be equal to the input SNR. The SNR measured in the datacubes is generally a bit smaller (∼80 per cent of the input SNR) due to the additional uncertainty introduced by the continuum subtraction process (Section 3.1), but this effect is not large. Throughout the paper, we quote SNR values as SNR per spaxel. Fig. 7 shows a comparison between H α maps without noise (left) and with noise added (right, SNR = 5).
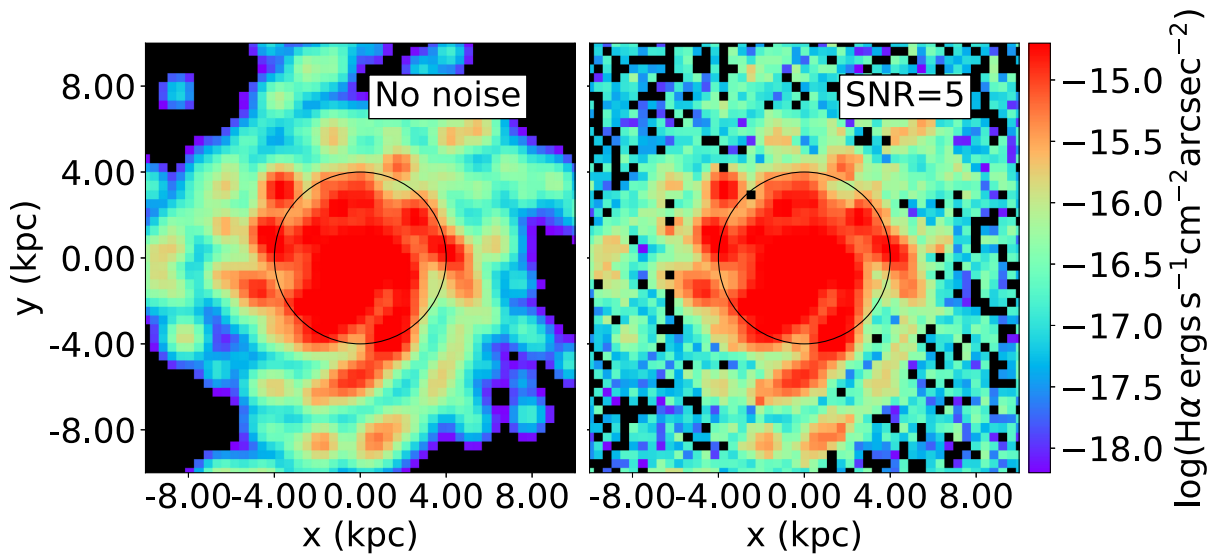
Figure 7. Same as Fig. 5, but this time comparing between noisy (SNR = 5; right) and without noise (left) case. Both images use a 1 arcsec PSF. The signal on the outskirts of the noisy image is relatively more affected than the core region. This is expected because the outer regions of the galaxy have fewer H II regions and hence poorer signal.

3 EXTRACTING OBSERVABLES

3.1 Emission-line fitting

Most established IFS data reduction pipelines, e.g. LZIFU for SAMI survey (Ho et al. 2016) and Pipe3D for CALIFA survey (Sánchez et al. 2016), are able to perform emission-line fitting but are custom-designed for the particular IFU survey. Moreover, the degree of sophistication that these pipelines operate at for certain steps such as the continuum fitting is not necessary for our simulation. The focus of our work is to investigate the relative change in emission-line ratios with PSF sizes. As such, we perform our own line fitting procedure without resorting to any conventional pipeline, which might require modifying a complex software that was designed for a specific kind of data.

In order to measure the emission-line fluxes, we first fit the continuum and subtract it out. The continuum fitting routine masks out a ± 150 km s^{-1} region around all expected spectral features. Then a rolling average is applied on the masked flux, followed by a smooth cubic spline interpolation to derive the stellar continuum. We estimate the continuum uncertainty as the root mean square (rms) deviation of the fitted continuum from the masked flux. As such, the continuum uncertainty along each spatial pixel is wavelength independent.

Next, we simultaneously fit all neighbouring spectral lines with one Gaussian profile per line using a non-linear least-squares method. A neighbour is defined as any line within ± 5 spectral resolution elements of its adjacent line. For the spectral resolution of 30 km s^{-1} used throughout this paper, the [N II] $\lambda 6584$ and H α lines are well resolved and hence fit individually. We constrain the fit to the centroid to a wavelength range $\lambda_0 \pm 3 + \delta z/(1 + z)$, where λ_0 is the line centre wavelength and $\delta z = 10^{-4}$. We similarly constrain the line width to lie within upper and lower bounds of 300 km s^{-1} and one spectral resolution element, respectively. The amplitudes of the Gaussian fits were allowed to vary freely. We propagate the continuum uncertainty and the noise in the spectra through to the fitting routine. Throughout this paper, we have assumed a spectral resolution of 30 km s^{-1}, implying that all the emission lines are

at least marginally resolved. Unlike some observational studies, we do not force the different emission lines to have a common width because we wish to isolate the effect of spatial resolution by reproducing the metallicity gradient with least possible constraints (i.e. maximum freedom) in the spectral fitting step. In principle, invoking constraints from atomic physics and forcing lines to have a common width may allow the use of low SNR spaxels, thereby maximally utilizing all the data. However, such assumptions do not recognize the fact that there could be multiple H II regions along each line of sight. For each spatial pixel, we measure the fluxes for the H α, [N II] $\lambda 6584$, and [S II] $\lambda\lambda 6717,30$ lines from the fitted Gaussian parameters, along with the corresponding flux uncertainties. Thus, we obtain 2D (spatial) maps for each of these emission lines. Fig. 8 shows example maps for a PSF of 1 arcsec and an SNR of 8.

In order to verify that our simulated maps provide a reasonable approximation of real galaxies, we can compare to an observed map from the SAMI galaxy survey (Green et al. 2018; Scott et al. 2018) selected to have a stellar mass ($M \sim 10^{11}$ M$_\odot$) and SFR ($\dot{M} \sim 1$ M$_\odot$ yr^{-1}) comparable to the G16 simulation (which is modelled on the MW). The bottom panel of Fig. 8 shows the emission line maps used in our work – [N II] $\lambda 6584$, Hα and [S II] $\lambda\lambda 6717,30$ – for SAMI galaxy N209807 (SAMI DR2: Scott et al. 2018). A comparison of the two upper and lower sets of panels demonstrates that we are able to reproduce the emission line flux values reasonably well. The limited spatial resolution of SAMI washes out the clumpy structure of the ISM and spiral arms in the massive galaxies. We present a similar low-resolution version of our emission-line maps from the simulations in Fig. 8. We thus demonstrate that our simulation is a reasonable model to test the effects of SNR and spatial resolution on metallicity gradients.

3.2 Metallicity diagnostic

In the main text, we use the Dopita et al. (2016, hereafter D16) diagnostic to derive the metallicity at each pixel; we present results for the alternative Kewley & Dopita (2002, hereafter KD02)
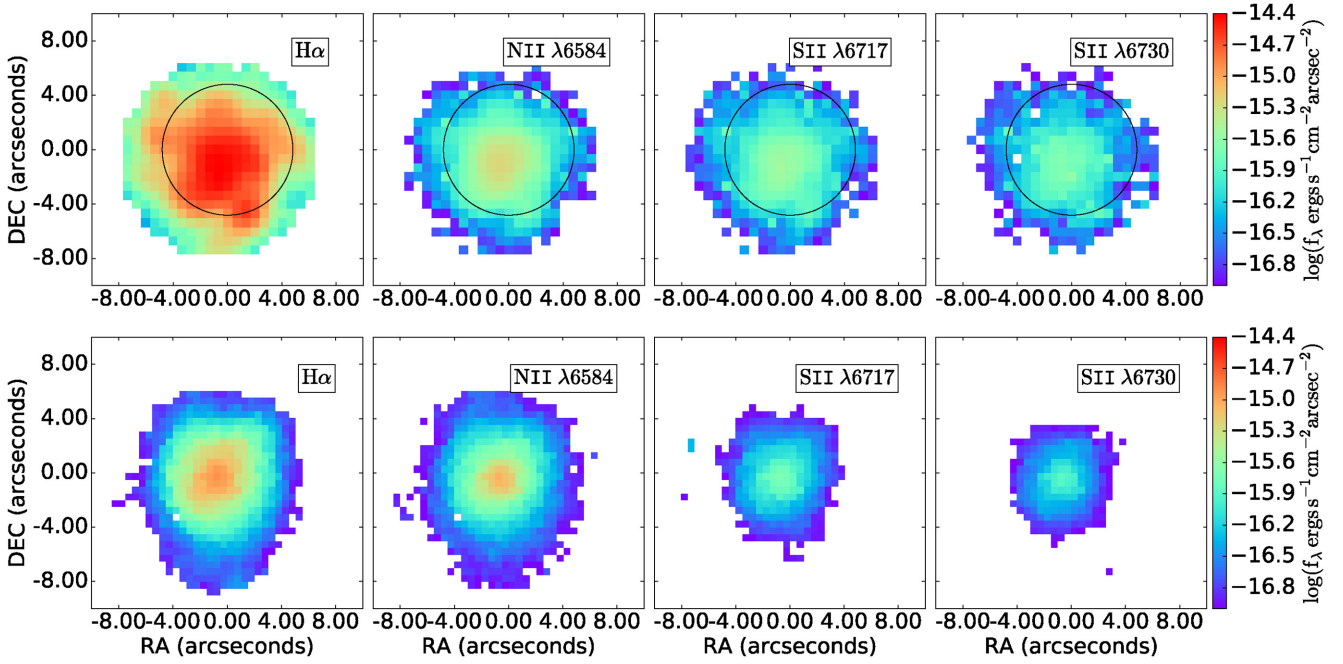
Figure 8. *Top*: Emission-line maps of H α (top left), [N II] λ6584 (top right), [S II] λ6717 (bottom left), and [S II] λ6730 (bottom right) derived by Gaussian fitting (Section 3.1) for a spatial resolution of 2 arcsec PSF (corresponding to SAMI; Green et al. 2018; Scott et al. 2018) and a mean SNR/spaxel of 8 at the scale length of the [N II] λ6584 map, indicated by the thin black circle. *Bottom*: Corresponding emission-line maps of an observed SAMI galaxy N209807 for comparison. The simulated maps in the top panel have been tailored to the SAMI footprint for ease of comparison.

diagnostic in Appendix A, and show that they are qualitatively similar. We analytically propagate the uncertainties in the line fluxes to obtain the corresponding uncertainty in metallicity. The D16 calibration uses the Hα, [N II] λ6584, [S II] λ6717 and [S II] λ6730 nebular emission lines and the relation

$$\log{(O/H)} + 12 = 8.77 + y + 0.45(y + 0.3)^5 \qquad (21)$$

where

$$y = \log{\frac{N\,II}{S\,II}} + 0.264\log{\frac{N\,II}{H\alpha}}. \qquad (22)$$

The main advantage of the D16 diagnostic is that reddening corrections are not important because the lines are closely spaced in wavelength. In reality, unresolved H II regions with different brightness and extinction are a concern for observational studies. This argues for using reddening-insensitive metallicity diagnostics, which is the approach we have followed here. We note, however, that simply using reddening-insensitive line ratios is insufficient to account for the differential extinction across H II regions. Accurately, quantifying the effect of differential extinction would require simultaneously capturing the structure of H II regions and the dusty shells around them which is beyond the scope of this paper. We therefore acknowledge the absence of dust as a potential limitation.

4 RESULTS

In this section, we present our results for the fiducial gas fraction ($f_g = 0.2$) simulation using the D16 metallicity diagnostic, assuming wind parameter $\Omega = 0.5$. In Appendix A, we present our detailed parameter study to demonstrate that our main result is robust to different values of f_g and Ω, and different choices of metallicity diagnostic. These choices impact the overall accuracy of the inferred metallicity gradient, but the variation with spatial resolution and SNR remains qualitatively unaffected. The full tables and figures

associated with Appendix A are provided as Supplementary Online material (SOM).

4.1 Model grid

We compute a grid of synthetic IFU datacubes, covering a range of four true metallicity gradients (-0.1, -0.025, -0.05, -0.01 dex kpc^{-1}), 14 PSF sizes [PSF $= (0.05, 0.15, 0.3, 0.5, 0.7, 0.8, 1.0,$ $1.2, 1.3, 1.5, 1.7, 2.0, 2.3, 2.5$ arcsec), approximately corresponding to (96.4, 32.1, 16.1, 9.6, 6.9, 6.0, 4.8, 4.0, 3.7, 3.2, 2.8, 2.4, 2.1, 1.9) beams per scale length and (0.04, 0.1, 0.2, 0.4, 0.6, 0.7, 0.8, 1.0, 1.1, 1.2, 1.4, 1.7, 1.9, 2.1) kpc, respectively, sampled at 2 pixels per beam] and seven values of intrinsic SNR of the datacube defined at the scale radius [SNR $= (1, 3, 5, 8, 10, 30, 300)$]. The range of SNR values comfortably covers the observed parameter space for nearby galaxy surveys, e.g. MaNGA (Mingozzi et al. 2020) and SAMI. We run our models up to large PSF sizes in order to sample the coarse end of the resolution satisfactorily. For each grid point, i.e. for each combination of the above parameters, we compute ten datacubes, the noise being drawn from a different random realization in each case. Throughout the paper, we report only the mean metallicity gradient measurement of the 10 realizations (and the standard deviations as the uncertainties) because the formal uncertainties of the fitted gradient are generally smaller than the variation between different realizations.

Fig. 9 illustrates the next step for a combination of two different PSF sizes (top and bottom) and SNR values (left and right). For each datacube, once we have the metallicity map (Section 3.2), we fit the pixels (grey points) with a simple linear [in log ($Z/Z_\odot$)] radial metallicity profile (blue line). All pixels are used for the fitting unless an SNR cut-off is imposed (see below). We propagate the uncertainties in the emission-line fluxes through the metallicity diagnostic in order to obtain uncertainties on the metallicity values.
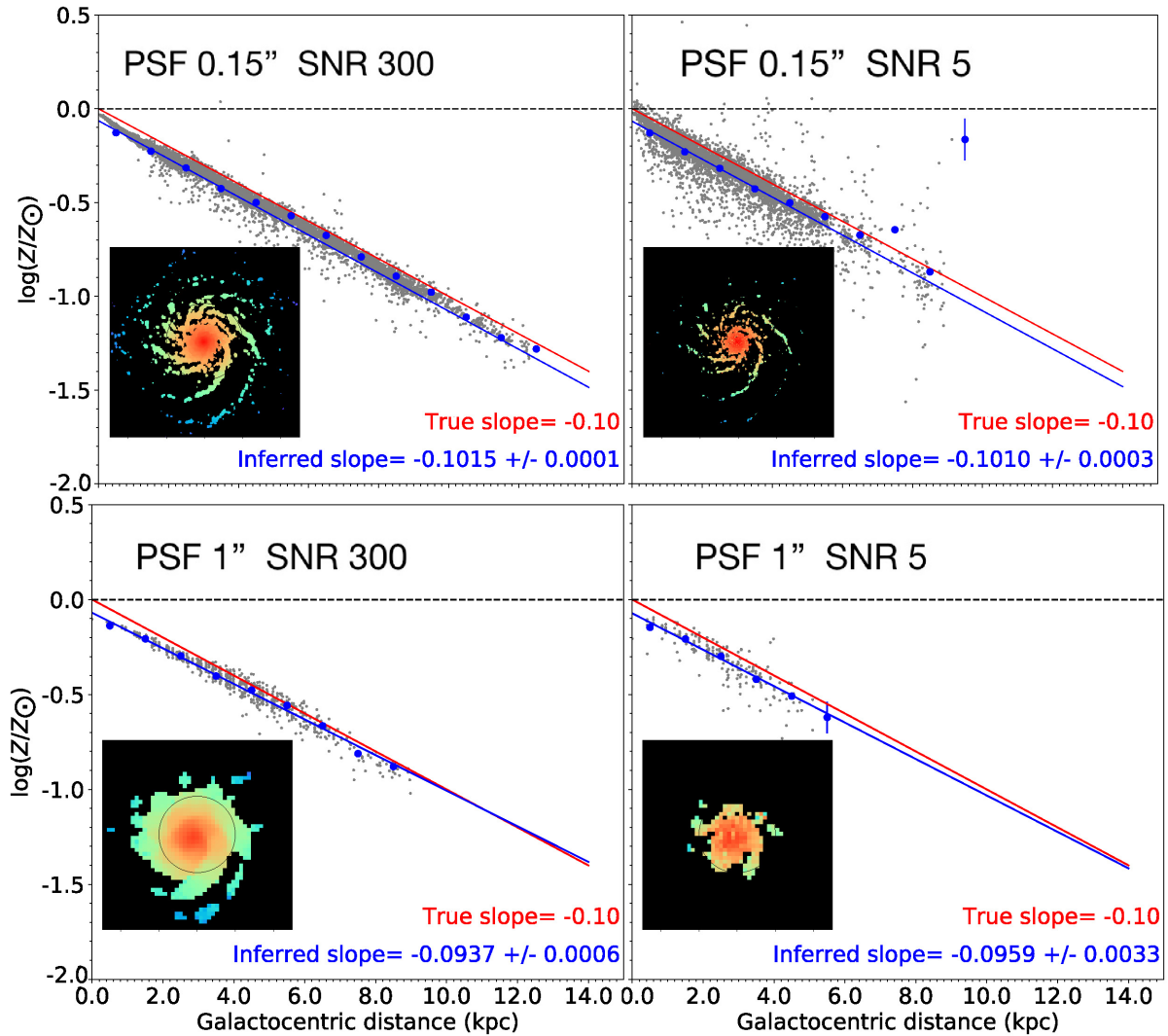
Figure 9. Sample results for four different combinations of resolution and SNR. In each panel, the grey points denote the metallicity of each spatial pixel as a function of the galactocentric radius; we compute error bars on each of these points, but in the plot we have suppressed them for clarity. In each panel, we have discarded pixels with SNR < 3, for each emission-line map involved in the metallicity diagnostic. The blue points with error bars are uncertainty-weighted mean metallicities in radial bins of 1 kpc. The red line denotes the true metallicity as a function of radius, while the blue line is the uncertainty-weighted best linear fit to the grey points. The slopes of the red and blue lines are quoted in each panel as the input and inferred metallicity gradients, in units of dex kpc^{-1}. The corresponding 2D metallicity map is shown in the insets. The spatial span of the 2D map is the same as shown in Fig. 2. The colour coding is based on metallicity and ranges from 0 to -1.2 in $\log(Z/Z_\odot)$ space. From top to bottom, the size of the PSF increases from 0.15 to 1 arcsec. From left to right: the SNR/spaxel deteriorates from 300 to 5. Therefore, the top left panel represents a close-to-ideal case, whereas the bottom right panel is the most realistic scenario.

The metallicity uncertainty in each pixel is then used to assign pixels relative weights for the radial fit.

Propagation of uncertainty is important even in the limiting case of very high SNR, because although the absolute uncertainties in the metallicity of each pixel would be small for very high intrinsic SNR of the data cube, it is the relative uncertainties of the outer-to-inner pixels that affect the fit.

4.2 Summary of results

Fig. 10 shows our main results, and Tables 1 and 2 quote the values for the full parameter space of our study. Our main finding is that coarser spatial resolution leads to an increasingly shallower metallicity gradient inferred in observations. We return to the question of why lower spatial resolution always has a flattening effect on the inferred metallicity gradient in Section 5.4.

Fig. 10 also shows that for a given spatial resolution, the accuracy of the inferred gradient saturates above a certain SNR, i.e. for the limiting case of infinite SNR, there is a best possible metallicity gradient that can be recovered at a given spatial resolution. However, this best possible value is achieved at an SNR ~ 30 for all practical scenarios. As described before in Section 2.7, we define SNR of a datacube as the mean SNR in the $(0.9\text{--}1.1) \times r_{\text{scale}}$ annulus of the [N II] $\lambda 6584$ emission-line map. The value of best possible accuracy depends on the true metallicity gradient and the metallicity diagnostic employed (see Appendix A for a discussion) because different diagnostics have varying sensitivities in the metallicity regime of interest.
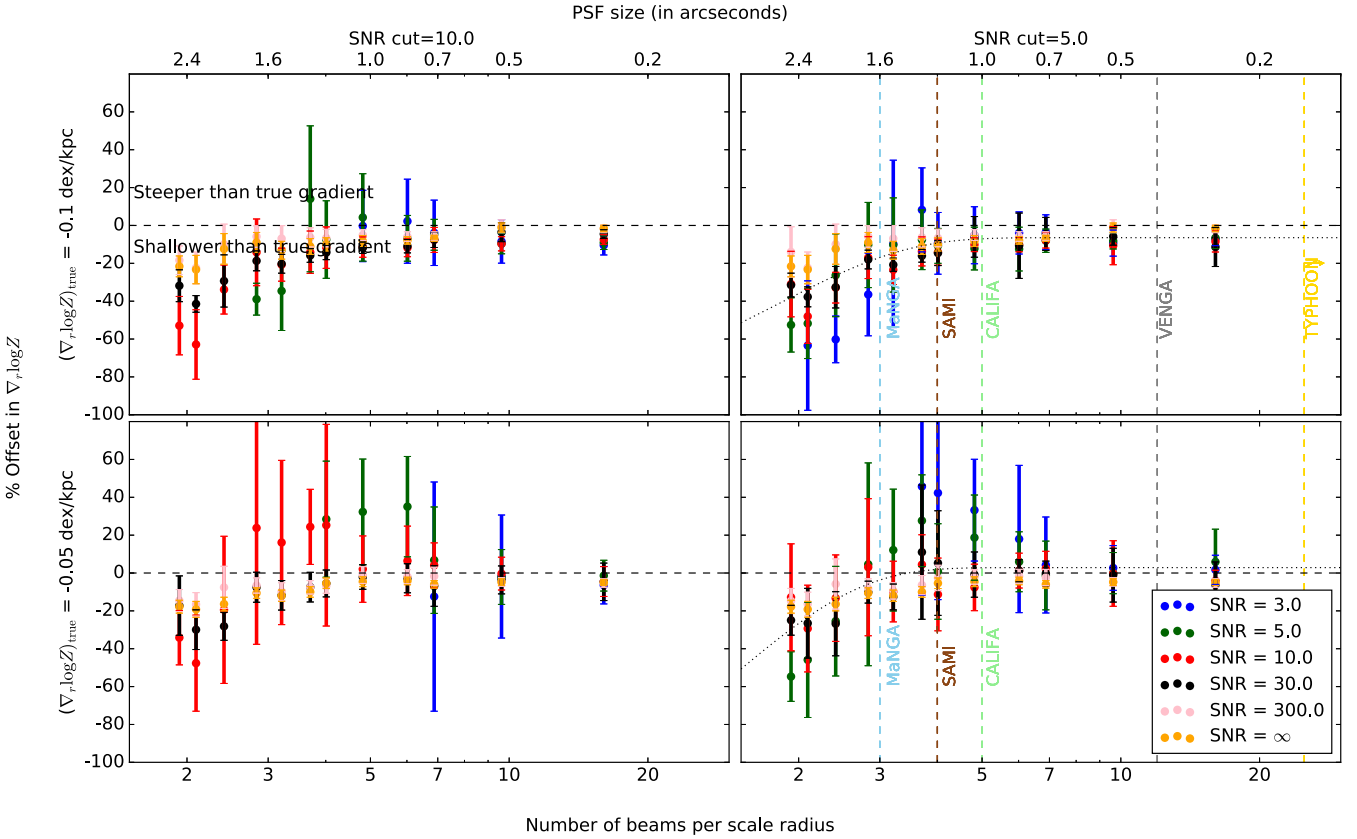
Figure 10. Error in inferred metallicity gradient as a function of spatial resolution and SNR. Each point and error bar show the mean and variance of our 10 realizations of the model. The *top* and *bottom* rows correspond to input gradients of -0.1 and -0.05 dex kpc^{-1}, respectively. The *left column* shows results for a fit using only pixels with an estimated SNR of 10 or more in each emission-line map, whereas the *right column* to an SNR cut-off of 5. The points in each panel show the relative offset (in percentage) between the inferred metallicity gradient and the input gradient, as a function of spatial resolution. The horizontal black dashed line indicates zero error, i.e. inferred metallicity gradient matches true gradient. Points above and below this line signify inferred gradients steeper and shallower than the true gradient, respectively. The coloured points correspond to input SNR values of 3–30, as indicated in the legend. Blue points on the right column imply datacubes that have a mean SNR = 3 at the scale length, but the measurements have been performed after discarding all pixels with SNR < 5. The black dotted curve is our best-fitting function (see Section 4.3). Vertical lines on the right column indicate typical spatial resolution levels of some of the current IFU surveys, with lower limits on the resolution marked with an arrow. We compute these typical resolution values by considering our model galaxy ($r_{\rm scale}$ = 4 kpc) at the mean redshift of the given survey.

We also find that, if one fits the metallicity gradient using all pixels, a lower SNR leads to a shallower inferred metallicity gradient (Fig. 10). While correct propagation of uncertainties that down-weights low SNR pixels in the outer parts of galaxies partly counters this effect, at low SNR this correction is imperfect, and the inferred metallicity gradient remains too shallow (right column of Fig. 9). One way to avoid this problem, which is widely practised by the community, is to discard the noisy pixels based on a threshold SNR cut-off, i.e. fit the metallicity gradient of the galaxy using only those pixels for which the SNR that we infer is above some threshold. When we implement this procedure using a threshold SNR of 5 for each emission line involved in the diagnostic, inferred metallicity gradients are close to the best possible accuracy ($\sim$10–15 per cent) even when the overall SNR is low, as seen in the right column of Fig. 10. We select the threshold of 5 because experimentation shows that this is sufficient to reproduce the best possible accuracy for most cases. However, this improvement has a price: for data cubes with low overall SNR, excluding pixels with SNR below 5 from the analysis might not leave enough pixels for a reliable fit, leading to large uncertainties. We can see the interplay of these effects in our model datacubes with the lowest overall SNR (SNR = 3). If we fit the metallicity gradients in these cases using all pixels (i.e. without

imposing an SNR cut-off), we infer gradients that are drastically different from the true values and exhibit a non-uniform trend with spatial resolution (blue data in left column of Fig. 10). This is because these low SNR (i.e. more noisy) datacubes lead to larger uncertainties during the fitting processes. Discarding pixels with SNR below 5 (the right column of Fig. 10) brings the results close to the best possible inferred gradient when the spatial resolution is high and thus there are a large number of independent pixels, ensuring that enough will have high SNR to allow reliable estimation of the gradient. However, at low spatial resolution applying a cut in SNR leaves so few pixels that the error in the mean result, and the dispersion between different noise realizations, is actually worse than if no SNR cut had been applied.

Fig. 11 summarizes all these effects by showing heat maps for our full parameter space, corresponding to the data in Tables 1 and 2. The colour map denotes the relative offset between the inferred and true gradients in percentage. Given an observed IFU datacube of some intrinsic SNR and PSF size, one can effectively read off (from Tables 1 and 2, or Fig. 11) by how much would the inferred metallicity gradient be offset from the true value. This is useful for planning large surveys, e.g. given the telescope resolution and source redshift, it is possible to calculate the exposure time required

Table 1. Relative offsets of the inferred gradient with respect to the input gradient for the full parameter space we studied, using the procedure of fitting the gradient using all pixels. Relative offsets in the table are expressed as a percentage of the true value, with the sign chosen so that shallower gradients correspond to negative values. Thus for example if the true gradient is −0.1 dex/kpc and the measured gradient is −0.09 dex/kpc, we report a value of −10 in the table. The SNR quoted in the first column denotes the input SNR. The output SNR i.e. the SNR measured at the scale length of the synthetic datacube is usually lower than the input SNR, due to the additional uncertainty introduci[ed] by continuum subtraction (Section 2.7). We fit the gradient using only pixels with SNR ≥ 10, in this case.

SNR	SNR cut-off = 10													
	Number of beams per scale radius													
	96.26	32.09	16.04	9.63	6.88	6.02	4.81	4.01	3.70	3.21	2.83	2.41	2.09	1.93
	Input gradient = −0.1 dex/kpc													
1.0	−7.5	−17.5	−5.0	–	–	–	–	–	–	–	–	–	–	–
3.0	−5.8	−7.5	−10.1	−8.0	−2.4	2.2	−0.3	–	–	–	–	–	–	–
5.0	−3.1	−5.0	−9.2	−9.4	−4.2	−6.8	4.2	−7.4	14.1	−34.6	−39.0	–	–	–
8.0	−4.2	−5.1	−7.0	−10.3	−10.1	−11.2	−12.0	−3.0	−5.7	−15.4	−15.0	−45.9	−65.5	−42.5
10.0	−1.9	−5.3	−7.5	−9.6	−6.6	−11.8	−11.4	−11.7	−14.1	−20.7	−14.2	−33.8	−62.9	−52.9
30.0	−3.0	−2.9	−4.5	−3.4	−7.5	−11.0	−11.8	−14.3	−15.9	−20.3	−18.7	−29.3	−41.5	−31.8
300.0	−3.4	−2.7	−1.8	−0.5	−5.5	−5.9	−5.1	−6.7	−6.3	−6.9	−5.9	−9.9	−22.4	−18.4
	Input gradient = −0.05 dex/kpc													
1.0	−12.0	−14.0	12.7	–	–	–	–	–	–	–	–	–	–	–
3.0	−1.5	−9.2	−7.2	−4.2	−17.9	–	–	–	–	–	–	–	–	–
5.0	−5.1	−6.8	−1.6	−1.4	10.1	35.1	32.3	28.4	131.8	–	–	–	–	–
8.0	−5.4	−3.1	−0.1	−0.1	3.3	17.4	10.0	3.3	23.2	−7.5	−26.6	−40.2	−58.6	–
10.0	−5.5	−0.7	−4.6	−0.4	5.9	6.4	2.1	25.3	24.4	16.1	23.8	−19.5	−47.6	−34.1
30.0	−6.5	−4.7	−4.7	−2.8	−6.3	−2.8	−2.1	−5.5	−7.5	−11.8	−7.5	−28.2	−29.9	−17.2
300.0	−6.9	−6.1	−5.3	−4.9	−1.7	−2.0	−1.1	−6.8	−7.6	−9.8	−6.6	−7.6	−16.6	−14.5
	Input gradient = −0.025 dex/kpc													
1.0	2.1	−11.2	–	–	–	–	–	–	–	–	–	–	–	–
3.0	−6.1	13.4	14.3	56.9	–	–	–	–	–	–	–	–	–	–
5.0	−10.0	−0.5	13.5	23.5	59.3	52.2	130.4	–	–	–	–	–	–	–
8.0	−9.3	−4.6	19.5	20.2	47.7	39.9	56.3	34.9	57.7	–	−100.8	–	–	–
10.0	−9.9	−0.4	15.7	20.3	15.1	36.5	35.7	36.0	52.5	60.1	−18.4	−25.7	−52.4	–
30.0	−10.0	−8.3	−2.8	9.0	2.4	11.7	17.5	10.0	11.2	−1.7	17.6	−11.4	−11.0	−27.2
300.0	−10.0	−8.9	−7.7	−7.3	−5.0	−5.8	−4.0	−1.1	−2.8	−2.7	5.6	−0.7	−7.2	−1.8
	Input gradient = −0.01 dex/kpc													
1.0	−16.5	1.1	–	–	–	–	–	–	–	–	–	–	–	–
3.0	12.4	27.7	47.5	−19.9	–	–	–	–	–	–	–	–	–	–
5.0	3.2	12.3	46.5	65.9	25.2	116.0	–	–	–	–	–	–	–	–
8.0	3.4	9.5	38.9	35.7	124.2	128.7	100.8	95.0	174.7	–	–	–	–	–
10.0	4.9	14.9	27.5	32.4	84.5	95.0	132.6	74.9	106.2	230.3	62.2	450.0	5.5	–
30.0	6.0	8.1	20.5	14.5	53.5	32.9	55.8	22.7	27.6	12.9	64.1	1.6	−9.8	7.2
300.0	6.0	8.2	9.1	8.6	10.0	13.9	5.7	9.0	9.6	5.3	27.1	10.0	2.0	13.0

to achieve a certain SNR and thereby achieve a target accuracy in the measured metallicity gradient.

4.3 Quantifying the effect of spatial resolution

In order to quantify the effect of the spatial resolution on the inferred metallicity gradient, we fit the data in Fig. 10 using a least square algorithm, with a function of the form

$$y(x) = a \left(\mathrm{erf} \left(\frac{\sqrt{\pi}x}{2b} \right) - 1 \right) + c, \tag{23}$$

where a and b denote the steepness and the 'knee' of the fit, respectively, c is the asymptotic offset of the fit from the true input gradient at high spatial resolution, x is the number of spatial resolution elements per scale radius, and y is the percentage offset of the inferred gradient from the true value. We choose this functional form because $\mathrm{erf}(z)$ limits to a constant value for $z \to \infty$ and approximates to a linearly increasing function for $z \ll 1$, which resembles the behaviour of our data. We define the 'knee' (b) as

the intersection of the linear (= $a(x/b\text{-}1) + c$) and constant (= c) parts of the fit function. If the scale length of an observed galaxy is resolved by b beams or fewer, it implies that the inferred metallicity gradient is considerably offset from the most accurate measurement possible, given the diagnostic.

We fit equation (23) to a data set consisting of all our inferred metallicity gradients at all SNRs and resolutions, derived using an SNR cut-off of 5 (i.e. the points shown in the right column of Fig. 10), and obtain best-fitting values $a = 100.5$, $b = 3.1$, $c = −8.9$ and $a = 123.0$, $b = 2.6$, $c = −2.2$ for true metallicity gradients of −0.1 and −0.05 dex kpc^{-1}, assuming fiducial values for all other parameters. Table 3 gives an overview of the different choices we investigated and the best-fitting parameters of the error in inferred gradient as a function of the spatial resolution for each case. The 'knee' of the function (parameter b) is similar for all cases, indicating that the dependence on resolution is qualitatively unaffected by the different choices of Ω, f_g, and diagnostic (see Appendix A for a detailed discussion). Using equation (23) and the values from Table 3, it is possible to estimate the accuracy of the

Table 2. Same as Table 1 but now fitting the gradient using only pixels with SNR ≥ 5. Missing values in the table correspond to cases where the imposed SNR cutoff leaves too few pixels to make it possible to fit a metallicity gradient.

SNR	SNR cut-off = 5													
	Number of beams per scale radius													
	96.26	32.09	16.04	9.63	6.88	6.02	4.81	4.01	3.70	3.21	2.83	2.41	2.09	1.93
	Input gradient = −0.1 dex/kpc													
1.0	−9.4	−12.7	−8.1	−17.1	−8.9	–	–	–	–	–	–	–	–	–
3.0	−4.8	−11.1	−6.9	−9.3	−3.5	−4.0	−5.2	−9.4	8.2	−9.7	−36.5	−60.2	−63.5	–
5.0	−3.2	−8.8	−6.7	−7.9	−7.7	−12.5	−12.2	−10.8	−7.6	−10.3	−10.3	−26.3	−51.8	−52.5
8.0	−6.9	−11.6	−5.6	−9.6	−7.7	−9.6	−9.7	−15.9	−11.1	−21.4	−18.2	−32.4	−40.7	−40.8
10.0	−4.2	−1.9	−8.2	−11.6	−7.3	−10.2	−10.2	−14.1	−13.7	−23.5	−17.0	−32.9	−48.0	−31.0
30.0	−3.6	−7.7	−11.1	−5.8	−3.8	−10.7	−6.0	−14.6	−16.0	−20.8	−18.0	−32.6	−37.7	−31.5
300.0	−3.4	−2.7	−1.8	−0.5	−5.5	−5.9	−5.0	−6.6	−6.3	−6.9	−5.8	−9.9	−22.4	−14.0
	Input gradient = −0.05 dex/kpc													
1.0	−4.4	4.7	−15.4	0.3	–	–	–	–	–	–	–	–	–	–
3.0	−3.4	−9.2	1.2	2.4	3.7	18.0	33.2	42.3	45.7	–	–	–	–	–
5.0	−5.4	0.6	4.5	0.3	−0.7	5.9	18.7	0.8	27.6	12.1	4.6	−25.4	−45.9	−54.7
8.0	−5.6	0.8	−1.9	−2.4	−0.0	−0.3	−0.9	−7.8	2.7	−8.2	−8.8	−9.2	−22.8	−10.6
10.0	−5.7	−3.1	−4.0	−1.4	3.0	1.3	−7.6	−11.3	4.4	−9.8	3.1	−13.3	−29.3	−12.8
30.0	−6.5	−5.4	−5.9	−1.2	−0.3	1.2	−0.8	5.3	11.0	−12.6	−10.1	−26.8	−26.5	−25.0
300.0	−6.9	−6.1	−5.3	−4.8	−2.3	−1.7	−2.1	−2.6	−6.4	−9.4	−8.3	−5.8	−16.4	−14.8
	Input gradient = −0.025 dex/kpc													
1.0	−4.1	−6.7	−11.8	–	–	–	–	–	–	–	–	–	–	–
3.0	−9.2	15.4	19.4	22.6	38.1	51.2	146.7	145.7	–	–	–	–	–	–
5.0	−10.4	0.1	21.2	25.1	31.5	47.2	53.8	61.1	75.1	47.4	45.2	94.0	−53.2	–
8.0	−9.5	−5.4	13.1	23.3	29.9	26.2	41.4	38.6	29.4	21.4	28.4	−1.6	−38.9	−84.6
10.0	−10.2	−2.4	12.2	9.6	25.1	14.6	31.7	20.8	23.7	27.3	27.8	1.8	3.2	15.2
30.0	−10.0	−8.4	−3.9	13.1	10.8	19.4	13.2	26.2	11.3	1.2	9.2	−11.6	−10.2	−14.6
300.0	−10.0	−8.9	−7.8	−7.4	−6.1	−5.0	−4.6	0.7	1.9	8.0	−0.8	0.3	0.2	−1.9
	Input gradient = −0.01 dex/kpc													
1.0	34.8	−4.5	80.2	–	–	–	–	–	–	–	–	–	–	–
3.0	3.0	20.1	34.6	83.3	153.0	96.0	76.3	–	–	–	–	–	–	–
5.0	2.5	11.3	74.2	36.8	109.3	121.8	220.2	151.6	224.7	187.1	20.3	−2.5	5.5	–
8.0	2.1	28.4	30.4	44.2	71.7	54.6	60.1	145.7	146.7	106.0	70.4	82.5	48.0	71.4
10.0	4.1	18.6	27.6	43.8	71.8	46.5	74.7	172.3	117.6	151.3	104.3	69.0	1.4	−51.7
30.0	6.0	6.8	35.2	27.5	32.7	40.2	25.6	18.4	19.4	16.4	16.6	−13.4	−31.5	7.4
300.0	6.0	8.0	9.2	8.5	10.9	13.8	7.5	23.2	29.2	40.7	15.0	12.2	3.1	19.0

inferred metallicity gradient for a given spatial resolution. However, we caution the reader that our models do not include contributions from the diffused ionized gas (DIG), a point to which we return in Section 5.1.1. As a rough rule of thumb, we find that the scale length should be resolved by more than 4 beams for the inferred metallicity gradient to be close to the most accurate value possible, for a given diagnostic.

5 DISCUSSION

5.1 Caveats

Here, we discuss the physical processes that we have not accounted for in our work and how they could potentially impact our results.

5.1.1 Diffused ionized gas

An important caveat of our work is that we do not account for the emission from the diffuse ionized gas (DIG). The DIG consists of low-density ISM illuminated by the ionizing photons that escape dense H II regions and appears to be ubiquitous across the ISM of galaxies (Haffner et al. 2009). Poetrodjojo et al. (in preparation)

show that DIG emission leads to an increase in [N II] λ6584/H α and [S II] λλ6717,30/H α emission-line ratios (Blanc et al. 2009; Zhang et al. 2017), an effect that is larger at coarser spatial resolutions. As such, DIG is likely to have a considerable impact on the D16 diagnostic, which makes use of these ratios. However, we have also tested our analysis with the KD02 [N II] λ6584/[O II] λλ3727,29 diagnostic (results in Section A3 of online material) which is less likely to be dominated by DIG effects. We intend to add the effects of DIG emission to our models in future work.

5.1.2 Dust

We do not explicitly account for dust in our model for the ionized gas bubble. The predicted emission-line fluxes for each H II region that we derive from the photoionization model grid are effectively the dereddened values. We use the D16 metallicity diagnostic which is robust to reddening by virtue of using emission lines that are narrowly spaced in wavelength. Using the D16 indicator ensures that our results will hold, to first order, in real world scenarios where dust is present. For instance, very small Balmer decrements ($\log (H_\alpha/H_\beta)_{\rm obs} \lesssim 3 \Rightarrow E(B − V) \lesssim 0.04$) have been observed in high spatial resolution ($\lesssim 100$ pc) IFU studies of face-on nearby
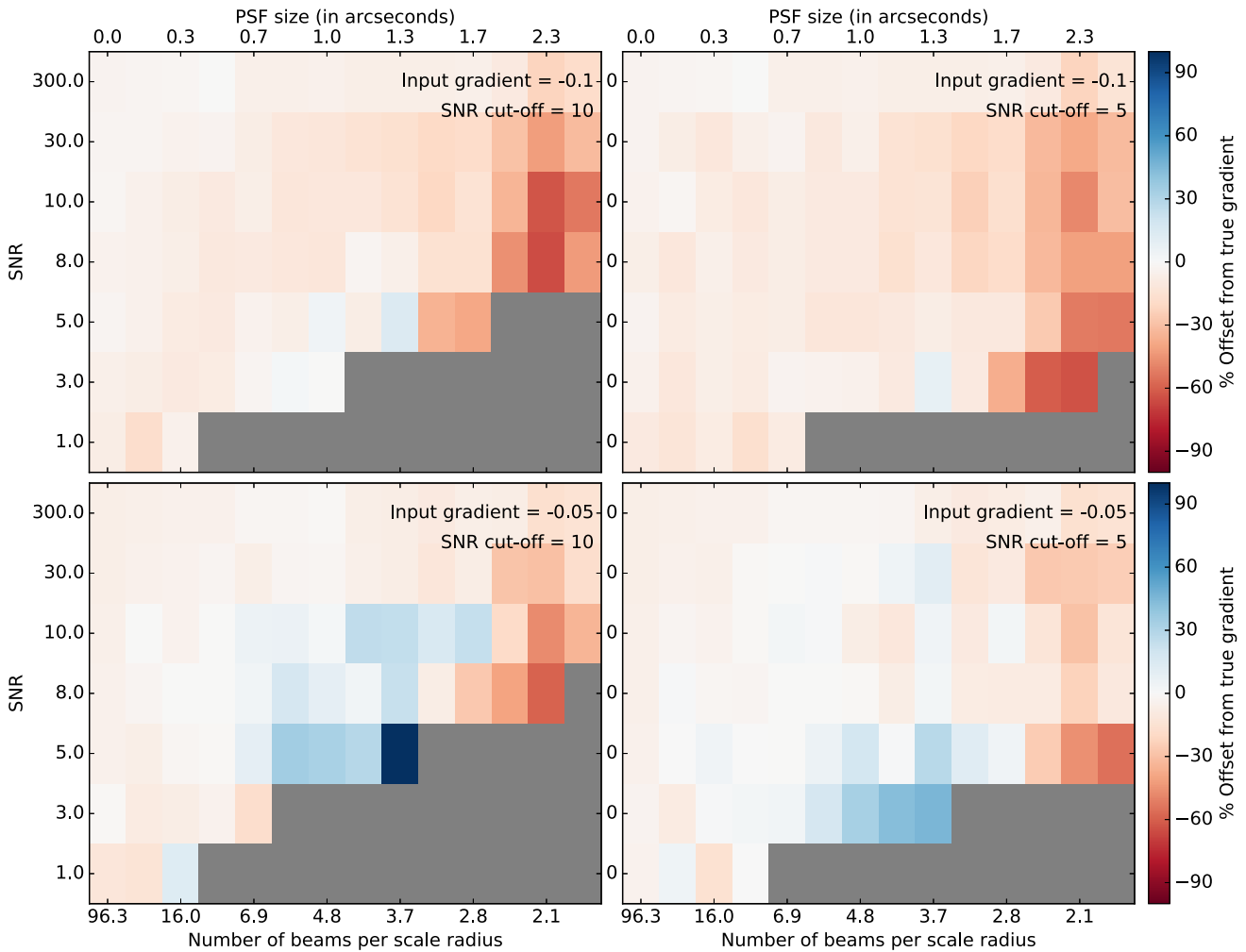
Figure 11. Error in inferred metallicity gradient as a function of resolution and SNR; the results shown here are a superset of those shown in Fig. 10. The top and bottom rows correspond to input metallicity gradients of −0.1 and −0.05 dex kpc^{-1}, respectively. The right column shows results for fits using only pixels with SNR of 5 or more in each emission line, while the left column shows results corresponding to SNR cut-off = 10. Grey cells correspond to combinations of resolution and SNR, whereby no pixels reached the threshold SNR, and thus no data were available to which to fit the metallicity gradient.

spiral galaxies (TYPHOON; private communication) which have negligible effect (≲0.001 dex) on the D16 diagnostic.

However, if multiple H II regions with different luminosities and reddening values lie along the line of sight, it could potentially lead to a second-order effect in the metallicity diagnostic. The precise magnitude of such an effect is out of the scope of this work but we present a qualitative argument here. For instance, in a scenario where H II regions are more embedded when they are younger and have harder spectra, not correcting for differential extinction can potentially lead to the metallicity being overestimated (by missing the young, dust-obscured, low metallicity H II regions) in relatively more dusty regions of the galaxy. The resultant effect on the inferred metallicity profile would depend on the spatial distribution of dust in the galaxy. If the dust distribution follows a smooth, radially declining profile, the central metallicity would be overestimated the most, resulting in artificial steepening of the gradient.

We do not model the Balmer absorption because we do not account for the underlying old stellar populations in the galaxy in determining the continuum. For this work, we consider only the young (<5 Myr) stellar associations that contribute to most of the ionizing flux that drives the H II regions. Although our continuum subtraction prior to line-fitting ensures that the overall shape of

the continuum does not impact the emission-line ratios, the lack of the old stellar population and the Balmer decrement potentially underestimates the uncertainties originating from continuum subtraction. Consequently, the uncertainties in the line fluxes and the metallicity gradients are potentially underestimated due to lack of dust.

5.1.3 Variation of wind strength

We parametrize the relative strength of stellar winds versus gas and radiation pressure via the quantity Ω. Quantitatively, Ω is the ratio of the volume occupied by the inner stellar wind cavity to that occupied by the ionized H II region shell. A high Ω implies a thin shell of ionized gas, whereas a low Ω value corresponds to an almost spherical ionized bubble with a negligible stellar cavity. The value of Ω is potentially important for line diagnostics because it changes the ionization parameter at fixed H II region size, and because it changes the amount of spherical divergence that the radiation field undergoes as it propagates through the H II region; Yeh & Matzner (2012) show that the value of Ω has observable effects on infrared line ratios, for example.

Table 3. Best-fit parameters for our model of metallicity gradient error as a function of spatial resolution (equation 23 in the main text), for all choices of diagnostic, and all values of Ω (wind strength parameter), f_g (simulation gas fraction), and $\nabla_r \log Z$ (true metallicity gradient). Note that the parameter c represents the minimum possible error that can be achieved for a given set of parameters, i.e., it is the value reached in the limit of infinite spatial resolution and SNR.

Diagnostic	Ω	f_g	$(\nabla_r \log Z)_{\text{true}}$	a	b	c
D16	0.5	0.2	− 0.1	92.5	2.4	− 5.7
D16	0.5	0.2	− 0.05	225.8	1.6	3.7
D16	0.5	0.1	− 0.05	77.3	3.2	− 1.8
D16	0.5	0.1	− 0.1	83.7	3.2	− 6.3
D16	5.0	0.2	− 0.1	95.8	2.6	− 4.6
D16	5.0	0.2	− 0.05	117.4	2.2	2.6
D16	0.05	0.2	− 0.1	124.9	2.1	− 6.9
D16	0.05	0.2	− 0.05	99.7	2.0	− 0.8
KD02	0.5	0.2	− 0.1	174.1	2.7	7.2
KD02	0.5	0.2	− 0.05	163.4	2.5	− 4.4
KD02	0.5	0.1	− 0.05	84.8	4.6	− 3.2
KD02	0.5	0.1	− 0.1	113.2	5.4	23.3
KD02	5.0	0.2	− 0.1	145.6	2.9	5.4
KD02	5.0	0.2	− 0.05	98.8	4.0	1.6
KD02	0.05	0.2	− 0.1	142.6	2.9	6.4
KD02	0.05	0.2	− 0.05	156.7	2.6	− 4.5

Our fiducial value, $\Omega = 0.5$, is primarily motivated by observations. In their analysis of starburst galaxies, Yeh & Matzner (2012) find that observed infrared line ratios are best explained by $\Omega \sim 0.5$. Similarly, Lopez et al. (2011, 2014) use X-ray observations of Magellanic Cloud H II regions to constrain the fractional contribution f_X of the wind bubble to the overall volume-averaged pressure; assuming that the H II region interior is isobaric, f_X and Ω are related simply by $\Omega = f_X/(1 - f_X)$ [see the appendix of Lopez et al. (2014) for further discussion]. They find that $f_X \sim$ 0.1–0.5 in the LMC, again consistent with our fiducial choice; for the SMC, $f_X < 0.1$, likely because the metallicity is too low for the stars to have substantial winds. Finally, in the MW, Draine (2011a) notes that the nearly spherical N49 bubble has a maximum surface brightness that is a factor of ≈ 2 higher than its central surface brightness, a ratio that is too large to be explained by radiation pressure effects alone, but is consistent with what is expected if the ionized gas is confined to a thick shell by a wind bubble. If we approximate the photoionized region as having uniform emissivity, so that the observed surface brightness variations are solely due to variations in the path-length through the shell along different lines of sight, then a maximum-to-central surface brightness ratio of 2 corresponds to a shell where the wind bubble occupies the inner $\Omega = 0.28$ of the volume, close to our fiducial value.

Estimates of Ω from theory are significantly more diverse. One-dimensional simulations often indicate that stellar winds should dominate the pressure of H II regions, implying large Ω (e.g. Fierlinger et al. 2016; Rahner et al. 2017; Silich & Tenorio-Tagle 2018). However, multidimensional simulations usually imply small values of Ω, due to efficient leakage of hot gas or mixing-induced radiative cooling (e.g. Mackey et al. 2015; Calura et al. 2015; Wareing, Pittard & Falle 2017). The 1D and multidimensional simulations are somewhat difficult to compare, because the 1D models explore a much larger range of parameter space, and often find winds to be important primarily in the parts of parameter space – very massive and dense clusters – that are most inaccessible to multidimensional simulations. Energetic analyses even of large H II regions tend to imply that there must be significant loss of wind

energy due to breakout or mixing-induced cooling (e.g. Harper-Clark & Murray 2009; Rosen et al. 2014; Rogers & Pittard 2014; Toalá & Arthur 2018), which would suggest $\Omega \lesssim 1$, but this is hard to quantify. For further discussion of this issue, we refer readers to the recent review by Krumholz et al. (2019).

Given that there is some uncertainty about the true value of Ω, it is worthwhile to explore how variations in its value could affect our results. We do so by performing two different tests. First, we construct model galaxies in which $\Omega = 0.05$ and 5.0 (compared to our fiducial choice $\Omega = 0.5$) independent of radius within the galaxy. Secondly, we construct a model galaxy in which Ω varies radially from 5.0 at the centre to 0.05 at $2R_e$.

Figs S3 and S4 in the SOM show the results for $\Omega = 0.05$ and 5.0, which correspond to a nearly spherical and a thin shell geometry, respectively. Comparing with Fig. 10 we see no qualitative difference in the results, and nearly identical dependence on resolution. The only significant effect of varying Ω is to slightly increase the absolute error, so that a few per cent error remains even in the limit of high SNR and resolution. The origin of this effect is easy to understand: the H II region models used to construct the diagnostic implicitly assume a value of Ω via the geometry they adopt. While the diagnostics attempt to minimize the effects of geometry, they cannot do so perfectly, leading to a persistent error if the geometry is not exactly what was assumed, even if the measured line strengths are perfect. This variation in the best possible accuracy produces an overall vertical offset to the data depending on Ω. As mentioned before, the vertical offset also depends on other factors, such as the DIG, that has not been accounted for in this work. However, variations in Ω have only very minor effects on the resolution dependence of the error. Quantitatively, when we use equation (23) to fit the resolution-dependent error in our recovered metallicity gradients for $\Omega = 0.05$ or $\Omega = 5$, the coefficients a and b that describe the shape and location of the 'knee' in the curve change by at most ≈ 10 per cent (see Table 3). We therefore conclude that, even if Ω takes on a wide range of values, our conclusions about the resolution dependence of metallicity errors change by at most ≈ 10 per cent. Full tables and heatmaps for cases with different Ω values are provided as SOM.

Fig. S7 in the SOM compares our model galaxy with Ω varying with radius to our fiducial case with fixed $\Omega = 0.5$. As with the case of varying Ω independent of radius, we find no qualitative difference between the two cases. We refer the readers to the SOM for details of the variable Ω models.

5.2 Dependence on metallicity diagnostic

Switching between metallicity diagnostics usually introduces an offset in the absolute value of the metallicity determined (Kewley & Ellison 2008). The absolute magnitude of such offset is not uniform across the metallicity range typically observed in galaxies. Therefore, different diagnostics can potentially introduce slight deviations to the inferred slope. To test the extent of this effect, we re-compute our results with the [N II] $\lambda 6584$/[O II] $\lambda\lambda 3727,29$ (N2O2) diagnostic following KD02 (as opposed to the D16 diagnostic used in the main text), keeping everything else unchanged. We note, however, that the absence of a treatment for dust could affect the N2O2 diagnostic due to the inclusion of the [O II] lines. In order to use the [O II] $\lambda\lambda 3727,29$ emission lines for the KD02 diagnostic, we produce an additional set of synthetic datacubes (and corresponding emission line maps) spanning the wavelength range of 3670–3680 Å following the same procedure as described in Sections 5.4 and 3.1.

Figs S8 and S9 (and Tables S7 and S8) of the SOM present the results using the KD02 diagnostic. We find no qualitative difference in our results on comparing Fig. S7 of SOM with Fig. 10. The figures show that the absolute offset of the accuracy of inferred gradient (parameter c in Table 3) depends on the diagnostic used. However, the accuracy as a function of the spatial resolution, including the 'knee' of the functional form (parameter b in Table 3), remains largely unchanged. The discrepancy in the vertical offset is due to different sensitivities of the different diagnostics in the metallicity regime concerned.

5.3 Comparison with previous work

Several authors have previously examined the question of how well observations at finite resolution and sensitivity can recover galaxy metallicity gradients. Most recently, Carton et al. (2017, hereafter C17) proposed a forward modelling approach to correct for the degrading effects of seeing and spatial binning on metallicity gradients. They find that coarser spatial resolution yields a systematically steeper metallicity gradient (fig. 4 of C17) which is contrary to what we observe in our models, and to what one would expect based on an intuitive understanding of how line ratio diagnostics operate. Metallicity diagnostics are non-linear functions of the ratio of a weaker nebular line to a stronger one. Consequently, in a galaxy where metallicity falls with radius, the true radial gradient of the weaker line tends to be steeper than the true radial gradient of the stronger line. Beam-smearing affects the line with the steeper gradient more severely, and so smearing tends to reduce the value of the diagnostic line ratio in the galaxy centres, and raise it in galaxy outskirts. This in turn leads one to infer an artificially shallower gradient. C17 point out that their contrary-to-expectation variation of the inferred gradient could arise from the systematics of their modelling. We can identify three significant differences between their approach and ours that might be responsible for this difference.

First and most simply, their choice of range of the PSF variation corresponds to ∼0.25–1 beams per scale length of the galaxy, which is towards the extreme coarse end of our range of spatial resolutions (Fig. 10). Secondly, C17 use an analytic prescription for the spatial distribution of star formation with a galaxy, while we use a distribution taken from a self-consistent, high-resolution hydrodynamic simulation. This leads to large differences in the morphology of emission: our galaxy has a clumpy, spiral structure similar to that of observed disc galaxies. C17 point out that a clumpy morphology can have significant impact on interpretation of metallicity gradients, but are unable to capture this effect with their method. Thirdly, C17 implement a noise model that depends solely on the emission line fluxes, with no sky contribution. As such, the spatial distribution of SNR within their model galaxy is uniform, which is not necessarily observed in IFU surveys. In addition to the flux dependence, the noise in observed galaxies is dictated by instrumental noise and contribution from the sky background. The latter two sources of *noise* have no real spatial variation, implying an overall radially declining *SNR* profile in observed emission-line maps. We account for this by implementing a wavelength-dependent noise model that takes into account both the instrumental read noise and the sky noise contribution, in addition to the flux dependence of noise. As we have shown in this paper, and as C17 have also discussed, accounting for the SNR is crucial to inferring metallicity gradients. We fit the noise-weighted metallicity of each pixel to a smooth radial profile, implying less relative weight on the pixels at the outskirts. But in the C17 noise model

the relative weight between pixels would effectively be uniform, which could potentially impact their metallicity gradient inference. Another potential cause for the steepening of metallicity gradient in C17 could be the systematic offsets in their models that tend to infer steeper output gradients for larger negative input gradients (their fig. B1) due to model degeneracies stemming from the finite extent of available photonization grids.

Our work shows better agreement with that of Yuan et al. (2013), who also find that coarser spatial resolution (or coarser annular binning) leads to flatter inferred gradients. They suggest this flattening is due to selective smearing of the low SNR pixels of the weak [N II] λ6584 line, thereby leading to an overestimate of the N II flux in the outskirts of the galaxy, flattening the gradient. Their argument suggests that we should see enhanced flattening for steeper intrinsic gradients, which is entirely consistent with the results shown in Fig. 10 and the argument made in Section 5.4. However, Yuan et al. also do not find any appreciable effect of SNR on the inferred gradient, whereas we derive systematically flatter gradients for lower SNR data cubes. This is likely because Yuan et al. impose an SNR cut-off on the N II line, which, as we show in the second column of Fig. 10, eliminates the SNR dependence of the inferred gradient.

5.4 Relevance to current IFU surveys and their interpretations

The vertical lines in the right-hand panel of Fig. 10 indicate the typical spatial resolution of some of the current IFU surveys. Galaxies within a stellar mass range of $9.8 \lesssim \log(M_\star/M_\odot) \lesssim 10.2$, in both SAMI and MaNGA surveys, have a typical spatial resolution of ∼3 beams per scale radius. Such resolution would yield a metallicity gradient ∼20–30 per cent shallower than the intrinsic value, modulo metallicity diagnostics used. The magnitude of this offset is highly dependent on the characteristic size of the disc galaxy being probed. The inferred gradient worsens rapidly with change in spatial resolution at the coarse-resolution end. Hence, a galaxy with 2 kpc scale length probed with SAMI or CALIFA would accommodate only 2 beams across its scale radius, thereby leading to metallicity gradients that are ∼50–60 per cent shallower. The SDSS-IV MaNGA survey (Bundy et al. 2015), with its typical 1–2 kpc resolution, resolves the disc scale lengths with 2–3 beams – thus yielding 20–40 per cent shallower gradients. The TYPHOON/PrISM survey (Seibert et al., in preparation) operates at <100 pc spatial resolution, implying >10 beams per scale length of typical disc galaxies. Our work demonstrates that at such fine resolution, the metallicity gradient measurement is limited by the intrinsic scatter in the metallicity diagnostic used, which depends on the intrinsic metallicity gradient of the galaxy. In other words, if the disc scale length is resolved by >6–7 beams that would yield the most-accurate-possible metallicity gradient. IFU surveys of nearby galaxies such as the VENGA survey (Blanc et al. 2013, ∼300 pc resolution) and the future MUSE/PHANGS survey (estimated ∼50–100 pc) have enough spatial resolution to overcome the PSF smearing effects on metallicity gradients.

The magnitude of the artificial 'flattening' of the metallicity gradient could potentially depend on the morphology of the galaxy, which is linked to the stellar mass. We note that our models are based on an MW type, spiral galaxy with a stellar mass of 10^{10} $M_\odot$. We therefore caution the reader that our quantitative measures will be applicable to large IFU surveys such as SAMI and MaNGA only after the samples have been appropriately trimmed to match our simulated galaxy. We will investigate the effect of morphology in a future work.

6 SUMMARY AND CONCLUSIONS

We produce synthetic IFU observations of face-on simulated galaxies in order to investigate how well observations of finite sensitivity and resolution are able to recover metallicity gradients in disc galaxies. Compared to previous investigations of this topic, our work features a significantly more realistic distribution of star formation within the galaxy, a substantially more realistic treatment of sources of noise, and a wider exploration of the parameter space of resolution and SNR. Our primary conclusions are as follows:

(i) Inferred metallicity gradients are generally shallower than the true value, with the error worsening with coarser spatial resolution and higher noise levels. This is because metallicity is a non-linear function of the emission-line ratios. A coarser PSF redistributes flux from the centre to the outskirts of the disc – an effect that is more pronounced for the emission lines that have a steeper intrinsic radial profile. This differential outcome of the smearing leads to the emission line ratios, and consequently the metallicity, being overestimated in the peripheral regions thereby flattening the metallicity gradient.

(ii) At a given resolution, there is a 'best possible accuracy' for the inferred gradient, no matter how high the SNR. The exact value of the best possible accuracy depends on the resolution, the true gradient and the metallicity diagnostic used. As a rough rule of thumb, however, in a galaxy where the scale length is resolved by 4 telescope beams, the best possible accuracy is a ∼10–15 per cent error in metallicity gradient.

(iii) Achieving the best possible performance at a given resolution requires careful treatment of errors. One must consistently propagate uncertainties all the way from measurement of the errors in individual IFU spaxels, through the steps of measuring line fluxes, converting these to metallicities, and finally fitting the overall gradient. Moreover, it is desirable to exclude pixels with low SNR from fits to the metallicity gradient, and we find that a cut-off at SNR = 5 yields gradient measurements very close to the best possible value. It is conceivable that a more accurate study can be done by performing forward modelling using a full Bayesian MCMC approach so that the weak SNR pixels no longer need to be discarded and we do not lose any data. We plan to investigate this in a future work. The principle motivation of the current work is to emulate the current practice in observational studies, most of which employ strong emission line diagnostics, to infer metallicity gradient for different spatial resolution.

(iv) With proper treatment of uncertainties, the dependence of the error in metallicity gradient on spatial resolution is relatively insensitive to the physical properties of H II region or galaxies (strength of stellar wind, gas fraction) or to the choice of metallicity diagnostic. These only affect the absolute error level, not its dependence on resolution.

In the future we plan to conduct follow up by studies by using the same analysis framework to investigate the effect of the instrumental properties on other observables e.g. star formation maps, ionization parameter, ISM pressure and other strong emission-line ratio diagnostics.

REFERENCES

Allen J. T. et al., 2015, MNRAS, 451, 2780
Belfiore F. et al., 2017, MNRAS, 469, 151
Bisbas T. G. et al., 2015, MNRAS, 453, 1324
Blanc G. A., Heiderman A., Gebhardt K., Evans, Neal J. I., Adams J., 2009, ApJ, 704, 842
Blanc G. A. et al., 2013, AJ, 145, 138
Bland-Hawthorn J., Gerhard O., 2016, ARA&A, 54, 529
Bresolin F., 2007, ApJ, 656, 186
Bresolin F., Gieren W., Kudritzki R.-P., Pietrzyński G., Przybilla N., 2002, ApJ, 567, 277
Bresolin F., Ryan-Weber E., Kennicutt R. C., Goddard Q., 2009, ApJ, 695, 580
Bundy K. et al., 2015, ApJ, 798, 7
Calura F., Few C. G., Romano D., D'Ercole A., 2015, ApJ, 814, L14
Carton D. et al., 2017, MNRAS, 468, 2140 (C17)
Cecil G. N., Croom S., The SAMI Galaxy Survey Team, 2014, American Astronomical Society Meeting Abstracts, Vol. 223, American Astronomical Society Meeting Abstracts #223. p. 246.03
Croom S. M. et al., 2012, MNRAS, 421, 872
da Silva R. L., Fumagalli M., Krumholz M., 2012, ApJ, 745, 145
Diaz A. I., 1989, in Beckman J. E., Pagel B. E. J., eds, Evolutionary Phenomena in Galaxies, Cambridge University Press, Cambridge, p. 377
Dopita M. A., Kewley L. J., Sutherland R. S., Nicholls D. C., 2016, Ap&SS, 361, 61 (D16)
Draine B. T., 2011a, ApJ, 732, 100
Draine B. T., 2011b, Physics of the Interstellar and Intergalactic Medium, Princeton University Press, Princeton, New Jersey
Federrath C. et al., 2017, in Crocker R. M., Longmore S. N., Bicknell G. V., eds, IAU Symp. Vol. 322, The Multi-Messenger Astrophysics of the Galactic Centre. p. 123
Fierlinger K. M., Burkert A., Ntormousi E., Fierlinger P., Schartmann M., Ballone A., Krause M. G. H., Diehl R., 2016, MNRAS, 456, 710
Fu J. et al., 2013, MNRAS, 434, 1531
Fujimoto Y., Krumholz M. R., Tachibana S., 2018, MNRAS, 480, 4025

Fujimoto Y., Chevance M., Haydon D. T., Krumholz M. R., Kruijssen J. M. D., 2019, MNRAS, 487, 1717

Garnett D. R., Shields G. A., 1987, ApJ, 317, 82

Garnett D. R., Shields G. A., Skillman E. D., Sagan S. P., Dufour R. J., 1997, ApJ, 489, 63

Goldbaum N. J., Krumholz M. R., Forbes J. C., 2016, ApJ, 827, 28 (G16)

Green A. W. et al., 2018, MNRAS, 475, 716

Groves B., Brinchmann J., Walcher C. J., 2012, MNRAS, 419, 1402

Haffner L. M. et al., 2009, Rev. Mod. Phys., 81, 969

Harper-Clark E., Murray N., 2009, ApJ, 693, 1696

Ho I. T. et al., 2018, A&A, 618, A64

Ho I.-T., Kudritzki R.-P., Kewley L. J., Zahid H. J., Dopita M. A., Bresolin F., Rupke D. S. N., 2015, MNRAS, 448, 2030

Ho I.-T. et al., 2016, Ap&SS, 361, 280

hoon Kim J. et al., 2013, ApJS, 210, 14

Jones E. et al., 2001, SciPy: Open Source Scientific Tools for Python, Available at: http://www.scipy.org/

Kennicutt R. C., Jr, Bresolin F., Garnett D. R., 2003, ApJ, 591, 801

Kennicutt, Robert C. J., Garnett D. R., 1996, ApJ, 456, 504

Kewley L. J., Dopita M. A., 2002, ApJS, 142, 35 (KD02)

Kewley L. J., Ellison S. L., 2008, ApJ, 681, 1183

Kewley L. J., Rupke D., Zahid H. J., Geller M. J., Barton E. J., 2010, ApJ, 721, L48

Krabbe A. C., Pastoriza M. G., Winge C., Rodrigues I., Ferreiro D. L., 2008, MNRAS, 389, 1593

Kreckel K. et al., 2019, ApJ, 887, 80

Kroupa P., 2001, MNRAS, 322, 231

Krumholz M. R., Burkhart B., 2016, MNRAS, 458, 1671

Krumholz M. R., Matzner C. D., 2009, ApJ, 703, 1352

Krumholz M. R., McKee C. F., Bland-Hawthorn J., 2019, ARA&A, 57, 227

Leitherer C. et al., 1999, ApJS, 123, 3

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

Mackey J., Gvaramadze V. V., Mohamed S., Langer N., 2015, A&A, 573, A10

Maiolino R., Mannucci F., 2019, A&AR, 27, 3

Martin P., Roy J.-R., 1994, ApJ, 424, 599

Mast D. et al., 2014, A&A, 561, A129

Mckinney W., 2010, in van der Walt S., Millman J., eds, Proc. 9th Python in Sci. Conf., p. 56

Mihos J. C., Hernquist L., 1994, ApJ, 427, 112

Mingozzi M. et al., 2020, A&A, 636, A42

Miralles-Caballero D., Díaz A. I., Rosales-Ortega F. F., Pérez-Montero E., Sánchez S. F., 2014, MNRAS, 440, 2265

Moffat A. F. J., 1969, A&A, 3, 455

Molina J., Ibar E., Swinbank A. M., Sobral D., Best P. N., Smail I., Escala A., Cirasuolo M., 2017, MNRAS, 466, 892

Moustakas J., Kennicutt, Robert C. J., Tremonti C. A., Dale D. A., Smith J.-D. T., Calzetti D., 2010, ApJS, 190, 233

Muñoz-Elgueta N., Torres-Flores S., Amram P., Hernandez-Jimenez J. A., Urrutia-Viscarra F., Mendes de Oliveira C., Gómez-López J. A., 2018, MNRAS, 480, 3257

Poetrodjojo H. et al., 2018, MNRAS, 479, 5235

Poetrodjojo H., D'Agostino J. J., Groves B., Kewley L., Ho I. T., Rich J., Madore B. F., Seibert M., 2019, MNRAS, 487, 79

Rahner D., Pellegrini E. W., Glover S. C. O., Klessen R. S., 2017, MNRAS, 470, 4453

Rich J. A., Torrey P., Kewley L. J., Dopita M. A., Rupke D. S. N., 2012, ApJ, 753, 5

Rogers H., Pittard J. M., 2014, MNRAS, 441, 964

Rosen A. L., Lopez L. A., Krumholz M. R., Ramirez-Ruiz E., 2014, MNRAS, 442, 2701

Rupke D. S. N., Kewley L. J., Barnes J. E., 2010, ApJ, 710, L156

Rupke D. S. N., Kewley L. J., Chien L. H., 2010, ApJ, 723, 1255

Sánchez S. F. et al., 2014, A&A, 563, A49

Sánchez S. F. et al., 2016, RMxAA, 52, 171

Sánchez-Menguiano L. et al., 2016, A&A, 587, A70

Sánchez-Menguiano L. et al., 2018, A&A, 609, A119

Scott N. et al., 2018, MNRAS, 481, 2299

Silich S., Tenorio-Tagle G., 2018, MNRAS, 478, 5112

Sillero E., Tissera P. B., Lambas D. G., Michel-Dansac L., 2017, MNRAS, 472, 4404

Sutherland R., Dopita M., Binette L., Groves B., 2013, Astrophysics Source Code Library, record ascl:1306.008

The Astropy Collaboration et al., 2013, A&A, 558, A33

Toalá J. A., Arthur S. J., 2018, MNRAS, 478, 1218

Torrey P., Cox T. J., Kewley L., Hernquist L., 2012, ApJ, 746, 108

Tremblin P. et al., 2014, A&A, 568, A4

Trujillo I., Aguerri J. A. L., Cepa J., Gutiérrez C. M., 2001, MNRAS, 328, 977

Turk M. J., Smith B. D., Oishi J. S., Skory S., Skillman S. W., Abel T., Norman M. L., 2011, ApJS, 192, 9

Van Der Walt S., Colbert S. C., Varoquaux G., 2011, Computing in Science and Engineering, 13, 22

Verdolini S., Yeh S. C. C., Krumholz M. R., Matzner C. D., Tielens A. G. G. M., 2013, ApJ, 769, 12

Vogt F. P. A., Dopita M. A., Borthakur S., Verdes-Montenegro L., Heckman T. M., Yun M. S., Chambers K. C., 2015, MNRAS, 450, 2593

Wareing C. J., Pittard J. M., Falle S. A. E. G., 2017, MNRAS, 465, 2757

Wuyts E. et al., 2016, ApJ, 827, 74

Yeh S. C. C., Matzner C. D., 2012, ApJ, 757, 108

Yuan T.-T., Kewley L. J., Rich J., 2013, ApJ, 767, 106

Zaritsky D., Kennicutt R. C., Jr, Huchra J. P., 1994, ApJ, 420, 87

Zhang K. et al., 2017, MNRAS, 466, 3217

Zhou L. et al., 2017, MNRAS, 470, 4573

Zinchenko I. A., Berczik P., Grebel E. K., Pilyugin L. S., Just A., 2015, ApJ, 806, 267

SUPPORTING INFORMATION

Supplementary data are available at *MNRAS* online.

Figure S1. Error in inferred metallicity gradient as a function of spatial resolution and SNR.

Figure S2. Error in inferred metallicity gradient as a function of resolution and SNR; the results shown here are a superset of those shown in Fig. S1.

Figure S3. Same as Fig. 9 of the main text but for $\Omega = 0.05$.

Figure S4. Same as Fig. 9 of the main text but for $\Omega = 5.0$.

Figure S5. Same as Figs 9 and 10 of the main text but for $\Omega = 0.05$.

Figure S6. Same as Fig. 10 of the main text but for $\Omega = 5.0$.

Figure S7. Similar to Fig. 8 of the main text.

Figure S8. Same as in Fig. 9 of the main text but for the KD02 metallicity diagnostic.

Figure S9. Same as in Fig. 10 of the main text but for the KD02 diagnostic.

Figure S10. Similar to Fig. 9 of the main text but here the all the HII regions have been treated individually, i.e., without merging.

Table S1. Relative offsets of the inferred gradient with respect to the input gradient for the full parameter space we studied, using the procedure of fitting the gradient using all pixels.

Table S2. Same as Table S1 but now fitting the gradient using only pixels with SNR ≥ 5.

Table S3. Same as Table 1 of the main text but for $\Omega = 0.05$.

Table S4. Same as Table S3 but now fitting the gradient using only pixels with SNR ≥ 5.

Table S5. Same as Table 1 of the main text but for $\Omega = 5.0$.

Table S6. Same as Table S5 but now fitting the gradient using only pixels with SNR ≥ 5.

Table S7. Same as Table 1 of the main text but for the KD02 diagnostic.

Table S8. Same as Table S7 but now fitting the gradient using only pixels with SNR ≥ 5.

Please note: Oxford University Press is not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.

APPENDIX A: FULL PARAMETER STUDY

In this section, we present our analyses for the full parameter space, and for different values Ω, gas fraction (f_g) and different diagnostics. Table 3 gives an overview of the different choices we investigated and best-fitting parameters of the error in inferred gradient as a function of the spatial resolution for each case. The purpose of this section is to demonstrate the robustness of our key results to the different choices of the above parameters. We now discuss the dependence of our result on each of these choices individually. The tables and figures associated with this discussion are available as SOM.

A1 Dependence on gas fractions

We find no qualitative difference in our results based on the gas fraction of the simulation used, on comparing Fig. 10 and Fig. S1 of the SOM. The difference in gas fraction yields an overall offset of the data. By virtue of a higher gas content, the fiducial gas fraction ($f_g = 0.2$) simulation had formed more star particles and consequently more H II regions than the LGF ($f_g = 0.1$) simulation. This implies the former case has a higher level of overall signal from the emission lines. Thus one would require a longer exposure time to reach the same SNR when observing the $f_g = 0.1$ galaxy than the fiducial, $f_g = 0.2$ case. However, when we compare results at fixed SNR, the value of f_g does not change the qualitative result

for the exact error level or the number of beams per scale radius required to achieve maximum accuracy.

A2 Dependence on the wind parameter

Figs S3–S6 and Tables S3–S6 in the SOM denote the resolution-dependent error in metallicity, heatmaps, and tabulated values corresponding to Figs 10 and 11 in the main text. We have discussed the impact of Ω in Section 5.1.3.

APPENDIX B: MERGING H II REGIONS

In the main text, we calculate the line emission from H II regions in close proximity by merging star particles in every 40 pc^3 cell by summing their luminosities and masses, implying that multiple star particles in close proximity would be responsible for driving one large H II region. We show that this produces a CMF in good agreement with observations. As an alternative approach, in this Appendix we do not merge H II regions at all; that is, we treat each ≈ 300 M$_\odot$ star particle as driving a single H II region at its position, regardless of its proximity to any other star particles. The CMF for this case is close to a δ function at 300 M$_\odot$. We then compute the emission-line fluxes and apply the remainder of our processing pipeline as described in the main text. We show the resulting resolution-dependent errors in the metallicity gradient in fig. S10 of the SOM. This may be compared with Fig. 10, the corresponding figure for our fiducial treatment. The results are nearly identical, agreeing to ≈ 1 per cent at all resolutions and in call cases. This demonstrates that our treatment of merging has no significant effect on the resolution dependence of the error that we infer.

This paper has been typeset from a TeX/LaTeX file prepared by the author.