Monthly Notices
of the
ROYAL ASTRONOMICAL SOCIETY

MNRAS **484,** 5278–5295 (2019)
Advance Access publication 2019 January 17

doi:10.1093/mnras/stz186

The galaxy stellar mass function and low surface brightness galaxies from core-collapse supernovae

Thomas M. Sedgwick,* Ivan K. Baldry, Philip A. James and Lee S. Kelvin

Astrophysics Research Institute, Liverpool John Moores University, IC2, Liverpool Science Park, 146 Brownlow Hill, Liverpool L3 5RF, UK

Accepted 2019 January 15. Received 2019 January 14; in original form 2018 October 11

ABSTRACT

We introduce a method for producing a galaxy sample unbiased by surface brightness and stellar mass, by selecting star-forming galaxies via the positions of core-collapse supernovae (CCSNe). Whilst matching ∼2400 supernovae from the SDSS-II Supernova Survey to their host galaxies using IAC Stripe 82 legacy coadded imaging, we find ∼150 previously unidentified low surface brightness galaxies (LSBGs). Using a sub-sample of ∼900 CCSNe, we infer CCSN-rate and star formation rate densities as a function of galaxy stellar mass, and the star-forming galaxy stellar mass function. Resultant star-forming galaxy number densities are found to increase following a power law down to our low-mass limit of $\sim 10^{6.4}$ $M_{\odot}$ by a single Schechter function with a faint-end slope of $\alpha = -1.41$. Number densities are consistent with those found by the EAGLE simulations invoking a Λ cold dark matter cosmology. Overcoming surface brightness and stellar mass biases is important for assessment of the sub-structure problem. In order to estimate galaxy stellar masses, a new code for the calculation of galaxy photometric redshifts, zMedIC, is also presented, and shown to be particularly useful for small samples of galaxies.

Key words: methods: statistical – supernovae: general – galaxies: distances and redshifts – galaxies: luminosity function, mass function – galaxies: star formation.

1 INTRODUCTION

The galaxy stellar mass function (GSMF) is a direct probe of galaxy evolution, as mass is known to be a primary driver of differences in galaxy evolution. For example, Kauffmann et al. (2003) find galaxy colours, star formation rates (SFRs), and internal structure all correlate strongly with stellar mass. It is argued by Thomas et al. (2010) that early-type galaxy formation is independent of environment and controlled solely by self-regulation processes, which depend only on intrinsic galaxy properties including mass. Pasquali et al. (2009) demonstrate that star formation and active galactic nucleus (AGN) activity show the strongest correlations with stellar mass. Past attempts to measure the low-redshift GSMF have established clear evidence of a low-mass upturn in galaxy counts, indicating that low-mass galaxies dominate the galaxy population by number at current epochs (Baldry et al. 2012, henceforth B12; see also Cole et al. 2001; Bell et al. 2003; Baldry, Glazebrook & Driver 2008; Li & White 2009; Kelvin et al. 2014).

The majority of cosmological simulations today invoke a Λ cold dark matter (ΛCDM) description of our Universe, due to its ability to simultaneously reproduce various observable properties of the Universe (Perlmutter et al. 1999; Bennett et al. 2013). Despite these successes, a major challenge to the ΛCDM model today is the 'sub-structure problem'. Numbers of dwarf galaxies as predicted by straightforward simulations are significantly larger than those observed, and as a consequence, so too is the overall number of galaxies on cosmological scales (Moore et al. 1999). This discrepancy in dwarf galaxy counts is reflected in the form of the GSMF.

The observed number density of dwarf galaxies increases down to $\sim 10^8$ $M_{\odot}$, below which the form of the GSMF is uncertain (B12). Cosmological simulations of galaxy evolution such as EAGLE (Schaye et al. 2015; Crain et al. 2015) and ILLUSTRIS (Genel et al. 2014) are now sophisticated enough to attempt to assess the GSMF into the dwarf regime. In contrast to observational results, the GSMF from these simulations sees number densities continue to increase in the dwarf mass regime, approximately following a power law.

This discrepancy may not be a fault of a ΛCDM description of our Universe, however, almost all galaxy surveys suffer from a combination of magnitude and surface brightness constraints (Cross & Driver 2002; Wright et al. 2017). Most dwarf systems (typically $\lesssim 10^8$ $M_{\odot}$; Kirby et al. 2013) have intrinsically lower surface brightnesses than their higher mass counterparts, and consequently, the lower mass end of the GSMF may be underestimated due to sample incompleteness, with lower surface brightness galaxies more likely to be missed by galaxy surveys. The current

* E-mail: T.M.Sedgwick@2013.ljmu.ac.uk

observed number densities of low-mass galaxies can be treated as a lower limit when constraining evolutionary models (Baldry et al. 2008).

Knowing the precise form of the GSMF is clearly crucial should we wish to use it as a diagnostic of galaxy evolution. Developing techniques to increase completeness of the low-mass end of the GSMF must be the focus should one wish to use it to assess the nature of the physics which controls this evolution.

Here, we develop and implement one such technique, using the Stripe 82 Supernova Survey (Sako et al. 2018, henceforth, S18) to produce a sample of galaxies located at the positions of core-collapse supernovae (CCSNe). As CCSNe peak at luminosities of 10^8–10^9 $L_\odot$, they can be used as pointers to their host galaxies, which may have been missed from previous galaxy surveys due to their low surface brightness: The Palomar Transient Factory (Law et al. 2009) located low surface brightness galaxies (LSBGs) when combining SN positions with imaging taken pre-supernova or long after SN peak epoch (Perley et al. 2016). As well as aiding the identification of LSBGs, a galaxy selection using a complete sample of supernovae may significantly reduce surface brightness and magnitude biases if the host galaxy is identified for each SN in a sample.

Given that the progenitor stars of core-collapse supernovae (CCSNe henceforth) have high masses, it is natural to use them as a tracer of recent star formation. The lower mass limit for zero-age main-sequence stars that end their lives as CCSNe has been closely constrained by numerous studies, with the review of Smartt (2009) presenting a consensus value of $8 \pm 1\,M_\odot$. The upper mass limit is much more uncertain, due to the possibility that the highest mass stars may collapse directly to black holes, with no visible explosion. However, it seems likely that stars at least as massive as $30\,M_\odot$ explode as luminous CCSNe; Botticella et al. (2017) adopt an upper mass limit of $40\,M_\odot$. The corresponding range of lifetimes of CCSN progenitors is then something like 6–40 Myr, for single star progenitors (see e.g. Maund 2017); mass-exchange in high-mass binary stars can extend these lifetimes (e.g. Smith & Tombleson 2015). Even with this extension, it is clear that on the time-scales relevant for studies of galaxy evolution, rates of CCSNe can be taken as a direct and virtually instantaneous tracer of the current rate of star formation.

Several studies have made use of CCSNe as an indicator of star formation in the local Universe. On the most local scales, both Botticella et al. (2012) and Xiao & Eldridge (2015) have compared CCSN rates and integrated SFRs within a spherical volume of radius 11 Mpc centred on the Milky Way, finding good agreement between observed and predicted numbers of SNe. A similar conclusion was also reached by Cappellaro, Evans & Turatto (1999), looking at a rather more extended (mean distance ∼40 Mpc) sample of SNe and host galaxies. Other studies have used CCSNe to probe star formation at intermediate redshifts, e.g. Dahlen et al. (2004), who investigated the increase in the cosmic SFR out to redshift ∼0.7, and Botticella et al. (2017) whose sample of 50 SNe mainly occurred in host galaxies in the redshift range 0.3–1.0. Pushing to still higher redshifts, Strolger et al. (2015) have investigated the cosmic SF history out to $z \sim 2.5$ using CCSNe within galaxies from the CANDELS (Grogin et al. 2011) and CLASH (Postman et al. 2012) surveys.

Supernovae have also been used to investigate SF in different environments and types of galaxies, e.g. in starbursts (Miluzio et al. 2013) and galaxies with AGNs (Wang, Deng & Wei 2010), and to determine the metallicity dependence of the local SF rate (Stoll et al. 2013).

By selecting galaxies using CCSNe and measuring galaxy stellar masses, the resultant number densities as a function of mass imply CCSN-rate densities as a function of mass ($\rho_{\rm CCSN}$) in units of $\rm yr^{-1}\ Mpc^{-3}$, under the assumption that the CCSN sample itself is complete. By assuming a relationship between CCSN rate and SFR, we are able to trace star formation rate densities (SFRDs) ($\rm M_\odot yr^{-1}\ Mpc^{-3}$). The well-established star-forming galaxy main sequence (Noeske et al. 2007; Davies et al. 2016; McGaugh, Schombert & Lelli 2017; Pearson et al. 2018) can then be used to determine typical star formation levels expected for a given stellar mass, to infer star-forming galaxy number densities ($\rm Mpc^{-3}$) as a function of galaxy stellar mass (the GSMF), such that $\rho_{\rm CCSN} \rightarrow \rho_{\rm SFR} \rightarrow$ GSMF.

A programme similar to this work was proposed by Conroy & Bullock (2015), who suggested that SNe detected by the Large Synoptic Survey Telescope from 2021 could be used as a statistical probe of the numbers and stellar masses of dwarf galaxies. This work can be seen as a precursor to such a study.

The structure of this work is as follows. Section 2 outlines in further detail the connections between CCSN-rate density, SFRD, and the GSMF, along with the assumptions required to form them. In Section 3, we present the data sets used. In Section 4, we outline our methodology for drawing from these complete SN and galaxy samples, unbiased by magnitude and surface brightness, as well as our procedure for obtaining photometric redshift estimates. Section 5 presents resultant SFRD and star-forming GSMF estimates obtained via a CCSN host galaxy selection, where comparison is drawn with existing SFRD and GSMF results, both observational and simulated.

2 CONVERTING CCSN-RATE DENSITY TO THE STAR-FORMING GALAXY STELLAR MASS FUNCTION

In this section, we represent mathematically the CCSN-rate density, the SFRD, and the GSMF. This is in order to define the connections between them, and hence, how we are able to arrive at an estimate for the GSMF by measuring the CCSN-rate density as a function of host galaxy stellar mass ($\mathcal{M}$).

For a volume-limited sample of galaxies, the binned GSMF is defined by

$$\Phi(\mathcal{M}) = \frac{1}{\Delta \log \mathcal{M}} \frac{N(\mathcal{M})}{V} \tag{1}$$

over a mass bin of width $\Delta \log \mathcal{M}$, where N is the number of galaxies in the bin, and V is the volume. In other words, the GSMF is the number of galaxies, per unit volume, per logarithmic bin of galaxy stellar mass.

The SFRD is often estimated for the entire galaxy population, particularly as a function of redshift (Madau & Dickinson 2014), but it can also be determined as a function of galaxy mass (Gilbank et al. 2010). This can then be given by

$$\rho_{\rm SFR}(\mathcal{M}) = \frac{1}{\Delta \log \mathcal{M}} \frac{\sum_{i=1}^{N} \mathcal{S}_i\, \mathcal{M}_i}{V}, \tag{2}$$

where $\mathcal{S}_i$ is the *specific* star formation rate (SSFR) for each of the N galaxies in a bin. In other words, the SFRD is the summed SFR, per unit volume, per logarithmic bin of stellar mass.

We can approximate the SFRD by considering that the majority of star formation in the Universe occurs on the galaxy main sequence (Noeske et al. 2007; Davies et al. 2016; McGaugh et al. 2017). This sequence represents the relation, and its scatter, of SFR versus mass

for typical star-forming galaxies. The SFRD can then be given by

$$\rho_{\rm SFR}(\mathcal{M}) \;=\; \frac{1}{\Delta \log \mathcal{M}} \frac{\overline{\mathcal{S}}(\mathcal{M})\, \mathcal{M}\, N_{\rm SF}(\mathcal{M})}{V} \tag{3}$$

where $N_{\rm SF}$ is number of star-forming galaxies in the bin, $\mathcal{M}$ is the mid-point mass (assuming $\Delta \log \mathcal{M} \ll 1$), and $\overline{\mathcal{S}}$ is the mean SSFR for star-forming galaxies.

By star-forming galaxy, we mean all galaxies that are not permanently quenched or virtually quenched with minimal residual star formation. In other words, these are the galaxies that are represented by the cosmological SFRD as a function of mass. Note that in the estimate of the mean $\mathcal{S}$, we should include galaxies that are in a quiescent phase but are otherwise representative of the typical star-forming population, and our CCSN host galaxy selection method naturally leads to an appropriate contribution from such galaxies. This is relevant for low-mass galaxies that undergo more variation in their SFR with time (see e.g. Skillman 2005; Stinson et al. 2007). The mean $\mathcal{S}$ should represent an average over duty cycles in this regime.

Comparing with equation (1), we note that the SFRD can then be rewritten in terms of the GSMF of star-forming galaxies $\Phi_{\rm SF}$ as follows:

$$\rho_{\rm SFR}(\mathcal{M}) \;=\; \overline{\mathcal{S}}(\mathcal{M})\, \mathcal{M}\, \Phi_{\rm SF}(\mathcal{M}). \tag{4}$$

By using a parametrization of SSFR with galaxy stellar mass (e.g. Noeske et al. 2007; Speagle et al. 2014), it is possible to estimate the GSMF for star-forming galaxies from the SFRD or vice versa.

Next we consider the observed CCSN-rate density. Here we define this as the rate of CCSNe observed over a defined volume of space (redshift and solid angle limited), per unit volume, per logarithmic bin of galaxy stellar mass. From a non-targeted Supernova Survey, like that of Sloan Digital Sky Survey (SDSS; Frieman et al. 2008, S18), this is given by

$$\rho_{\rm CCSN,obs}(\mathcal{M}) \;=\; \frac{1}{\Delta \log \mathcal{M}} \frac{n_{\rm CCSN,obs}(\mathcal{M})}{\tau\, V}, \tag{5}$$

where $n_{\rm CCSN,obs}$ is the number of observed CCSNe associated with galaxies in the bin, and τ is the effective rest-frame time over which CCSNe could be identified. The time period of the supernova survey, in the average frame of the host galaxies (τ), is shorter than that in the observed frame (t), such that $\tau = t/(1+\overline{z})$.

The relationship between the CCSN rate and SFRD is then given by

$$\rho_{\rm CCSN,obs}(\mathcal{M}) \;=\; \rho_{\rm SFR}(\mathcal{M})\, \overline{\epsilon}(\mathcal{M})\, \overline{\mathcal{R}}(\mathcal{M}), \tag{6}$$

where $\overline{\mathcal{R}}$ is the mean ratio of CCSN rate to SFR, which is equivalent to the number of CCSNe per mass of stars formed; and $\overline{\epsilon}$ is the mean efficiency of detecting supernovae. For an apparent-magnitude-limited supernova survey, the latter function accounts for varying brightnesses and types of supernova occurring in star-forming galaxies of a given stellar mass, and the variation in extinction along different lines of sight to the supernovae.

By combining these relations we arrive at

$$\Phi_{\rm SF} \;=\; \frac{\rho_{\rm CCSN,obs}}{\overline{\epsilon}\, \overline{\mathcal{R}}\, \overline{\mathcal{S}}\, \mathcal{M}}, \tag{7}$$

which explicitly relates CCSN-rate density to the star-forming GSMF. The connection is given in terms of three functions of galaxy stellar mass: $\overline{\mathcal{S}}$ is the SSFR relation of the galaxy main sequence; $\overline{\mathcal{R}}$ is the number of CCSNe per unit mass of stars formed; and $\overline{\epsilon}$ is the efficiency of detecting CCSNe, which depends on their luminosity function, and also non-intrinsic effects of sample selection and survey strategy, in particular, the limiting CCSN detection magnitude. The basic premise is that these should be a weak function of galaxy stellar mass. We investigate the effects of varying ε on estimated CCSN-rate densities in Section 5.2, and of varying $\overline{\mathcal{S}}$ on the GSMF in Section 5.4.

3 DATA

This study makes use of three data sets, all of which are data products of the SDSS. SDSS is a large-area imaging survey of mainly the North Galactic cap, with spectroscopy of $\sim 10^6$ galaxies and stars, and $\sim 10^5$ quasars (York et al. 2000). The survey uses a dedicated, wide-field, 2.5 m telescope (Gunn et al. 2006) at Apache Point Observatory, New Mexico. A 142 megapixel camera, using a drift-scan mode (Gunn et al. 2006), gathers data in five optical Sloan broad-band filters, *ugriz*, approximately spanning the range from 3000 to 10 000 Å, on nights of good seeing. Images are processed using the software of Lupton et al. (2001) and Stoughton et al. (2002). Astrometric calibrations are achieved by Pier et al. (2003). Photometric calibrations are achieved using methods described by Hogg et al. (2001) and Tucker et al. (2006) via observations of primary standard stars observed on a neighbouring 0.5 m telescope (Smith et al. 2002).

We make use of data associated with the Stripe 82 region, a 275 deg^2 equatorial region of sky (Baldry et al. 2005). The region spans roughly $20^{\rm h} < {\rm RA} < 4^{\rm h}$ and $-1.26^\circ < {\rm Dec.} < 1.26^\circ$. Between 1998 and 2004, the region was scanned $\sim$80 times. A further $\sim$200 images were taken between 2005 and 2007, as part of the SDSS-II Supernova Survey (Frieman et al. 2008, S18).

The SDSS-II Supernova Survey data release outlined in S18 forms the basis of the supernova sample used for this study. 10 258 transient sources were identified using repeat *ugriz* imaging of the region, with light curves and follow-up spectra used for transient classifications, all of which are utilized in this study to produce an SN sample, with great care taken to ensure its completeness and the removal of non-SN transients.

We aim to produce a galaxy sample selected via the SNe which they host. Host galaxies for many of the transient sources were already identified as part of the Supernova Survey. However, in this study we revisit host galaxy identification for two reasons: (i) There is now access to deeper, coadded SDSS imaging with which to search for the host galaxy. (ii) There is often a natural bias towards assigning a transient to a higher surface brightness galaxy when one or more lower surface brightness galaxy is nearby. Our method of transient-galaxy matching is designed specifically to address this bias.

To form this galaxy sample, we make use of both single-epoch imaging and multiple epoch SDSS imaging. Single-epoch imaging published as part of SDSS-IV DR14 forms our initial galaxy sample (referred in this study as the SDSS sample for simplicity), and the sample of stars used for the removal of variable stars from our SN sample, as outlined in Sections 4.1.1 and 4.2.1, respectively. Galaxy and star classification is described in section 4.4.6 of Stoughton et al. (2002).

We then turn to coadds of multiple epoch imaging. The IAC Stripe 82 legacy project (Fliri & Trujillo 2016) performs median stacking of existing legacy Stripe 82 data, with additional complex sky-subtraction routines applied thereafter, in order to reach the extremely faint limits of the data ($\sim$28.5 mag arcsec^2 to 3σ for 10×10 arcsec^2). The IAC Stripe 82 legacy catalogue hence forms a deeper sample of objects used for this investigation. Approximately 100 single-epoch images are median stacked per SN region, to

produce the deeper imaging crucial for LSBG detection. From this coadded imaging we aim to identify additional LSBGs not found by the SDSS sample. IAC Stripe 82 image mosaicking and postage stamp creation about the positions of our SNe, crucial for host galaxy identifications, were completed using the Cutout and Mosaicking Tool, part of the ARI Survey Imaging Tools, created by one of the authors (LSK). We compare the completeness of the SDSS sample with the sample found by Fliri & Trujillo (2016) from this coadded data, as well with a SEXTRACTOR implementation designed as a bespoke search for CCSN host galaxies, using the same data (Section 4.2.3), in order to demonstrate the sensitivity of results to sample incompleteness.

We also require redshift estimates for our SN-galaxy pairs. Approximately 480 of the SN candidates have spectra of their own, from 10 sources outlined in Frieman et al. (2008). We also use host galaxy spectra for our SN-galaxy pairs once the host galaxy has been confidently identified. The galaxy spectra utilized stem from three main sources within SDSS. These are the SDSS-II Legacy (York et al. 2000), SDSS-II Southern (Baldry et al. 2005), and SDSS-III BOSS/SDSS-IV eBOSS surveys (Dawson et al. 2013, 2016). The latter contains spectroscopy for galaxies identified as the hosts of 3743 of the 10 258 SN candidates in S18, approximately a third of which were identified as non-supernovae as a result. Supernova redshifts are used in cases where both are available. Photometric redshifts of galaxies are calculated from the coadded photometry in cases where no spectroscopic redshift is available for an SN-galaxy pair, as outlined in Section 4.3.2.

4 METHODOLOGY

4.1 Selection of the supernova sample

4.1.1 Star removal

In order to produce a sample of CCSNe, we first focus on removing non-supernovae from the SDSS-II Supernova Survey sample. 10 258 transient sources were found by S18. We build on their classification attempts by first removing those transient sources categorized as variable stars (objects detected over multiple observing seasons) and AGNs (identified spectroscopically via their broad hydrogen lines).

The main SN classifications of S18 are Type II, Type Ib/c, and Type Ia SNe. We ultimately wish to remove Type Ia SNe to obtain a CCSN sample in order to trace SFRs. However, we keep likely Type Ia SNe in the sample at this stage to search for LSBGs and to increase the size of our training sample used for the estimations of galaxy redshifts, as described in Section 4.3.2. At this stage, the sample consists of 6127 transients. Of these objects, 1809 are spectroscopically confirmed SNe and a further 2305 are, photometrically, deemed very likely to be supernovae, via a combination of Bayesian, nearest-neighbour and light-curve fit probabilities (see S18 for a full description). Those remaining are classified as 'Unknown'. However, these objects may still be supernovae. For several of these objects it may simply be unclear from the photometry what type of supernova is being seen. For instance, if probabilities derived from the three aforementioned techniques give a reasonable likelihood for more than one of Type Ia, Ib/c, or II, the object will be classified as 'Unknown'.

We match transient positions with all objects of the SDSS-IV DR14 PhotoPrimary catalogue located in the Stripe 82 region with Galactic extinction-corrected r-band magnitude < 22.0 (Petrosian, psf or model) ($\sim 10^7$ objects). We refer to this as the SDSS catalogue

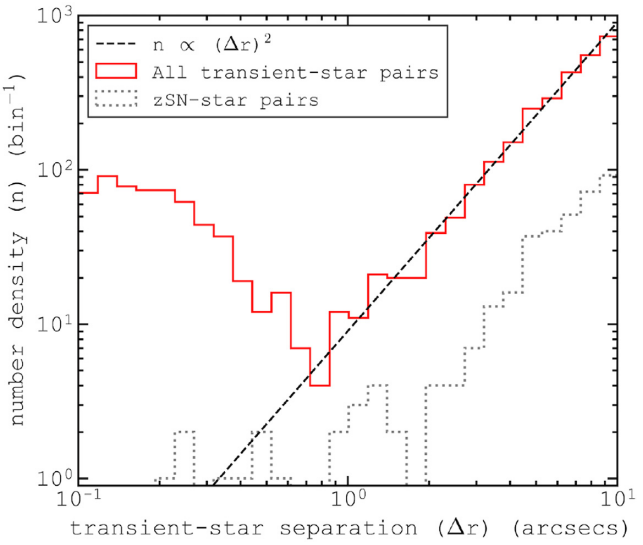

Figure 1. Number densities of transient-star pairings as a function of transient-star sky separation. The red solid line represents all possible transient-star pairings. The grey dotted line represents all spectroscopically confirmed SN-star pairings. The dashed line shows the $n = 2$ power law expected to be followed by unassociated transient-star pairs. For a transient-star separation $\lesssim$1 arcsec transients are likely to be stars.

in the remainder of this work, for simplicity. Additional variable stars are found in the SN sample by computing the separations between SDSS stars and all transients without a spectroscopic SN classification. Variable stars are identified as those objects found within 1 arcsec of an SDSS star. This 1 arcsec transient-star separation cut-off was chosen following inspection of Fig. 1.

The counts of non-associated transient-star pairs rise as the square of their separation. Additional counts arise below a separation of approximately 1 arcsec due to genuine association between the transient and SDSS object, and the detection is deemed to have arisen from the star. We remove 718 stars from the supernova sample in this manner, leaving a sample of 5549 transients. Most of these transients are likely to be SNe, but some may be quasi-stellar objects (QSOs). Redshift information helps to distinguish SNe and QSOs. However, not all of these transients have spectroscopic redshifts, as outlined in Section 3. We therefore keep all remaining transients in the sample at this stage, until each source is matched to its host galaxy, for which a spectroscopic redshift may be available.

4.1.2 SN apparent peak magnitude cut

With stars removed, we turn our attention to the completeness of the sample. The r-band supernova peak magnitude ($r_{\rm SN,peak}$) distribution is found to follow a power law up to $\sim$21.8, beyond which the slope decreases. A power-law increase to counts is expected due to the increased volume sampled as mean SN apparent magnitude becomes fainter. A deviation from this power law for $r_{\rm SN,peak} > 21.8$ hence indicates incompleteness. We henceforth implement a cut to include only supernovae brighter than $r_{\rm SN,peak} = 21.8$. Approximately 25 per cent of the aforementioned removed stars are brighter than this cut. Table 1 shows the SN sample sizes following the removal of stars and spectroscopically confirmed AGNs. SN counts are also given as a function of SN type. The $r_{\rm SN,peak} = 21.8$ SN apparent magnitude cut reduces the sample to 2931 SNe.

Table 1. Transient counts as a function of type, built from the SDSS-II Supernova Survey sample of 10 258 transients. Counts are divided into (i) those rejected by a magnitude cut, $r_{\rm SN,peak} > 21.8$, because of sample incompleteness for fainter transients; (ii) those brighter than $r_{\rm SN,peak} < 21.8$ but rejected as variable stars or AGNs; and (iii) those that are selected for our sample. Variable star counts are shown as the summation of those classified by Sako et al. (2018), by host matching to single-epoch SDSS imaging, and to Stripe 82 legacy coadded imaging. AGN counts are shown as the summation of those classified by Sako et al. (2018), by host matching to single-epoch SDSS imaging, and from redshifts that indicate the host is a QSO (see Sections 4.2.1 to 4.2.3). Counts are also subdivided into those classified using spectroscopy and those using photometry.

	Transient type	spec	phot	Total
$r_{\rm SN,peak} > 21.8$ (5257)	Ia	301	302	603
	Ib/c	17	12	29
	II	149	813	962
	SL	0	0	0
	Unknown	0	884	884
	Variable star	0	2416	2416
	AGN	363	0	363
$r_{\rm SN,peak} < 21.8$, rejected (2545)	Variable star	0	1342 + 185 + 294	1821
	AGN	543 + 98 + 83	0	724
$r_{\rm SN,peak} < 21.8$, selected (2456)	Ia	966	267	1233
	Ib/c	51	7	58
	II	274	207	481
	SL	3	0	3
	Unknown	0	681	681
Total	–	2848	7410	10 258

4.2 SN host galaxy identification

4.2.1 SDSS catalogue

Following supernova sample completeness checks, we aim to determine the correct host galaxy for each supernova. First, similar to the aforementioned transient-star matching, for each supernova we find the separation from each SDSS galaxy within a 130 arcsec radius. We then size-normalize this separation to be in units of the galaxy's Petrosian radius. To do this, we take the following steps:

(i) **Flag unreliable radii:** Galaxy Petrosian radii calculations are deemed reliable if *all* of the following SDSS flag criteria (Lupton et al. 2001) are met: (a) `NOPETRO = 0`; (b) `petroradErr_r > 0`; (c) `clean = 1`; (d) `petroR90Err_r/petroR90_r < 1`.

(ii) **Winsorize Petrosian radii:** Winsorization is the limiting of extreme values to reduce the effects of potentially spurious outliers (Hastings et al. 1947). We set a minimum radius of 2 arcsec, and if the radius is flagged as unreliable, set a maximum radius of 10 arcsec. This radius maximum prevents a galaxy with an unphysically large radius measurement from being matched to a distant, unassociated supernova. The radius minimum ensures SN-galaxy matches are not missed due to an underestimation of radius. 2 arcsec is chosen as a minimum as it approximates the radius of the lowest stellar mass galaxies known to be in SDSS Stripe 82 at the lowest redshifts in our sample (see Section 4.3.1), whereas the maximum of 10 arcsec corresponds to the 90th percentile of radius in the SDSS Stripe 82 galaxy sample.

(iii) **Account for galaxy eccentricity:** Axial ratio b/a from an exponential fit is Winsorized to $b/a > 0.5$. From axial ratio and Petrosian radius, we calculate $r_{\rm gal,proj}$: the length on the sky projected from the galaxy centre, to the edge of the galaxy ellipse, in the direction of the supernova.

SN-galaxy separation is then normalized to units of this projected galaxy radius, and for each SN, the three galaxies with the lowest normalized separations are taken as the three most likely host candidates. The Petrosian radius is chosen for this method due to the robustness of measurements over a large redshift range (Stoughton et al. 2002). To improve confidence in the most likely host galaxy for each SN, we then consider the following three factors:

(i) **Is the normalized separation reasonable?** A separation of <1.25 galaxy radii is deemed as a likely association, based on a similar analysis as seen in Fig. 1. If no galaxy lies within 1.25 radii of the SNe, the host is flagged as ambiguous.

(ii) **Are SN and galaxy redshift compatible?** We use the 10th and 90th percentiles of expected SN absolute magnitude, according to Richardson et al. (2014, henceforth, R14). Different distributions are used for each SN type. We then draw from these and the observed SN apparent magnitude a range of possible redshifts. Should the SN and galaxy redshift appear inconsistent, the host is ambiguous.

(iii) **Is the Petrosian radius reliable?** (see above).

Each SN region is inspected using IAC Stripe 82 legacy coadded *gri* imaging (Fliri & Trujillo 2016), with the above flags used to aid the search for a host galaxy. Fig. 2 shows example *gri* imaging used for inspection, with supernova position and host galaxy candidates labelled. At this stage, we only allow SNe to be assigned to galaxies in the SDSS catalogue, such that we can test the performance of the SDSS single-epoch imaging inferred sample for the task of host identification.

Following manual inspection of the coadded *gri* imaging, our procedure finds that 86 per cent of $r_{\rm SN,peak} < 21.8$ supernovae are matched to an $r < 22.0$ SDSS galaxy. 96 per cent of these are ultimately assigned to their normalized-nearest galaxy. When deciding the normalized-nearest galaxy to each SN without accounting for galaxy ellipticity, this fraction is reduced to ∼94 per cent. Approximately 2 per cent are matched to the second normalized-nearest. Reasons why the normalized-nearest galaxy is not the SN host include that fact that the galaxy extraction pipeline of SDSS sometimes catalogues a galaxy's star-forming region or the galaxy bulge as a galaxy in its own right. Another reason comes in cases of extreme galaxy eccentricities or irregular galaxy morphologies. Only ∼0.1 per cent of SNe are assigned to the third normalized-

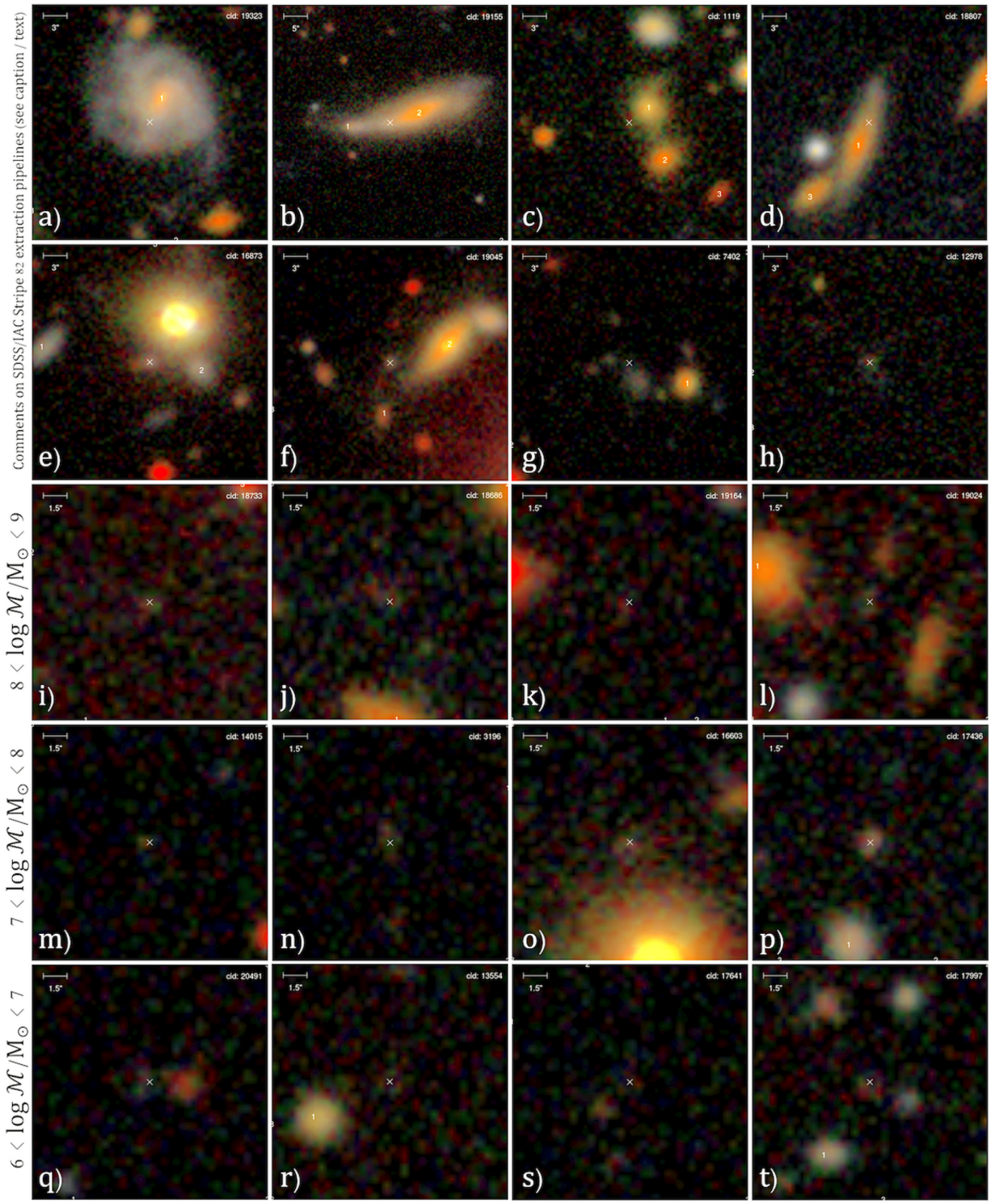

Figure 2. Example postage stamps of IAC Stripe 82 legacy coadded imaging, centred on SN positions, used to inspect each SN region to identify its host galaxy. SN catalogue ID, as listed in the SDSS-II Supernova Survey is indicated. Labels 1, 2, and 3 in each stamp indicate the central positions of the three normalized-nearest SDSS galaxies to the SN. **Row 1:** Chosen to show (a) a straightforward SN-galaxy match, (b) the successful resolution of satellite galaxies from their primaries, as well as a particularly ambiguous case, (c) resolved galaxies in group environments, and (d) a case involving extreme morphology. For (a) to (d), the normalized-nearest galaxy is chosen as the host galaxy. **Row 2**: Example SNe associated with newly identified LSBGs in IAC imaging. Uncatalogued SN hosts are missed by the SDSS and IAC catalogues due to (g) and (h) low surface brightness alone; (f) being a close-in satellite or the lower luminosity constituent of a merger; or (e) a bright neighbour. **Rows 3–5:** Examples of newly identified LSBGs. Rows bin by mean galaxy stellar mass from a Monte Carlo technique (see Section 4.3.3).

Table 2. SEXTRACTOR configurations used for a bespoke search for LSBGs. Parameters shown are the only parameters varied for each configuration, from the setup of Fliri & Trujillo (2016) (here labelled IAC_S82). 2391 identified galaxies using the setup IAC_S82 are matched to our SN catalogue. SEXTRACTOR configurations are implemented in the order (1) to (6), and photometry from the first configuration to detect the galaxy is used. Additional SN-galaxy matches with successive configurations are indicated.

SEXTRACTOR configuration	DETECT_MINAREA (pixels^2)	DETECT_THRESH	FILTER_NAME	DEBLEND_NTHRESH	DEBLEND_MINCONT	Additional SN host galaxy detections
(1) IAC_S82	4	2	default.conv	16	0.002	0
(2) default_3.0	3	1	default.conv	32	0.001	63
(3) tophat_2.5	3	1	tophat_2.5_3x3.conv	64	0.0001	28
(4) mexhat_2.5	3	1	mexhat_2.5_7x7.conv	64	0.0001	23
(5) tophat_1.5	3	1	tophat_1.5_3x3.conv	64	0.0001	18
(6) gauss_1.5	2	0.5	gauss_1.5_3x3.conv	64	0.0001	9

nearest galaxy. The remaining ∼2 per cent are assigned to another $r < 22.0$ SDSS galaxy outside of the three normalized-nearest galaxies. The most common reason for this is again the case of bright galaxies with well-resolved stellar/dust components, and in particular, where a bright bulge was catalogued as the galaxy, and the disc was missed by the extraction pipeline altogether.

We emphasize the importance of taking the described precautions before assigning SNe to host galaxies, as it is more likely that SNe belonging to dwarf or satellite galaxies are assigned to the wrong host. Matching without caution would almost certainly cause a biased selection towards brighter galaxies which would ultimately reflect in our GSMF.

Ninety-eight of the SDSS galaxies matched to SN candidates were previously classified QSOs. The AGNs were then assumed to be the source of these galaxies' transients. Removing these leaves at this stage 2422 SN candidates with hosts found in single-epoch SDSS imaging.

4.2.2 *IAC Stripe 82 legacy catalogue*

To find the host galaxy for the remaining 14 per cent of SNe, attention is turned to the deeper IAC Stripe 82 legacy catalogue (Fliri & Trujillo 2016), extracted from the same coadded *gri* imaging used for the above manual inspection. Data are formed from the median stacks of ∼100 images per supernova region, with individual epoch imaging stemming from both the SDSS main data release (Stoughton et al. 2002) and from repeat visits of the supernova regions between 2005 and 2007 as part of the SDSS SN survey (S18). Both the median stacks and SN peak-magnitude-epoch imaging are inspected to ensure that the point sources are not visible in the former, and are visible in the latter, such that neither supernova nor host galaxy are spurious detections.

The increased signal to noise achieved by the coadded imaging showed that 294 of the objects matched to SN candidates which were classified by the single-epoch imaging as galaxies were more likely to be stars. The corresponding SN candidates were removed accordingly, leaving at this stage 2128 galaxies identified from single-epoch imaging in the SN host sample.

The SDSS and IAC catalogues were linked, by matching their objects within 2 arcsec. Removing objects in common, we then repeat the previous matching procedure using Kron-based magnitudes and radii in the place of Petrosian measurements, due to the increased flexibility of Kron-based magnitudes in capturing a large fraction of the object flux (Kron 1980). Note that these objects have photometry only, and do not have redshift estimates. Therefore, we do not consider the possible redshift range, nor the trustworthiness of the radius measurement in our matching procedure.

Nevertheless, ∼1 per cent and ∼64 per cent of the previously unmatched supernovae candidates are matched to a star or to a galaxy from the IAC catalogue, respectively. This still leaves 146 supernovae not matched to either a galaxy or a star in either catalogue. We hence turn to our own SEXTRACTOR procedure, applied to the same coadded imaging, to attempt to locate the remaining host galaxies.

4.2.3 *Bespoke LSBG identification*

Care is taken to obtain photometry for the remaining host galaxies, whilst simultaneously obtaining photometry for the previously matched objects which is consistent with that found by SDSS and IAC. To do this we consider several SEXTRACTOR parameter setups (Bertin & Arnouts 1996), using different detection thresholds, minimum aperture sizes, and smoothing filters, as shown in Table 2. *ugriz* magnitudes are calculated using a fixed *r*-band defined, elliptical Kron aperture for all bands.

Each of these SEXTRACTOR setups is more prone to extracting spurious objects than the last, due to its increasingly generous extraction threshold. However, as we only keep detections corresponding to the likely supernova host, and as the supernova's presence is confirmed by visual inspection, the spurious detections are not deemed problematic.

To maximize the reliability of our photometry and to minimize the number of these false detections, we first generate a galaxy catalogue using the first SEXTRACTOR setup, i.e. the most reliable. This was the setup used by Fliri & Trujillo (2016), and we successfully reproduce their catalogue. We then complete the aforementioned SN-galaxy matching procedure using this catalogue. If the host galaxy is not found, we turn to the next SEXTRACTOR setup, to create a catalogue of the previously missed objects, and so on, for each of our setups, until a galaxy match is found for each SN. In the case of 5 SN, none of our SEXTRACTOR setups could automatically detect the galaxy host. In these cases, we assume the presence of a host galaxy and a 2.5 arcsec circular aperture is used for the galaxy's photometry, centred on the *r*-band SN position. Indeed, these five galaxies, along with their supernovae, may be spurious, but can still be used to assess uncertainties on the form of the GSMF.

Using our procedure, the photometry of the host galaxy is the photometry derived from the SEXTRACTOR setup which first located it. The galaxies which required several extraction attempts are therefore subject to the largest uncertainties. However, these uncertainties are folded into our results and can help constrain the form of the GSMF considerably.

In the case that the object is detected in the *r* band but is not detected in another band, the magnitude is set to three times the sky noise in the aperture for that band. However, we attempt to avoid using these magnitudes wherever possible. Only in the case of six galaxies are we forced to use these magnitudes, due to the lack of reliable photometry in other bands.

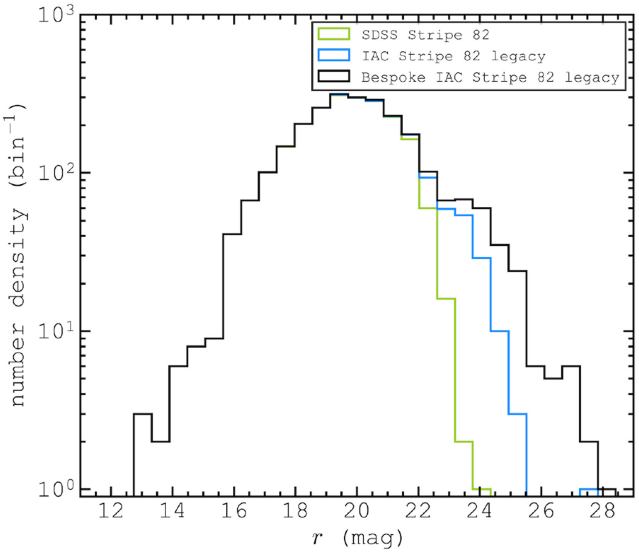

Figure 3. Galaxy number densities as a function of Kron *r*-band magnitude calculated from a bespoke search of IAC Stripe 82 legacy coadded imaging for SN host galaxies. Galaxies selected using our SN sample are shown in black; in blue, the same but only including galaxies known from the IAC Stripe 82 legacy catalogue; in green, the same but only including galaxies known from the SDSS Stripe 82 galaxy catalogue.

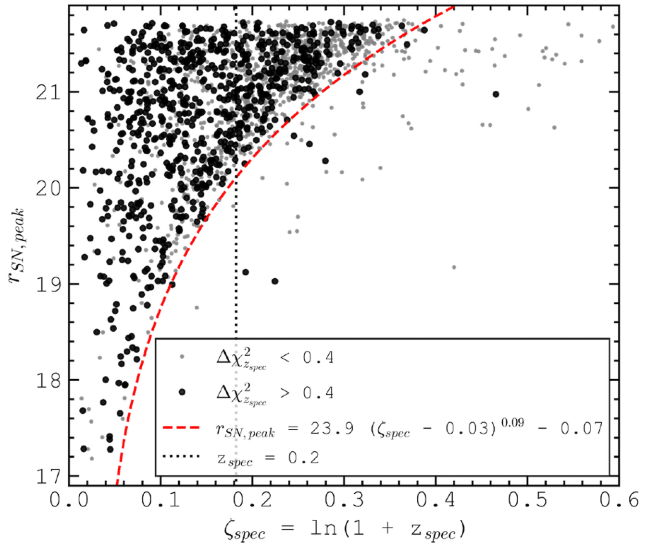

Figure 4. SN host galaxy spectroscopic redshift, expressed as $\zeta_{\rm spec} = \ln(1 + z_{\rm spec})$, versus supernova peak *r*-band apparent magnitude ($r_{\rm SN,peak}$), corrected for Galactic extinction. Black circles represent host galaxies with $\Delta\chi^2_{z_{\rm spec}} > 0.4$, representing more confident redshift estimates. Grey circles represent host galaxies with spectroscopic redshifts not satisfying this criterion. The dashed red line is the inferred line of maximum redshift as a function of $r_{\rm SN,peak}$, used to assess the validity of host galaxy identification.

Row 2 of Fig. 2 shows example IAC Stripe 82 legacy coadded imaging in which the supernova is centred on a newly identified LSBG from our bespoke search. Uncatalogued SN hosts were missed by the SDSS and IAC catalogues due to (i) low surface brightness alone; (ii) a bright neighbour; or (iii) being a close-in satellite or the lower luminosity constituent of a merger.

The different SEXTRACTOR setups shown in Table 2 were found to overcome different problems faced by the SDSS/IAC extraction pipelines. For instance, a 'Mexican hat'-type smoothing filter was particularly useful for identifying dwarf satellite galaxies close to brighter companions.

4.2.4 Summary

For the same reasons given for a galaxy being missed by the SDSS and IAC catalogues, we find that ∼2 per cent of hosts were incorrectly identified by S18. The vast majority of galaxies constituting this 2 per cent are more massive than the true SN host, and thus their inclusion in our galaxy sample would cause an underestimation of dwarf galaxy counts.

Fig. 3 shows the galaxy magnitude distributions of the 3 SN-matched catalogues presented in this work, giving a clear comparison of their depth. Recall that the SDSS sample was selected to include only $r < 22.0$ (extinction-corrected) galaxies. Thus, the best direct comparison of *r*-band magnitude depth is between the IAC data set and that of the present work. The (90th, 99th) percentiles of galaxy counts come at *r*-band magnitudes of (22.0, 23.8) for the IAC sample and (22.8, 25.3) for our bespoke sample, respectively.

We are able to test the success of our galaxy matching by using a relationship between Galactic extinction-corrected SN peak apparent magnitude ($r_{\rm SN,peak}$) and redshift. Fig. 4 shows this relationship for those SNe matched to a galaxy for which spectroscopic redshift ($z_{\rm spec}$) is known. Redshift is plotted as $\zeta = \ln(1 + z)$ (Baldry 2018). For Type Ia SNe, there is an approximate maximum SN redshift as a function of SN peak apparent magnitude. CCSNe do not exhibit a relationship that is sufficiently sharp to set a strict maximum redshift, but still follow the same underlying distribution in $r_{\rm SN,peak}$–ζ space. We can hence test if a galaxy match is reasonable by verifying that the galaxy spectroscopic redshift lies below the maximum redshift expected for that supernova.

For all host galaxies with redshifts exceeding their predicted maximum, given by the red demarcation line in Fig. 4, we perform further inspection of the coadded imaging, to check for incorrect identification of SN hosts. However, in no case do we find a more sensible host galaxy according to this extra criterion.

Not only can Fig. 4 be used to test galaxy matches, it can also be used as a further check that all of our transient objects are indeed SNe. The demarcation line implies that the maximum redshift for the faintest SNe in our bespoke sample, at $r_{\rm SN,peak} = 21.8$, is $z = 0.48$. This is hence the maximum trustworthy redshift for SNe in our sample. Bolton et al. (2012) classify objects as QSOs from their spectra. Cross-matching classifications with our host galaxies, we find a sudden rise in QSO classification at $z = 0.48$, with $\gtrsim$80 per cent of $z > 0.48$ galaxies classified as QSOs. We thus remove from the SN sample 83 transients assigned to a galaxy with $z_{\rm spec} > 0.48$, citing the fact that the transient is AGN in nature and not an SN, leaving a final sample of 2456 host galaxies for CCSNe or Type Ia SNe.

Fig. 5 shows SN-galaxy separation in arcseconds plotted against Kron galaxy magnitude. The same overall distribution is followed for each of the three galaxy sub-samples: SDSS galaxies, IAC galaxies, and our newly identified LSBGs. SN-galaxy separation increases towards brighter magnitudes due to the correlation of galaxy radius with galaxy magnitude. Normalizing separation by galaxy radius, no correlation is found between magnitude and separation. This helps confirm that the SN-galaxy separations found for our LSBGs are reasonable.

To summarize, we have successfully located the host galaxy for each of the supernovae in the sample, with great care taken to ensure

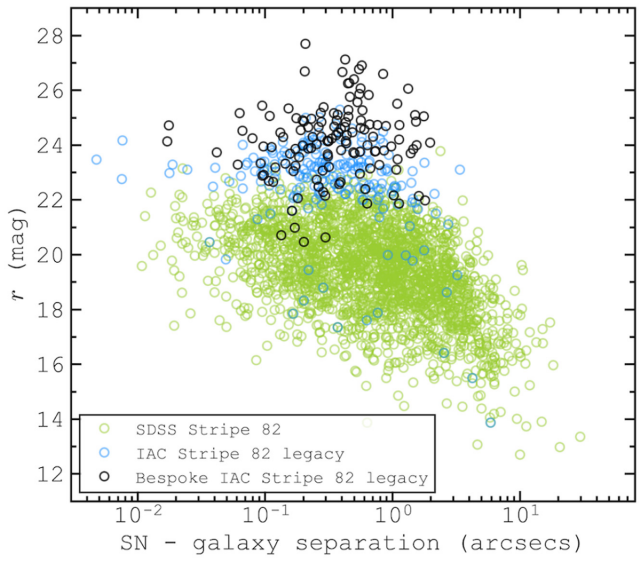

Figure 5. SN-galaxy separation in arcseconds versus galaxy Kron *r*-band magnitude calculated from our bespoke search of IAC Stripe 82 legacy coadded imaging for SN host galaxies. In black is shown galaxies found using our bespoke SN host search only; in blue, those found by the IAC Stripe 82 legacy survey; and in green, those found by the SDSS Stripe 82 survey.

the correct host is chosen. Following several steps taken to remove contamination to the SN sample from AGNs and variable stars, a sample of 2456 SN host galaxies is obtained. These galaxies were identified from single-epoch SDSS imaging (2046), the standard pipeline of multi-epoch Stripe 82 legacy imaging (262), and our bespoke search for the hosts within the multi-epoch imaging (148). We can now use the resultant galaxy sample, free of incompleteness, for the remainder of our analysis.

4.3 Stellar masses from observed colours and redshifts

Following the determination of the sample we now require galaxy stellar masses, in order to calculate CCSN-rate densities, SFR densities, and star-forming galaxy number densities as a function of galaxy stellar mass. It is often useful to define an approximate stellar mass that can be obtained from absolute magnitudes and colours (Bell et al. 2003). This works reasonably well because of the correlation between mass-to-light ratio and rest-frame colour (Taylor et al. 2011). Following Bryant et al. (2015), one can also effectively fold in the *k*-correction, and estimate a stellar mass from distances and observed magnitudes as follows:

$$\log \mathcal{M} = -0.4i + 0.4\mathcal{D} + f(z) + g(z)(g-i)_{\rm obs}, \quad (8)$$

where $\mathcal{D}$ is the distance modulus, and f and g are two functions to be determined.

To calibrate the observed-magnitude–$\mathcal{M}$ relation, we used the GAMA stellar masses that were determined by spectral energy distribution fitting (Taylor et al. 2011; Baldry et al. 2018). The data were binned in redshift over the range $0.002 < z < 0.35$, in 18 regular intervals of redshift. The values for f and g were determined for each bin, and finally $f(z)$ and $g(z)$ were fitted by polynomials. The coefficients obtained were (1.104, –1.687, 9.193, –15.15) for f and (0.8237, 0.5908, –12.84, 26.40) for g, with the constant terms first. Note because these are cubic functions, they cannot be reliable extrapolated to $z > 0.35$.

Clearly this prescription requires a galaxy redshift in order to calculate stellar mass. We use the most reliable redshift available for our galaxies, as explained below.

4.3.1 *Spectroscopic redshifts*

Where available, the most reliable redshifts for our galaxies are spectroscopic. The order of preference for the spectroscopic redshift ($z_{\rm spec}$) used is as follows:

(i) Galaxy redshifts obtained by either the SDSS-II Legacy (York et al. 2000), SDSS-II Southern (Baldry et al. 2005), or SDSS-III BOSS/SDSS-IV eBOSS surveys (Dawson et al. 2013, 2016), and derived via a χ^2 minimization method described in Bolton et al. (2012) (∼60 per cent of the galaxy sample).

(ii) In the absence of galaxy spectra, we use supernova spectroscopic redshifts, from the various sources outlined in S18 (a further ∼8 per cent of the sample).

(iii) With neither available, in cases where we are confident that the host galaxy which was missed by SDSS is tidally interacting with a galaxy possessing a spectroscopic redshift, we use that spectroscopic redshift (∼1 per cent of the sample).

Approximately 70 per cent of our galaxy sample have some form of spectroscopic redshift. For the remaining galaxies we turn to photometric redshift estimations, described below.

4.3.2 *Photometric redshifts using zMedIC*

Approximately 65 per cent of the photometry-only galaxies have SDSS KF method (kd-tree nearest-neighbour fitted) photometric redshift ($z_{\rm phot}$) estimates (Csabai et al. 2007; Beck et al. 2016). However, for galaxies only found in the IAC catalogue or from our bespoke SEXTRACTOR implementations, no galaxy redshift is available. A supernova photometric redshift is only known for ∼50 per cent of these galaxies. We set out to calculate galaxy photometric redshifts for the remainder of the sample using a new empirical method. We ultimately find this method to be the most reliable photometric redshift estimator of the three, and hence use our own photometric redshifts for all of the photometry-only galaxies.

For our photometric redshift determination, we use the IAC Stripe 82 legacy imaging (Fliri & Trujillo 2016) along with all available spectroscopic redshifts in Stripe 82. 2.5 arcsec circular-aperture-derived *ugriz* colours (used to maximize signal to noise) and their errors are utilized for the method, corrected for Galactic extinction as a function of position, using the extinction maps of Schlegel, Finkbeiner & Davis (1998).

We use a training set to calculate an empirical relationship between galaxy colours and redshift. The training set consists of a random 50 per cent of the galaxies for which all of the following conditions are met:

(i) SDSS/BOSS spectroscopic galaxy redshift is known.

(ii) The difference between the χ^2 values of the most likely and second most likely redshift, i.e. $\Delta\chi^2$, is >0.4. The most likely redshifts were determined from best-fitting redshift templates (Dawson et al. 2013).

(iii) Galaxy 2.5 arcsec aperture magnitudes have S/N > 10 in each of the optical Sloan filters, corresponding to colour errors of $\lesssim$0.15 mag.

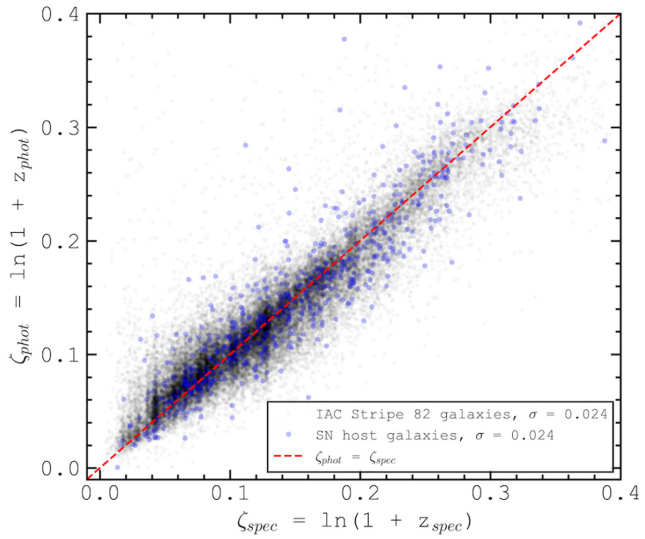

Figure 6. Spectroscopic versus photometric redshift estimates from zMedIC, derived using 2.5 arcsec circular-aperture-derived optical galaxy colours. Redshifts are represented as $\zeta = \ln(1+z)$. The validation set of ∼22 000 galaxies from the IAC Stripe 82 legacy catalogue is shown as black circles. The validation set of ∼400 SN host galaxies found from a bespoke search of the same IAC coadded imaging is shown as cyan circles. All redshifts are calculated using the zMedIC coefficients found from the training set of ∼22 000 galaxies. Weighted rms deviations (σ) in ζ are indicated.

Conditions (ii) to (iii) ensure the spectroscopic redshifts and colours we train our colour–z relation on are reliable. The resulting training and validation sets each consist of galaxies.

Equation (9) (below) is used to relate galaxy colours to redshift, giving each colour a coefficient. The optimal coefficients are those which yield an output $z_{\rm phot}$ which resembles the known $z_{\rm spec}$ for our training sample galaxy. The coefficients k_1 and k_2 in equation (9) are used as scaling values, to normalize and linearize the relationship between colours and $z_{\rm phot}$. We name this code zMedIC (**z Me**asure**d** via **I**teration of **C**oefficients).

$$\zeta_{\rm phot} = (c_1(u-g) + c_2(g-r) + c_3(r-i) + c_4(i-z) + k_1)^{k_2}\,. \quad (9)$$

Optimal coefficients are found by producing random sets of numbers to iteratively approach the set which minimizes the $\zeta_{\rm spec}$ root-mean-square (rms) deviation, given by equation (10).

$$\sigma = \sqrt{\frac{\sum_{i=1}^{n} w_i(\zeta_{\rm phot} - \zeta_{\rm spec})^2}{\sum_{i=1}^{n} w_i}}. \quad (10)$$

The weights w_i are chosen to give larger weight to confident $z_{\rm spec}$ measurements and more precise colours, and to even out weighting across redshift space. To ensure the latter, we divide the data into ζ bins of width 0.05, normalizing weights with a value, $\mathcal{F}_{\rm bin}$, different for each bin, such that the sum of weights is the same in each bin. Thus, the final form of the weights w_i is $1/(\chi^2_{z_{\rm spec}}\mathcal{F}_{\rm bin}\Delta(i-z))$. Due to the observed upper limit to redshift as a function of SN apparent magnitude, given by the red demarcation line in Fig. 4, we add a constraint such that the set of coefficients do not result in more than 10 per cent of the sample lying above this redshift limit.

Fig. 6 shows $\zeta_{\rm spec}$ versus $\zeta_{\rm phot}$ for each r-band magnitude bin, and the corresponding values of σ for our validation set. The best-fitting coefficients of equation (9) are shown in Table 3. Note that this colour–z relation is trained on both star-forming and quiescent galaxies. Removing quiescent galaxies from the training set to better reflect our SN hosts, we find no significant change to the optimal coefficients. Furthermore, training instead on only the SN sample of host galaxies for which the above training set selection criteria are met (∼400 galaxies), coefficients are similarly unaffected. We also test the effect of binning galaxies by r-band apparent magnitude, to check for potential improvements to σ. No benefit to σ is found when dividing the large training sample of ∼22 000 galaxies into magnitude bins, but for the much smaller sample of ∼400 CCSN host galaxies, binning by magnitude is found to reduce σ for the $r \leq 19.5$ galaxies by ∼20 per cent.

Table 3. Best-fitting values for the coefficients in equation (9), as inferred from a training set of ∼22 000 galaxies, with coefficients calculated separately for three bins of r-band galaxy magnitude, and for the entire training sample (labelled 'All').

	c_1	c_2	c_3	c_4	k_1	k_2
All	–0.21	0.32	–0.11	–0.03	0.18	0.90
$r \leq 17.5$	–0.17	0.55	0.13	–0.38	0.18	1.97
$17.5 < r \leq 19.5$	–0.15	0.35	–0.22	–0.12	0.13	0.74
$r > 19.5$	–0.19	0.38	–0.07	0.06	0.12	0.87

The redshift parametrization of equation (9) can be modified to deal with non-detections. The optimal coefficients are calculated with every possible combination of colours. For instance, if a non-detection is found for a galaxy in the g-band only, we are able to use the optimal coefficients associated with the colours $(u-r)$, $(r-i)$, and $(i-z)$, in order to infer photometric redshift. The most common filter with non-detections for our host galaxies is the u band. Removing $(u-g)$ colour increases σ by ∼20 per cent and ∼2 per cent for $r \leq 19.5$ and $r > 19.5$.

The method works over the redshift range of $0 < z < 0.4$. Beyond this upper limit, there are too few reliable spectroscopic redshifts in our sample for an assessment, and the chosen relationship for all galaxies may be expected to break down due to evolutionary effects and more complex k-corrections.

To emphasize, we find that equation (9) yields the most accurate redshift estimates when using galaxy colours alone. The coefficients used are specific to 2.5 arcsec circular-aperture colours. The rms deviation values between photometric and spectroscopic redshifts for this method are comparable to those using the SDSS KF method (Csabai et al. 2007; Beck et al. 2016), and are substantially better than rms deviations using an SN light-curve method (Guy et al. 2007).

With redshifts estimated, we now remove likely Type Ia supernovae from our sample to leave only likely CCSNe. In Section 5.1, we compare the effects of assumptions for the nature of the unknown-type SNe. It is likely that the unknown-type fraction of the sample consists of both CCSNe and Type Ia SNe, and it is discussed how we attempt to circumvent this problem.

4.3.3 *Monte Carlo assessment of uncertainties*

The CCSN-rate densities, SFRDs and hence GSMFs we will derive are sensitive to redshift estimates in two ways: (i) galaxy stellar masses are estimated using redshifts (see equation 8) and (ii) incompleteness of the sample is a function of redshift. Redshifts are required to volume limit the sample. As small numbers of $\log(\mathcal{M}/{\rm M}_\odot) \lesssim 8.0$ galaxies can significantly change the form of the SFRD, and hence the GSMF, we must take care when utilizing photometric redshifts.

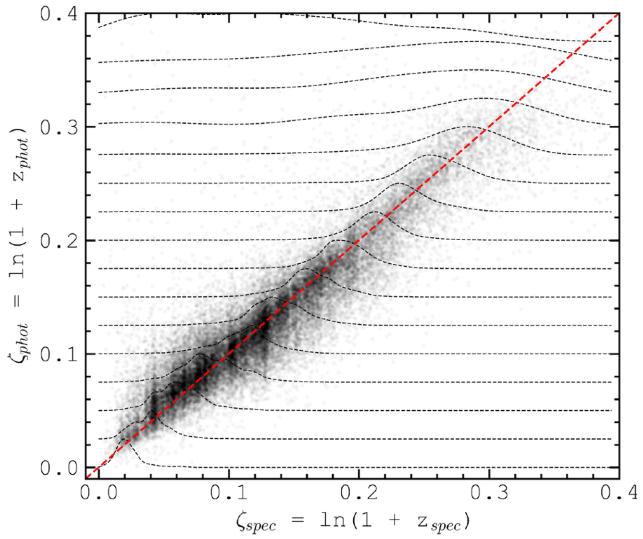

Figure 7. As in Fig. 6, but showing only the validation set of approximately ∼22 000 IAC Stripe 82 legacy galaxies. PDFs of spectroscopic redshift as a function of photometric redshift are superimposed; these were used to model photometric redshift uncertainties. PDFs are calculated from a multi-Gaussian kde, in $\zeta_{\rm phot}$ bins of width 0.025. Redshift PDFs as a function of photometric redshift estimation are implemented into a 1000-iteration Monte Carlo method to assess galaxy stellar mass uncertainties.

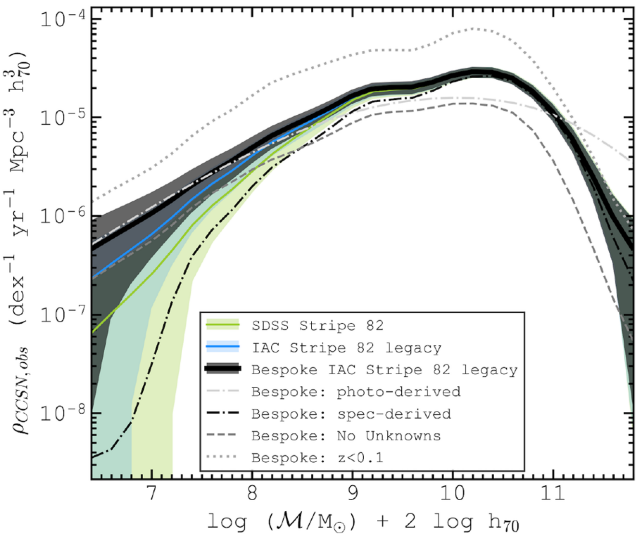

Figure 8. Volumetric CCSN-rate densities ($z < 0.2$) as a function of host galaxy stellar mass. The black solid line shows galaxies derived from a bespoke search for CCSN host galaxies in IAC Stripe 82 legacy imaging. The blue line shows CCSN host galaxies known from the IAC Stripe 82 legacy galaxy catalogue. The green line shows CCSN host galaxies known from the SDSS Stripe 82 galaxy catalogue. Shaded regions indicate the 1σ of standard deviation from a Monte Carlo method and Poisson errors. The dot–dashed grey line shows the bespoke sample but with all galaxy masses derived using photometric redshifts, while the black dot–dashed line shows all galaxy masses derived using spectroscopic redshifts (with galaxies omitted where spectroscopic redshifts are unavailable). The grey dashed line shows the bespoke sample but with all unknown-type SNe removed. The grey dotted line shows the bespoke sample but with a redshift cut of $z < 0.1$.

To circumvent this problem, we turn to a Monte Carlo technique. First, to assess the uncertainty in our photometric redshifts we divide the $\zeta_{\rm spec}$ versus $\zeta_{\rm phot}$ space into bins along the $\zeta_{\rm phot}$ axis. We then make a histogram of counts along the $\zeta_{\rm spec}$ axis for each $\zeta_{\rm phot}$ bin, smoothed via a multi-Gaussian kernel-density estimation (kde) (Parzen 1962). As such, we have the $\zeta_{\rm spec}$ distribution as a function of $\zeta_{\rm phot}$. We can use each kde as a probability density function (PDF) for a particular $\zeta_{\rm spec}$ distribution as a function of our $\zeta_{\rm phot}$ input. These distributions are shown superimposed on to the combined $\zeta_{\rm spec}$ versus $\zeta_{\rm phot}$ distribution in Fig. 7. $\zeta_{\rm phot}$ is found to be systematically greater than $\zeta_{\rm spec}$ at the highest values ($\zeta_{\rm phot} \gtrsim 0.3$). The PDF is designed to statistically account for this systematic when implemented into the Monte Carlo technique.

We then run a 1000-iteration Monte Carlo code where each photometric-galaxy's redshift is replaced by a value drawn from the PDFs of redshift shown in Fig. 7. For spectroscopic galaxies, the spectroscopic redshift is used, and varied for each Monte Carlo iteration according to its error. For each iteration we convert a galaxy's z-estimate to a luminosity distance assuming an $h = 0.7, \Omega_m = 0.3, \Omega_\Lambda = 0.7$, flat cosmology. g and i 2.5 arcsec circular-aperture-derived magnitudes, as well as elliptical Kron-aperture-derived i-band magnitude, are varied with each iteration in accordance with their uncertainties. This allows an estimate of the galaxy's stellar mass for each iteration using equation (8).

To volume limit the sample to obtain an SFRD, a $z < 0.2$ cut is made following each Monte Carlo iteration. This cut is chosen to limit the effects of galaxy evolution at higher redshifts, to match to the redshift cut of G10, for the most direct comparison of SFRDs, and to limit the effects of extinction on SN counts. At higher redshifts, more SNe are near the detection limit of the survey, where a small amount of extinction can make the SNe undetectable. Limiting the number of SN detections sensitive to extinction decreases our results' reliance on the extinction model. Using this cut, the number density of CCSNe as a function of host galaxy mass is estimated, with galaxy stellar mass bins 0.2 dex in width. We then take the mean of bin counts and the standard deviation of bin counts over the 1000 iterations, for each bin.

5 RESULTS AND DISCUSSION

5.1 Observed CCSN-rate densities

Using equation (5), we can convert number statistics of CCSNe as a function of host galaxy stellar mass into volumetric CCSN-rate densities, given effective SN rest-frame survey time τ and survey volume V. CCSN-rate densities are corrected for cosmological time dilation effects on survey time period as a function of redshift.

Fig. 8 shows volumetric $z < 0.2$ CCSN number densities as a function of host galaxy mass, derived from our Monte Carlo technique. Based on the redshift cut, sky coverage and the effective span of the Stripe 82 SN survey [$\tau \sim 270/(1+\bar{z})$ d], $\sim 10^{-7}$ CCSN yr^{-1} Mpc^{-3} h_{70}^{3} corresponds to 1 observed CCSN. As a result, we do not assess densities below $\log(\mathcal{M}/{\rm M}_\odot) = 6.4$, as below this mass the mean number counts per bin descend below 1 for the full sample of galaxies found from a bespoke search for SN hosts in IAC Stripe 82 legacy imaging. For this sample we find CCSN number densities to decrease as a power law for $\log(\mathcal{M}/{\rm M}_\odot) \lesssim 9.0$. To show the effects of selection bias we also calculate CCSN number densities using only those CCSNe assigned to hosts in the SDSS and IAC galaxy catalogues. As expected, consistency is found at higher masses, whilst a deviation in CCSN counts is found at lower host galaxy masses ($\log(\mathcal{M}/{\rm M}_\odot) < 9.0$) due to decreased sample completeness. A double peak in number density is observed, consisting of a primary peak at $\log(\mathcal{M}/{\rm M}_\odot) \sim 10.8$ and

a secondary peak at $\log(\mathcal{M}/\mathrm{M}_{\odot}) \sim 9.4$. Using the alternative stellar mass prescription of G10 in our Monte Carlo method, a modification of the prescription introduced by Kauffmann et al. (2003), we find this double peak to persist.

Error bars incorporate the uncertainties in redshift and in the stellar mass parametrization, as well as the uncertainty in the nature of the unknown-type transients. We have built on the work of S18 to deduce that these objects are very likely to be supernovae. However, each of these objects could be CC or Type Ia SNe. We use volume-limited SN number statistics, calculated by R14, to derive a ratio of CCSNe to Type Ia SNe. This gives a predicted percentage of unknown-type SNe that are Type Ia, and for each Monte Carlo iteration, this percentage of unknown-type SNe, selected at random, are removed from the sample.

Fig. 8 shows the effect of removing this fraction of unknown-type SNe from the sample. Comparing with the full sample, no strong correlation is found between the percentage of SNe that are unknown-type and galaxy stellar mass. As such, a removal of a percentage of unknown-type SNe to attempt to remove Type Ia's effectively corrects number densities by a constant amount, rather than modifying the CCSN-rate density distribution with mass. Changing the percentage of unknown-type SNe removed does not affect the presence of the double peak in CCSN-rate density.

Also shown is the sub-sample of galaxies for which spectroscopic redshift is known. Lower mass galaxies are less likely to have been selected for spectroscopic analysis. The low-mass limit of the CCSN-rate density is therefore dominated by galaxies with photometric redshifts.

To test the performance of zMedIC in producing reliable redshift estimates, we observe the effect of calculating all galaxy masses using our photometric redshifts. We see in Fig. 8, that zMedIC-derived redshifts, depicted by the grey dot–dashed series, are able to reproduce the fundamental shape of the CCSN number densities as a function of mass, but that fine features such as the double peak are not reproduced.

Also plotted in Fig. 8 is the bespoke sample's CCSN-rate densities using a redshift cut of $z < 0.1$. CCSN-rate densities are increased when using this cut, by a factor of $\sim$3. This is because we do not yet have a truly volume-limited sample, due to the supernova magnitude cut. This draws our attention to the need for corrections for SN detection efficiency, ε, as discussed in Section 5.2.

To test if the double peak in observed CCSN-rate density as a function of galaxy mass arises due to a particular SN type, we plot Fig. 9, separating the CCSN sample into Type II and tType Ib/c SN. Those with spectroscopically confirmed SN types are also plotted in isolation. For all SN types and for probable and confirmed SNe types, the double peak remains, with or without a Monte Carlo variation of masses.

5.2 Corrected CCSN-rate densities

In Section 4.1.2, we found that it is unlikely that any $r_{\mathrm{SN,peak}} < 21.8$ SNe are missed by the instrumentation described in S18 over the observing seasons, and that by locating the host galaxy for each of these SNe, surface brightness/mass biases are significantly reduced. However, the magnitude $r_{\mathrm{SN,peak}}$, which controls whether an SN is contained in our sample, may be a function of redshift, the type of CCSN that we are observing, and, most importantly for this analysis, host galaxy stellar mass.

(i) **In the case of galaxy mass**, higher mass galaxies may contain more dust than lower mass galaxies. G10 uses the Balmer decrement

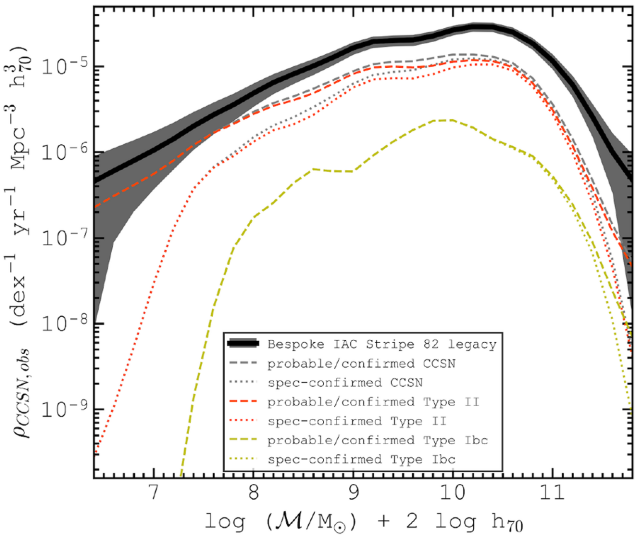

Figure 9. As Fig. 8, but also showing the full CCSN host sample subdivided into probable/confirmed CCSNe (grey dashed) and confirmed CCSNe only (grey dotted), where probable indicates those classified via photometry only, and where confirmed indicates those classified via spectroscopy. These series are subdivided further by SN type: probable/confirmed and confirmed Type II SN (red dashed, red dotted) and probable/confirmed and confirmed Type Ib/c SN (yellow dashed, yellow dotted).

to estimate the dependence of $A_{\mathrm{H}\alpha}$ on galaxy stellar mass. We test the assumption that rest-frame r-band extinction for an SN line of sight in its host galaxy follows the same extinction-mass relation as the H α line. Alternatively, extinction in CCSN environments within these galaxies may not be so strongly dependent on the extinction inferred from the Balmer decrement. In our sample, mean CCSN colours at peak epoch do not show any notable correlation with host mass. For Type Ia SNe, which are known to peak at approximately the same $(B - V)$ colour (Nugent, Kim & Perlmutter 2002), there is no correlation between colour and host mass (although environments may differ for Type Ia and CCSNe).

(ii) **In the case of redshift**, CCSNe will be fainter with distance due to inverse square dimming to flux. They will also generally experience additional dimming to r-band magnitude with redshift, in our low-redshift regime, due to the shape of their spectra. k-corrections are therefore necessary to represent the higher redshift portion of the sample correctly. CCSN k-corrections are estimated using a Type IIP template spectrum from Gilliland, Nugent & Phillips (1999) at peak magnitude for all SNe in our sample, and following equation 1 of Kim, Goobar & Perlmutter (1996).

To investigate these effects, we consider how SN detection efficiency depends on host galaxy mass and redshift. Using absolute r-band magnitude distributions of SNe as a function of SN type, we can estimate the probability that each SN is brighter than $r_{\mathrm{SN,peak}} = 21.8$, given its redshift and extinction. These probabilities lead to detection efficiencies, ε. Each CCSN contributes $1/\varepsilon$ counts to the number density within its galaxy stellar mass bin, leading to corrected CCSN-rate densities. Detection efficiencies are calculated for each Monte Carlo iteration to account for the uncertainties in redshift, mass and SN type that efficiencies depend on.

In order to estimate detection efficiency we first require an assumption about SN absolute magnitudes. The distributions used are derived from the volume-limited distributions calculated by R14, who observe approximately Gaussian distributions for each

SN type. These SN types are finer classifications than made in S18: a Type II SN as classified by S18 could be any one of IIP, IIL, IIn, or IIb in R14. Therefore, if each of the four sub-types follow different Gaussian distributions in absolute magnitude, we assume an absolute magnitude for all Type II SNe which is the sum of these four Gaussians, whilst preserving the relative counts for each sub-type. This is done similarly for Type Ib and Type Ic SNe, which come under 'Type Ib/c' in S18. The *r*-band absolute magnitude distributions used are derived from the *B*-band distributions of R14, converted using the prescription of Jester et al. (2005) for stars with $R_c - I_c < 1.15$. Assuming $(B - V) = 0.0$, as found to be typical for Type Ia SNe at peak magnitude by Nugent et al. (2002), we estimate that $M_r \sim M_B + 0.1$ for our CCSNe.

If the absolute magnitude distribution of an SN type can be approximated as a Gaussian with a mean $\overline{M}$ (and standard deviation σ), then for an individual SN of that type, the mean apparent magnitude $\overline{m}$ as a function of redshift and extinction is given by equation (11), where k_{rr} is the *r*-band *k*-correction for the SNe, $A_{r,\mathrm{h}}$ is the host galaxy *r*-band extinction and $\mathrm{A}_{r,\mathrm{MW}}$ is the Galactic *r*-band extinction along the line of sight.

$$\overline{m} = \overline{M} + 5\log d_{\mathrm{L}}(z) + k_{rr} + A_{r,\mathrm{h}} + A_{r,\mathrm{MW}}. \quad (11)$$

Again, we take the sum of the Gaussian distributions to obtain distributions for Type Ib/c and Type II SNe. By integrating under the summed Gaussian distributions, the efficiency of detection, ε, as a function of redshift and extinction, correcting for the effects of an SN peak magnitude cut at $r_{\mathrm{SN,peak}} = 21.8$ is then estimated by equation (12), where N_i are the predicted relative numbers of each SN type, used to weight the sum of the n Gaussians.

$$\epsilon = \frac{1}{2} - \frac{1}{2\sum_i^n N_i} \sum_i^n N_i \mathrm{erf}\left(\frac{\overline{m_i} - 21.8}{\sqrt{2}\sigma_i}\right). \quad (12)$$

Detection efficiency is clearly a function of SN type. We test the effects of bias in the SN classifications of S18 by varying the ratio of types Ia, Ib/c, and II SNe in the unknown-type fraction of the sample. However, it is found that any effects are of second-order importance. Therefore, we simply assume that the unknown-type SNe follow the volume-limited type ratios of R14.

For each Monte Carlo iteration, unknown-type SNe are reassigned an SN type, and are thus either removed from the sample as a Type Ia, or are given an absolute magnitude drawn from the distribution associated with either a Type Ib/c or Type II SN, which enables an estimate of their detection efficiency for each iteration.

Equations (11) and (12) are sensitive to assumptions for host galaxy extinction, $A_{r,\mathrm{h}}$. Significant uncertainty exists around the relationship between CCSN extinction and host galaxy mass. Fig. 10 shows corrected CCSN-rate density versus redshift, for different values of $A_{r,\mathrm{h}}$. In comparison to the observed values, clearly there are larger corrections at higher redshift and with higher assumed $A_{r,\mathrm{h}}$.

Also shown in Fig. 10 is the inferred CCSN-rate density derived from the star formation history of Madau & Dickinson (2014), assuming different values for $\mathcal{R}$, the expected number of stars that explode as CCSNe per unit mass of stars formed. Madau & Dickinson (2014) assume a Salpeter IMF for the star formation history. Using this IMF with initial masses 0.1–100 $\mathrm{M}_\odot$ and assuming that all stars with initial masses 8–40 $\mathrm{M}_\odot$ result in CCSNe, then $\log \mathcal{R} = -2.17$. We find that using values of $A_{r,\mathrm{h}}$ from ∼0.3 to 0.6 reproduces the evolution of CCSN density with redshift, with $\log \mathcal{R}$ in the range −2.2 to −1.8. We adopt $\log \mathcal{R} = -1.9$.

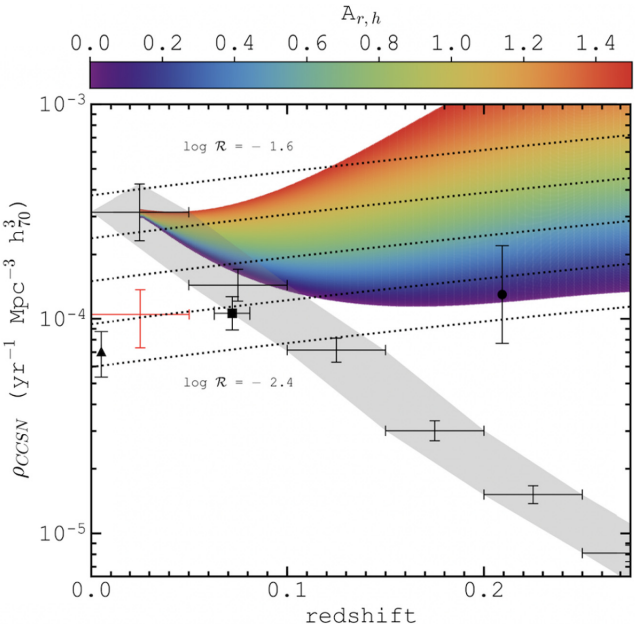

Figure 10. Volumetric CCSN-rate density versus redshift, calculated in redshift bins of 0.05, and as a function of SN detection efficiency corrections. The grey shaded region depicts observed CCSN-rate densities from our CCSN sample, uncorrected for SN detection efficiencies, bounded by 1σ Poisson + Monte Carlo + Cosmic Variance errors. Colour-shaded density values are corrected for SN detection efficiency. The colour bar represents the different corrected CCSN rates obtained when assuming different values of CCSN host galaxy extinction, $A_{r,\mathrm{h}}$. Dashed lines represent expected CCSN-rate histories derived from the star formation history of Madau & Dickinson (2014), assuming different values of log $\mathcal{R}$ in the range –1.6 to –2.4 (see the text). Triangular, square, and circular points represent CCSN-rate densities obtained by Li et al. (2011), Taylor et al. (2014), and Botticella et al. (2008), respectively. The red point shows how our observed $z < 0.05$ SN counts are reduced when objects inferred to have peak absolute magnitudes > -15 are discounted as CCSNe.

Note though that the measured $z < 0.05$ rate is higher than expected. We find that there is a significant excess of SNe with very faint peak *r*-band absolute magnitudes ($M_{r,\mathrm{sn}} > -15$), only found for this redshift bin. R14 predict the fraction of CCSNe with $M_{r,\mathrm{sn}} > -15$ to be negligible. One explanation for this excess could be a contamination of the SN sample at these lowest redshifts from outbursts of Luminous Blue Variables, which can exhibit similar light-curve properties to Type IIn supernovae (see e.g. Goodrich et al. 1989; Van Dyk et al. 2000). The red point of Fig. 10 shows the effect of removing objects inferred to have $M_{r,\mathrm{sn}} > -15$ on the $z < 0.05$ CCSN counts. This corrected value is in agreement with Li et al. (2011) and Taylor et al. (2014).

The effects of corrections for SN detection bias on CCSN number densities as a function of galaxy stellar mass are shown in Fig. 11 (with $\mathrm{A}_{r,\mathrm{h}} = 0.5$). CCSN-rate densities ($z < 0.2$) are now higher than the uncorrected values by a factor of ∼2 at the low-mass limit, and by a factor of ∼3 at $\log(\mathcal{M}/\mathrm{M}_\odot) = 10.6$, the position of the primary peak. Even though host galaxy extinction is set constant with galaxy stellar mass, this larger factor at higher masses indicates the effects of a weak SN-type dependence on galaxy stellar mass.

The uncertainties take account of the efficiency corrections in Fig. 11. Even with these uncertainties, corrected CCSN-rate densities using the full sample of CCSN hosts show a power-law decrease with decreasing stellar mass, down to the low-mass limit of the sample. This is not evident from the samples that do not use the

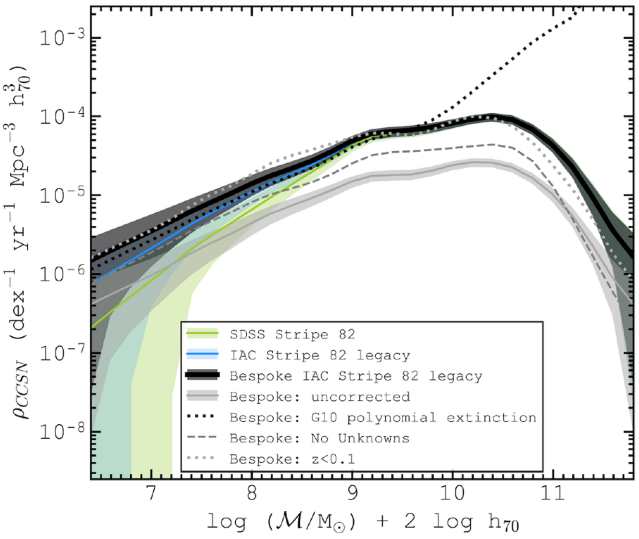

Figure 11. Volumetric $z < 0.2$ CCSN-rate densities as a function of host galaxy stellar mass, corrected for the effects of SN detection efficiencies. All series are as described in Fig. 8. All results assume constant host galaxy extinction with mass, $A_{r,h} = 0.5$ mag, except the black dotted line which assumes polynomially increasing host galaxy extinction with mass (as in G10). The grey solid line and shaded region represent CCSN-rate densities uncorrected for SN detection efficiencies and their 1σ errors, respectively.

bespoke LSBG detections. The 1σ levels indicated by the shaded regions suggest non-zero number densities from single-epoch SDSS imaging only down to $\log(\mathcal{M}/\mathrm{M}_\odot) = 7.2$, and down to $\log(\mathcal{M}/\mathrm{M}_\odot) = 6.8$ for IAC Stripe 82 coadded imaging. The $z < 0.1$ densities now approximate the $z < 0.2$ densities across the mass range, indicating the validity of corrections.

The black dotted line of Fig. 11 shows CCSN-rate densities assuming increasing extinction with mass as in G10. High-mass counts appear to be significantly overestimated compared to predictions for the SFRD (see Section 5.3). This discrepancy is not surprising because G10 uses the Balmer decrement and we expect this to be an overestimate of typical CCSN extinction for two reasons: (i) Balmer line production is weighted towards higher luminosity stars, and therefore younger phases of star clusters (∼1–5 Myr, G10); (ii) SNe occur at the end of a star's life, counts are weighted by number in the IMF, and therefore SNe typically occur later in the life of star clusters (∼10–40 Myr). While we still expect some increase of extinction with host galaxy mass due to the interstellar medium, we assume a constant extinction with mass for the remaining results of this work.

5.3 Star formation rate density

To convert CCSN-rate densities into SFRDs, we require an assumption for the CCSN rate per unit of star formation, $\mathcal{R}$, as discussed in Section 2. Using a value of $\log \mathcal{R} = -1.9$, we obtain the SFRDs plotted in Fig. 12, as a function of galaxy stellar mass. Our SFRDs are consistent with the G10 SFRDs ($\mathrm{SFR_{tot}}$) between $9.0 < \log(\mathcal{M}/\mathrm{M}_\odot) < 11.0$, the stellar mass range for which the G10 galaxy sample is expected to be complete.

Using our bespoke search for the CCSN hosts in IAC Stripe 82 legacy imaging, results are sufficiently constrained to deduce a power-law decrease to SFRDs with decreasing galaxy stellar mass, down to the low-mass limit. Our method allows for an estimation of SFRs 1.6 dex lower in stellar mass than achieved by G10, who calculated SFRs directly from galaxy emission lines.

5.4 The galaxy stellar mass function

For the previous results of this work, no assumptions are required for the volumetric numbers of galaxies at each stellar mass. On the contrary, by making assumptions about SSFR levels with mass, volumetric galaxy number densities as a function of mass can be estimated.

The star-forming GSMF can be derived from CCSN-rate densities using equation (7), requiring an assumption for mean specific CCSN-rate variation with galaxy stellar mass ($\mathcal{M}$). Specific CCSN rates are expected to trace $\mathcal{S}$ irrespective of $\mathcal{M}$ (see Section 1). CCSN rates may in fact be the most direct tracers of SFRs. Studies which attempt to measure both $\mathcal{S}$ and specific CCSN rates find consistent slopes with $\mathcal{M}$ (see e.g. Graur, Bianco & Modjaz 2015). This gives confidence in our assumption of a constant $\mathcal{R}$ value with $\mathcal{M}$, as assumed to estimate SFRDs in Section 5.3. With this similar trend in mind we can use the more numerous studies of $\mathcal{S}$ versus $\mathcal{M}$ to suggest the sensible range of assumptions for specific CCSN rate versus mass, required to derive the star-forming GSMF.

There exist conflicting results in the literature for the variation of $\mathcal{S}$ versus $\mathcal{M}$. Whilst some studies find much higher efficiencies towards lower masses (Zheng et al. 2007; Karim et al. 2011; Li et al. 2011; Graur et al. 2015), where typically $\mathcal{S} \propto \mathcal{M}^{-0.5}$, others find much shallower trends consistent with a constant $\mathcal{S}$ (see e.g. Bell et al. 2007; James, Prescott & Baldry 2008; Wuyts et al. 2011, and notably, S18).

The majority of these studies do not probe down to the masses studied in this work. Uncertainty exists around whether the $\mathcal{S}$ versus $\mathcal{M}$ relations found for massive galaxies apply to the dwarf regime down to $\log(\mathcal{M}/\mathrm{M}_\odot) = 6.4$. McGaugh et al. (2017) find a distinct star-forming main sequence for $7 \lesssim \log(\mathcal{M}/\mathrm{M}_\odot) \lesssim 10.0$, with results consistent with a constant $\mathcal{S}$. Rate simulations of Graur et al. (2015) predict a tanh-like function to specific CCSN rate with galaxy mass, with constant specific rates in the dwarf regime and at the highest galaxy masses ($>10^{11}$), and with decreasing $\mathcal{S}$ with increasing mass for the masses in between.

In this work, we are most interested in the shape of the GSMF in the dwarf regime. High-mass ($\log(\mathcal{M}/\mathrm{M}_\odot) \gtrsim 9.0$) galaxy counts in this work are subject to uncertainties related to the modelling of host galaxy extinction in massive galaxies, and the star-forming GSMF is already well constrained at high masses by several independent works finding consistent number densities (e.g. Baldry et al. 2008; Li & White 2009; Kelvin et al. 2014). Consider instead the mass range of $8.0 < \log(\mathcal{M}/\mathrm{M}_\odot) < 9.0$. The variation of SF galaxy counts with mass in this range is also well constrained, yet the effects of varying assumptions for host galaxy extinction with mass are significantly smaller than at higher masses. Hence, we would expect that the $\mathcal{S}$ versus $\mathcal{M}$ relation which produces a slope consistent with well-constrained number densities from previous studies is the most reliable relation.

We find the best-fitting $\mathcal{S}$ versus $\mathcal{M}$ log–log relation for $8.0 < \log(\mathcal{M}/\mathrm{M}_\odot) < 9.0$ to B12 star-forming galaxy number densities (computed from GAMA data; Driver et al. 2011) to have a slope of $-0.08^{+0.08}_{-0.10}$. We can express equation (7) in the following form:

$$\phi_{\mathrm{SF}} \propto \rho_{\mathrm{CCSN}} \mathcal{M}^{-\gamma}. \quad (13)$$

For a constant ratio of CCSN rate to SF rate with mass, our best-fitting $\mathcal{S}$ versus $\mathcal{M}$ slope leads to $\gamma = 0.92^{+0.08}_{-0.10}$. We find a constant value of γ with mass is able to give good agreement with

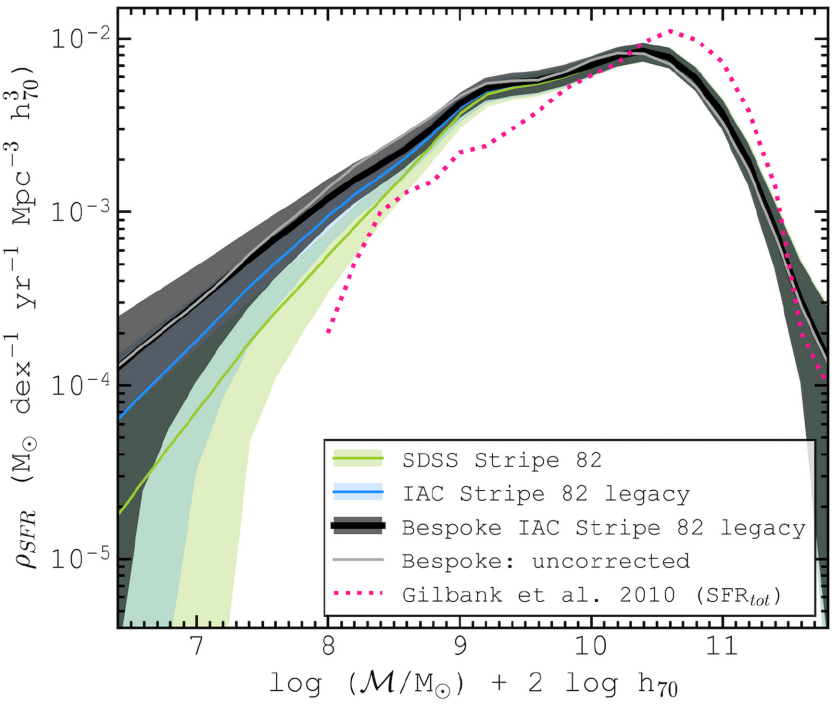

Figure 12. Volumetric $z < 0.2$ SFRDs as a function of host galaxy stellar mass. The black line depicts galaxies selected from a bespoke search for CCSN host galaxies in IAC Stripe 82 legacy imaging. The blue line shows galaxies from the IAC Stripe 82 legacy galaxy catalogue. The green line shows galaxies from the SDSS Stripe 82 galaxy catalogue. Shaded regions indicate 1σ of standard deviation from Monte Carlo, Poisson, and cosmic variance errors. The grey line depicts the same as the black solid line but uncorrected for SN detection efficiencies. The magenta dotted line indicates G10 'SFR$_{\rm tot}$' SFRDs.

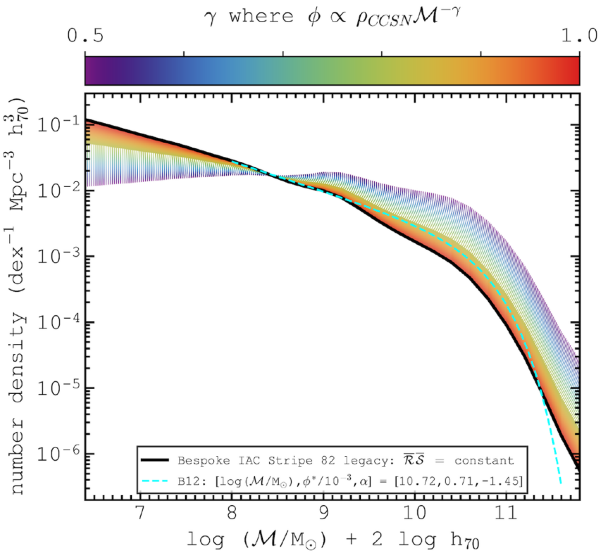

Figure 13. The $z < 0.2$ star-forming GSMF, as a function of the parameter $\gamma(\mathcal{R}, \mathcal{S})$. The black line shows the GSMF-derived assuming constant specific CCSN rate with stellar mass, i.e. $\gamma = 1.0$. The solid region shows star-forming galaxy number densities corresponding to the 1σ uncertainty level on the best-fitting value of γ to the B12 star-forming GSMF (the cyan dashed line) in the range $8.0 < \log(\mathcal{M}/\mathrm{M}_\odot) < 9.0$. The hatched region shows number densities derived using lower values of γ.

B12 star-forming galaxy number densities, across their full range of masses. We show the effects of different values of γ on the derived star-forming GSMF in Fig. 13. This shows the typical range for $\mathcal{S}$ versus $\mathcal{M}$ found in the literature.

Given our focus is on low-mass galaxies, we assume constant $\mathcal{S}$ with mass ($\gamma = 1.0$), and normalize number densities to those of B12 in the mass range $8.0 < \log(\mathcal{M}/\mathrm{M}_\odot) < 9.0$, to derive the star-forming galaxy number densities (Mpc^{-3} dex^{-1}) for $6.4 < \log(\mathcal{M}/\mathrm{M}_\odot) < 11.8$. This is shown in Fig. 14 with the Monte Carlo, Poisson and cosmic variance uncertainties. We once more show the results based on the three galaxy samples derived using SDSS Stripe 82 galaxies, IAC Stripe 82 legacy galaxies, and those from our bespoke SN host search of IAC Stripe 82 legacy imaging.

Using the full sample of CCSN host galaxies, we observe a continuation of a power-law rise in galaxy number density with decreasing mass. When selecting host galaxies from the IAC and SDSS catalogues, incompleteness is found below masses of $\log(\mathcal{M}/\mathrm{M}_\odot) \sim 9.0$, and zero number densities cannot be ruled out below masses of $\log(\mathcal{M}/\mathrm{M}_\odot) = 6.8$ and $\log(\mathcal{M}/\mathrm{M}_\odot) = 7.2$, respectively.

We find that our result for the full sample is consistent with the Schechter function fit to B12 star-forming galaxy number densities, when extrapolating the fit below $\log(\mathcal{M}/\mathrm{M}_\odot) = 8.0$. B12 use a fit with parameters [log ($\mathcal{M}$/M$_\odot$), $\phi*/10^{-3}$ dex^{-1} Mpc^{-3}, α] = [10.72, 0.71, –1.45], which is plotted in Fig. 14. For our method, using $\gamma = 1.0$, we find the best-fitting parameters to be [10.54, 1.32, –1.41], obtained using a Levenberg–Marquardt algorithm applied to the full mass range shown in Fig. 14.

We also compare with the GSMF derived from G10's SFRD assuming constant SSFR with galaxy mass. Between $8.4 < \log(\mathcal{M}/\mathrm{M}_\odot) < 11.0$ the gradient of number densities with mass is consistent with the B12 equivalent, giving support to this assumption.

Although we show the star-forming GSMF rather than that of all galaxies, it is expected that the low-mass population is dominated by star-forming galaxies (B12). Therefore, by constraining the

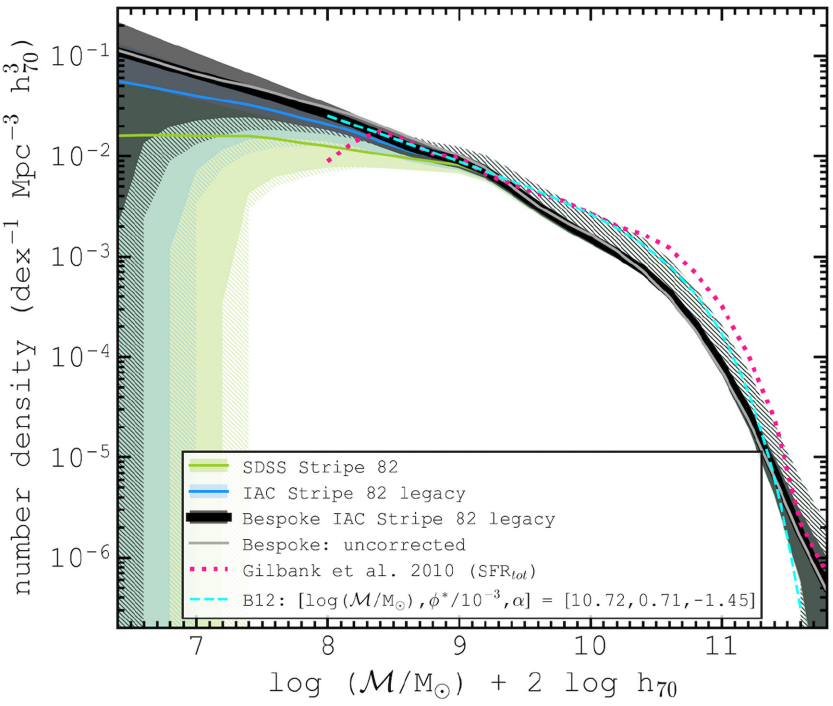

Figure 14. The $z < 0.2$ star-forming GSMF: star-forming galaxy number densities as a function of stellar mass, in logarithmic units of solar mass, as derived from corresponding SFRDs (see Fig. 12). The cyan dashed line represents the Schechter fit to star-forming galaxy number densities obtained by B12. Hatched regions represent additional uncertainties on top of observational uncertainties concerning the optimal model of SSFR versus galaxy mass (see Fig. 13).

low-mass end of the star-forming GSMF we are putting strong constraints on the form of the total GSMF. We find our low-mass number densities to be consistent with those from the EAGLE simulations (Schaye et al. 2015) assuming a standard ΛCDM cosmology. Given that a GSMF with incompleteness in the dwarf regime could be mistaken as evidence for tension with the standard cosmology, it is clear that our overcoming of surface brightness and stellar mass biases is crucial for an assessment of the sub-structure problem. Consistency is also found with the low-mass galaxy number densities of Wright et al. (2017), derived from a method used to estimate and correct for surface brightness incompleteness. We find an upper limit of 0.1 dex^{-1} Mpc^{-3} at 10^7 M$_\odot$ that is on the low end of their results, which do not distinguish between star-forming and quenched galaxies.

6 SUMMARY AND CONCLUSIONS

Using an SNe sample (Sako et al. 2018) as pointers to their host galaxies, approximately 150 newly identified LSBGs have been located in IAC Stripe 82 legacy coadded imaging (Fig. 2). This results in a significant improvement to magnitude depth of a CCSN-selected galaxy sample (Fig. 3). A galaxy selection using a complete sample of CCSNe removes surface brightness bias.

In order to estimate stellar masses of host galaxies without spectroscopic redshifts, we use a colour-based code, called zMedIC, that uses the strong correlations between Sloan *ugriz* colours and redshift. The parametrization used is shown to work for a sample containing both star-forming and quiescent galaxies at $z < 0.4$ (Fig. 6). In order to assess uncertainties on CCSN-rate densities as a function of galaxy stellar mass, we employ a Monte Carlo method to fold in errors on photometric redshift (Fig. 7) and *ugriz* photometry. The observed CCSN-rate densities as a function of mass are shown to peak at $\sim 10^{10.5}$ M$_\odot$ and drop by a factor of $\sim$100 down to $10^{6.5}$ M$_\odot$ (Fig. 8).

We use a model to correct CCSN-rate densities for SN detection efficiencies (ε) that depend on host galaxy extinction, Galactic extinction, SN type (Richardson et al. 2014) and redshift. Corrected CCSN-rate densities are consistent with expectations derived from the cosmic star formation history (Madau & Dickinson 2014) assuming $\log \mathcal{R} \simeq -2$ (Fig. 10), where $\mathcal{R}$ is the expected number of stars that explode as CCSNe per solar mass of stars formed.

By assuming a value for $\mathcal{R}$, the corrected CCSN-rate density as a function of stellar mass (Fig. 11) can then be scaled to the SFRD (Fig. 12). The SFRD is consistent with the emission-line derived SFRD at high masses (Gilbank et al. 2010) but our method extends the measurement to lower masses. By assuming a constant SSFR ($\mathcal{S}$), and scaling appropriately (to Baldry et al. 2012), we convert the SFRD to the star-forming GSMF (Fig. 14). Derived star-forming galaxy number densities are found to rise as a power law with decreasing galaxy stellar mass down to the lowest assessed masses of $\log(\mathcal{M}/\mathrm{M}_\odot) = 6.4$; and even at the lower end of our estimated uncertainties, there is no turnover in the number density of star-forming galaxies per unit log mass at least down to $\log(\mathcal{M}/\mathrm{M}_\odot) \sim 7$.

We have demonstrated a method which significantly reduces tension between observations and the simulated predictions of galaxy number densities derived via a standard ΛCDM cosmology. The lower detection limit to galaxy stellar mass for the SN method outlined in this paper depends on the area, depth and the observing period of the SN survey. A future sample derived from LSST time-series and coadded imaging could significantly increase the number of reliably identified CCSN hosts at $z < 0.2$. As part of this, deep multiband photometry is crucial for constraining photometric redshifts and stellar masses of LSBGs. This would

enable measurements to even lower masses, and future work would also enable a more detailed assessment of the functional form of $\mathcal{R}$, ε, and $\mathcal{S}$ on the conversion from observed CCSN-rate density to SFRD and then to GSMF.

ACKNOWLEDGEMENTS

TMS acknowledges funding from STFC and LJMU.

Funding for the SDSS-II, SDSS-III, and SDSS-IV has been provided by the Alfred P. Sloan Foundation, the National Science Foundation, the U.S. Department of Energy Office of Science, the National Aeronautics and Space Administration, the Japanese Monbukagakusho, the Max Planck Society, the Higher Education Funding Council for England, and the Participating Institutions. SDSS-IV acknowledges support and resources from the Center for High-Performance Computing at the University of Utah. The SDSS web site is www.sdss.org.

SDSS-II, SDSS-III, and SDSS-IV are managed by the Astrophysical Research Consortium for the Participating Institutions of the SDSS Collaboration including the American Museum of Natural History, University of Arizona, Astrophysical Institute Potsdam, University of Basel, the Brazilian Participation Group, Brookhaven National Laboratory, University of Cambridge, the Carnegie Institution for Science, Carnegie Mellon University, Case Western Reserve University, University of Chicago, the Chilean Participation Group, University of Colorado Boulder, Drexel University, Fermilab, University of Florida, the French Participation Group, the German Participation Group, Harvard-Smithsonian Center for Astrophysics, Harvard University, the Institute for Advanced Study, Instituto de Astrofísica de Canarias, the Japan Participation Group, Johns Hopkins University, the Joint Institute for Nuclear Astrophysics, the Kavli Institute for Particle Astrophysics and Cosmology, Kavli Institute for the Physics and Mathematics of the Universe (IPMU)/University of Tokyo, the Korean Scientist Group, the Chinese Academy of Sciences (LAMOST), Lawrence Berkeley National Laboratory, Leibniz Institut für Astrophysik Potsdam (AIP), Los Alamos National Laboratory, Max-Planck-Institut für Astronomie (MPIA Heidelberg), Max-Planck-Institut für Astrophysik (MPA Garching), Max-Planck-Institut für Extraterrestrische Physik (MPE), the Michigan State/Notre Dame/JINA Participation Group, Universidad Nacional Autónoma de México, National Astronomical Observatories of China, New Mexico State University, New York University, University of Notre Dame, Observatário Nacional/MCTI, Ohio State University, University of Oxford, Pennsylvania State University, University of Pittsburgh, University of Portsmouth, Princeton University, Shanghai Astronomical Observatory, the Spanish Participation Group, University of Tokyo, the United States Naval Observatory, University of Utah, Vanderbilt University, United Kingdom Participation Group, University of Virginia, the University of Washington, University of Wisconsin, Vanderbilt University, and Yale University.

REFERENCES

Baldry I. K. et al., 2005, MNRAS, 358, 441
Baldry I. K. et al., 2012, MNRAS, 421, 621(B12)
Baldry I. K. et al., 2018, MNRAS, 474, 3875
Baldry I. K., 2018, preprint(arXiv:1812.05135)
Baldry I. K., Glazebrook K., Driver S. P., 2008, MNRAS, 388, 945
Beck R., Dobos L., Budavári T., Szalay A. S., Csabai I., 2016, MNRAS, 460, 1371
Bell E. F., McIntosh D. H., Katz N., Weinberg M. D., 2003, ApJS, 149, 289
Bell E. F., Zheng X. Z., Papovich C., Borch A., Wolf C., Meisenheimer K., 2007, ApJ, 663, 834
Bennett C. L. et al., 2013, ApJS, 208, 20
Bertin E., Arnouts S., 1996, A&AS, 117, 393
Bolton A. S. et al., 2012, AJ, 144, 144
Botticella M. T. et al., 2008, A&A, 479, 49
Botticella M. T. et al., 2017, A&A, 598, A50
Botticella M. T., Smartt S. J., Kennicutt R. C., Cappellaro E., Sereno M., Lee J. C., 2012, A&A, 537, A132
Bryant J. J. et al., 2015, MNRAS, 447, 2857
Cappellaro E., Evans R., Turatto M., 1999, A&A, 351, 459
Cole S. et al., 2001, MNRAS, 326, 255
Conroy C., Bullock J. S., 2015, ApJ, 805, L2
Crain R. A. et al., 2015, MNRAS, 450, 1937
Cross N., Driver S. P., 2002, MNRAS, 329, 579
Csabai I., Dobos L., Trencséni M., Herczegh G., Józsa P., Purger N., Budavári T., Szalay A. S., 2007, Astron. Nachr., 328, 852
Dahlen T. et al., 2004, ApJ, 613, 189
Davies L. J. M. et al., 2016, MNRAS, 461, 458
Dawson K. S. et al., 2013, AJ, 145, 10
Dawson K. S. et al., 2016, AJ, 151, 44
Driver S. P. et al., 2011, MNRAS, 413, 971
Fliri J., Trujillo I., 2016, MNRAS, 456, 1359
Frieman J. A. et al., 2008, AJ, 135, 338
Genel S. et al., 2014, MNRAS, 445, 175
Gilbank D. G., Baldry I. K., Balogh M. L., Glazebrook K., Bower R. G., 2010, MNRAS, 405, 2594(G10)
Gilliland R. L., Nugent P. E., Phillips M. M., 1999, ApJ, 521, 30
Goodrich R. W., Stringfellow G. S., Penrod G. D., Filippenko A. V., 1989, ApJ, 342, 908
Graur O., Bianco F. B., Modjaz M., 2015, MNRAS, 450, 905
Grogin N. A. et al., 2011, ApJS, 197, 35
Gunn J. E. et al., 2006, AJ, 131, 2332
Guy J. et al., 2007, A&A, 466, 11
Hastings C., Mosteller F., Tukey J. W., Winsor C. P., 1947, Ann. Math. Stat., 18, 413
Hogg D. W., Finkbeiner D. P., Schlegel D. J., Gunn J. E., 2001, AJ, 122, 2129
James P. A., Prescott M., Baldry I. K., 2008, A&A, 484, 703
Jester S. et al., 2005, AJ, 130, 873
Karim A. et al., 2011, ApJ, 730, 61
Kauffmann G. et al., 2003, MNRAS, 341, 33
Kelvin L. S. et al., 2014, MNRAS, 444, 1647
Kim A., Goobar A., Perlmutter S., 1996, PASP, 108, 190
Kirby E. N., Cohen J. G., Guhathakurta P., Cheng L., Bullock J. S., Gallazzi A., 2013, ApJ, 779, 102
Kron R. G., 1980, ApJS, 43, 305
Law N. M. et al., 2009, PASP, 121, 1395
Li C., White S. D. M., 2009, MNRAS, 398, 2177
Li W., Chornock R., Leaman J., Filippenko A. V., Poznanski D., Wang X., Ganeshalingam M., Mannucci F., 2011, MNRAS, 412, 1473
Lupton R., Gunn J. E., Ivezić Z., Knapp G. R., Kent S., 2001, in Harnden F. R., Jr., Primini F. A., Payne H. E., eds, ASP Conf. Ser. Vol. 238, Astronomical Data Analysis Software and Systems X. Astron. Soc. Pac., San Francisco, p. 269
Madau P., Dickinson M., 2014, ARA&A, 52, 415
Maund J. R., 2017, MNRAS, 469, 2202
McGaugh S. S., Schombert J. M., Lelli F., 2017, ApJ, 851, 22
Miluzio M. et al., 2013, A&A, 554, A127
Moore B., Ghigna S., Governato F., Lake G., Quinn T., Stadel J., Tozzi P., 1999, ApJ, 524, L19
Noeske K. G. et al., 2007, ApJ, 660, L43
Nugent P., Kim A., Perlmutter S., 2002, PASP, 114, 803
Parzen E., 1962, Ann. Math. Stat., 33, 1065
Pasquali A., van den Bosch F. C., Mo H. J., Yang X., Somerville R., 2009, MNRAS, 394, 38
Pearson W. J. et al., 2018, A&A, 615, A146

Perley D. A. et al., 2016, ApJ, 830, 13
Perlmutter S. et al., 1999, ApJ, 517, 565
Pier J. R., Munn J. A., Hindsley R. B., Hennessy G. S., Kent S. M., Lupton R. H., Ivezić, Ž., 2003, AJ, 125, 1559
Postman M. et al., 2012, ApJS, 199, 25
Richardson D., Jenkins Robert L. I., Wright J., Maddox L., 2014, AJ, 147, 118(R14)
Sako M. et al., 2018, PASP, 130, 064002(S18)
Schaye J. et al., 2015, MNRAS, 446, 521
Schlegel D. J., Finkbeiner D. P., Davis M., 1998, ApJ, 500, 525
Skillman E. D., 2005, New Astron. Rev., 49, 453
Smartt S. J., 2009, ARA&A, 47, 63
Smith N., Tombleson R., 2015, MNRAS, 447, 598
Smith J. A. et al., 2002, AJ, 123, 2121
Speagle J. S., Steinhardt C. L., Capak P. L., Silverman J. D., 2014, ApJS, 214, 15
Stinson G. S., Dalcanton J. J., Quinn T., Kaufmann T., Wadsley J., 2007, ApJ, 667, 170
Stoll R., Prieto J. L., Stanek K. Z., Pogge R. W., 2013, ApJ, 773, 12
Stoughton C. et al., 2002, AJ, 123, 485
Strolger L.-G. et al., 2015, ApJ, 813, 93
Taylor M. et al., 2014, ApJ, 792, 135
Taylor E. N. et al., 2011, MNRAS, 418, 1587
Thomas D., Maraston C., Schawinski K., Sarzi M., Silk J., 2010, MNRAS, 404, 1775
Tucker D. L. et al., 2006, Astron. Nachr., 327, 821
Van Dyk S. D., Peng C. Y., King J. Y., Filippenko A. V., Treffers R. R., Li W., Richmond M. W., 2000, PASP, 112, 1532
Wang J., Deng J. S., Wei J. Y., 2010, MNRAS, 405, 2529
Wright A. H. et al., 2017, MNRAS, 470, 283
Wuyts S. et al., 2011, ApJ, 742, 96
Xiao L., Eldridge J. J., 2015, MNRAS, 452, 2597
York D. G. et al., 2000, AJ, 120, 1579
Zheng X. Z., Bell E. F., Papovich C., Wolf C., Meisenheimer K., Rix H.-W., Rieke G. H., Somerville R., 2007, ApJ, 661, L41

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