The physical properties of star-forming galaxies in the low-redshift Universe

J. Brinchmann,[1,2]⋆† S. Charlot,[1,3] S. D. M. White,[1] C. Tremonti,[4,5] G. Kauffmann,[1] T. Heckman[4] and J. Brinkmann[6]

[1]*Max-Planck-Institut für Astrophysik, Karl-Schwarzschild-Str. 1, 85740 Garching bei München, Germany*
[2]*Centro de Astrofísica da Universidade do Porto, Rua das Estrelas - 4150-762 Porto, Portugal*
[3]*Institut d'Astrophysique du CNRS, 98 bis Boulevard Arago, F-75014 Paris, France*
[4]*Department of Physics and Astronomy, Johns Hopkins University, Baltimore, MD 21218, USA*
[5]*Steward Observatory, University of Arizona, Tucson, AZ 85721, USA*
[6]*Apache Point Observatory PO Box 59, Sunspot, NM 88349-0059, USA*

ABSTRACT

We present a comprehensive study of the physical properties of $\sim 10^5$ galaxies with measurable star formation in the Sloan Digital Sky Survey (SDSS). By comparing physical information extracted from the emission lines with continuum properties, we build up a picture of the nature of star-forming galaxies at $z < 0.2$. We develop a method for aperture correction using resolved imaging and show that our method takes out essentially all aperture bias in the star formation rate (SFR) estimates, allowing an accurate estimate of the total SFRs in galaxies. We determine the SFR density to be $1.915^{+0.02}_{-0.01}$ (random)$^{+0.14}_{-0.42}$ (systematic) $h_{70}10^{-2}$ M$_\odot$ yr^{-1} Mpc^{-3} at $z = 0.1$ (for a Kroupa initial mass function) and we study the distribution of star formation as a function of various physical parameters. The majority of the star formation in the low-redshift Universe takes place in moderately massive galaxies (10^{10}–10^{11} M$_\odot$), typically in high surface brightness disc galaxies. Roughly 15 per cent of all star formation takes place in galaxies that show some sign of an active nucleus. About 20 per cent occurs in starburst galaxies. By focusing on the SFR per unit mass we show that the present to past average SFR, the Scalo b-parameter, is almost constant over almost three orders of magnitude in mass, declining only at $M_* > 10^{10}$ M$_\odot$. The volume averaged b parameter is $0.408^{+0.005}_{-0.002}$ (random)$^{+0.029}_{-0.090}$ (systematic)h_{70}^{-1}. We use this value to constrain the star formation history of the Universe. For the concordance cosmology the present-day Universe is forming stars at at least 1/3 of its past average rate. For an exponentially declining cosmic star formation history this corresponds to a time-scale of $7^{+0.7}_{-1.5}$ Gyr. In agreement with other work we find a correlation between b and morphological type, as well as a tight correlation between the 4000-Å break (D4000) and b. We discuss how D4000 can be used to estimate b parameters for high-redshift galaxies.

Key words: surveys – galaxies: abundances – galaxies: evolution – galaxies: fundamental parameters – galaxies: general.

1 INTRODUCTION

One of the most active areas of research during the last decade has been the study of the present and past star formation histories of galaxies. This includes detailed studies of the stellar population of our own Galaxy (e.g. Rocha-Pinto et al. 2000) and of Local Group dwarfs, both individually and as relic evidence of the global star formation history of the Universe (Hernandez, Gilmore & Valls-Gabaud 2000; Hopkins, Irwin & Connolly 2001b). While potentially very powerful, the techniques used in these studies are difficult to extend much beyond the Local Group.

At somewhat larger distances, efforts have gone towards understanding the global properties of more massive galaxies and the dependence of star formation on mass and/or morphological classification (e.g. Kennicutt 1983; Kennicutt, Tamblyn & Congdon 1994; Gavazzi & Scodeggio 1996; Boselli et al. 2001). This has been accompanied by careful mapping of both the star formation and

⋆Present address: Centro de Astrofísica da Universidade do Porto.
†E-mail: jarle@astro.up.pt

© 2004 RAS

the gas distribution in galaxies (e.g. Young et al. 1995; Martin & Kennicutt 2001; Wong & Blitz 2002). Based on these studies, a picture has emerged in which more massive galaxies undergo a larger fraction of their star formation at early times than less massive ones. This has often been said to present a challenge for existing models of galaxy formation.

These studies have traditionally been carried out with relatively small samples of local galaxies. For example, the comparatively large sample of Gavazzi et al. (2002) contains 369 spiral galaxies. With the advent of large fibre-fed surveys such as the 2dF (Colless et al. 2001) and the Sloan Digital Sky Survey (SDSS; York et al. 2000), it is now possible to extend such studies dramatically in size. Large galaxy surveys have most recently been exploited by Hopkins et al. (2003), who carried out a comprehensive study of different star formation rate (SFR) indicators, and by Nakamura et al. (2003) who studied the local Hα luminosity density.

Alongside these careful studies of individual local galaxies, a global picture of the star formation history of the Universe has been assembled through careful studies of star-forming galaxies as a function of look-back time (Cowie 1988; Songaila, Cowie & Lilly 1990; Lilly et al. 1996; Madau et al. 1996; Madau, Pozzetti & Dickinson 1998; Hogg 2001). The main emphasis of this work has been the star formation history of the Universe as a whole. Some workers have also tried to study the evolution of different morphological types (Guzman et al. 1997; Brinchmann et al. 1998). These studies have adopted a wide variety of star formation indicators (Flores et al. 1999; Glazebrook et al. 1999; Afonso, Cram & Mobasher 2000; Hopkins et al. 2001a; Sullivan et al. 2001; Glazebrook et al. 2003). Despite differences between different indicators it is clear that there has been a decline in the rate of star formation since $z \sim 1$, although there is still no consensus on how rapid this decline was. There is also no clear consensus on the star formation history before $z = 2$. Studies in the literature give answers ranging from a clear increase back to $z = 8$ (Lanzetta et al. 2002), through a constant rate of star formation back to $z = 5$ (e.g. Baldry et al. 2002; Giavalisco et al. 2004), to a decline at high z (e.g. Madau et al. 1998).

One of the key unanswered questions is what physical parameter(s) drive changes in the SFR in individual galaxies. Although an increasing body of evidence points to turbulence being a key influence on star formation on small scales (e.g. MacLow et al. 2004, and references therein), the origin of this turbulence and its relevance for star formation on galaxy-sized scales is unclear. It is therefore of major interest to investigate empirically what global quantities correlate strongly with star formation activity. It is logical, however, to expect that the SFR is a function of many variables. To investigate this dependence it is necessary to study the SFR density function in various projections (see also the approach by Lanzetta et al. 2002). This is also becoming a favoured approach to analyse theoretical models (e.g. Springel & Hernquist 2003; Yang, Mo & van den Bosch 2003).

We will pursue this goal here by trying to tie together detailed information on individual galaxies with their contribution to the overall cosmic star formation activity. This is possible both through studying a snapshot in time and by comparing current and past average star formation distributions. On a global scale, this provides a nearly model-independent complement to the studies of the cosmic spectrum by Baldry et al. (2002) and Glazebrook et al. (2003).

This kind of study requires a survey that spans a large range in galaxy properties and with accurate photometric and spectroscopic information. The SDSS (York et al. 2000) is an ideal candidate for this kind of work. Its accurate photometry allows good characterization of the selection function and the structural parameters of

galaxies. The large data base of accurately calibrated spectra allows a much more careful study of the spectral properties of galaxies than has hitherto been possible.

With such unprecedented quality of data it is pertinent to improve modelling over what has generally been done before. One aspect in particular that is generally missing in earlier work is a rigorous treatment of line formation and dust. This is especially important for the SDSS as it spans a large range in luminosity, so the physical parameters of the galaxy can be reasonably expected to vary. In this paper we have therefore built on the work of Charlot et al. (2002) and fit all strong emission lines simultaneously to simultaneously constrain these physical properties.

Our paper is structured as follows. We start in Section 2 by briefly describing the data used. Section 3 gives an overview of the models and the modelling needed for our analysis. This includes a discussion of the relationship between Hα luminosity and SFR. Section 4 discusses in some detail the influence of active galactic nuclei (AGN) activity on our estimates of SFRs. Section 5 discusses the very important question of aperture effects in a fibre survey such as the SDSS and sets out our method for removing these. With these preliminaries out of the way, we discuss the local SFR density in Section 6 and give an inventory of star formation in Section 7. The specific SFR and its variation with mass is discussed in Section 8 and the variation with other physical parameters is covered in Section 9. The paper is concluded in Section 10. We use a cosmology with $H_0 = 70$ km s^{-1} Mpc^{-1}, $\Omega_M = 0.3$ and $\Omega_\Lambda = 0.7$. Where appropriate we define $h = H_0/100$ km s^{-1} Mpc^{-1}. We use the Kroupa (2001) universal initial mass function (IMF). To convert to the popular choice of a Salpeter (1955) IMF between 0.1 and 100 M$_\odot$, one should multiply our SFR estimates by 1.5, the ratio of mass in the two IMFs for the same amount of ionizing radiation. We will also make use of stellar mass estimates below, and to convert these to a Salpeter IMF it is again necessary to multiply the masses by 1.5.

2 THE DATA USED

The SDSS (York et al. 2000; Stoughton et al. 2002) is an ambitious survey to obtain spectroscopic and photometric data across π sr. The survey is conducted using a dedicated 2.5-m telescope at Apache Point Observatory. The photometry is obtained using drift-scanning with a unique CCD camera (Gunn et al. 1998), allowing near-simultaneous photometry in five bands (u, g, r, i, z, Fukugita et al. 1996; Hogg et al. 2001; Smith et al. 2002). The resulting data are reduced in a dedicated photometric pipeline, PHOTO (Lupton et al. 2001) and astrometrically calibrated (Pier et al. 2003). We use the results of PHOTO v5.4. To estimate total magnitudes we use the CMODEL magnitudes recommended for use by the DR2 Paper (Abazajian et al. 2004). This is a weighted combination of the exponential and de Vaucouleurs fits to the light profile provided by PHOTO and has the advantage over Petrosian magnitudes that they have a higher signal-to-noise ratio (cf. the discussion in Baldry et al. 2004). We have also performed all calculations with the Petrosian total magnitudes and found that the results in the paper change if these are used although we get a few more outliers. Fibre magnitudes in g, r, i are measured directly from the observed spectrum and the r-band magnitude is normalized to the fibre magnitude of PHOTO.[1] When estimating fixed-frame magnitudes we follow Blanton et al.

---

[1] For those galaxies where we cannot measure fibre magnitudes from the spectra we adopt the fibre magnitudes from PHOTO. This has no effect on the results in this paper.

(2003a) and *k*-correct our magnitudes to $z = 0.1$ to minimize the errors on the *k*-correction. We will also adopt their notation and refer to these redshifted magnitudes as $^{0.1}g$, $^{0.1}r$ and $^{0.1}i$.

The spectroscopic observations are obtained using two fibre-fed spectrographs on the same telescope, with the fibre placement carried out using an efficient tiling algorithm (Blanton et al. 2003). The spectroscopic data are reduced in two independent pipelines SPECTRO1D (SubbaRao et al. in preparation) and SPECBS (Schlegel et al. in preparation) – we use the SPECBS redshifts, but the differences between the two pipelines are entirely negligible.

The data set used in the present study is based on the main SDSS galaxy sample, described by Strauss et al. (2002). We use a subset of Blanton et al.'s (2003b) Sample 10, consisting of 149 660 galaxies with spectroscopic observations, $14.5 < r < 17.77$ and $0.005 < z < 0.22$. This encompasses slightly more galaxies than the SDSS Data Release 1 (Abazajian et al. 2003). The overall continuum properties of this sample are discussed in some detail by Kauffmann et al. (2003a,b) and the sample selection and spectrum pipeline analysis are discussed by Tremonti et al. (2004). The spectrophotometric calibration we use will also be used in Data Release 2 (Abazajian et al. 2004). We make extensive use of stellar masses for the galaxies taken from Kauffmann et al. (2003a). To be consistent with the latter we will throughout use the narrow definition of the 4000-Å break from Balogh et al. (1998) and denote this as D4000.

As detailed in Tremonti et al. (2004), we have chosen to reanalyse the 1D spectra using our own optimized pipeline. This allows us to take more care in the extraction of emission line fluxes than is done in the general purpose SDSS pipeline. The key differences for the work presented here are as follows: we remove the smear exposures so that the spectra are uniformly exposed within a 3-arcsec fibre[2] and we use the latest high-resolution population synthesis models by Bruzual & Charlot (2003) to fit the continuum using a non-negative linear-least-squares routine (Lawson & Hanson 1974). This provides an excellent fit to the continuum and lets us perform continuum subtraction with unprecedented accuracy. The fitting procedure automatically accounts for weak metal absorption under the forbidden lines and for Balmer absorption (see e.g. Bruzual & Charlot 2003, for examples of the quality of fit).

The ability to extract very weak line emission turns out to be very important when studying trends with mass (and hence increasing continuum contribution) and when estimating dust attenuation. We will see below that it is also important to identify possible AGN contamination of the emission lines.

A crucial aspect of the SDSS for studies of the nearby galaxy population is that it has a well-defined and well-studied selection function (Strauss et al. 2002) and covers a large range in absolute magnitude. This means that we can use the galaxy sample not only to study individual galaxies, but also to construct distribution functions for volume-limited samples with high accuracy. Note that one can usually extract *trends* from a judiciously chosen, but statistically poorly characterized, sample but the *distribution* of a parameter requires a statistically well-defined sample. We will therefore focus our attention on distribution functions throughout our analysis.

We will limit our study to galaxies with $0.005 < z < 0.22$. The lower limit reflects our wish to include galaxies of the lowest possible luminosity. At the same time, we wish to avoid redshifts where deviations from Hubble flow are substantial and the accessible volume is very small. At the low-redshift limit, a galaxy with the mass-to-light ratio appropriate for a 13.6-Gyr old stellar population and
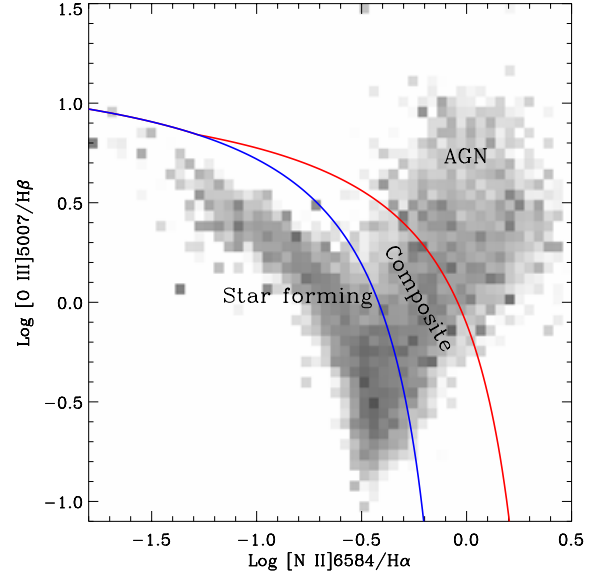

Figure 1. The distribution of the galaxies in our sample in the BPT line-ratio diagram. The two lines shows the division of our sample into the three subsamples discussed in the text. An unweighted version of this diagram can be seen in Fig. 9. The galaxies plotted here have S/N > 3 in all four lines.

$r = 17.77$ would have a stellar mass just below 10^8 solar masses. We should therefore be complete to this mass limit.[3] Our sample can thus be used to reconstruct the properties of a volume limited sample of galaxies with $M_* > 10^8 \, M_\odot$. As we will see below, we also include the vast majority of all star formation in the nearby Universe.

2.1 Sample definitions

We will attempt to treat the entire sample in a uniform way in our SFR calculations. It will, however, prove necessary in the following discussion to define several subsamples of objects based on their emission line properties. These are defined on the basis of the Baldwin, Phillips & Terlevich (1981, hereafter BPT) diagram shown in Fig. 1. The diagram has been divided into three regions which we discuss further below.

Although a classification can be made on the basis of Fig. 1 regardless of the signal-to-noise ratio (S/N) in the lines, we find that requiring S/N > 3 for all lines is useful. Below this limit, a rapidly increasing fraction of galaxies have negative measured line fluxes and the non-symmetric distribution of galaxies along the *y*-axis of the BPT diagram leads to classification biases. We have adopted the following subsamples (cf. Table 1).

All. The set of all galaxies in the sample regardless of the S/N of their emission lines.

SF. The star-forming galaxies. These are the galaxies with S/N > 3 in all four BPT lines that lie below our most conservative AGN rejection criterion. As discussed in Section 4, they are expected to have a very low (<1 per cent) contribution to Hα from AGN.

C. The objects with S/N > 3 in all four BPT lines that are between the upper and lower lines in Fig. 1. We refer to these as composite galaxies. Up to 40 per cent of their Hα luminosity might come from

---

[2] This is the default procedure in DR2.

[3] Dust attenuation should be of minor importance in this estimate given the observed low attenuation in old stellar systems.

Table 1. Basic data about the subsamples discussed in the text.

Subsample	Number	Per cent	Per cent of mass density
All	146994	100.0	100.0
SF	39141	26.6	20.6
C	14372	9.8	11.3
AGN	8836	6.0	7.1
Low S/N LINER	11752	8.0	9.5
Low S/N SF	29115	19.8	20.0
UnClass	43778	29.8	31.6

an AGN. The lower line is taken from Kauffmann et al. (2003c) and is a shifted version of the upper line. It does lie very close to an empricial determination of the SF class (see Section 4 below) but we have kept the Kauffmann et al. line to be consistent with that work. No results below are affected by this choice.

AGN. The AGN population consists of the galaxies above the upper line in Fig. 1. This line corresponds to the theoretical upper limit for pure starburst models so that a substantial AGN contribution to the line fluxes is required to move a galaxy above this line. The line has been taken from equation (5) in Kewley et al. (2001b), but our models have an identical upper limit.

Low S/N AGNs. A minimum classification for AGN galaxies is that they have [N II]6584/Hα > 0.6 (and S/N > 3 in both lines) (e.g. Kauffmann et al. 2003c). It is therefore possible to classify these even if [O III]5007 and/or Hβ have too low S/N to be useful (cf. Fig. 3, later). A similar approach is taken by Miller et al. (2003). In general, we will include these galaxies together with the AGN class.

Low S/N SF. After we have separated out the AGN, the composites and the low S/N AGNs, we have thrown out most galaxies with a possible AGN contribution to their spectra. The remaining galaxies with S/N > 2 in Hα are considered low S/N star formers. We can still estimate the SFR of these galaxies from their line strengths, even though we cannot use the full modelling apparatus described in the following section.

Unclassifiable. Those remaining galaxies that are impossible to classify using the BPT diagram. This class is mostly made up of galaxies with no or very weak emission lines.

Note that this is merely a 'nuclear' classification – it does not make any statement about the properties of the parts of the galaxy outside the region sampled by the fibre. In particular, we expect the Unclassifiable category to include a substantial number of galaxies where we only sample the central bulge – for such galaxies there may be considerable amounts of star formation outside the fibre. We will return to this point when discussing aperture corrections below.

Fig. 2 shows the distribution of log S/N in each line considered for the galaxies in our sample. All galaxies falling outside the plotted range have been included in the three bins at the extremes of the plot. The thick solid line shows the overall S/N distribution and the distributions for each of the different classes defined above are shown with different line styles as indicated by the legend in the first panel. For clarity, the low S/N AGNs have been grouped with the AGN.

These distributions are of considerable interest to understand possible biases in our classification scheme. The most notable results from these panels can be summarized as follows.

(i) The SF & C classes are limited by the strength of the [O III]5007 line (see the sharp edge for the thick dashed line at

S/N = 3 in the [O III]5007-panel). This is because the low S/N SF&C galaxies tend to be in the bottom right of the BPT diagram where [O III]5007<Hβ.

(ii) Because the AGN class has [O III]5007 > Hβ, the main limitation for the AGN class turns out to be the S/N in Hβ. This is why using a AGN definition that only depends on Hα and [N II]6584 can help us identify a larger number of AGN.

(iii) There is clear evidence that there is a considerable number of star-forming galaxies that are thrown out because they have low S/N in [O III]5007. The inclusion of the low S/N SF class is therefore very beneficial. As Table 1 shows this almost doubles the number of galaxies for which we can use Hα to determine the SFR.

Our sample covers a large range in galaxy luminosity and the distributions shown in Fig. 2 suppress this dimension. A complementary view of these data is shown in Fig. 3. This shows the trend in the median S/N of the different lines as a function of the stellar mass of the galaxies. We have suppressed the results for the [O II]3727 line as it falls out of the sample at low redshift and also lies in a region of the spectrum that has rather different noise characteristics to that where the other lines reside. It shows a similar trend to to [O III]5007.

The most important point in this figure is that the fraction of galaxies for which we can assign a nuclear classification is a declining function of mass. With any fixed S/N cut, the unclassifiable category will be biased towards more massive galaxies.

The decline in S/N in the lines with mass is seen at $M_* > 10^{10}$ M$_\odot$. For lower masses, we note that the S/N in Hα is constant over a large range in mass. We will return to this point later. [O III]5007 changes from being the strongest line at low mass to being the weakest at high mass. This is caused by the decreasing electron temperature. At high mass, most of the cooling takes place in mid-infrared (MIR) fine-structure lines, depressing the [O III]5007 flux strongly. At the very highest masses the trend reverses because of the higher fraction of AGN in the most massive galaxies (cf. Kauffmann et al. 2003c). The [N II]6584 and [S II]6716 lines increase in strength as a function of mass because of the increasing metallicity (see the discussion in CL01). At high masses, the [N II]6584 line is similar in strength to the Hα line, consistent with an increasing fraction of LINERs at high masses (cf. Kauffmann et al. 2003c).

3 MODELLING

Our aim is to use the data presented in the previous section to constrain the physical properties of the galaxies in our sample. The main emphasis in this paper is on the SFR. In the following we will sometimes distinguish between SFRs derived directly from the emission lines as discussed in this section and those derived indirectly from the 4000-Å break as discussed in Section 4.1. The former we will refer to as SFR$_e$ and the latter as SFR$_d$. In the majority of the paper we will use SFR$_e$ for some galaxies and SFR$_d$ for others and in that case we will use SFR for simplicity, the details will be given towards the end of Section 5.

We will build on the methodology of Charlot et al. (2002, hereafter C02) and model the emission lines in our galaxies with the Charlot & Longhetti (2001, CL01) models. These combine galaxy evolution models from Bruzual & Charlot (1993, version BC02) with emission line modelling from CLOUDY (version C90.04; Ferland 1996; see CL01 for details). These are similar in spirit to other models recently presented in the literature (Kewley et al. 2001a; Moy, Rocca-Volmerange & Fioc 2001; Zackrisson et al. 2001;
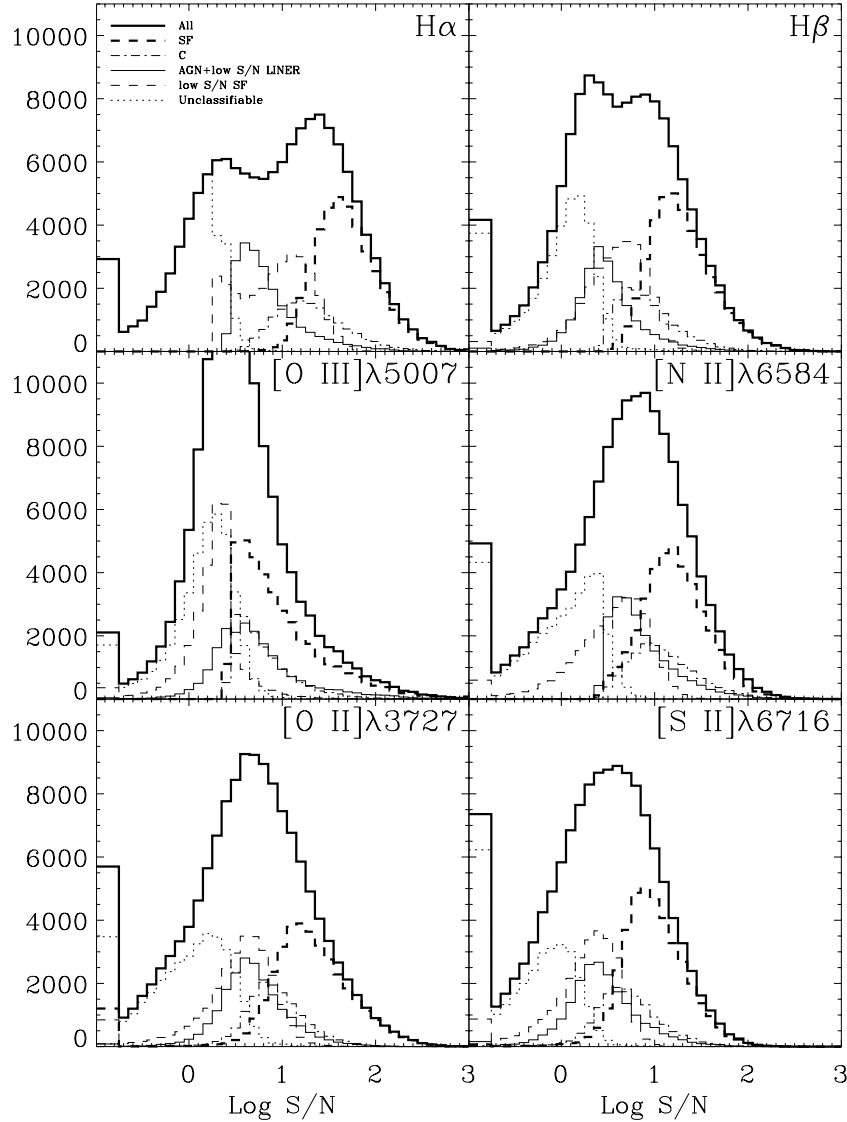
Figure 2. The distribution of signal-to-noise ratio for the six lines that we include in our fits. The thick solid line shows the overall distribution and the distributions for the different subcategories are shown with different line styles as indicated in the legend. The objects with S/N in the emission line outside the displayed range are included in the extreme three bins of the distributions.
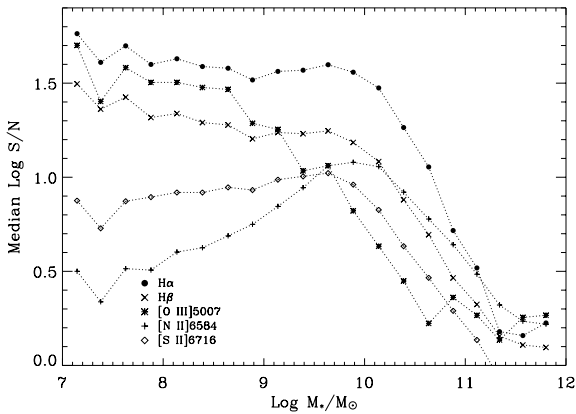

Figure 3. The median signal-to-noise ratio for five of our lines that can be observed over the entire redshift range spanned by our sample. Notice in particular how [O III]5007 goes from being very strong at low masses to being the weakest line at high masses.

Bressan, Silva & Granato 2002; Panuzzo et al. 2003), but they contain a physically motivated dust model and are optimized for integrated spectra as opposed to individual H II regions. The modelling is discussed in more detail in Charlot et al. (in preparation, hereafter Paper I, see also C02); here we will only discuss those aspects important for our current undertaking.

We have generated a model grid using the following *effective* (i.e. galaxy-wide) parameters: metallicity Z, the ionization parameter at the edge of the Strömgren sphere U, the total dust attenuation in the V-band τ_V, and the dust-to-metal ratio of the ionized gas ξ. The parameter ranges and grid sizes adopted are given in Table 2. As we will only study relative line fluxes in this paper, we are not sensitive to stellar age, star formation history and the relative attenuation by dust in the stellar birthclouds (i.e. giant molecular clouds) and the ambient interstellar medium (ISM). In total our grid contains $\sim 2 \times 10^5$ models.

Given the major uncertainty in traditional star formation estimates due to dust attenuation (Sullivan et al. 2001; Hopkins et al. 2001a; Bell & Kennicutt 2001; Calzetti 1997), we should briefly summarize

Table 2. The model grid calculated for the present work. This is calculated for a constant star formation history at $t = 10^8$ yr – see text for details.

Parameter	Description	Range
Z	The metallicity	$-1 < \log Z/Z_\odot < 0.6$ in 24 steps
U	The ionization parameter	$-4.0 < \log U < -2.0$ in 33 steps
τ_V	The total dust attenuation	$0.01 < \tau_V < 4.0$ in 24 steps
ξ	The dust-to-metal ratio	$0.1 < \xi < 0.5$ in 9 steps

our dust treatment which follows that of Charlot & Fall (2000, hereafter CF00). The CF00 model provides a consistent model for ultraviolet (UV) to far-infrared (FIR) emission. It does this by incorporating a multicomponent dust model where the birth clouds of young stars have a finite lifetime. This ensures a consistent picture for the attenuation of continuum and line emission photons. In addition, the model deals with angle-averaged, effective quantities, which depend on the combined optical properties and spatial distribution of the dust.

Each model in our grid has been calculated with a particular dust attenuation and these attenuated line ratios are compared directly to the observed spectrum. This means that all lines contribute to the constraint on the dust attenuation. To first approximation, however, our dust corrections are based on the Hα/Hβ ratio. Note that we do *not* interpret the derived dust attenuations as coming from a simple foreground screen, a procedure which has been rightfully criticized as simplistic (e.g. Witt & Gordon 2000; Bell & Kennicutt 2001; Serjeant, Gruppioni & Oliver 2002).

This model grid is somewhat larger than that used by C02, but the main difference from C02 is that we adopt a Bayesian approach to calculate the likelihood of each model given the data. This is similar to the approach used by Kauffmann et al. (2003a) in their calculation of stellar masses. If, for notational simplicity, we enumerate the models with a single index, j, we can write the likelihood of a model given the data, $P(M_j| \{L_i\})$, as

$$\ln P(M_j|\{L_i\}) = -\frac{1}{2} \sum_{i=0}^{N_{\text{lines}}} \left(\frac{L_i - AM_{i,j}}{\sigma_i} \right)^2, \qquad (1)$$

where the sum is over all lines observed for that galaxy, L_i is the luminosity of line i, A is an overall scaling factor and M is the model grid – in practice this is a 4D array. The prior adopted is given in Table 2 and is essentially a maximum ignorance, flat prior for all parameters. The choice of prior has very little effect for the high S/N galaxies as long as we span the range of physical parameters in the sample. However, as we will explain below, it can make a difference for low S/N galaxies.

We fit in linear flux units rather than to flux ratios, because the approximation of Gaussianity is more appropriate there. (By comparing the observed ratio of the [O III] lines at 4959 and 5007 Å with the theoretical value,[4] we find that the errors on our flux measurements become non-Gaussian below S/N $\approx$ 2. As we limit our star-forming sample to S/N > 3, this is not a major concern).[5]

The important advantage of our approach is that it generates the full likelihood distribution of a given parameter and from this we

[4] The likelihood distribution of the ratio of two Gaussians is, of course, not Gaussian and we compare with the appropriate distribution for the ratio of two Gaussians.
[5] Few galaxies have S/N <3 in [O II]3727 or [S II]6716 when the S/N > 3 in the other lines (1–2 per cent for [O II]3727 and $\sim$4 per cent for [S II]6716). These do not affect our fits as the other lines dominate the fits for these objects.
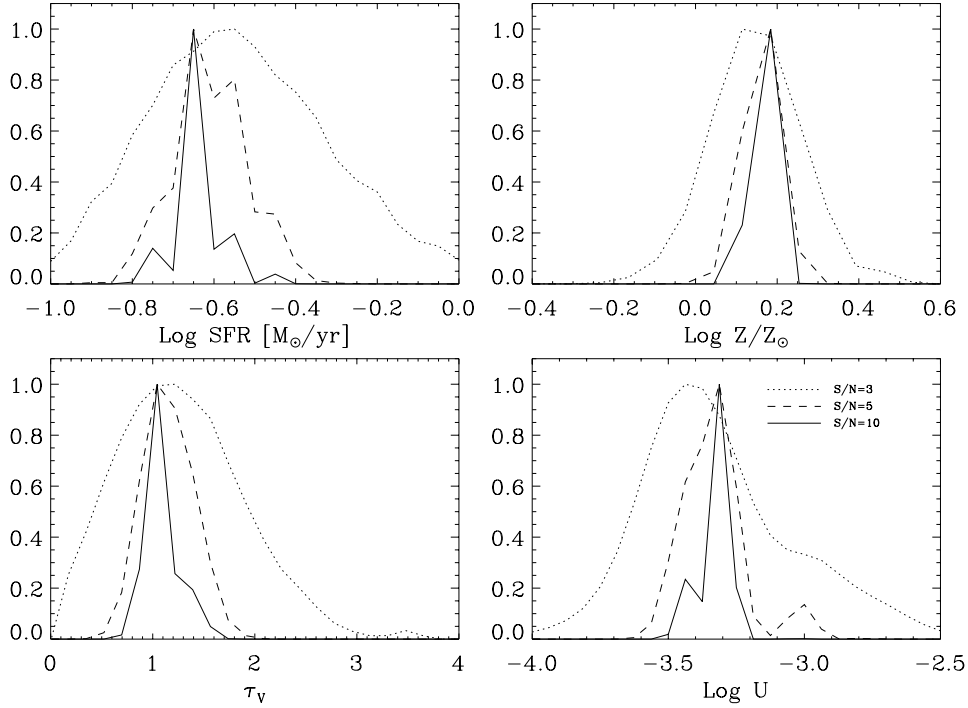
can rigorously define confidence intervals for our estimates. For high S/N objects, these likelihood distributions are symmetric around a best-fitting value. In this situation it is possible to summarize the distribution using the best-fitting value and the spread around this value. However, for lower S/N galaxies, the distributions may be double-peaked or non-symmetric and it is necessary to keep the entire likelihood distribution in further calculations. The main disadvantage of working with full distributions is that they are somewhat cumbersome to manipulate. The basic theory is covered in introductory statistics books (e.g. Rice 1995), but for the benefit of the reader we have collected the results most important for the present paper in Appendix A.

Fig. 4 shows the result of taking a high S/N spectrum and decreasing the S/N of the emission lines. Each likelihood distribution is the result of 100 realizations (assuming Gaussian noise on the line flux). The figure shows S/N $=2$, 5 and 15, where the S/N is that of the weakest line of those contributing to the BPT diagram; for this galaxy this is Hβ.

The distributions broaden as expected when the S/N decreases. There is also a slight shift in the estimates with decreasing S/N – in particular the estimate of the SFR$_e$ goes up by 0.1 dex in the mode, somewhat more in average. Although the magnitude and sign of this shift depends on the particulars of the line fluxes of the galaxy, it is an indication that our apparatus does not function well at low S/N. For this reason, and because we cannot reliably classify galaxies at low S/N, we only use the results of the model fits for the SF class.
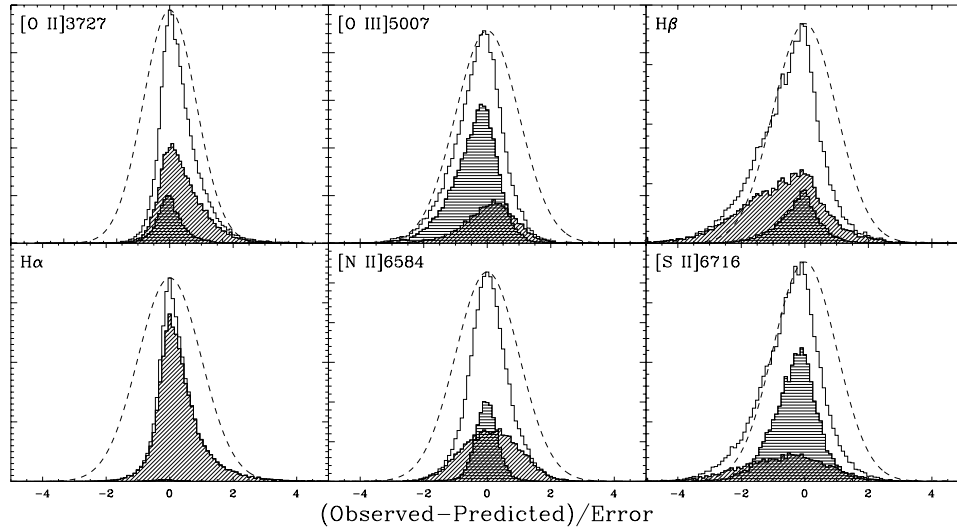
Fig. 5 shows the distribution of the difference between the observed value of the line fluxes and that of the the best-fitting model, divided by the error for all galaxies in the SF class. The agreement is excellent except perhaps that the models slightly overpredict the Hβ flux for the highest S/N objects. This could be caused by uncertainties in the continuum subtraction, but it has negligible effect on our SFR$_e$ estimates.

This modelling approach is necessary to get accurate SFR$_e$ estimates because the SDSS spans a range in galaxy properties such as Hubble type, mass and emission-line characteristics. Previous studies of star formation activity (see Kennicutt 1998 for a review) using optical spectra have usually assumed a fixed conversion factor between Hα luminosity and SFR$_e$. Dust attenuation is estimated from Hα/Hβ assuming a fixed unattenuated Case B ratio. The fixed conversion factor appears to be a reasonable approximation when stellar absorption in the Balmer lines is taken into account (e.g. Charlot et al. 2002). However, as is well known, and shown quantitatively by CL01, it is not correct in detail. The approximation neglects the effect of metallicity and ionization state. It also neglects the diffuse emission in galaxies. More careful modelling is required in order to avoid creating biases in the SFR$_e$ estimates as function of metallicity or stellar mass.

However, in many situations, in particular at high redshift, detailed modelling in the way we have done is very hard to carry out. It is therefore of considerable interest to look at the relationship between the Hα luminosity and SFR$_e$ found in our study.

3.1 The relationship between Hα luminosity and SFR$_e$

We will follow CL01 and denote the ratio of observed Hα luminosity to SFR$_e$ as, $\eta_{\text{H}\alpha}$:

$$\eta_{\text{H}\alpha} = L_{\text{H}\alpha}/\text{SFR}_e. \qquad (2)$$

In this form the Hα luminosity is assumed to be uncorrected for dust attenuation. We now wish to study how this parameter varies with the physical properties of the galaxy. It will prove convenient to split

Figure 4. The likelihood distributions derived for model parameters and the SFR_e as a function of the S/N of the weakest line (here Hβ) for one high S/N galaxy. Each distribution is the average of 100 realizations with that S/N. The top left shows the likelihood distribution of log SFR_e the top right the logarithm of the gas metallicity in units of solar. The bottom left-hand panel shows the dust attenuation in the emission lines at V and finally the bottom right shows the logarithm of the ionization parameter at the edge of the Strömgren sphere. The galaxy is SDSS J234342.75−001255.9 at $z = 0.026$.

Figure 5. The distribution of the difference between the observed value of the line fluxes and that of the the best-fit model, divided by the error for all galaxies in the SF class. The smooth dashed curve shows a unit Gaussian for reference. The horizontally shaded histogram shows the results for the galaxies that have S/N in the line between 3 and 10. The obliquely shaded histogram is for galaxies with S/N > 20 in the line.

this into the conversion factor between *unattenuated* Hα luminosity and SFR_e which we will refer to as $\eta_{H\alpha}^0$ and the dust attenuation at Hα in mag, $A_{H\alpha}$, which we will take as

$$A_{H\alpha} \approx 0.96\tau_V, \tag{3}$$

This assumes an attenuation law of the form $\tau(\lambda) \propto \lambda^{-0.7}$, which is a good approximation to the CF00 model over the wavelength range we consider.

Fig. 6 shows the likelihood distributions for $A_{H\alpha}$ in five bins in stellar mass. These have been constructed by co-adding the individ-

ual likelihood distributions for $A_{H\alpha}$ for all SF galaxies in the relevant mass range. We note that ignoring the metallicity-dependence of the Case B Hα/Hβ ratio would lead to an overestimate of the dust attenuation by up to ~ 0.5 mag for the most metal rich galaxies. The most striking result from Fig. 6 is the clear increase in dust content at high stellar masses. Tremonti et al. (2004) has shown that there is a strong correlation between metallicity and mass for star-forming galaxies. One would therefore expect that the main physical reason for the correlation between dust attenuation and stellar mass is an increase in the metallicity of the emitting gas. It is also interesting
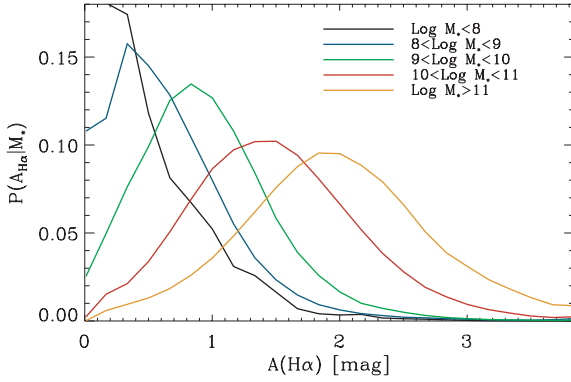
Figure 6. The likelihood distributions of the dust attenuation as a function of stellar mass. These are derived from the model fits to the SF class.
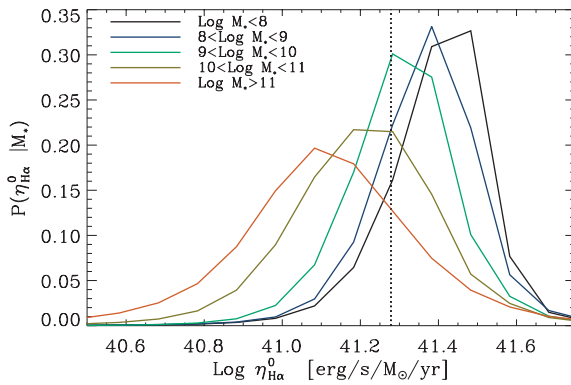

Figure 7. The likelihood distribution of the conversion factor between Hα luminosity and SFR$_e$, $\eta_{H\alpha}^0$. These were again derived from the model fits to the SF class.

that the width of the $A_{H\alpha}$ distribution increases strongly as one goes to more massive galaxies. Some of this is due to the more massive galaxies having lower S/N emission lines and hence more extended likelihood distributions. However, note that we require S/N > 3 for all four lines in the BPT diagram, so it is unlikely that this is the main cause.

Fig. 7 shows the corresponding likelihood distributions for $\eta_{H\alpha}^0$. We have compared this to the conversion factor advocated by Kennicutt (1998) ($\eta_{H\alpha}^0$ (Kennicutt) $= 10^{41.28}$ erg s^{-1} M$_\odot$ yr^{-1} for our adopted IMF). This value is indicated as the vertical dotted line in Fig. 7. Clearly the Kennicutt (1998) conversion factor is a very good *typical* value and compares well with the median value for our sample log $\eta_{H\alpha}^0 = 41.27$. However, the peak values of log $\eta_{H\alpha}^0$ do vary by nearly 0.4 dex, with the most massive/most metal rich galaxies producing less Hα luminosity for the same SFR$_e$ than low mass/metal poor galaxies. This correlation is also likely to be driven in main by the changes in gas-phase metallicity with mass, but as discussed by CL01 and Charlot et al. (2002) a changing escape fraction for ionizing photons is also likely to contribute to these changes in $\eta_{H\alpha}^0$.

3.2 A comparison to other estimators

These trends with mass suggest that any simple recipes for conversion of Hα luminosity into SFR$_e$ are likely to show some biases with stellar mass. To quantify these possible biases, it is useful to compare our SFR$_e$ estimates with those obtained from previous methods to convert observed Hα flux to SFRs (see also Charlot et al. 2002).

We noted above that the Kennicutt (1998) conversion factor is a good typical value so we will use this in what follows, i.e.

$$\text{SFR}_e = L_{H\alpha}^0/10^{41.28}, \tag{4}$$

where $L_{H\alpha}^0$ is the estimated emission luminosity of Hα. Following the discussion above we will also fix the Case B Hα/Hβ ratio to be 2.86 (appropriate for an electron density of $n = 100$ cm^{-3} and electron temperature $T_e = 10^4$ K; Osterbrock 1989), where relevant below. We have also adopted the Seaton (1979) attenuation law in the following. These choices have been made to ensure that we will use a method to estimate SFRs which is very similar to those commonly used in the literature (e.g. Sullivan et al. 2000). The residual difference between methods therefore lie in the estimates of dust attenuation and stellar absorption under the Balmer lines. This leads to the following list of methods of SFR$_e$ estimation.

(1) Assume that the observed flux can be converted to a luminosity without any correction for dust or continuum absorption. This underestimates the SFR$_e$, but might be reasonable for strongly star-forming low-metallicity galaxies.

(2) Correct for dust attenuation by assuming a default attenuation of $A_V = 1$ mag in the emission line and correct for absorption at Hα by assuming a typical stellar absorption, EW$_{abs}$ Hα, of 2 Å. If the assumptions are correct in the mean for the sample this is likely to give a good estimate of the SFR$_e$ values, but the scatter around the mean value might be large depending on the sample.

(3) Improve on the dust corrections by using Hα/Hβ (or some other Balmer line combination) to determine the dust attenuation using a fixed dust-free Case B reference value. We can make two assumptions about the absorption underlying Hβ, as follows.

(a) We can assume that EW$_{abs}$ Hα = EW$_{abs}$ Hβ, this is the most common assumption in the literature.
(b) EW$_{abs}$ Hα = 0.6 EW$_{abs}$ Hβ. This is a better assumption for the BC03 models and presumably for our galaxies as well.

(4) Finally, with good spectra such as the SDSS it is possible to perform a careful estimate of the absorption at Hα and Hβ, which is what we have done for the present paper. These fluxes can be used to determine the dust attenuation as above and thence the SFR$_e$.

We compare these four estimates of SFR$_e$ values with our best estimates in Fig. 8. To produce this figure we took all SF galaxies with S/N in Hβ > 10 (to ensure that dust attenuations from Hα/Hβ are reliable) and compared each of the methods 1–4 above to the mode of our likelihood distribution for the SFR$_e$ values. The figure shows the median of the logarithm of the ratio of the two SFR$_e$ estimates in bins of 0.2 dex in stellar mass.

It is immediately clear that methods 1, 3(b) and 4 all agree well with our estimates for low-mass galaxies. This is expected because the equivalent width of stellar absorption is very small compared to that of the emission line (the median ratio is 2–6 per cent at log $M_* < 10$) and the dust attenuation is very low in these low metallicity galaxies. However, it is clear that the assumption (method 3a) of equal absorption under Hα and Hβ leads to an overestimate of the dust attenuation and hence an overestimate of the SFR$_e$. Method 2 does *not* perform well at low masses either, but this is again expected, as an average dust attenuation of 1 mag at V is too high for these galaxies so again we overestimate the SFR$_e$.

As expected, method 1 starts to diverge from the other estimators when going to more massive and dusty systems (cf. Fig. 6). Likewise, method 3(a) gives systematically discrepant results as explain above. Method 2 is similar to method 1 but agrees with the more

sophisticated estimators in a different mass range since the underlying assumptions are more appropriate for more massive systems.

Finally the remaining methods, 3(b) and 4, agree fairly well with the results of the full fits except at the highest masses. This is because the assumption of a fixed η_{H}^0 and Case B ratio for all masses is flawed as we noted in the discussion of Fig. 7. However, it is interesting to note that the changes in the Case B ratio and in $\eta_{\mathrm{H}\alpha}^0$ with mass nearly cancel when one estimates the SFR_e over a large range in mass. This happy coincidence comes about because the standard Case B ratio will tend to overestimate the unattenuated $H\alpha$ flux, whereas ignoring the variation on $\eta_{\mathrm{H}\alpha}^0$ will require a larger $H\alpha$ flux for the same SFR_e. Thus it is a fairly good approximation to use a fixed Case B ratio and $\eta_{\mathrm{H}\alpha}^0$ value, but one should never use only one and then apply the relations derived here. Also, it is necessary to include the uncertainty in the Case B ratio when quoting uncertainties, this uncertainty can only be accurately assessed using the full model fits.

At the very highest masses it is however worth noting that ignoring these factors might lead to up to a factor of 3 systematic underestimate of the SFR_e values. It is therefore important to adjust the conversion factor and Case B ratio to that expected for the sample galaxies in which case the simple methods will work well.

We should also note that similar trends are seen with the emission equivalent width of $H\alpha$, with the result that methods 2, 3(b) and 4 all give SFR_e estimates in good agreement with the full model fits when the $H\alpha$ emission line has an equivalent width $\gtrsim 20$ Å.

As remarked above, there are a large number of galaxies with significant $H\alpha$ which cannot be classified in the BPT diagram, the low S/N SF class. We would nevertheless still like to use the $H\alpha$ line to constrain the SFR_e of these galaxies. We do this by adopting the average likelihood distribution for the set of high S/N SF galaxies with masses that lie within a factor of 3 of the mass of the object in question. The range is increased if it contains less than 50 galaxies.

3.3 Additional uncertainties

There are several residual uncertainties in our determination of SFR_e that warrant discussion. First, we have considered only one IMF, the Kroupa (2001) universal IMF. The emission lines we consider here depend on the massive end of the IMF. As long as this does not vary dramatically, the effect of changing the IMF is mostly to change the overall normalization of our SFR_e values. To correct from our choice of IMF to the Salpeter (1955) IMF between 0.1 and 100 $M_\odot$, one should multiply our SFR_e estimates by 1.5. This is the ratio of mass in the two IMFs for the same amount of ionizing radiation.

Secondly, our model assumes that essentially no ionizing radiation escapes the galaxies (see CL01). This assumption is necessary in order to relate $H\alpha$ luminosity to SFR_e and seems reasonable for our sample of relatively massive galaxies. Even local starburst galaxies have escape fractions <10 per cent (Leitherer et al. 1995; Heckman et al. 2001, see also CL01).

Finally, our model uses angle-averaged, 'effective' parameter values, whereas any particular galaxy is only seen from one angle. For a given galaxy, there may be a mismatch between the modelling and the observations. We have verified that no significant systematic is present with respect to the inclination of the galaxies, but a residual scatter is likely to remain. This can not be quantified without rigorous modelling of the spatially resolved properties of galaxies.

As explained in Paper I, additional uncertainties come from the mix of stellar populations within the fibre, from uncertainties related to interpolation of our model grid (1–2 per cent), and from uncertainties in stellar tracks and theoretical population modelling (~ 2 per cent). We include these uncertainties below by adding a 4 per cent uncertainty in quadrature.

The largest source of uncertainty in the estimates of SFR_e values within in the fibre, however, is that due to the possible contamination of our spectra by other sources of ionizing radiation. We now turn to a discussion of these.

4 OTHER SOURCES OF EMISSION LINE FLUX

As with any other non-resolved spectroscopic survey of galaxies, the emission line fluxes of the galaxies in our sample will contain a component due to sources not directly connected to star formation. These include planetary nebulae (PNe), supernova remnants (SNRs), the diffuse ionized gas (DIG) and active galactic nuclei (AGN). The contributions of SNR, PNe and DIG are discussed in Appendix B, where we show that they are of minor importance for our work. Here we will quantify the possible level of AGN contribution in each of the classes defined in Section 2.1. We will show that, for the Composite and AGN classes, blind application of the modelling machinery described in the previous section is likely to lead to incorrect results. Note that a comprehensive study of the AGN activity in our sample is presented in Kauffmann et al. (2003c).

There are at least three methods commonly employed to deal with the presence of AGN in a galaxy survey, as follows.

(i) Remove galaxies with AGNs by cross-correlating the sample with published AGN catalogues (e.g. Condon, Cotton & Broderick 2002; Serjeant et al. 2002).

(ii) Identify galaxies with non-stellar ionizing spectra using a diagnostic diagram, typically the so-called BPT diagram introduced by Baldwin et al. (1981). This is possible so long as one can detect the lines required for classification (see Section 2.1). At $z \gtrsim 0.5$, where $H\alpha$ is redshifted out of the optical, this method is less accurate (see however Rola, Terlevich & Terlevich 1997; Tresse et al. 1996).

(iii) Finally AGN can be subtracted in a statistical manner. This is necessary when no other methods are applicable (e.g. Tresse & Maddox 1998). The method is, of course, sensitive to evolution in AGN activity with redshift and it is therefore necessary to include the related uncertainties in the error budget for the derived SFRs (e.g. Tresse et al. 2002; Flores et al. 1999).

All of these methods tend to classify galaxies as either AGN or starbursts, but as shown by Kauffmann et al. (2003c), many AGN have young stellar populations and ongoing star formation.

Our analysis will be based on the BPT diagram. The left-hand panel of Fig. 9 shows the BPT diagram as a 2D histogram for galaxies in our sample with S/N > 3 in all lines. The shading is based on the square root of the number of galaxies in each bin and the grid spacing is 0.05 along the x- and y-axes. The lines delineate our AGN, C and SF classes. The cross shows the location of an 'average' AGN which is constructed by taking the mean of the luminosities of all AGN galaxies with sigma-clipping of outliers using $\sigma = 3.5$. We use this AGN below, the results are robust to the exact average AGN used.

Kewley et al. (2001b) advocate the use of several diagnostic diagrams to rigorously isolate AGNs. However, diagnostic diagrams involving [S II]6716/$H\alpha$ or [O I]6300/$H\alpha$ are much less effective at assessing the degree to which AGN may contaminate our estimated SFR_e. This is because a change in AGN fraction moves a galaxy almost parallel to the SF locus in these diagrams. In the BPT diagram of Fig. 9, a change in AGN fraction moves the galaxy perpendicular to this locus.

The upper line in Fig. 9 shows the uppermost region that can *in principle* be reached by systems ionized by normal stellar populations. However, real SF galaxies (and ionization-bounded H II
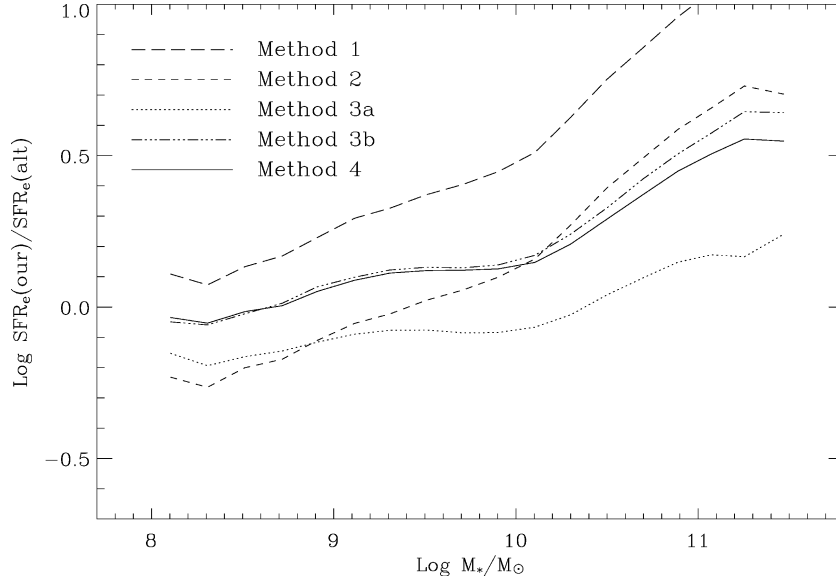
Figure 8. Our estimates compared with the simpler SFR_e estimates discussed in the text as a function of stellar mass. We have used all SF galaxies with S/N Hβ >10. The figure shows the logarithm of the ratio of our estimate of the SFR_e to that obtained by the alternative method, so a value of zero corresponds to equality. See text for details.
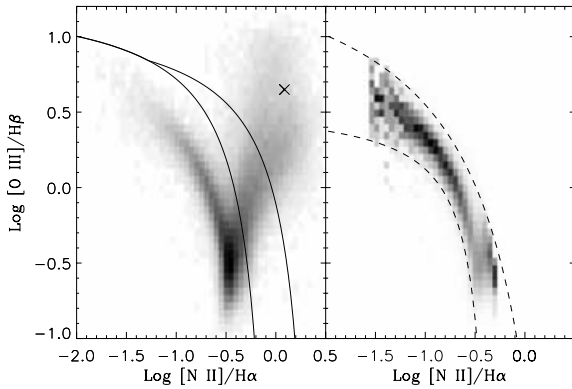
Figure 9. Left: the standard BPT excitation diagram for those galaxies in our sample with S/N > 3 in all lines used here (70189 in total). The grey-scale shows the square root of the number of galaxies in each bin of 0.05 in the *x*- and *y*-axes. The cross shows the location of the average AGN used to assess the AGN contribution to the Hα flux. The lines are the same as those in Fig. 1. Right: the same plot after removal of the composites and AGN (leaving 42 262 galaxies; cf. Table 1) and after normalizing the distribution in each bin along the *x*-axis to 1. The thick dashed lines show fits to intervals containing 95 per cent of the galaxies as described in the text.
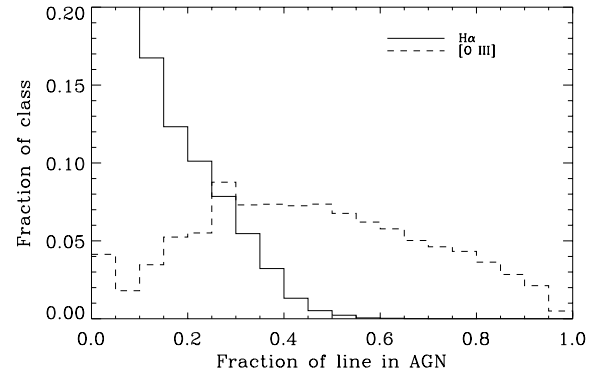

Figure 10. The fraction of [O III]5007 and Hα luminosity estimated to come from the AGN in the composite galaxies (see text for details). The solid line refers to Hα and the dashed to [O III]5007.

regions) populate a narrow sequence in the BPT diagram, because there are strong correlations between ionization parameter, dust-depletion and metallicity. A galaxy with both an AGN and ongoing star formation will be located to the right of this line (see also Kewley et al. 2001b).

The right-hand panel in Fig. 9 shows the BPT diagram for the SF class, but this time as a conditional density diagram (i.e. the distribution in each bin along the x-axis has been normalized to unity). This distribution is very similar to that found for H II regions in nearby galaxies (e.g. Bresolin & Kennicutt 2002; Kennicutt et al. 2000). The dashed lines show fits to the intervals enclosing 95 per cent of the distribution. The fits use only the data with log [N II]6584/Hα <

−0.5. This ensures that the fits are the same if the AGN and C classes are included in the distribution.

We will take the upper dashed line in the right-hand panel as the upper limit for our pure star formation sequence. For each galaxy above this limit, we ask what AGN contribution needs to be subtracted to move the galaxy below the limit. The average AGN we use is indicated with a cross in the left-hand panel in Fig. 9. The results are fairly robust to the exact location of this average AGN until one gets close to the AGN region.

In Fig. 10 we show the results of this exercise for the composite class. The solid histogram shows the fraction of galaxies that have a given fraction of their Hα flux coming from an AGN. The dashed histogram shows the same for [O III]5007. At the top of the AGN plume in Fig. 9, all the [O III]5007 flux can be expected to come from an AGN, but we do not expect more than ∼40 per cent of the Hα flux to have a non-stellar origin for any galaxy.[6]

---

[6] This number depends somewhat on the average AGN spectrum adopted.

In summary, we find that 11 per cent of the (observed) Hα luminosity density in the composites is likely to come from an AGN. For [O III]5007 AGNs contribute 41 per cent of the luminosity density in the composites. This means that using the Hα luminosity for the Composite class without any correction for AGN activity will only mildly overestimate the SFR_e.

Because different lines are affected by AGN in different ways, model fitting may give unreliable results. We therefore estimate the in-fibre star formation in the Composite and AGN classes using their measured D4000 values, as discussed below.

We should also comment that the pure SF sequence differs slightly from the definition of the SF class, but we have checked that no more than 0.2 per cent of the Hα luminosity density in the SF galaxies is expected to be of AGN origin, which is well below our systematic uncertainties and we will ignore this henceforth. The AGN contribution to the [O III]5007-luminosity of the SF galaxies can also be shown to be well below the level necessary to bias our model fits.

4.1 Estimating the SFR/M_* from D4000

For the Composite galaxies and AGN in the sample, we cannot use the model fits to determine the SFR_e values, because the line fluxes are likely to be affected by the AGN component. In principle it is possible to subtract off an AGN component, but in practice this is a very uncertain procedure. We have therefore decided to use the measured D4000 value to estimate the SFRs and will denote this SFR_d.

This method uses the relationship between SFR_e/M_* and D4000 shown in Fig. 11 to estimate the specific SFR_d of a galaxy, and from there its total SFR_d. This relation shown in the figure has been constructed using the derived SFR_e likelihood distributions and the measured D4000 values (we assume Gaussian errors on the latter). The contours show the sum of the probability distribution function (PDF) in each bin. The upper solid line shows the average of the distribution at a given D4000 and the lower solid line the mode. The dashed lines show the 95 per cent limits at a given D4000. To derive the likelihood distribution of the SFR_d of a given galaxy, we convolve the likelihood distribution in Fig. 11 with the likelihood distribution of D4000 for that galaxy.

To minimize the possibility of biases, we did not make a S/N cut when constructing Fig. 11, but we did throw out all AGN, C and low S/N AGN. To test the reliability of this procedure, we have
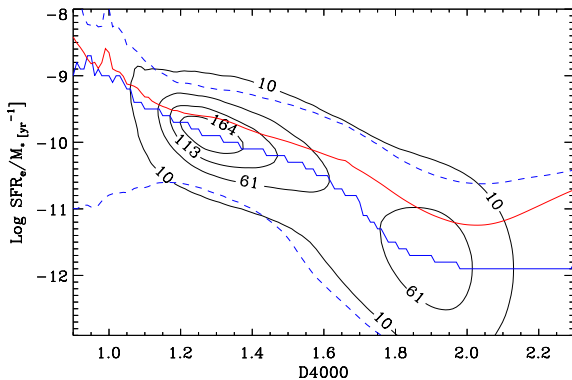

Figure 11. The observed relationship between D4000 and specific SFR_e (both inside the fibre). The contour lines are equally spaced from 10 galaxies per bin to the maximum. The bin size is 0.01 × 0.1 in the units given in the plot. The upper solid line shows the average at a given D4000 and the lower solid line shows the mode of the distribution. The dashed lines show the limits containing 95 per cent of the galaxies at a given D4000.

checked that if we co-add the spectra of the low S/N galaxies in the SF region of the BPT diagram and run these spectra through our pipeline, we retrieve the same SFR_e values (within a couple of per cent) that are obtained if we co-add the SFR_d values derived for individual galaxies using the D4000 calibration outlined here.

In what follows we will use SFR_d for the AGN, C and Unclassifiable classes and SFR_e for the SF and low S/N SF classes as our best estimate of the fibre SFR, and we will refer to this as the SFR.

5 APERTURE EFFECTS

The SDSS is a fibre-based survey. At the median redshift of the survey, the spectra only sample ≈1/3 of the total galaxy light. The presence of radial gradients in galaxy properties can therefore lead to substantial uncertainties when one wants to correct to total quantities. The implications of these biases have been discussed in some detail by several authors (e.g. Kochanek, Pahre & Falco 2001; Baldry et al. 2002; Gómez et al. 2003; Nakamura et al. 2003; Pérez-González et al. 2003).

We now discuss our method for aperture correction which uses the resolved colour information available for each galaxy in the SDSS. We will show that by using this empirically based aperture correction we can remove the aperture bias.

Our aperture correction scheme focuses on the likelihood distribution $P(SFR/L_i|colour)$, i.e. the likelihood of the specific star formation rate for a given set of colours. It is essential to keep the entire likelihood distribution in order to avoid biases in our estimates, as the measured distributions of SFR/L_i for certain colours turn out to be multipeaked. We choose to normalize to the $^{0.1}i$-band luminosity. This is the reddest band with uniformly small photometric uncertainties. We have verified that the results do not depend significantly on the choice of photometric band.

In theory we could construct likelihoods for SFR/L_i using all the colour information available. This has the disadvantage that even with our large sample, many bins have very few galaxies. The method is then sensitive to outliers. It turns out, however, that $^{0.1}(g-r)$ and $^{0.1}(r-i)$ contain all the useful information necessary for our purposes. Adding other colours does not significantly improve our constraints, this is also true if we use the model magnitudes which give somewhat higher S/N in the u and z magnitudes (Baldry et al. 2004).

We therefore construct $P(SFR/L_i|colour)$ on a grid with bins of size 0.05 in $^{0.1}(g-r)$ and 0.025 in $^{0.1}(r-i)$. We add together the likelihood distributions for SFR/L_i (inside the fibre) for all galaxies in each bin (we exclude AGN and low S/N AGN, including them with their SFRs estimated from D4000 does not change the results). The result of this procedure is illustrated in Fig. 12. This shows $P(SFR/L_i|colour)$ for a subset of the full $^{0.1}(g-r)$, $^{0.1}(r-i)$ grid. The colour is indicated as $^{0.1}(g-r)/^{0.1}(r-i)$ in the top left-hand corner of each panel and increases to the right for $^{0.1}(g-r)$ and upwards for $^{0.1}(r-i)$. The number of galaxies in each bin is also indicated in each panel as well as the number of galaxies that have colours in this bin *outside* the fibre (see below). The thick and thin bars on the x-axis show the average and median SFR/L_i for that bin, respectively. Because of the non-symmetric nature of these distributions, the average differs noticeably from the median.

The shape of the distributions varies strongly with colour and it is often not well represented by a Gaussian. In addition, higher values of SFR/L_i often occur at *redder* $^{0.1}(r-i)$ colours. This is because the emission lines around Hα contribute significantly to the $^{0.1}i$ flux during an episode of star formation. The likelihood distribution also becomes very wide for galaxies with $^{0.1}(g-r) \gtrsim 0.7$, $^{0.1}(r-i) \lesssim 0.5$
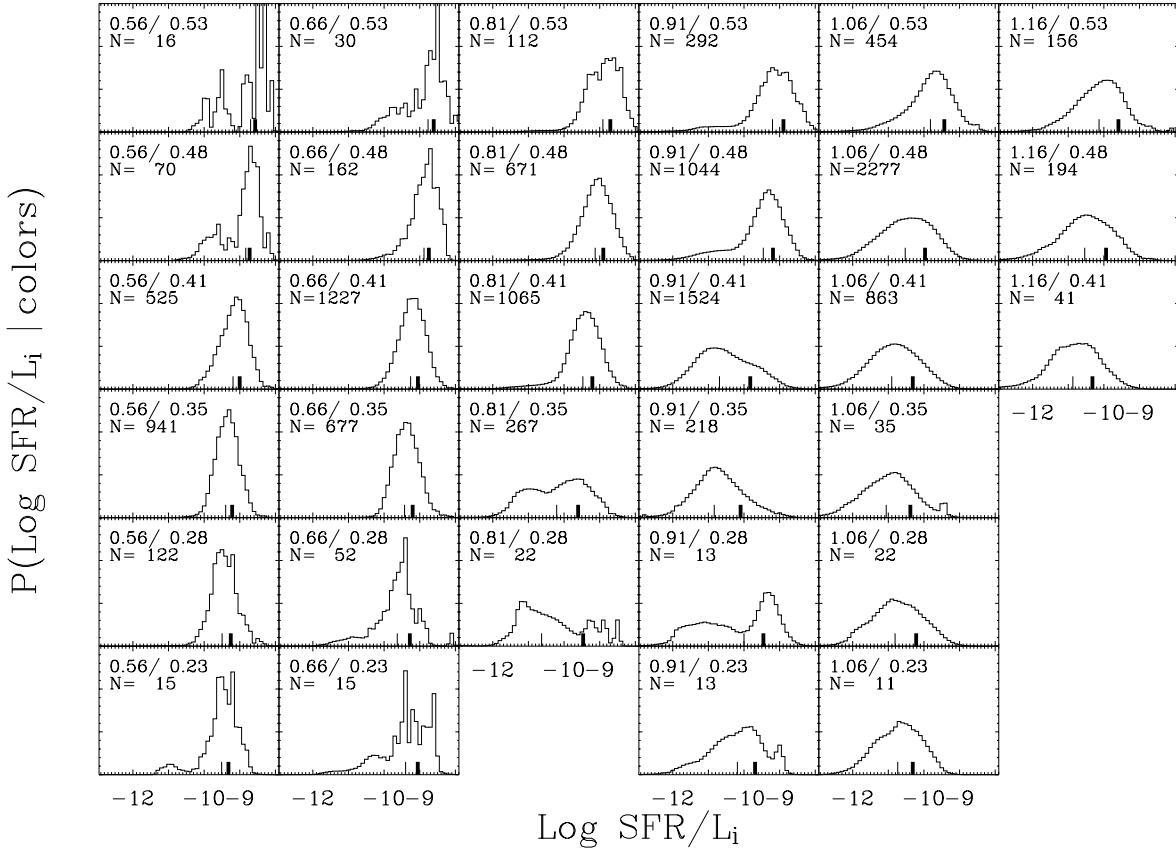
Figure 12. The likelihood distribution of log SFR/L_i in bins of $^{0.1}(g - r)$ and $^{0.1}(r - i)$. The $^{0.1}(g - r)$ colour increases to the right and is indicated as the first number in the top left-hand corner of each panel. The second number gives the $^{0.1}(r - i)$ colour and this increases upwards in the figure. The thick black mark on the x-axis shows the average SFR/L_i whereas the thin mark shows the median. Note that as one moves towards redder colours the distributions become highly non-symmetrical and there is a substantial difference between the average and the median. The number of objects contributing to the distribution shown in a given panel is indicated in each panel as N = X. Note also that this is a subset of the full 41×45 grid.

and their star formation rates are constrained to no better than a factor of 10. This poor constraint on SFR/L_i is simply due to the degeneracy between age, metallicity and dust. This region of colour space is populated both by dusty star-forming galaxies and by galaxies with old stellar populations.

We might consider improving our estimates by including the estimated gas-phase metallicity as a constraint. We cannot implement this procedure for the full sample, however, because gas-phase metallicities can only be determined for the SF class. Including the stellar mass as an additional constraint gives little apparent improvement.

We have also verified that k-correction uncertainties do not appreciably affect our aperture correction estimates. We did this by calculating $P(SFR/L_i|\text{colour})$ from galaxies with appropriate colours in a bin of width $\Delta z = 0.01$ in redshift around the galaxy in question. This method uses only observed colours but agrees very well with the procedure using k-corrected colours. The disadvantage of the model is that for a subset of the data there are not enough galaxies with the appropriate colours to construct reliable likelihood distributions. We have therefore settled for the simplest solution; we use only $^{0.1}(g - r)$, $^{0.1}(r - i)$ to constrain SFR/L_i.

We calculate the colour of the galaxy outside the fibre by subtracting off the fibre magnitudes from the CMODEL (i.e. total) magnitudes and we convolve this with $P(SFR/L_i|\text{colour})$ to get an estimate of SFR/L_i outside the fibre. We require that at least five galaxies

contribute to the estimate of $P(SFR/L_i|\text{colour})$ in a given bin. For the bluest colours, there are bins where no empirical estimate of $P(SFR/L_i|\text{colour})$ exists, as few galaxies in the survey have very blue fibre colours. For these bins we use the closest bin. We have tested that this procedure does not introduce any significant bias into our estimates. For a few galaxies (176) this procedure failed because of photometric problems, these have been excluded from the further analysis. No noticeable differences would be found if they were included with an average aperture correction for their mass.

We would like to emphasize that our method of aperture correction is purely empirical. Our main assumption is that the *distribution* of SFR/L_i for given $^{0.1}(g - r)$, $^{0.1}(r - i)$ colours is similar inside and outside the fibre. This assumption and the reliability of this procedure should clearly be tested using resolved spectroscopy. We are carrying out such a program and will report on it in an upcoming paper.

For now, the easiest way to test our assumption is to calculate $P(SFR/L_i|\text{colours})$ in bins of $(z/z_{max})^2$, where z_{max} is the highest redshift at which the galaxy in question would pass the sample selection criteria. This way we can compare $P(SFR/L_i|\text{colours})$ derived from galaxies where the fibre only samples the innermost regions with that derived from galaxies where the fibre sample a larger fraction of the galaxy. We have done this in eight bins in $(z/z_{max})^2$ and compared the resulting predicted log SFR/L_i for the sample galaxies in each of these bins with that derived from the
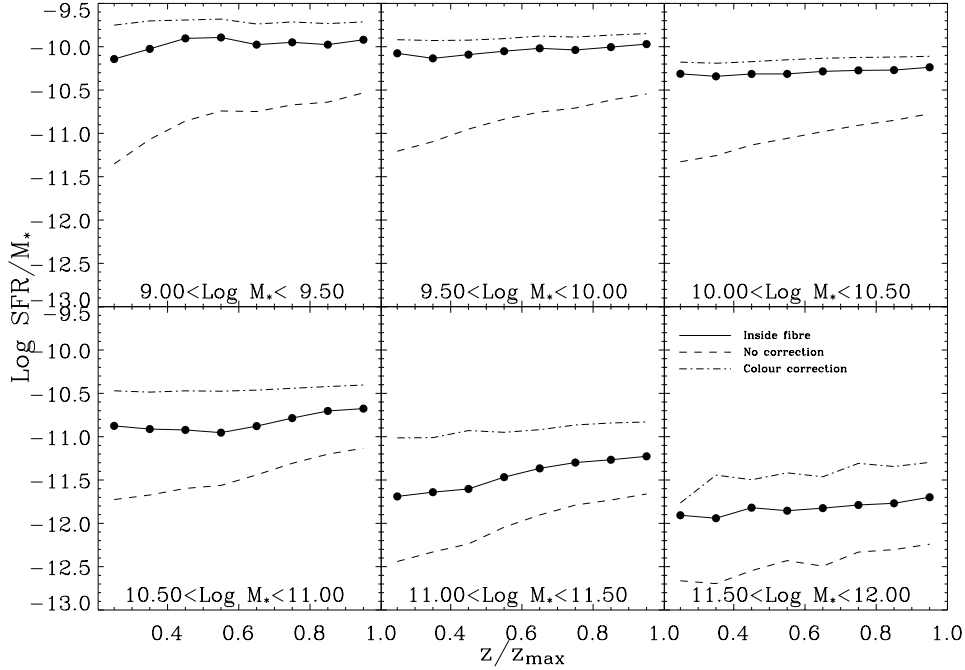
Figure 13. The value of the specific SFR, SFR/M_* as a function of z/z_{max}. The solid line with dots shows the trend within the fibre, i.e. both M_* and SFR calculated within the fibre and a clear aperture effect is seen. The dashed line shows the result when assuming that *all* the SFR takes place inside the fibre, i.e. M_* is for the galaxy as a whole and the SFR for the fibre only, and the dash–dotted line shows the result from our aperture corrections. Notice that we very nearly take out all aperture biases.

total sample. There is a spread, reflecting the intrinsic scatter in the method, but for $(z/z_{max})^2 \gtrsim 0.4$ individual bins gave results in agreement with the overall distributions. This indicates that the distribution of SFR/L_i at a given $^{0.1}(g - r)$, $^{0.1}(r - i)$ is fairly similar at different radii as long as we sample $\gtrsim 20$ per cent of the total r-band light [the median fraction of light within the fibre for galaxies with $(z/z_{max})^2 \gtrsim 0.4$]. This is the case for the majority of the sample.

Fig. 13 shows that our method removes essentially all the aperture bias from the sample. This figure shows the median SFR/M_* as a function of z/z_{max} in six different mass bins. All galaxies have been included. The specific SFR is estimated from the measured value of D4000 for the AGN and C classes (cf. Section 4.1). The likelihood distributions of SFR/M_* have been co-added at each bin in z/z_{max}.

In the absence of aperture effects (and evolution), this plot should show no trend with z/z_{max}. Note that evolution is unlikely to provide more than 0.1 dex change over the range plotted in these diagrams. The dashed lines in Fig. 13 show the trend if one assumes that *all* star formation activity takes place within the fibre but we use the stellar mass for the whole galaxy. This will clearly underestimate the true SFR/M_*. Fig. 13 shows that this quantity decreases towards lower values of z/z_{max}, as expected.

The solid line with dots in the diagram shows SFR/M_* when both quantities are calculated inside the fibre and it is clear that there are still strong aperture effects for galaxies with $\log M_* \gtrsim 10.5$. This is expected since these galaxies often have prominent bulges, in which the specific SFR is expected to be low. For lower mass galaxies, the aperture problems are considerably smaller and it is clear that a simple scaling of the fibre SFR by the r-band flux, as done for example by Hopkins et al. (2003), is an acceptable aperture correction. This is also the case for galaxies that show no sign of star formation activity when more than a third of the total r-band light is sampled. It is not correct for other galaxies.

Finally, the dot–dashed lines show the result after applying our aperture corrections. Our correction has removed the aperture bias, at least in a statistical way. We note that the spread in SFR/M_* at a given z/z_{max} is substantial, typically 2 orders of magnitude. We have also tested whether the *distribution* of points is the same at all z/z_{max} after application of our aperture corrections. We find that it is, and that Fig. 13 is an appropriate summary of the process even though it only shows the median.

At this point, it is useful to summarize the SFR indicators that are used for each subclass. Inside the fibre we have the following situation:

(i) for the **SF** class we use the results of the full model fits;

(ii) for the **Low S/N SF** class we use the observed Hα luminosity and convert this into a SFR using the $\eta_{H\alpha}$ distribution functions derived for the SF class (Section 3.1); and

(iii) for the **AGN**, **Composite** and **Unclassifiable** classes we use the D4000 value to estimate SFR/M_* and SFR using the method outlined in Section 4.1.

Outside the fibre, we always use the colour method to estimate the SFR.

The uncertainties on our individual galaxy SFR estimates are summarized in Fig. 14. To construct this figure we have calculated the 68 per cent confidence interval for the log SFR estimate for each galaxy and calculated the spread in this value in bins of log SFR/M_* (where the latter is taken to be the median of the distribution, but the results are the same with the mode or the average). All galaxies with log SFR/$M_* < -12$ have been placed in one bin at log SFR/$M_* = -13$.

The solid points show the median 68 per cent confidence interval in dex for the estimate of log SFR *inside the fibre* for each bin in log SFR/M_*. In the case of Gaussian errors this would correspond to 2σ. The error bars show the 68 per cent spread in error estimates within

Table 3. The best estimate of the star formation density at $z = 0.1$ from our survey. We show the mode and the 68 per cent confidence intervals as well as the percentage contribution to the total for the different classes. The total for $\beta = 0$ (no evolution) is shown on the first row, the others use $\beta = 3$.

| Sample	Fibre			Total			Per cent of total
			ρ_{SFR} [10^{-2} M$_\odot$ yr^{-1} Mpc^{-3}]				
	16 per cent	Mode	84 per cent	16 per cent	Mode	84 per cent	
Total ($\beta = 0$)	0.490	0.496	0.509	1.840	1.855	1.870	
Total ($\beta = 3$)	0.507	0.509	0.528	1.962	1.974	1.998	100.00
SF	0.317	0.320	0.324	1.001	1.018	1.024	51.58
Low S/N SF	0.063	0.064	0.065	0.445	0.448	0.464	22.69
Composites	0.078	0.080	0.092	0.203	0.207	0.210	10.48
AGN	0.026	0.026	0.028	0.082	0.084	0.086	4.26
Unclassifiable	0.019	0.019	0.020	0.213	0.216	0.228	10.96
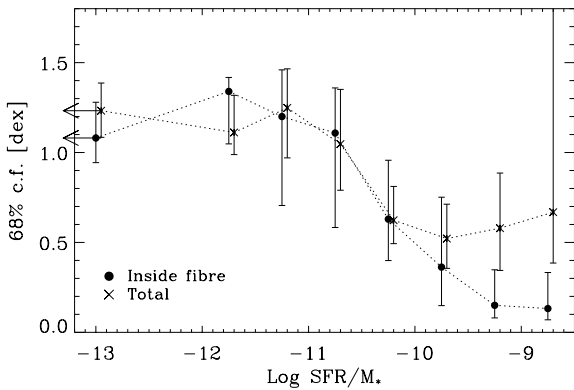
Figure 14. The distribution of the errors on the SFR estimates. The line with solid points show the uncertainties derived from the likelihood distributions inside the fibre. The line with crosses points shows the same for the likelihood distributions for the total SFR derived as described in the text. We have grouped together all galaxies with log SFR/M_* < -12 into one point and placed this at log SFR/$M_* = -13$. The x-axis shows log SFR/M_*, measured inside the fibre for the solid points and total for the crosses. The points for the total SFR/M_* are offset slightly for clarity.

that bin. The crosses show the same for the log SFR estimated after aperture corrections (i.e. total) and are slightly offset for clarity.

Most of the trends are easy to understand. First, the SFR estimates have lowest uncertainty at high SFR/M_* because the S/N is high in most lines and tight constraints can be obtained. For low values of SFR/M_*, the SFR inside the fibre is estimated from the D4000 value. This causes the error estimates to rise because D4000 provides a poorer constraint on SFR/M_*. Finally, the error decreases when the D4000 is so large that there is no longer any room for significant star formation.

The uncertainties on the total log SFRs are larger than for log SFR measured inside the fibre at high SFR/M_* because the aperture corrections are significantly more uncertain than the SFR estimates from the spectra. The spread in the uncertainty estimates is also increasing for increasing SFR/M_* because the outer parts of these galaxies are very blue and few galaxies are available for the calibration of the aperture corrections at these very blue colours (cf. Fig. 12). The median uncertainties on the total SFR/M_* are moderate, however, as the aperture corrections are in average well constrained as Fig. 12 shows.

At low values of SFR/M_* the total and fibre estimates have comparable uncertainties.[7] This is because the aperture corrections for these galaxies are calibrated by the fibre estimates with similar and broad likelihood distributions. The total log SFR therefore gets a very similar likelihood distribution. Some galaxies with poorly constrained log SFR in the fibre might nevertheless get better constrained log SFR total because they have bluer outsides than insides and hence better constrained aperture corrections.

The best test of our error estimates for the fibre SFRs is repeat observations. Approximately 5000 of our targets have been observed at least twice during the SDSS survey. We have used these duplicate observations to check that the errors we derive for the fibre SFRs are consistent with the dispersion between repeat observations.

6 THE LOCAL SFR DENSITY

We now turn to estimating the total SFR density of the local Universe. As discussed above, we have a likelihood distribution for the SFR for each galaxy. This includes both uncertainties in the model fit and the aperture correction. We also need to make a correction for evolution in the SFR over the 2.5 Gyr spanned by our redshift limits. We do this by assuming that the SFRs of all galaxies evolve as:

$$SFR(z) \propto (1 + z)^{\beta}. \qquad (5)$$

For the *integrated* SFR density, the favoured evolution exponent up to $z \sim 1$ is currently $\beta \sim 3$ (Hogg 2001; Glazebrook et al. 2003), although substantial uncertainty exists. We will therefore adopt this form for individual galaxies as well. We will calculate our results at $z = 0.1$. The differences between $\beta = 0$ and $\beta = 3$ are then fairly small ($\sim$6 per cent, see Table 3 below). Note that a *single* evolutionary correction for *individual* galaxies is likely to be wrong. As we will see below, there are clear differences in the star formation histories of galaxies of different mass, so although the *integrated* star formation density evolves as a power-law, the individual SFRs cannot show the same evolution. Fortunately, most of the results in the following sections are not dependent on the evolution within the sample.

Our results for the total SFR density are summarized in Table 3. This table gives the mode of the likelihood distribution of ρ_{SFR} as well as the 68 per cent confidence intervals. We show the result for

---

[7] If we had plotted the uncertainty on the SFR rather than log SFR this would of course not be the case.

$\beta = 0$ on the top line and the following lines all assume $\beta = 3$ – the relative contributions of the different classes do not depend on β. To estimate the star formation density of the survey, we add up the likelihood distributions, using $1/V_{\max}$ as a weight (Felten 1976). As discussed in Appendix A, the summation is carried out with a Monte Carlo method. To calculate errors on the final SFR density, we carry out 100 bootstrap repetitions. For each repetition we draw 149 660 objects from our sample with replacement which makes up the bootstrap sample. For this sample we carry out 30 Monte Carlo summations as discussed in Appendix A – the results do not change when these numbers are increased. We then calculate the mean, mode, median and confidence intervals from the distribution of results obtained for each bootstrap repetition and Monte Carlo summation. The confidence intervals reflect only random errors. The systematic uncertainties are substantially larger.

It is, of course, difficult to estimate accurately the systematic uncertainties. The first is the uncertainty in the estimator used to calculate the SFR inside the fibre for the Composite, AGN and Unclassifiable class. We have experimented with various estimators before settling on D4000. They all agree fairly well with a spread of 2 per cent which has to be taken as a systematic uncertainty.

Our SFR density estimate is sensitive to our aperture correction scheme. We remarked above that our main assumption is that the distribution of $\log \mathrm{SFR}/L_i$ is the same at a given colour inside the fibre as outside. This is not necessarily so, something which is hinted at in the clear difference between the mode, median and average for the likelihood distributions. These indicate that the likelihood distribution for a given colour might be composed of at least two components and the balance between these could be different in the outer regions of the galaxy, cf. the double-peaked nature of some distributions in Fig. 12. It is therefore possible to take the difference between the average (the formally correct estimate to use) and the mode as indicating the importance of this assumption. We have therefore carried out the aperture corrections also using the mode as an estimator of $\log \mathrm{SFR}/L_i$. In this case we find a SFR density 15 per cent lower than that in Table 3.

In addition there is a 6 per cent difference between $\beta = 0$ and $\beta = 3$ which we take as an additional systematic uncertainty of ± 3 per cent. Finally, the differences between the linear interpolation scheme that we use to draw numbers from the likelihood distribution and higher order schemes also produce a scatter of 1–2 per cent per cent in the calculation of the average. We will therefore adopt $^{+6}_{-21}$ per cent as indicative of the expected range of the systematic uncertainties.

This gives our best estimate for the total SFR density at $z = 0.1$ as

$$\rho_{\mathrm{SFR}} = 1.915^{+0.02}_{-0.01}(\mathrm{rand.})^{+0.14}_{-0.42}(\mathrm{sys.}), \tag{6}$$

in units of $10^{-2} \, h_{70} \, \mathrm{M_\odot \, yr^{-1} \, Mpc^{-3}}$. The random errors correspond to the 68 per cent confidence interval, while the larger systematic error is the expected total range. We have here taken the average of the $\beta = 3$ and $\beta = 0$ estimates because this difference is included in the systematic uncertainties.

We have calculated the SFR density in bins of absolute $^{0.1}r$-band magnitude and this is plotted as solid circles in Fig. 15. For comparison, we plot the r-band luminosity density for our sample (see Blanton et al. 2003a, for a more in-depth discussion) scaled to the same peak value as the SFR density (crosses). If we fit both distributions with a (Schechter 1976) function of the form

$$\phi(L)\,\mathrm{d}L = \phi_0 \left(\frac{L}{L_*} \right)^{\alpha_{\mathrm{FS}}} e^{-L/L_*} \frac{\mathrm{d}L}{L_*}. \tag{7}$$
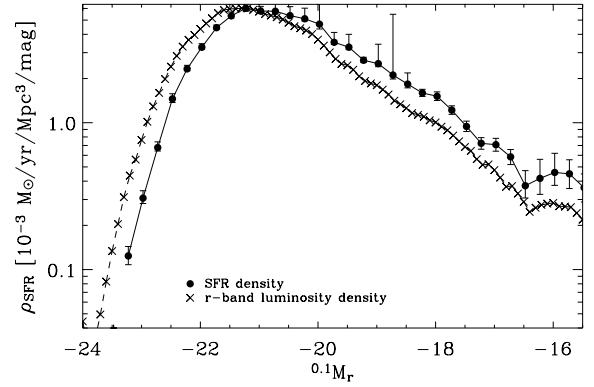
Figure 15. The star formation density as a function of r-band absolute magnitude compared with the r-band luminosity density scaled to have the same peak value. The error bars on the SFR density are derived from bootstrap resampling. The error bars on the luminosity density estimate are suppressed for clarity.

we find that the faint-end slopes, α_{FS}, are the same for the two functions, but that $Ł_*$ is 0.27 mag fainter for the star formation density. The error bars on the SFR density are derived from bootstrap resampling of the sample and do not include systematic errors. These will to good approximation only affect the overall normalization, and not the shape.

The curves are very similar in shape. At faint magnitudes where the majority of galaxies are strongly star-forming the r-band light is a good tracer of the SFR activity, but low luminosity galaxies are somewhat more important for the overall SFR density than for the r-band luminosity density. At the bright end of the distribution, there is less star formation per unit r-band luminosity. We will see later that this is even more prominent when plotted against the stellar mass.

Very little of the total SFR density occurs outside our sample (see also Glazebrook et al. 2003, their fig. 10). When integrating up the Schechter function, we find that only 1–2 per cent of the total SFR density originates outside our selection limits. We have therefore decided *not* to correct our values for incompleteness.

Fig. 16 places our estimate together with other recent estimates from the literature. In order to make this comparison we have adjusted the SFR density reported to our IMF. All values have been corrected to $h = 0.7$, but no attempt has been made to adjust the values for small differences in the assumed cosmological model.

We have indicated different methods of estimating the SFR using different colours in the figure. It is clear that radio determinations (Haarsma et al. 2000; Condon et al. 2002; Serjeant et al. 2002) typically give the highest estimates, although alternative calibrations might bring these down somewhat (Hopkins et al. 2003).

Our value seems to be in good agreement with other Hα-based estimates (Gallego et al. 1995; Tresse & Maddox 1998; Glazebrook et al. 2003). In particular our value is in very good agreement with that derived for the SDSS by Glazebrook et al. (2003) in their analysis of the cosmic spectrum. We show the full range of their Hα-based estimate for the local SFR in Fig. 16 and it is clear we are in very good agreement despite the different approaches taken. It might seem surprising that our error-bar is no smaller than many other determinations, which have used samples that are two orders of magnitude smaller than ours. This is mostly due to the adoption of a rather conservative systematic error for the aperture corrections.
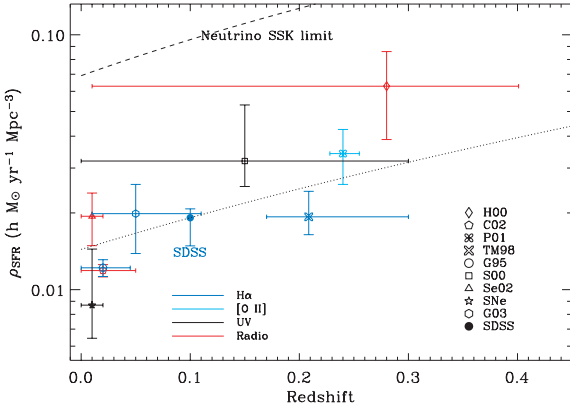
Figure 16. The star formation density in the present study. The filled circle is our preferred estimate with the systematic and random errors included in the error bar. This is compared with several other recent determinations of the local SFR. The papers indicated in the legend are: H00: Haarsma et al. (2000), C02: Condon et al. (2002), P01: Pascual et al. (2001), TM98: Tresse & Maddox (1998), G95: Gallego et al. (1995), S00: Sullivan et al. (2000), Se02: Serjeant et al. (2002), SNe: SFR derived from SNe are as reported by Fukugita & Kawasaki (2003) who are also responsible for the upper limits from neutrino observations at SuperKamiokande, G03: Glazebrook et al. (2003) for whom we show the full range of the Hα-derived local SFR. The dotted line shows a $(1 + z)^3$ evolution of ρ_{SFR} for comparison. All values have been corrected to the Kroupa (2001) universal IMF and $h = 0.7$.

Finally, to close off this section, we comment on the contributions to the overall ρ_{SFR} from the different classes in Table 3. It is noteworthy that despite only contributing about 23 per cent of the total stellar mass in the local Universe, the SF class is the dominant class in terms of SFR density. Over 50 per cent of ρ_{SFR} comes from this class. Likewise, it is interesting that up to 15 per cent of the total SFR takes place in galaxies that show signs of AGN activity (cf. Kauffmann et al. 2003c).

7 THE OVERALL PROPERTIES OF LOCAL SF GALAXIES

We now study the dependence of the SFRs on other physical parameters of the galaxies. Most of the results we show below will be familiar, as the general trends have been known for a long time. The main importance of what follows is that for the first time it is possible to derive the full distribution functions for these quantities. This adds considerable quantitative information to what was known before. One important point is that the results shown in this section are only very weakly dependent on the systematic uncertainties discussed previously. We will therefore ignore these in what follows.

We start with Fig. 17, which shows the SFR distribution as a function of stellar mass for the SF, low S/N SF and unclassifiable classes. The figure has been volume weighted and normalized in bins of stellar mass so it shows the conditional probability of SFR given a stellar mass. The clear correlation between SFR and stellar mass over a significant range in $\log M_*$ is noticeable and will be a recurring theme in what follows. Note that that at $\log M_*/M_\odot \gtrsim 10$, the distribution of SFRs broadens significantly and the correlation between stellar mass and SFR breaks down.

In addition to the correlations between different galaxy parameters, it is instructive to study 1D projections of our multidimensional distribution functions. This enables us to address the question of *how much* of the total SFR density takes place in different kinds of galax-
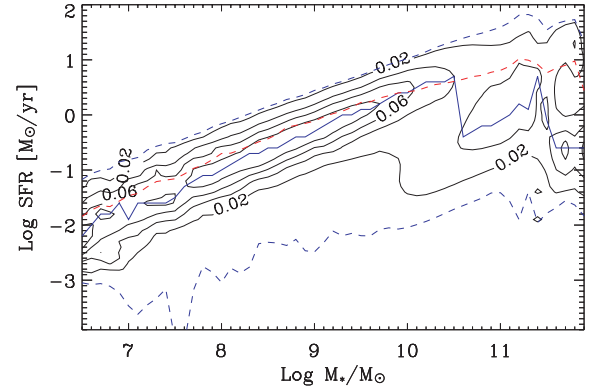

Figure 17. The relationship between the stellar mass and the SFR (both inside the fibre) for all galaxies with no AGN contribution. The figure has been volume weighted and normalized in bins of stellar mass. The contours are therefore showing the conditional likelihood of SFR given a stellar mass. The bin size is 0.1×0.1 in the units given in the plot. The red line shows the average at a given stellar mass, whereas the blue line shows the mode of the distribution. The dashed lines show the limits containing 95 per cent of the galaxies at a given stellar mass.

ies. To do this we carry out the same Monte Carlo summation that we used to derive the total SFR density, but this time in bins of mass, concentration, size, surface density, etc. The results of this exercise are shown in Figs 18 and 19. These show the star formation density per bin for a set of eight different physical quantities. The bin size is indicated in each panel. The shaded grey area shows the 68 per cent (1σ) confidence interval determined from the Monte Carlo and bootstrap summations.

The top left-hand panel in Fig. 18 shows ρ_{SFR} as a function of the central concentration of the galaxies, defined as the ratio of the radius containing 90 per cent of the r-band light to that containing 50 per cent. As discussed by Shimasaku et al. (2001) and Strateva et al. (2001), this quantity correlates quite well with morphological type for the galaxies in the SDSS. It related to other similar quantitative measures of galaxy concentration discussed previously (e.g. Morgan 1958; Doi, Fukugita & Okamura 1993; Abraham et al. 1994). The plot shows that the ρ_{SFR} distribution is broad and that it peaks around the value for a pure disc $R_{90}/R_{50} = 2.33$. The SFR density contributions split into individual galaxy subclasses are shown by thin lines as described in the legend. We see that the SF class contributes primarily to the disc-like part of the distribution. The star formation in AGN occurs in galaxies that are more bulge-like.

The top right-hand panel shows ρ_{SFR} as a function of galaxy stellar mass. We can observe several interesting features here, as follows.

(i) A considerable amount of star formation takes place in low mass galaxies. The best-fitting Schechter function has a faint-end slope of -0.45 and a characteristic mass of $\log M_*/M_\odot = 10.95$ with a possible transition to a steeper slope of $\alpha_{FS} \approx -0.55$ at $\log M_*/M_\odot < 9.8$. Approximately 50 per cent of the total SFR takes place at $M_* > 2 \times 10^{10}\ M_\odot$ and about 90 per cent at $M_* > 10^9\ M_\odot$.

(ii) The SF class almost completely dominates the SFR budget at masses less than about $10^{10}\ M_\odot$.

(iii) At masses $>10^{10}\ M_\odot$ the majority of SFR takes place in galaxies low S/N SF class with important contributions estimated from galaxies that either cannot be classified or which show signs of AGN in their fibre spectra.
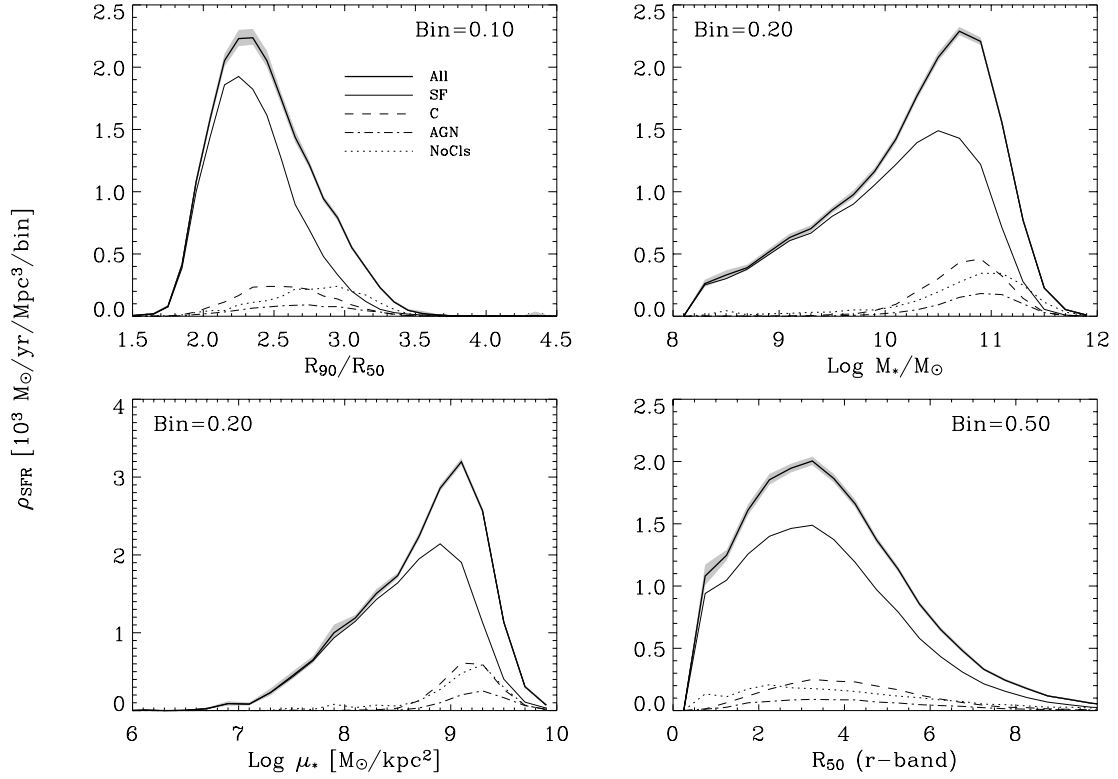
Figure 18. The total star formation density, ρ_{SFR}, as a function of Top left: the central concentration of the galaxies. As discussed in section 7 this is a proxy for a proper morphological classification. Top right: the stellar mass and Bottom left: the stellar surface mass density. The final panel shows ρ_{SFR} as a function of the galaxy half-light radius. The shaded region in each panel shows the 68 per cent confidence region. The units are given on the left, and the bin size is given in each panel.
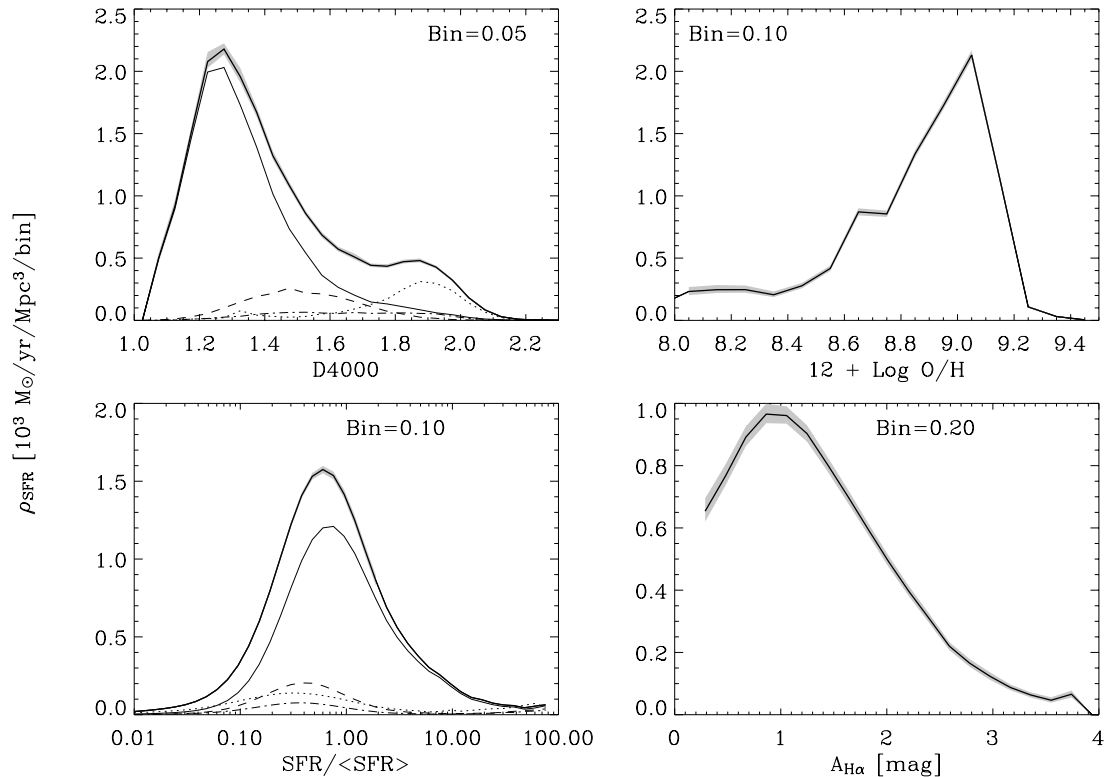
Figure 19. Like the previous figure, but this time we show parameters determined inside the fibre along the *x*-axis and the distribution functions are therefore relative to the SFR density *inside* the fibre. The top left-hand panel shows the distribution with respect to the 4000-Å break, the gas-phase oxygen abundance is at the top right, the present-to-past average star formation rate is at the bottom left, and finally the dust attenuation at Hα is in the bottom right-hand panel.

Many of the same trends can also be seen in the bottom left-hand panel which shows ρ_{SFR} as a function of the stellar surface mass density (Kauffmann et al. 2003b). Taken together with the top left-hand panel, it is clear that the majority of the SFR comes from high-surface brightness spirals. These typically have half-light radii around 3.0 kpc (slightly smaller than the Milky Way).

Turning to Fig. 19, we now plot quantities that are determined inside the fibre aperture on the x-axis. We caution that because we only know these quantities within the fibre aperture, they may not be representative of the galaxy as a whole. The top right-hand panel shows that most of the star formation takes place in galaxies with low D4000. 12 per cent of the total SFR density comes from galaxies with D4000 > 1.8, 2 per cent with D4000 > 2.0. Note that the star formation in these galaxies comes entirely from their outer parts and they are probably spiral systems with significant bulges. The bottom left-hand panel shows that the majority of star formation is taking place in galaxies that have a present-to-past average star formation rate between 0.3 and 0.5.

The top right-hand panel shows the distribution as a function of gas-phase oxygen abundance. Because the metallicity is derived from the fits to the CL01 models, we only show results for the SF class. It is interesting that most stars are forming in galaxies with solar or slightly super-solar metallicities, depending on the overall calibration of the metallicity scale (e.g. Kennicutt, Bresolin & Garnett 2003). Finally, the bottom right-hand panel shows the star formation density as a function of the attenuation at Hα (again, only for the SF class). This shows a peak around 1.0 mag of extinction in Hα, in good agreement with other determinations (e.g. Kennicutt 1983; Sullivan et al. 2000; Hopkins et al. 2003; Nakamura et al. 2003, C02). Notice that we calculate the SFR weighted distribution function of $A_{H\alpha}$, whereas most earlier studies have focused on a sample averaged value. It is clear from Fig. 6 that any such aver-

aging will depend strongly on the sample used. This has also been pointed out by Hopkins et al. (2003), who show that a radio-selected subsample of the SDSS has considerably higher attenuation than the average SDSS galaxy.

We now compare these SFR distributions with number and mass density distributions. Number distributions as function of many photometric properties and as a function of local density have been discussed in detail by Blanton et al. (2003b) and by Hogg et al. (2003). Likewise, the mass distributions have been discussed by Kauffmann et al. (2003a,b).

Fig. 20 is analogous to Fig. 18 and shows the distribution of ρ_{SFR} together with ρ_{M_*}, the mass density distribution, and ρ_N, the number density distribution. In these plots we have normalized the distributions to the total *within the plot limits*. We remark that the spikes seen in several of the number density distributions are due to a handful of galaxies that are extremely close to our selection limits with very high volume corrections. The bootstrap errors on these spikes are large.

Fig. 20 shows that although high-concentration galaxies are few in number, they contain the majority of the mass (see the discussion by Kauffmann et al. 2003a). As discussed previously, most of the star formation occurs in disc-like systems. The top right-hand panel shows that the majority of galaxies are small (cf. Blanton et al. 2003b) and that the galaxies dominating the star formation budget and those dominating the mass budget are fairly similar in size. The bottom left-hand panel in the figure shows the dependence on mass. We see that star formation is typically occurring in galaxies that are somewhat less massive than those that contain the bulk of the stellar mass in the local Universe. As we noted previously, most of the star formation is taking place in high surface brightness galaxies. However, the the bottom right-hand panel shows that the star formation is distributed over a wider range in stellar surface density than the
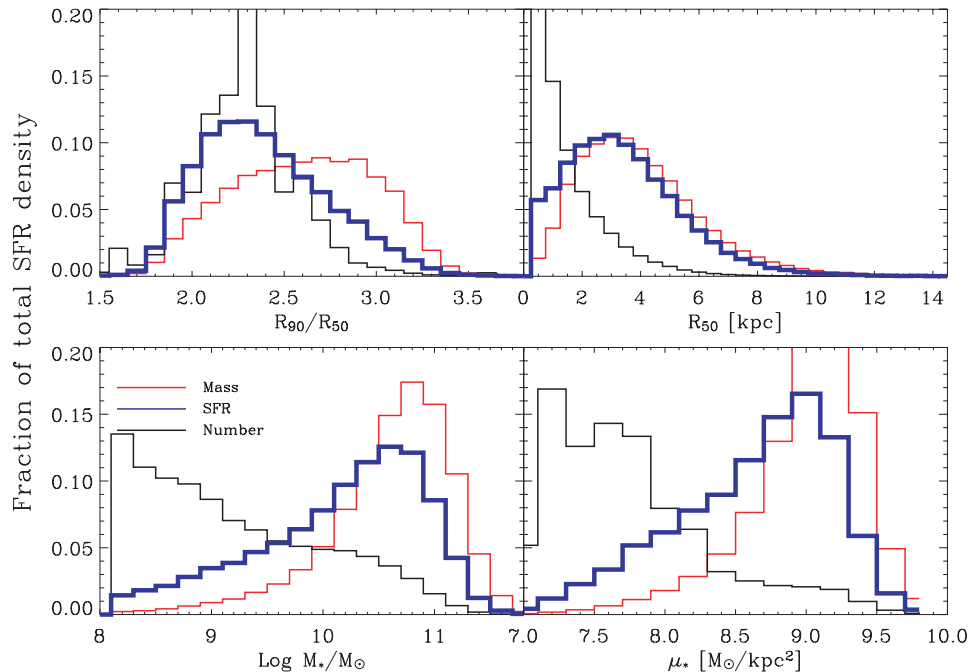

Figure 20. The contribution to the total number, mass and star formation density as a function of various galaxy parameters. The SF density is shown in blue, the mass density in red and the number density in black. Top left: the contribution to the different densities as a function of the concentration of the galaxies. Top right: the same, but as a function of the half-light radii of the galaxies. Lower left: the density contributions as a function of log stellar mass. Lower right: the contributions as a function of log of the stellar surface density in M$_\odot$ kpc^{-2}.
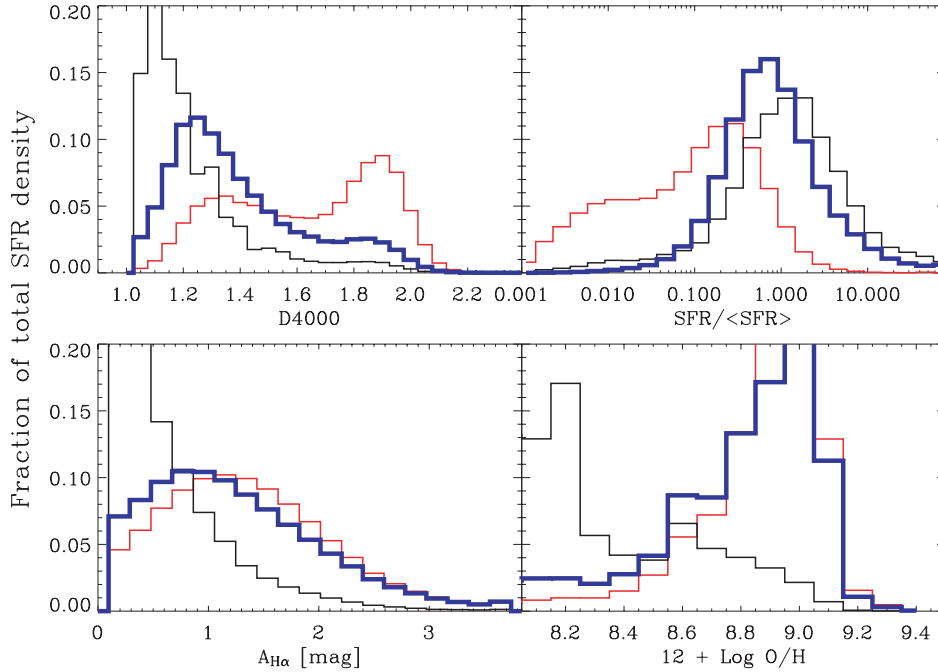
Figure 21. Like the previous figure, but this time showing the star formation density versus the 4000-Å break (upper left-hand panel) the present to past-average SFR (upper right), the dust attenuation at Hα in mag (lower left) and the gas-phase oxygen abundance (lower right). The latter two use only the SF class.

stellar mass. This is presumably because a significant fraction of the stellar mass is located in ellipticals, which are not currently forming stars.

Fig. 21 is analogous to Fig. 19 and shows properties measured within the fibre aperture. As discussed previously, the majority of the star formation is occurring in galaxies with low D4000. It is nevertheless intriguing that the majority of galaxies in the Universe have even lower D4000 than those that dominate the SFR density. A very similar result is illustrated in the top right-hand panel, which contrasts the contributions to ρ_N, ρ_{SFR} and ρ_{M_*} from galaxies with different present-to-past average star formation rates.

The bottom panels show the density distributions as a function of the attenuation at Hα (left) and as a function of gas phase oxygen abundance (right). These quantities are only calculated for the SF class. We see that the majority of SF galaxies have low metallicity and dust attenuation, two facts that are likely to be closely connected. Most of the mass and most of the star formation are found in galaxies with a wide distribution of $A_{H\alpha}$, centred around 1.0 mag with a spread of ∼0.7 mag for the SFR weighted distribution (see also the related study by Hopkins et al. 2003). For the $A_{H\alpha}$ distribution weighted by stellar mass, we find an average value of ∼1.3 mag and a similar spread. The difference reflects the fact that the most massive galaxies are more dusty than the galaxies dominating the SFR budget.

8 THE SPECIFIC STAR FORMATION RATE

The total SFR of a galaxy is interesting and for certain physical questions it is the relevant quantity. However, given the strong correlation between SFR and mass (cf. Fig. 17), it is clear that by normalizing the SFR by the stellar mass, one can more easily study the relationship between star formation activity and the physical parameters of the galaxies. In this section we will turn our attention towards the star formation rate per unit mass, the specific star formation rate. For compactness we will refer to the specific SFR as r_{SFR} below.

The specific star formation rate is closely related to several other important physical quantities. It defines a characteristic time-scale of star formation,

$$r_{SFR} = \tau_{SFR}^{-1} = \frac{\dot{M}_*}{M_*}. \tag{8}$$

Because the supernova rate is proportional to the star formation rate, the specific star formation rate represents the current input of supernova energy per star in the galaxy. Likewise, if the stellar mass is proportional to the mass of the ISM, it is related to the porosity of the ISM in the galaxy (Clarke & Oey 2002). Finally r_{SFR} does not depend on cosmology and is insensitive to the IMF as long as the high-mass slopes are comparable.

The star formation per unit mass has either explicitly or implicitly been used in numerous studies of field galaxies at $z < 1$ (e.g. Brinchmann & Ellis 2000; Cowie et al. 1996; Guzman et al. 1997; Ellis 1997, and references therein) by using equivalent widths of Hα or [O II]3727. In the local Universe, it has however often been rephrased in terms of the present to past-average star formation rate (e.g. Kennicutt 1983; Scalo 1986; Kennicutt et al. 1994; Boselli et al. 2001; Gavazzi et al. 2002). This is a suggestive quantity, which immediately gives an indication of the past star formation history of the galaxy and its relation to present-day activity. The main disadvantage of the present-to-past average SFR is that it involves more assumptions than the specific star formation rate, as we will see below.

The conversion between the two can be succinctly summarized as

$$b \equiv \frac{SFR}{\langle SFR \rangle} = r_{SFR} T (1 - R), \tag{9}$$

where R is the fraction of the total stellar mass initially formed that is returned to the ISM over the lifetime of the galaxy, and T is the time over which the galaxy has formed stars [$T = t_H(z) - t_{form}$, where $t_H(z)$ is the Hubble time at redshift z and t_{form} is the time of formation].

Given a choice of IMF, it is possible to reliably estimate R using stellar evolution theory. We use the median R estimated for our galaxies, which is $R \sim 0.5$. For T it is customary to take $T = t_H$, i.e. assume that all galaxies started to form stars at $t = 0$. We will adopt this convention below (with $T = t_H(0.1)$). This assumption is sensible when comparing average b values for large samples, but might be more questionable for individual galaxies. In general, this means that the b values we calculate will be upper limits.

Throughout this discussion, we will use b and r_{SFR} interchangeably, depending on which is more appropriate. As highlighted by Rocha-Pinto et al. (2000), the b value is a convenient parametrization of 'burstiness' and $b = 1$ provides a convenient divide between galaxies consistent with a constant or declining SFR over the age of the Universe and those that have a higher SFR at present than they have typically had in the past.

Before turning to the results for our sample, we note that there are at least two reasonable definitions for the *average* r_{SFR} and one must take care to specify which one is discussing.

The average r_{SFR} value of a sample of N galaxies is simply

$$r_{SFR}^g = \left\langle \frac{SFR}{M_*} \right\rangle = \frac{1}{\sum_i w_i} \sum_i w_i \frac{SFR_i}{M_{*,i}}, \tag{10}$$

where the weights w_i can be unity or e.g. $1/V_{max}$, depending on the question asked. This is the appropriate quantity to quote when talking about the typical galaxy. In practice we calculate this using the full likelihood distributions of r_{SFR} for each galaxy. We also usually choose to quote the median rather than the average as this is binning insensitive (cf. discussion in Appendix A).

r_{SFR}^g can be very different from the r_{SFR} value for a volume-limited sample:

$$r_{SFR}^V = \frac{\rho_{SFR}}{\rho_{M_*}} = \frac{\sum_i w_i SFR_i}{\sum_i w_i M_*}, \tag{11}$$

which is the appropriate quantity to quote for the the star formation history of the Universe as a whole. To calculate r_{SFR}^V using our likelihood distributions, we employ the same Monte Carlo summation technique used in the calculation of ρ_{SFR} in Section 6. It is important to include the stellar mass in this Monte Carlo code, because, as mentioned in Appendix A, neglecting to do so will bias our results. (To investigate the size of this bias we carried out the calculation of r_{SFR}^V with and without the full likelihood distribution of $\log M_*$, and found that ignoring the uncertainty in $\log M_*$ resulted in an overestimate of r_{SFR} of about 15 per cent.)

We begin by calculating b^V for the sample as a whole. This is a straightforward application of the results in the previous section and we show this quantity in Table 4. We see that for $h = 0.7$, the present-day Universe is forming stars at slightly higher than 1/3 of its past average rate, depending on the time-span one averages over in equation (9). Once again, we have calculated the likelihood distribution using the Monte Carlo method discussed in Appendix A

with 100 bootstrap repetitions and 30 Monte Carlo sums for each repetition. The resulting errors are also indicated in Table 4, but as with the total SFR density in Section 6 the random errors are substantially smaller than the systematic ones. Following the discussion in Section 6, we adopt the same estimates of systematic errors on the total SFRs and find a present-to-past average SFR for the local Universe of

$$b^V(z = 0.1) = 0.408^{+0.005}_{-0.002}(\text{rand.})^{+0.029}_{-0.090}(\text{sys.}), \tag{12}$$

with random errors corresponding to the 68 per cent confidence interval.

It is interesting to compare b^V for the different classes we have defined. If we look at quantities determined inside the fibre, it follows almost by definition that the SF class has the highest b^V value and the unclassifiable category the lowest. If we look at the values derived from the total SFR, the SF class still dominates, but it is interesting to note that the global b value for this class is the same as the central one. For the other classes one typically finds that the global b is higher than that inside the fibre. This is, of course, partly a sample selection issue: galaxies classed as SF are those that have relatively more star formation inside the fibre. It is also interesting to note that galaxies showing indications of AGN activity form an intermediate category between the Composites and the unclassifiable. This is again logical because AGN would have been classified as composites if they had substantially higher SFR.

We now study how the specific star formation rate varies between galaxies of different stellar mass. We split the galaxy sample into nine bins in *total* stellar mass. We sum up the likelihood distributions of r_{SFR}, i.e. we calculate r_{SFR}^g without the normalization in Equation (10). We show the resulting distributions of r_{SFR}^g in Fig. 22. The histograms are normalized to the total number of galaxies within that bin. No volume correction has been applied, but the difference to that shown is very small except in the lowest mass bins. (Note that we might not be complete in mass in the lowest bin.) The contribution to the distributions that is located outside the plotted range has been distributed on the two lowest and two highest bins respectively.

The histogram for galaxies with $10.0 < \log M_* < 10.5$ is repeated in each panel for reference. The shaded histograms show the distributions of those galaxies classified as recent starbursts based on their D4000 and $H\delta_A$ values by Kauffmann et al. (2003a). The thick black line shows the median volume weighted r_{SFR} value (r_{SFR}^g), and the red line shows r_{SFR}^V. The latter is higher as the total SFR is dominated by galaxies with extreme values of r_{SFR}.

Fig. 22 shows that there is a reasonably good agreement between the recent burst estimates of Kauffmann et al. (2003a), which are sensitive to activity in the last Gyr, and the activity indicated by emission lines. Thus galaxies that are currently undergoing elevated star formation activity (relative to the Hubble time) have typically done so for at least a Gyr. It is also clear that the *fraction* of galaxies

Table 4. The best estimate of the present-to-past average star formation rate, b^V. We have averaged the values form $\beta = 3$ and $\beta = 0$ as appropriate for our systematic error estimates.

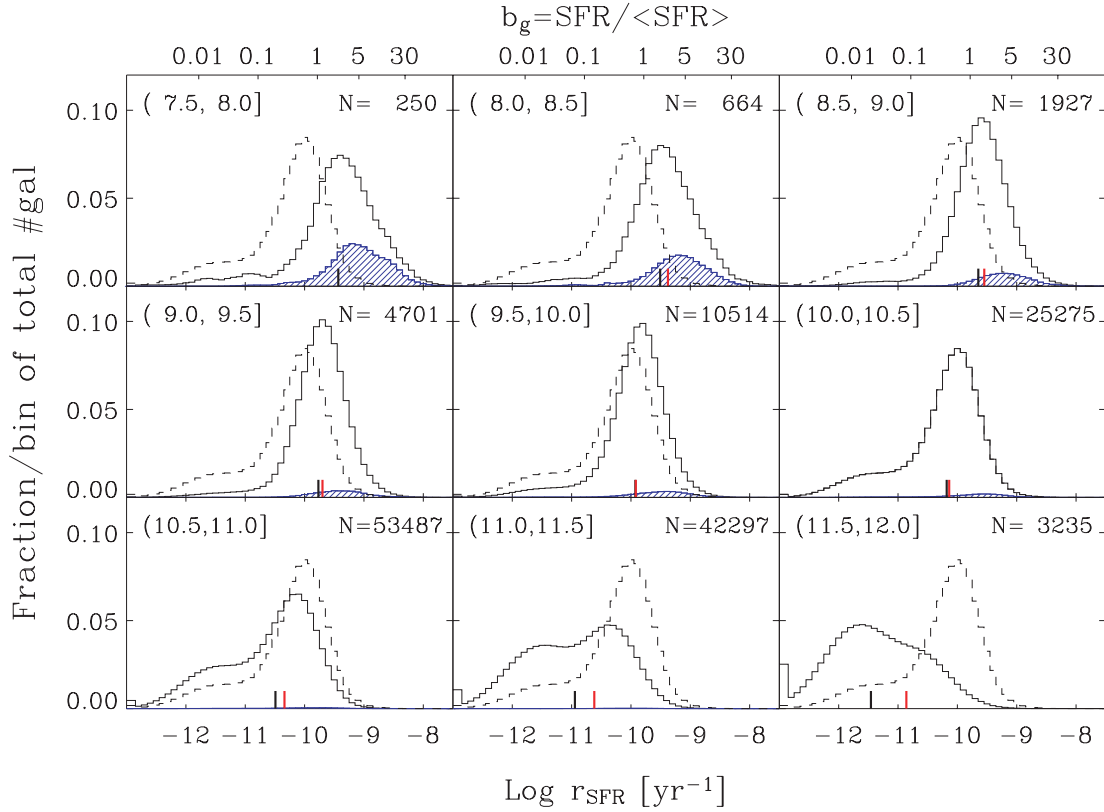| Sample	Fibre			Total		
	16 per cent	Mode	84 per cent	16 per cent	Mode	84 per cent
Total	0.313	0.314	0.325	0.406	0.408	0.413
SF	0.994	0.997	1.011	0.881	0.888	0.898
Low S/N SF	0.198	0.201	0.202	0.421	0.432	0.437
Composites	0.396	0.406	0.464	0.362	0.365	0.373
AGN	0.210	0.216	0.232	0.229	0.233	0.239
Unclassifiable	0.030	0.031	0.032	0.131	0.133	0.138

Figure 22. The observed specific star formation as a function of total stellar mass in the mass ranges indicated in the top left hand corner of each panel. Shaded areas show recent starbursts, see text for details. Each histogram is normalized to the total number of galaxies in the mass range, given as N in each panel. The dashed histogram is the distribution in the $\log M_* \in (10.0, 10.5)$ bin and is included for reference. The thick black bar shows the average b value, b^g, in each mass range, and the thick red bar shows b^V. The latter is not shown in the first panel because we might not be statistically complete there. All objects with $\log \text{SFR}/M_* \leqslant -12$ have been placed as a Gaussian around $\log \text{SFR}/M_* = -12$ as discussed in the text.

that show clear signatures of recent starbursts declines rapidly with mass. This can either be due to a lower gas fraction in more massive galaxies and/or the fact that periods of elevated star formation may be shorter in duration in more massive galaxies.

Fig. 22 also shows that the peak of the distribution shifts rather little over a large range in mass ($\log M_* \lesssim 10.5$). The peak value is very close to $b = 1$. In other words, most galaxies with $\log M_* \lesssim 10.5$ are forming stars at a rate that is consistent with their past average (i.e. at a constant rate). Above few $\times 10^{10} \, \text{M}_\odot$, we see a clear change in the shape of the distribution of b^g with a transition towards a low specific SFRs. This coincides well with the transition mass defined by Kauffmann et al. (2003a). Also note that the second peak showing up at $r_{\text{SFR}} \sim 10^{-12}$ at high mass is broader than the one at high r_{SFR} because the SFRs are less well constrained there.

Despite this near-constancy of the peak, there is a marked change in the shape of the tails. At low masses there is a substantial tail towards higher b-values, i.e. a trend towards more bursty star formation activity. This changes to a substantial tail towards low b-values at high masses. Many massive galaxies are now forming stars at a depressed rate compared to their past average.

9 THE BIRTHRATE PARAMETER AS A FUNCTION OF PHYSICAL PARAMETERS

Section 7 focused on the distribution of the SFR density as a function of different physical parameters. We will now study the relation between the specific SFR and these same quantities. This will help us

understand which factors control/are controlled by the star formation activity in individual galaxies.

Most previous studies have focused on the trend with mass and Hubble type. A key question has been which of these parameters correlate better with b. The studies of local spirals by Gavazzi & Scodeggio (1996) and Boselli et al. (2001) indicated that the driving variable was the stellar mass content. Other studies have made more of the correlation with Hubble type (e.g. Kennicutt et al. 1994). Here we will improve on these studies with a sample that is several orders of magnitude larger than the previous studies and with more careful and internally consistent modelling. We will not discuss the dependence on local density here. Recent studies of the correlation between star formation and density can be found in Gómez et al. (2003) and Balogh et al. (1998, 2001).

Throughout this analysis, we will show results using b^g and b^V derived from our *total* SFR estimates. In Fig. 23, we show the trend of b with central concentration/Hubble type. The left-hand panel in Fig. 23 contrasts b^g (solid line) with b^V (crosses). The solid line shows the median of the unweighted b^g calculated from the right-hand panel using the formalism discussed in Appendix A. We do not show the confidence intervals, but as is evident from the full likelihood distributions, they are wide. The crosses show b^V in each bin in R_{90}/R_{50}. This is calculated using Monte Carlo summation and is basically the ratio of the black and red curves in Fig. 20. In general b^V is larger than b^g because r_{SFR} has substantial tails towards large values. b^V is therefore dominated by galaxies away from the median.
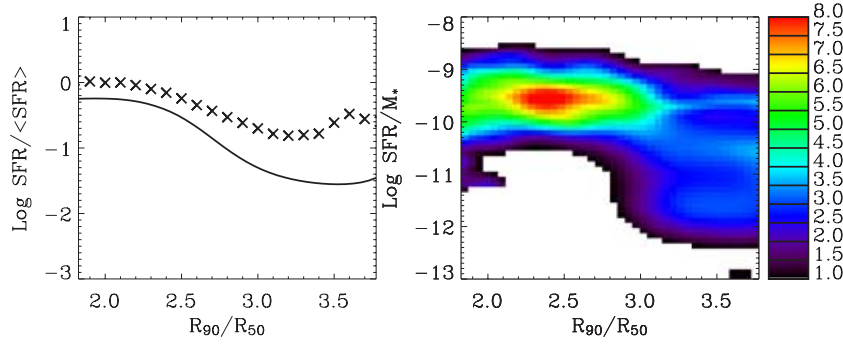
Figure 23. The specific SFR as a function of concentration. The left-hand panel contrasts b^g with b^V. The continuous line in this plot shows the median of the unweighted b^g (right-hand panel) and the crosses show b^V calculated from the data in Fig. 20. The right-hand panel shows the (log of the) observed likelihood distribution of r_{SFR} with respect to the concentration parameter, R_{90}/R_{50}, calculated as described in the text. The shading shows the conditional likelihood distribution (volume corrected) given a value for R_{90}/R_{50}. The contributions to the likelihood distributions below the plotted range have been put in the two lowest bins in r_{SFR}.

The right-hand panel of Fig. 21 shows the conditional likelihood distribution of r_{SFR} given the concentration parameter, $C = R_{90}/R_{50}$, for the local Universe. This has been constructed by assuming that $P(C)$ is independent of $P(r_{SFR})$ and assuming a 5 per cent error on C (this latter assumption is not important for the results). Each distribution has been weighted by $1/V_{max}$. The distribution in the right-hand panel then follows from a straight addition of all the individual $P(C, r_{SFR})$ and normalization of the distribution in each bin of C to produce the conditional likelihood distribution. As in Fig. 22 contributions to the likelihood distributions outside the plotted range in r_{SFR} has been placed in the two lowest bins in r_{SFR}.

These two panels illustrate the bimodal nature of the distribution of galaxies. This has recently been described and quantified in a number of papers (e.g. Strateva et al. 2001; Blanton et al. 2003b; Kauffmann et al. 2003b; Baldry et al. 2004). As can be seen, galaxies split into two basic populations: concentrated galaxies with low specific star formation rates, and low-mass, less concentrated galaxies with high specific star formation rates. The distribution at high concentration is broad however because there is also a fairly important contribution from concentrated, actively SF galaxies at very low redshift. When the figure is plotted for the sample directly, i.e. without volume weighting, the two regimes are much more clearly separated.

Fig. 24 shows, with the same layout, the distribution of b as a function of stellar mass. We see that although there is a clear decline in b with increasing stellar mass, the trend in b^g is weak over a large range in stellar mass until it suddenly drops at around $10^{11}\, h_{70}^{-2}\, M_\odot$.
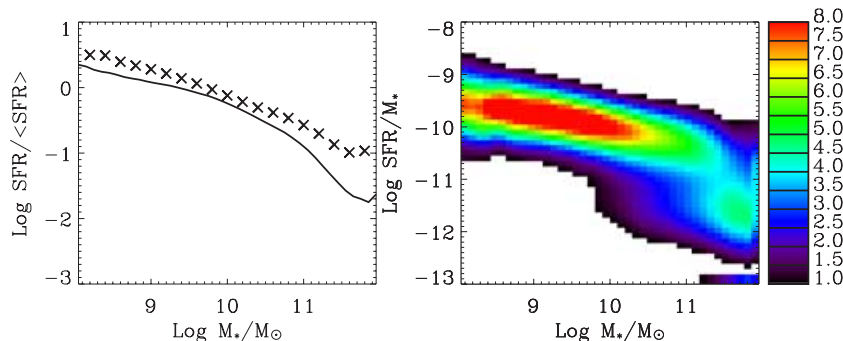
The b^V value shows a much smoother trend. This is caused by the large spread in r_{SFR} at given mass. If we compare Figs 24 and 23, we see that the Hubble type correlates somewhat less well with the b parameter than the stellar mass does.

Fig. 25 shows b as a function of stellar mass surface density, μ_*. The connection between this and M_* has been discussed in detail by Kauffmann et al. (2003b). Those authors show that μ_* correlates well with M_*, except at the very highest masses where μ_* reaches a saturation value. It is therefore not surprising that we see a very similar trend with μ_* as we saw with M_* in Fig. 24. When we plot the figure without volume weighting it is very similar but the low SFR peak becomes more prominent at high μ_* than at high M_*, suggesting that the surface density of stars is more important than the stellar mass in determining when star formation is turned off.

Fig. 26 again presents a similar view. This time the b parameter is plotted as a function of absolute r-band magnitude at $z = 0.1$. The correlation here is fairly good. For galaxies fainter than about $M_r = -20$, the b_g parameter increases by about 40 per cent mag^{-1}.

Fig. 27 shows the relationship between b and the 4000-Å break. The left-hand panel shows this to be a fairly tight relation – one reason why Kauffmann et al. (2003a) were so successful in constraining the M/L ratio for galaxies using the 4000-Å break and the $H\delta_A$ index. We note that this relation does have the potential to constrain star formation in high-redshift galaxies, as the Balmer break can be measured out to considerably higher redshift than Hα. From Fig. 27 we see that a measurement of D4000 allows one to constrain b_g to within a factor of 3 (1σ) for D4000 < 1.6. Although this is not very precise, it should be weighed against the fact that it is very

Figure 24. Similar to Fig. 23 but this time showing b as a function of the stellar mass.
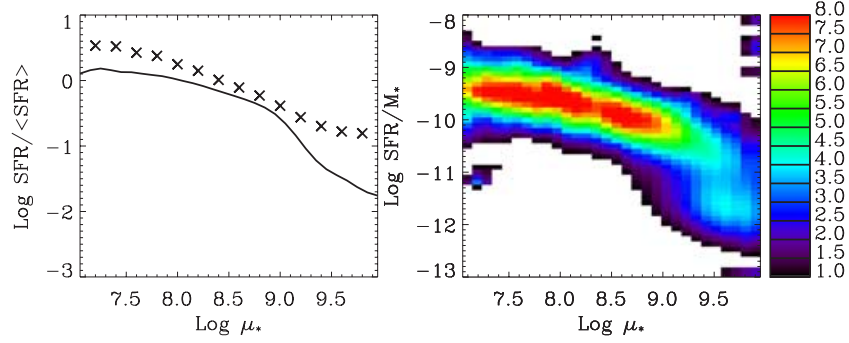
Figure 25. Similar to Fig. 23 but this time showing *b* as a function of stellar mass surface density.
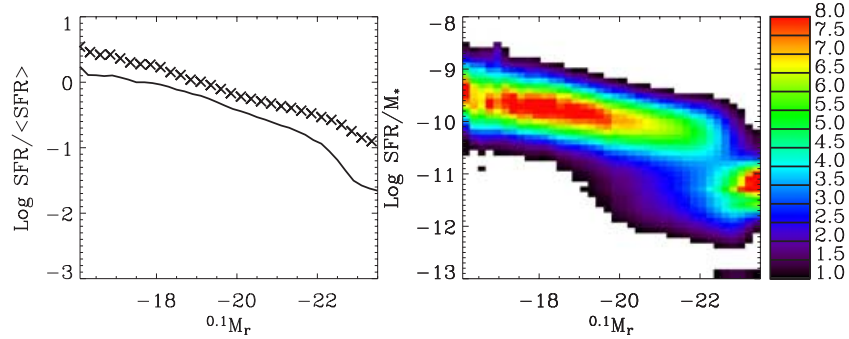

Figure 26. Similar to Fig. 23 but this time showing *b* as a function of $^{0.1}M_r$ mag.
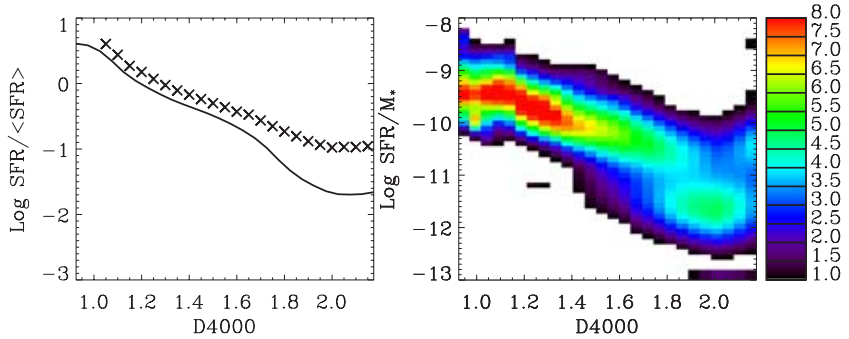

Figure 27. Similar to Fig. 23 but this time showing *b* as a function of D4000. Notice the clear correlation, despite a substantial scatter. In this figure the solid line shows b^g determined in the fibre whereas the crosses are for the total.

difficult to estimate SFRs to better than within a factor of 2 with any technique. It is however important to caution that the D4000 index intrinsically depend on the SFR in the last Gyr rather than the current SFR, as well as the metallicity. It is therefore necessary to assume that the present SFR is similar to the average activity in the last Gyr to make recourse to the relation in Fig. 27.

We caution that D4000 is measured within the fibre, whereas our SFR estimates are for the entire galaxy. If we make the same plot using SFR estimated within the fibre, we still recover a tight correlation between b^g and D4000. Moreover, we find that b^g and b^V now coincide. It is galaxies that have high D4000 within the fibre but significant star formation outside which make b^V considerably higher than b^g at high D4000 in Fig. 27.

We finish by showing *b* versus R_{50} in Fig. 28. It is clear that the low- and high-*b* galaxies co-exist over the entire range in R_{50}.

However, there is a general trend for higher b^V at the smallest sizes and for large galaxies to have typically lower values of b^V.

9.1 Comparison with previous studies

Before we turn to a discussion of these results, it is enlightening to compare our results with previous work in more detail. The main comparison samples are those presented by Kennicutt et al. (1994) and Gavazzi et al. (2002) in their studies of star formation histories in spiral galaxies. They calculate *b*-parameters for a range in morphological types and in Fig. 29 we plot this distribution of b versus approximate morphological type for our sample. We have taken the relation between RC3 T-type and concentration from Shimasaku et al. (2001) to convert concentration measures into approximate T-types. We fitted a linear relation between *T* and *C* from their
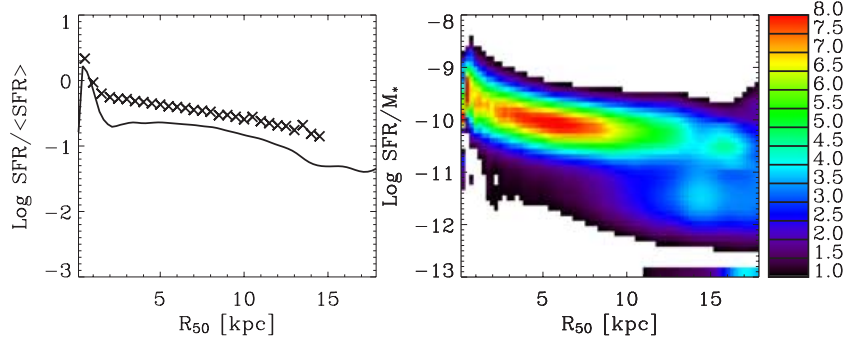
Figure 28. Similar to Fig. 23 but this time showing b as a function of the half-light radius of the galaxies.
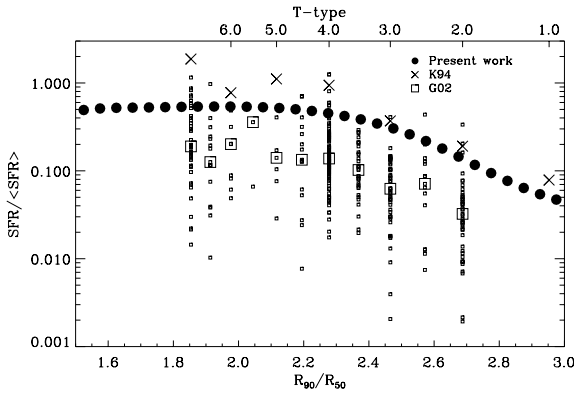

Figure 29. The present to past-average star formation rate for the sample plotted against central concentration in the r-band. The top x-axis shows the mean T-type implied by the concentration according to the recipe described in the text. The crosses show the sample of Kennicutt et al. (1994) and the boxes that of Gavazzi et al. (2002).

table 3 which is indicated on the upper axis. In the same plot we show the results of Kennicutt et al. (1994, their table 4) as crosses and those of Gavazzi et al. (2002) as squares, with big squares corresponding to the median value (BCGs are not shown). The data we show for our study are the unweighted b^g values which should be appropriate for comparison to local studies.

There is clearly a considerable uncertainty in the x-axis, but the general agreement in trends and turn-over between the different samples indicates that the conversion between C and T-type derived by Shimasaku et al. (2001) seems to be fairly reliable. It is also clear that the values from Gavazzi et al. (2002) are considerably lower than ours. This may be due in part to differences in galaxy density in the two samples, but is more likely due to differences in the method used to determine the stellar mass. Gavazzi et al. use an average dark matter mass to stellar mass ratio to derive stellar masses. We have chosen not to attempt to adjust for differences in methodology, as our main interest is in comparing the overall trend.

10 DISCUSSION

The preceding sections have provided a detailed inventory of the star formation activity in the local Universe. One possibly surprising result is that the dominant contribution to the SFR density comes from galaxies with on-going star formation in their inner few kpc (the SF and low S/N SF classes). These galaxies contribute $\sim$72 per cent of the total SFR density, whereas galaxies with signs of AGN

(Composites and AGN) contribute about 15 per cent of the total. The remaining 12 per cent of ρSFR is located in galaxies we are unable to classify based on their fibre spectra. These are almost certainly galaxies with substantial bulges, in which stars are forming in the outer disc. A thorough examination of this question has to be postponed to a different article, but it is clear that the relative importance of star formation in the outer disc is a function of the mass or luminosity of the galaxy. When we look at narrow redshift ranges we find that the median aperture corrections increase by a factor of 2 from stellar masses around 10^9 $M_\odot$ to $M_* \sim 10^{11}$ $M_\odot$. The natural question is how this relates to star formation in bulges as compared to discs, but this will have to await a study of the resolved properties of our galaxies.

We found that galaxies in the SF class have SFR profiles that follow the r-band light. For these galaxies, it is possible, and accurate, to remove aperture bias by simple scaling with the r-band flux (cf. Hopkins et al. 2003).

We have, for the first time, been able to calibrate the conversion factor from Hα to SFR as a function of mass and we show that the standard Kennicutt (1998) conversion factor is a very good average correction for most galaxies in the local Universe. By lucky coincidence, trends in the conversion factor and in the intrinsic Case B ratio of Hα to Hβ tend to cancel each other out. As a result, it will be sufficient to take a fixed Case B ratio and a fixed $\eta_{H\alpha}^0$ for most studies, even though neither of the two assumptions is correct.

One of the main questions we want to address is what excites large-scale star formation in galaxies. Seen from this angle the different parameters we have studied fall into two categories. First there is the incidental set – those parameters which are controlled by star formation activity rather than the other way around. Among these are D4000 and colour. Of more interest are those parameters that may have some direct effect on star formation such as M_*, μ_*, Hubble type and size.

We have found that the stellar mass correlates well with the SFR, a result found previously for a much smaller sample by Gavazzi & Scodeggio (1996). This is quite likely because even though the gas *fractions* of galaxies decline with increasing mass, the total gas masses still increase as a function of mass (Boselli et al. 2001).

Because of this overall correlation, the SFR density distribution functions are rather similar to the mass density distribution functions discussed by Kauffmann et al. (2003a). We find that the majority of star formation takes place in high surface brightness disc-dominated galaxies spanning a range of concentrations. These span a broad range of masses, peaking for galaxies slightly less massive than 10^{11} $M_\odot$ and with half-light radii around 3 kpc. The SF gas has typically solar or slightly super-solar metallicity.

Our most striking result is contained in Fig. 18, which shows that the majority of star formation in the local Universe takes place in massive galaxies. This is a robust prediction of hierarchical models and the shape of our distribution matches that of Springel & Hernquist (2003) (although these authors show ρ_{SFR} versus the halo mass). This clearly warrants a more in-depth comparison and we will follow this up in a forthcoming paper.

Another interesting question is whether star formation in the local Universe is predominantly bursty or relatively quiescent. As we saw in Fig. 24, the answer depends strongly on the mass of the galaxy. It is safe to say, however, that the vast majority of star formation in the local Universe takes place in galaxies that are not undergoing extreme starbursts. Quantitatively, we find that 28 per cent of ρ_{SFR} comes from galaxies with SFR/$\langle$SFR$\rangle > 1$ and 50 per cent at SFR/$\langle$SFR$\rangle > 0.54$, but only 3 per cent of ρ_{SFR} comes from galaxies with SFR/$\langle$SFR$\rangle > 10$. The traditional definition of star bursts is however $b \geqslant 2$–3 and with that definition around 20 per cent of the local SFR density takes place in star bursts. We should caution that our method of inferring the SFR from optical emission lines is likely to miss part of the SFR occurring in extremely dusty FIR-dominated galaxies. Given that such systems are very rare at low redshifts (e.g. Sanders & Mirabel 1996), this does not significantly affect our results.

Although IR-luminous galaxies are rare in the local Universe, most local star formation occurs in galaxies with substantial optical attenuation in their central parts. It is important to realize that this quantity might suffer from aperture biases so the galaxy averaged attenuation could be lower. The relationship between dust attenuation and SFR activity has been discussed by many authors. The most recent discussion is given by Hopkins et al. (2003) for a subsample of our SDSS data. Our results are in good agreement with theirs. Our main result is that the SFR-weighted dust attenuation at Hα is $A_{H\alpha} \sim 1.0 \pm 0.5$ mag, with a good correlation between dust attenuation and SFR. The exact relationship between dust and star formation remains unclear. We see evidence for a strong correlation between stellar mass and dust attenuation, presumably reflecting a correlation between metallicity and dust content. The correlation between SFR and dust may simply reflect a more fundamental correlation between dust and metallicity. This question clearly merits further investigation. Another consequence of the moderately strong dust attenuation present in most galaxies is that the colours of the galaxies that dominate the SFR density have intermediate values – this is why we needed a second colour to get a good constraint on star formation activity in Section 5.

To isolate the influence of other factors that may regulate the star formation, it is clearly necessary to take out the dominant trend with mass. One should probably normalize by the mass of the gas reservoir, but as this information is not available we have normalized by stellar mass. Previous studies by Gavazzi et al. (2002) and Boselli et al. (2001) have shown that SFR/M_* correlates with mass. We confirm this trend. The median value of r_{SFR} is fairly constant over a wide range at low masses and then decreases smoothly at masses greater than $M_* \sim 10^{10}\,M_\odot$. The simplest interpretation is that low-mass galaxies form stars at an approximately constant rate, whereas the star formation history is increasingly peaked towards earlier times for more massive galaxies.

These trends have arguably been seen before. They agree with the 'downsizing' scenario of Cowie et al. (1996). What has not been apparent up to now is the fact that the extremes of the distribution change noticeably. As we go up in mass, there are fewer and fewer galaxies with very high r_{SFR} and a corresponding tail towards low r_{SFR} starts to appear.

It is clear from the previous section that no single factor stands out as controlling r_{SFR}. This is not entirely surprising, as the properties we have used are simple and somewhat limited. It might turn out that more detailed properties correlate better with b, for example lopsidedness (Rudnick, Rix & Kennicutt 2000), local density (Gómez et al. 2003) or merging state (e.g. Le Fèvre et al. 2000). We will study these relations further in future work.

We mentioned previously that the r_{SFR}^V value for the sample as a whole provides a strong constraint on the star formation history of the Universe. This has been discussed in some detail by Boselli et al. (2001), but that study did not use a SFH that is compatible with look-back studies – it assumed an exponential star formation history for all galaxies. Our b^V value with our cosmology corresponds to an effective $\tau = 7^{+0.7}_{-1.5}$ Gyr (including systematic errors) if one considers the whole Universe to have an exponential SFH.

We will however adopt the parametrization of the SFH of the Universe used by Baldry et al. (2002). In general form this reads

$$\text{SFR} \propto \begin{cases} (1+z)^\alpha & \text{for } z \in [z_d, z_c] \\ (1+z)^\beta & \text{for } z < z_d \\ 0 & \text{for } z > z_c, \end{cases} \qquad (13)$$

where we will adopt $z_d = 1$, $z_c = 5$ to be consistent with Baldry et al. (2002) and Glazebrook et al. (2003). Those authors showed that the local 'cosmic spectrum' is well fitted by a model with $\beta = 3$ and $\beta = 0$.

To calculate b^V for the SFH defined by equation (13) we simply integrate over time and define

$$\langle \text{SFR} \rangle = \frac{1}{t_0} \int_0^{t_0} \text{SFR}(t)\,dt, \qquad (14)$$

with t_0 being the age of the Universe at $z = 0.1$. This, together with equation (13), then gives the present to past-average SFR for the model

$$b_{\text{model}}^V = t_0 \frac{\text{SFR}(t_0)}{\int_0^{t_0} \text{SFR}(t)\,dt}. \qquad (15)$$

It is interesting to note that because of the way we measure M_* and the SFR, there is a dependence on the Hubble parameter in our b value calculations. That dependence cancels out in equation (13). To constrain α and β we thus need a prior on H_0. We will take the simple approach and choose $h = 0.7$. We use the constraint on b^V given in equation (12).

Fig. 30 shows the resulting constraints on on α and β. Because the b^V value only constrains the $z = 0.1$ value of the SFR and the integral under the curve, there is a strong degeneracy even with a fixed functional form. This degeneracy cannot be lifted without independent constraints, but it is still of considerable interest. Most noticeably it is clear that any really strong evolution, $\beta > 4$, at $z < 1$ is excluded by the current data.

The dotted lines in Fig. 30 show the effect of systematic uncertainties on these constraints (for simplicity we have symmetrized these). The dashed lines show the purely statistical errors (1σ). It is clear that if we could remove all systematic uncertainties, the method would give a very powerful constraint on the star formation history of the Universe. It is therefore worth asking what is required to further reduce the systematic uncertainties.

A thorough test of our aperture corrections using resolved spectroscopy would substantially reduce the systematic uncertainty in our SFR estimates. To reduce uncertainties due to evolution, one could limit oneself to a narrow range in redshift, but that would mean a corresponding increase in the uncertainty in extrapolating
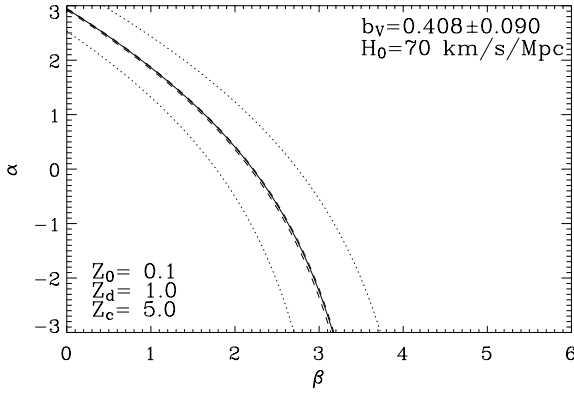
Figure 30. The constraint on α and β (see equation 13) from the b^V value calculated for our sample. The solid line shows the locus of $b^V = 0.400$, whereas the dashed line shows the 68 per cent confidence limits from the likelihood distributions and the dotted the locus of our conservative systematic error estimates.

to faint galaxies. Alternatively, it might be possible to directly constrain the evolution inside the sample, but that is a rather delicate operation due to the complex aperture effects. We have postponed this to a later paper.

Systematic uncertainties notwithstanding, the results presented here provide remarkably good constraints on the cosmic star formation history using only data at $z < 0.22$. It is clear that a similar study carried out at higher redshift will be of great interest and provide a very interesting complement to recent studies of the mass-assembly of the Universe (Brinchmann & Ellis 2000; Drory et al. 2001; Dickinson et al. 2003; Fontana et al. 2003; Rudnick et al. 2003).

In general it is our firm belief that future studies of high-redshift data should try as far as possible to study the evolution of the full distribution functions in question rather than summarizing the distributions to a single value. As we have seen, important changes might occur in the tails of the distributions with little influence on the median value. These physically interesting details can only be extracted through the use of distribution functions. It has been standard to study luminosity functions. With the techniques introduced in the present paper it is now possible to extend the analysis to a wide variety of distribution functions. This has been possible due to the unprecedented data set provided by the SDSS survey.

REFERENCES

Abazajian K. et al., 2003, AJ, 126, 2081
Abazajian K. et al., 2004, AJ, accepted
Abraham R. G., Valdes F., Yee H. K. C., van den Bergh S., 1994, ApJ, 432, 75
Afonso J., Cram L., Mobasher B., 2000, ApJ, 536, 68
Baldry I. K. et al., 2002, ApJ, 569, 582
Baldry I. K., Glazebrook K., Brinkmann J., Ivezic Z., Lupton R. H., Nichol R. C., Szalay A. S., 2004, ApJ, 600, 681
Baldwin J. A., Phillips M. M., Terlevich R., 1981, PASP, 93, 5
Balogh M. L., Christlein D., Zabludoff A. I., Zaritsky D., 2001, ApJ, 557, 117
Balogh M. L., Schade D., Morris S. L., Yee H. K. C., Carlberg R. G., Ellingson E., 1998, ApJ, 504, L75
Bell E. F., Kennicutt R. C., 2001, ApJ, 548, 681
Blanton M. R. et al., 2003a, ApJ, 592, 819
Blanton M. R. et al., 2003b, ApJ, 594, 186
Blanton M. R., Lin H., Lupton R. H., Maley F. M., Young N., Zehavi I., Loveday J., 2003, AJ, 125, 2276
Boselli A., Gavazzi G., Donas J., Scodeggio M., 2001, AJ, 121, 753
Bresolin F., Kennicutt R. C., 2002, ApJ, 572, 838
Bressan A., Silva L., Granato G. L., 2002, A&A, 392, 377
Brinchmann J. et al., 1998, ApJ, 499, 112
Brinchmann J., Ellis R. S., 2000, ApJ, 536, L77
Bruzual A. G., Charlot S., 1993, ApJ, 405, 538
Bruzual G., Charlot S., 2003, MNRAS, 344, 1000
Calzetti D., 1997, AJ, 113, 162
Charlot S., Fall S. M., 2000, ApJ, 539, 718
Charlot S., Longhetti M., 2001, MNRAS, 323, 887
Charlot S., Kauffmann G., Longhetti M., Tresse L., White S. D. M., Maddox S. J., Fall S. M., 2002, MNRAS, 330, 876
Ciardullo R., Feldmeier J. J., Jacoby G. H., Kuzio de Naray R., Laychak M. B., Durrell P. R., 2002, ApJ, 577, 31
Clarke C., Oey M. S., 2002, MNRAS, 337, 1299
Colless M. et al., 2001, MNRAS, 328, 1039
Condon J. J., Cotton W. D., Broderick J. J., 2002, AJ, 124, 675
Cowie L. L., 1988, in Kaiser N., Lasenby A. N., eds, Post-Recombination Universe Protogalaxies. Kluwer, Dordrecht, p. 1
Cowie L. L., Songaila A., Hu E. M., Cohen J. G., 1996, AJ, 112, 839
Dickinson M., Papovich C., Ferguson H. C., Budavári T., 2003, ApJ, 587, 25
Doi M., Fukugita M., Okamura S., 1993, MNRAS, 264, 832
Drory N., Bender R., Snigula J., Feulner G., Hopp U., Maraston C., Hill G. J., de Oliveira C. M., 2001, ApJ, 562, L111
Ellis R. S., 1997, ARA&A, 35, 389
Felten J. E., 1976, ApJ, 207, 700
Ferland G., 1996, Hazy: A Brief Introduction to CLOUDY. Internal Report, Univ. Kentucky
Flores H. et al., 1999, ApJ, 517, 148
Fontana A. et al., 2003, ApJ, 594, L9
Fukugita M., Kawasaki M., 2003, MNRAS, 340, L7
Fukugita M., Ichikawa T., Gunn J. E., Doi M., Shimasaku K., Schneider D. P., 1996, AJ, 111
Gallego J., Zamorano J., Aragon-Salamanca A., Rego M., 1995, ApJ, 455, L1
Gavazzi G., Scodeggio M., 1996, A&A, 312, L29

Gavazzi G., Boselli A., Pedotti P., Gallazzi A., Carrasco L., 2002, A&A, 396, 449

Giavalisco M. et al., 2004, ApJ, 600, L103

Glazebrook K., Blake C., Economou F., Lilly S., Colless M., 1999, MNRAS, 306, 843

Glazebrook K. et al., 2003, ApJ, 587, 55

Gómez P. L. et al., 2003, ApJ, 584, 210

Gunn J. E. et al., 1998, AJ, 116, 3040

Guzman R., Gallego J., Koo D. C., Phillips A. C., Lowenthal J. D., Faber S. M., Illingworth G. D., Vogt N. P., 1997, ApJ, 489, 559

Haarsma D. B., Partridge R. B., Windhorst R. A., Richards E. A., 2000, ApJ, 544, 641

Heckman T. M., Sembach K. R., Meurer G. R., Leitherer C., Calzetti D., Martin C. L., 2001, ApJ, 558, 56

Hernandez X., Gilmore G., Valls-Gabaud D., 2000, MNRAS, 317, 831

Hogg D. W., 2001, PASP, submitted (astro-ph/0105280)

Hogg D. W., Finkbeiner D. P., Schlegel D. J., Gunn J. E., 2001, AJ, 122, 2129

Hogg D. W. et al., 2003, ApJ, 585, L5

Hoopes C. G., Walterbos R. A. M., 2003, ApJ, 586, 902

Hopkins A. M., Connolly A. J., Haarsma D. B., Cram L. E., 2001a, AJ, 122, 288

Hopkins A. M., Irwin M. J., Connolly A. J., 2001b, ApJ, 558, L31

Hopkins A. et al., 2003, ApJ, 599, 971

Kauffmann G. et al., 2003a, MNRAS, 341, 33

Kauffmann G. et al., 2003b, MNRAS, 341, 54

Kauffmann G. et al., 2003c, MNRAS, 346, 1055

Kennicutt R. C., 1983, ApJ, 272, 54

Kennicutt R. C., 1998, ARA&A, 36, 189

Kennicutt R. C., Tamblyn P., Congdon C. E., 1994, ApJ, 435, 22

Kennicutt R. C., Bresolin F., French H., Martin P., 2000, ApJ, 537, 589

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

Kewley L. J., Dopita M. A., Sutherland R. S., Heisler C. A., Trevena J., 2001a, ApJ, 556, 121

Kewley L. J., Heisler C. A., Dopita M. A., Lumsden S., 2001b, ApJS, 132, 37

Kochanek C. S., Pahre M. A., Falco E. E., 2001, ApJ, submitted (astro-ph/0011458),

Kroupa P., 2001, MNRAS, 322, 231

Lanzetta K. M., Yahata N., Pascarelle S., Chen H., Fernández-Soto A., 2002, ApJ, 570, 492

Lawson C. L., Hanson R. J., 1974, Solving Least Squares Problems. Prentice-Hall, New Jersey

Le Fèvre O. et al., 2000, MNRAS, 311, 565

Lehnert M. D., Heckman T. M., 1994, ApJ, 426, L27

Leitherer C., Ferguson H. C., Heckman T. M., Lowenthal J. D., 1995, ApJ, 454, L19

Lilly S. J., Le Fevre O., Hammer F., Crampton D., 1996, ApJ, 460, L1

Lupton R., Gunn J. E., Ivezić Z., Knapp G. R., Kent S., Yasuda N., 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

MacLow M. M., Klessen R. S., 2004, Rev. Mod. Phys., 76, 125

Madau P., Ferguson H. C., Dickinson M. E., Giavalisco M., Steidel C. C., Fruchter A., 1996, MNRAS, 283, 1388

Madau P., Pozzetti L., Dickinson M., 1998, ApJ, 498, 106

Martin C. L., Kennicutt R. C., 2001, ApJ, 555, 301

Mathis J. S., 2000, ApJ, 544, 347

Miller C. J., Nichol R. C., Gomez P., Hopkins A., Bernardi M., 2003, ApJ, 597, 142

Morgan W. W., 1958, PASP, 70, 364

Moy E., Rocca-Volmerange B., Fioc M., 2001, A&A, 365, 347

Nakamura O., Fukugita M. et al., 2003, ApJ, submitted

Oey M. S., Dopita M. A., Shields J. C., Smith R. C., 2000, ApJS, 128, 511

Osterbrock D. E., 1989, Astrophysics of Gaseous Nebulae and Active Galactic Nuclei. Univ. Sci. Books, New York

Pérez-González P. G., Zamorano J., Gallego J., Aragón-Salamanca A., Gil de Paz A., 2003, ApJ, 591, 827

Panuzzo P., Bressan A., Granato G. L., Silva L., Danese L., 2003, A&A, 409, 99

Pascual S., Gallego J., Aragón-Salamanca A., Zamorano J., 2001, A&A, 379, 798

Pier J. R., Munn J. A., Hindsley R. B., Hennessy G. S., Kent S. M., Lupton R. H., Ivezić Ž., 2003, AJ, 125, 1559

Rice J. A., 1995, Mathematical Statistics and Data Analysis, 2nd edn. Duxbury Press, Pacific Grove, CA

Rocha-Pinto H. J., Scalo J., Maciel W. J., Flynn C., 2000, A&A, 358, 869

Rola C. S., Terlevich E., Terlevich R. J., 1997, MNRAS, 289, 419

Rudnick G., Rix H., Kennicutt R. C., 2000, ApJ, 538, 569

Rudnick G. et al., 2003, ApJ, 599, 847

Salpeter E. E., 1955, ApJ, 121, 161

Sanders D. B., Mirabel I. F., 1996, ARA&A, 34, 749

Scalo J. M., 1986, Fundamentals of Cosmic Physics, 11, 1

Schechter P., 1976, ApJ, 203, 297

Seaton M. J., 1979, MNRAS, 187, 73

Serjeant S., Gruppioni C., Oliver S., 2002, MNRAS, 330, 621

Shimasaku K. et al., 2001, AJ, 122, 1238

Smith J. A. et al., 2002, AJ, 123, 2121

Songaila A., Cowie L. L., Lilly S. J., 1990, ApJ, 348, 371

Springel V., Hernquist L., 2003, MNRAS, 339, 312

Stoughton C. et al., 2002, AJ, 123, 485

Strateva I. et al., 2001, AJ, 122, 1861

Strauss M. A. et al., 2002, AJ, 124, 1810

Sullivan M., Treyer M. A., Ellis R. S., Bridges T. J., Milliard B., Donas J., 2000, MNRAS, 312, 442

Sullivan M., Mobasher B., Chan B., Cram L., Ellis R., Treyer M., Hopkins A., 2001, ApJ, 558, 72

Tremonti C. et al., 2004, ApJ, submitted

Tresse L., Maddox S. J., 1998, ApJ, 495, 691

Tresse L., Rola C., Hammer F., Stasinska G., Le Fevre O., Lilly S. J., Crampton D., 1996, MNRAS, 281, 847

Tresse L., Maddox S. J., Le Fèvre O., Cuby J.-G., 2002, MNRAS, 337, 369

Wang J., Heckman T. M., Lehnert M. D., 1997, ApJ, 491, 114

Witt A. N., Gordon K. D., 2000, ApJ, 528, 799

Wong T., Blitz L., 2002, ApJ, 569, 157

Yang X., Mo H. J., van den Bosch F. C., 2003, MNRAS, 339, 1057

York D. G. et al., 2000, AJ, 120, 1579

Young J. S. et al., 1995, ApJS, 98, 219

Zackrisson E., Bergvall N., Olofsson K., Siebert A., 2001, A&A, 375, 814

APPENDIX A: MANIPULATION OF LIKELIHOOD DISTRIBUTIONS

Throughout this paper we use the full likelihoods of all parameters. This means that in many situations it is necessary to combine likelihood distributions. For the benefit of the reader we summarize some of the most important results here. We will use standard notation and write e.g. $f_Z(z)$ for the likelihood of a value z in the likelihood distribution Z. Thus we represent the addition of two likelihood distributions as $Z = X + Y$. This sum is discussed in most introductory statistics textbooks and the well known result is that

$$f_Z(z) = \sum_i f_{XY}(x_i, z - x_i) \stackrel{\text{indep}}{=} \sum_i f_X(x_i) f_Y(z - x_i), \qquad \text{(A1)}$$

where the latter equality is valid in the case of independent variables.

In the present work we do not often use the straight addition of distributions. The general case is not normally given in statistics textbooks. It is however easy to derive. We will focus our attention on the cumulative distribution function (CDF) $F(z \leqslant \hat{z})$. In the continuous case this is given as

$$F(z \leqslant \hat{z}) = \int_{\mathcal{R}} \int f_{XY}(x, y)\, dx\, dy, \qquad \text{(A2)}$$

where $\mathcal{R}$ is the region where the constraint $z \leqslant \hat{z}$ is fulfilled. If we write the general composition of two likelihood distributions as $Z = g(X, Y)$ then the constraint $z \leqslant \hat{z}$ translates into $z \leqslant g(x, y)$. For well-behaved distributions we can invert this to write $y \leqslant h(x, z)$. In general we can then write

$$F(z \leqslant \hat{z}) = \int_{-\infty}^{\infty} \int_{-\infty}^{h(x,z)} f_{XY}(x, y)\, \mathrm{d}x\, \mathrm{d}y. \tag{A3}$$

A simple change of variable presents itself and by setting $y = h(x, u)$ we get

$$F(z \leqslant \hat{z}) = \int_{-\infty}^{\infty} \int_{-\infty}^{z} f_{XY}(x, h(x, u))J(h)\, \mathrm{d}x\, \mathrm{d}u, \tag{A4}$$

with $J = \partial h / \partial u$ being the Jacobian of the variable transformation. By taking the derivative of this we can get the likelihood distribution $f_Z(z)$:

$$f_Z(z) = \frac{\partial F}{\partial z} = \int_{-\infty}^{\infty} f_{XY}(x, h(x, z))|J(h)|\, \mathrm{d}x, \tag{A5}$$

which can be evaluated for any given $g(X, Y)$. The extension to discrete distribution functions is straightforward (see below).

As an example we can take the addition of two likelihood distributions for the SFR which we need for the aperture corrections. Here we have

$$Z = \log_{10}(10^X + 10^Y), \tag{A6}$$

as the likelihood distributions are in log space. Thus calculating the Jacobian and writing it as a discrete distribution

$$f_Z(z_j) = \sum_i f_X(x_i) f_Y\left(\log_{10}(10^{z_j} - 10^{x_i})\right) \frac{10^{z_j}}{10^{z_j} - 10^{x_i}}. \tag{A7}$$

This formalism extends fairly straightforwardly to the iterated composition of many (independent) likelihood distributions

$$f_Z(z) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f_{X_1}(x_1) f_{X_2}[h_2(x_1, x_2)] \dots$$
$$f_{X_N}[h_N(x_{N-1}, x_N)] \left[\prod_{i=2}^{N} \left| \frac{\partial h_i}{\partial x_i} \right| \right] \mathrm{d}^N x_i, \tag{A8}$$

but this is not a practical approach for arbitrary likelihood distributions due to combinatorial explosion (the sum/integral has to be done over all combinations of variables that combine to a particular value of z). Instead it is easier in this situation to calculate the combined likelihood distribution in an approximate manner by carrying out a Monte Carlo sampling of the distributions. In this case we first form the CDF, $F(x)$, for each galaxy. We then draw, for each galaxy, a random number, p, between 0 and 1 and solve $F(x) = p$ numerically using linear interpolation.

We use the latter approach for instance when evaluating the star formation density of the local Universe. The main disadvantage of a Monte Carlo approach is that one needs to carry out a large number of repetitions to ensure a sufficient sampling of the final likelihood distribution and to estimate the confidence intervals with good accuracy.

Before continuing, we would like to emphasize a technical issue which is of considerable importance for some combinations of likelihood distributions: ignoring the likelihood distribution of a particular variable (in effect setting its error to zero) will in general *bias* the estimate of any combination of this variable with any other unless they are both Gaussians. A specific example relevant for our work is the ratio of two variables whose likelihood distributions are kept in log space.

In particular, the calculation of r_{SFR} in Section 8 falls into this category. Here we have $Z = 10^Y / 10^X$ where X is the likelihood distribution of the stellar mass and Y that of the SFR. If we now for illustrative purposes assume that these are Gaussian in *log* space: $\log \mathrm{SFR} \sim N(\log \mathrm{SFR}_0, \sigma_{\mathrm{SFR}})$ and likewise for $\log M_*$ then a straightforward calculation gives us that the mean of Z

$$\langle z \rangle = \int_0^{\infty} z f_Z(z)\, \mathrm{d}z, \tag{A9}$$

can be written

$$\langle z \rangle = e^{\log \mathrm{SFR}/M_*} + \left(\sigma_{M_*}^2 + \sigma_{\mathrm{SFR}}^2 \right)/2 \ln 10. \tag{A10}$$

Clearly setting $\sigma_{M_*} = 0$ would bias the average value for the specific SFR in this situation. The exact amount depends on the details of the likelihood distribution and the bias could vanish entirely. In practice (as mentioned in Section 8) we find that for the calculation of the specific SFR, ignoring the uncertainty in M_* would bias our results about 15 per cent high.

Finally we would like to mention the various ways one can summarize a distribution and what we typically use (see also Kauffmann et al. 2003a). The average is the simplest, as it is calculated as

$$\langle x \rangle = \sum_i x_i f_X(x_i), \tag{A11}$$

because f_X is normalized to a total of one. The mode is equally simple:

$$x_{\mathrm{mode}} = x_i \text{ such that } \max_j f_X(x_j) = f_X(x_i). \tag{A12}$$

In the event of a tie we use the smallest value satisfying the criterion. These two, but in particular the mode, are affected by the binning chosen for the likelihood distributions.

More commonly used in the present paper are the percentiles of the distributions, defined in the standard way as

$$F_X(x_p) = p, \tag{A13}$$

where F_X is the CDF and p is the percentile – $p = 0.5$ gives the median, $p = 0.16$ and $p = 0.84$ give the upper and lower '1σ' limits chosen, respectively, and x_p is the value to solve for.

To avoid severe binning effects (in particular for the extremes of the distributions) we calculate these percentiles by solving equation (A13) using linear interpolation between grid points. In this situation we interpolate the CDF and this is a robust procedure. For the average a straight application of equation (A11) would require interpolation of the probability density function (PDF) which might depend sensitively on the interpolation scheme. Instead we calculate the average by interpolating the CDF and drawing 10^4 realizations of this for each galaxy and calculating the average from this. We have experimented with different numbers of realizations and interpolation schemes and found that the estimate of the average so calculated is robust to these changes within the accuracy needed for our project.

APPENDIX B: MINOR CONTRIBUTIONS TO THE EMISSION LINE SPECTRA

We will here discuss, briefly, the contribution of SNRs and PNe to the overall emission line flux of galaxies. The SNRs are interesting in their own right, but due to their short lifetime and the fact that they co-exist with star-forming regions they are unlikely to affect the line radiation significantly. This is borne out in studies of resolved H II regions in the LMC (Oey et al. 2000), which show that SNRs only affect the emission line spectra at a fairly low level.

As discussed by CL01, the contribution of PNe to the line radiation can be ignored in galaxies with active star formation, however, in the current work we have galaxies with very low Hα luminosities and it is therefore necessary to ask if these fluxes can come from PNe. It is straightforward to assess this, as the PNe O III luminosity function seems to be universal (e.g. Ciardullo et al. 2002 and references therein). We can therefore use this to determine the O III flux from the PNe population in a given galaxy. Using the PNe LF in Ciardullo et al. (2002), we find that the maximum contribution to the total O III flux of the sample by PNe is of the order of 1.5 per cent. This is in general totally negligible in comparison with other sources of uncertainty, but it is worth noting that we do predict that among galaxies with very weak emission lines, S/N < 5 in O III, around 20 per cent of the galaxies could have PNe contributing more than 20 per cent of the O III flux thereby complicating studies of galaxies with very weak emission lines. Because the O III/Hα ratio in PNe is in general considerably larger than for star-forming galaxies, the effect on Hα, and hence the star formation estimates, is therefore entirely negligible except at S/N < 1 in Hα; it has no effect whatsoever on our results here.

A significant amount, typically 30–50 per cent, of the emission line flux in galaxies comes from the DIG (Lehnert & Heckman 1994; Wang et al. 1997; Hoopes & Walterbos 2003). This is implicitly taken into account by the CL01 models because, as CL01 show, the combination of the (low-ionization) DIG and the (higher ionization) H II regions is typically represented by a lower effective ionization parameter (see also Mathis 2000). Although we do not explicitly incorporate leaky H II regions as done by e.g. Hoopes & Walterbos (2003), we can produce similar line ratios by changing ξ. Thus the models do not need any extra component to take into account DIG, and as long as the DIG is heated by OB stars as seems necessary, its presence should not bias our SFR estimates. The interpretation of the fitted ionization parameter and dust-to-metal ratio is more complex and we will discuss this issue in more detail in an upcoming paper.

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