Radio continuum size evolution of star-forming galaxies over 0.35 < z < 2.25

E. F. Jiménez-Andrade1,2, B. Magnelli1, A. Karim1, G. Zamorani3, M. Bondi4, E. Schinnerer5, M. Sargent6, E. Romano-Díaz1, M. Novak5,7, P. Lang3, F. Bertoldi1, E. Vardoulaki1, S. Toft8, V. Smolčić7, K. Harrington1,2, S. Leslie5, J. Delhaize7,9, D. Liu5, C. Karoumpis1,2, J. Kartaltepe10, and A. M. Koekemoer11

1 Argelander Institut für Astronomie, Universität Bonn, Auf dem Hügel 71, 53121 Bonn, Germany
e-mail: ericja@astro.uni-bonn.de
2 International Max Planck Research School of Astronomy and Astrophysics at the Universities of Bonn and Cologne, Bonn, Germany
3 INAF – Osservatorio di Astrofisica e Scienza dello Spazio di Bologna, Via Gobetti 93/3, 40129 Bologna, Italy
4 INAF – Istituto di Radioastronomia, Via Gobetti 101, 40129 Bologna, Italy
5 Max Planck Institute for Astronomy, Königstuhl 17, 69117 Heidelberg, Germany
6 Astronomy Centre, Department of Physics and Astronomy, University of Sussex, Brighton BN1 9QH, UK
7 Department of Physics, University of Zagreb, Bijenička cesta 32, 10002 Zagreb, Croatia
8 Cosmetic Dawn Center (DAWN), Niels Bohr Institute, University of Copenhagen, Lyngbyvej 2, Copenhagen 2100, Denmark
9 Department of Astronomy, University of Cape Town, Private Bag X3, Rondebosch 7701, South Africa
10 School of Physics and Astronomy, Rochester Institute of Technology, 84 Lomb Memorial Drive, Rochester, NY 14623, USA
11 Space Telescope Science Institute, 3700 San Martin Dr., Baltimore, MD 21218, USA

Received 31 January 2019 / Accepted 28 March 2019

ABSTRACT

To better constrain the physical mechanisms driving star formation, we present the first systematic study of the radio continuum size evolution of star-forming galaxies (SFGs) over the redshift range 0.35 < z < 2.25. We use the VLA COSMOS 3 GHz map (noise rms = 2.3 μJy beam−1, θbeam = 0.75 arcsec) to construct a mass-complete sample of 3184 radio-selected SFGs that reside on and above the main sequence (MS) of SFGs. We constrain the overall extent of star formation activity in galaxies by applying a 2D Gaussian model to their radio continuum emission. Extensive Monte Carlo simulations are used to validate the robustness of our measurements and characterize the selection function. We find no clear dependence between the radio size and stellar mass, M∗, of SFGs with 10.5 ≤ log(M∗/M⊙) ≤ 11.5. Our analysis extends Effective radius of SFGs on (above) the MS to Reff = 1.5 ± 0.2 (1.0 ± 0.2) kpc near the MS. The size of the stellar component of galaxies is larger than the extent of the radio continuum emission by a factor \( \frac{1}{2} \approx 2 \) at \( z = 0.5 \), indicating star formation is enhanced at small radii. The galactic-averaged star formation rate surface density \( \Sigma_{\text{ff}} \) scales with the distance to the MS, except for a fraction of MS galaxies (≈10%) that harbor starburst-like \( \Sigma_{\text{ff}} \). These “hidden” starbursts might have experienced a compaction phase due to disk instability and/or a merger-driven burst of star formation, which may or may not significantly offset a galaxy from the MS. We thus propose to use \( \Sigma_{\text{ff}} \) and distance to the MS in conjunction to better identify the galaxy population undergoing a starbursting phase.

Key words. galaxies: evolution – galaxies: high-redshift – galaxies: structure – galaxies: starburst – radio continuum: galaxies

1 Introduction

Most galaxies follow a tight correlation in the star formation rate (SFR) – stellar mass (M∗) plane known as the main sequence (MS) of star-forming galaxies (SFGs; e.g., Brinchmann et al. 2004; Noeske et al. 2007; Elbaz et al. 2007; Salim et al. 2007; Daddi et al. 2007; Pannella et al. 2009, 2015; Magdis et al. 2010; Peng et al. 2010; González et al. 2010; Rodighiero et al. 2011; Karim et al. 2011; Wuyts et al. 2011; Bouwens et al. 2012; Whitaker et al. 2012, 2014; Rodighiero et al. 2014; Renzini & Peng 2015; Schreiber et al. 2015, 2017). This relation holds over ~90% of the cosmic history of the Universe (e.g., Stark et al. 2013; González et al. 2014; Steinhardt et al. 2014; Salmon et al. 2015) and has a slope and normalization that increase with redshift, while its dispersion of only 0.3 dex remains nearly constant throughout cosmic time (see Speagle et al. 2014; Pearson et al. 2018, and references therein).

Although most galaxies have an implied SFR that scatters within a factor two around the MS, some do show a significantly higher SFR. Those objects also exhibit a higher gas content, shorter gas depletion times (e.g., Genzel et al. 2015; Schinnerer et al. 2016; Tacconi et al. 2013, 2018), and higher dust temperatures (e.g., Magnelli et al. 2014). Likewise, the stellar-light radial distribution is different in these two galaxy populations; while MS galaxies are closely approximated by exponential disks (e.g., Bremer et al. 2018), those above (and below) it exhibit a higher central mass concentration
A crucial parameter for verifying these scenarios is the size of a galaxy. Recent studies have explored the structural properties of SFGs by mapping their stellar component (e.g., van der Wel et al. 2014; Shibuya et al. 2015; Mowla et al. 2018). However, the size of the overall star-forming component has been poorly explored. This is partially due to observationally expensive high-resolution infrared (IR)/radio observations, which have been limited to relatively small samples of SFGs (e.g., Rujopakarn et al. 2016; Miettinen et al. 2017; Murphy et al. 2017; Elbaz et al. 2018). While large and representative samples of SFGs can be obtained from ultraviolet (UV)/optical observations, they are affected by dust extinction, rendering size measurement difficult (e.g., Elbaz et al. 2011; Nelson et al. 2016a). To better understand the mechanisms that regulate star formation in galaxies we need a statistically significant, mass-complete sample of radio-selected SFGs over cosmic time, and a dust-extinction-corrected photometric redshift.

A114, page 2 of 20
infrared data from the Spitzer Large Area Survey (SPLASH) within the Hyper-Suprime-Cam Spitzer legacy program.

Photometric redshifts were derived with LePhare (Arnouts et al. 2002; Ilbert et al. 2006) using a set of 31 templates of spiral and elliptical galaxies from Polletta et al. (2007), and 12 templates of young blue SFGs using the Bruzual & Charlot (2003) models. Through a comparison with spectroscopic redshift samples in the COSMOS field, Laigle et al. (2016) derived a photometric redshift precision of $\sigma_{z} / (1 + z_{0}) = 0.007$ and a small catastrophic failure fraction of $\eta = 0.5\%$ for $z_{0} < 3$.

Stellar masses were also derived with LePhare using a library of synthetic spectra from the Stellar Population Synthesis model of Bruzual & Charlot (2003). A Chabrier (2003) initial mass function, an exponentially declining and delayed star formation history (SFH) and solar/half-solar metallicities were considered. The stellar masses used here correspond to the median of the inferred probability distribution function. A 90% completeness limit of $10^{3.5} (10^{10} M_{\odot})$ was achieved up to $z = 0.35 (2.25)$.

3. Data analysis

We measure the size and flux density of radio sources directly from the VLA COSMOS 3 GHz mosaic, i.e., in the image plane, and further revise those estimates using extensive Monte Carlo simulations. While these sizes and fluxes could also be estimated in the $uv$-plane, this is impractical due to the large data volume of the VLA COSMOS 3 GHz survey.

3.1. Source extraction

The advent of large radio astronomical surveys has stimulated the development of robust source extraction algorithms such as bloscat (Hales et al. 2012) and PyBDSF (Mohan & Rafferty 2015). Here, we use PyBDSF as it provides parametric information of the source morphology such as the deconvolved major axis FWHM ($\theta_{M}$), that is,

$$\theta_{M} = \left( \frac{(\theta_{\text{F}})^{2} - (\theta_{\text{beam}})^{2}}{2} \right)^{1/2},$$

where $\theta_{\text{beam}}$ is the FWHM of the synthesized beam of the VLA COSMOS 3 GHz map (0.75 arcsec) and $\theta_{\text{syn}}$ the observed/convolved major axis FWHM.

The PyBDSF algorithm characterizes the radio source properties as follows. First, it identifies peaks of emission above a given threshold (thresh_pix) that are surrounded by contiguous pixels, i.e., islands, with emission greater than a minimum value (thresh_isl). Second, it fits multiple Gaussians to each island depending on the number of the peaks identified within the VLA COSMOS 3 GHz mosaic ($0.75$ arcsec) and $\theta_{\text{F}}$ the convolved major axis FWHM.

PyBDSF characterizes the radio source properties as follows. First, it identifies peaks of emission above a given threshold (thresh_pix) that are surrounded by contiguous pixels, i.e., islands, with emission greater than a minimum value (thresh_isl). Second, it fits multiple Gaussians to each island depending on the number of the peaks identified within the VLA COSMOS 3 GHz mosaic ($0.75$ arcsec) and $\theta_{\text{F}}$ the convolved major axis FWHM. PyBDSF then associates with an active galactic nucleus (AGN), and not star formation, we rely on the results from Smolčić et al. (2017b). They characterized the host galaxy of radio sources in the VLA COSMOS 3 GHz map by identifying their optical/near-IR/mid-IR counterparts from the $i$-band selected catalog (optical; Capak et al. 2007), the COSMOS2015 catalog (near-IR; Laigle et al. 2016), and the Spitzer COSMOS (S-COSMOS) Infrared Array Camera (3.6 µm-selected, IRAC; Sanders et al. 2007). Based on this multiwavelength counterpart association, a sample of AGN and SFGs was assembled.

AGN host galaxies were identified as such, and excluded from our sample, if the following criteria were met:

- the intrinsic $[0.5–8]$ keV X-ray luminosity is greater than $L_{X} = 10^{42}$ erg s$^{-1}$ (e.g., Szokoly et al. 2004);
- the flux throughout the four IRAC bands (3.6, 4.5, 5.8, and 8) displays a monotonic rise and follows the criterion proposed by Donley et al. (2012);
- an AGN component significantly improves the fitting of their optical to millimeter spectral energy distribution (SED, as in Da Cunha et al. 2008; Berta et al. 2013; Delvecchio et al. 2014, 2017);
- $M_{\text{SMDUV}} - M_{r}$, i.e., rest-frame near-ultraviolet (NUV) minus $r+$ band, is greater than 3.5 (Ilbert et al. 2010);
- the observed radio emission $L_{\text{1.4GHz}}$ exceeds that expected from the host galaxy $S_{\nu,\text{IR}}$ (estimated via IR SED fitting, Delvecchio et al. 2017).

Excluding AGN hosts through all these criteria yields a highly clean sample of SFGs. Within the redshift range probed in this work ($0.35 < z < 2.25$), we find that 4216 galaxies match with our catalog of 9223 radio sources and have available stellar mass estimates in the COSMOS2015 catalog. While most of them (3248, i.e., 77%) are classified as SFGs, 968 galaxies (23%) exhibit one or more of the above-mentioned signatures of AGN activity. Since comparing the radio size evolution of AGN and SFGs is beyond the scope of this work, we refer the reader to Bondi et al. (2018) who presented a similar analysis using the VLA COSMOS 3 GHz map and following the same AGN-SFGs classification scheme used here.

We note that out of the 3248 radio-selected SFGs, 64 (2%) of them are fitted with multiple Gaussians by PyBDSF, suggesting a more complex and/or extended morphology. Since modeling such systems in our Monte Carlo simulations (Sect. 3.3) is challenging, we exclude them from the analysis. We verified, however, that none of the relations/results reported thereafter are affected, within the uncertainty, by the inclusion of these multi-component sources. Our final sample, therefore, comprises 3184 radio-selected SFGs over the redshift range $0.35 < z < 2.25$, in which a mass-complete sample of log($M_{\ast}/M_{\odot}$) $\geq 10.5$ SFGs can be assembled (Sect. 3.5).

3.2. AGN rejection

To identify galaxies in which the radio continuum emission is associated with an active galactic nucleus (AGN), and not star formation, we use these matched sources to enhance the pureness of our radio source catalog.

3.3. Accuracies and limitations of our size and flux density measurements

In this section, we describe the Monte Carlo (MC) simulations used to characterize the biases associated with size and flux determination of SFGs in the sample. This approach is based...
on the injection of mock sources, following a realistic flux and size distribution, into noise maps that accurately represent the original dataset (e.g., Casey et al. 2014, Sect. 3.2). After retrieving these sources from the maps with PyBDSF, we compare the input and output properties and hence address these particular questions: (a) what are the minimum/maximum source sizes we can detect in the VLA COSMOS 3 GHz mosaic at a given flux density? and (b) how reliable are our measurements for a given intrinsic flux density and FWHM?

These MC simulations require a mock sample that follows the intrinsic, yet unknown, flux density ($S_{\text{int}}$) and angular size ($\theta_M$) distributions of SFGs. For this purpose, we use previous constraints on the $\mu$Jy radio source population as presented in Smolčić et al. (2017a). First, we approximate the observed flux density distribution of this mock sample with a single power-law model ($N \propto S_{\text{int}}^{-0.8}$). Second, we assume that their angular size is linked to their total flux density (Windhorst et al. 1990; Richards 2000) as $\theta_{\text{median}}[\text{arcsec}] = 1.8S_{\text{int}}^{0.6} [\text{mJy}]$ (Bondi et al. 2003; Smolčić et al. 2017a).

The input sample comprises \( \sim 7 \times 10^5 \) sources modeled with a single Gaussian component. We explore the parameter space where $\theta_M^{\text{in}}$ ranges from 0.03–12 arcsec (with ellipticity $e = 0.25$, 0.5, 0.75, and 1 and random position angle) and $10^{-5}$ Jy < $S_{\text{int}}$ < $10^{-1.5}$ Jy, which is the observed range of retrieved VLA-COSMOS 3 GHz galaxies (see Fig. 1). These mock galaxies were convolved with the synthesized beam and randomly injected into the mosaic in purely noise dominated regions, i.e., those areas where no original source is found within 36 x 36 arcsec$^2$. They were subsequently retrieved with PyBDSF using the same parameters described in Sect. 3.1 and cross-matched with the input mock catalog (within a circle of 1 arcsec radius). The ratio of the number of successfully retrieved mock sources to the original mock sources injected in the map, in each $[S_{\text{int}}^{\text{in}}, \theta_M^{\text{in}}]$ bin, represents the completeness (see Fig. 1).

3.3.1. Selection function, maximum recovered size

To constrain the maximum detectable size of a galaxy as a function of redshift, stellar mass, and $\Delta \log(\text{SSFR})_{\text{KMS}}$, we explore the angular size of mock sources that were resolved by PyBDSF. The completeness levels in the $\theta_M^{\text{in}}$ versus $S_{\text{int}}^{\text{in}}$ plane (Fig. 1) reveal that the maximum recovered deconvolved FWHM ($\theta_M^{\text{max}}$), for sources within $10^{-5}$ Jy < $S_{\text{int}}$ < $10^{-3}$ Jy, strongly depends on $S_{\text{int}}$: i.e., a higher flux density increases the possibility of detecting extended sources$^1$. Thus, at a given redshift, faint galaxies are preferentially detected if they are compact, while bright starbursting systems are detected even if they are extended. This selection function (i.e., completeness level of 10%) is further discussed in Sects. 4.1 and 4.2.

3.3.2. Upper limit for the size of unresolved sources

A total of 665 SFGs (21%) from our sample are unresolved ($\theta_M^{\text{out}} = 0$ arcsec) by PyBDSF. In order to assign an upper limit to their intrinsic angular size ($\theta_M^{\text{in}} < \theta_{\text{beam}}$), we explore the input size of mock sources retrieved as unresolved in the MC simulations. In Fig. 2, we plot their distribution in the $\theta_M^{\text{in}}$–$S_{\text{int}}^{\text{in}}$ plane. Most of the sources retrieved as unresolved by PyBDSF are, as expected, at the faint and compact end of the parameter space tested here. Based on their angular size distribution (Fig. 2, right panel), we find that around 90% of them satisfy the condition: $\theta_M^{\text{in}} \leq \theta_{\text{beam}}$.

---

$^1$ Not all bins at the bright/compact end exhibit a 100% completeness. We attribute this result to the minimum number of pixels in an island (minpix_isl = 9) used to retrieve the radio sources with PyBDSF. Negative noise fluctuations might hinder the detection of islands of emission above this threshold. Certainly, we verified that using $\text{minpix_isl} = 6$ yields a higher completeness at the bright/compact end. Even so, we adopted $\text{minpix_isl} = 9$ to be consistent with the original VLA COSMOS 3 GHz catalog (Smolčić et al. 2017a).
Fig. 3. Systematic errors and uncertainties for the FWHM (left two panels) and flux density (right two panels) of mock sources in the $S^\text{out}_{\text{int}}-S^\text{out}_{\text{in}}$ plane covered by 2519 resolved SFGs studied in this work. Contour levels showing the distribution of these sources are at 1, 5, 50, 250 sources per bin. Positive (negative) values of $\mu^i_{\text{ij}}$ and $\sigma^i_{\text{ij}}$ indicate that the measured quantity is overestimated (underestimated). Values of $\sigma^0_i$ and/or $\sigma^j_i$ higher (lower) than 1 suggest that the uncertainty of the measured parameter is being underestimated (overestimated).

We thus define $\theta_{\text{beam}} = 0.75$ arcsec as the upper limit for the size of the 665 unresolved SFGs in our sample.

3.3.3. How reliable are the retrieved FWHM and flux density?

It is well known that noise fluctuations boost the flux of faint sources detected in sensitivity-limited astronomical surveys (e.g., Hogg & Turner 1998; Copin et al. 2005; Casey et al. 2014). It is expected that a similar effect takes place when determining the size of faint and compact sources. Therefore, in a pioneering effort, we use the MC simulations to correct both the FWHM and flux density (and associated uncertainties) provided by PyBDSF. We proceeded as follows:

1. We create a catalog containing all mock sources retrieved by PyBDSF. Hence, it contains information about the input ($S^\text{in}_{\text{int}}$, $\theta^\text{in}_{\text{M}}$) and output parameters ($S^\text{out}_{\text{int}}, \theta^\text{out}_{\text{M}}$).
2. All the sources in the catalog are binned in the $S^\text{out}_{\text{int}}-\theta^\text{out}_{\text{M}}$ plane (as shown in Fig. 1). For all objects in each bin, we estimate $r^i_0 \equiv (\theta^\text{out}_{\text{M}} - \theta^\text{in}_{\text{M}})/\sigma^0_i$ and/or $r^j_S \equiv (S^\text{out}_{\text{int}} - S^\text{in}_{\text{int}})/\sigma^j_S$, where $\sigma^0_i$ and $\sigma^j_S$ are the uncertainties provided by PyBDSF.
3. We derive the mean ($\mu$) and standard deviation ($\sigma$) of the $r^i_0$ and $r^j_S$ distributions (Fig. 3). While the value of $\mu$ quantifies systematic biases (e.g., “flux boosting”), $\sigma$ evaluates whether the uncertainties given by PyBDSF are under- or overestimated. In the ideal case where the measured properties and uncertainties are an appropriate description of the input mock sources, the mean ($\mu$) and standard deviation ($\sigma$) of the distribution should be 0 and 1, respectively. Nevertheless, for both FWHM and flux density, $\mu$ is generally greater than zero (Fig. 3), meaning that PyBDSF tends to overestimate the size and flux density of mock sources. The value of $\sigma$ is also heterogeneous across the $\theta^\text{out}_{\text{M}}-S^\text{out}_{\text{int}}$ plane (Fig. 3); $\sigma^i_{\text{ij}}$ and/or $\sigma^j_i$ higher (lower) than 1 suggests that the uncertainty provided by PyBDSF is being underestimated (overestimated).
4. Under the condition that all $r^i_0$ and $r^j_S$ distributions should have a mean of zero and dispersion of 1, the corrected source properties ($\theta^\text{out}_{\text{M}}, S^\text{out}_{\text{int}}$) and associated uncertainties ($\sigma^0_{\text{ij}}, \sigma^j_S$) are given by

$$\begin{align*}
\theta^\text{out}_{\text{M}} &= \theta^\text{in}_{\text{M}} + \mu^i_{\text{ij}} \times \sigma^0_i, \\
S^\text{out}_{\text{int}} &= S^\text{in}_{\text{int}} + \mu^j_S \times \sigma^j_S,
\end{align*}$$

and

$$\begin{align*}
\sigma^0_i &= \sigma^i_{\text{ij}} \times \sigma^0_i, \\
\sigma^j_S &= \sigma^j_{ij} \times \sigma^j_S,
\end{align*}$$

(4)

where $\mu^i_{\text{ij}}, \mu^j_S, \sigma^i_{\text{ij}},$ and $\sigma^j_i$ are the mean and standard deviations of the $r^i_0$ and $r^j_S$ distributions in each bin; here $i = 1 \ldots m$ and $j = 1 \ldots n$, with $m$ and $n$ the numbers of columns and rows used to grid the $\theta^\text{out}_{\text{M}}-S^\text{out}_{\text{int}}$ plane.

5. After applying our corrections to all mock resolved sources, we retrieve the distribution of $r^i_0$ and $r^j_S$. By fitting a single-Gaussian component, we find $\mu = 0$ and $\sigma = 1.0$ for both distributions (Fig. 4). This assures that the corrected flux densities and FWHM, as well as their associated uncertainties, are a good description of the input mock sources. It should be noted that for a small fraction of mock sources, our corrected FWHM is still being underestimated; this gives rise to a wing in the $r^i_0$ distribution (Fig. 4). We verified that these outliers are mainly located at the extended and bright end of the $\theta^\text{out}_{\text{M}}-S^\text{out}_{\text{int}}$ plane.

6. To correct the measured flux density of unresolved sources, we compare the input and output flux density of mock sources retrieved as unresolved by PyBDSF (see Fig. A.1). We then derive flux boosting factors as a function of $S/N$; at $S/N = 5$ the flux density is overestimated by 17%, while at $S/N > 7$ the effect of flux boosting is negligible.

7. We verified that the corrections and the completeness do not strongly depend on the input angular size and flux density distribution used in the MC simulations. A uniform distribution (equal number of sources per bin in the $\theta^\text{out}_{\text{M}}-S^\text{out}_{\text{int}}$ plane) yields correction factors that are consistent with those obtained from a realistic input distribution.

After validating our method, we then derived the corrected flux density and size of SFGs in our sample. In Fig. 5, we compare the flux and FWHM before and after revision in order to illustrate the effect of our corrections. Both flux and size measurements appear to be overestimated for faint radio sources. This result is expected as positive noise fluctuations enhance the flux density on a pixel-by-pixel basis and, consequently, the amplitude and variance of a 2D Gaussian model are magnified. This
phenomenon translates into a flux boosting factor of $\sim 20\%$ at the faint end (see right panel of Fig. 5), which is comparable with the uncertainty on the flux density of a $5\sigma$ radio source detection. On the other hand, “size boosting” seems to be ubiquitous for faint and compact sources that have a deconvolved FWHM smaller than the size of the synthesized beam (see left panel of Fig. 5). This can be attributed to the large uncertainties associated with the deconvolution process of slightly resolved and faint radio sources.

As a consistency test, we compare our corrected flux density measurements with those reported by Smolčić et al. (2017b), which were derived following a non-parametric approach with blobcat. Considering both resolved and unresolved sources (see Fig. A.2), we found that the two quantities are, on average, consistent.

3.4. From flux and size measurements to SFR and effective size estimates

We estimate the total SFR by adding the estimates from the infrared (SFRIR) and uncorrected UV emission (SFRUV), allowing us to account for the dust obscured and unobscured star formation activity. We use the Kennicutt (1998) calibration and the infrared–radio correlation (e.g., Magnelli et al. 2015; Delhaize et al. 2017) to derive SFRIR as

$$\text{SFRIR} = f_{\text{IMF}} 10^{-24} \varphi_{\text{IMF}} L_{1.4\text{GHz}} \text{[W Hz$^{-1}$]}, \quad (5)$$

where $f_{\text{IMF}} = 1.72$ for a Salpeter initial mass function (IMF) and $\varphi_{\text{IMF}}$ is parametrized as a function of redshift (for SFGs only) as $\varphi_{\text{IMF}} = (2.83 \pm 0.02) \times (1 + z)^{-0.15 \pm 0.01}$ (Delhaize et al. 2017). The value of $L_{1.4\text{GHz}}$, on the other hand, can be derived from the observer-frame 3 GHz fluxes ($S_{\nu_{\text{obs}}} \text{[W Hz$^{-1}$ m$^{-2}$]}$) through

$$L_{1.4\text{GHz}} = \frac{4\pi D_L(z)^2}{(1 + z)^{3-\alpha}} \left(\frac{1.4}{3}\right)^{-\alpha} S_{3\text{GHz}}, \quad (6)$$

where $D_L$ is the luminosity distance in meters and $\alpha$ is the spectral index of the synchrotron power law ($S_{\nu} \propto \nu^{-\alpha}$) of 0.8 (Condon 1992).

We also use the near-UV (NUV) emission of galaxies from the COSMOS2015 catalog (Laigle et al. 2016) to estimate SFRUV as follows (Kennicutt & Evans 2011):

$$\text{SFRUV} = 10^{-43.17} L_{\text{NUV}} \text{[erg s$^{-1}$]} \quad (7)$$

Finally, to compare our radio continuum size estimates with those derived from the optical/UV, we convert our $\theta_{\text{eff}}$ measurements into effective radius ($R_{\text{eff}}$), i.e., the radius enclosing half of the total flux density. To this end, we assume that most of our galaxies are star-forming disks with an exponentially declining surface brightness distribution. This is consistent with the average Sérsic index of $n \sim 1$ for MS galaxies (e.g., Nelson et al. 2016b) and luminous sub-mm selected galaxies (SMGs; Hodge et al. 2016), preferentially located above the MS. Under this assumption, Murphy et al. (2017) have analytically proven that for slightly resolved radio sources (with $R_{\text{eff}} \approx \theta_{\text{beam}}$), $\theta_{\text{eff}}$ and $R_{\text{eff}}$ can be related by

$$\theta_{\text{eff}} = 2.430 R_{\text{eff}}. \quad (8)$$

3.5. Final sample

We distributed the 3184 SFGs in our sample in five redshift bins following those presented by Laigle et al. (2016): [0, 0.35), [0.35, 0.65], [0.65, 0.9], [0.9, 1.35], [1.35, 1.7], and [1.75, 2.25]. This allows us to directly use the stellar mass completeness limits (per redshift bin) of the COSMOS2015 catalog, and hence assemble a mass-complete sample of radio-selected SFGs in the COSMOS field. The number of SFGs per redshift bin is nearly homogeneous (with a median of $\sim 650$ sources). Given the small comoving volume probed by COSMOS at low redshift and the selection function that restricts our parameter space to compact starburst galaxies, we are not able to explore the size evolution of SFGs in the redshift regime below $z = 0.35$.

In Fig. 6, we present the sample of 3184 SFGs in the SFR–$M_*–z$ plane. The bulk of the radio-selected SFGs is consistent with the position and dispersion of the MS of SFGs, as given by Schreiber et al. (2015). At the low-mass end, however, our radio detection limit biases our sample towards the starburst population. Since we aim to statistically analyze the size distribution of SFGs on and above the MS, we need to focus on the high-mass end. For this purpose, we define a mass-limit ($M_{\text{lim}}$) for each redshift bin, above which we are able to consistently probe both SFGs on ($\Delta \log(M/M_\odot) < 0.3$) and above the MS ($\Delta \log(M/M_\odot) > 0.3$). By considering systems with $M_* > M_{\text{lim}}$, we are also able to assemble a mass-complete sample of radio-selected SFGs, given that in all redshift bins $M_{\text{lim}}$ is higher than the stellar mass completeness limit of the COSMOS2015 catalog.

4. Results

In this section, we explore the dependence of the radio continuum size ($R_{\text{cont}}$) on the stellar mass, distance to the MS and redshift. We carefully address these relations while keeping in mind the completeness and size biases mentioned in Sect. 3.3 and that our analysis is restricted to $M_* > M_{\text{lim}}$, i.e., the part of the parameter space where the sample of SFGs on and above the MS is complete. We also verified that the trends presented below remain even if we use uncorrected measurements (see Appendix B).
4.1. Radio continuum size versus stellar mass

The stellar mass-size relation in galaxies (e.g., Furlong et al. 2017; Allen et al. 2017) is thought to be linked to the physical processes that regulate galaxy assembly, such as galaxy minor and major mergers and gas accretion (e.g., Khochfar & Silk 2006, 2009; Dekel et al. 2009; Oser et al. 2010; Gómez-Guijarro et al. 2018). Thus, it is a fundamental ingredient to understand galaxy evolution.

Here, we attempt to characterize the stellar mass–radio size relation up to $z = 2.25$. We thus explore the scatter of SFGs in the $R_{\text{MS}}-M_\star$ plane by deriving their density distribution per stellar mass bin (0.5 dex width; Fig. 7). We use 10,000 Monte Carlo trial model runs to take into account the dispersion introduced by the uncertainties and upper limits of $R_{\text{MS}}$ for resolved and unresolved sources, respectively. Based on our MC simulations (Fig. 2), the size of unresolved sources can be drawn from a uniform distribution in log space within the range $[0.1, R_{\text{lim}}^\text{ms}]$ kpc, where $R_{\text{lim}}^\text{ms}$ is the upper limit for the source size. We also derived the median size of SFGs through the Kaplan–Meier (KM) estimator (Kaplan & Meier 1958), which allows us to take into account the upper limits for the size of unresolved sources. We find that the two methods, MC realizations and KM estimator, yield consistent results (Fig. 7). In all redshift bins, the size distribution of SFGs remains constant over the range of stellar mass probed here, where the median size differs by less than 25% (see Table C.1). Qualitatively, this result is consistent with the shallow slope ($\alpha_{\text{opt/UV}}$) of the stellar mass and optical/UV size relation of SFGs ($\alpha_{\text{opt/UV}} \sim 0.2$; e.g., van der Wel et al. 2014; Mowla et al. 2018). Finally, we checked that this relation remains if we use two separate samples of SFGs: one composed of galaxies on the MS ($-0.3 \leq \Delta \log(\text{SSFR})_{\text{MS}} \leq 0.3$) and another above it ($\Delta \log(\text{SSFR})_{\text{MS}} > 0.3$).

We still have to consider that the last result might be affected by our selection function. As mentioned in Sect. 3.3, galaxies are preferentially detected if they are compact, especially at the faint end. This could yield a misleading stellar mass–radio size relation, as low-mass SFGs are fainter than their massive counterparts (due to the MS slope, Fig. 6). To quantify this possible bias, we use the output of our MC simulation to estimate the maximum recoverable size as a function of stellar mass as follows. At a given redshift bin and for each mass, we infer the SFR of a galaxy with $\Delta \log(\text{SSFR})_{\text{MS}} = 0$. Then we convert this SFR into flux density using the central redshift of the bin. Finally, this flux is associated with a maximum recoverable size using our 10% completeness limit in the $\theta_{\text{MS}}^2-S_{\text{ms}}^\text{int}$ plane (Sect. 3.3.1). As observed in Fig. 7, this selection function hinders the detection of extended SFGs with stellar mass below and near $M_{\star}^\text{lim}$; however, it does not affect the parameter space above $\log(M_\star/M_\odot) = 10.5$. Hence, the negligible dependence of the stellar mass on the radio size of SFGs with $\log(M_\star/M_\odot) > 10.5$ remains unaffected by our selection.

4.2. Radio continuum size of SFGs on and above the main sequence

Since both the size and $\Delta \log(\text{SSFR})_{\text{MS}}$ of SFGs can be discussed within the context of gas accretion and merger-driven star formation (e.g., Elbaz et al. 2011, 2018; Lang et al., in prep.), it is essential to characterize their interplay in detail. We therefore take advantage of our mass-complete sample of radio-selected SFGs to systematically explore their size distribution as a function of $\Delta \log(\text{SSFR})_{\text{MS}}$ and cosmic time (Fig. 8). We recall that we consider SFGs with $M_\star > M_{\star}^\text{lim}$, which is the region of the parameter space where we can consistently probe galaxies on and above the MS.

Similarly to the previous section, we derive the median size of SFGs per $\Delta \log(\text{SSFR})_{\text{MS}}$ bin following a MC approach and using the KM estimator. As observed in Fig. 8, the two methods agree well and reveal a trend where SFGs with higher $\Delta \log(\text{SSFR})_{\text{MS}}$ values are compact, in particular at lower redshifts where this tendency is more pronounced. The median size of $z \sim 0.5$ ($z \sim 2$) MS galaxies is 4 (2) times larger than those with $\Delta \log(\text{SSFR})_{\text{MS}} > 0.9$ (see Table C.2). We note that although SFGs on the MS are preferentially extended (median $R_{\text{eff}} \sim 1.5 \pm 0.2$ kpc), some can be as compact as galaxies with elevated $\Delta \log(\text{SSFR})_{\text{MS}}$.

To verify that our selection function does not bias these trends, we estimate the maximum recoverable size as a function of $\Delta \log(\text{SSFR})_{\text{MS}}$. At a given redshift bin and for each $\Delta \log(\text{SSFR})_{\text{MS}}$, we infer the SFR of a galaxy with...
Fig. 6. Upper panels: sample of 3184 SFGs in the SFR−M_⋆−z plane, color-coded for the number of sources per bin. Shown are the positions of the MS of SFGs (black lines) and the associated dispersion (dashed black lines) given by Schreiber et al. (2015). Vertical red lines show the mass limit above which we consistently probe SFGs on and above the MS. Middle and lower panels: median size and star formation surface density, respectively, of the 3184 SFGs in the sample. We use 1000 MC realizations to include the values for unresolved sources, which are drawn from the distributions described in Sects. 4.2 and 4.4.

M_⋆ = M_{\text{lim}}^{\text{MS}} evaluated at the central redshift of the bin. We then convert this SFR into flux density and associate it to a maximum recoverable size using our completeness in the θ_{\text{rec}}−S_{\text{lim}} plane (Sect. 3.3.1). As inferred from Fig. 8, while the selection function does not affect the size distribution of SFGs above the MS, it does hamper the detection of extended SFGs on and below the MS. We note that retrieving these missing systems would only strengthen the anti-correlation between the median size of SFGs and log(SFR)_{MS}.

The size dichotomy of SFGs on and above the MS is consistent with the results of Elbaz et al. (2011) and Rujopakarn et al. (2016). They did not report, however, that compact galaxies can be “hidden” within the MS (Fig. 8), which is in agreement with recent high-resolution observations of z ~ 2 SFGs (e.g., Elbaz et al. 2018; Lang et al. in prep.). This effect is related to the stacking approach followed by Elbaz et al. (2011), which can only provide median quantities, and the small sample of Rujopakarn et al. (2016), which might be affected by incompleteness.

In general, these findings support the emerging consensus where the global star-forming region of SFGs on the MS is preferentially but not exclusively extended, while SFGs above the MS are always more compact systems.

4.3. Size of SFGs in different wavelengths and its evolution with redshift

We now explore the radio continuum size evolution of SFGs over the redshift range 0.35 < z < 2.25 to better constrain the
processes regulating the growth of galaxies. In addition, through the comparison of the size–redshift relation as traced by stellar light, dust, and supernova remnants, we investigate where and how new stars are formed in galaxies. To this end, we select galaxies from our final sample (Sect. 3.5) with $\log(M_*/M_\odot) > 10.5$, which is the only mass bin consistently probed across the redshift range explored here. For all the redshift bins, we then derive the median $R_{\text{eff}}$ (via the KM estimator) of SFGs on and above the MS and $(\Delta \log(\text{SSFR})_{\text{MS}} > 0.3)$, respectively. As illustrated in Fig. 9, the radio continuum size of both SFG populations remains nearly constant across cosmic time. By using a parametrization of the
form $R_{\text{eff}} \propto (1 + z)^\alpha$, we find a slope of only $-0.26 \pm 0.08$ ($0.12 \pm 0.14$) for SFGs on (above) the MS. As expected from the results in Sect. 4.2, the median size of SFGs on the MS ($R_{\text{eff}} = 1.5 \pm 0.2$ kpc) is significantly larger than for those above it ($R_{\text{eff}} = 1.0 \pm 0.2$ kpc).

The size evolution presented here is still influenced by two factors that cannot be characterized with the available data:

- First, our radio continuum size estimates differ across each rest-frame frequency, from 4 GHz at $z = 0.35$ to 9.7 GHz at $z = 2.25$. Since higher energy electrons lose energy more rapidly, their propagation throughout the galaxy is more limited than their low-energy counterparts (e.g., Kohayashi et al. 2004); 3 GHz emitting electrons, in particular, are expected to diffuse $\sim 25\%$ farther into the ISM than those at 10 GHz (Murphy et al. 2017). Thus, the observed radio continuum synchrotron emission would tend to be more concentrated as the redshift decreases. This phenomenon does not significantly affect the trends presented above as a 25\% larger radio size at $z = 2.25$ would lead to $\alpha \sim -0.10 \pm 0.25$ for SFGs on (above) the MS.

- Second, given the limited number of resolution elements across the SFGs in the sample, we cannot assess their radio continuum surface brightness distribution and determine $R_{\text{eff}}$. Therefore, we converted the deconvolved FWHM to $R_{\text{eff}}$ (following Murphy et al. 2017) assuming that most of our SFGs follow an exponentially declining surface brightness distribution (with Sérsic index $n = 1$). Naturally, such a conversion might deviate from the true $R_{\text{eff}}$ on a galaxy-by-galaxy basis, especially for SFGs above the MS that tend to have cuspier light profiles (e.g., Wuyts et al. 2011). For slightly resolved sources, like the ones presented here, we do not expect major changes in $R_{\text{eff}}$ if the Sérsic index is larger than 1. For example, even for 0.2 arcsec resolution dust-continuum observations, Elbaz et al. (2018) report that $R_{1/2} \equiv 0.5 \times \text{FWHM}$ and $R_{\text{eff}}$ are both equally good proxies for the half-light radius, either leaving the index free or fixed to 1.

4.3.1. Comparison with other radio continuum size estimates

Bondi et al. (2018) have independently derived the $R_{\text{eff}}$ of radio sources detected in the VLA COSMOS 3 GHz map (Smolčić et al. 2017a), including AGN and SFGs up to $z \sim 3$. They assembled a sample of SFGs that is complete in $L_\text{3 GHz}$ over $0.8 < z < 3$ with median log($M_*/M_\odot$) = 10.6; no distinction was made between on and off MS galaxies for this sample, the size of SFGs marginally increases with cosmic time, from median $\sim 1.4$ kpc at $z = 2.1$ to $R_{\text{eff}} \sim 1.6$ kpc at $z = 0.8$. This is in agreement with the shallow size evolution of MS galaxies in our sample (which corresponds to $\sim 90\%$ of all SFGs) with median $R_{\text{eff}} = 1.5 \pm 0.1$ kpc, within the same redshift range and comparable stellar mass. We note that despite the independent methodologies used to measure radio sizes in the $\mu$Jy regime, and different selection criteria, our median sizes are consistent. Bondi et al. (2018) used, in particular, the original and convolved images (up to a resolution of 2.2 arcsec) of the VLA COSMOS 3 GHz mosaic and took the flux density provided by blobcat as a prior in their 2D Gaussian fitting procedure. This flux prior limits the effect of size boosting, leading to comparable size measurement between our two studies. On the other hand, by using the VLA COSMOS 3 GHz map, Miettinen et al. (2017) reported a median $R_{\text{eff}} \sim 1.9$ kpc for 152 SMGs over the redshift range of $1 \leq z \leq 6$. The discrepancy of $\sim 35\%$ between the Miettinen et al. (2017) value of $R_{\text{eff}}$ and that reported here at $z = 2.25$ and in Bondi et al. (2018) is likely driven by the different selection criteria.

The angular size of the $\mu$Jy radio sources has also been recently explored in different extragalactic deep fields. At the same frequency, 3 GHz, it was found that $z \sim 1$ SFGs in the Lockman-Hole have a median effective radius of $\sim 1.0$ kpc (Cotton et al. 2018). This agrees with the median size of SFGs above the MS derived here (see Fig. 9). Murphy et al. (2017) have reported that $z \sim 1.2$ GOODS-N SFGs have a median $R_{\text{eff}}$ of only $\sim 0.5$ kpc at 10 GHz. This small size could be associated with selection effects as the high-resolution 10 GHz observations (0.22 arcsec) are sensitive to smaller angular scales. Additionally, as stated by Murphy et al. (2017), the discrepancy between 3 GHz and 10 GHz radio sizes is driven by the frequency-dependent cosmic ray diffusion. This physical phenomenon could partially explain the large median $R_{\text{eff}}$ of 2.3 $\pm$ 0.6 kpc at 1.4 GHz of $z \sim 1.5$ SFGs (in the Hubble Deep Field; Lindroos et al. 2018), which is $\sim 1.5$ times larger than the median size at 3 GHz of galaxies in our sample. On the contrary, the larger energy loss rate at higher frequencies cannot explain the large median $R_{\text{eff}}$ of $\sim 2.7$ kpc (FWHM $\sim 0.8$ arcsec) at 5.5 GHz reported for SFGs at similar redshift (in GOODS-N; Guidetti et al. 2017). Finally, we note that (apart from the frequency and resolution of the observations) a more general issue about size determination is related to the surface brightness limit of each survey. As inferred from Fig. 1, a lower rms sensitivity hampers the maximum detectable angular size, biasing the sample towards more compact radio sources.
4.3.2. Comparison with FIR, optical, and Hα sizes

It has been reported that the FIR size of $z \sim 2$ SFGs is, on average, ~1.5 kpc (e.g., Rujopakarn et al. 2016; Elbaz et al. 2018, Lang et al. in prep.), which is consistent with the median radio size of SFGs reported here. Extinction-corrected Hα radial profiles tracing the global star-forming region of $z \sim 1.4$ SFGs yield a median effective radius of $<1$ kpc (with $9.8 < \log(M_*/M_\odot) < 11$; Nelson et al. 2016a), comparable with the radio sizes of SFGs above the MS. In contrast, the median effective radius derived from uncorrected Hα emission is $4.2 \pm 0.1$ kpc for SFGs at $z \sim 1$ and similar stellar mass (Nelson et al. 2012). This disparity is related to the high dust content in massive galaxies; if star formation is highly obscured at small radii, Hα emission is overestimated and the inferred effective radius will thus be larger (e.g., Möllerhoff et al. 2006; Nelson et al. 2016a,b).

In Fig. 9, we also present a comparison of the size of SFGs as observed from radio continuum and optical to UV throughout cosmic time. We use the size evolution derived by Shibuya et al. (2015) via broadband optical imaging with the Hubble Space Telescope (HST). In this case, we adopt the median fit obtained for the UV bright galaxies ($-24 < M_UV < -21$) corresponding to the stellar mass range of $10^9 \lesssim \log(M_*/M_\odot) < 11$, which is consistent with the mass range of SFGs used in this work. Given that Shibuya et al. (2015) masked star-forming clumps, their size estimates can be used as a proxy for the stellar mass distribution of galaxies. As revealed by Fig. 9, the overall star-forming region of SFGs (on and above the MS) is more compact than their stellar component. In particular, at $z = 0.5$ (2) the optical to UV emission is $\sim 2$ (1.3) times more extended than the radio continuum. Since these HST-based estimates are not corrected for extinction, it is likely that (similar to what has been shown for Hα sizes) the optical to UV effective radius is overestimated. Given that dust extinction becomes substantial in high-redshift galaxies (e.g., Leslie et al. 2018), we would expect that their optical to UV size is overestimated by a larger fraction than those galaxies at lower redshifts. Correcting for this effect could alleviate the discrepancy between the extent of the stellar and star-forming component of galaxies at high redshift.

In summary, radio continuum, FIR, and extinction-corrected Hα emission suggest that star formation occurs in smaller regions relative to the total stellar component (e.g., Simpson et al. 2015; Rujopakarn et al. 2016; Fujimoto et al. 2017; Elbaz et al. 2018; Karoumpis et al., in prep.). Here, in particular, we find that while the radio continuum size remains nearly constant, that inferred from optical to UV emission increases with cosmic time. In Sect. 5.2, we discuss this finding further within the context of bulge growth.

4.4. Cosmic evolution of $\Sigma_{\text{SFR}}$

From numerical simulations, SFGs are expected to experience a compaction phase while the SFR increases and they move towards the upper end of the MS (e.g., Tacchella et al. 2016). This scenario can now be tested through our radio continuum size estimates. We thus derive the galactic-average SFR surface density, $\Sigma_{\text{SFR}} = \text{SFR}/2\pi R_{\text{eff}}$, and use it to evaluate how concentrated the star formation activity in galaxies is as they evolve across the MS (see Figs. 6 and 10). In calculating $\Sigma_{\text{SFR}}$, we assume that the total SFR ($\text{SFR}_{\text{tot}} + \text{SFR}_{\star}$) is confined within the radio continuum-based $R_{\text{eff}}$. This simplification is valid as the UV-based SFR is considerably low for massive, high-redshift SFGs (e.g., Buat et al. 2012); therefore, the star formation activity is mainly traced by the radio continuum (unobscured emission).

As in Sects. 4.1 and 4.2, we follow a MC approach to derive the distribution of $\Sigma_{\text{SFR}}$ per $\Delta\log(\text{SFR})_{\text{MS}}$ bin (Fig. 10); again, only galaxies with $M_\star > M_{\text{lim}}^\star$ are included in the analysis. In this case, the $\Sigma_{\text{SFR}}$ for unresolved sources is drawn from a uniform logarithmic distribution within the range $[\Sigma_{\text{SFR}}(R_{\text{MS}}), \Sigma_{\text{SFR}}(0.1\,\text{kpc})]$. We verify the reliability of this approach by using the KM estimator (Table C.3). At all redshifts and for both methods, there is a positive relation between $\Sigma_{\text{SFR}}$ and $\Delta\log(\text{SFR})_{\text{MS}}$ that can be described with a power law,

$$
\log(\Sigma_{\text{SFR}}) = \alpha \times \log(\text{SFR})_{\text{MS}} + \beta,
$$

where $\alpha$ and $\beta$ are the slope and normalization, respectively. We use a least-squares (Levenberg–Marquardt) algorithm to fit a linear function to $\Delta\log(\text{SFR})_{\text{MS}}$ and derive the best-fitting values for $\alpha$ and $\beta$. This procedure is done for each MC realization, while restricting our parameter space to $\Delta\log(\text{SFR})_{\text{MS}} > -0.3$ where our sample is complete. The final values for $\alpha$ and $\beta$ shown in Fig. 10, correspond to the median (and 16th and 84th percentiles) after executing 1000 MC trial model runs. While the normalization of the $\log(\Sigma_{\text{SFR}}) - \Delta\log(\text{SFR})_{\text{MS}}$ relation decreases with redshift, the value of $\alpha$ reveals that this trend becomes shallower at higher redshift. At $z \sim 2$, the difference between $\Sigma_{\text{SFR}}$ of galaxies on and above the MS is smaller than in the local Universe.

Spatially resolved studies of local SFGs have also reported more centrally peaked radial profiles of $\Sigma_{\text{SFR}}$ as the distance to the MS increases (Ellison et al. 2018). It has been found that the FIR surface density evolves across the MS with a logarithmic slope of 2.6 (Lutz et al. 2016), which is consistent with the value we derived at $z \sim 0.5$ ($\alpha = 2.6$) using the radio continuum emission. Likewise, from Hα resolved maps it has been shown that $z \sim 1$ SFGs follow this relation with $\alpha \sim 1.1$ (Magdis et al. 2016), which is ~50% lower than that reported in this work. This tendency can also be inferred from the $\Sigma_{\text{SFR}}$, $M_\star$, and SFR of $1 \lesssim z \lesssim 3$ SFGs reported by Genzel et al. (2010) and Tacconi et al. (2013). As presented in Fig. 10, these SFGs also exhibit higher $\Sigma_{\text{SFR}}$ as the distance to the MS increases, yet their $\Sigma_{\text{SFR}}$ values appear systematically lower than those reported here. This could be a consequence of the optical-/UV/Hα- or CO-based size estimates used by the authors, which are larger than those obtained from radio continuum emission (Sect. 4.3.2). If we scale their $\Sigma_{\text{SFR}}$ values by using their $R_{\text{opt}}/R_{\text{radio}}$ ratios, they increase by a factor of $\sim 0.7$ ($0.4$) dex at redshift 1.5 (2), which would alleviate this observed discrepancy. Finally, in Fig. 10 we present the sample of SFGs from Elbaz et al. (2018) and Lang et al. (in prep.). Although $z \sim 1.5$ SFGs also follow a positive relation between $\log(\Sigma_{\text{SFR}})$ and $\Delta\log(\text{SFR})_{\text{MS}}$, those at $z \sim 2$ are widely scattered. As reported by Elbaz et al. (2018), $z \sim 2$ SFGs with starburst-like $\Sigma_{\text{SFR}}$ are also found close to or within the MS. According to our results, there is a fraction of MS galaxies for which $\Sigma_{\text{SFR}}$ is significantly higher than expected from the $\log(\Sigma_{\text{SFR}})-\Delta\log(\text{SFR})_{\text{MS}}$ relation. These high $\Sigma_{\text{SFR}}$ MS galaxies are present at all redshifts and comprise less than 10% of the MS galaxy population (see contour levels of Fig. 10).

5. Discussion

5.1. Cold gas accretion versus merger mode of star formation

We revealed that most SFGs ($\log(M_\star/M_\odot) > 10.5$) follow a linear relation in the $\log(\Sigma_{\text{SFR}})-\Delta\log(\text{SFR})_{\text{MS}}$ plane over the
redshift range $0.35 < z < 2.25$. To the first order, these results can be discussed within the context of the Kennicutt–Schmidt (KS) relation ($\Sigma_{\text{SFR}} \propto \Sigma_{\text{gas}}$; Kennicutt 1998). We therefore use the scaling relations of Genzel et al. (2015) to derive the typical molecular gas mass of galaxies at three different $\Delta \log(\text{SSFR})_{\text{MS}}$ bins, namely $[-0.3,0.3]$, $[0.3,0.6]$, and $[0.9, 2.0]$ (see Fig. C.2). Then we assume that our radio size estimates (Table C.2) also trace the extent of the molecular gas reservoir; these estimates are subsequently used to approximate the galactic averaged molecular gas density ($\Sigma_{\text{mol gas}} = M_{\text{mol gas}}/2\pi R_{\text{eff}}$). This information is combined with the $\Sigma_{\text{SFR}}$ values presented in Table C.3, allowing us to approximate the shape of the KS relation (see Fig. 11). It is reassuring that our data points, which cover a wide range in redshift and $\Delta \log(\text{SSFR})_{\text{MS}}$, agree within the uncertainties with the KS relation presented by Genzel et al. (2010). Moreover, our data points are consistent with the scenario wherein low- and high-redshift SFGs follow a similar molecular gas–star formation relation (Bouché et al. 2007; Genzel et al. 2010). By considering SFGs with $-0.3 \leq \Delta \log(\text{SSFR})_{\text{MS}} \leq 0.9$, we derive a super-linear slope of $1.3 \pm 0.1$. If SFGs with $\Delta \log(\text{SSFR})_{\text{MS}} > 0.9$ are included, the slope becomes steeper ($1.5 \pm 0.1$), which indicates that SFGs evolve towards a more efficient regime of star formation as $\Delta \log(\text{SSFR})_{\text{MS}}$ increases. This is consistent with the small size, and hence higher $\Sigma_{\text{SFR}}$, of galaxies above the MS (Fig. 8), which could be the result of gas-rich mergers (e.g., Moreno et al. 2015; Wellons et al. 2015) and/or violent disk instability (VDI; e.g., Dekel & Burkert 2014; Tacchella et al. 2016; Wang et al. 2019).

Beyond the broad picture of galaxy evolution discussed above, we also reported the discovery of MS galaxies harboring starburst-like $\Sigma_{\text{SFR}}$ conditions (Fig. 10). This result echoes, in particular, that of Elbaz et al. (2018), who reported the presence of hidden starbursts within the MS at $z \sim 2$. Then the fundamental question arises: what is the physical mechanism responsible for high-$\Sigma_{\text{SFR}}$ MS galaxies? First, these systems could be a result of large cold gas reservoirs distributed over small disk radii (Wang et al. 2019) that, due to disk instability episodes (e.g., Dekel & Burkert 2014), yield high $\Sigma_{\text{SFR}}$. If the gas replenishment occurs when a galaxy lies at the lower envelope of the MS (i.e., $\Delta \log(\text{SSFR})_{\text{MS}} \sim -0.3$), the SFR enhancement might not suffice to bring the galaxy above the MS. Second, high-$\Sigma_{\text{SFR}}$ MS galaxies could be explained in the context of merger-driven
bursts of star formation where, depending on the gas content, mergers could not significantly increase the SFR and offset the galaxy from the MS. (e.g., Fensch et al. 2017; Wang et al. 2019). This is in agreement with the observational evidence of merging activity in galaxies on and above the MS (e.g., Kartaltepe et al. 2012; Ellison et al. 2018; Calabrò et al. 2018; Wang et al. 2019; Cibinel et al. 2019; Kartaltepe, priv. comm.).

In light of these findings, SFR arises as a remarkable proxy for identifying starburst galaxies, where star formation is triggered by either mergers or VDI that lead to high SFR. As an illustrative case, here we evaluate $\Sigma_{\text{SFR max}} = M_{\text{max}} / \Sigma_{\text{SFR}}(\Delta \log(\Sigma_{\text{SFR MS}}) = 0.7)$\(^2\) at each redshift bin (Sect. 4.4), and adopt it as a threshold to identify starbursting systems. Under this definition, it is possible to decompose the bimodal distribution of SFGs along $\Delta \log(\Sigma_{\text{SFR MS}})$ (e.g., Sargent et al. 2012) into main sequence ($<\Sigma_{\text{SFR}}$) and starburst ($>\Sigma_{\text{SFR}}$) contributions (see Fig. 12). The first, and more dominant, distribution is centered at $\Delta \log(\Sigma_{\text{SFR MS}}) = 0$ and represents the population of galaxies forming stars through a secular mode of star formation (e.g., Dekel et al. 2009; Sellwood 2014). The distribution of non-merger-induced and merger-induced starbursts exhibits an enhanced tail at high $\Delta \log(\Sigma_{\text{SFR MS}})$ and, consequently, its median lies at

\[ \Delta \log(\text{SFR})_{\text{MS}} > 0. \]  

We note that this $\Sigma_{\text{SFR max}}$-based scheme brings the galaxy-pair and merger rate in better agreement with the fraction of high-redshift starbursts (see Fig. 13), given that a $\Delta \log(\text{SFR})_{\text{MS}}$-based definition misses the merger-induced starbursts hidden within the MS (Elbaz et al. 2018; Cibinel et al. 2019).

### 5.2. Is the centrally concentrated star formation in galaxies evidence of bulge growth?

The finding of compact radio continuum emission of SFGs on and above the MS further support the evidence of star formation enhancement at small radii (e.g., Simpson et al. 2015; Rujopakarn et al. 2016; Nelson et al. 2016; Fujimoto et al. 2017; Elbaz et al. 2018). Interestingly, while the extent of the stellar component increases with cosmic time, the overall region where most stars are formed remains nearly constant (see Fig. 9). This might indicate that fresh star-forming gas is constantly fueled towards the center of galaxies, either due to VDI, minor or major mergers, and/or tidal interactions (e.g., Larson 2003; Rupke et al. 2010; Sillero et al. 2017; Ellison et al. 2018; Muñoz-Elgueta et al. 2018). Regardless of the dominant mechanism driving the formation of stars in galaxies (on and above the MS), violent and secular galaxy evolutionary channels lead to the formation of a bulge (e.g., Kormendy & Kennicutt 2004; Fisher 2006; Hopkins et al. 2009; Brooks et al. 2016; Tonini et al. 2016). Ultimately, the presence of a dominant bulge could stabilize the gas disk against gravitational instabilities, and thus prevent the formation of stars (e.g., Lang et al. 2014, and references therein).

In this context, we hypothesize that the centrally concentrated star formation activity of most SFGs in the sample might reflect the growth of the bulge, which might precede the quenching of the galaxy from the inside out (e.g., Ellison et al. 2018). At this late evolutionary stage the bulge of massive galaxies is fully quenched, while star formation activity still takes place at large radii (e.g., Tacchella et al. 2015; Rowlands et al. 2018). Spatially resolved studies of low- and high-mass SFGs at high redshift are needed to verify such a scenario, and will allow us to understand how star formation, and hence stellar mass, is distributed in galaxies across cosmic time.

### 6. Summary

We presented the first systematic study of the radio continuum size evolution of SFGs over $0.35 < z < 2.25$. We used a mass-complete sample of 3184 radio-selected SFGs, detected in the VLA COSMOS 3 GHz map (Smolčić et al. 2017a), and performed extensive Monte Carlo simulations to characterize our selection function and validate the robustness of our measurements. We found the following:

- The radio continuum size shows no clear dependence on the stellar mass of SFGs with $10^{10.5} \lesssim M_*/M_\odot < 11.5$, which is the mass range where our sample is not affected by our selection function;
- MS galaxies are preferentially (but not exclusively) extended, while SFGs above the MS are more compact systems; the median size of SFGs on (above) the MS is $R_{\text{eff}} = 1.5 \pm 0.2 (1.0 \pm 0.2) \text{kpc}$. Using the parametrization of the form $R_{\text{eff}} \propto (1 + z)^\alpha$, we found that the median size remains nearly constant with cosmic time, with $\alpha = -0.26 \pm 0.08 (0.12 \pm 0.14)$ for SFGs on (above) the MS;
- The median radio size of SFGs is smaller (by a factor 1.3–2) than that inferred from optical to UV emission that traces their stellar component (Shibuya et al. 2015). Overall, these results are consistent with compact radio continuum, FIR, and
extinction-corrected Hα emission ($\lesssim 1.5$ kpc; e.g., Nelson et al. 2016a; Rujopakarn et al. 2016; Murphy et al. 2017; Cotton et al. 2018; Elbaz et al. 2018; Lindroos et al. 2018).

- Most SFGs follow a linear relation in the log ($\Sigma_{SFR}$ – $\Delta \log (SSFR)$) plane, consistent with previous studies of SFGs in the local Universe (Lutz et al. 2016) and at $z \sim 1$ (Magdis et al. 2016). While its normalization increases with redshift, its slope becomes steeper at lower redshifts (from $\alpha = 1.5$ at $z \sim 2$ to 2.6 at $z \sim 0.5$);

- There is a fraction ($\lesssim 10\%$) of MS galaxies harboring starburst-like $\Sigma_{SFR}$, consistent with recent evidence of hidden starbursts within the MS at $z \sim 2$ (Elbaz et al. 2018).

Overall, our results suggest that SFGs with enhanced star formation undergo a compaction phase. These systems could be explained in the context of disk instability and/or merger-driven burst of star formation that, depending on the gas content, offset the galaxy from the MS in different proportions (e.g., Fensch et al. 2017; Wang et al. 2019). Since using $\log (SSFR)$ alone prevents us from identifying those starburst galaxies hidden within the MS, we recommend using $\Sigma_{SFR}$ as well to better identify starbursting systems. Having constraints on $\Sigma_{SFR}$ is the first step towards the characterization of the KS relation at high redshift. Exploring in detail the gas content and optical morphology of SFGs in our sample is the subject of an upcoming manuscript.

Acknowledgements. E.F.J.A, B.M., A.K., F.B., E.V., E.R.D., K.H., and C.K. acknowledge the support of the Collaborative Research Center 956, subproject A1, funded by the Deutsche Forschungsgemeinschaft (DFG). E.F.J.A. thanks A. Mikler and J. Erler for their encouragement and insightful comments. E.V. acknowledges funding from the DFG grant BE 1837/13-1. P.L. acknowledges funding from the European Research Council (ERC) under the European Union’s Horizon 2020 Research and Innovation Programme (grant agreement No. 694343). S.T. acknowledges support from the ERC Consolidator Grant funding scheme (project ConText, grant No. 684179). The Cosmic Dawn Center is funded by the Danish National Research Foundation. J.D. acknowledges the financial assistance of the South African Radio Astronomy Observatory (SARAO; www.ska.ac.za). This work is based on observations with the National Radio Astronomy Observatory, which is a facility of the National Science Foundation operated under cooperative agreement by Associated Universities, Inc., and is also based on data products from observations made with ESO Telescopes at the La Silla Paranal Observatory under ESO program ID 179.A-2005 and on data products produced by TERAPIX and the Cambridge Astronomy Survey Unit on behalf of the UltraVISTA consortium.

References
Bourke, S., Mooley, K., & Hallinan, G. 2014, ASP Conf. Ser., 485, 367
Appendix A: Flux boosting

Fig. A.1. Flux boosting for unresolved sources as a function of S/N estimated from MC simulations. The solid and dashed red lines show the 50th percentile, and the 14th and 84th percentiles of the distribution as a function of S/N. The vertical blue dashed line indicates our detection threshold (S/N = 5).

Fig. A.2. Comparison between the flux density of 3184 SFGs in the sample (resolved and unresolved) derived from PyBDSF (corrected) and those reported by Smolčić et al. (2017a). The solid and dashed red lines show the 50th percentile, and 14th and 84th percentiles of the distribution as a function of the flux density reported in this study.
Appendix B: Size evolution of SFGs using uncorrected FWHM and flux density

The extensive Monte Carlo simulations performed in this work indicate that the FWHM and flux density are being overestimated for most of the radio sources in the sample (see Fig. 5). Using these values, however, does not systematically affect the trends and relations presented in this work, as initially inferred from Fig. B.1. First, this approach also leads to a flat relation between the median radio size and stellar mass of SFGs (Fig. B.2). Second, we find a similar radio size and $\Sigma_{\text{SFR}}$ dichotomy between SFGs on and above the MS (Figs. B.1 and B.3). Using uncorrected measurements leads to a smaller fraction of MS galaxies with starburst-like $\Sigma_{\text{SFR}}$ (Fig. B.4), given that the size of faint MS galaxies is overestimated and, consequently, $\Sigma_{\text{SFR}}$ becomes smaller. We note that regardless of the use of corrected or uncorrected values, the fraction of starburst-like $\Sigma_{\text{SFR}}$ systems remains unclear, due to the presence of MS galaxies that are unresolved in the VLA COSMOS 3 GHz map.

Fig. B.1. Upper panels: sample of 3184 SFGs in the SFR–$M_\star$–z plane color-coded for the number of sources per bin. Solid and dashed black lines show the position of the MS of SFGs and the associated dispersion given by Schreiber et al. (2015). Vertical red solid lines show the mass-limit above which we consistently probe SFGs on and above the MS. Middle and lower panels: median size and star formation surface density, respectively, of the 3184 SFGs in the sample. We use 1000 MC realizations to include the values for unresolved sources, which are drawn from the distributions described in Sects. 4.2 and 4.4.
Fig. B.2. Sample of 3184 SFGs in the $R_{\text{eff}}-M_*-z$ plane. The blue solid line shows the maximum size, corresponding to the 10% completeness level, that can be observed for a galaxy with $\Delta \log(\text{SSFR}) = 0$ evaluated at the central redshift value per bin. Blue dashed lines show the mass-limit above which we consistently probe SFGs on and above the MS. Black arrows show the upper limits for the size of unresolved sources. Lower panels: density distribution per stellar mass bin (0.5 dex width) of SFGs in the $R_{\text{eff}}-M_*-z$ plane. Contour levels are at the 10, 20, 30, 40, 50, 60, 70, 80, and 90th percentiles. The median size derived via the KM estimator is shown by the dark magenta circles; the error bars correspond to the 16th and 86th percentiles of the distribution.

Fig. B.3. Upper panels: star-forming galaxies with $M_* > M_*^{\text{lim}}$ in the size--$\Delta \log(\text{SSFR})--z$ plane. The blue solid line shows the maximum size, corresponding to the 10% completeness level, that can be observed for a galaxy with $M_* = M_*^{\text{lim}}$ evaluated at the central redshift value per bin. Black arrows show the upper limits for the size of unresolved sources. Lower panels: density distribution per $\Delta \log(\text{SSFR})$ bin (0.5 dex width) of SFGs in the size--$\Delta \log(\text{SSFR})--z$ plane. Contour levels are at the 10, 20, 30, 40, 50, 60, 70, 80, and 90th percentiles. The median size derived via the KM estimator is shown by the dark magenta circles; the error bars correspond to the 16th and 86th percentiles of the distribution.
Fig. B.4. Upper panels: star-forming galaxies with $M_\star > M_{\lim}^\star$ in the $\Sigma_{\text{SFR}} - \Delta \log(\text{SSFR}) - z$ plane. Dashed blue lines show the $\Delta \log(\text{SSFR})$ limit of $-0.3$ above which our sample of SFGs is complete in terms of distance to the MS. Black arrows show the lower limits for the size of unresolved sources.

Middle panels: density distribution per $\Delta \log(\text{SSFR})$ bin (0.4 dex width) of SFGs in the $\Sigma_{\text{SFR}} - \Delta \log(\text{SSFR}) - z$ plane. Contour levels are at the 10, 20, 30, 40, 50, 60, 70, 80, and 90th percentiles. The median size derived via the KM estimator is shown by the dark magenta circles; the error bars correspond to the 16th and 86th percentiles of the distribution.

Lower panels: power law describing the $\Sigma_{\text{SFR}} - \Delta \log(\text{SSFR})$ relation (dotted black line). The slope and normalization are given in the lower right corner. The median size derived via the KM estimator is also shown by the dark magenta circles. For comparison, we present the compilation of SFGs from Genzel et al. (2010, black filled circles), Tacconi et al. (2013, black empty squares), Elbaz et al. (2018, red pentagons), and Lang et al. (in prep., orange triangles). Black arrow illustrates the factor to be considered when comparing $\Sigma_{\text{SFR}}$ of galaxies in our sample with that reported by Genzel et al. (2010) and Tacconi et al. (2013).
### Appendix C: Tables

#### Table C.1. Radio continuum size as a function of the stellar mass of SFGs.

<table>
<thead>
<tr>
<th>( \log(M_*/M_\odot) ) (dex)</th>
<th>0.35 &lt; z &lt; 0.65</th>
<th>0.65 &lt; z &lt; 0.95</th>
<th>0.95 &lt; z &lt; 1.30</th>
<th>1.30 &lt; z &lt; 1.75</th>
<th>1.75 &lt; z &lt; 2.25</th>
</tr>
</thead>
<tbody>
<tr>
<td>[10.0, 10.5]</td>
<td>1.5_{-0.8}^{+1.2}</td>
<td>1.5_{-0.9}^{+1.4}</td>
<td>...^{(a)}</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>[10.5, 11.0]</td>
<td>1.8_{-1.1}^{+1.6}</td>
<td>1.5_{-0.7}^{+1.4}</td>
<td>1.4_{-0.7}^{+1.6}</td>
<td>1.3_{-0.8}^{+1.2}</td>
<td>1.2_{-0.5}^{+1.4}</td>
</tr>
<tr>
<td>[11.0, 11.5]</td>
<td>1.5_{-0.6}^{+1.4}</td>
<td>1.3_{-0.8}^{+1.4}</td>
<td>1.7_{-0.9}^{+1.4}</td>
<td>1.2_{-0.5}^{+2.2}</td>
<td>1.5_{-0.8}^{+1.2}</td>
</tr>
</tbody>
</table>

Notes. The effective radius is given in kpc. The uncertainties correspond to the 16th and 84th percentiles of the size distribution per stellar mass bin. \(^{(a)}\)No values are given for the mass bins that are strongly affected by incompleteness.

#### Table C.2. Radio continuum size as a function of the distance to the MS of SFGs.

<table>
<thead>
<tr>
<th>( \Delta \log(SSFR)_{\text{MS}} ) (dex)</th>
<th>0.35 &lt; z &lt; 0.65</th>
<th>0.65 &lt; z &lt; 0.95</th>
<th>0.95 &lt; z &lt; 1.30</th>
<th>1.30 &lt; z &lt; 1.75</th>
<th>1.75 &lt; z &lt; 2.25</th>
</tr>
</thead>
<tbody>
<tr>
<td>[−0.3, 0.3]</td>
<td>1.7_{-1.0}^{+1.3}</td>
<td>1.5_{-0.8}^{+1.3}</td>
<td>1.4_{-0.8}^{+1.5}</td>
<td>1.5_{-0.9}^{+1.2}</td>
<td>1.5_{-0.7}^{+1.2}</td>
</tr>
<tr>
<td>[0.3, 0.9]</td>
<td>1.0_{-0.7}^{+2.0}</td>
<td>1.0_{-0.6}^{+2.0}</td>
<td>1.3_{-0.7}^{+1.7}</td>
<td>0.9_{-0.5}^{+0.6}</td>
<td>0.9_{-0.3}^{+1.0}</td>
</tr>
<tr>
<td>&gt;0.9 (^{(a)})</td>
<td>0.4_{-0.1}^{+0.8}</td>
<td>0.4_{-0.2}^{+0.1}</td>
<td>0.5_{-0.2}^{+0.1}</td>
<td>0.7_{-0.2}^{+0.2}</td>
<td>0.8_{-0.3}^{+0.1}</td>
</tr>
</tbody>
</table>

Notes. The effective radius is given in kpc. The uncertainties correspond to the 16th and 84th percentiles of the size distribution per \( \Delta \log(SSFR)_{\text{MS}} \) bin. The minimum stellar mass probed throughout the different redshift bins is \( \log(M_*/M_\odot) = 9.9, 10.2, 10.5, 10.5, \) and 10.7. \(^{(a)}\)The highest \( \Delta \log(SSFR)_{\text{MS}} \) bin is centered at \( \Delta \log(SSFR)_{\text{MS}} = 1.2, 1.05, 1.05, 1.45, 1.45 \).

#### Table C.3. Star formation surface density (\( \Sigma_{\text{SFR}} \)) as a function of distance to the MS of SFGs.

<table>
<thead>
<tr>
<th>( \Delta \log(SSFR)_{\text{MS}} ) (dex)</th>
<th>0.35 &lt; z &lt; 0.65</th>
<th>0.65 &lt; z &lt; 0.95</th>
<th>0.95 &lt; z &lt; 1.30</th>
<th>1.30 &lt; z &lt; 1.75</th>
<th>1.75 &lt; z &lt; 2.25</th>
</tr>
</thead>
<tbody>
<tr>
<td>[−0.3, 0.3]</td>
<td>0.8_{-0.6}^{+2.9}</td>
<td>1.7_{-1.2}^{+7.0}</td>
<td>4_{-3}^{+15}</td>
<td>8_{-3}^{+27}</td>
<td>12_{-9}^{+29}</td>
</tr>
<tr>
<td>[0.3, 0.9]</td>
<td>5_{-4}^{+31}</td>
<td>15_{-14}^{+123}</td>
<td>14_{-12}^{+49}</td>
<td>63_{-30}^{+194}</td>
<td>96_{-66}^{+135}</td>
</tr>
<tr>
<td>&gt;0.9 (^{(a)})</td>
<td>412_{-380}^{+900}</td>
<td>1140_{-900}^{+190}</td>
<td>1435_{-1200}^{+100}</td>
<td>690_{-250}^{+610}</td>
<td>2860_{-970}^{+100}</td>
</tr>
</tbody>
</table>

Notes. \( \Sigma_{\text{SFR}} \) is given in \( M_\odot \text{yr}^{-1} \text{kpc}^{-2} \). The uncertainties correspond to the 16th and 84th percentiles of the \( \Sigma_{\text{SFR}} \) distribution per \( \Delta \log(SSFR)_{\text{MS}} \) bin. The minimum stellar mass probed throughout the different redshift bins is \( \log(M_*/M_\odot) = 10.0, 10.2, 10.5, 10.5, \) and 10.7. \(^{(a)}\)The highest \( \Delta \log(SSFR)_{\text{MS}} \) bin is centered at \( \Delta \log(SSFR)_{\text{MS}} = 1.2, 1.05, 1.05, 1.45, 1.45 \).