



## Measuring size evolution of distant, faint galaxies in the radio regime

Downloaded from: <https://research.chalmers.se>, 2025-12-05 03:11 UTC

Citation for the original published paper (version of record):

Lindroos, L., Knudsen, K., Stanley, F. et al (2018). Measuring size evolution of distant, faint galaxies in the radio regime. *Monthly Notices of the Royal Astronomical Society*, 476(3): 3544-3554.  
<http://dx.doi.org/10.1093/mnras/sty426>

N.B. When citing this work, cite the original published paper.

# Measuring size evolution of distant, faint galaxies in the radio regime

L. Lindroos,<sup>1</sup> K. K. Knudsen,<sup>1</sup>★ F. Stanley,<sup>1</sup> T. W. B. Muxlow,<sup>2</sup> R. J. Beswick,<sup>2</sup>  
J. Conway,<sup>1</sup> J. F. Radcliffe<sup>2,3,4</sup> and N. Wrigley<sup>2</sup>

<sup>1</sup>Department of Space, Earth and Environment, Chalmers University of Technology, Onsala Space Observatory, SE-439 92 Onsala, Sweden

<sup>2</sup>Jodrell Bank Centre for Astrophysics/e-MERLIN, The University of Manchester, M13 9PL, UK

<sup>3</sup>ASTRON, the Netherlands Institute for Radio Astronomy, Postbus 2, NL-7990 AA Dwingeloo, the Netherlands

<sup>4</sup>Kapteyn Astronomical Institute, University of Groningen, NL-9747 AD Groningen, the Netherlands

Accepted 2018 February 14. Received 2018 February 14; in original form 2016 August 19

## ABSTRACT

We measure the evolution of sizes for star-forming galaxies as seen in 1.4 GHz continuum radio for  $z = 0$ –3. The measurements are based on combined VLA+MERLIN data of the Hubble Deep Field, and using a  $uv$ -stacking algorithm combined with model fitting to estimate the average sizes of galaxies. A sample of  $\sim 1000$  star-forming galaxies is selected from optical and near-infrared catalogues, with stellar masses  $M_{\odot} \approx 10^{10}$ – $10^{11} M_{\odot}$  and photometric redshifts 0–3. The median sizes are parametrized for stellar mass  $M_{*} = 5 \times 10^{10} M_{\odot}$  as  $R_e = A \times (H(z)/H(1.5))^{\alpha_z}$ . We find that the median radio sizes evolve towards larger sizes at later times with  $\alpha_z = -1.1 \pm 0.6$ , and  $A$  (the median size at  $z \approx 1.5$ ) is found to be  $0''.26 \pm 0''.07$  or  $2.3 \pm 0.6$  kpc. The measured radio sizes are typically a factor of 2 smaller than those measure in the optical, and are also smaller than the typical  $H\alpha$  sizes in the literature. This indicates that star formation, as traced by the radio continuum, is typically concentrated towards the centre of galaxies, for the sampled redshift range. Furthermore, the discrepancy of measured sizes from different tracers of star formation, indicates the need for models of size evolution to adopt a multiwavelength approach in the measurement of the sizes star-forming regions.

**Key words:** techniques: interferometric – galaxies: statistics – galaxies: structure – radio continuum: galaxies.

## 1 INTRODUCTION

As galaxies grow in mass over cosmic time, they also grow in size. Several studies use the high angular resolution of the *Hubble Space Telescope* (*HST*) to study high-redshift galaxies, and find that even at a fixed stellar mass ( $M_{*}$ ) the sizes of galaxies grow significantly with time (e.g. Giavalisco, Steidel & Macchetto 1996; van der Wel et al. 2014). In particular, the size evolution of quiescent galaxies with little ongoing star formation has garnered much interest, as the evolution is generally much faster for these galaxies compared to star-forming galaxies (e.g. van der Wel et al. 2014). Several mechanisms have been proposed for why quiescent galaxies evolve in size, such as mergers and mass loss in the centre. Major mergers are generally considered unlikely as they would result in some galaxies with significantly higher masses than local massive elliptical galaxies, as well as the fact that they are rare (Lotz et al. 2011). Minor mergers on the other hand are more frequent, and can inflate the size of galaxies without significantly increasing the mass (e.g. Bezanson et al. 2009; Bezanson, van Dokkum & Franx 2012; Pacifici et al. 2016). Central mass loss of galaxies can cause an

evolution of galactic sizes. Stars move to adapt to the shallower potential, resulting in increase of the galactic size (e.g. Fan et al. 2010).

However, the size evolution of quiescent galaxies is difficult to study in isolation, as star-forming galaxies continuously transform into quiescent through quenching. Some studies have found that there may be no need for size evolution in quiescent galaxies, but that the observed evolution of the median is the result of larger galaxies joining the population due to quenching (e.g. Valentinuzzi et al. 2010; Poggianti et al. 2013). Larger surveys such as SDSS find too few compact massive quiescent galaxies to be consistent with this picture (e.g. Trujillo et al. 2009), but it is clear that the sizes and morphologies of star-forming galaxies before quenching is an important parameter to understand the size evolution of the population of quiescent galaxies.

Tacchella et al. (2016) simulations indicate that sizes of star-forming galaxies evolve rapidly prior to full quenching, due to *inside-out quenching* of star formation. In this process, star formation is quenched first in the centre of the galaxies, leading to higher specific star formation rate (SFR) density at the outskirts of the galaxies. In the Tacchella et al. (2016) model of inside-out quenching, the majority of galaxies in a given mass range will undergo quenching at the same time, which will create an observational

★ E-mail: kirsten.knudsen@chalmers.se

signature if the star formation rate surface density is observed for high-redshift galaxies. Such a signature was detected for a sample of  $z \sim 2$  star-forming galaxies observed in H $\alpha$  (Tacchella et al. 2015), however, the sample of galaxies is small.

In this paper, we aim to provide a statistical picture of the typical morphology of star formation in high-redshift star-forming galaxies. This provides insight into the role star formation plays in the size evolution of star-forming galaxies, and to address models such as the proposed inside-out growth model from Tacchella et al. (2016). We will achieve this by using stacking methods on large samples of galaxies in the Great Observatories Origins Survey North (GOODS-N) field, which has been surveyed by both the Multi-Element Radio-Linked Interferometer Network (MERLIN) and the Very Large Array (VLA) at 1.4 GHz. Faint galaxies in this data set have been studied (e.g. Beswick et al. 2008), however, none have focused on the sizes of the faint galaxies. In this study, we will use the new stacking algorithm developed by Lindroos et al. (2015). This algorithm has been found to be effective to derive sizes (Lindroos et al. 2016), using model fitting in the  $uv$  domain on the stacked sources. Using the long baselines of MERLIN, this will allow us to resolve the star formation within the stacked galaxies.

This paper is structured as follows. In Section 2, we describe the photometric data and our sample selection, as well as the MERLIN and VLA data and reduction. In Section 3, we describe our stacking algorithm and the procedure used to determine sizes. In Section 4, we present the stacking results. In Section 5, we compare our size measurements with earlier measurements and discuss the implications for galaxy evolution. Finally, in Section 6 we summarize the results.

We assume a standard cosmology with  $H_0 = 67.3 \text{ km s}^{-1} \text{ Mpc}^{-1}$ ,  $\Omega_\Lambda = 0.685$ , and  $\Omega_m = 0.315$  (Planck Collaboration XVI, 2014). All magnitudes are in AB (Oke 1974) unless otherwise specified.

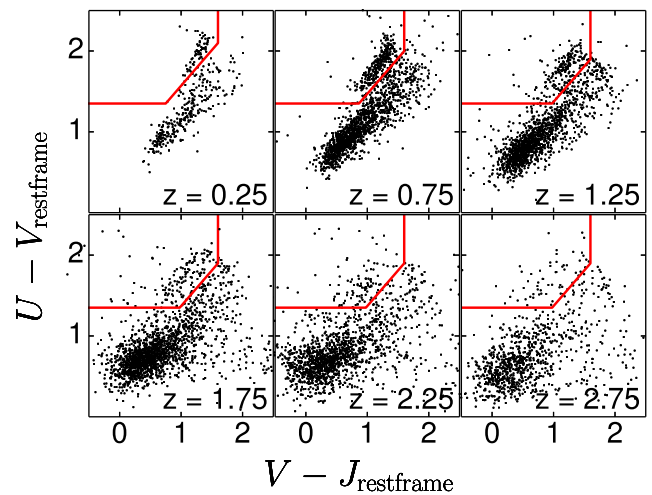
## 2 SAMPLE SELECTION

GOODS-N is a well-studied field, with rich multiwavelength data. The recent 3D-HST survey (Brammer et al. 2012) adds further *HST* near-infrared data to the field. Skelton et al. (2014) combined this data with earlier studies to provide a rich photometric catalogue. For the GOODS-N, the catalogue is based on *HST* observations in 7 bands (F435, F606W, F775W, F850LP, F125W, F140W, F160W), complemented by ground-based observation for 11 bands ( $U$ ,  $B$ ,  $G$ ,  $V$ ,  $R$ ,  $I$ ,  $z$ ,  $J$ ,  $H$ ,  $K$ ), and *Spitzer* observations in the 4 IRAC bands. Sources were identified from the combination of the F125W, F140W, and F160 bands, with a typical resolution of  $0''.18$ .

### 2.1 Redshifts, rest-frame colours, and stellar masses

For the  $\sim 32\,500$  sources identified in the field, 2081 sources have spectroscopic redshifts. For the sources with no spectroscopic redshift, we use photometric redshifts determined by Skelton et al. (2014), using the EAZY code<sup>1</sup> (Brammer, van Dokkum & Coppi 2008). Using these redshifts Skelton et al. (2014) also calculate the rest-frame colours for several bands, including the  $U$ ,  $V$ , and  $J$  band which we use.

By using FAST<sup>2</sup> (Kriek et al. 2009) with the Bruzual & Charlot (2003) stellar-synthesis models and a Chabrier (2003) initial mass



**Figure 1.** Rest frame  $U - V$  and  $V - J$  colours for all sources in the GOODS-N with  $M_* > 10^9 M_\odot$  and  $z < 3$ . The lines indicate the criteria used to separate early-type galaxies from star-forming galaxies, with early-type galaxies in the upper left corner.

**Table 1.** Number of galaxies in each sub sample for star-forming galaxies.

$z$	$\log_{10} M_*/M_\odot$		
	10–10.5	10.5–11	11–11.5
0.0–0.5	42	12	1
0.5–1.0	176	58	12
1.0–1.5	182	71	9
1.5–2.0	163	60	5
2.0–2.5	177	66	21
2.5–3.0	98	31	8

function (IMF), Skelton et al. (2014) estimate the stellar mass, dust attenuation, and SFR for all detected galaxies in GOODS-N. The stellar masses are generally robust, however, the dust attenuation and SFR are uncertain as there is only photometry from optical to near-IR (see e.g. Wuyts et al. 2012).

### 2.2 Main sample

We select all galaxies with  $z < 3$  and  $M_* > 10^{10} M_\odot$  (a total of 1933 galaxies). Following the criteria from Williams et al. (2009), we select the star-forming galaxies, based on rest-frame colours in  $U - V$  and  $V - J$ , shown in Fig. 1. Our study predominantly focuses on galaxies with 1.4GHz flux densities below  $100 \mu\text{Jy}$ , where star-forming galaxies are shown to be the dominant population (e.g. Padovani et al. 2015). However, we exclude 173 galaxies ( $\sim 10$  percent of the sample) that are detected in the *Chandra* survey of the GOODS-N field (Alexander et al. 2003), in order to minimize contamination from active galactic nuclei (AGNs) (e.g. Richards et al. 2007). The final sample is further sub-divided based on  $M_*$  and  $z$ , resulting in a total of 18 subsamples with number of sources according to Table 1.

### 2.3 sBzK sample

For comparison to our main sample, as well as to previous results obtained with ALMA (for BzK galaxies; Lindroos et al. 2016), we select a sample of sBzK galaxies, selected from the Skelton et al. (2014) catalogue. The sBzK sample was selected based on the criteria of Daddi et al. (2004),  $(B - z) - (z - K) \geq -0.2$

<sup>1</sup> The Easy and Accurate Redshift from Yale (EAZY), <https://github.com/gbrammer/ezay-photoz/>.

<sup>2</sup> <http://w.astro.berkeley.edu/mariska/FAST.html>

and  $K < 24$ . sBzK galaxies selected in this manner are actively star-forming galaxies (with SFRs  $\sim 200 M_{\odot} \text{yr}^{-1}$ ), at redshifts of  $1.4 < z < 2.5$ , that due to dust reddening may have not been selected by a rest-frame UV selection (e.g. Daddi et al. 2004).

An additional sBzK sample available is that of Yuma et al. (2011). This sample was selected differently from the one above, using photometry extracted based on  $K$ -band detections. If we limit this sample to the redshift range 1.4–2.5, we find that it has an 80 per cent overlap in sources with our sBzK selection. Furthermore, all except six sources of the Yuma et al. (2011) sample have at least one counterpart in the Skelton et al. (2014) catalogue within 1 arcsec. Since the two samples have a large overlap in sources, and Skelton et al. (2014) uses additional data for the  $z$ -band photometry which was not used by Yuma et al. (2011), we use the sBzK sample selected from the Skelton et al. (2014) catalogue described above.

### 3 RADIO DATA

We study the radio continuum emission of our sample using combined data from VLA and MERLIN.

The MERLIN data are described in detail in Muxlow et al. (2005). The data were observed using the MERLIN telescope from February 1996 to April 1997, for a total of 18.23 days. The survey covers  $\sim 100 \text{ arcmin}^2$  centred on the GOODS-North ( $12^{\text{h}}36^{\text{m}}49.4, +62^{\circ}12'58''00$  in J2000). The data is centred on 1.42 GHz with a bandwidth of 16 MHz (32 channels in dual circular polarization), with a typical depth of  $\sim 7 \mu\text{Jy beam}^{-1}$ .

The MERLIN data were combined with VLA data from Richards (2000), with a total of 50 h of observations. The VLA data was observed in the A configuration, and was centred on a frequency of 1.4 GHz with a typical bandwidth of 44 MHz (14 channels in dual polarization).

As part of the original data analysis by Muxlow et al. (2005) the brightest sources in the field were imaged. All sources brighter than 0.5 mJy were cleaned and the models removed from the data. Also several bright nearby sources outside the field in the VLA data were also imaged, and removed from the data to reduce the impact on the noise. Muxlow et al. (2005) report a naturally weighted beam size of  $0''.2$  in the centre of the field for the combined data. In this study, however we primarily rely on model fitting in the  $uv$  domain to measure sizes, where the accuracy depends on the signal-to-noise ratio (SNR) of the source. Based on the beam size we expect an accuracy of  $\sim 0''.04$  for the size of a source with  $\text{SNR} = 10$  (Martí-Vidal, Pérez-Torres & Lobanov 2012).

#### 3.1 Uncertainties of alignment between optical and radio positions

To measure the offset between the optical Skelton et al. (2014) catalogue positions, and the coordinate system of our radio observations, we study the positions of our main sample in the VLA and MERLIN maps. Most of the sources are faint as expected and, in this analysis, we focus on sources with SNR greater than 5. This results in a sample of 43 and 7 sources for the VLA and MERLIN, respectively. For each source, we fit the position in VLA and MERLIN data using model fitting in the  $uv$ -domain.

We measure the median offset in right ascension and declination, and estimate the uncertainties using case resampling (bootstrapping, see section 4.4). Based on this we find the magnitude of the median offset to be  $0''.0 \pm 0''.04$  for VLA and  $0''.0 \pm 0''.03$  for MERLIN. We also perform case resampling using the standard deviation to measure random offsets between radio and optical positions. Such

offsets can result if the radio emission does not originate from the same region of a galaxy as the optical emission. We find that the random offsets are  $0''.21 \pm 0''.02$  for VLA and  $0''.08 \pm 0''.02$  for MERLIN. This is similar to the expected uncertainties for the fitted radio positions at the given SNR (Martí-Vidal et al. 2012), indicating that the typical misalignment between the optical and radio emission is very small.

#### 3.2 Primary beam model

To account for primary-beam attenuation in this work we use two different primary-beam models. The VLA data is corrected using the VLA primary beam model present in CASA version 4.4, that makes use of an Airy disc model. For the MERLIN data, the primary beam is modelled as a third-order polynomial with

$$P(l, m) = \left( \sum_{k=0}^3 g_k \left( \frac{(l^2 + m^2)v^2}{\text{arcmin}^2 \text{GHz}^2} \right)^k \right)^{-1}, \quad (1)$$

where  $g_0 = 0.0$ ,  $g_1 = -3.237 \times 10^{-3}$ ,  $g_2 = 6.286 \times 10^{-6}$ ,  $g_3 = -5.278 \times 10^{-9}$ ,  $(l, m)$  are directional cosine relative to the pointing centre, and  $v$  is the frequency (Wrigley 2011; Wrigley et al. in preparation). The typical primary beam attenuation correction applied to these data is 6 per cent and 9 per cent for the VLA and MERLIN data, respectively.

### 4 STACKING ROUTINE AND MODELLING

We use the  $uv$ -stacking approach described in Lindroos et al. (2016), which performs all operations directly on the visibility data. The stacking is performed using the  $uv$ -stacking algorithm described Lindroos et al. (2015). This allows us to combine the data from VLA and MERLIN in the  $uv$  domain. This is especially important when combining two arrays with different  $uv$  coverages, as the  $uv$  stacking approach allows us to apply correct weighting to all baselines after stacking.

#### 4.1 Subtraction of bright sources

Bright sources can significantly impact the flux density of faint sources in interferometric data. To ensure that the stacked results are robust, bright sources are modelled and removed from the data. Sources brighter than 0.5 mJy were individually modelled in the original data reduction, and these sources were removed as part of the original reduction for both VLA and MERLIN. To further decrease the impact of bright sources, the field was imaged using CLEAN in Common Astronomy Software Applications package<sup>3</sup> (CASA) and cleaned down to a threshold of  $50 \mu\text{Jy beam}^{-1}$ , and the clean model was subtracted from the  $uv$  data. This results in a noise in the centre of the field of  $\sim 5 \mu\text{Jy beam}^{-1}$ . The subtracted sources were checked against the positions of our main sample to ensure that no target sources were subtracted.

After subtracting the bright source, we performed a stack of 200 random positions to check for residuals of bright sources which could influence the stacking. The average visibility in the MERLIN baselines was  $0.28 \mu\text{Jy}$ , indicating that the impact of residuals on the data is not systematically increasing or decreasing flux density in stacking.

<sup>3</sup> <http://casa.nrao.edu>

#### 4.2 $uv$ stacking method

The stacking method used in this work is based on moving the target sources into the phase centre of the data, correcting for primary-beam attenuation, and combining the sources in the  $uv$  plane. The primary-beam attenuations were corrected for using the models described in 3.2. The algorithm does not duplicate visibilities for each source, but adds the phase rotated and primary-beam corrected visibilities directly. This results in a new  $uv$  data set with the same  $uv$  coverage as the original data, but with significantly lower noise and the stacked source located at the phase centre. At this point, the stacked source can either be imaged using interferometric techniques, or modelled directly in the  $uv$  domain. A more detailed explanation can be found in Lindroos et al. (2015). In this work, we use a model fitting approach to achieve maximum accuracy on the sizes of the sources.

#### 4.3 Model fitting

To estimate the flux densities and sizes of our sources from the stacked data, we use model fitting. Before model fitting the data are binned by baseline length into 100 bins with an equal number of visibilities in each bin. VLA and MERLIN data is binned separately resulting in a total of 200 bins. Each model is fitted by minimizing  $\chi^2$  computed as

$$\chi^2(\text{sample}, V_{\text{model}}) = \sum_{i=1}^{200} W_i (V_i - V_{\text{model},i})^2, \quad (2)$$

where  $V_i$  is the average visibility in bin  $i$  for the stacked data of the sample,  $V_{\text{model}}$  is the model used,  $V_{\text{model},i}$  is the model evaluated for the median baseline length of bin  $i$ , and  $W_i$  is the sum of the corresponding visibility weights.

The Sérsic model is typically used to measure optical sizes of galaxies, and we want to use this distribution as we are interested in comparing our radio sizes to the optical sizes. However, the Sérsic profile is designed for an individual galaxy, while in this study we are dealing with stacked measures. To account for the variation of sizes within each sample, we convolved the Sérsic profile with a distribution of sizes. Based on the sizes measured by van der Wel et al. (2014) at  $1.25\mu\text{m}$ , we studied the distribution of sizes for our sample. The distribution were found to be well fitted by a Rayleigh distribution. This results in a Rayleigh–Sérsic model defined as

$$V_{\text{rs}(\Phi, \sigma, n)}(u, v) = \Phi \int_0^\infty R_\sigma(t) V_s(u, v, t, n) dt, \quad (3)$$

where  $\Phi$  is the total flux density,  $(u, v)$  are the projected baselines,  $\sigma$  is the source size,  $R_\sigma(t) (= t/\sigma^2 e^{-t^2/\sigma^2})$  is a Rayleigh distribution, and  $V_s(u, v, R_e, n)$  is the Fourier transform of a Sérsic intensity profile with effective radius  $R_e$  and Sérsic index  $n$ . This model has three free parameters:  $\Phi$ ,  $\sigma$ , and  $n$ . We find the best fit for these three parameters using the model fitting algorithm described here. From the parameter  $\sigma$ , we can calculate the median size of the fitted galaxies as

$$\text{median}(R_e) = \sqrt{2 \ln 2} \sigma. \quad (4)$$

To simplify comparison with other results, we report the median effective radius in place of  $\sigma$ .

#### 4.4 Estimates of uncertainties

The uncertainties were primarily estimated using the bootstrapping method, using a Monte Carlo variant of the case resampling method

(Chihara & Hesserberg 2011). This estimates the uncertainties of a given measure  $f$  (such as the median), for  $N$  data points  $x_1, x_2, \dots, x_N$ . In the paper, we perform the case resample as follows:

- (i) Resample the data points with replacement, i.e. select  $N$  random integers  $(l_1, l_2, \dots, l_N)$  between 1 and  $N$ .
- (ii) Calculate the measure for the resampled set:  $Y = f(x_{l_1}, x_{l_2}, \dots, x_{l_N})$ .
- (iii) Repeat step (i) and (ii) for 1000 times to obtain a set  $Y_1, Y_2, \dots, Y_{1000}$ .

We refer to  $Y_1, Y_2, \dots$  as the bootstrapping distribution. From the bootstrapping distribution, we calculate a cumulative distribution function (CDF), and we report error range as where the CDF crosses 0.159 and 0.841, equivalent to  $\pm 1\sigma$  for a normal distribution.

The bootstrapping method is performed for each sub-sample with a stacked peak SNR  $>4$ . The resampling is performed over the sources of each sample, and the stacking and model fitting is done for each resampled set of sources.

An advantage with the bootstrapping method is that it will include errors from sample variance in the estimate. If the variance in the population is large and the sample size is small this can contribute to higher noise. We compare the bootstrapping distributions to the  $\chi^2$  distribution (see appendix A), in order to establish if sample variance is an important effect in the data. The similar results between the two error estimates indicate that variance within the samples is not an important factor.

#### 4.5 Star formation rate

Our sample is selected to exclude most AGN, and as such the bulk of the flux density is assumed to originate from star formation. To calculate the SFR of our stacked galaxies, the measured flux densities are converted to luminosities with a  $K$ -correction, where the radio emission is assumed to have a spectral index of  $-0.8$  (e.g. Condon 1992), i.e.

$$L_{1.4\text{ GHz}} = S_{1.4\text{ GHz}} 4\pi d_L^2(z) \times (1+z)^{0.2}, \quad (5)$$

where  $L_{1.4\text{ GHz}}$  is the 1.4 GHz rest-frame luminosity,  $S_{1.4\text{ GHz}}$  is the 1.4 GHz observer-frame flux density, and  $d_L$  is the luminosity distance. The 1.4 GHz luminosity is converted to SFR using the Bell (2003) correlation scaled to a Chabrier (2003) IMF:

$$\text{SFR} = k_{\text{SFR}} L_{1.4\text{ GHz, restframe}}, \quad (6)$$

where  $k_{\text{SFR}} = 3.18 \times 10^{-22} \text{M}_\odot \text{yr}^{-1} (\text{W Hz}^{-1})^{-1}$ . Bell (2003) argue that low-luminosity galaxies need to be treated separately; however, none of the derived luminosities in this study fall below their low-luminosity cut-off ( $L_c = 6.4 \times 10^{21} \text{W Hz}^{-1}$ ).

#### 4.6 Combined model for evolution

In this paper, we aim to measure how the sizes and flux densities (as a tracer of the SFR), of the galaxies evolve with redshift and stellar mass. To do this, we fit combined models for the size ( $R_e(z, M_*)$ ) and SFR ( $\text{SFR}(z, M_*)$ ), respectively, that include evolution with redshift. Based on the typical evolutionary models used for optical sizes (e.g. van der Wel et al. 2014), we parametrize the combined model for size as:

$$R_e(z, M_*) = A \left( \frac{M_*}{5 \times 10^{10} \text{M}_\odot} \right)^{\alpha_M} \left( \frac{H(z)}{H(1.5)} \right)^{\alpha_z}, \quad (7)$$



**Table 2.** Fitted flux density ( $\Phi$ ) in bins of stellar mass and redshift. The fitted parameters are estimated using a Rayleigh–Sérsic model (described in section 4.3), and errors are estimated using bootstrapping.

$z$	$\log_{10} M_*/M_\odot$		
	10–10.5	10.5–11	11–11.5
0.0–0.5	$12.4^{+4.3}_{-2.7}$ $\mu\text{Jy}$	$75.2^{+61.3}_{-39.6}$ $\mu\text{Jy}$	$144.3^{+146.6}_{-114.5}$ $\mu\text{Jy}$
0.5–1.0	$7.7^{+1.4}_{-1.2}$ $\mu\text{Jy}$	$13.4^{+1.6}_{-1.5}$ $\mu\text{Jy}$	$15.7^{+7.2}_{-6.7}$ $\mu\text{Jy}$
1.0–1.5	$5.3^{+1.3}_{-1.4}$ $\mu\text{Jy}$	$11.4^{+1.6}_{-1.5}$ $\mu\text{Jy}$	$20.0^{+3.2}_{-4.5}$ $\mu\text{Jy}$
1.5–2.0	$2.9^{+1.3}_{-0.9}$ $\mu\text{Jy}$	$9.1^{+1.6}_{-2.1}$ $\mu\text{Jy}$	$18.0^{+5.7}_{-6.6}$ $\mu\text{Jy}$
2.0–2.5	$2.7^{+10.0}_{-1.3}$ $\mu\text{Jy}$	$8.0^{+1.8}_{-2.0}$ $\mu\text{Jy}$	$23.7^{+5.3}_{-4.3}$ $\mu\text{Jy}$
2.5–3.0	$3.5^{+11.4}_{-2.3}$ $\mu\text{Jy}$	$8.8^{+1.6}_{-1.7}$ $\mu\text{Jy}$	$21.0^{+6.3}_{-5.9}$ $\mu\text{Jy}$

where  $H(z)$  is the Hubble parameter at redshift  $z$ , and  $A$ ,  $\alpha_M$ , and  $\alpha_z$  are free parameters. The combined model for SFR is modelled in parallel as

$$\text{SFR}(z, M_*) = B \left( \frac{M_*}{5 \times 10^{10} M_\odot} \right)^{\beta_M} \left( \frac{z+1}{2.5} \right)^{\beta_z}, \quad (8)$$

where  $B$ ,  $\beta_M$ , and  $\beta_z$  are free parameters.

We fit the above to the median  $z$  ( $z_i$ ), and median  $M_*$  ( $M_j$ ), of each subsample. Where indexes  $i$  and  $j$ , correspond to the redshift bin and  $M_*$  bin, respectively, and can be described as  $0.5(i-1) < z < 0.5i$  and  $10^{10+0.5(j-1)} < M_* < 10^{10+0.5j}$  (e.g. (i,j)=(1,1) corresponds to subsample in redshift bin  $z = 0-0.5$  and  $M_*$  bin  $\log_{10} M_*/M_\odot = 10-10.5$ , see Table 1).

Making use of equations (2) and (3), we calculate a total  $\chi^2$  for the main sample (MS):

$$\chi^2 = \sum_{i=1}^6 \sum_{j=1}^3 \chi^2(\text{MS}_{i,j}, V_{\text{rs}}(\Phi_{i,j}, \sigma_{i,j}, 1)), \quad (9)$$

where  $\Phi_{i,j}$  is the flux density given by the equation:

$$\Phi_{i,j} = \frac{\text{SFR}(z_i, M_j)}{4\pi(z_i + 1)^{0.2} d_L^2(z_i) k_{\text{SFR}}}, \quad (10)$$

and  $\sigma_{i,j}$  is the size:

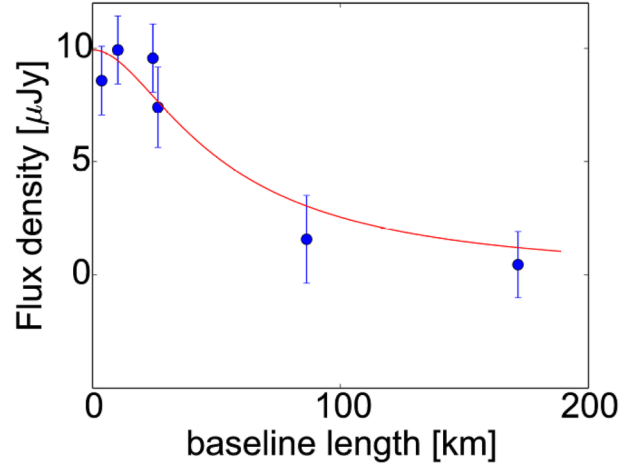
$$\sigma_{i,j} = \frac{R_c(z_i, M_j)}{\sqrt{2 \ln 2} d_A}, \quad (11)$$

where  $\chi^2(\text{MS}_{i,j}, V_{\text{rs}}(\Phi_{i,j}, \sigma_{i,j}))$  is  $\chi^2$  for sub-sample  $\text{MS}_{i,j}$ ,  $d_A$  is the angular distance at  $z_i$ , and  $d_L$  is the luminosity distance at  $z_i$ .

We exclude those sub-samples that have peak SNR below 4, these correspond to  $(i, j) = (1, 2), (1, 3), (2, 3), (4, 3), (5, 1)$ , and  $(6, 1)$  (see Table 3). The Sérsic index  $n$  is fixed to 1, based on the typical brightness profile of star-forming galaxies at optical wavelengths (van der Wel et al. 2014). We also fix  $\alpha_M = 0.22$ , based on optical observations of the sizes of star-forming galaxies (e.g. Morishita, Ichikawa & Kajisawa 2014; van der Wel et al. 2014). This results in a model with five free parameters:  $\alpha_z$ ,  $\beta_z$ ,  $\beta_M$ ,  $A$ , and  $B$ . Details on the error estimates for the combined models are given in Appendix A.

## 5 STACKING RESULTS

The measured sizes and flux densities for each bin of the main sample using the Rayleigh–Sérsic model can be found in Tables 2 and 3. A typical model fit to the stacked data is shown in Fig. 2. The Sérsic index  $n$  was fixed to 1 for each subsample. Out of our main sample, six subsamples have a peak brightness lower than four times the noise. The fitted sizes for these low SNR subsamples, given in



**Figure 2.** A typical model fit for a stacked data set, shown for the stacked set where  $1.5 < z < 2.0$  and  $10^{10.5} M_\odot < M_* < 10^{11} M_\odot$ . Visibilities are binned as a function of baseline length. The red line shows the Rayleigh–Sérsic fit to the data. The visibilities in the image are binned in 3 bins each for VLA and MERLIN, while the model was fit to the full 200 bins.

parenthesis in Table 3, are significantly less robust. These sources are excluded from further analysis.

The bootstrapping distributions and  $\chi^2$  plots are shown in appendix A. For the samples with  $\text{SNR} > 4$ , the error estimates show similar results. This indicates that the sample size is sufficiently large to draw conclusions on the average properties of the larger population of star-forming galaxies.

### 5.1 Evolution with redshift

Using the combined model described in Section 4.6, we estimate the evolution of sizes. In Figs 3 and 4, we show the modelled evolution for size and flux density for our main sample, overlaid with the results for each subsample. The main result of this paper is  $A = 2.3^{+0.6}_{-0.5}$  kpc and  $\alpha_z = -1.1^{+0.7}_{-0.5}$ . This indicates that radio sizes evolve with redshift, and that typical sizes in the radio regime are approximately half of those measured in the optical regime (e.g. van der Wel et al. 2014).

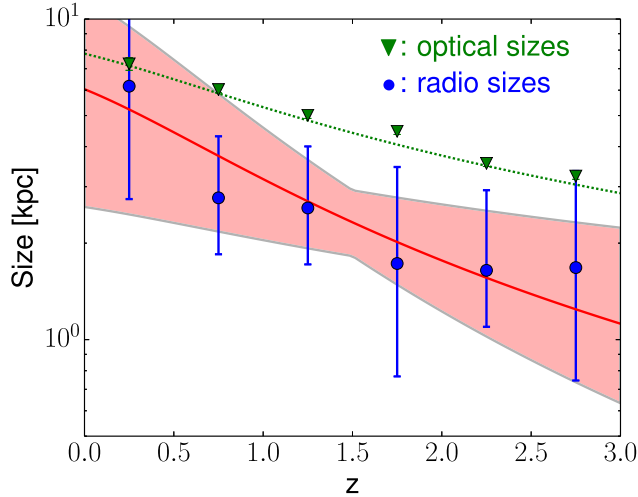
We also fit the SFR evolution in parallel and find  $B = 56 \pm 3 M_\odot \text{ yr}^{-1}$ ,  $\beta_M = -0.2 \pm 0.1$ , and  $\beta_z = 3.2 \pm 0.4$ . This is consistent with previous measures of the SFR evolution for similar galaxies. For comparison, we show Whitaker et al. (2014) and Karim et al. (2011) in Fig. 4.

If we assume that SFR of our galaxies follow the measurements from Karim et al. (2011), we can fit a model where only the parameters of equation 7 are allowed to vary. In this case, we can allow  $\alpha_M$  to vary as the data is sufficient to constrain the three parameters. Doing this, we find  $A = 2.18 \pm 0.3$  kpc,  $\alpha_M = 0.1 \pm 0.2$ , and  $\alpha_z = -0.6^{+0.5}_{-0.4}$ . This value for  $\alpha_z$  is somewhat smaller than the previously found value; however, within statistical uncertainties it is consistent with both our measurements, and the optical measurements from van der Wel et al. (2014) and Morishita et al. (2014).

In the model fits, we fixed the Sérsic index  $n$  to 1, however, if we change  $n$  it does not significantly impact the result. For example, with  $n = 2.0$  we find  $A = 2.3^{+0.5}_{-0.4}$  and  $\alpha_z = -1.2^{+0.8}_{-0.7}$ , for the combined model. For comparison, we also performed a fit where the Rayleigh–Sérsic profile was replaced by a simple Gaussian, which is commonly used at radio and sub-millimetre (submm) wavelengths to estimate sizes. This results in  $A = 2.0^{+0.5}_{-0.4}$  kpc and

**Table 3.** Fitted sizes ( $R_e$ ) in bins of stellar mass and redshift, both angular size and linear size calculated based on median redshift in each bin. The fitted parameters are estimated using a Rayleigh–Sérsic model (described in Section 4.3), and errors are estimated using bootstrapping. The numbers within parenthesis indicate subsamples with peak SNR  $< 4$ .

$z$	$10-10.5$		$\log_{10} M_*/M_\odot$ 10.5–11		11–11.5	
0.0–0.5	$0''.47^{+0.17}_{-0.25}$	$1.91^{+0.69}_{-1.01}$ kpc	$(1''.97^{+2.12}_{-0.95})$	$(7.99^{+8.60}_{-3.85}$ kpc)	$(2''.3^{+1.7}_{-1.9})$	$(9.33^{+6.89}_{-7.71}$ kpc)
0.5–1.0	$0''.28^{+0.13}_{-0.11}$	$2.12^{+0.99}_{-0.83}$ kpc	$0''.29^{+0.06}_{-0.06}$	$2.20^{+0.45}_{-0.45}$ kpc	$(0''.33^{+0.15}_{-0.11})$	$(2.51^{+1.14}_{-0.83}$ kpc)
1.0–1.5	$0''.14^{+0.15}_{-0.07}$	$1.20^{+1.29}_{-0.60}$ kpc	$0''.26^{+0.07}_{-0.07}$	$2.23^{+0.60}_{-0.60}$ kpc	$0''.15^{+0.06}_{-0.04}$	$1.29^{+0.52}_{-0.34}$ kpc
1.5–2.0	$0''.21^{+0.35}_{-0.11}$	$1.83^{+3.04}_{-0.96}$ kpc	$0''.16^{+0.07}_{-0.05}$	$1.37^{+0.61}_{-0.43}$ kpc	$(0''.24^{+0.15}_{-0.09})$	$(2.09^{+1.30}_{-0.78}$ kpc)
2.0–2.5	$(0''.66^{+3.56}_{-0.57})$	$(5.58^{+30.1}_{-4.82}$ kpc)	$0''.13^{+0.07}_{-0.04}$	$1.10^{+0.59}_{-0.34}$ kpc	$0''.35^{+0.17}_{-0.10}$	$2.96^{+1.44}_{-0.85}$ kpc
2.5–3.0	$(0''.85^{+3.68}_{-0.56})$	$(6.88^{+29.8}_{-4.53}$ kpc)	$0''.15^{+0.04}_{-0.03}$	$1.21^{+0.32}_{-0.24}$ kpc	$0''.29^{+0.05}_{-0.05}$	$2.34^{+0.40}_{-0.40}$ kpc



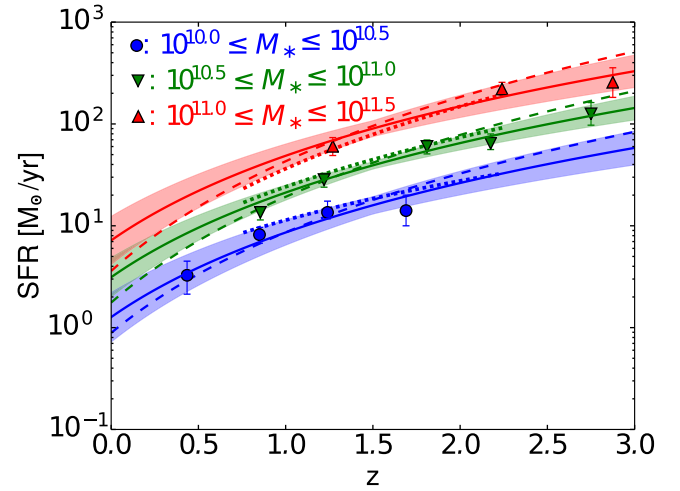
**Figure 3.** Sizes measured for star-forming galaxies at a stellar mass of  $5 \times 10^{10}$ . The blue circles indicate sizes measured at 1.4 GHz and the green squares indicate the size measurements from van der Wel et al. (2014) at rest-frame wavelength of 500 nm. The red solid line indicate the best fit to the 1.4 GHz sizes and the green dotted line indicate the best fit to the optical sizes from van der Wel et al. (2014). The red area indicates the  $1\sigma$  uncertainties of the fitted model.

$\alpha_z = 1.0 \pm 0.6$ . Note that this is the HWHM size of the Gaussian, which no longer corresponds directly the median sizes of the sources in the sample. However, the small difference to median effective radius indicates that the choice of profile is not highly significant for the main result.

## 5.2 sBzK-selected galaxies

To allow direct comparison with other studies we also stack and model sBzK galaxies. We find a flux density of  $2.1^{+0.5}_{-0.3}$   $\mu$ Jy and a size of  $0''.23^{+0.16}_{-0.09}$  for the full sBzK comparison sample. If we limit the sample to  $K < 21$ , we find  $\Phi = 5.7^{+0.7}_{-0.7}$   $\mu$ Jy and  $\sigma = 0''.23^{+0.09}_{-0.09}$ , and if we limit the sample to  $K < 20$  we find  $\Phi = 13.5^{+2.8}_{-1.3}$   $\mu$ Jy and  $\sigma = 0''.29^{+0.13}_{-0.06}$ . If we limit the sample to  $K < 20$ , we have a median  $z$  of 1.83 and a median  $M_*$  of  $10^{10.8} M_\odot$ , this results in a physical size of  $2.4 \pm 1.0$  kpc. This is similar to the main sample, where the modelled size is 2.0 kpc at  $z = 1.83$  and  $M_* = 10^{10.8} M_\odot$ .

To compare these results with the stacked sizes measured at submm wavelength using the Atacama Large Millimetre/sub-millimetre Array (ALMA) by Lindroos et al. (2016), we measured the sizes using a Gaussian profile for the sBzK sample with



**Figure 4.** Estimated SFRs for each redshift and stellar mass bin of the main sample with errors estimated from variations in  $\chi^2$ . The solid line shows our fitted model, with the areas in corresponding colour indicating the typical  $1\sigma$  uncertainties of the models. The dashed line indicates the SFR estimated by Karim et al. (2011). The stellar mass bins used by Karim et al. (2011) are a little different from the bins used in this work, to account for this the SFR was scaled to the median stellar mass of each bin with a spectral index of 0.6. The dotted lines indicate the SFR estimates from Whitaker et al. (2014) for the median stellar mass of each bin. The redshifts of the bins are centred on the median redshift of the galaxies in each bin.

$K_{\text{VEGA}} < 20$ :  $\Phi = 12.9^{+1.6}_{-1.6}$   $\mu$ Jy and size =  $0''.34^{+0.11}_{-0.11}$ . This is the same method used in Lindroos et al. (2016) and the same sample selection, where they found the HWHM size to be  $0''.35 \pm 0''.1$ .

## 6 DISCUSSION

The stacking results presented show that the radio sizes follow a similar evolution to the optical sizes, with significantly smaller sizes at higher redshift. They also show that the radio sizes are typically around half compared to the optical sizes. In this section, we discuss what properties of star-forming galaxies can be determined from these measurements, as well as implications for galaxy evolution.

### 6.1 Radio continuum as a tracer of the distribution of star formation

As an integrated measure for the SFR, radio continuum has been shown to work well across a wide range of redshifts (e.g. Condon 1992; Karim et al. 2011). We note that our measured SFR evolution

agrees well with previous measurements for the same sample, i.e. the Whitaker et al. (2014) measurements based on *HST*/WFC3 imaging combined with stacked *Spitzer*/MIPS 24  $\mu\text{m}$  data.

For the measured sizes of the star-forming galaxies, there are several factors that could influence the result.

(i) The SFR surface density of the galaxies is assumed to follow a Sérsic brightness profile with  $n = 1.0$ . This is consistent with the existing measurements of the SFR surface density for star-forming galaxies (e.g. Tacchella et al. 2015). We note that the size measured is not very sensitive to the choice of model. Changing the Sérsic index to 2.0, the typical fitted sizes change to within 10 per cent. Using a Gaussian profile the measured result is equivalent, with the median effective radius of the Sérsic profile comparable to the HWHM.

(ii) The radio synchrotron emission is emitted when electrons travel through the magnetic field in the galaxy. If the density of the interstellar medium is low, this will typically imply a low magnetic field strength (e.g. Yoast-Hull, Gallagher & Zweibel 2015). This could allow the electrons to escape from the star formation region. The escape fraction from the galaxy can not be very high, or we would expect the radio continuum to fail as a SFR tracer. However, it is possible that the electrons deposit the energy further from the centre compared to where they originate. Murphy et al. (2008) studied local galaxies and found that the radio surface density was equal to the FIR surface density convolved with a kernel with radius 1–2 kpc.

(iii) The stacking positions are based on *HST*  $z$ -band images. If the *HST* coordinates are offset compared to the radio positions, this could lead to errors in the estimated sizes. Based on brightest sources in the sample we quantify this effect (see Section 3.1), and the effect is found to be small.

(iv) It is possible that the optical  $z$ -band emission is not centred on the same position as the radio emission. Such mis-alignment would be random in direction, and could lead to an increase in the measured sizes in stacking. This effect is also quantified for the brightest sources in the sample (see Section 3.1), and the effect is found to be small compared to the measured sizes.

All the above mentioned effects would typically cause the measured sizes to be too large compared to the true sizes. An upper limit on the effect is provided by the smallest size measured by our method (0'.13).

## 6.2 Size evolution

Our model fits show evidence for size evolution of the average radio sizes from  $z = 3$  to  $z = 0$ . Previous measurements of the size evolution exists at optical wavelength. For example, van der Wel et al. (2014) measured the sizes of star-forming galaxies using a very similar sample to the main sample in this study. They use *HST* to determine the size at 500 nm and found  $R_e = 7.8 \left( \frac{H(z)}{H_0} \right)^{-0.66}$  kpc. Within statistical uncertainties our fitted  $\alpha_z = -1.1 \pm 0.6$  is consistent with this evolution. However, the slope of the evolution is not well constrained by our data.

The radio size measurements of this study is based on a single observer-frame frequency, which results in a considerable difference in rest-frame frequency for the highest and lowest redshifts. Recent studies indicate that radio sizes of  $\mu\text{Jy}$  sources are significantly smaller at  $\sim 10$  GHz compared to at  $\sim 1.4$  GHz (Murphy et al. 2017). However, the sizes of submm sources measured at 3 GHz are likely from the same parent distribution as the sizes at  $\sim 1.4$  GHz

(Miettinen et al. 2017). The above results may indicate that there is a dependency of size measurements on the rest-frame frequency in which they are measured, and this may be a stronger feature in fainter galaxies and/or at larger frequency separations. The highest rest-frame frequency (5.7 GHz at  $z \approx 3$ ) of this study is significantly below the  $\sim 10$  GHz of the Murphy et al. (2017) study; however, it is possible that the size measured at  $z = 3$  in this study is smaller than if a rest-frame frequency of 1.4 GHz had been used. This could lead to a the size evolution measured by our model, that is stronger compared to the true evolution of the sizes of the star-forming region in the galaxies. Currently, we do not have sufficient data to quantify this effect, and future studies spanning a larger part of the radio spectrum will be important. Similar size corrections are applied to sizes measured at 500 nm, where colour gradients cause galaxies to appear smaller at shorter wavelengths (Szomoru et al. 2011; Wuyts et al. 2012). For optical sizes, the colour gradients and the corrections vary depending on galaxy selection (van der Wel et al. 2014), indicating the importance for such corrections to be calibrated for galaxies with the same selection criteria as used in the study.

While the evolution of the radio sizes is not strongly constrained by our data, the average size at  $z > 1$  is much better constrained, with a median size of  $2.3 \pm 0.5$  kpc at  $z = 1.5$ . Compared to the sizes measured by van der Wel et al. (2014), this is approximately half ( $7.8 \left( \frac{H(1.5)}{H_0} \right)^{-0.66} = 4.4$  kpc). Morishita et al. (2014) also measured the sizes of star-forming galaxies at 500 nm, and report somewhat smaller sizes. They found a median effective radius of  $3.09 \pm 0.75$  kpc for star-forming galaxies with  $z$  between 1.25 and 1.5 and stellar masses greater than  $10^{10.5} M_\odot$ . In general, the radio sizes are significantly smaller than the 500 nm sizes.

## 6.3 Comparison to previous size measurements of star formation

Previous studies of sizes at radio wavelength have been limited to galaxies with higher flux densities. Biggs & Ivison (2008) measured the sizes of 12 sub-millimetre galaxies (SMGs) with 1.4 GHz flux densities greater than  $50 \mu\text{Jy}$ , and a median size (HWHM) of  $0'.32 \pm 0'.1$  or 2.5 kpc for the 8 SMGs with photometric redshift. Considering the typical stellar masses of SMGs are around  $7 \times 10^{10} M_\odot$  (e.g. Hainline et al. 2011), this is in close agreement with our size measurement. Other size measurements of the extent of the star formation in SMGs have been performed using submm emission. Simpson et al. (2015) measured the sizes of a sample of 23 SMGs and found a median size of  $1.2 \pm 0.1$  kpc (HWHM) and Ikarashi et al. (2015) measure the size of a sample of 13 SMGs and found a median size of  $0.67 \pm 0.13$  kpc (effective radius). We can conclude the typical sizes measured for SMGs are generally smaller compared to most galaxies from the full population of star-forming galaxies at similar redshifts and stellar masses (see also Rujopakarn et al. 2016). However, this is not very surprising as the mode of star formation in SMGs is generally believed to be very different (e.g. Hainline et al. 2011).

Lindroos et al. (2016) measured the sizes of star-forming galaxies selected with sBzK criteria, using the same stacking technique as this paper. They found a typical size (HWHM) of  $0'.35 \pm 0'.07$  or  $3.0 \pm 0.7$  kpc at the median  $z \approx 2$ . We here measure the radio sizes for sBzK detected at various depths in  $K$  band. If we limit our sample to  $K_{\text{VEGA}} > 20$ , which results in a selection criteria which is identical to the one used in Lindroos et al. (2016), we find a typical size of  $2.4 \pm 1.0$  kpc (FWHM). This agrees within the uncertainties.



Several measurements exist using  $H\alpha$  to probe ongoing star formation in high-redshift galaxies. These surveys have typically found that sizes derived using  $H\alpha$  maps ( $H\alpha$  sizes) are larger than optical sizes for massive star-forming galaxies with  $M_* > 10^{10} M_\odot$  (e.g. Nelson et al. 2013; Wuyts et al. 2014; Nelson et al. 2016; Tacchella et al. 2015). In particular, Nelson et al. (2016) derive  $H\alpha$  sizes using stacking on a sample of star-forming galaxies with  $z \approx 1$ , and find  $H\alpha$  sizes larger or equal to optical sizes for  $M_* \approx 10^9$ – $10^{11} M_\odot$ . Dust attenuation in galaxies is anti-correlated with radius (Wuyts et al. 2012), which could cause  $H\alpha$  sizes to be too large compared to sizes derived from a star formation tracer less affected by dust attenuation. Nelson et al. (2016) argue that this effect is likely weak at the low mass end as dust attenuation is less for lower mass galaxies, however, this could contribute to the difference seen between radio and  $H\alpha$  sizes for star-forming galaxies with  $M_* \approx 5 \times 10^{10} M_\odot$ . Interestingly, the submm sizes of  $H\alpha$  selected star-forming galaxies were found to be extremely compact, at half their optical sizes, and comparable to those of SMGs (Tadaki et al. 2017).

Looking at the shape of the radial profiles of surface brightness, Nelson et al. (2016) is in good agreement with our results, finding that on average the radial profiles are well fitted by smooth Sérsic profiles with  $n \approx 1$ . Additionally, Hodge et al. (2016) found Sérsic profiles with  $n = 0.9 \pm 0.2$  for luminous SMGs, again within error with our results. Tacchella et al. (2015) and Genzel et al. (2014) have found  $H\alpha$  rings in some star-forming galaxies at  $z \sim 2$ . However, the results of this paper and Nelson et al. (2016) indicate that such systems are not very common.

#### 6.4 The potential implication for inside-out quenching

The aim of this study was to measure the average radio sizes of radio-faint massive star-forming galaxies. Our results indicate that, for star-forming galaxies, the star formation, as traced by the radio, is typically concentrated in the inner half of the galaxies, with effective radii that are smaller on average than those traced by optical and  $H\alpha$ . At face value, this seems in contrast to the signature of inside-out quenching found by studies tracing star formation through  $H\alpha$  emission (e.g. Tacchella et al. 2015; Nelson et al. 2016). This signature has up to now been traced using the specific-SFR (sSFR) profiles as traced by  $H\alpha$ , but in our statistical study we do not derive the sSFR profile and thus cannot directly compare our galaxy profiles to those expected for galaxies undergoing inside-out quenching (e.g. Tacchella et al. 2016). However, we would expect that the fitted radio sizes of our sample would be larger than the fitted optical sizes at lower redshift if the galaxies had similar inside-out growth to those observed by previous studies (e.g. Tacchella et al. 2015).

The fact that the radio sizes measured in this study, are more compact than those measured in the optical and  $H\alpha$ , as well as the fact that even the submm sizes of  $H\alpha$  selected galaxies are significantly more compact than their respective radio sizes (Tadaki et al. 2017), highlights the need for models such as that of Tacchella et al. (2016) to include information from multiple star formation tracers. Furthermore, an important factor that should be considered in such studies is the variable dust attenuation across the galaxies, that could cause some of the observed disagreements between size measurements.

## 7 CONCLUSIONS

We have carried out a study of the size evolution of star-forming galaxies in the redshift range  $z = 0$ – $3$  using combined 1.4 GHz

MERLIN and VLA observations of the Hubble Deep Field-North. We used a sample of  $\sim 1000$  galaxies selected to have  $M_* = 10^{10}$ – $10^{11.5} M_\odot$ . Using a  $uv$ -stacking approach and model fitting in the  $uv$  domain we have been able to determine the radio sizes as a function of  $z$ . The main results of this analysis are: (i) Radio sizes are smaller at  $z \approx 3$  compared to  $z \approx 0.5$ . (ii) Radio sizes are significantly smaller compared to sizes derived from optical broad-band filters, with sizes typically smaller by a factor of 2. We parametrize the median effective radius for  $M_* = 5 \times 10^{10} M_\odot$  as  $R_e = A \times (H(z)/H(1.5))^{\alpha_z}$ , and find  $A = 2.3 \pm 0.6$  kpc and  $\alpha_z = -1.1 \pm 0.6$ .

Radio emission can be used as a proxy for the SFR, and the results indicate (with some uncertainties discussed in Section 6.1), that the SFR is higher in the centre of star-forming galaxies compared to at larger radii. This is in agreement with measurements using  $H\alpha$  as a SFR tracer (e.g. Nelson et al. 2016); however these studies generally find higher sSFR at larger radii (e.g. Tacchella et al. 2015; Nelson et al. 2016). Such results from  $H\alpha$  have previously been interpreted as evidence for inside-out quenching of galaxies. However, the implications of a variable dust attenuation across the galaxies have not yet been explored. Submm interferometers such as ALMA and Northern Extended Millimeter Array (NOEMA) can map dust emission directly, and the results of this paper indicate the importance of such observation to robustly map star formation in high-redshift galaxies. Further insight can be gained from observations with the upgraded MERLIN telescope (e-MERLIN) with larger bandwidth, which will allow size measurements of lower mass galaxies as well as more detailed mapping of high mass star-forming galaxies.

The small radio sizes derived could also indicate that mechanisms other than inside-out growth, such as minor mergers and central mass loss, play an important role in the size evolution of star-forming galaxies. This emphasizes the need for mechanisms in galaxy evolution which can redistribute the stellar mass, while also maintaining an exponential stellar-mass surface density profile for star-forming galaxies.

## ACKNOWLEDGEMENTS

We thank the anonymous referee for their valuable contribution to the improvement of this paper. KK acknowledges support from the Swedish Research Council. The National Radio Astronomy Observatory is a facility of the National Science Foundation operated under cooperative agreement by Associated Universities, Inc. MERLIN is a national facility operated by the University of Manchester on behalf of the Science and Technology Facilities Council (STFC).

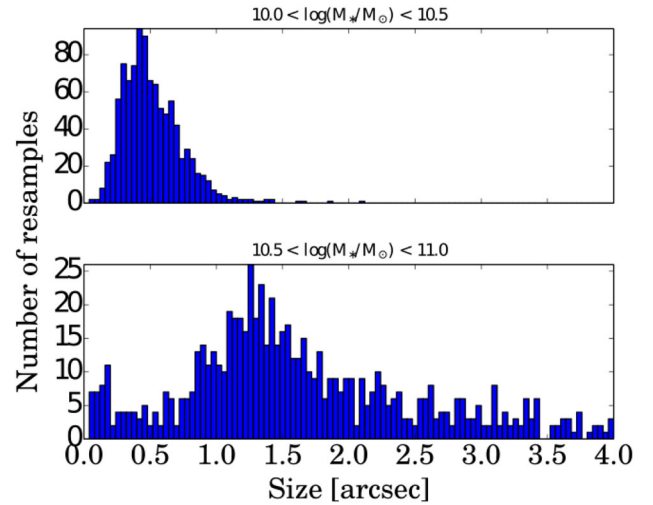
## REFERENCES

- Alexander D. M. et al., 2003, *AJ*, 126, 539
- Bell E. F., 2003, *ApJ*, 586, 794
- Beswick R. J., Muxlow T. W. B., Thrall H., Richards A. M. S., Garrington S. T., 2008, *MNRAS*, 385, 1143
- Bezanson R., van Dokkum P. G., Tal T., Marchesini D., Kriek M., Franx M., Coppi P., 2009, *ApJ*, 697, 1290
- Bezanson R., van Dokkum P., Franx M., 2012, *ApJ*, 760, 62
- Biggs A. D., Ivison R. J., 2008, *MNRAS*, 385, 893
- Brammer G. B., van Dokkum P. G., Coppi P., 2008, *ApJ*, 686, 1503
- Brammer G. B. et al., 2012, *ApJS*, 200, 13
- Bruzual G., Charlot S., 2003, *MNRAS*, 344, 1000
- Chabrier G., 2003, *PASP*, 115, 763
- Chihara L., Hesserberg T., 2011, *Mathematical Statistics with Resampling and R*. John Wiley & Sons, Inc, NJ

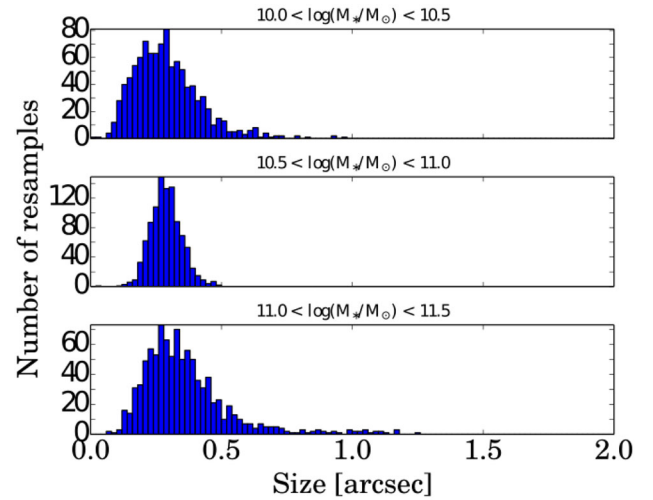
Condon J. J., 1992, *ARA&A*, 30, 575  
Daddi E., Cimatti A., Renzini A., Fontana A., Mignoli M., Pozzetti L., Tozzi P., Zamorani G., 2004, *ApJ*, 617, 746  
Fan L., Lapi A., Bressan A., Bernardi M., De Zotti G., Danese L., 2010, *ApJ*, 718, 1460  
Genzel R. et al., 2014, *ApJ*, 785, 75  
Giavalisco M., Steidel C. C., Macchetto F. D., 1996, *ApJ*, 470, 189  
Hainline L. J., Blain A. W., Smail I., Alexander D. M., Armus L., Chapman S. C., Ivison R. J., 2011, *ApJ*, 740, 96  
Hodge J. A. et al., 2016, *ApJ*, 833, 103  
Ikarashi S. et al., 2015, *ApJ*, 810, 133  
Karim A. et al., 2011, *ApJ*, 730, 61  
Kriek M., van Dokkum P. G., Labbé I., Franx M., Illingworth G. D., Marchesini D., Quadri R. F., 2009, *ApJ*, 700, 221  
Lindroos L., Knudsen K., Vlemmings W., Conway J., Martí-Vidal I., 2015, *MNRAS*, 446, 3502  
Lindroos L. et al., 2016, *MNRAS*, 462, 1192  
Lotz J. M., Jonsson P., Cox T. J., Croton D., Primack J. R., Somerville R. S., Stewart K., 2011, *ApJ*, 742, 103  
Martí-Vidal I., Pérez-Torres M. A., Lobanov A. P., 2012, *A&A*, 541, A135  
Miettinen O. et al., 2017, *A&A*, 602, A54  
Morishita T., Ichikawa T., Kajisawa M., 2014, *ApJ*, 785, 18  
Murphy E. J., Helou G., Kenney J. D. P., Armus L., Braun R., 2008, *ApJ*, 678, 828  
Murphy E. J., Momjian E., Condon J. J., Chary R.-R., Dickinson M., Inami H., Taylor A. R., Weiner B. J., 2017, *ApJ*, 839, 35  
Muxlow T. W. B. et al., 2005, *MNRAS*, 358, 1159  
Nelson E. J. et al., 2013, *ApJ*, 763, L16  
Nelson E. J. et al., 2016, *ApJ*, 828, 27  
Oke J. B., 1974, *ApJS*, 27, 21  
Pacifi C. et al., 2016, *ApJ*, 832, 79  
Padovani P., Bonzini M., Kellermann K. I., Miller N., Mainieri V., Tozzi P., 2015, *MNRAS*, 452, 1263  
Planck Collaboration XVI, 2014, *A&A*, 571, A16  
Poggianti B. M., Moretti A., Calvi R., D'Onofrio M., Valentinuzzi T., Fritz J., Renzini A., 2013, *ApJ*, 777, 125  
Richards E. A., 2000, *ApJ*, 533, 611  
Richards A. M. S. et al., 2007, *A&A*, 472, 805  
Rujopakarn W. et al., 2016, *ApJ*, 833, 12  
Simpson J. M. et al., 2015, *ApJ*, 799, 81  
Skelton R. E. et al., 2014, *ApJS*, 214, 24  
Szomoru D., Franx M., Bouwens R. J., van Dokkum P. G., Labbé I., Illingworth G. D., Trenti M., 2011, *ApJ*, 735, L22  
Tacchella S. et al., 2015, *Science*, 348, 314  
Tacchella S., Dekel A., Carollo C. M., Ceverino D., DeGraf C., Lapiner S., Mandelker N., Primack J. R., 2016, *MNRAS*, 458, 242  
Tadaki K.-i. et al., 2017, *ApJ*, 834, 135  
Trujillo I., Cenarro A. J., de Lorenzo-Cáceres A., Vazdekis A., de la Rosa I. G., Cava A., 2009, *ApJ*, 692, L118  
Valentinuzzi T. et al., 2010, *ApJ*, 712, 226  
van der Wel A. et al., 2014, *ApJ*, 788, 28  
Whitaker K. E. et al., 2014, *ApJ*, 795, 104  
Williams R. J., Quadri R. F., Franx M., van Dokkum P., Labbé I., 2009, *ApJ*, 691, 1879  
Wrigley N., 2011, Master's thesis, Univ. Manchester  
Wuyts S. et al., 2012, *ApJ*, 753, 114  
Wuyts E. et al., 2014, *ApJ*, 789, L40  
Yoast-Hull T. M., Gallagher J. S., Zweibel E. G., 2015, *MNRAS*, 453, 222  
Yuma S., Ohta K., Yabe K., Kajisawa M., Ichikawa T., 2011, *ApJ*, 736, 92

## APPENDIX A: MODEL FITTING AND ERROR ESTIMATES

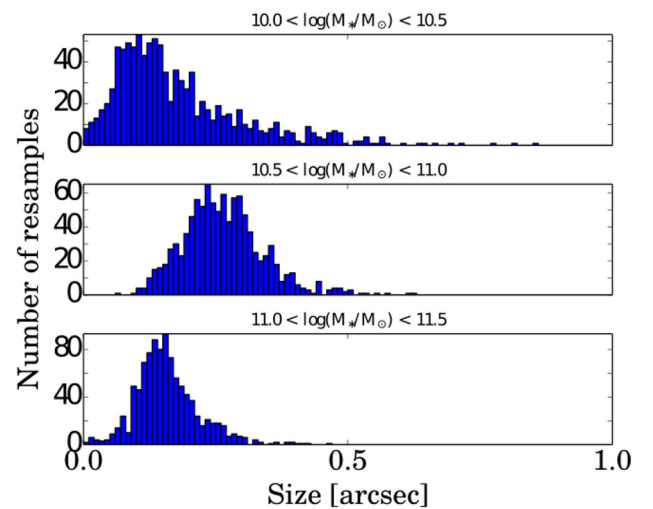
In Figs A1–A6, we show bootstrapping distribution of the size for all subsamples. Note that bootstrapping for the highest mass sample for  $z < 0.5$  was not performed as only one source was present in this subsample. This subsample was not used in any further analysis,



**Figure A1.** Bootstrapping distribution of sizes for galaxies with  $0.0 < z < 0.5$ .



**Figure A2.** Bootstrapping distribution of sizes for galaxies with  $0.5 < z < 1.0$ .



**Figure A3.** Bootstrapping distribution of sizes for galaxies with  $1.0 < z < 1.5$ .

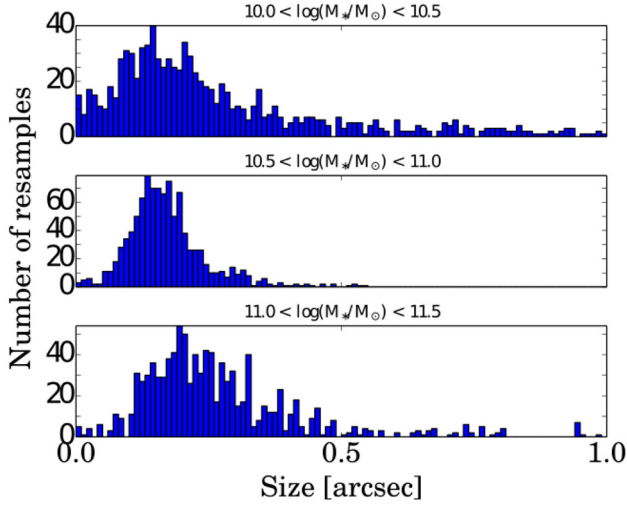


Figure A4. Bootstrapping distribution of sizes for galaxies with  $1.5 < z < 2.0$ .

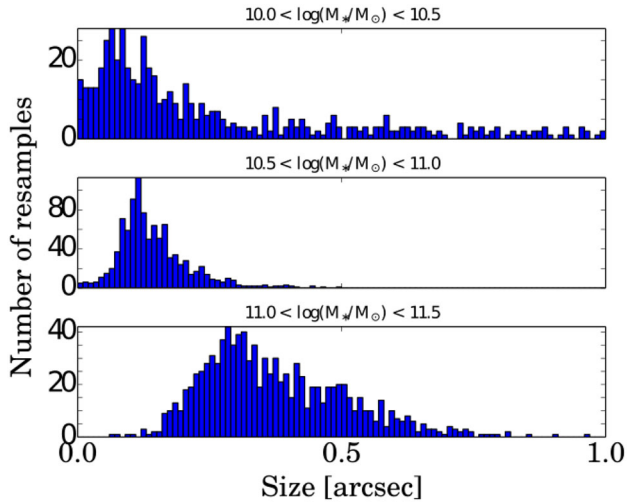


Figure A5. Bootstrapping distribution of sizes for galaxies with  $2.0 < z < 2.5$ .

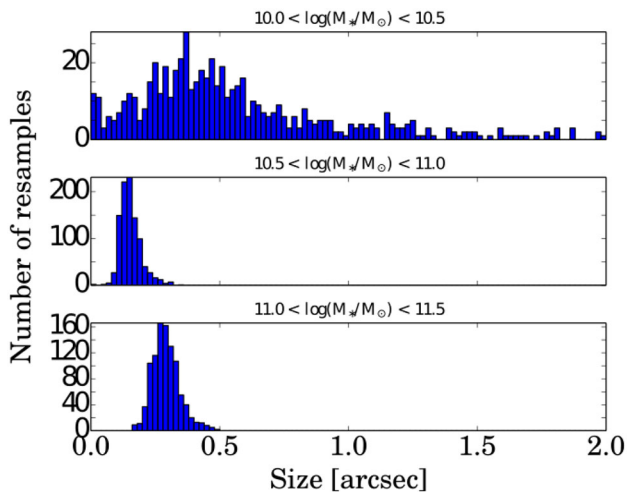


Figure A6. Bootstrapping distribution of sizes for galaxies with  $2.5 < z < 3.0$ .

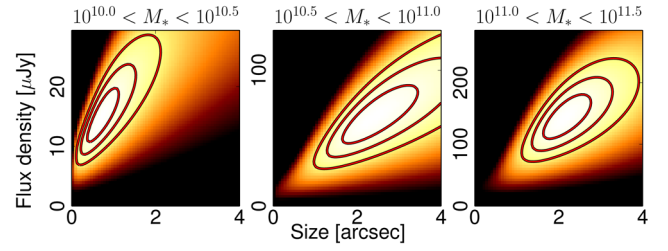


Figure A7. Variation for  $\chi^2$  for the stacked galaxies of the main sample with  $z$  between 0.0 and 0.5. The contours show a confidence interval of 68.3 per cent, 95.5 per cent, and 99.7 per cent (e.g. 1, 2, and  $3\sigma$  for a Gaussian distribution).

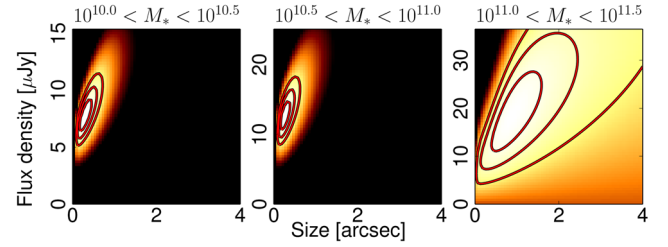


Figure A8. Variation for  $\chi^2$  for the stacked galaxies of the main sample with  $z$  between 0.5 and 1.0. The contours show a confidence interval of 68.3 per cent, 95.5 per cent, and 99.7 per cent (e.g. 1, 2, and  $3\sigma$  for a Gaussian distribution).

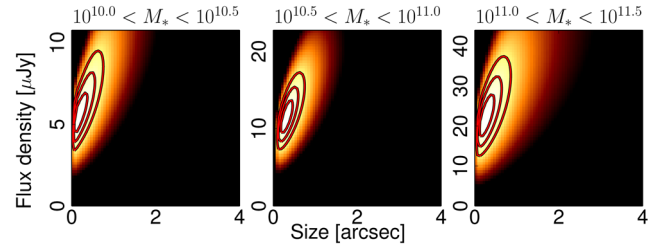
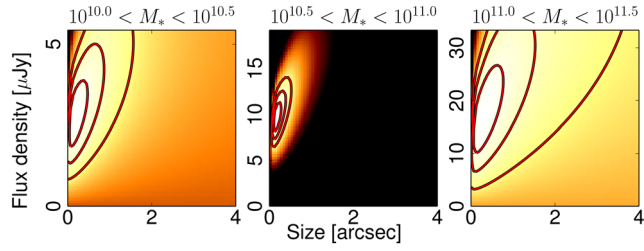


Figure A9. Variation for  $\chi^2$  for the stacked galaxies of the main sample with  $z$  between 1.0 and 1.5. The contours show a confidence interval of 68.3 per cent, 95.5 per cent, and 99.7 per cent (e.g. 1, 2, and  $3\sigma$  for a Gaussian distribution).

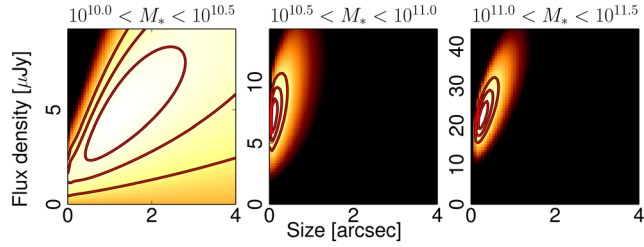
and the presented errors in the paper are based on the  $\chi^2$  variations. For comparison, we also calculate errors based on the largest deviations of parameters with  $\chi^2_{\text{sample}} < 2.31 + \min_{\Phi, \sigma} \chi^2_{\text{sample}}(\Phi, \sigma)$ , i.e. a confidence interval of 68 per cent (e.g.  $1\sigma$  for a Gaussian distribution). In Figs A7–A12, we show the  $\chi^2$  variation for each subsample.

### A1 Combined model

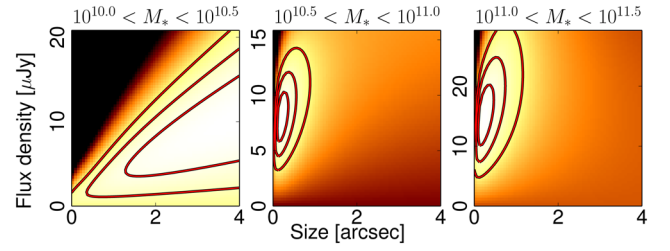
We estimate errors by varying the input parameters on a fixed input grid. The errors are reported as the largest deviations of parameters with  $\chi^2 < 5.89 + \min_{\alpha_z, \beta_z, \beta_M, A, B} \chi^2(\alpha_z, \beta_z, \beta_M, A, B)$ , i.e. a confidence interval of 68 per cent for our model with five free parameters.



**Figure A10.** Variation for  $\chi^2$  for the stacked galaxies of the main sample with  $z$  between 1.5 and 2.0. The contours show a confidence interval of 68.3 per cent, 95.5 per cent, and 99.7 per cent (e.g. 1, 2, and  $3\sigma$  for a Gaussian distribution).



**Figure A11.** Variation for  $\chi^2$  for the stacked galaxies of the main sample with  $z$  between 2.0 and 2.5. The contours show a confidence interval of 68.3 per cent, 95.5 per cent, and 99.7 per cent (e.g. 1, 2, and  $3\sigma$  for a Gaussian distribution).



**Figure A12.** Variation for  $\chi^2$  for the stacked galaxies of the main sample with  $z$  between 2.5 and 3.0. The contours show a confidence interval of 68.3 per cent, 95.5 per cent, and 99.7 per cent (e.g. 1, 2, and  $3\sigma$  for a Gaussian distribution).

This paper has been typeset from a  $\text{\LaTeX}$  file prepared by the author.