



<b>Publication Year</b>	2021
<b>Acceptance in OA</b>	2025-02-20T10:28:08Z
<b>Title</b>	The interstellar medium of quiescent galaxies and its evolution with time
<b>Authors</b>	Magdis, Georgios E., Gobat, Raphael, Valentino, Francesco, Daddi, Emanuele, ZANELLA, Anita, Kokorev, Vasily, Toft, Sune, Shuowen, Jin, Whitaker, Katherine E.
<b>Publisher's version (DOI)</b>	10.1051/0004-6361/202039280
<b>Handle</b>	<a href="http://hdl.handle.net/20.500.12386/36094">http://hdl.handle.net/20.500.12386/36094</a>
<b>Journal</b>	ASTRONOMY & ASTROPHYSICS
<b>Volume</b>	647

# The interstellar medium of quiescent galaxies and its evolution with time

Georgios E. Magdis<sup>1,2,3</sup>, Raphael Gobat<sup>4</sup>, Francesco Valentino<sup>1,3</sup>, Emanuele Daddi<sup>5</sup>, Anita Zanella<sup>6</sup>, Vasily Kokorev<sup>1,3</sup>, Sune Toft<sup>1,3</sup>, Shuowen Jin<sup>7,8</sup>, and Katherine E. Whitaker<sup>9,1</sup>

<sup>1</sup> Cosmic Dawn Center (DAWN), Copenhagen, Denmark  
e-mail: geoma@space-dtu.dk

<sup>2</sup> DTU-Space, Technical University of Denmark, Elektrovej 327, 2800 Kgs. Lyngby, Denmark

<sup>3</sup> Niels Bohr Institute, University of Copenhagen, Lyngbyvej 2, 2100 Copenhagen Ø, Denmark

<sup>4</sup> Instituto de Física, Pontificia Universidad Católica de Valparaíso, Casilla 4059, Valparaíso, Chile

<sup>5</sup> CEA, IRFU, DAp, AIM, Université Paris-Saclay, Université Paris Diderot, Sorbonne Paris Cité, CNRS, 91191 Gif-sur-Yvette, France

<sup>6</sup> Istituto Nazionale di Astrofisica, Vicolo dell'Osservatorio 5, 35122 Padova, Italy

<sup>7</sup> Instituto de Astrofísica de Canarias (IAC), 38205 La Laguna, Tenerife, Spain

<sup>8</sup> Universidad de La Laguna, Dpto. Astrofísica, 38206 La Laguna, Tenerife, Spain

<sup>9</sup> Department of Astronomy, University of Massachusetts, Amherst, MA 01003, USA

Received 28 August 2020 / Accepted 18 December 2020

## ABSTRACT

We characterise the basic far-IR properties and the gas mass fraction of massive ( $\langle \log(M_*/M_\odot) \rangle \approx 11.0$ ) quiescent galaxies (QGs) and explore how these evolve from  $z = 2.0$  to the present day. We use robust, multi-wavelength (mid- to far-IR and sub-millimetre to radio) stacking ensembles of homogeneously selected and mass complete samples of  $\log(M_*/M_\odot) \geq 10.8$  QGs. We find that the dust to stellar mass ratio ( $M_{\text{dust}}/M_*$ ) rises steeply as a function of redshift up to  $z \sim 1.0$  and then remains flat at least out to  $z = 2.0$ . Using  $M_{\text{dust}}$  as a proxy of gas mass ( $M_{\text{gas}}$ ), we find a similar trend for the evolution of the gas mass fraction ( $f_{\text{gas}}$ ), with  $z > 1.0$  QGs having  $f_{\text{gas}} \approx 7.0\%$  (for solar metallicity). This  $f_{\text{gas}}$  is three to ten times lower than that of normal star-forming galaxies (SFGs) at their corresponding redshift but  $\geq 3$  and  $\geq 10$  times larger compared to that of  $z = 0.5$  and local QGs. Furthermore, the inferred gas depletion time scales are comparable to those of local SFGs and systematically longer than those of main sequence galaxies at their corresponding redshifts. Our analysis also reveals that the average dust temperature ( $T_d$ ) of massive QGs remains roughly constant ( $\langle T_d \rangle = 21.0 \pm 2.0$  K) at least out to  $z \approx 2.0$  and is substantially colder ( $\Delta T_d \approx 10$  K) compared to that of SFGs. This motivated us to construct and release a redshift-invariant template IR SED, that we used to make predictions for ALMA observations and to explore systematic effects in the  $M_{\text{gas}}$  estimates of massive, high- $z$  QGs. Finally, we discuss how a simple model that considers progenitor bias can effectively reproduce the observed evolution of  $M_{\text{dust}}/M_*$  and  $f_{\text{gas}}$ . Our results indicate universal initial interstellar medium conditions for quenched galaxies and a large degree of uniformity in their internal processes across cosmic time.

**Key words.** galaxies: evolution – galaxies: ISM – galaxies: formation

## 1. Introduction

The existence of massive, passively evolving galaxies at various cosmic epochs is now well established thanks to observational efforts that span more than two decades, from first discoveries (e.g., Daddi et al. 2000, 2005; Cimatti et al. 2004; Toft et al. 2005, 2007; Kriek et al. 2006) to the latest spectroscopic confirmations (e.g., Toft et al. 2012, 2017; Gobat et al. 2012; Whitaker et al. 2013; Glazebrook et al. 2017; Schreiber et al. 2018b; Tanaka et al. 2019; Valentino et al. 2020; D'Eugenio et al. 2020; Stockmann et al. 2020). This population of early type galaxies (ETGs), or more generally of quiescent galaxies (QGs)<sup>1</sup>, experiences low (or negligible) levels of star formation, resides well below the main sequence (MS) of star formation (e.g., Noeske et al. 2007; Daddi et al. 2007; Magdis et al. 2010; Speagle et al. 2014; Schreiber et al. 2015) at their corresponding redshift, and is characterised by red colours, indicative of old and evolved stellar populations. With a large fraction of

$z \geq 2$  massive galaxies ( $M_* > 10^{11} M_\odot$ ) experiencing a downfall in their star formation activity (e.g., Whitaker et al. 2012; Davidzon et al. 2017), the abundance of QGs progressively increases at later times. At the same time, recent studies have attested that a fraction of them are already in place by  $z \sim 4$  (e.g., Glazebrook et al. 2017; Schreiber et al. 2018b; Tanaka et al. 2019; Valentino et al. 2020).

While the mere detection of passive and massive galaxies out to  $z \sim 3-4$  poses in itself challenges in our theories of galaxy evolution, their existence is perplexing at a more basic level: The very mechanisms responsible for triggering and/or sustaining the cessation of star formation at any redshift remain an unsolved puzzle (for a recent overview, see Man & Belli 2018). Several competing scenarios that attempt to explain quenching have been put forward, promoting mechanisms that prevent gas from cooling (halo quenching, active galactic nuclei and stellar feedback, strangulation, e.g., Cattaneo et al. 2006; Peng et al. 2015; Henriques et al. 2015), as well as gas expulsion via feedback (outflows, e.g., DiMatteo et al. 2005; Hopkins et al. 2006) and gas stabilisation (morphological quenching,

<sup>1</sup> In this paper we use the terms ETG and QG interchangeably.

e.g., Martig et al. 2009). While these plausible quenching mechanisms invoke different processes at different physical scales (from dark matter halos to nuclear activity), it is evident that they all have some common parameters at their core: the gas mass budget, the physical conditions, and the dynamical state of the interstellar medium (ISM). This should not come as a surprise. Since gas and dust are the agents of star formation, it quite naturally follows that they should also be at the heart of its cessation, or in other words, for quenching. Consequently, progress in the field necessitates a careful study of the ISM of QGs, very similar to what has already been achieved for star-forming galaxies (SFGs).

To this end, a large volume of studies have scrutinised the ISM properties of local QGs. By exploring their far-IR (FIR) and millimetre (mm) dust continuum emission and/or targeting atomic and molecular lines (e.g., low transition CO lines), these studies have revealed low, but not negligible, amounts of dust (e.g., Gomez et al. 2010; Rowlands et al. 2012; Smith et al. 2012; Agius et al. 2013; Werner et al. 2014; Boselli et al. 2014; Lianou et al. 2016) and gas (e.g., Saintonge et al. 2011a,b; Young et al. 2011; Cappellari et al. 2013; Davis et al. 2014; Boselli et al. 2014; French et al. 2015; Alatalo et al. 2016) in local QGs and overall FIR characteristics that are in contrast to those of local SFGs. In particular, the dust mass fraction ( $M_{\text{dust}}/M_*$ ) of local QGs is found to vary between  $10^{-5}$  and  $10^{-4}$ , and their molecular gas fraction ( $f_{\text{gas}} = M_{\text{gas}}/M_*$ ) from 0.3 to 1%, with both quantities being  $\sim 2$  orders of magnitude lower than those of normal galaxies in the local Universe. Similarly, their typical ‘luminosity-weighted’ dust temperatures ( $T_{\text{d}}$ ) are  $\sim 20$  K (e.g., Smith et al. 2012), on average 3–5 K colder than those of normal SFGs at  $z = 0$ .

On the other hand, the exploration of the properties of the ISM in QGs beyond the local Universe is still in its infancy, in part due to their strenuous selection (and spectroscopic confirmation), but more importantly due to their faint FIR nature. In spite of these challenges though, some ‘brave’ attempts to measure the  $M_{\text{gas}}$  of distant QGs and post-starburst (pSB) galaxies have already been carried out (e.g., Sargent et al. 2015; Suess et al. 2017; Rudnick et al. 2017; Hayashi et al. 2018; Spilker et al. 2018; Bezanson et al. 2019; Williams et al. 2021). These studies of small and inhomogeneous (in terms of selection, and physical properties, e.g.,  $M_*$ ) samples of distant QGs, have so far provided only loose (if any) constraints in their gas mass budget and its evolution with time, placing their  $f_{\text{gas}}$  anywhere between 15% and  $< 3\%$ . The diverse  $f_{\text{gas}}$  measurements emerging from various studies demonstrate the need for a systematic study of the ISM for a carefully and uniformly selected sample of QGs in various redshifts bins.

At the same time, the shape of the FIR spectral energy distributions (SEDs) of distant QGs remains even less explored. Indeed, while reasonable  $M_{\text{gas}}$  and  $M_{\text{dust}}$  estimates can be achieved with single line (e.g., CO) or single-band dust continuum observations, a full characterisation of their IR SEDs requires multi-wavelength photometry probing both the Wien part and the Rayleigh-Jeans (R-J) tail, a prohibitively expensive task in terms of observing time, even for the limited number of available QGs that benefit from lensing magnification. As such, any progress on this front can only be achieved through stacking (e.g., Viero et al. 2013; Man et al. 2016; Gobat et al. 2018, G18 hereafter).

Finally, it is worth recalling that the detailed study of the FIR properties of representative and complete samples of SFGs at various cosmic epochs has advanced our understanding of galaxy evolution through scaling relations and evolutionary

tracks that have been used to guide simulations and theoretical models (e.g., Dekel et al. 2009; Popping et al. 2014, 2017; Narayanan et al. 2015; Lagos et al. 2015; Davé et al. 2017, 2020). Far-IR studies have revealed that the gas and dust mass fraction of SFGs have sharply declined over time, with SFGs having an ISM mass budget larger by a factor of  $\sim 10$  at  $z \approx 2$  compared to that of SFGs of similar  $M_*$  in the local Universe (e.g., Daddi et al. 2008, 2010; Geach et al. 2011; Magdis et al. 2012, 2017; Tacconi et al. 2018; Liu et al. 2019). At the same time high- $z$  SFGs, despite their larger gas fractions, are characterised by shorter gas depletion time scales ( $\tau_{\text{dep}}$ ) indicative of a higher star formation efficiency at earlier times (e.g., Tacconi et al. 2018; Liu et al. 2019). Similarly, the  $T_{\text{d}}$  of SFGs also increases with redshift (e.g., Magdis et al. 2012, 2017; Magnelli et al. 2014; Schreiber et al. 2018a; Cortzen et al. 2020), mirroring an evolution in their star formation activity, star formation surface density, and the strength of their radiation field. As a final example, we refer to the construction of representative FIR SEDs of SFGs at various redshifts (e.g., Elbaz et al. 2011; Magdis et al. 2012; Schreiber et al. 2018a), which have greatly facilitated models, predictions, and extrapolations.

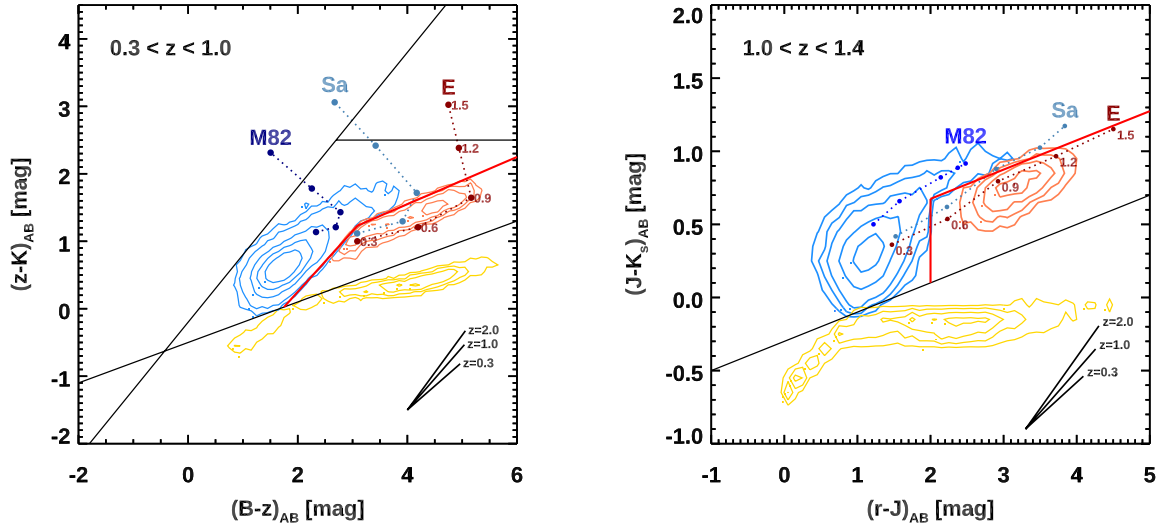
These successes motivated us to undertake a similar effort, but this time for the population that consists of the ending point of the evolutionary path of SFGs, in other words, for QGs. In this work we aim to characterise the basic FIR properties ( $L_{\text{IR}}$ ,  $T_{\text{d}}$ ,  $M_{\text{dust}}$ ,  $M_{\text{gas}}$ ,  $f_{\text{gas}}$ ,  $\tau_{\text{dep}}$ ), and trace their evolution with time, for a robust and controlled sample of QGs in various redshift bins. For this task we employed a multi-wavelength, mid- to far-IR and millimetre to radio stacking analysis of homogeneously selected and mass complete ( $\log(M_*/M_{\odot}) \gtrsim 10.8$ ) QGs samples in the  $z = 0.3$ – $1.5$  redshift range, drawn from the COSMOS field. To extend the redshift coverage of our work we further complement our analysis with the  $1.5 < z < 2.0$  stacked sample of QGs presented in Gobat et al. (2018) and which is constructed using the same selection criteria as those employed in our study.

The paper is organised as follows: in Sect. 2 we present the selection criteria for our samples of QGs in various redshift bins. In Sect. 3 we present the stacking methodology and the subsequent photometry. In Sect. 4 we present a detailed analysis of the FIR properties of our stacked ensembles and in Sect. 5 we explore their evolution with time. In Sect. 6 we discuss the implications of our findings and provide an outlook for future IR studies of distant QGs with (sub-)millimetre facilities. Finally, in Sect. 7 we summarise the main findings of our work. Throughout the paper we adopt a Salpeter (1955) initial mass function (IMF) and  $H_0 = 70 \text{ km s}^{-1} \text{ Mpc}^{-1}$ ,  $\Omega_{\text{M}} = 0.3$ , and  $\Omega_{\Lambda} = 0.7$ .

## 2. Selection

We selected QGs in the COSMOS field (Scoville et al. 2007) based on the latest public photometric catalogue by Laigle et al. (2016). First, we removed stars (i.e. TYPE=1 flag) and set a magnitude cut at  $K_s = 24.5$  mag, the  $3\sigma$  depth in  $2''$  diameter apertures in the ‘deep’ UltraVista stripes (Laigle et al. 2016). We further excluded X-ray *Chandra*-detected sources from the latest catalogues by Civano et al. (2016) and Marchesi et al. (2016), thus removing active galactic nuclei and strongly star-forming objects. Then, following the approach in G18, we separated red QGs from blue star-forming sources by combining the rest-frame *NUVrJ* criterion (Arnouts et al. 2007; Ilbert et al. 2013) with the observed *BzK* (Daddi et al. 2004) and (a custom) *rJK* diagrams.

*The BzK criterion.* The *BzK* diagram was originally designed to select galaxies at  $1.4 \lesssim z \lesssim 2.5$ , but it can also be successfully



**Fig. 1.** Redshift ranges for an efficient use of the  $BzK$  and  $rJK$  colour diagrams. *Left:*  $BzK$  diagram for the selection of  $0.3 \leq z < 1.0$  QGs. The yellow, blue, and red contours represent the loci of stars, SFGs, and QGs at  $0.3 \leq z < 1.0$  selected by the  $NUVrJ + BzK$  selection. Sources with a  $3\sigma$  detection at  $24\mu\text{m}$  in Jin et al. (2018) have been excluded. The black lines represent the original cuts by Daddi et al. (2004) to select galaxies at  $1.4 \leq z \leq 2.5$ . The red lines show our cut at low redshifts (Eq. (1)). For reference, the tracks show the synthetic colours for the 13 Gyr elliptical (E), early-type spirals (Sa), and M82-like templates by Polletta et al. (2007), as marked in the labels. The colours are computed for  $0.3 < z < 1.5$  in step of  $\Delta z = 0.3$ . No evolutionary effects are included. The black segments show the reddening vectors at  $z = 0.3, 1, \text{ and } 2$ . *Right:*  $rJK$  diagram for the selection of  $1.0 \leq z < 1.4$  QGs. The yellow, blue, and red contours represent the loci of stars, SFGs, and QGs at  $1.0 \leq z < 1.4$  selected by the  $NUVrJ + rJK$  selection. The remaining lines are coded as in the left panel.

used at lower redshift (Bielby et al. 2012). We computed the  $B - z$ ,  $z - K$  colours using magnitudes in  $2''$  diameters apertures, following McCracken et al. (2010, 2012).

We applied a  $+0.05$  mag correction to the  $z - K$  colour to better match the stellar locus from the Lejeune et al. (1997) stellar models reported in the original selection from Daddi et al. (2004). Red QGs at  $0.3 \leq z < 1.0$  fall in the portion of the  $B - z$ ,  $z - K$  plane delimited by:

$$\begin{cases} z - K > 0.3 \times (B - z) - 0.5 \\ z - K \leq 0.35 \times (B - z) + 0.15 \\ z - K \leq 0.9 \times (B - z) - 1.55. \end{cases} \quad (1)$$

We selected galaxies detected at  $5\sigma$  in the  $K_s$  and  $z^+$  bands, and retained sources with lower limits on the  $B - z$  colour compatible with the cuts.

*The  $rJK$  criterion.* We adapted the principle of the  $BzK$  diagram (i.e. two bands bluer and one band redder than the Balmer/4000 Å jumps at the redshift of reference) to classify galaxies in the range  $1.0 \leq z < 1.4$ . We computed the  $r - J$ ,  $J - K$  colours using magnitudes measured in  $2''$  diameter apertures. We selected red quiescent objects as:

$$\begin{cases} J - K > 0.2 \times (r - J) - 0.3 \\ J - K < 0.2 \times (r - J) + 0.275 \\ r - J > 2.0. \end{cases} \quad (2)$$

We selected sources detected at  $5\sigma$  in the  $K_s$  and  $J$  band, and retained objects with lower limits on the  $r - J$  colour compatible with the cuts.

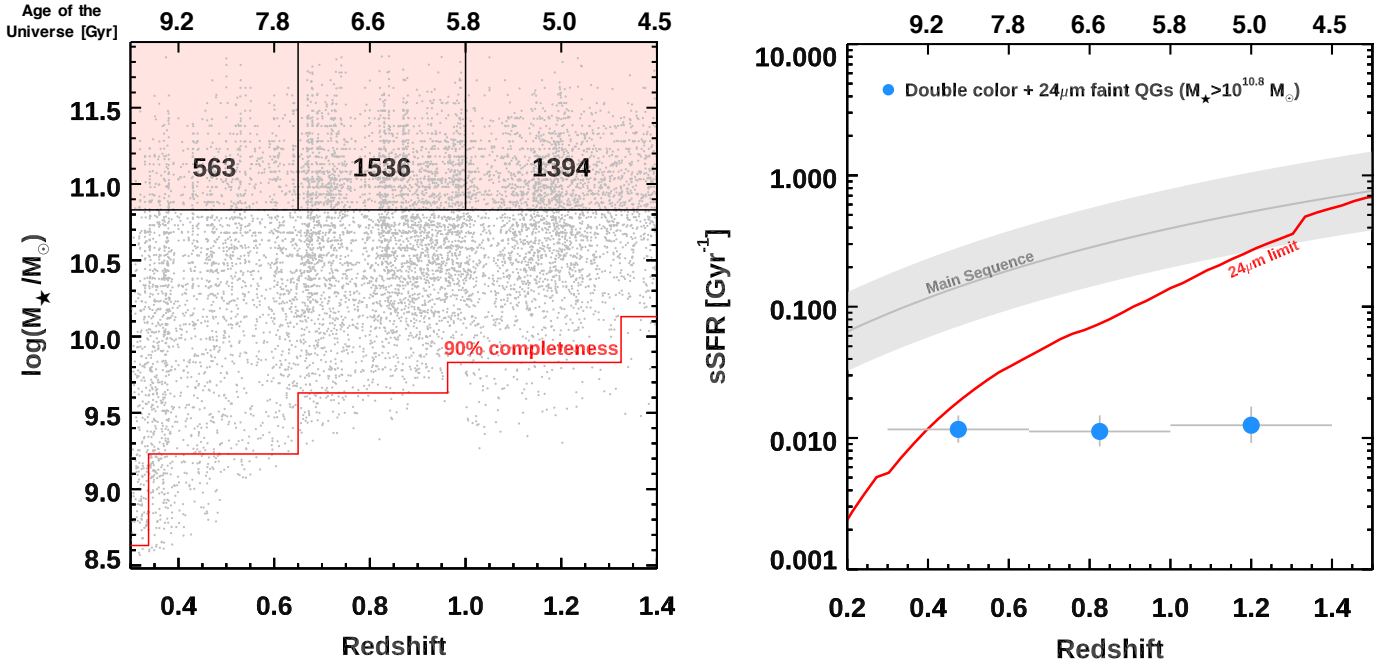
We show the observed colour diagrams and the loci occupied by red QGs at  $0.3 \leq z \leq 1.4$  in Fig. 1. For guidance, we show the synthetic colours for a series of templates of different galaxy types from Polletta et al. (2007; see Bielby et al. 2012 for a similar exercise with Bruzual & Charlot 2003 stellar population models). We computed the colours at different redshifts, but

without including evolution effects. The locus of 13 Gyr-old QGs falls entirely within the region occupied by  $z \lesssim 1$  QGs in the  $BzK$  diagram and  $z \lesssim 1.4$  QGs in the  $rJK$  one. Templates typical of strongly star-forming galaxies lie entirely outside the quiescent region in both diagrams. Early-type spirals initially fall in the star-forming portions of the colour diagrams, later entering the quiescent region.

The addition of the  $BzK$  and  $rJK$  criteria removes more efficiently the red dusty galaxies contaminating the sample of quiescent objects selected with the  $NUVrJ$  diagram only. Sources without a constraint on the observed colours occupy regions of the  $NUVrJ$  diagram with bluer  $r - J$  and redder  $NUV - r$  colours, partially overlapping with the final quiescent sample only below  $z < 1$ . On the other hand, galaxies excluded by the  $BzK$  and  $rJK$  criteria lie close to the line separating the star-forming population. A 2D Kolmogorov-Smirnov test (Fasano & Franceschini 1987) suggests that both the unconstrained and excluded galaxies are drawn from a parent distribution different from the one of the final quiescent sample.

The validity of the double-colour selection is further supported by the lower fraction of *Spitzer*/*MIPS*  $24\mu\text{m}$   $3\sigma$  detected red galaxies when using  $NUVrJ + BzK$  and  $NUVrJ + rJK$  colour selections ( $\sim 15\%$ ), rather than a simple  $NUVrJ$  diagram ( $\sim 25\%$ ). For the comparison we use the latest  $24\mu\text{m}$  catalogue by Jin et al. (2018), reaching an rms of  $\sim 15\mu\text{Jy}$ . Figure 2 shows that we can trace SFRs down to several times below the MS parameterised as in Schreiber et al. (2015). To compute the SFR limit in Jin et al. (2018), we rescaled the redshift-dependent, MS templates by Magdis et al. (2012) to match the observed  $24\mu\text{m}$  depth and then integrated over the wavelength range  $8 - 1000\mu\text{m}$  to derive the total infrared luminosity ( $L_{\text{IR}}$ ). We smoothed over the width of the *Spitzer*/*MIPS*  $24\mu\text{m}$  filter to avoid the strong impact of the PAH features.

We excluded all the  $24\mu\text{m}$  detected sources in the Jin et al. (2018) catalogue from our sample of QGs. We remark that the sample in the highest redshift bin could still potentially be



**Fig. 2.** Selection and binning. *Left:* grey points showing the final sample of double-colour selected and  $24\ \mu\text{m}$  faint QGs in the redshift vs. stellar mass plane. The rose area and the black grid mark the limits of the bins used to group galaxies for stacking. The number of stacked sources per bin are displayed. The red line indicates the 90% stellar mass completeness limit for QGs in the ‘deep’ UltraVISTA sample (Laigle et al. 2016), adjusted for the different IMF. *Right:* blue points showing the position of the stacked massive QGs ( $M_* > 10^{10.8} M_\odot$ ) in the redshift vs. sSFR(*UV*) plane. The error bars on the redshift indicate the width of the bins. Only the statistical uncertainties on sSFR(*UV*) are shown on the y-axis. The solid red line indicates the  $3\sigma$  detection limit for *Spitzer*/*MIPS* at  $24\ \mu\text{m}$  ( $\sigma = 15\ \mu\text{Jy}$ , Jin et al. 2018) converted to SFR using the MS templates from Magdis et al. (2012) and assuming  $M_* = 10^{11.2} M_\odot$ , matching the median stellar mass of our sample in each redshift bin. The grey shaded area and the grey line indicate the location of the MS and its 0.3 dex scatter, computed for a galaxy with  $M_* = 10^{11.2} M_\odot$ .

**Table 1.** Physical properties of the stacked quiescent samples.

Redshift	$N$	$\log(M_*)$ [ $M_\odot$ ]	SFR( <i>UV</i> ) <sup>(†)</sup> [ $M_\odot\ \text{yr}^{-1}$ ]	$\log(L_{\text{IR}})$ [ $L_\odot$ ]	$T_d$ [K]	$\langle U \rangle$	$\log(M_{\text{dust}})$ [ $M_\odot$ ]	$\log(M_{\text{gas}}^{\text{sol}})$ [ $M_\odot$ ]	$f_{\text{gas}}^{\text{sol}}$ [%]
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
$0.30 < z < 0.65$	563	$11.17 \pm 0.15$	$1.7^{+0.3}_{-0.2}$	$9.40 \pm 0.05$	$20.0 \pm 4.0$	$0.6 \pm 0.3$	$7.51^{+0.20}_{-0.24}$	$9.46^{+0.20}_{-0.24}$	$2.0 \pm 1.0$
$0.65 < z < 1.00$	1536	$11.20 \pm 0.17$	$1.7^{+0.3}_{-0.2}$	$9.89 \pm 0.05$	$19.0 \pm 4.0$	$0.6 \pm 0.3$	$8.01^{+0.20}_{-0.25}$	$9.97^{+0.20}_{-0.25}$	$5.9 \pm 3.0$
$1.00 < z < 1.40$	1394	$11.16 \pm 0.15$	$1.9^{+0.4}_{-0.3}$	$10.20 \pm 0.05$	$22.0 \pm 3.0$	$1.1 \pm 0.4$	$8.05^{+0.20}_{-0.24}$	$10.08^{+0.20}_{-0.24}$	$7.3 \pm 3.0$
$1.50 < z < 2.20$ <sup>(‡)</sup>	997	$11.04 \pm 0.15$	$1.5^{+0.4}_{-0.4}$	$10.35 \pm 0.13$	$23.5 \pm 2.0$	$1.8 \pm 1.2$	$8.00^{+0.16}_{-0.19}$	$9.96^{+0.16}_{-0.19}$	$8.4 \pm 3.0$

**Notes.** Column 1: Redshift bin. Column 2: Number of objects in the redshift bin. Column 3: mean stellar mass. Column 4: mean SFR(*UV*). Column 5: total infrared luminosity from the SED fitting with Draine & Li (2007) models integrated over 8 – 1000  $\mu\text{m}$ . Column 6: ‘luminosity-weighted’  $T_d$  from the SED modelling with a single-temperature modified black body. Column 7: mean intensity of the radiation field from Draine & Li (2007) models. Column 8: dust mass. Column 9: molecular gas fraction assuming a solar gas-to-dust conversion factor of GDR = 92. All quantities are computed using a Salpeter (1955) IMF. <sup>(†)</sup>Mean and interquartile range. <sup>(‡)</sup>Stacked photometry from Gobat et al. (2018) that for consistency was re-fitted here with the same method and setup as for the rest of the redshift bins.

contaminated by a fraction of galaxies closer to the MS, given the looser constraints on the IR emission. However, most of the contamination in the quiescent sample is due to dusty starbursts (or transitioning galaxies), rather than blue objects on the MS with undetectable IR emission. Indeed, absent IR detections, the median sSFR(*UV*) of the final stacked samples of massive quiescent objects is constant over the redshift interval under consideration. Therefore, the median location with respect to the MS decreases with redshift, given the steady increase in the normalisation of the former (Fig. 2).

For the purpose of stacking, we further excluded the 1% of the sample that lay too close to the edges of the IR image mosaics. Finally, we binned our sample by redshift, as shown

in Fig. 2. We split the galaxies selected with the *rJK* colour at  $z = 0.65$ , the mid-point of the redshift interval over which this criterion is effective. In this work we focus on a complete sample of massive galaxies at  $0.3 < z < 1.5$  ( $\log(M_*/M_\odot) \geq 10.83$ )<sup>2</sup>, matching the median stellar mass of similarly selected objects at  $\langle z \rangle = 1.8$  from G18. This allows us to consistently trace the evolution of the dust content over 10 Gyr of cosmic evolution. We report the properties of the stacked sample in Table 1.

<sup>2</sup> This corresponds to a  $\log(M_*/M_\odot) \geq 10.6$  mass cut in the original Laigle et al. (2016) catalogue, which is built assuming a Chabrier IMF.

### 3. Source stacking and extraction

Our stacking and source extraction methodology broadly follows that described in G18, namely: For each redshift bin, we extracted cutouts, centered on each galaxy, from publicly available FIR imaging of the COSMOS field at  $24\ \mu\text{m}$  (Le Floch et al. 2009),  $100\ \mu\text{m}$ , and  $160\ \mu\text{m}$  (Lutz et al. 2011),  $250\ \mu\text{m}$ ,  $350\ \mu\text{m}$ , and  $500\ \mu\text{m}$  (Oliver et al. 2012),  $850\ \mu\text{m}$  (Geach et al. 2017),  $1.1\ \text{mm}$  (Aretxaga et al. 2011),  $10\ \text{cm}$  (Smolčić et al. 2017), and  $20\ \text{cm}$  (Schinnerer et al. 2010). For each band, we then combined the cutouts to create a median image, ignoring masked (e.g., zero-coverage) regions, and estimated the variance of each pixel through bootstrap resampling of the data with half the sample size. We note that these variance cutouts do not show any appreciable structure (i.e. they are flat), suggesting that there is little to no contribution from clustering to the noise (e.g., Béthermin et al. 2012).

As in G18, we model the emission in each median image with a combination of three components: point-like emission from the central QG, an autocorrelation term due to source overlap within the sample, and extended emission from SFGs associated with the central QGs. This includes both true SF satellites of the QGs and the two-halo term arising from clustering with more massive SFGs. For simplicity, we hereafter refer to it as the ‘satellite halo’. Finally, we also include a flat background term. As the point-source and autocorrelation terms arise from the same galaxy population their amplitudes are tied and we determine their relative scaling beforehand. To model the combined central point-like emission and autocorrelation signal, we first created simulated images of the COSMOS field for each redshift bin, consisting of a blank map populated by model or empirical beams (depending on the data) located at the position of each QG and normalised to the same unitary flux. This latter approximation was used because the scatter of the FIR flux of our QGs is unknown by construction of the sample. Consequently, we do not associate an uncertainty with the shape of the autocorrelation signal. The simulated sources were then extracted using the same procedure as described above. This yields median cutouts that are essentially indistinguishable from point-sources in the case of higher-resolution data (e.g., at  $24\ \mu\text{m}$ ) but deviate substantially from it when the beam is large, due to the blending of sources. In particular, the SPIRE ( $250\ \mu\text{m}$ ,  $350\ \mu\text{m}$ , and  $500\ \mu\text{m}$ ) ‘point’ stacks have therefore extended wing-like emission and higher peak values than a simple beam, while the  $850\ \mu\text{m}$  and  $1.1\ \text{mm}$  point stacks show an opposite effect, with lower peak values, arising from the deep negative rings in their PSFs. This additional step, which accounts for the intrinsic clustering of QGs, is necessary here since we did not eliminate projected pairs during selection, as in G18. Indeed, excluding from the sample all objects that partially overlap in the lowest-resolution maps (SPIRE) would more than halve it<sup>3</sup>.

On the other hand, the distribution of FIR emission from satellite halos is estimated using multi-wavelength catalogs available for the COSMOS field (Muzzin et al. 2013b; Laigle et al. 2016). Satellites were selected in the catalogue as objects that both satisfy the *UVJ* criterion (Williams et al. 2009) for SFGs and have compatible photometric redshifts, namely  $z_{L68} < z_{\text{cen}} < z_{H68}$ , where  $z_{\text{cen}}$  is the redshift of the central QG and  $z_{*68}$  are the lower and upper 68% confidence limits to the redshift of satellites, respectively. We estimated the SFR of each satellite candidate by fitting its rest-frame ultraviolet SED with constant

star formation history stellar population templates (Bruzual & Charlot 2003), with and without dust extinction. The difference between these two estimates yields an obscured SFR surface density that, when convolved with the instrumental beam in each band, is assumed to be proportional to the FIR flux distribution from the satellite halo. More details are given in Appendix A.

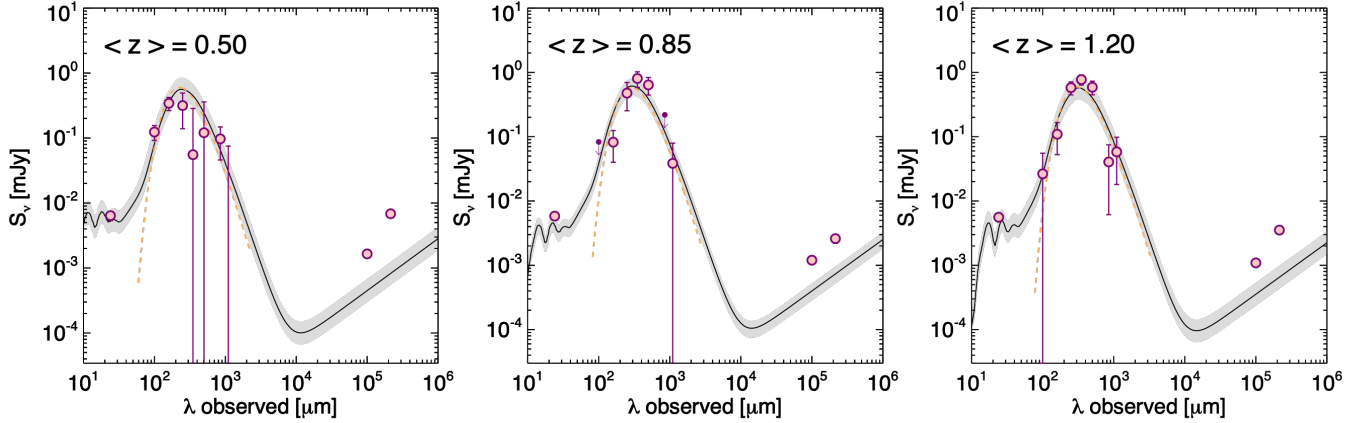
To extract the FIR SED of the QGs in each redshift bin, we fitted the cutouts in each band with combinations of these two components plus a flat background term. We let the scaling of the satellite emission vary freely at  $24\ \mu\text{m}$ ,  $100\ \mu\text{m}$ , and  $160\ \mu\text{m}$ , where the resolution is high enough to properly decompose the observed signal into point-like and extended components.

Satellite fluxes in these three bands were then fitted with a set of redshift-dependent MS IR templates (Magdis et al. 2012) that capture the average value and the scatter of  $T_d$  and  $\langle U \rangle$  of MS galaxies at a given redshift (e.g., Magdis et al. 2012; Béthermin et al. 2015). From these templates (and their scatter) we extrapolated satellite fluxes (and their uncertainties) at  $\geq 250\ \mu\text{m}$ , where both the point-like and halo emission are extended and the large beams preclude a non-degenerate decomposition. For these bands, the scaling of the satellite model was fixed to the extrapolated flux, with only the background and central source being fitted. The satellite fluxes thus removed from the FIR SED correspond to  $\log(L_{\text{IR}}/L_{\odot}) = 10.40 \pm 0.09$ ,  $10.70 \pm 0.07$ , and  $9.30 \pm 0.10$  for the high-, intermediate-, and low-redshift bins, respectively. These estimates are consistent with the predicted  $L_{\text{IR}}$  infrared from the obscured SFR of the satellites as inferred by the *UV/optical* modelling ( $\log(L_{\text{IR}}/L_{\odot}) = 10.47 \pm 0.13$ ,  $10.77 \pm 0.09$ , and  $9.70 \pm 0.38$ ). Finally, we note that while these  $L_{\text{IR}}$  values are comparable to or higher than the  $L_{\text{IR}}$  of QGs, the sum of the autocorrelation and the QG component still dominates the  $\geq 250\ \mu\text{m}$  bands due to a combination of lower  $T_d$  and large autocorrelation term.

The fits of the various components were done with pixel weights derived from the stacked variance maps described at the beginning of this section. However, these cannot be used to estimate accurate uncertainties on the extracted fluxes as the pixel-to-pixel variance is necessarily correlated in the FIR due to confusion. Therefore, rather than considering the (small) formal errors on scaling of the combined point-source and autocorrelation models, uncertainties on central fluxes were estimated by fitting instrumental beams at random positions in the cutouts, well clear of both the central source and satellite emission. To these we added in quadrature the error of both the satellite fluxes and background, in the corresponding band. We note that the autocorrelation term alone is free of uncertainty by construction, the sample being set and the beam models in  $\geq 250\ \mu\text{m}$  bands (where it is not negligible) being theoretical and provided without associated errors. The uncertainties for each decomposition component are presented in Table A.1.

Finally, the measured central QG fluxes were rescaled down to account for contamination due to random clustering in the image using the corrections described in Béthermin et al. (2015). We note that the background term might not necessarily be flat, due to lensing from the QGs’ host dark matter halo. Magnification bias is typically dominated by area increase, so its net result is a deficit in faint source counts with respect to the field. Assuming that the QGs are central to their halo, this should lead to a lower effective background towards the centre of the cutouts, thus slightly underestimated central fluxes. However, its magnitude depends on the shape of the luminosity distribution of background sources, for which we have little to no information. Furthermore, this effect can be mild even for massive galaxy clusters (e.g., Umetsu et al. 2014). In this case, where the host

<sup>3</sup> As a sanity check we have validated that an analysis excluding the pairs yields consistent results albeit with significantly higher uncertainties.



**Fig. 3.** Observed median IR SEDs of massive QGs. The magenta circles mark the observed photometry, with arrows indicating  $3\sigma$  upper limits for bands where the measured flux density was negative. We note that the radio points at 10 and 20 cm were not included in the modelling. The black line and the grey shaded areas indicate the best-fit model from the *Draine & Li (2007)* templates and the associated uncertainty. The dashed orange line is the best fit adopting a single  $T_d$  modified black body. The *left*, *middle*, and *right* panels show the results for the redshift intervals  $0.3 < z < 0.65$ ,  $0.65 < z < 1.0$ , and  $1.0 < z < 1.4$ , respectively. Salient properties of the sample are reported in Table 1.

**Table 2.** Stacked photometry of the QG population, as derived through flux density decomposition.

Redshift	24 $\mu\text{m}$	100 $\mu\text{m}$	160 $\mu\text{m}$	250 $\mu\text{m}$	350 $\mu\text{m}$	500 $\mu\text{m}$	850 $\mu\text{m}$	1100 $\mu\text{m}$	3.0 GHz	1.4 GHz
(1)	[ $\mu\text{Jy}$ ]	[mJy]	[mJy]	[mJy]	[mJy]	[mJy]	[mJy]	[mJy]	[ $\mu\text{Jy}$ ]	[ $\mu\text{Jy}$ ]
$\langle 0.50 \rangle$	$6.40 \pm 1.47$	$0.12 \pm 0.04$	$0.34 \pm 0.08$	$0.32 \pm 0.21$	$0.06 \pm 0.23$	$0.12 \pm 0.23$	$0.1 \pm 0.04$	$< 0.225^{(\dagger)}$	$1.65 \pm 0.12$	$6.84 \pm 0.63$
$\langle 0.85 \rangle$	$5.82 \pm 0.61$	$< 0.08^{(\dagger)}$	$0.08 \pm 0.04$	$0.47 \pm 0.22$	$0.80 \pm 0.21$	$0.64 \pm 0.19$	$< 0.21^{(\dagger)}$	$0.04 \pm 0.04$	$1.20 \pm 0.07$	$2.60 \pm 0.41$
$\langle 1.20 \rangle$	$5.60 \pm 0.79$	$0.30 \pm 0.30$	$0.11 \pm 0.05$	$0.57 \pm 0.13$	$0.77 \pm 0.14$	$0.59 \pm 0.14$	$0.04 \pm 0.03$	$0.06 \pm 0.04$	$1.09 \pm 0.07$	$3.53 \pm 0.43$

**Notes.**  $^{(\dagger)}$  $3\sigma$  upper limits following the prescription of *Béthermin et al. (2020)*.

halos are smaller by a factor of  $\sim 10$  (*G18*), we do not expect it to be significant. The final photometry of our stacked QGs is presented in Table 2 while the various decomposition components for each redshift bin and each broadband photometry are presented in Appendix A.

#### 4. SED modelling

To derive the FIR properties of the stacked samples, we employed the dust models of *Draine & Li (2007; DL07)*, with the standard parameterisation used in previous studies (e.g., *Magdis et al. 2012*). Namely, we considered diffuse ISM models with radiation field intensities ranging from  $0.1 \leq U_{\min} \leq 50$ , while fixing the maximum radiation field to  $U_{\max} = 10^9$ . We allowed for  $\gamma$  (the fraction of dust exposed to starlight with intensities ranging from  $U_{\min}$  to  $U_{\max}$ ) to vary between  $0 \leq \gamma \leq 0.5$  with a step of 0.01 and adopted Milky Way dust models with  $q_{\text{PAH}}$  (i.e. the fraction of the dust mass in the form of PAH grains) from 0.4% to 4.6%. To account for the artificial SED broadening due to the redshift distribution of the stacked sources, we constructed an average SED for each DL07 model by co-adding each original DL07 model at each redshift within the bin, weighted by the redshift distribution of stacked sources. The redshift-weighted, ‘smoothed’ models were shifted to the median redshift of each bin and were fit to the data, yielding best-fit  $L_{\text{IR}}$ ,  $M_{\text{dust}}$ , and  $\langle U \rangle$  values.

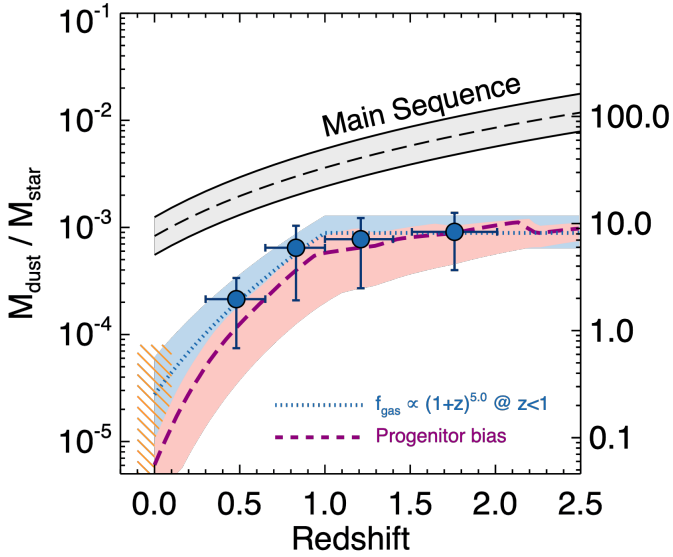
To estimate the uncertainties of the derived parameters we bootstrapped and fitted with same methodology 1000 artificial realisations of the real SEDs, by perturbing the extracted photometry within  $1\sigma$  for the detections ( $S/N \geq 3$ ) and within flux  $+3\sigma$  for the rest, following the methodology presented in

*Béthermin et al. (2020)*. As uncertainty for each derived parameter from the real data, we adopted the standard deviation of the values of the corresponding parameter inferred from the perturbed SEDs. As a sanity check we also introduced a bootstrapping component in our analysis by randomly excluding one observed data point at each realisation. We also confirm that the data sets of available detections (which include at least one data point at  $\lambda_{\text{rest}} \geq 220 \mu\text{m}$  for all four stacked ensembles, as well as a detection at  $24 \mu\text{m}$  and at least one detection in the Wien part of the SED), along with the upper limits, ensure against possible systematic effects in the derivation of  $M_{\text{dust}}$  and  $L_{\text{IR}}$  for all redshift bins (see Appendix D). Finally, to derive ‘luminosity-weighted’  $T_d$  we complemented our analysis by considering a single-temperature optically thin modified blackbody (MBB) models with a fixed effective dust emissivity index of  $\beta = 1.8$ .

The best fit models to the data ( $24 \mu\text{m}$  to 1.1 mm for DL07 and  $\lambda_{\text{rest}} > 50 \mu\text{m}$  to 1.1 mm for MBB), along with the range of SEDs from the random realisations, are shown in Fig. 3. The best fit parameters and their corresponding uncertainties are summarised in Table 1. While the radio data points are not included in the fit, we add to each the best-fit DL07 model a power-law radio slope with a spectral index  $\alpha = 0.8$  and a normalisation given by the FIR–radio correlation (e.g., *Delhaize et al. 2017*), to investigate (and visualise) possible radio excess in our stacked ensembles.

#### 5. Results

Using the derived parameters from the SED modelling in the previous section, we attempted to characterise the dust and gas content of QGs and investigate possible evolutionary trends. For the



**Fig. 4.** Evolution of the dust to stellar mass ratio and of the gas fraction of QGs. The filled blue circles mark the  $M_{\text{dust}}/M_*$  of the stacked QGs in the  $0.3 < z < 2.0$  range, including the results from G18. The grey shaded region depicts the trend for MS galaxies from Kokorev et al. (2021). The orange shaded region captures the range of  $M_{\text{dust}}/M_*$  for local QGs, as drawn from the literature (see text). The dotted blue line (blue shaded region) corresponds to the best fit (scatter) to the  $0.0 < z < 1.0$   $M_{\text{dust}}/M_*$  data of QGs, with a functional form of  $M_{\text{dust}}/M_* \propto (1+z)^{5.0}$ . The bold, dashed purple line (shaded region) depicts the evolution (scatter) of  $M_{\text{dust}}/M_*$  (and of  $f_{\text{gas}}$ ) as derived from the progenitor bias analysis of Gobat et al. (2020), which is discussed in Sects. 5.2 and 6. The conversion of  $M_{\text{dust}}/M_*$  to  $f_{\text{gas}}$  assumes solar metallicity and  $\text{GDR} = 92$ .

rest of the analysis we will be referring to the  $\langle z \rangle = 0.50, 0.85$ , and  $1.20$  stacks as low- $z$ , mid- $z$ , and high- $z$  quiescent sample respectively. We also complemented our study with the stacked ensemble of G18 at  $\langle z \rangle = 1.75$ , which has been constructed and analysed using similar sample selection, stacking, photometry, and SED modelling techniques as those adopted here. Overall, this work covers the average properties of QGs in the redshift range  $0.3 < z < 2.0$  with a homogeneous and systematic approach.

### 5.1. Basic FIR properties

The available mid-IR to millimetre photometry that spans from observed 24- to  $1100 \mu\text{m}$  allows for a robust determination of the total 8– $1000 \mu\text{m}$  infrared luminosities ( $L_{\text{IR}}$ ), dust masses ( $M_{\text{dust}}$ ), mean radiation field (equivalent of dust mass-weighted luminosity or temperature), parametrised as  $\langle U \rangle \propto M_{\text{dust}}/L_{\text{IR}}$ , and luminosity-weighted dust temperatures ( $T_{\text{d}}$ ). Our SED fitting yields low infrared outputs for our stacked ensembles at all redshifts with  $\log(L_{\text{IR}}/L_{\odot}) = 9.40\text{--}10.30$ , dust masses in the range of  $\log(M_{\text{dust}}/M_{\odot}) = 7.50\text{--}8.15$ , low mean radiation field  $\langle U \rangle = 0.6\text{--}1.2$ , and cold  $T_{\text{d}} < 23$  K dust temperatures.

In comparison to previous studies that presented stacked SEDs of QGs at similar redshifts and equivalent  $M_*$  bins, our analysis yields significantly lower  $L_{\text{IR}}$  and  $M_{\text{dust}}$  estimates. For example, fitting the photometry presented in Man et al. (2016) with same methodology as for our data, we find  $\log(L_{\text{IR}}/L_{\odot}) = 9.8\text{--}11.2$  and  $\log(M_{\text{dust}}/M_{\odot}) = 8.0\text{--}8.6$  for the stacked ensembles of QGs at  $z = 0.4\text{--}1.6$ . These values are 0.2–0.9 dex larger than the corresponding parameters inferred from our analysis for

our sample. These discrepancies stem from the more conservative sample selection that we adopted in this work (double-colour criterion and exclusion of individually detected  $24 \mu\text{m}$  sources). In addition, the adopted stacking and, especially, the treatment and correction for the contribution of blended satellites further affects the comparison with Man et al. (2016, see also G18). Nevertheless, and in agreement with Man et al. (2016), we do find a radio excess between the observed radio fluxes and those inferred by the infrared-radio correlation of SFGs by Delhaize et al. (2017), indicative of AGN activity and ‘radio-mode’ feedback in massive QGs (e.g., Gobat et al. 2018; Barišić et al. 2017).

In order to test for the possible sensitivity of these recovered parameters to the selection criteria described in Sect. 2, as well as possible contamination of quiescent samples by dusty star forming objects, we performed a series of tests using subsamples obtained by either applying an additional 0.1 mag colour margin or by dividing the low-, mid-, and high- $z$  bins into blue and red halves. Applying the same stacking, source extraction, and modelling procedures, as described in Sects. 3 and 4, to these yields SEDs that are entirely consistent with the full low-, mid-, and high- $z$  samples. This suggests that our analysis is robust with respect to selection effects. More details are given in Appendix B.

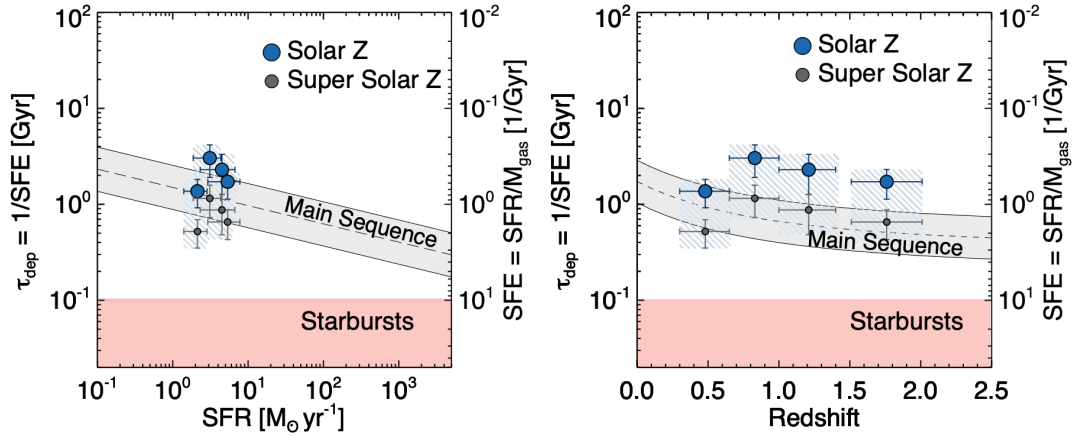
Finally, we inferred gas masses ( $M_{\text{gas}}$ ) estimates by converting the derived  $M_{\text{dust}}$  to  $M_{\text{gas}}$  through the metallicity-dependent dust to gas mass ratio (GDR) equation of Magdis et al. (2012);  $\log(\text{GDR}) = 10.54 - 0.99 \times (12 + \log(\text{O}/\text{H}))$ . We considered a range of metallicities  $Z_{\odot} \leq Z \leq 2 Z_{\odot}$  ( $8.66 \leq 12 + \log(\text{O}/\text{H}) \leq 9.1$  in the Pettini & Pagel 2004 scale), appropriate for the SFR and  $M_*$  of our samples (e.g., Mannucci et al. 2010). For these metallicities the GDR varies between  $35 \leq M_{\text{gas}}/M_{\text{dust}} \leq 92$ . For the rest of our analysis we use the  $M_{\text{gas}}$  estimates based on a solar metallicity as benchmark and those for a super-solar metallicity as lower limits. Therefore, throughout the paper,  $M_{\text{gas}} = M_{\text{gas}}^{\text{sol}}$  (and  $f_{\text{gas}} = f_{\text{gas}}^{\text{sol}}$ ), unless otherwise stated. We note that since the  $M_{\text{dust}}/M_{\text{gas}} - Z$  relation of Magdis et al. (2012) is almost linear (slope of 0.99), we can easily rescale  $M_{\text{gas}}$  estimates between solar and an arbitrary metallicity  $Y$  by using:

$$\log[M_{\text{gas}}^Y] = \log[M_{\text{gas}}^{\text{sol}}] + [8.66 - Y], \quad Y = 12 + \log(\text{O}/\text{H}). \quad (3)$$

We also recall that the GDR –  $Z$  relation yields  $M_{\text{gas}}$  estimates that trace the sum of the atomic ( $M_{\text{HI}}$ ) and the molecular ( $M_{\text{H}_2}$ ) gas mass (including a  $\times 1.36$  He contribution). As such,  $M_{\text{gas}} = M_{\text{HI}} + M_{\text{H}_2}$ . While the exact state of the cold gas in our QGs cannot be characterised, the  $\langle M_{\text{H}_2}/M_{\text{HI}} \rangle > 1$  ratios reported in the literature for  $z > 0.5$  and for high stellar surface densities (e.g., Blitz & Rosolowsky 2006; Bigiel et al. 2008; Obreschkow et al. 2009) motivate us to make the simplified assumption that  $M_{\text{H}_2} \gg M_{\text{HI}}$  and, thus,  $M_{\text{H}_2} \approx M_{\text{gas}}$ . We note though, that this estimate is an upper limit to the molecular gas mass of our samples.

### 5.2. Evolution of dust and gas mass fractions

In Fig. 4 we bring together the  $M_{\text{dust}}$  estimates for our stacked ensembles as well as a range of literature  $M_{\text{dust}}$  estimates for local QGs of comparable  $M_*$  (Smith et al. 2012; Boselli et al. 2014; Lianou et al. 2016; Michałowski et al. 2019) and explore the evolution of dust mass fraction,  $M_{\text{dust}}/M_*$ , as a function of redshift. Our data indicate that the  $M_{\text{dust}}/M_*$  remains roughly constant from  $z = 2.0$  to  $z = 1.0$ , followed by sharp decline towards the present day, with a functional form of  $M_{\text{dust}}/M_* \propto (1+z)^{5.00\text{--}1.94}$  (for  $z < 1.0$ ). As shown in Fig. 4, a similar



**Fig. 5.** Gas depletion timescales. *Left*: gas depletion time scale ( $t_{\text{dep}}$ ) as a function of SFR. The blue (grey) circles correspond to the stacked samples of QGs for the case of solar (super solar) metallicity, assuming  $\text{GDR} = 92$  (35). The dashed grey line and the grey shaded region depict the trend for MS galaxies from Sargent et al. (2014) with a scatter of 0.23 dex. The magenta shaded region marks the area with  $\tau_{\text{dep}} < 100$  Myr, typical of star-bursting galaxies. *Right*: gas depletion time scale as a function of redshift. The symbols are the same as in the left panel. The dashed grey line and the grey shaded region depict the trend of Liu et al. (2019) for MS galaxies at fixed  $M_* = 2 \times 10^{11} M_{\odot}$  and its corresponding scatter of 0.23 dex.

trend characterises the evolution of the dust mass fraction in SFGs Kokorev et al. (2021), albeit with a shallower slope. Evidently, QGs exhibit systematically lower dust mass fractions with respect to SFGs at all redshifts, indicative of dust consumption, destruction or expulsion during the quenching phase. However, the flat evolution for QGs between  $z = 2.0$ – $1.0$  and the monotonically decreasing  $M_{\text{dust}}/M_*$  for SFGs at all redshift, results in a minimum deviation between the two populations at  $z \sim 1.0$ .

The evolution of  $M_{\text{dust}}/M_*$  we find here also mirrors the evolution of the gas fraction of QGs. Under the assumption of constant solar metallicity for massive QGs at all redshifts, we can directly convert the  $M_{\text{dust}}/M_*$  presented in the left y-axis of Fig. 4 to  $f_{\text{gas}}$ . This approach yields a constant  $f_{\text{gas}} \approx 7$ – $8\%$  between  $z = 2.0$ – $1.0$  that subsequently drops to  $f_{\text{gas}} \approx 2\%$  at  $z = 0.5$  and  $f_{\text{gas}} \leq 1\%$  in the local Universe. These values are in agreement with independent  $M_{\text{gas}}$  estimates or upper limits inferred by CO observations of individual QGs and post-starburst galaxies at various redshifts (e.g., French et al. 2015; Sargent et al. 2015; Rudnick et al. 2017; Suess et al. 2017; Spilker et al. 2018; Bezanson et al. 2019). We note that for a super solar metallicity of  $2 \times Z_{\odot}$  the  $f_{\text{gas}}$  values presented in Fig. 4 would correspond to  $f_{\text{gas}}^Z \approx 0.4 \times f_{\text{gas}}^{\text{sol}}$ . In Sect. 6, we will interpret the observed redshift evolution of  $M_{\text{dust}}/M_*$  and  $f_{\text{gas}}$  of QGs in terms of progenitor bias and ageing of the stellar population.

### 5.3. SFR and gas depletion time scales

The gas masses inferred in the previous section, when coupled with the corresponding SFRs of the stacked ensembles, can be used to define the star formation efficiency ( $\text{SFE} = \text{SFR}/M_{\text{gas}}$ ) or, inversely, the gas depletion time scale ( $\tau_{\text{dep}} = 1/\text{SFE}$ ). To derive the total SFR in each redshift bin, we considered the average unobscured component as derived from the optical photometry ( $\text{SFR}_{UV}$ ) and the obscured star formation by converting  $L_{\text{IR}}$  to  $\text{SFR}_{\text{IR}}$  using the Kennicutt (1998) scaling relation. We caution the reader that for QGs a considerable fraction of  $L_{\text{IR}}$  could arise from dust heated by old rather than young stellar populations (e.g., Hayward et al. 2014), as assumed by the Kennicutt  $\text{SFR}$ - $L_{\text{IR}}$  relation. Therefore, the estimates of  $\text{SFR}_{\text{tot}} = \text{SFR}_{\text{IR}} + \text{SFR}_{UV}$  should be regarded as upper limits. To

constrain the fraction of  $L_{\text{IR}}$  arising from dust heated by old stars, we fit each galaxy’s optical-NIR SED with composite stellar population models derived from Bruzual & Charlot (2003) templates. For simplicity, we used delayed exponential star formation histories, starting at  $z_{\text{in}} = 10$  and with time scales  $\tau = 100 \text{ Myr} - 3 \text{ Gyr}$ , each truncated 1 Gyr before observation. Finally, we assumed a Noll et al. (2009) attenuation curve with variable slope and computed the luminosity difference of the best-fit model before and after attenuation. We find that this absorbed energy does not make up more than  $\sim 30\%$  of  $L_{\text{IR}}$  at  $z \gtrsim 1$  but could represent up to 100% of the infrared luminosity of  $z \sim 0.5$  QGs. Given the uncertainties and the small IR output indicated by our stacks, we chose to consistently consider  $\text{SFR}_{\text{tot}}$  at all redshift bins.

The emerging (maximum) SFRs are listed in Table 1 and they range from  $\sim 2.0 M_{\odot} \text{yr}^{-1}$  in the lowest redshift bin ( $\langle z \rangle = 0.5$ ) to  $\sim 6.5 M_{\odot} \text{yr}^{-1}$  at  $\langle z \rangle = 1.8$ . These average SFRs place our stacked ensembles  $\geq 10\times$  below the MS at their corresponding redshifts, confirming their passive nature. At the same time,  $\tau_{\text{dep}}$ , ranges from 1.3 to 3.0 Gyr for the case of solar metallicity with a mean  $\langle \tau_{\text{dep}} \rangle = 2.1 \pm 0.8 \text{ Gyr}$  and 0.5–1.1 Gyr for a super solar metallicity. These  $\tau_{\text{dep}}$  estimates are plotted as a function of SFR and redshift in the two panels of Fig. 5 along with the corresponding tracks for MS galaxies (Sargent et al. 2014; Liu et al. 2019). We stress again that since we are using the maximum SFR estimates the quoted  $\tau_{\text{dep}}$  should be regarded as lower limits for each metallicity. In agreement with Spilker et al. (2018), we find that distant QGs, for their SFRs, have a  $\tau_{\text{dep}}$  that is broadly consistent with that of local MS galaxies, but longer than that of MS galaxies at their corresponding redshift. We note that while our data hint upon an increase in  $\tau_{\text{dep}}$  from  $z = 2.0$  to  $z = 0.8$  followed by a decline towards later cosmic epochs (Fig. 5, right), the values are overlapping within their uncertainties yielding a non significant correlation. Possible metallicity effects as well as the uncertainties linked to the conversion of  $L_{\text{IR}}$  to SFR, will be investigated in detail in a future study.

### 5.4. SED and $T_d$ evolution

The star formation activity and the dust heating photons per unit ISM mass, both parametrised above in terms of SFR,  $M_{\text{dust}}$ , and

SFE, should also be imprinted in the shape of the FIR SED and the characteristic  $T_d$  of QGs. Indeed, a large body of literature that has focused on the evolution of  $T_d$  as a function of redshift for MS and star-bursting galaxies alike (e.g., Hwang et al. 2010; Magdis et al. 2012, 2017; Magnelli et al. 2014; Schreiber et al. 2018a; Jin et al. 2019; Cortzen et al. 2020) has established that the increase in sSFR and SFE at higher redshifts is also followed by a trend of warmer SEDs and increasing  $T_d$  at least out to  $z \sim 4$ . At the same time, SEDs and the effective  $T_d$  of distant QGs as a function of redshift are far less explored.

In Fig. 6 we plot the  $T_d$  of QGs derived from our MBB fits as a function of redshift, along with literature data and trends for MS galaxies. As expected, QGs exhibit colder  $T_d$  with respect to star forming galaxies at all redshifts, indicative of a lower star formation activity. Also, at first reading, the  $T_d$  estimates for our quiescent ensembles are overlapping within their uncertainties, with an average  $\langle T_d \rangle = 21.0 \pm 2.0$  K. Notably, in the redshift range probed by our data, SFGs have witnessed a decrease in their  $T_d$  by  $\sim 8$  K, indicative of a steep evolution that is not observed for QGs.

However, the hint of small  $T_d$  decrease in QGs between  $z = 2.0$  and  $z = 0.3$  ( $\Delta T_d \approx 3$  K), even if not statistically vigorous, provides the grounds for a tempting speculation of a weak evolution driven by their size evolution. Indeed, a size evolution of  $R \propto (1+z)^{-1.48}$  (van der Wel et al. 2014) yields a size increase of a factor of  $\sim 0.5$  between  $z = 1.8$  and  $0.8$ . Similarly, following the analysis of Chanial et al. (2007),  $T_d^{4+\beta} \propto [\Sigma_{\text{IR}}]^\gamma$ , where  $\gamma = \frac{0.4}{1.4}$  and  $\Sigma_{\text{IR}} \propto L_{\text{IR}}/R^2$ . For fixed  $L_{\text{IR}}$  (or SFR), as in the case for our stacked samples in this redshift range, it follows that  $T_d(z_1)/T_d(z_2) \propto [R(z_2)/R(z_1)]^{2 \times \gamma / (4+\beta)}$ . Anchoring this equation at  $T_d(z = 1.8) = 23.5$  K, as inferred by our analysis and adopting  $\beta = 1.8$ , provides an evolutionary track that is in broad agreement with our measurements for the  $0.3 < z < 2.0$  stacks. We conclude that a weak (if any) evolution of  $T_d$  in QGs can be understood in terms of size growth. At the same time an average  $\langle T_d \rangle = 21.0 \pm 2.0$  K provides a statistically robust estimate for a characteristic  $T_d$  for the ISM of QGs over the last ten billion years.

In addition, the small variance in  $T_d$  and  $L_{\text{IR}}/M_{\text{dust}}$ , suggests that massive QGs at various redshifts share a common FIR SED shape, albeit with a varying normalisation that mirrors the evolution of  $L_{\text{IR}}$  and of the ISM mass budget as described in the previous sections. To produce an average SED, we bring the stacked photometry of each redshift bin at rest-frame, normalise it so that all SEDs are anchored at the same reference luminosity of  $1.0 \times 10^{10} L_\odot$ , and fit with DL07 and MBB models following the methodology presented in Sect. 4. The observed points (and upper limits) along with the best-fit models are presented Fig. 6 (right). The emerging average SED of QGs has  $T_d = 20$  K, and  $L_{\text{IR}}/M_{\text{dust}} \approx 90 L_\odot/M_\odot$ , in direct contrast to that of typical  $z > 0$  star forming galaxies. As an example, we consider the average SED of  $z = 1.0$  MS galaxies from Magdis et al. (2012), which clearly peaks at shorter  $\lambda$  with a  $T_d \approx 30$  K and has approximately a factor of  $\sim 10\times$  larger infrared energy output per unit dust mass ( $L_{\text{IR}}/M_{\text{dust}} \approx 1000 L_\odot/M_\odot$ ). Finally, we stress that the radio data are not included in the fit but are presented to demonstrate the radio-excess that we have already briefly discussed in a previous section. The average SED of QGs (normalised to  $L_{\text{IR}} = 1.0 \times 10^{10} L_\odot$  and  $M_{\text{dust}} = 1.11 \times 10^8 M_\odot$ ) is publicly available<sup>4</sup>.

## 6. Discussion

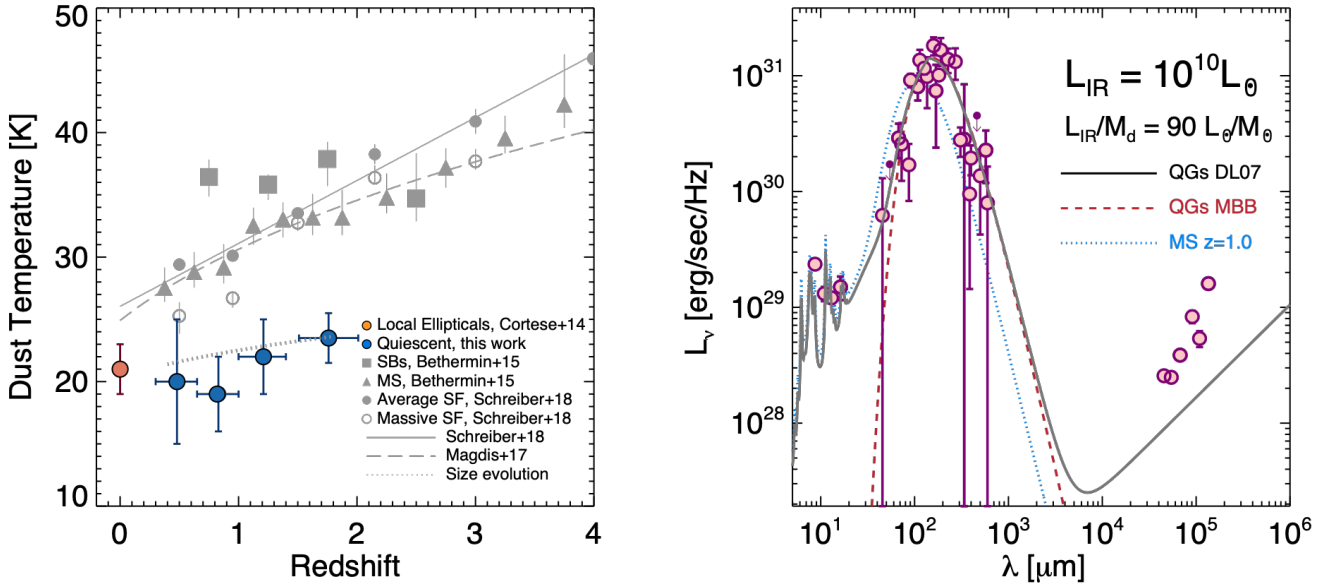
In the previous sections, we characterised the mass budget and the properties of the ISM of QGs at various redshifts. Here we offer an explanation for the origin and the drivers of the observed trends.

To interpret the evolution of  $M_{\text{dust}}$  in our quiescent sample, we first need to consider the expected redshift evolution of a population of QGs. At  $\log(M_*/M_\odot) \sim 11.0$ , their co-moving number density increases significantly between  $z \sim 4$  and  $z \sim 1$ , with little evolution after that (e.g., Muzzin et al. 2013a; Davidzon et al. 2017). This implies that an unbiased sample of QGs will contain a large fraction of newly formed objects at any redshift  $z \gtrsim 1$  while, few new objects enter the  $z < 1$  population. For simplicity, let us neglect dry merging events and equate the formation time of a QG with the quenching time of its precursor. In this case, the average age of the quenching event will vary little between quiescent galaxy samples at  $z \gtrsim 1$  and increase with cosmic time at  $z \lesssim 1$ . In this context, the apparent non-evolution of the dust fraction between  $z \sim 2$  and  $z \sim 1$ , where newly quenched galaxies constitute a large part of the quiescent population, suggests that whatever quenching mechanisms are at play, they leave behind very similar ISM conditions independently of redshift. Conversely, the steep decrease in the dust fraction between  $z \sim 1$  and  $z = 0$  is consistent with a consumption (or destruction) without replenishment of the dust and gas as few to no new QGs enter the massive population in this redshift range. Combining both regimes yields an apparent evolution very similar to our observed constraints, as shown in Fig. 4. Here we estimated the median gas fraction of the whole population of QGs as a function of redshift by assuming only the QG production rate from stellar mass functions, an initial gas fraction (constant with redshift) of  $\sim 10\%$  immediately following quenching, and a closed-box consumption with a time scale of  $\sim 2$  Gyr after that (Gobat et al. 2020). The track derived in this manner reproduces our data well, especially when considering that the latter are likely slightly biased towards dustier objects at lower redshift (see Fig. 2), which would flatten the observed trend.

This ‘progenitor-bias’ explanation for the observed evolution of  $f_{\text{gas}}$  and  $M_{\text{dust}}/M_*$  would also require an anti-correlation between these quantities and the average quenching time, or more simply, the stellar age. Indeed, in a recent study, Michałowski et al. (2019), presented an exponential decline of the  $M_{\text{dust}}/M_*$  with age for a sample of galaxies at a fixed cosmic epoch ( $\langle z \rangle = 0.13$ ) and a range of luminosity-weighted stellar ages between  $9.0 < \log(\text{age}/\text{yr}) < 10$ . The authors report a decline of  $M_{\text{dust}}/M_*$  with age that follows  $e^{-\text{age}/\tau}$  and a dust lifetime  $\tau \approx 2.5$  Gyr. While we do not have any information about the actual stellar age of the galaxies in our stacks, the effect of varying age can be mimicked by the our different redshift bins and under the assumption that the passive galaxies at  $z > 0.9$  will, on average, remain passive at later times. Indeed the time interval between  $z = 0.9$  and  $z = 0.5$  corresponds to  $\approx 2.3$  Gyr. According to the formula of Michałowski et al. (2019), and assuming an average stellar age of 2–3 Gyr for the QGs at  $z = 0.9$ , we should expect a drop in  $\log(M_{\text{dust}}/M_*)$  by  $\sim 0.4$ – $0.6$  dex, in very good agreement with our observations. At the same time, as described above, the flat behaviour of the  $M_{\text{dust}}/M_*$  (and  $f_{\text{gas}}$ ) at  $z > 1.0$  could then be explained by the younger, and roughly constant stellar age of the QGs in the higher redshift bins (e.g.,  $\log[\text{Age}/\text{yr}] = 9.0$  at  $z = 2.0$ – $4.0$ , Toft et al. (2017), Valentino et al. (2020)).

Finally, the rather constant FIR SED shape (and  $T_d$ ) of QGs at all redshifts, indicate that  $L_{\text{IR}}$  and  $M_{\text{dust}}$  roughly scale

<sup>4</sup> <http://www.georgiosmagdis.com/software/>



**Fig. 6.** Dust temperatures and average SED. *Left:* the evolution of  $T_d$  in QGs. The filled blue circles mark the luminosity-weighted  $T_d$  for the stacked QGs at  $0.3 < z < 2.0$ , including the results from G18. Filled grey triangles and squares indicate SFGs on the MS and strong starbursts from Béthermin et al. (2015). Filled and open grey circles represent the mass-averaged values and the most massive galaxies ( $11.0 < \log(M_*/M_\odot) < 11.5$ , Salpeter IMF) for the star-forming sample of Schreiber et al. (2018a), respectively. The solid and dashed grey lines indicate the best fit models from Schreiber et al. (2018a) and Béthermin et al. (2015), extending the parameterisation of Magdis et al. (2012). The dotted grey line indicates the expected  $T_d$  track for a size evolution of QGs,  $R \propto (1+z)^{-1.48}$ . The track is normalised to the  $T_d$  of the  $z = 1.8$  sample. *Right:* composite and template FIR SED for  $z > 0$  QGs. The magenta circles (arrows) correspond to the photometry ( $3\sigma$  upper limits) of the stacked ensembles presented in Fig. 3 normalised to  $L_{\text{IR}} = 10^{10} L_\odot$  (and  $M_{\text{dust}} = 1.11 \times 10^8 M_\odot$ ). The DL07 and MBB model fits to the SED are shown with grey and red dashed lines. The dotted blue line correspond to the typical SED of  $z = 1.0$  MS galaxies from Magdis et al. (2012), normalised to the same  $L_{\text{IR}}$ . The QG template is publicly available at [www.georgiosmagdis.com/software/](http://www.georgiosmagdis.com/software/).

together irrespective of the time of quenching. If indeed  $L_{\text{IR}}$  is a proxy of SFR and  $M_{\text{dust}}$  a proxy of  $M_{\text{gas}}$  this means that the decrease in  $M_{\text{gas}}$  is closely followed by a decrease in SFR. In this case, the universality of the initial ISM conditions, as advocated by the evolution of ISM mass budget (in terms of  $f_{\text{gas}}$  and  $M_{\text{dust}}/M_*$ ), is followed by a regularity in the internal processes within quenched galaxies.

We should stress that while the progenitor bias offers an appealing empirical explanation for the observed evolution of the ISM properties of QGs, it bares little, if any, information about the underlying physics of the origin and the maintenance of quenching. For the latter, the radio excess present in our SEDs as well as in previous studies (e.g., Man et al. 2016; D’Eugenio et al. 2020), point towards ‘radio-mode’ feedback from AGN-driven jets as compelling quenching mechanism that can both prevent gas accretion due to heating of the halo gas and expel a fraction of the cold gas reservoir (e.g., Croton et al. 2006; Gobat et al. 2018). This radio-mode feedback would also need to be stronger at higher redshift to compensate for the higher (putative) gas fractions seen at these epochs. Using the average 1.4 GHz excess, with respect to the FIR-radio correlation, and the evolution of the AGN luminosity function (Novak et al. 2018), we indeed estimate that the AGN duty cycle for our QGs decreases from  $>50\%$  at  $z \sim 1.8$  (see G18) to  $\sim 15\%$  in the low- $z$  redshift bin (see Appendix C).

### 6.1. Outlook and ALMA predictions

From the discussion above, it stems that the  $M_{\text{dust}}$  (and subsequently  $M_{\text{gas}}$  and  $f_{\text{gas}}$ ) of QGs depends both on redshift and on time elapsed since the quenching event. So while our  $M_{\text{dust}}$  and  $M_{\text{gas}}$  estimates should be valid for the average QGs population at

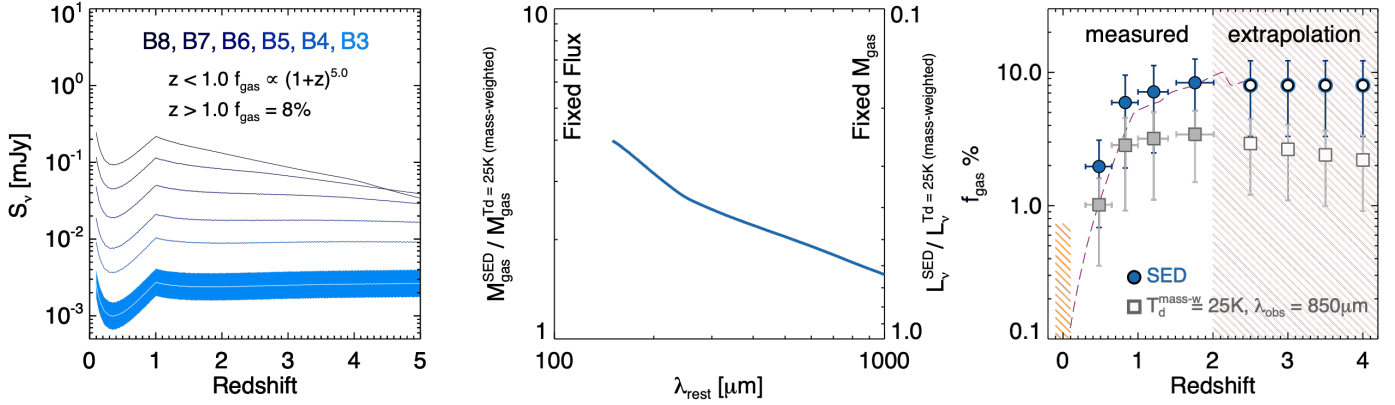
a given redshift, this might not be the case for individual sources due to possible age variations. Clearly, the way forwards necessitates focused ISM studies for large and representative samples of individual QGs at various cosmic epochs. In this regard, the IR SED template presented in Sect. 5, can facilitate a series of useful predictions.

The main characteristic of the template is the ratio of its native infrared luminosity ( $L_{\text{IR}}^0$ ) to its native dust mass ( $M_{\text{dust}}^0$ ), which is expressed as  $L_{\text{IR}}^0/M_{\text{dust}}^0 = 90 L_\odot/M_\odot$ . Consequently, for any arbitrary normalisation of the template, it follows that for this template,  $L_{\text{IR}} = 90 \times M_{\text{dust}}$  [in  $L_\odot$ ]. Similarly,  $M_{\text{gas}}$ , for a fixed metallicity  $Z$ , scales linearly with  $M_{\text{dust}}$ , with  $M_{\text{gas}} = \text{GDR}(Z) \times M_{\text{dust}}$  (e.g., Leroy et al. 2011; Magdis et al. 2012). Therefore, the template can be rescaled to any arbitrary  $M_{\text{dust}}$  that corresponds to a given  $M_{\text{gas}}$  (or  $f_{\text{gas}}$ ). Indeed, since  $f_{\text{gas}} = M_{\text{gas}}/M_* = \text{GDR}(Z) \times M_{\text{dust}}/M_*$  we simply have to multiply the SED by  $N = M_{\text{dust}}/M_{\text{dust}}^0$ , which takes the form of:

$$N = \frac{f_{\text{gas}} \times M_*}{\text{GDR}(Z) \times M_{\text{dust}}^0}. \quad (4)$$

Assuming that our SED is representative of massive QGs with a typical  $\log(M_*/M_\odot) = 11.2$  at all redshifts and by adopting a universal solar metallicity that corresponds to  $\text{GDR}(Z_\odot) = 92$  (e.g., Leroy et al. 2011; Magdis et al. 2012), the only free parameter in the SED normalisation is  $f_{\text{gas}}$ . Then we can use our normalised template SED to calculate the emitted flux density at the central wavelength of various ALMA bands at various redshifts for any adopted  $f_{\text{gas}}$  after implementing CMB effects as prescribed in daCunha et al. (2013).

As an example, we examine the case where  $f_{\text{gas}}$  is evolving with redshift as  $f_{\text{gas}} \propto (1+z)^{5.0}$  at  $z < 1.0$  and then remains flat



**Fig. 7.** ALMA outlook and systematics. *Left:* ALMA flux densities in various bands for an evolving  $f_{\text{gas}} \propto (1+z)^{5.0}$  at  $0.0 < z < 1.0$  and fixed  $f_{\text{gas}} = 8\%$ , at  $z > 1.0$ . The predicted fluxes are derived using the average DL07 SED presented in Fig. 6(right) and are corrected for CMB effects. The blue shaded region depicts the range of predicted flux densities in ALMA Band 3 assuming a 50% uncertainty in the adopted  $f_{\text{gas}}$ . *Middle:* ratio of  $M_{\text{gas}}$  estimates, as inferred for a fixed monochromatic luminosity at a given rest-frame wavelength from (1) scaling the average SED of QGs presented in Fig. 6(right),  $M_{\text{gas}}^{\text{SED}}$ , and (2) the prescription of Scoville et al. (2017),  $M_{\text{gas}}^{T_d=25\text{K}}$  (mass-weighted). Similarly, the right y-axis depicts the ratio of the monochromatic luminosities predicted by the two methods for a fixed  $M_{\text{gas}}$ . *Right:* the effect of the adopted  $M_{\text{gas}}$ -method on the recovered evolution of  $f_{\text{gas}}$ . Filled blue circles correspond to the values presented in Fig. 4 using the full SED analysis ( $f_{\text{gas}}^{\text{SED}}$ ). Filled grey squares correspond to the values derived from the Scoville et al. (2017) prescription, assuming monochromatic observations at  $\lambda_{\text{obs}} = 850\mu\text{m}$ , ( $f_{\text{gas}}^{850}$ ). The empty points (and the colour-shaded region) are an extrapolation to higher redshifts under the assumption of fixed  $f_{\text{gas}}^{\text{SED}} = 8\%$  at  $z > 2.0$ .

with  $f_{\text{gas}} = 8\%$  at  $z \geq 1.0$  (Sect. 5). The flux density tracks as a function of redshift for various ALMA bands are presented in Fig. 7(left). We stress that these tracks are specifically tailored around massive ( $\log(M_*/M_\odot) = 11.20$ ) QGs with a solar metallicity. However, the recipe described above is also valid for any arbitrary gas phase metallicity and any  $M_*$  provided that our template is also representative of less (or more) massive QGs.

We emphasise that the conversion of R-J flux densities ( $f_{\text{ALMA}}$ ) to  $M_{\text{gas}}$ , and vice versa, is heavily dependent on the adopted method (and template). Indeed, at fixed  $f_{\text{ALMA}}$ , the commonly adopted monochromatic prescription of Scoville et al. (2017) yields systematically lower  $M_{\text{gas}}$  estimates with respect to those inferred in here. As shown in Fig. 7(middle) the tension between the two methods becomes more prominent at  $\lambda_{\text{rest}}$  closer to the peak of the SED and progressively eases moving further into the R-J tail. This is also demonstrated in Fig. 7(right) where we show how applying the Scoville et al. (2017) method at  $\lambda_{\text{obs}} = 850\mu\text{m}$  would affect the recovered evolution of  $f_{\text{gas}}$ . We note that while we lack observations at  $z > 2.0$  we extrapolate to higher redshifts under the assumption of flat  $f_{\text{gas}}$  evolution at  $z > 2$  to highlight how discrepancy (for fixed  $\lambda_{\text{obs}}$ ) between the two methods increases at higher redshifts.

The tension in the  $M_{\text{gas}}$  (and  $f_{\text{gas}}$ ) estimates between the two methods, arises predominantly from the assumed  $T_d$  of QGs. We recall that the Scoville et al. (2017) recipe is calibrated on SFGs, assuming a fixed ‘mass-weighted’  $T_d = 25\text{K}$ . This corresponds to a  $\sim 7\text{--}10\text{K}$  warmer ‘luminosity-weighted’  $T_d$  compared to that of our template. Similarly, as discussed above, our QG template is  $\sim 10\text{K}$  colder with a  $\sim 10$  times lower  $L_{\text{IR}}/M_{\text{dust}}$  compared to that of a typical SFG at  $z = 1.0$ . It then naturally follows that with respect to the Scoville et al. (2017) approach (or a  $z > 0$  SFG template) our method predicts lower  $f_{\text{ALMA}}$  and subsequently a significantly longer integration time to detect dust emission from a QG of a given  $M_{\text{gas}}$  (e.g.,  $\sim \times 6$  for a galaxy at  $z = 3$  and  $\lambda_{\text{obs}} = 1200\mu\text{m}$ ). This, could explain the scarce dust continuum detections among the limited number of high- $z$  QGs that have been followed up with ALMA so far.

Another point of concern is selection biases, especially in terms of  $M_*$ . Williams et al. (2021), reports no CO[2-1] detection for a sample of massive ( $\log(M_*/M_\odot) \approx 11.7$  (for Salpeter IMF) QGs at  $z \sim 1.5$ , placing stringent upper limits in their  $f_{\text{gas}} \leq 1.0\text{--}3.5\%$ , that are lower compared to our estimates. However, we note that this sample is  $\sim 0.5\text{--}0.7$  dex more massive with respect to the stacked ensembles studied here. If the  $f_{\text{gas}}$  of high- $z$  QGs decreases as a function of  $M_*$ , similarly to what is observed for SFGs and local ETGs ( $f_{\text{gas}} \propto M_*^{-(0.40\text{--}0.55)}$  e.g., Saintonge et al. 2011a; Magdis et al. 2012), then rescaling our  $f_{\text{gas}}$  estimates accordingly would significantly ease the tension (see also Fig. 6 of Williams et al. 2021). Nevertheless, we caution the reader that (1) the dust method to infer  $M_{\text{gas}}$ , when applied to QGs, could in principle recover a sizeable amount of HI that is untraceable by the CO lines and (2) DL07 or similar dust emission models could be a poor representation of the dust grain composition of distant QGs.

Finally, we discuss the impact of the CMB on the predicted flux densities, an effect that was recently observationally confirmed in a sample of ‘cold’  $z > 3$  SFGs (Jin et al. 2019; Cortzen et al. 2020). Our analysis suggests that while  $f_{\text{ALMA}}$  should remain unaffected up to  $z \sim 2.0$  in all ALMA bands, the ‘dimming’ of the dust continuum emission that can be measured against the CMB is not negligible at higher redshifts and lower frequencies. This is demonstrated by the flat evolution of the flux density tracks in Bands 3–6, that clearly deviate from the expected rising trend due to the negative  $k$ -correction (Fig. 7left). Indeed, our analysis suggests that the recoverable continuum emission drops by a factor of 1.4 at  $z = 3\text{--}5$  for Band 5 and by a factor of  $\sim 1.6\text{--}2.0$  at  $z = 4\text{--}5$  for Band 3 observations. Admittedly, the analysis presented in this work does not extend beyond  $z \sim 2$ , and therefore the actual IR SED (and  $T_d$ ) of QGs at  $z > 2$  is still unconstrained. However, if the template SED presented here is also valid at  $z > 2.0$ , CMB-corrections should be implemented both in the observing strategy and in the interpretation of dust continuum and spectral line observations of  $z > 3.0$  QGs in the (sub)-mm.

## 6.2. Caveats

As pointed out in the previous sections, this analysis is not free of caveats which are inherent to the stacking method. First, decomposing the stacked signal into various components is a challenging task. Indeed, ‘satellite’ fluxes in the SPIRE bands, and their corresponding uncertainties, depend on the adopted templates used for extrapolation. While we have considered an uncertainty on the SED of the SF component, our analysis does not account for spatially correlated variations of  $T_{\text{dust}}$ . This would effectively make the satellite emission term frequency-dependent. On the other hand, the average stellar mass of SFGs in the satellite halo varies, with respect to their distance to the QGs, by less than a factor of 2. This suggests that the population of SF satellites is fairly homogeneous.

Furthermore, the combined photometry in the available bands can place constraints in the shape of the SED and in the inferred FIR properties, under the assumption that the FIR emission of QGs is well-represented by the adopted DL07 models. This is of particular importance, especially since none of our stacked ensembles has a formal detection ( $S/N > 3$ ) in the R-J tail ( $\lambda_{\text{rest}} > 350 \mu\text{m}$ ). With these important considerations in mind we conclude that future infrared and millimetre observations of large and representative samples at  $\leq 1-2''$  resolution, are necessary to capture unequivocally the ISM mass budget and ISM conditions of high- $z$  QGs.

## 7. Conclusions

In this work we presented a robust multi-wavelength stacking analysis of a carefully selected sample of massive QGs from the COSMOS field and in three redshift bins,  $\langle z \rangle = 0.5, 0.9, 1.2$ . We further complemented our sample with the stacking ensemble of Gobat et al. (2018), at  $\langle z \rangle = 1.8$ , constructed following the same selection criteria as those adopted in the current study, as well as with literature data for local QGs. By modelling the stacked photometry, we drew the characteristic FIR SEDs, inferred the ISM mass budget, and explored the evolution of the FIR properties of massive QGs over the last ten billion years. Our results are summarised as follows:

- The  $M_{\text{dust}}/M_*$  ratio rises steeply as a function of redshift up to  $z \sim 1.0$  and then remains flat at least out to  $z = 2.0$ . The evolution of  $f_{\text{gas}}$  follows (by construction) a similar trend, with a normalisation that depends on the assumed gas phase metallicity of the QGs. For solar metallicity,  $f_{\text{gas}}$  increases from 2% to 8% between  $z = 0.5$  and  $z = 1.0$ . The evolution of  $f_{\text{gas}}$  in our QG samples can then be interpreted as a combination of progenitor bias at  $z > 1$ , a uniformity in ISM conditions among newly quenched galaxies, and a closed-box consumption of the ISM after quenching, with a depletion time of  $\sim 2$  Gyr.
- The gas depletion time scales of massive QGs at all redshifts are comparable to that of local SFGs and systematically longer than that of MS galaxies at their corresponding redshifts.
- The dust temperature of massive QGs remains roughly constant with  $T_{\text{d}} = 21 \pm 2$  K out to  $z = 2.0$ , with only a weak (if any) evolution towards a marginally higher  $T_{\text{d}}$  at higher redshifts, which can be explained as the consequence of the average size evolution of QGs. This motivated us to construct and make publicly available a template IR SED of massive  $z > 0$  QGs.
- Based on our template SED, we provide predictions for the flux densities of the continuum emission in the R-J tail of

passive galaxies as a function of redshift, and highlight the need of accounting for CMB effects at  $z > 3$ . Finally, we argue that  $M_{\text{gas}}$  prescriptions calibrated on SFGs could lead to an overestimate of the predicted dust continuum flux density of QGs in the (sub)-mm bands and discuss the implications for ALMA observing strategies.

After accounting for progenitor bias, we are ultimately left with the picture of a weak to flat evolution of initial ISM conditions in QGs. That is, our current constraints do not require that the gas fraction of QGs immediately after quenching, nor the depletion time after that, change significantly in the 10 Gyr between  $z = 0$  and  $z \sim 2$ . This apparent regularity in the internal properties of QGs across most of the history of the Universe mirrors the consistency of star formation physics in MS galaxies throughout cosmic time. We conclude that this uniformity of QGs should extend to earlier  $z > 2.0$  epochs, which so far lack strong constraints in this regard.

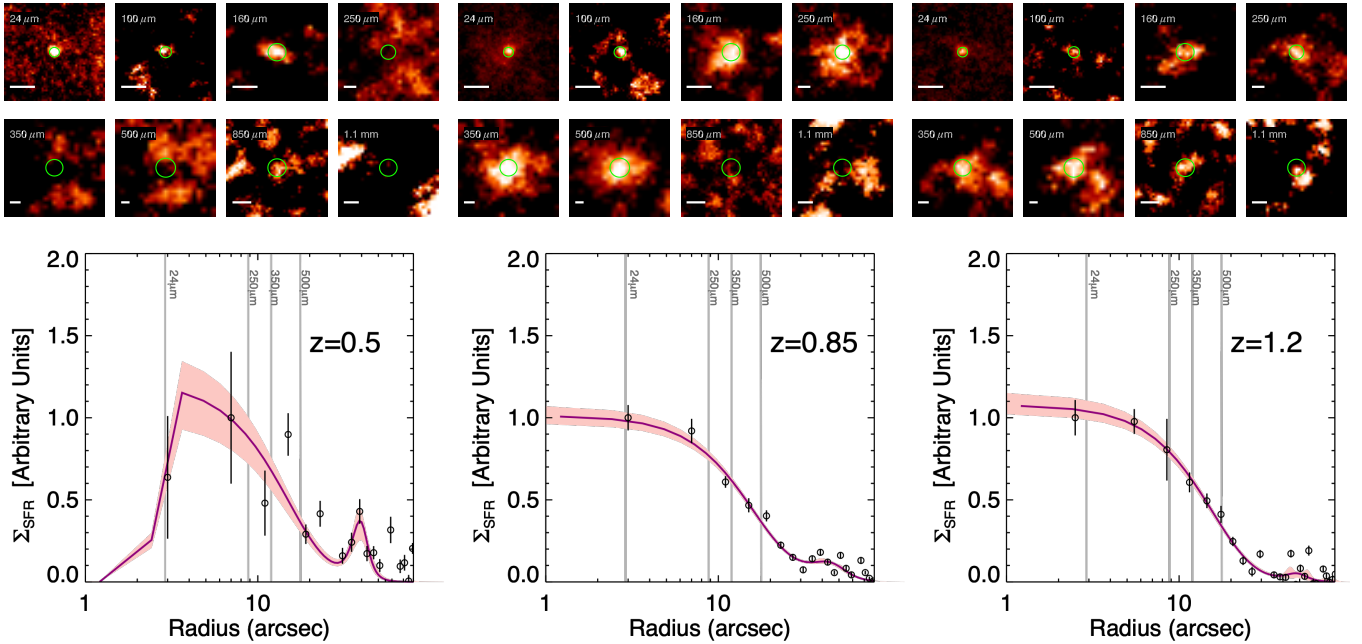
*Acknowledgements.* GEM and FV acknowledge the Villum Fonden research Grant 13160 “Gas to stars, stars to dust: tracing star formation across cosmic time” and the Cosmic Dawn Center of Excellence funded by the Danish National Research Foundation under then Grant No. 140. FV acknowledges support from the Carlsberg Foundation research grant CF18-0388 “Galaxies: Rise And Death”. S. J. acknowledges financial support from the Spanish Ministry of Science, Innovation and Universities (MICIU) under AYA2017-84061-P, co-financed by FEDER (European Regional Development Funds). K.E.W. wishes to acknowledge funding from the Alfred P. Sloan Foundation.

## References

- Agius, N. K., Sansom, A. E., Popescu, C. C., et al. 2013, *MNRAS*, **431**, 1929  
 Alatalo, K., Lisenfeld, U., Lanz, L., et al. 2016, *ApJ*, **827**, 106  
 Aretxaga, I., Wilson, G. W., Aguilar, E., et al. 2011, *MNRAS*, **415**, 3831  
 Arnouts, S., Walcher, C. J., Le Fèvre, O., et al. 2007, *A&A*, **476**, 137  
 Barišić, I., van der Wel, A., Bezanson, R., et al. 2017, *ApJ*, **847**, 72  
 Béthermin, M., Le Floch, E., Ilbert, O., et al. 2012, *A&A*, **542**, A58  
 Béthermin, M., Daddi, E., Magdis, G., et al. 2015, *A&A*, **573**, A113  
 Béthermin, M., Fudamoto, Y., Ginolfi, M., et al. 2020, *A&A*, **643**, A2  
 Bezanson, R., Spilker, J., Williams, C. C., et al. 2019, *ApJ*, **873**, L19  
 Bielby, R., Hudelot, P., McCracken, H. J., et al. 2012, *A&A*, **545**, A23  
 Bigiel, F., Leroy, A., Walter, F., et al. 2008, *AJ*, **136**, 2846  
 Blitz, L., & Rosolowsky, E. 2006, *ApJ*, **650**, 933  
 Boselli, A., Cortese, L., Boquien, M., et al. 2014, *A&A*, **564**, A66  
 Bruzual, G., & Charlot, S. 2003, *MNRAS*, **344**, 1000  
 Cappellari, M., McDermid, R. M., Alatalo, K., et al. 2013, *MNRAS*, **432**, 1862  
 Cattaneo, A., Dekel, A., Devriendt, J., Guiderdoni, B., & Blaizot, J. 2006, *MNRAS*, **370**, 1651  
 Chaniai, P., Flores, H., Guiderdoni, B., et al. 2007, *A&A*, **462**, 81  
 Cimatti, A., Daddi, E., Renzini, A., et al. 2004, *Nature*, **430**, 184  
 Civano, F., Marchesi, S., Comastri, A., et al. 2016, *ApJ*, **819**, 62  
 Cortzen, I., Magdis, G. E., Valentino, F., et al. 2020, *A&A*, **634**, L14  
 Croton, D. J., Springel, V., White, S. D. M., et al. 2006, *MNRAS*, **365**, 11  
 da Cunha, E., Groves, B., Walter, F., et al. 2013, *ApJ*, **766**, 13  
 Daddi, E., Cimatti, A., Pozzetti, L., et al. 2000, *A&A*, **361**, 535  
 Daddi, E., Cimatti, A., Renzini, A., et al. 2004, *ApJ*, **617**, 746  
 Daddi, E., Renzini, A., Pirzkal, N., et al. 2005, *ApJ*, **626**, 680  
 Daddi, E., Dickinson, M., Morrison, G., et al. 2007, *ApJ*, **670**, 156  
 Daddi, E., Dannerbauer, H., Elbaz, D., et al. 2008, *ApJ*, **673**, L21  
 Daddi, E., Bournaud, F., Walter, F., et al. 2010, *ApJ*, **713**, 686  
 Davé, R., Rafieferantsoa, M. H., Thompson, R. J., & Hopkins, P. F. 2017, *MNRAS*, **467**, 115  
 Davé, R., Crain, R. A., Stevens, A. R. H., et al. 2020, *MNRAS*, **497**, 146  
 Davidzon, I., Ilbert, O., Laigle, C., et al. 2017, *A&A*, **605**, A70  
 Davis, T. A., Young, L. M., Crocker, A. F., et al. 2014, *MNRAS*, **444**, 3427  
 Dekel, A., Birnboim, Y., Engel, G., et al. 2009, *Nature*, **457**, 451  
 Delhaize, J., Smolčić, V., Delvecchio, I., et al. 2017, *A&A*, **602**, A4  
 D’Eugenio, C., Daddi, E., Gobat, R., et al. 2020, *ApJ*, **892**, L2  
 DiMatteo, T., Springel, V., & Hernquist, L. 2005, *Nature*, **433**, 604  
 Draine, B. T., & Li, A. 2007, *ApJ*, **657**, 810  
 Elbaz, D., Dickinson, M., Hwang, H. S., et al. 2011, *A&A*, **533**, A119  
 Fasano, G., & Franceschini, A. 1987, *MNRAS*, **225**, 155  
 French, K. D., Yang, Y., Zabludoff, A., et al. 2015, *ApJ*, **801**, 1

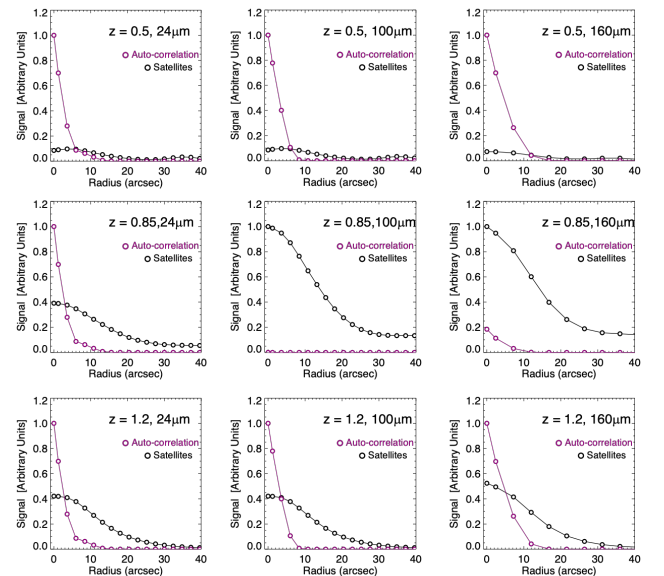
- Geach, J. E., Smail, I., Moran, S. M., et al. 2011, *ApJ*, 730, L19
- Geach, J. E., Dunlop, J. S., Halpern, M., et al. 2017, *MNRAS*, 465, 1789
- Glazebrook, K., Schreiber, C., Labbé, I., et al. 2017, *Nature*, 544, 71
- Gobat, R., Strazzullo, V., Daddi, E., et al. 2012, *ApJ*, 759, L44
- Gobat, R., Daddi, E., Magdis, G., et al. 2018, *Nat. Astron.*, 2, 239
- Gobat, R., Magdis, G., D'Eugenio, C., & Valentino, F. 2020, *A&A*, 644, L7
- Gomez, H. L., Baes, M., Cortese, L., et al. 2010, *A&A*, 518, L45
- Hayashi, M., Tadaki, K.-I., Kodama, T., et al. 2018, *ApJ*, 856, 118
- Hayward, C. C., Lanz, L., Ashby, M. L. N., et al. 2014, *MNRAS*, 445, 1598
- Henriques, B. M. B., White, S. D. M., Thomas, P. A., et al. 2015, *MNRAS*, 451, 2663
- Hopkins, P. F., Somerville, R. S., Hernquist, L., et al. 2006, *ApJ*, 652, 864
- Hwang, H. S., Elbaz, D., Magdis, G., et al. 2010, *MNRAS*, 409, 75
- Ilbert, O., McCracken, H. J., Le Fèvre, O., et al. 2013, *A&A*, 556, A55
- Jin, S., Daddi, E., Liu, D., et al. 2018, *ApJ*, 864, 56
- Jin, S., Daddi, E., Magdis, G. E., et al. 2019, *ApJ*, 887, 144
- Kennicutt, R. C. Jr 1998, *ARA&A*, 36, 189
- Kokorev V., Magdis G. E., Davidzon I., et al. 2021, *ApJ*, submitted
- Kriek, M., van Dokkum, P. G., Franx, M., et al. 2006, *ApJ*, 649, L71
- Lagos, C. d. P., Crain, R. A., Schaye, J., et al. 2015, *MNRAS*, 452, 3815
- Laigle, C., McCracken, H. J., Ilbert, O., et al. 2016, *ApJS*, 224, 24
- Le Floe'h, E., Aussel, H., Ilbert, O., et al. 2009, *ApJ*, 703, 222
- Lejeune, T., Cuisinier, F., & Buser, R. 1997, *A&AS*, 125
- Leroy, A. K., Bolatto, A., Gordon, K., et al. 2011, *ApJ*, 737, 12
- Lianou, S., Xilouris, E., Madden, S. C., & Barmby, P. 2016, *MNRAS*, 461, 2856
- Liu, D., Schinnerer, E., Groves, B., et al. 2019, *ApJ*, 887, 235
- Lutz, D., Poglitsch, A., Altieri, B., et al. 2011, *A&A*, 532, A90
- Magdis, G. E., Rigopoulou, D., Huang, J. S., & Fazio, G. G. 2010, *MNRAS*, 401, 1521
- Magdis, G. E., Daddi, E., Béthermin, M., et al. 2012, *ApJ*, 760, 6
- Magdis, G. E., Rigopoulou, D., Daddi, E., et al. 2017, *A&A*, 603, A93
- Magnelli, B., Lutz, D., Saintonge, A., et al. 2014, *A&A*, 561, A86
- Man, A., & Belli, S. 2018, *Nat. Astron.*, 2, 695
- Man, A. W. S., Greve, T. R., Toft, S., et al. 2016, *ApJ*, 820, 11
- Mannucci, F., Cresci, G., Maiolino, R., Marconi, A., & Gnerucci, A. 2010, *MNRAS*, 408, 2115
- Marchesi, S., Civano, F., Elvis, M., et al. 2016, *ApJ*, 817, 34
- Martig, M., Bournaud, F., Teyssier, R., & Dekel, A. 2009, *ApJ*, 707, 250
- McCracken, H. J., Capak, P., Salvato, M., et al. 2010, *ApJ*, 708, 202
- McCracken, H. J., Milvang-Jensen, B., Dunlop, J., et al. 2012, *A&A*, 544, A156
- Michałowski, M. J., Hjorth, J., Gall, C., et al. 2019, *A&A*, 632, A43
- Muzzin, A., Wilson, G., Demarco, R., et al. 2013a, *ApJ*, 767, 39
- Muzzin, A., Marchesini, D., Stefanon, M., et al. 2013b, *ApJ*, 777, 18
- Narayanan, D., Turk, M., Feldmann, R., et al. 2015, *Nature*, 525, 496
- Noeske, K. G., Weiner, B. J., Faber, S. M., et al. 2007, *ApJ*, 660, L43
- Noll, S., Pierini, D., Cimatti, A., et al. 2009, *A&A*, 499, 69
- Novak, M., Smolčić, V., Schinnerer, E., et al. 2018, *A&A*, 614, A47
- Obreschkow, D., Croton, D., De Lucia, G., Khochfar, S., & Rawlings, S. 2009, *ApJ*, 698, 1467
- Oliver, S. J., Bock, J., Altieri, B., et al. 2012, *MNRAS*, 424, 1614
- Peng, Y., Maiolino, R., & Cochrane, R. 2015, *Nature*, 521, 192
- Pettini, M., & Pagel, B. E. J. 2004, *MNRAS*, 348, L59
- Polletta, M., Tajer, M., Maraschi, L., et al. 2007, *ApJ*, 663, 81
- Popping, G., Somerville, R. S., & Trager, S. C. 2014, *MNRAS*, 442, 2398
- Popping, G., Somerville, R. S., & Galametz, M. 2017, *MNRAS*, 471, 3152
- Rowlands, K., Dunne, L., Maddox, S., et al. 2012, *MNRAS*, 419, 2545
- Rudnick, G., Hodge, J., Walter, F., et al. 2017, *ApJ*, 849, 27
- Saintonge, A., Kauffmann, G., Kramer, C., et al. 2011a, *MNRAS*, 415, 32
- Saintonge, A., Kauffmann, G., Wang, J., et al. 2011b, *MNRAS*, 415, 61
- Salpeter, E. E. 1955, *ApJ*, 121, 161
- Sargent, M. T., Daddi, E., Béthermin, M., et al. 2014, *ApJ*, 793, 19
- Sargent, M. T., Daddi, E., Bournaud, F., et al. 2015, *ApJ*, 806, L20
- Schinnerer, E., Sargent, M. T., Bondi, M., et al. 2010, *ApJS*, 188, 384
- Schreiber, C., Pannella, M., Elbaz, D., et al. 2015, *A&A*, 575, A74
- Schreiber, C., Glazebrook, K., Nanayakkara, T., et al. 2018a, *A&A*, 618, A85
- Schreiber, C., Elbaz, D., Pannella, M., et al. 2018b, *A&A*, 609, A30
- Scoville, N., Aussel, H., Brusa, M., et al. 2007, *ApJS*, 172, 1
- Scoville, N., Lee, N., Vanden Bout, P., et al. 2017, *ApJ*, 837, 150
- Smith, M. W. L., Gomez, H. L., Eales, S. A., et al. 2012, *ApJ*, 748, 123
- Smolčić, V., Novak, M., Bondi, M., et al. 2017, *A&A*, 602, A1
- Speagle, J. S., Steinhardt, C. L., Capak, P. L., & Silverman, J. D. 2014, *ApJS*, 214, 15
- Spilker, J., Bezanson, R., Barišić, I., et al. 2018, *ApJ*, 860, 103
- Stockmann, M., Toft, S., Gallazzi, A., et al. 2020, *ApJ*, 888, 4
- Suess, K. A., Bezanson, R., Spilker, J. S., et al. 2017, *ApJ*, 846, L14
- Tacconi, L. J., Genzel, R., Saintonge, A., et al. 2018, *ApJ*, 853, 179
- Tanaka, M., Valentino, F., Toft, S., et al. 2019, *ApJ*, 885, L34
- Toft, S., van Dokkum, P., Franx, M., et al. 2005, *ApJ*, 624, L9
- Toft, S., van Dokkum, P., Franx, M., et al. 2007, *ApJ*, 671, 285
- Toft, S., Gallazzi, A., Zirm, A., et al. 2012, *ApJ*, 754, 3
- Toft, S., Zabl, J., Richard, J., et al. 2017, *Nature*, 546, 510
- Umetsu, K., Medezinski, E., Nonino, M., et al. 2014, *ApJ*, 795, 163
- Valentino, F., Tanaka, M., Davidzon, I., et al. 2020, *ApJ*, 889, 93
- van der Wel, A., Franx, M., van Dokkum, P. G., et al. 2014, *ApJ*, 788, 28
- Viero, M. P., Moncelsi, L., Quadri, R. F., et al. 2013, *ApJ*, 779, 32
- Werner, N., Oonk, J. B. R., Sun, M., et al. 2014, *MNRAS*, 439, 2291
- Whitaker, K. E., Kriek, M., van Dokkum, P. G., et al. 2012, *ApJ*, 745, 179
- Whitaker, K. E., van Dokkum, P. G., Brammer, G., et al. 2013, *ApJ*, 770, L39
- Williams, R. J., Quadri, R. F., Franx, M., van Dokkum, P., & Labbé, I. 2009, *ApJ*, 691, 1879
- Williams, C. C., Spilker, J. S., Whitaker, K. E., et al. 2021, *ApJ*, 908, 54
- Young, L. M., Bureau, M., Davis, T. A., et al. 2011, *MNRAS*, 414, 940

## Appendix A: Cutouts and satellite density



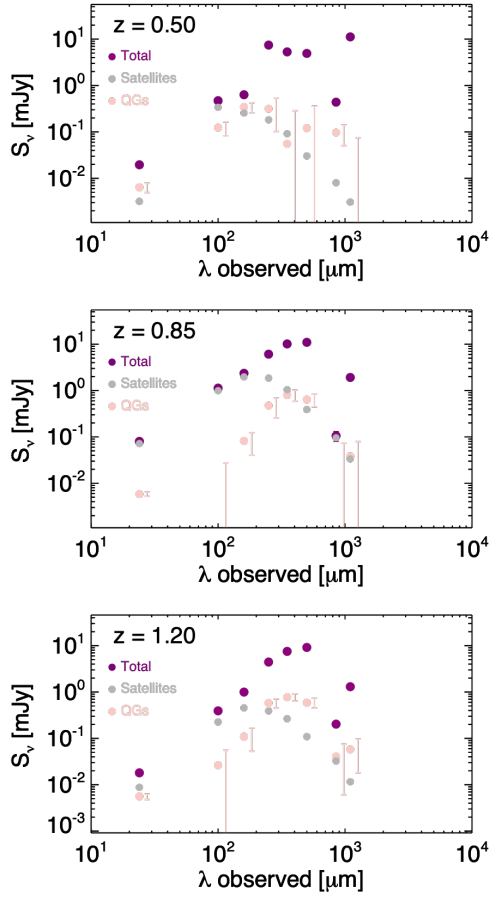
**Fig. A.1.** Stacked cutouts and SFR surface density profiles. *Top:* median stacked cutouts at the position of the QGs in the three redshift bins, at  $24\ \mu\text{m}$  (*Spitzer/MIPS*),  $100$  and  $160\ \mu\text{m}$  (*Herschel/PACS*),  $250$ ,  $350$  and  $500\ \mu\text{m}$  *Herschel/SPIRE*,  $850\ \mu\text{m}$  (*JCMT/SCUBA-2*) and  $1.1\ \text{mm}$  (*ASTE/AzTEC*). The white line in each panel has a length of  $15''$ , while the green circles show the beam FWHM in each band. *Bottom:* Normalised, average SFR surface density ( $\Sigma_{\text{SFR}}$ ) profile of the satellites of QGs (circles) in various redshift bins. The error bars correspond to the standard deviation of the surface density in each bin. The pink lines and the shaded regions show the best-fit ‘quenched’  $\beta$ -models to  $\Sigma_{\text{SFR}}$  and their uncertainties. The vertical grey lines denote the half-width at half-maximum of the MIPS and SPIRE beams.

Galaxies associated with the QGs in our samples (hereafter ‘satellites’) were selected from the [Laigle et al. \(2016\)](#) multi-wavelength catalogue as star forming based on the *UVJ* criterion and with photometric redshifts compatible within their uncertainties with that of each QG. The rest-frame *UV* SEDs of satellites were then fitted with 100 Myr-old composite stellar population models of constant star formation, with and without dust attenuation ([Noll et al. 2009](#)), to derive both extinction-corrected and uncorrected SFRs. We then took the difference between those two estimates as the obscured SFR of satellites, and these values were averaged in radial bins to construct for each sample (Fig. A.1, bottom panels). The obscured SFR density distributions thus constructed were fitted with a nine-parameter combination of two  $\beta$ -profiles of the form  $A \left[ 1 + \left( \frac{r - r_0}{r_c} \right)^2 \right]^{-\beta}$ , the first fixed at  $r_0 = 0$ , and a multiplicative quenching term of the form  $\min(r/r_q, 1)^{\alpha_q}$ . The latter is only appreciable in the case of the low-redshift bin, as shown in Fig. A.1. The two  $\beta$ -models describe the one- and two-halo terms of satellites, and the latter increases noticeably in amplitude with decreasing redshift. This is not surprising as galaxies with a fixed stellar mass will be less dominant within their halo at low than at high redshift. Finally, the functional profiles were circularised and normalised to  $1 L_{\odot}$  when fitting the stacked cutouts. The 2D QG (point source + autocorrelation) and circularised satellite signals are left to vary freely at  $24$ ,  $100$  and  $160\ \mu\text{m}$ , where the instrumental beam is small enough that their different shapes can be resolved (see Fig. A.2). At longer wavelengths the satellite signal is extrapolated from the  $24$ ,  $100$  and  $160\ \mu\text{m}$  amplitudes using [Magdis et al. \(2012\)](#) templates, as described in Sect. 3. After subtracting the central source, autocorrelation signal, and satellite profile, we find that the median residuals are  $\lesssim 6\%$ .



**Fig. A.2.** 1D profiles of the auto-correlation and satellite signal, showing their respective amplitudes (in arbitrary units) at  $24$ ,  $100$  and  $160\ \mu\text{m}$ . The autocorrelation signal includes the central point source since one fitted amplitude is used for both (see Sect. 3).

This method makes the implicit assumptions that the SED-derived obscured SFR and FIR emission of satellites have the same spatial distributions, and that satellites below the detection threshold of the multi-wavelength [Laigle et al. \(2016\)](#) catalogue are similarly distributed as their brighter counterparts. On the other hand, we note that the median mass of satellites



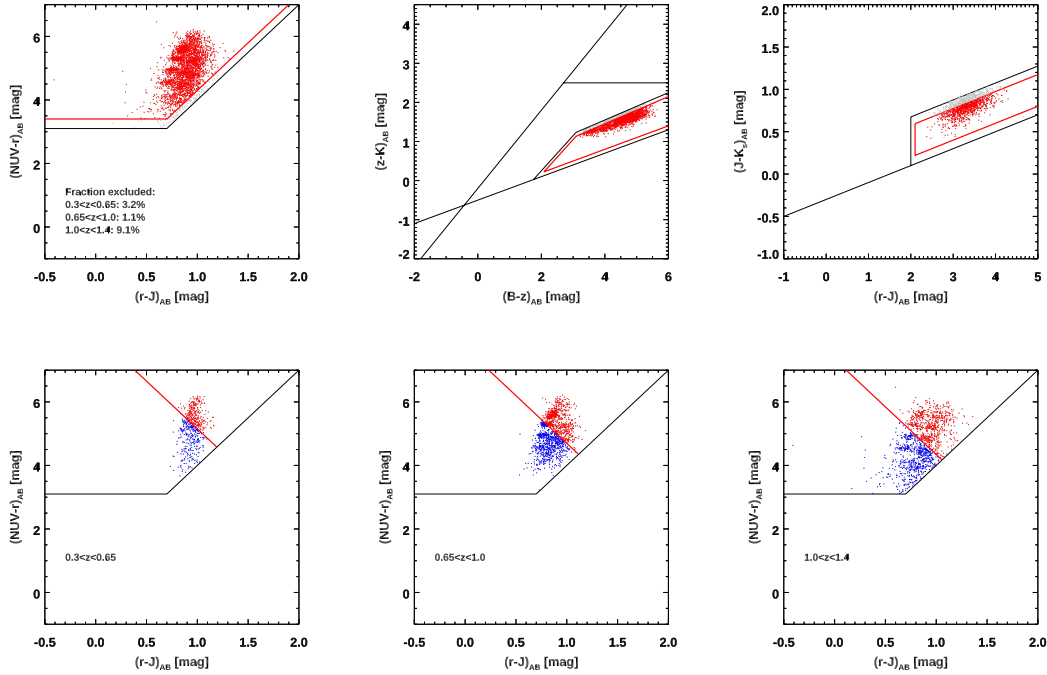
**Fig. A.3.** Background-subtracted and bias-corrected FIR SEDs, before and after decomposition in SF and QG components: total signal (purple), SF satellites (grey), and QG with error bars on the side (light pink). Only the point-source component of the QG SED is shown (that is, after subtraction of the autocorrelation term). As described in Sect. 3, the satellite points are fitted to the data at  $\leq 160 \mu\text{m}$  and extrapolated from Magdis et al. (2012) templates at longer wavelengths.

increases somewhat at radii  $r < 10''$ , which suggests that sub-threshold (i.e. lower mass) satellites contribute less signal at small radii. This might result in a slight underestimate of the FIR flux of the central QG. Fig. A.3 shows, for the three redshift bins, the total measured FIR signal (after background subtraction, and bias correction as in Béthermin et al. 2015) and its decomposition in SF satellite and QG components (the latter after subtraction of the autocorrelation term). The values of each component in each band along with the corresponding uncertainties are listed in Table A.1.

**Table A.1.** Stack decomposition.

Redshift	24 $\mu\text{m}$ [mJy] (2)	100 $\mu\text{m}$ [mJy] (3)	160 $\mu\text{m}$ [mJy] (4)	250 $\mu\text{m}$ [mJy] (5)	350 $\mu\text{m}$ [mJy] (6)	500 $\mu\text{m}$ [mJy] (7)	850 $\mu\text{m}$ [mJy] (8)	1100 $\mu\text{m}$ [mJy] (9)
(0.50)	19.57 $\pm$ 0.10	0.47 $\pm$ 0.02	0.63 $\pm$ 0.07	7.43 $\pm$ 0.21	5.30 $\pm$ 0.22	4.91 $\pm$ 0.23	0.43 $\pm$ 0.04	11.16 $\pm$ 0.07
(0.85)	80.07 $\pm$ 0.45	1.12 $\pm$ 0.02	2.35 $\pm$ 0.04	6.08 $\pm$ 0.12	10.15 $\pm$ 0.15	10.97 $\pm$ 0.18	0.10 $\pm$ 0.07	1.91 $\pm$ 0.04
(1.20)	18.06 $\pm$ 0.59	0.39 $\pm$ 0.02	0.99 $\pm$ 0.04	4.43 $\pm$ 0.12	7.52 $\pm$ 0.13	9.15 $\pm$ 0.14	0.20 $\pm$ 0.03	1.30 $\pm$ 0.03
(0.50)	3.18 $\pm$ 1.00	0.33 $\pm$ 0.03	0.25 $\pm$ 0.02	0.18 $\pm$ 0.02	0.09 $\pm$ 0.01	0.03 $\pm$ 0.01	0.008 $\pm$ 0.001	0.003 $\pm$ 0.001
(0.85)	72.22 $\pm$ 0.42	0.99 $\pm$ 0.02	1.98 $\pm$ 0.01	1.85 $\pm$ 0.18	1.04 $\pm$ 0.15	0.39 $\pm$ 0.06	0.09 $\pm$ 0.02	0.03 $\pm$ 0.01
(1.20)	8.78 $\pm$ 0.53	0.22 $\pm$ 0.02	0.45 $\pm$ 0.03	0.39 $\pm$ 0.04	0.02 $\pm$ 0.04	0.10 $\pm$ 0.02	0.032 $\pm$ 0.01	0.011 $\pm$ 0.002
Background								
(0.50)	6.40e-03 $\pm$ 5.4e-05	1.23e-01 $\pm$ 1.4e-05	3.24e-01 $\pm$ 4.5e-05	2.86e-01 $\pm$ 4.5e-02	3.22e-02 $\pm$ 3.4e-05	1.11e-01 $\pm$ 4.1e-05	1.06e-01 $\pm$ 1.3e-05	2.66e-7 $\pm$ 7.0e-05
(0.85)	5.820e-03 $\pm$ 3.5e-05	0.00 $\pm$ 9.6e-06	1.95e-01 $\pm$ 2.9e-05	5.10e-01 $\pm$ 2.2e-02	1.01e+00 $\pm$ 2.8e-05	6.51e-01 $\pm$ 1.2e-02	2.11e-8 $\pm$ 7.6e-06	3.15e-02 $\pm$ 4.2e-05
(1.20)	5.38e-04 $\pm$ 2.9e-05	4.29e-03 $\pm$ 8.7e-06	2.57e-02 $\pm$ 3.3e-05	9.09e-01 $\pm$ 2.3e-02	4.89e-03 $\pm$ 2.7e-05	-2.18e-01 $\pm$ 1.2e-02	4.18e-03 $\pm$ 8.3e-06	-1.60e-02 $\pm$ 4.6e-05

## Appendix B: Robustness of the sample selection



**Fig. B.1.** Sample selection tests. *Top:*  $NUVrJ$ ,  $BzK$ , and  $rJK$  colour diagrams with a conservative 0.1 mag margin (red lines) added to prevent contamination from reddened star forming objects. *Bottom:* low-, mid-, and high- $z$  samples divided into blue and red sub-samples in  $NUVrJ$  colour space.

To investigate the dependency of our results on the selection criteria used to build the three samples, and in particular to conclusively exclude a non-quiescent origin for the FIR signal found, we performed the analysis on sub-samples drawn from the main one using the following a ‘conservative’ colour cut that adds a 0.1 mag margin to the  $NUVrJ$ ,  $rJK$ , and  $BzK$  quiescent loci used to select QGs (Fig. B.1, top). This was intended to further minimise the possibility of contamination of the sample by dusty SFGs close to the edge of the quiescent region. This yielded 909, 1241, and 415 objects in the high-, mid-, and low- $z$  redshift bins, respectively, corresponding to a reduction by 35%, 19%, and 26%. Finally, supplementary ‘blue’ and ‘red’ cuts, in which the quiescent loci in  $NUVrJ$  space were split perpendicular to the dividing line so as to generate approximately equal-size sub-samples (Fig. B.1, bottom) we applied.

Stacked cutouts were produced and fluxes were extracted following the method described in Sect. 3, while the resulting FIR SEDs were modelled as in Sect. 4. Despite the lowered S/N due to the reduced sample sizes, we found that the SEDs of the sub-samples were consistent, within their uncertainties, with those of the full samples. This is further confirmed by the parameters produced by the fit, which do not deviate appreciably with those listed in Table 1. In particular, those derived from the blue and red sub-samples are entirely consistent with each other. Consequently, we conclude that the latter are not the product of either contamination by non-quiescent objects, or due a particular population of QGs within the main sample.

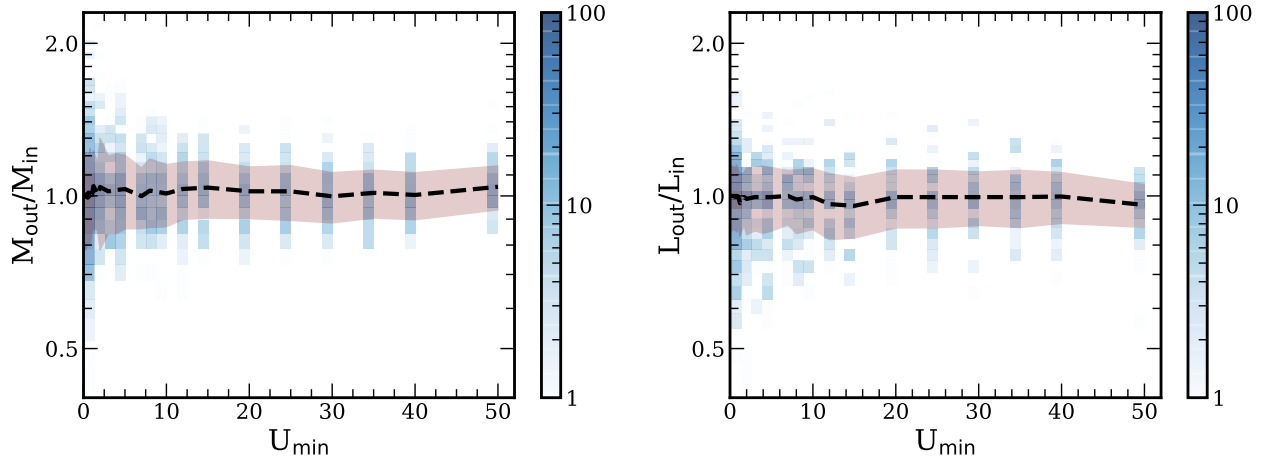
### Appendix C: Estimate of the AGN duty cycle

After subtracting the star-forming radio slope, estimated from the FIR-radio correlation (Delhaize et al. 2017), the average

excess at 1.4 GHz in the high-, mid-, and low- $z$  bins corresponds to intrinsic luminosities of  $2.2 \times 10^{22}$ ,  $7.0 \times 10^{21}$ , and  $5.2 \times 10^{21} \text{ W Hz}^{-1}$ , respectively. We also add to these the  $5.4 \times 10^{22} \text{ W Hz}^{-1}$  1.4 GHz excess from the  $z \sim 1.8$  sample of G18. To translate these excess fluxes extracted from median stacks into an AGN fraction, we considered the evolving 1.4 GHz AGN luminosity functions of Novak et al. (2018), evaluated between  $10^{22}$  and  $10^{28} \text{ W Hz}$ . We first noted that the G18, high-, mid-, and low- $z$  excesses correspond respectively to  $>1$ ,  $0.59 \pm 0.9$ ,  $0.19 \pm 0.04$ , and  $0.14 \pm 0.02$  times the integrated AGN luminosity function. However, this does not account for the sampling of the redshift bins nor the non-zero noise of the original data used to make the stacks (about  $0.015 \mu\text{Jy beam}^{-1}$ ; Schinnerer et al. 2010), which can impact the median flux. We performed a set of 1000 simulations based on random samplings of the Novak et al. (2018) luminosity function, with sample sizes and redshift distributions corresponding to our three redshift bins. For each simulation the duty cycle was then varied from 0 to 100%, with ‘off’ AGNs set to zero luminosity. Finally, we added Gaussian noise to each object based on the r.m.s. of the data, converted to intrinsic luminosities. This yielded duty cycles of  $0.50^{+0.18}_{-0.20}$ ,  $0.32^{+0.07}_{-0.08}$ ,  $0.18^{+0.04}_{-0.05}$ , and  $0.23^{+0.05}_{-0.04}$  for the G18, high- $z$ , mid- $z$ , and low- $z$  quiescent stacks, respectively.

### Appendix D: Simulation of FIR parameters

To investigate the uncertainties and possible systematic effects in the derivation of the  $M_{\text{dust}}$  and  $L_{\text{IR}}$  estimates introduced by the available photometry of our stacked SEDs as a function of the intrinsic shape of the SED, we performed bootstrapping simulations. We first built the full set of DL07 models ranging from  $U_{\text{min}} = 0.1$  to 50 with a fixed  $L_{\text{IR}}$  and  $M_{\text{dust}}$ . We then red-



**Fig. D.1.** Simulations on the derivation of  $M_{\text{dust}}$  and  $L_{\text{IR}}$ . The ratio of derived over input  $M_{\text{dust}}$  (left) and  $L_{\text{IR}}$  (right) as a function of intrinsic  $\langle U_{\text{min}} \rangle$  for a range of simulated DL07 models mimicking the dynamical range and S/N of the available photometry for our stacked SEDs. For each  $U_{\text{min}}$  template the data are colour coded by the density of the points. The pink shaded region encloses the 68% of the derived ratios. Noticeably, the range of the output/input ratios, and thus the corresponding uncertainties are larger for colder templates ( $U_{\text{min}} < 5.0$ ).

shifted each template to the central redshift of each of our redshift bins and measured its synthetic photometry in all the available bands covered by our stacking analysis. To mimic our observations from the emerging set of 8 photometric data points ( $24 \mu\text{m}$ – $1.1 \text{ mm}$ ), we added a random fluctuation in each band of up to 20% and randomly selected and placed at  $S/N = 3$  four data points between  $24$ – $1100 \mu\text{m}$ . In the process, we ensure that the data set included one data point at observed  $24 \mu\text{m}$  and one data point in the R-J tail, but not longer than  $\lambda_{\text{rest}} \approx 220 \mu\text{m}$ , as is the case for all of our observed stacks. The remaining bands were placed at  $S/N = 1$  and treated as upper limits. Following this

procedure, we constructed 100 sets of synthetic photometry for each  $U_{\text{min}}$  and performed SED fitting using the same methodology as that applied to our observations. The comparison between in output and input  $M_{\text{dust}}$  and  $L_{\text{IR}}$  for each template is presented in Fig. D.1. It is evident that the uncertainties become more severe for the coldest templates. This is expected since for the coldest templates with the lowest  $U_{\text{min}}$  the  $\lambda_{\text{rest}}^{\text{max}} \approx 220 \mu\text{m}$  is moving closer to the peak of the SED, leaving the R-J tail less constrained, with implications on the derived  $M_{\text{dust}}$ . Nevertheless, we see that for the combination of the data points considered here, there are no systematic effects in the derivation of  $M_{\text{dust}}$  and  $L_{\text{IR}}$ .