Astronomy & Astrophysics

Near infrared spectroscopy and star-formation histories of $3 \leq z \leq 4$ quiescent galaxies*

C. Schreiber[1,12], K. Glazebrook[2], T. Nanayakkara[1,2], G. G. Kacprzak[2], I. Labbé[1,2], P. Oesch[3], T. Yuan[4], K.-V. Tran[5,6], C. Papovich[7], L. Spitler[8,9,10], and C. Straatman[11]

[1] Leiden Observatory, Leiden University, 2300 RA Leiden, The Netherlands
[2] Centre for Astrophysics and Supercomputing, Swinburne University of Technology, Hawthorn, VIC 3122, Australia
[3] Observatoire de Genève, 1290 Versoix, Switzerland
[4] Research School of Astronomy and Astrophysics, The Australian National University, Cotter Road, Weston Creek, ACT 2611, Australia
[5] Australia Telescope National Facility, CSIRO Astronomy and Space Science, PO Box 76, Epping, NSW 1710, Australia
[6] School of Physics, University of New South Wales, Sydney, NSW 2052, Australia
[7] George P. and Cynthia W. Mitchell Institute for Fundamental Physics and Astronomy, Department of Physics and Astronomy, Texas A&M University, College Station, TX 77843, USA
[8] Research Centre for Astronomy, Astrophysics & Astrophotonics, Macquarie University, Sydney, NSW 2109, Australia
[9] Department of Physics & Astronomy, Macquarie University, Sydney, NSW 2109, Australia
[10] Australian Astronomical Observatory, 105 Delhi Rd., Sydney, NSW 2113, Australia
[11] Max-Planck Institut für Astronomie, Königstuhl 17, 69117 Heidelberg, Germany
[12] Department of Physics, University of Oxford, Keble Road, Oxford OX1 3RH, UK
     e-mail: corentin.schreiber@physics.ox.ac.uk

ABSTRACT

We present Keck–MOSFIRE H and K spectra for a sample of 24 candidate quiescent galaxies at $3 < z < 4$, identified from their rest-frame UVJ colors and photometric redshifts in the ZFOURGE and 3DHST surveys. With median integration times of one hour in H and five in K, we obtain spectroscopic redshifts for half of the sample, using either Balmer absorption lines or nebular emission lines. We confirm the high accuracy of the photometric redshifts for this spectroscopically-confirmed sample, with a median $|z_{\rm phot} - z_{\rm spec}|/(1 + z_{\rm spec})$ of 1.2%. Two galaxies turn out to be dusty Hα emitters at lower redshifts ($z < 2.5$), and these are the only two detected in the sub-mm with ALMA. High equivalent-width [O III] emission is observed in two galaxies, contributing up to 30% of the K-band flux and mimicking the UVJ colors of an old stellar population. This implies a failure rate of only 20% for the UVJ selection at these redshifts. Lastly, Balmer absorption features are identified in four galaxies, among the brightest of the sample, confirming the absence of OB stars. We then modeled the spectra and photometry of all quiescent galaxies with a wide range of star-formation histories. We find specific star-formation rates (sSFR) lower than $0.15\,{\rm Gyr}^{-1}$ (a factor of ten below the main sequence) for all but one galaxy, and lower than $0.01\,{\rm Gyr}^{-1}$ for half of the sample. These values are consistent with the observed Hβ and [O II] luminosities, and the ALMA non-detections. The implied formation histories reveal that these galaxies have quenched on average 300 Myr prior to being observed, between $z = 3.5$ and 5, and that half of their stars were formed by $z \sim 5.5$ with a mean SFR $\sim 300\,M_\odot\,{\rm yr}^{-1}$. We finally compared the UVJ selection to a selection based instead on the sSFR, as measured from the photometry. We find that galaxies a factor of ten below the main sequence are 40% more numerous than UVJ-selected quiescent galaxies, implying that the UVJ selection is pure but incomplete. Current models fail at reproducing our observations, and underestimate either the number density of quiescent galaxies by more than an order of magnitude, or the duration of their quiescence by a factor two. Overall, these results confirm the existence of an unexpected population of quiescent galaxies at $z > 3$, and offer the first insights on their formation histories.

Key words. techniques: spectroscopic – galaxies: evolution – galaxies: high-redshift – galaxies: statistics

1. Introduction

In the present-day Universe, clear links have been observed between the stellar mass of a galaxy, the effective age of its stellar population, its optical colors, its morphology, and its immediate environment. The most massive galaxies, in particular, tend to be located in galaxy over-densities (e.g., clusters or groups), have old stellar populations and little on-going star formation, and display red, featureless spheroidal light profiles

with compact cores (e.g., Baldry et al. 2004). These different observables have been used broadly to identify galaxies belonging to this population, sometimes interchangeably, and be it from their morphology [early-type galaxies (ETGs), spheroids, ellipticals], their colors [red or red sequence galaxies, extremely red objects (EROs), luminous red galaxies (LRGs)], their star formation history [old, quiescent, evolved, passive, or passively evolving galaxies (PEGs)], their mass [massive galaxies], their environment [bright cluster galaxy (BCGs), central galaxy], or any combination thereof.

However, these links tend to dissolve at earlier epochs. While massive galaxies always seem to have red optical colors, at higher redshifts this is increasingly caused by dust

obscuration rather than old stellar populations (e.g., Cimatti et al. 2002; Dunlop et al. 2007; Spitler et al. 2014; Martis et al. 2016). Similarly, the proportion of star-forming objects among massive galaxies, compact galaxies, or within over-dense structures was larger in the past (e.g., Butcher & Oemler 1978; Elbaz et al. 2007, 2018; van Dokkum et al. 2010; Brammer et al. 2011; Barro et al. 2013, 2016; Wang et al. 2016). When exploring the evolution of galaxies through cosmic time, it is therefore crucial not to assume that the aforementioned observables independently map to the same population of objects, and to precisely define which population is under study.

In the present paper, we aim to constrain and understand the emergence of massive galaxies with low levels of on-going star formation, which we will dub hereafter quiescent galaxies (QGs), in opposition to star-forming galaxies (SFGs). In our view, for a galaxy to qualify as quiescent its star formation rate (SFR) needs not be strictly zero, but remain significantly lower than the average for SFGs of similar masses at the same epoch. In other words, these galaxies must reside below the so-called star-forming main sequence (MS; Elbaz et al. 2007; Noeske et al. 2007). If found with SFRs sufficiently lower than that expected for an MS galaxy, say below the MS by an order of magnitude or three times the observed MS scatter (e.g., Schreiber et al. 2015), these galaxies must have experienced a particular event in their history which suppressed star formation (either permanently or temporarily). At any given epoch, this is equivalent to selecting galaxies with a low specific SFR (sSFR $= \mathrm{SFR}/M_*$).

Regardless of how they are defined, the evolution of the number density of QGs has been a long standing debate, and has proven an important tool to constrain galaxy evolution models (see Daddi et al. 2000; Glazebrook et al. 2004, 2017; Cimatti et al. 2004, discussions therein, and below). After two decades of observations, solid evidence now show that QGs already existed in significant numbers in the young Universe at all epochs, now up to $z \sim 4$ (e.g., Franx et al. 2003; Glazebrook et al. 2004, 2017; Cimatti et al. 2004; Kriek et al. 2006, 2009; Gobat et al. 2012; Hill et al. 2016), and that their number density has been rising continuously until the present day (e.g., Faber et al. 2007; Ilbert et al. 2010, 2013; Brammer et al. 2011; Muzzin et al. 2013; Stefanon et al. 2013; Tomczak et al. 2014; Straatman et al. 2014). Spectroscopic observations confirmed their low current SFRs from faint or absent emission lines, their old effective ages (mass- or light-weighted) of more than half a Gyr from absorption lines, or their large masses from kinematics (e.g., Kriek et al. 2006, 2009; van de Sande et al. 2013; Hill et al. 2016; Belli et al. 2017a,b; Glazebrook et al. 2017). High-resolution imaging from the *Hubble* Space Telescope (HST) simultaneously showed that distant QGs also display de Vaucouleurs (1948)-type density profiles, and effective radii getting increasingly larger with time possibly as a result of dry merging (e.g., van Dokkum et al. 2008; Newman et al. 2012; Muzzin et al. 2012; van der Wel et al. 2014b).

The existence of these galaxies in the young Universe poses a number of interesting and still unanswered questions. Chief among them is probably the fact that, according to our current understanding of cosmology, galaxies are not closed boxes but are continuously receiving additional gas from the intergalactic medium through infall (e.g., Press & Schechter 1974; Audouze & Tinsley 1976; Rees & Ostriker 1977; White & Rees 1978; Tacconi et al. 2010). While the specific infall rate should go down with time as the density contrast in the Universe sharpens and the merger rate decreases (e.g., Lacey & Cole

1993), gas flows still remain large enough to sustain substantial star formation in massive galaxies, where feedback from supernovae is inefficient (e.g., Benson et al. 2003), and also in clusters (Fabian 1994). A mechanism must therefore be invoked in massive galaxies, either to remove this gas from the galaxies, or to prevent it from cooling down to temperatures suitable for star formation. To produce observationally-identifiable QGs, this mechanism must act over at least the lifetime of OB stars, a few tens of Myr, and should be allowed to persist over longer periods to explain their observed ages of up to several Gyr (e.g., Kauffmann et al. 2003b). This mechanism has been dubbed "quenching" (e.g., Bower et al. 2006; Faber et al. 2007).

Nowadays, the most favored actor for quenching is the feedback that slow and fast-growing supermassive black holes can apply on their host galaxies (e.g., Silk & Rees 1998; Bower et al. 2006; Croton et al. 2006; Hopkins et al. 2008; Cattaneo et al. 2009). During the fastest accretion events (e.g., during a galaxy merger), the energetics of these active galactic nuclei (AGNs) is such that they are capable of driving powerful winds and remove gas from the galaxy, resulting in so-called quasar-mode feedback. However this mechanism alone cannot prevent star formation over long periods of time. Indeed, the expelled gas eventually re-enters the galaxy. This gas must first cool down (hence form stars) before reaching the galaxy's center, fueling black hole growth, and triggering a new quasar event. There is therefore a need to introduce a heating source to prevent the gas infalling on quiescent galaxies from cooling (this need was first identified in the core of galaxy clusters, e.g., Blanton et al. 2001). This long-lasting, less violent mechanism could then maintain the quiescence established by a quasar episode.

Lower levels of accretion onto central black holes can fulfill this role, by injecting energy into the halo of their host galaxy with jets (see Croton et al. 2006). However this is not the sole possible explanation. In particular, "gravitational heating" of infalling gas in massive dark matter halos can have the same net effect (Birnboim & Dekel 2003; Dekel & Birnboim 2008), while stabilization of extended gas disks by high stellar density in bulges can also prevent star formation on long timescales (Martig et al. 2009).

While all of these phenomena have been shown to play some role in quenching galaxies, it remains unknown which (if any) is the dominant process. For example, recent simulations show that the QG population up to $z \sim 2$ can be reproduced without the violent feedback of AGNs and instead simply shutting off cold gas infall, leaving existing gas to be consumed by star formation (Gabor & Davé 2012; Davé et al. 2017). Furthermore, the observation of significant gas reservoirs in higher redshifts QGs, as well as SFGs transitioning to quiescence, suggests that quenching is not simply caused by a full removal of the gas, but is accompanied (and, perhaps, driven) by a reduced star-formation efficiency (e.g., Davis et al. 2014; Alatalo et al. 2014, 2015; French et al. 2015; Schreiber et al. 2016; Suess et al. 2017; Lin et al. 2017; Gobat et al. 2018). A complete census of QGs across cosmic time and a better understanding of their star formation histories are required to differentiate these different mechanisms.

Because of their low sSFR and the lack of young OB stars, QGs necessarily have red optical colors. For this reason they are usually identified from said colors, as seen in broadband photometry either directly with observed bands (e.g., Franx et al. 2003; Daddi et al. 2004; Labbé et al. 2005) or by computing rest-frame colors when the redshift is known (e.g., Faber et al. 2007; Williams et al. 2009; Ilbert et al. 2010). However, dusty SFGs can contaminate such color-selected samples: while

quiescent galaxies are red, red galaxies are not necessarily quiescent. The rate of contamination probably depends on the adopted method and the quality of the data. Selection methods based on a single color (such as color-magnitude diagrams) were very successful in the local Universe, but suffer from high contamination at higher redshifts owing to the increasing prevalence of dusty red galaxies (e.g., Labbé et al. 2005; Papovich et al. 2006). Two-color criteria were later introduced to break the degeneracy between dust and age to first order, and allow the construction of purer samples (Williams et al. 2009; Ilbert et al. 2010). Compared to full spectral modeling coupled to a more direct sSFR selection, these color criteria are less model-dependent, particularly so in deep fields where the wavelength coverage is rich and interpolation errors are negligible. Because they are so simple to compute, observational effects are also simpler to understand. But as a trade of, the comparison with theoretical models is harder than with a more direct sSFR selection, since it requires models to predict synthetic photometry.

Recently, a number of QGs were identified at $z > 3$ with such color selection technique (Straatman et al. 2014; Mawatari et al. 2016). Their observed number density significantly exceeds that predicted by state-of-the-art cosmological simulations, with and without AGN feedback (e.g., Wellons et al. 2015; Sparre et al. 2015; Davé et al. 2016), and requires a formation channel at $z > 5$ with SFRs larger than observed in the mostly dust-free Lyman-break galaxies (LBGs; e.g., Smit et al. 2012, 2016). However, at the time the accuracy of color selections of QGs had not been tested beyond $z \sim 2$, and spectroscopic confirmation of their redshifts and properties was needed to back up these unexpected results.

For this reason, we have designed several observing campaigns to obtain near-infrared spectra of these color-selected $z > 3$ massive QGs with Keck–MOSFIRE. The first results from this data set were described in Glazebrook et al. (2017; hereafter G17), where we reported the spectroscopic confirmation of the most distant QG at $z \sim 3.7$, the first at $z > 3$, using Balmer absorption lines. While flags were raised owing to the detection of sub-millimeter emission toward this galaxy by ALMA (Simpson et al. 2017), we later demonstrated this emission originates from a neighboring dusty SFG, and provided a deep upper limit on obscured star formation in the QG (Schreiber et al. 2018b). The confirmed redshift and quiescence of this galaxy (ZF-COS-20115, nicknamed "Jekyll") provided the first definite proof that QGs do exist at $z > 3$, and the fact that these were found in cosmological surveys of small area (a fraction of a square degree) implies they are not particularly rare.

In this paper, we describe the observations and results for the entire sample of galaxies observed with MOSFIRE. Using this sample, we derived statistics on the completeness and purity of the UVJ color selection at $z > 3$, and used this information to derive updated number densities and star formation histories for QGs at these early epochs, to compare them against models.

In Sect. 2, we describe our observations and sample, including in particular the sample selection, the spectral energy distribution (SED) modeling, and the reduction of the spectra. In Sect. 3 we describe our methodology for the analysis of the spectra, and make an inventory of the observed spectral features, the line properties, and the measured redshifts. In particular, Sect. 3.7 discusses the revised UVJ colors. In Sect. 4 we discuss the quiescence and inferred star-formation histories for the galaxies with MOSFIRE spectra. In Sect. 5, we build on the results of the previous sections to update the number density of quiescent galaxies, using the full ZFOURGE catalogs, and dis-

cuss the link between the UVJ selection and the specific SFR. Lastly, Sect. 6 compares our observed number densities and star formation histories to state-of-the-art galaxy evolution models, while Sect. 7 summarizes our conclusions and lists possibilities for future works.

In the following, we assumed a ΛCDM cosmology with $H_0 = 70 \, \mathrm{km \, s^{-1} \, Mpc^{-1}}$, $\Omega_\mathrm{M} = 0.3$, $\Omega_\Lambda = 0.7$ and a Chabrier (2003) initial mass function (IMF) to derive physical parameters from the photometry and spectra. All magnitudes are quoted in the AB system, such that $m_\mathrm{AB} = 23.9 - 2.5 \log_{10}(S_\nu \, [\mu \mathrm{Jy}])$.

2. Sample selection and observations

This section describes the sample of galaxies we analyzed in this paper, the new MOSFIRE observations, the associated data reduction, and the analysis of the spectra.

2.1. Parent catalogs

The sample studied in this paper consists of $3 < z < 4$, massive ($M_* \geq 2 \times 10^{10} \, M_\odot$) galaxies identified using photometric redshifts. The UVJ color-color diagram was then used to separate star-forming and quiescent galaxies (Williams et al. 2009). The galaxies were selected either from the ZFOURGE or 3DHST catalogs (Skelton et al. 2014; Straatman et al. 2016) in the CANDELS fields EGS/AEGIS, GOODS–South, COSMOS, and UDS (Grogin et al. 2011; Koekemoer et al. 2011). All fields include a wide variety of broadband imaging ranging from the U band up to the *Spitzer* 8 μm channel. This includes in particular (5σ limiting magnitudes quoted for EGS, GOODS-S, COSMOS, and UDS, respectively): deep *Hubble* imaging in the F606W ($R < 26.8, 27.4, 26.7, 26.8$); F814W ($I < 26.4, 27.2, 26.5, 26.8$); F125W ($J < 26.3, 26.1, 26.1, 25.8$); F160W ($H < 26.1, 26.4, 25.8, 25.9$); deep K_s or K-band imaging ($K < 24, 24.8, 25, 24.9$); and deep *Spitzer* 3.6 and 4.5 μm imaging ($[3.6] < 25.2, 24.8, 25.1, 24.6$). The photometry in these catalogs was assembled with the same tools and approaches, namely aperture photometry on residual images cleaned of neighboring sources (see Skelton et al. 2014; Straatman et al. 2016).

The ZFOURGE catalogs supersede the 3DHST catalogs by bringing in additional medium bands from $\lambda = 1.05 - 1.70 \, \mu\mathrm{m}$ and deeper imaging in the K_s band (obtained with the *Magellan* FourStar camera). The additional near-infrared filters allow a finer sampling of the Balmer break at $z \sim 2 - 3$, and more accurate photometric redshifts. However, they only cover a $11' \times 11'$ region within each of the southern CANDELS fields (GOODS–South, UDS, and COSMOS). We thus used the higher quality data from ZFOURGE whenever possible, and resorted to the 3DHST catalogs outside of the ZFOURGE area. In both cases, we only used galaxies with a flag use = 1. In ZFOURGE, we used an older version of the use flags than that provided in the DR1. Indeed, the latter were defined to be most conservative, in that they flag all galaxies which are not covered in all FourStar bands, those missing HST imaging, or those too close to star spikes in optical ground-based imaging (Straatman et al. 2016). This would effectively reduce the covered sky area by excluding galaxies which, albeit missing a few photometric bands, are otherwise well characterized. Instead, we adopted the earlier use flags from Straatman et al. (2014), which are more inclusive.

After the sample was assembled, a few source-specific adjustments were applied to the catalog fluxes. For ZF-COS-17779, we discarded the CFHT photometry which had negative fluxes with high significance (although inspecting the images

did not uncover any particular issue). For 3D-EGS-26047 we removed the WirCAM J band which was incompatible with the flux in the surrounding passbands (including the Newfirm J medium bands), and for which the image showed some artifacts close to the source. For 3D-EGS-40032, we discarded the Newfirm photometry because the galaxy was at the edge of the FOV; the noise in the image at this location is higher but the error bars reported in the catalog were severely underestimated, visual inspection of the image revealed no detection. For 3D-EGS-31322, we removed the *Spitzer* 5.8 and 8 μm fluxes which were abnormally low; the galaxy is located in a crowded region, and the photometry in these bands may have been poorly de-blended. These modifications are minor, and do not impact our results significantly. Lastly, for ZF-COS-20115 (the G17 galaxy) we used the photometry derived in Schreiber et al. (2018b), where the contamination from a dusty neighbor (*Hyde*) was removed. This reduced the stellar mass of ZF-COS-20115 by 30% and had no impact on its inferred star formation history (see Schreiber et al. 2018b).

In this paper, our main focus is placed on quiescent galaxies observed with MOSFIRE (this sample is described later in Sect. 2.4). However, to place these galaxies in a wider context, we also considered all massive galaxies in the parent sample at $3 < z < 4$. For this purpose, we only used the ZFOURGE catalogs (in GOODS–South, COSMOS, and UDS) since they have data of similarly high quality, and are all K_s-selected (while the 3DHST catalogs were built from a detection image in F125W+F140W+F160W). To further ensure reliable photometry, we only considered galaxies with $K_s < 24.5$; the impact of this magnitude cut on the completeness is addressed in the next section. We visually inspected the SEDs and images of all galaxies with $M_* > 10^{10} M_\odot$ to reject those with problematic photometry (3% of the inspected galaxies). In the end, the covered area was 139, 150, and 153 arcmin^2 in GOODS–South, COSMOS, and UDS, respectively.

2.2. Initial photometric redshifts and galaxy properties

The photometric redshifts (z_{phot}), rest-frame colors ($U - V$ and $V - J$), and stellar masses (M_*) provided in the ZFOURGE and 3DHST catalogs were computed with the same softwares, namely EAzY (Brammer et al. 2008) and FAST (Kriek et al. 2009), albeit with slightly different input parameters. These values were used to build the MOSFIRE masks in the different observing programs. However, to ensure the most homogeneous data set for our analysis, we recomputed redshifts, colors, and masses for all galaxies once the sample was compiled, using a uniform setup for all fields and taking advantage of all the available photometry. This setup is described below.

Photometric redshifts and rest-frame colors were obtained with the latest version of EAzY[1] and the galaxy template set "eazy_v1.3", which includes in particular a "old and dusty" and a "high-equivalent-width emission line" template. These additional templates were also used in the original ZFOURGE catalogs, but not in 3DHST. We also did not enable the redshift prior based on the K-band magnitude since this prior is based on models which do not reproduce the high redshift mass functions (see discussion in Sect. 6). The resulting scatter in photometric redshifts was 5% when comparing our new redshifts to that published by ZFOURGE for the entire catalog at $z > 3$, and 7% for the quiescent galaxies (described later in Sect. 2.4).

Stellar masses and SFRs were re-computed using FAST++[2] v1.2 with the same setup as in Schreiber et al. (2018b), but with more refined star-formation histories. Briefly, we assumed $z = z_{phot}$, the Bruzual & Charlot (2003) stellar population model, the Chabrier (2003) initial mass function (IMF), and the dust screen model of Calzetti et al. (2000) with A_V up to 6 mag. The only notable difference with the published ZFOURGE and 3DHST catalogs is that we assumed a more elaborate functional form for the star formation history (SFH), which consisted of two main phases: an exponentially rising phase followed by an exponentially declining phase, both with variable e-folding times τ_{rise} and τ_{decl}, respectively:

$$\text{SFR}_{base}(t) \propto \begin{cases} e^{(t_{burst} - t)/\tau_{rise}} & \text{for } t > t_{burst}, \\ e^{(t - t_{burst})/\tau_{decl}} & \text{for } t \leq t_{burst}, \end{cases} \quad (1)$$

where t is the "lookback" time ($t = 0$ is the point in time when the galaxy is observed, $t > 0$ is in the galaxy's past). This was performed assuming $z = z_{phot}$ initially, and later on with $z = z_{spec}$ (Sect. 4.1). Varying the lookback time t_{burst} that separates these two epochs, this allowed us to describe a large variety of SFHs, including rapidly or slowly rising SFHs, constant SFHs, and rapidly or slowly quenched SFHs (see Schreiber et al. 2018b for a more detailed description of this model). Allowing rising SFHs in particular can prove crucial to properly characterize massive SFGs at high redshift (Papovich et al. 2011). We varied t_{burst} from 10 Myr to the age of the Universe (at most 2 Gyr at $z > 3$), and τ_{rise} and τ_{decl} from 10 Myr–3 Gyr, all with logarithmic steps (0.05 dex for t_{burst}, 0.1 dex for τ_{rise} and τ_{decl}).

In addition, following Ciesla et al. (2016, 2017) and G17, we decoupled the current SFR from the past history of the galaxy by introducing a free multiplicative factor to the instantaneous SFR within a short period, of length t_{free}, preceding observation:

$$\text{SFR}(t) = \text{SFR}_{base}(t) \times \begin{cases} 1 & \text{for } t > t_{free}, \\ R_{SFR} & \text{for } t \leq t_{free}. \end{cases} \quad (2)$$

We considered values of t_{free} ranging from 10 to 300 Myr, and values of R_{SFR} ranging from 10^{-2} to 10^5 (i.e., either abrupt quenching or bursting), with logarithmic steps of 0.5 and 0.2 dex, respectively. We emphasize that this additional parameter is not directly linked to quenching, as a galaxy may still have a very low sSFR from Eq. (1) alone (see Fig. 1). In fact, as discussed later in Sect. 4.2, this additional freedom had little impact on the quiescent galaxies beside marginally increasing the uncertainty on the SFH, however we find it is necessary to properly reproduce the bulk properties of the star-forming galaxies. In particular, without this extra freedom the mean sSFR of main-sequence galaxies was too low by a factor of about three compared to stacked *Herschel* and ALMA measurements (this is also an issue affecting the SFRs provided in the original ZFOURGE and 3DHST catalogs).

This model is illustrated in Fig. 1, and the parameters with their respective bounds are listed in Table 1. Over 200 million models were considered for each galaxy, and the fit could be performed on a regular desktop machine in less than a day thanks to the numerous optimizations in FAST++. The adopted parametrization described above may seem overly complex, and indeed most of the free parameters in Eqs. (1) and (2) have little chance to be constrained accurately. This was not our goal however, since we eventually marginalized over all these parameters to compute more meaningful quantities, such as the current
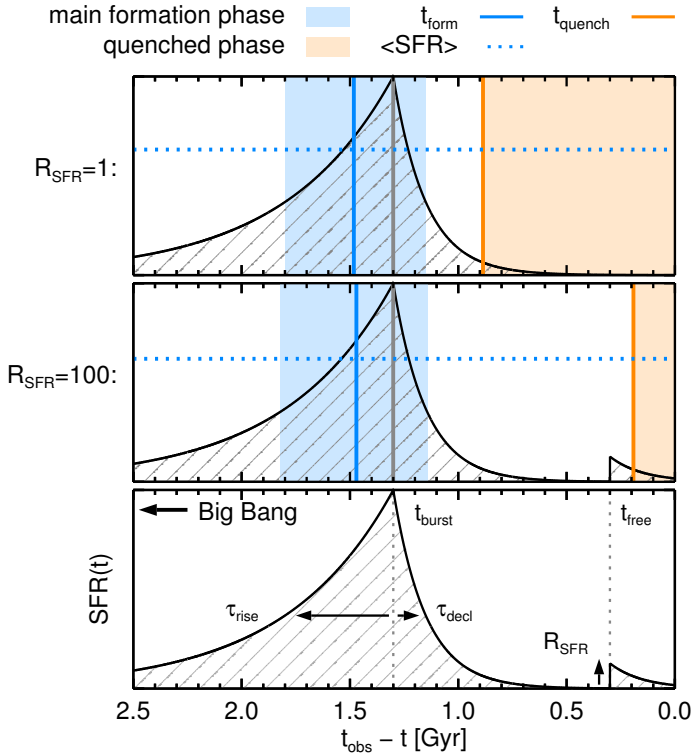

Fig. 1. Illustration of the adopted star formation history parametrization (*bottom*) and the marginalized parameters (*middle* and *top*). We show the time of peak SFR (solid gray line, here coinciding with t_{burst}), the star-forming phase surrounding it (shaded in pale blue), and the mean SFR during this phase (horizontal blue dotted line). We also display the time of quenching t_{quench} (orange solid line) and the following quenched phase (shaded in pale orange). Finally, the time at which the galaxy had formed 50% its stars (t_{form}) is shown with a blue solid line.

Table 1. List of model parameters in our SED modeling.

Parameter (unit)	Low	Up	Step	Values
t_{burst} (Gyr)	0.01	$t_H(z)$	0.05^a	45–50
τ_{rise} (Gyr)	0.01	3	0.1^a	26
τ_{decl} (Gyr)	0.01	3	0.1^a	26
R_{SFR}	10^{-2}	10^5	0.2^a	36
t_{free} (Myr)	10	300	0.5^a	4
A_V (mag)	0	6	0.02	61
Z		$Z_\odot$		
z		z_{phot} or z_{spec}		
IMF		Chabrier (2003)		
Attenuation curve		Calzetti et al. (2000)		
Stellar population		Bruzual & Charlot (2003)		

Notes. $^{(a)}$ Logarithmic step, in dex.

SFR and stellar mass, and non-parametric quantities describing the SFH (see Fig. 1 and Sect. 4.1). The point of introducing such complexity is therefore to allow significant freedom on the SFH, to avoid forcing too strong links between the current and past SFR, as well as to obtain accurate error bars on the aforementioned quantities. A similar approach was adopted in G17.

We then compared our best-fit values to that initially given in the ZFOURGE and 3DHST catalogs. Considering all galaxies at $3 < z_{phot} < 4$ and $M_* > 10^{10} M_\odot$, we find a scatter in stellar masses of 0.07 dex with a median increase of +0.04 dex (our new masses

are slightly larger), while the scatter in SFR is 0.34 dex and a median increase of +0.26 dex (our SFRs are substantially larger).

To estimate the completeness in mass of our sample resulting from our $K_s < 24.5$ magnitude cut, we binned galaxies in sSFR and computed in each bin the 80th percentile of the mass-to-light ratio in K, $\langle M_*/L_K \rangle$ (where L_K is the luminosity in the observed K_s band and M_* is the best-fit stellar mass obtained with FAST++). We note that this method accounts for changes in M/L caused both by variations in stellar populations as well as variation in dust obscuration. Since galaxies with low sSFR tend to be less obscured at fixed mass (Wuyts et al. 2011), these two effects work in opposite directions and can lead to a weaker evolution of M/L with sSFR. In practice, we find $\langle M_*/L_K \rangle = 1.6 \, M_\odot/L_\odot$ for sSFR $= 10^{-3} \, \mathrm{Gyr}^{-1}$, and $0.24 \, M_\odot/L_\odot$ for sSFR $= 10 \, \mathrm{Gyr}^{-1}$. Our adopted magnitude cut of $K_s < 24.5$ implies $M_* > 2.3 \times 10^{10} \langle M_*/L_K \rangle$ at $z = 3.5$, hence a 80% completeness down to $3.7 \times 10^{10} \, M_\odot$ for sSFR $< 10^{-3} \, \mathrm{Gyr}^{-1}$ (this is consistent with the value obtained in Straatman et al. 2014), and a factor seven lower at sSFR $= 10 \, \mathrm{Gyr}^{-1}$.

2.3. MOSFIRE masks and runs

MOSFIRE (McLean et al. 2012) is a multi-object infrared spectrograph installed on the Keck I telescope, on top of Mauna Kea in Hawaii. Its field of view of $6' \times 3'$ can be used to simultaneously observe up to 46 slits per mask, with a resolving power of $R \sim 3500$ in a single band ranging from Y (0.97 μm) to K (2.41 μm). The data presented here make use only of the H and K bands.

The quiescent galaxies studied in this paper were observed by four separate MOSFIRE programs comprising 10 different masks, listed in Table 2. All masks contained a bright "slit star", detected in each exposure, which was used a posteriori to measure the variations of seeing, alignment, and effective transmission with time (see Appendix B). Slits were configured with the same width of 0.7″ (except for the mask COS-Y259-A which had 0.9″ slits), and masks were observed with the standard "ABBA" pattern, nodding along the slit by ±1.5″ around the target position. Individual exposures lasted 120 and 180s in the H and K bands, respectively.

The first program was primarily targeting $z \sim 3.5$ quiescent galaxies (PI: Glazebrook), and observed one mask in EGS, one mask in COSMOS, and one mask in UDS (masks COS-W182, UDS-W182, and EGS-W057). Each mask was observed in the H and K filters, with on-source integration times ranging from 0.3 to 3.9 h in H, and 2.4−7.2 h in K. The masks were filled in priority with quiescent galaxy candidates identified in Straatman et al. (2014; or from the 3DHST catalogs for EGS), and our MOSFIRE observations for the brightest of these galaxies were already discussed in G17. The remaining slits were filled with massive $z \sim 4$ star-forming galaxies, and $z \sim 2$ galaxies; these fillers are not discussed in the present paper, and were only used for alignment correction and data quality tests. The SEDs of all galaxies were visually inspected, and this determined their relative priorities in the mask design.

The second and third programs (PIs: Oesch, Illingworth) were more broadly targeting massive galaxies at $2 < z < 3.6$ identified in the 3DHST catalogs, and quiescent galaxies were not prioritized over star-forming ones (see van Dokkum et al. 2015). These programs consisted of multiple masks in EGS, COSMOS and UDS, however all the quiescent candidates in EGS were at $z < 3$. We thus only used a total of three masks in COSMOS, and three masks in UDS (masks COS-Y259-A, COS-Y259-B, UDS-Y259-A, UDS-Y259-B, COS-U069, and UDS-U069). Only one

Table 2. MOSFIRE masks used in this paper.

| Mask	PI	Observing date		Integration time		Average seeing		Quiescent
		H	K	H	K	H	K	candidates
COS-W182	Glazebrook	2016-Feb-26, 27	2016-Jan-8, 2016-Feb-27	3.9 h	7.2 h	0.75″	0.61″	5
COS-U069	Illingworth	2014-Dec-16	2014-Dec-16	0.3 h	3.6 h	0.80″	0.55″	2
COS-Z245	Kewley	–	2017-Feb-14	–	1.6 h	–	0.61″	2
COS-Y259-A	Oesch	–	2014-Dec-13	–	3.3 h	–	0.71″	1
COS-Y259-B	Oesch	–	2014-Dec-14	–	2.0 h	–	0.57″	1
EGS-W057	Glazebrook	2017-Feb-13, 14	2016-Feb-26, 27	0.8 h	4.8 h	0.63″	0.65″	6
UDS-W182	Glazebrook	2016-Jan-8	2016-Jan-8	0.3 h	2.4 h	0.69″	0.65″	4
UDS-U069	Illingworth	–	2014-Dec-16	–	4.7 h	–	0.66″	1
UDS-Y259-A	Oesch	–	2014-Dec-13	–	4.9 h	–	0.63″	5
UDS-Y259-B	Oesch	–	2014-Dec-14	–	4.0 h	–	0.75″	4
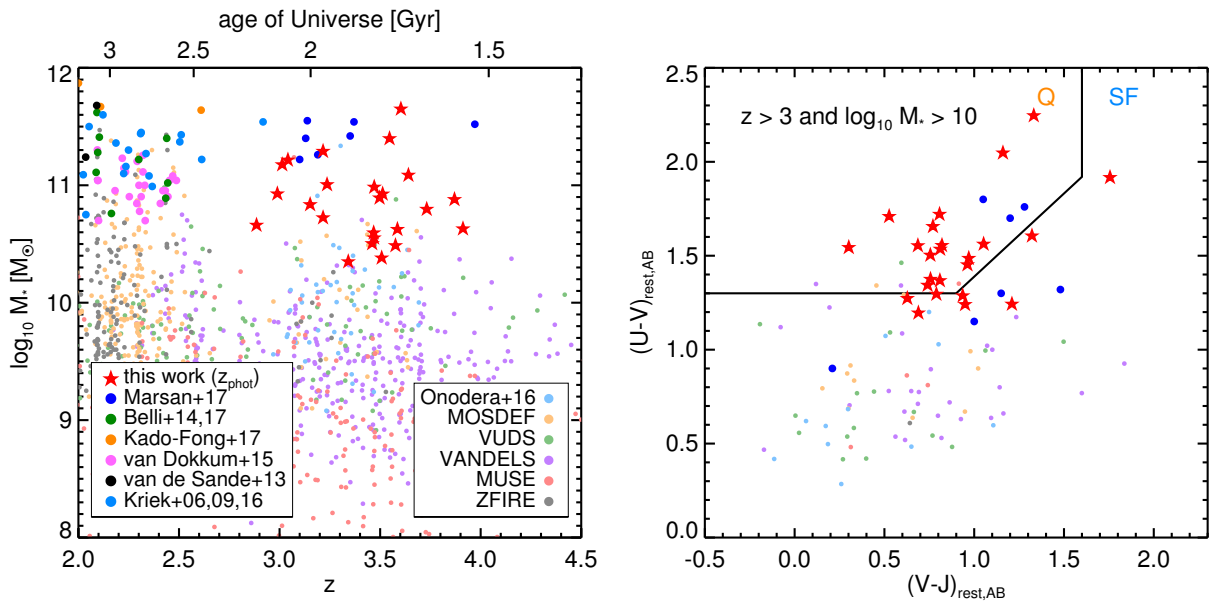
Fig. 2. *Left*: stellar mass as a function of redshift for galaxies with public spectroscopic redshifts from the literature (circles) and galaxies from our sample with photometric redshifts (red stars, using photometric redshifts). We show the sample of massive $z > 3$ galaxies from Marsan et al. (2017) in dark blue, the sample of Onodera et al. (2016) in light blue, the quiescent $z \sim 2$ galaxies of Belli et al. (2014, 2017a) in dark green and Kado-Fong et al. (2017) in orange, the compact $z \sim 2$ star-forming galaxies of van Dokkum et al. (2015) in pink, the quiescent galaxies observed in Kriek et al. (2006, 2009, 2016) in medium blue, the quiescent galaxies from van de Sande et al. (2013) in black, the galaxies observed by MOSDEF (Kriek et al. 2015) in orange, the galaxies observed by VUDS (Tasca et al. 2017) in green, the galaxies observed by VANDELS (McLure et al. 2017) in purple, the galaxies in the MUSE deep fields (Inami et al. 2017) in light pink, and the targets of the ZFIRE program in gray (Nanayakkara et al. 2016). *Right*: UVJ color-color diagram for a subset of galaxies shown on the left, limited to $z > 3$ and $M_* > 10^{10} M_\odot$. The $(U - V)$ and $(V - J)$ colors were computed in the rest frame, in the AB system. The black line delineates the standard dividing line between quiescent (Q) and star-forming (SF) galaxies, as defined in Williams et al. (2009).

mask was observed in the H band for 0.3 h, and all masks were observed in K with integration times ranging from 2.0 to 4.9 h.

The fourth and last program is the MOSEL emission line survey (PI: Kewley). This program observed several masks, in which massive $z \sim 4$ galaxies from ZFOURGE were only observed as fillers. Only two quiescent galaxy candidates were actually observed in one mask of the COSMOS field (mask COS-Z245), where 1.6 h was spent observing in the K band. One of them was the galaxy described in G17, for which the red end of the K was observed to cover the absent [O III] emission line.

2.4. Observed sample

From the MOSFIRE masks described in the previous section, we extracted all the galaxies with $z_{phot} > 2.8$, $M_* \geq 2 \times 10^{10} M_\odot$

and UVJ colors satisfying the Williams et al. (2009) criterion with a tolerance threshold of 0.2 mag. The resulting 24 quiescent galaxy candidates are listed in Table A.1, and their properties determined from the photometry alone (Sect. 2.2) are listed in Table A.2. The photometric redshifts ranged from $z_{phot} = 2.89$ up to 3.91, and stellar masses ranged from $M_* = 2.3 \times 10^{10}$ to $4.5 \times 10^{11} M_\odot$, as illustrated in Fig. 2. The broadband SEDs and best fit models using z_{phot} are shown in Fig. 3.

Some of our targets were observed in multiple MOSFIRE masks, and have accumulated more exposure time than the rest of the sample. In particular, ZF-COS-20115 (already described in G17) was observed for a total of 14.4 h in the K band and 4.2 h in H. Other galaxies have exposure times ranging from 1.6 h to 7.3 h in the K band, and zero to 3.9 h in H. The resulting line sensitivities are discussed in Sect. 2.7.
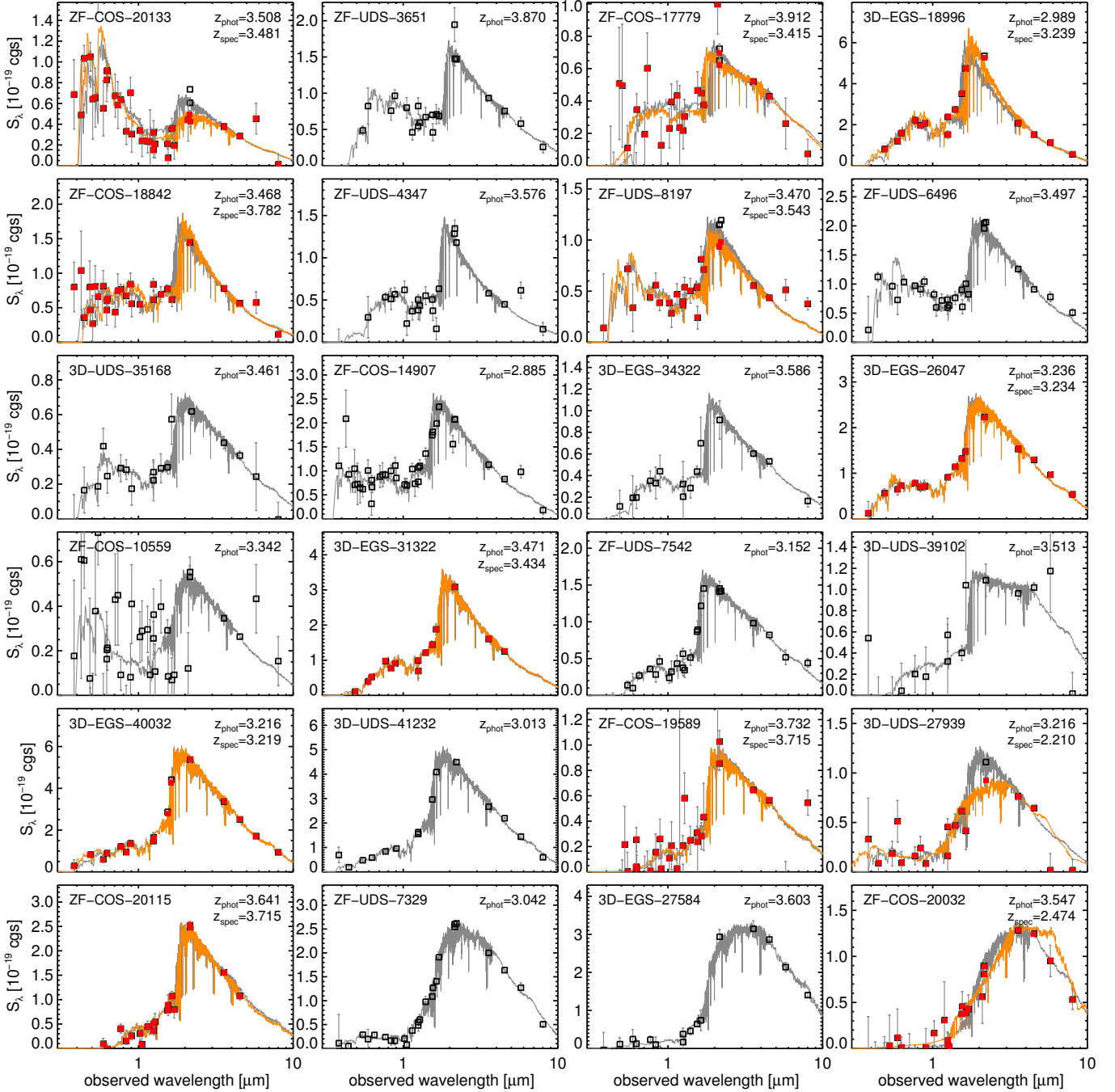
Fig. 3. Spectral energy distributions of the galaxies in our target sample, sorted by increasing observer-frame $z - K$ color (rest-frame $NUV - g$ at $z = 3.5$). The observed photometry is shown with open black squares and gray error bars, and the best-fitting stellar continuum template from FAST++ obtained assuming $z = z_{phot}$ is shown in gray in the background. For galaxies with a measured spectroscopic redshift (see Sect. 3.3), we display the best-fitting template at $z = z_{spec}$ in orange, and the photometry corrected for emission line contamination with red squares.

In Fig. 2 we compare this sample to recent spectroscopic campaigns targeting high-redshift galaxies. With the exception of the sample studied in Marsan et al. (2017), massive galaxies at $z > 3$ have so far received very limited spectroscopic coverage, and the situation is even worse for quiescent galaxies. Priority is often given to lower mass, bluer galaxies, for which redshifts can be more easily obtained with emission lines. Indeed, we checked that, despite being selected in the well studied CANDELS fields, none of our targets were observed by the largest spectroscopic programs (MOSDEF, VUDS, and VANDELS). The only exception is ZF-COS-20115 which was observed by MOSDEF for 1.6 h in K; we did not attempt to combine these data with our own given that this galaxy was already observed for 14 h and such a small increment would not bring significant improvement.

Combining data from these different programs, the collected MOSFIRE data have a non-trivial selection function. In some programs, galaxies were prioritized based on how clean their SEDs looked, which can bias our sample toward those quiescent candidates with the best photometry, or those with a more pronounced Balmer break. In addition, samples drawn from the 3DHST catalogs also tend to have lower redshifts and brighter magnitudes than that drawn from the ZFOURGE catalogs, as could be expected based on the different selection bands and
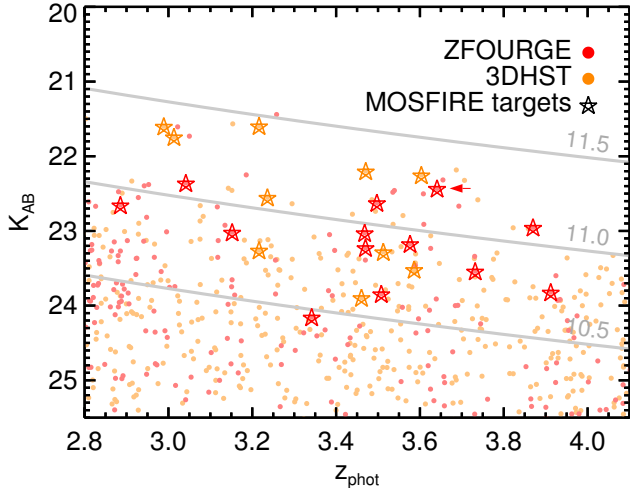
Fig. 4. K-band magnitude as a function of photometric redshift for UVJ quiescent galaxies (with a 0.2 mag threshold on the UVJ diagram). The quiescent galaxies from ZFOURGE are shown in red, while those from 3DHST in UDS and EGS are shown in orange. The stars indicate the galaxies for which we collected MOSFIRE spectra. The red arrow indicates the position of the $z = 3.7$ galaxy first discussed in G17. The gray solid lines show the K-band magnitude corresponding to different stellar masses, 3×10^{10}, 10^{11}, and $3 \times 10^{11} \, M_\odot$ (assuming $M_*/L_K = M_\odot/L_\odot$).

depths in these two catalogs. Yet, as shown in Fig. 4, the combined sample does homogeneously cover the magnitude-redshift or mass-redshift space for quiescent galaxies, within $3 < z < 4$ and $M_* > 4 \times 10^{10} \, M_\odot$ (or $K < 23.5$). We thus considered this spectroscopic sample to be fairly representative of the overall UVJ-quiescent population at these redshifts.

2.5. Reduction of the spectra

The reduction of the raw frames into 2D spectra was performed using the MOSFIRE pipeline as in Nanayakkara et al. (2016). However, since we were mostly interested in faint continuum emission, we performed additional steps in the reduction to improve the signal-to-noise and the correction for telluric absorption. The full procedure is described in Appendix B, and can be summarized as follows.

All masks were observed with a series of standard ABBA exposures, nodding along the slit. For each target, rather than stacking all these exposures into a combined 2D spectrum, we reduced all the individual "A − B" exposures separately and extracted a 1D spectrum for each pair of exposures. These spectra were optimally extracted with a Gaussian profile of width determined by the time-dependent seeing (hence, assuming the galaxies are unresolved), and were individually corrected for telluric correction and effective transmission using the slit star. Using the slit star rather than a telluric standard observed during the same night, we could perform the telluric and transmission correction for each exposure separately, rather than on the final data. This correction included slit loss correction, calibrated for point sources (see next section for the correction to total flux). The individual spectra were then optimally combined, weighted by inverse variance, to form the final spectrum. This approach allows to automatically down-weight exposures with poorer seeing. Flux uncertainties in each spectral element were determined by bootstrapping the exposures, and a binning of three spectral elements was adopted to avoid spectrally-correlated noise. This resulted in an average dispersion of $\lambda/\Delta\lambda \sim 3000$, which is close to the nominal resolution of MOSFIRE with 0.7″ slits.

Further binning or smoothing were used for diagnostic and display purposes, but all the science analysis was performed on these $\lambda/\Delta\lambda \sim 3000$ spectra. For this and in all that follows, binning was performed with inverse variance weighting, in which regions of strong OH line residuals were given zero weight.

2.6. Rescaling to total flux

Our procedure for the transmission correction includes the flux calibration, as well as slit loss correction. However, because the star used for the flux calibration is a point source, the slit loss corrections are only valid if our science targets are also point-like (angular size $\ll 0.6″$, the typical seeing, see Table 2). If not, additional flux is lost outside of the slit and has to be accounted for.

We estimated this additional flux loss by analyzing the H and K broadband images of our targets, convolved with a Gaussian kernel if necessary to match our average seeing (see Nanayakkara et al. 2016). We simulated the effect of the slit by measuring the broadband flux $S_{\rm slit}$ in a rectangular aperture centered on each target and with the same position angle as in the MOSFIRE mask, and by measuring the "total" flux in a 2″ diameter aperture, $S_{\rm tot}$. Since our transmission correction already accounted for slit loss for a point source, we also measured the fraction of the flux in the slit for a Gaussian profile of width equal to the seeing, $f_{\rm PSF,slit}$. We then computed the expected slit loss correction for extended emission as $S_{\rm tot} \times f_{\rm PSF,slit}/S_{\rm slit}$. The obtained values ranged from 1.0 (no correction) to 1.8 with a median of 1.2, and were multiplied to the 1D spectra.

We then compared the broadband fluxes from the ZFOURGE or 3DHST catalogs against synthetic fluxes generated from our spectra, integrating flux within the filter response curve of the corresponding broadband. Selecting targets which have a synthetic broadband flux detected at $>10\sigma$, we find that our corrections missed no more than 30% of the total flux, with an average of 10%. For fainter targets, this number reached at most 150%, and the highest values are found for the three faintest targets of the EGS-W057 mask (3D-EGS-26047, 3D-EGS-27584 and 3D-EGS-34322). While one of these three is intrinsically faint and thus has an uncertain total flux, the other two were expected to be detected with a synthetic broadband S/N of 19 and 25, but we find only 9 and 7, respectively. This may suggest a misalignment of the slits for these particular targets. To account for this and other residual flux loss, we finally rescaled all our spectra to match the ZFOURGE or 3DHST photometry. We only performed this correction if the continuum was detected at more than 5σ in the spectrum, to avoid introducing additional noise.

We noted that one galaxy's average flux in the H band was negative (3D-EGS-27584), and we also observed a strong negative trace in its stacked 2D spectrum. Because this galaxy is close (1.5″) to a bright $z \sim 1$ galaxy, we suspect that some of the bright galaxy's flux contaminated the H band. Regardless of the cause, this H-band spectrum was unusable. However, and perhaps owing to the neighboring galaxy being fainter in K, the K-band spectrum appeared unaffected and the target galaxy's continuum was well detected; we thus kept it in our sample and simply discarded the H-band spectrum.

2.7. Achieved sensitivities

The achieved spectral sensitivities and S/N in coarse 70 Å bins ($\sim 1000 \, {\rm km \, s^{-1}}$) are listed for all our targets in Table A.3. We describe in more detail the derivation of these uncertainties and their link to spectral binning in Appendix B.3. Because our sample is built from masks with different exposure times, the average sensitivity can vary from one galaxy to the next. In practice,

the median sensitivity (1σ, [min, median, max]) ranges between $[0.4, 0.7, 0.9] \times 10^{-19}\,\mathrm{erg\,s^{-1}\,cm^{-2}\,\mathring{A}^{-1}}$ in H band, and $[0.2, 0.5, 0.9] \times 10^{-19}\,\mathrm{erg\,s^{-1}\,cm^{-2}\,\mathring{A}^{-1}}$ in K, which resulted in continuum S/N of $[0.4, 1.3, 7.1]$ and $[0.7, 3.6, 12]$, respectively (these ranges reflect variation within our sample, and not variations of sensitivity within a given spectrum).

In terms of line luminosity at $z = 3.5$, assuming a width of $\sigma = 300\,\mathrm{km\,s^{-1}}$, these correspond to 3σ detection limits of $[0.5, 0.9, 1.1] \times 10^{42}\,\mathrm{erg\,s^{-1}}$ in H band, and $[0.3, 0.6, 1.1] \times 10^{42}\,\mathrm{erg\,s^{-1}\,cm^{-2}\,\mathring{A}^{-1}}$ in K. For Hβ in K and assuming no dust obscuration, this translates into 3σ limits on the SFR of $[4, 9, 16]\,M_\odot\,\mathrm{yr^{-1}}$ (see Sect. 4.3 for the conversion to SFR). For the massive galaxies in our sample, this is a factor $[0.04, 0.11, 0.20]$ times the main sequence SFR. With $A_V = 2$ mag, this is increased to a factor $[0.39, 0.98, 1.8]$. Therefore, on average, our spectra are deep enough to detect low levels of unobscured star-formation, or obscured star-formation in main-sequence galaxies.

Finally, given the observed K-band magnitudes of our targets and considering the median uncertainties listed above, these spectra allow us to detect lines contributing at least $[0.3, 1.1, 3.8]\%$ of the observed broadband flux (resp. [min,median,max] of our sample). This suggests we should be able to determine, in all our targets, if emission lines contribute significantly to their observed Balmer breaks. However this is assuming a constant uncertainty over the entire K band, which is optimistic. Indeed, a fraction of the wavelength range covered by the MOSFIRE spectra is rendered unexploitable because of bright sky line residuals.

To quantify this effect for each galaxy, we set up a line detection experiment in which we simulated the detection of a single line, of which we varied the full width from $\Delta\lambda = 100-1000\,\mathrm{km\,s^{-1}}$, and the central wavelength λ_0 within the boundaries of the K filter passband. In each case, we computed the line flux required for the line to contribute $f = 10\%$ to the observed broadband flux, accounting for the broadband filter transmission at the line's central wavelength. For simplicity, here we assume that the line has a tophat velocity profile and that the filter response does not vary over the wavelength extents of the line. By definition,

$$S_{\mathrm{BB}} = \frac{\int \mathrm{d}\lambda\, R(\lambda)\, S_\lambda(\lambda)}{\int \mathrm{d}\lambda\, R(\lambda)} \qquad (3)$$

where S_{BB} is the observed broadband flux density (e.g., in $\mathrm{erg\,s^{-1}\,cm^{-2}\,\mathring{A}^{-1}}$), $R(\lambda)$ is the broadband filter response, and $S_\lambda(\lambda)$ is the spectral energy distribution of the galaxy. Decomposing S_λ into a line and a continuum components, and with the above assumptions, we can extract the line peak flux density

$$S_{\mathrm{line}}(f, \lambda_0, \Delta\lambda) = f\, S_{\mathrm{BB}}\, \frac{\int \mathrm{d}\lambda\, R(\lambda)}{\Delta\lambda\, R(\lambda_0)}. \qquad (4)$$

For each galaxy, we then compared this line flux against the observed error spectrum, and computed the fraction of the K passband where such a line could be detected at more than 5σ significance. At fixed integrated flux, narrower lines should have a higher peak flux and thus be easier to detect, but they can also totally overlap with a sky line and become practically undetectable, contrary to broader lines. As we show below, in practice these two effects compensate such that the line detection probability does not depend much on the line width. We find that narrow lines ($100\,\mathrm{km\,s^{-1}}$) can be detected over $[73, 82, 92]\%$ of the K passband, while broad lines ($500\,\mathrm{km\,s^{-1}}$)

can be detected over $[77, 86, 96]\%$ (resp. [min,median,max] of our sample). Therefore the probability of missing a bright emission line, which we adopted as the average probability for the narrow and broad lines, is typically 15% per galaxy. The highest value is 27% (3D-UDS-35168) and is in fact more caused by lack of overall sensitivity toward the red end of the K band rather than by sky lines. We used these numbers later on, when estimating detection rates, by attributing a probability of missed emission line to each galaxy.

2.8. Archival ALMA observations

We cross matched our sample of quiescent galaxies with the ALMA archive and find that nine galaxies were observed, all in Band 7 except ZF-COS-20115 which was also observed in Band 8. The majority (ZF-COS-10559, ZF-COS-20032, ZF-COS-20115, ZF-UDS-3651, ZF-UDS-4347, ZF-UDS-6496, and 3D-UDS-39102) were observed as part of the ALMA program 2013.1.01292.S (PI: Leiton), which we introduced in Schreiber et al. (2017). ZF-COS-20115 was also observed in Band 8 in 2015.A.00026.S (PI: Schreiber; Schreiber et al. 2018b), ZF-UDS-6496 was also observed in 2015.1.01528.S (PI: Smail), while 3D-UDS-27939 and 3D-UDS-41232 were observed in 2015.1.01074.S (PI: Inami).

We measured the peak fluxes of all galaxies on the primary-beam-corrected ALMA images, and determined the associated uncertainties from the pixel RMS within a $5''$ diameter annulus around the source. Parts of the programs 2015.1.01528.S and 2015.1.01074.S were observed at high resolution (FWHM of $0.2''$) which may resolve the galaxies, therefore we re-reduced the images from these two programs with a tapering to $0.4''$ and $0.7''$ resolution, respectively, before measuring the fluxes (these were the highest values we could pick while still providing a reasonable sensitivity of about 0.3 mJy RMS). For ZF-COS-20115 we used the flux reported in Schreiber et al. (2018b), after deblending it from its dusty neighbor, resulting in a non-detection. In total, two quiescent galaxiy candidates were thus detected, ZF-COS-20032 and 3D-UDS-27939, with no significant spatial offset ($<0.2''$). As we show below, these are dusty redshift interlopers for which we detected Hα emission; we kept them in our analysis regardless, since they provide important statistics on the rate of interlopers. Since both galaxies are spatially extended, we used their integrated flux as measured from (u, v) plane fitting using uvmodelfit (as in Schreiber et al. 2017). Excluding ZF-COS-20032, 3D-UDS-27939, and ZF-COS-20115, the stacked ALMA flux of the remaining galaxies is 0.07 ± 0.11 mJy (using inverse variance weighting), indicating no detection. The collected fluxes are listed in Table A.1.

3. Redshifts and line properties

Here we describe the newly obtained spectroscopic redshifts, how they were measured, and how they compare to photometric redshifts. We also discuss the properties of the identified emission and absorption lines, and what information they provide on the associated galaxies.

3.1. Redshift identification method and line measurements

The spectra were analyzed with slinefit[3] to measure the spectroscopic redshifts. Using this tool, we modeled the observed spectrum of each galaxy as a combination of a stellar contin-

---

[3] https://github.com/cschreib/slinefit

uum model and a set of emission lines. The continuum model was chosen to be the best-fit FAST++ template obtained at $z = z_{phot}$ (see Sect. 2.2). The emission lines were assumed to have a single-component Gaussian velocity profiles, and to share the same velocity dispersion. The line doublets of [O III] and [N II] were fit with fixed flux ratios of 0.3, [O II] with a flux ratio of one, and [S II] with a flux ratio of 0.75, otherwise the line ratios were left free to vary. Emission lines with a negative best-fit flux were assumed to have zero flux, and the fit was repeated without these lines; we therefore assumed that the only allowed absorption lines had to come from the stellar continuum model from FAST++. This continuum model was convolved with a Gaussian velocity profile to account for the stellar velocity dispersion σ_*. Based on the empirical relation with the stellar mass observed at $z \sim 2$ in Belli et al. (2017a), we assumed $\log_{10}(\sigma_*/(km\,s^{-1})) = 2.4 + 0.33 \times \log_{10}(M_\star/10^{11}\,M_\odot)$.

The photometry was not used in the fit. Since we took particular care in the flux and telluric calibration of our spectra, we did not fit any additional color term to describe the continuum flux, a method sometimes introduced to address shortcomings in the continuum shape of observed spectra (e.g., Cappellari & Emsellem 2004). Even without such corrections, the reduced χ^2 of our fits are already close to unity (Table A.4), indicating that the quality of the fits are excellent and further corrections are not required. Furthermore, as discussed below, all the spectroscopic redshifts we measured are anchored on emission or absorption features anyway, which are not affected by such problems.

For each source, we systematically explored a fixed grid of redshifts covering $2 < z < 5$ in steps of $\Delta z = 0.0003$, fitting a linear combination of the continuum model and the emission lines and computing the χ^2. The redshift probability distribution was then determined from (e.g., Benítez 2000)

$$P(z) \propto \exp\left[-\frac{\chi^2(z) - \chi^2_{min}}{2\,C}\right]. \tag{5}$$

The constant C is an empirical rescaling factor described below. From this $P(z)$, we then estimated the probability p that the true redshift lies within ±0.01 of the best-fit redshift, namely:

$$p = \int_{-0.01}^{+0.01} du\, P(z_{peak} + u). \tag{6}$$

We considered as "robust", "uncertain" and "rejected" spectroscopic identifications those for with we computed $p > 90\%$, $50\% < p < 90\%$ and $p < 50\%$, respectively. The reliability of this classification is assessed in the next section.

Since not all our targets were expected to have detectable emission lines, we ran `slinefit` twice: with and without including emission lines. Doing so solved cases where the redshift got hooked on spurious positive flux fluctuations while the continuum was otherwise well detected (e.g., for 3D-EGS-31322). Comparing the outcome of this run to the run with emission lines, we kept the redshift determination with the highest p value.

To speed up computations and avoid unphysical fits, we first performed the fit only including the brightest emission lines, namely [S II], Hα, [N II], [O III], Hβ and [O II], and only allowing two velocity dispersion values: $\sigma = 60\,km\,s^{-1}$, which is essentially unresolved, and $300\,km\,s^{-1}$, the expected dispersion for galaxies of these masses. Once the redshift was determined, we ran `slinefit` again fixing $z = z_{spec}$, leaving the velocity dispersion free to vary from $\sigma = 60$ to $1000\,km\,s^{-1}$, and

adding fainter lines to the fit, namely [Ne III]$_{3689}$, [Ne IV]$_{2422}$, [Ne V]$_{3426}$, Mg II$_{2799}$, He II$_{4686}$, [O I]$_{6300}$, He I$_{5876}$, and [O III]$_{4363}$. From this run we computed the velocity dispersions, total fluxes and rest-frame equivalent widths of all lines. Uncertainties on all these parameters were determined from Monte Carlo simulations where the input spectrum was randomly perturbed within the uncertainties. We note that since we fit the lines jointly with a stellar continuum model, our line fluxes were automatically corrected for stellar absorption.

To make sure that our redshifts and line properties were not biased because the continuum models were obtained at $z = z_{phot}$ rather than $z = z_{spec}$, in a second step we re-launched the entire procedure described above, this time using the best-fit stellar continuum model obtained at $z = z_{spec}$ (see Sect. 4.1). The best-fit redshifts did not change significantly, except for one galaxy (ZF-UDS-6496, $z_{spec} = 3.207$ became 2.033) for which the redshift was anyway rejected ($p < 50\%$). No galaxy changed its classification category (e.g., from robust to uncertain) in the process, while line fluxes and equivalent widths changed by at most 2%. The differences were thus insignificant, but for the sake of consistency we used the results of this second run in all that follows.

3.2. Accuracy of the derived redshifts

In ideal conditions, namely if our search method was perfect and the noise in each spectral element of the spectrum was uncorrelated, Gaussian, and with an RMS equal to the corresponding value in the uncertainty spectrum, then the constant C in Eq. (5) should be set to one. However any of these conditions may be untrue, in which case we could attempt to compensate by setting $C > 1$ (which would effectively broaden the probability distribution). The reduced χ^2 we obtain are very close to one (see Table A.4), which should be a sign that our uncertainty spectra are in good agreement with the observed noise. However the reduced χ^2 is always dominated by the noise of the highest frequency (in the Fourier sense, i.e., one spectral element), while the continuum spectral features useful for the redshift determination actually span multiple spectral elements. Therefore this constant C has a different sensitivity to the noise properties compared to the reduced χ^2.

We thus calibrated C by simulating redshift measurements of artificial galaxies of various K-band magnitude added to pure sky spectra. We find that setting $C = 2$ is required to obtain accurate p values, as illustrated in Fig. 5.

In an attempt to investigate the source of this correction, we also performed an identical test on mock spectra produced with ideal Gaussian noise. Despite the ideal noise, we find that a correction is still required, with $C = C_{ideal} = 1.25$. This suggests that part of the needed correction is intrinsic to our redshift measurement method, and not related to the quality of the data. If we decompose $C = C_{ideal} \times C_{noise}$, we find $C_{noise} = 1.6$, which would be equivalent to stating that our uncertainty spectrum is underestimating the noise (on the relevant scales) by $\sqrt{C_{noise}} = 26\%$. This value is close to our estimate of the residual correlated noise in Appendix B.

Finally, we compared this automatic identification method to visual identification: all the redshifts that were visually identified (looking mostly for the [O III] doublet and Balmer absorption lines) were recovered with $p > 90\%$, except 3D-EGS-31322 for which $p = 84\%$. In addition, the automatic identification allowed us to obtain additional redshifts for galaxies with no detectable emission lines and with weak continuum emission, albeit with a reduced (but quantified) reliability.
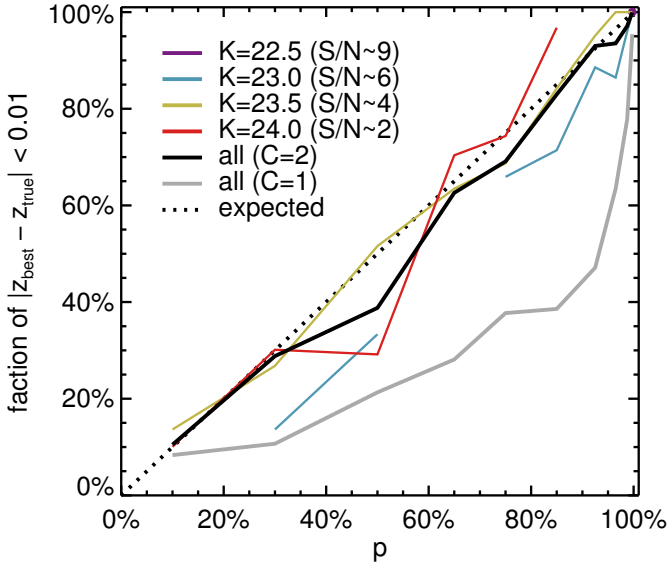
Fig. 5. Calibration of the criterion for redshift reliability, p, using simulated spectra. The p value quantifies the probability that the measured redshift lies within $\Delta z = 0.01$ of the true redshift. The x-axis shows the p value estimated from the $P(z)$ of the simulated spectra, and the y-axis is the actual fraction of the simulated redshift measurements that lie within $\Delta z = 0.01$ of the true redshift. The line of perfect agreement is shown wish a dashed black line. The relation obtained with $C = 2$ (see text) is shown with colored lines for simulated spectra of different K-band magnitude (the S/N given in parentheses corresponds to 70 Å bins), and for all magnitudes combined in black. All simulated galaxies with $K = 22$ had an estimated $p \sim 100\%$ and are therefore shown as a single data point in the *top-right* corner. The relation for all magnitudes and $C = 1$ is shown in gray for comparison.

3.3. Measured redshifts

A condensed overview of the outcome of the automatic redshift search is provided in Fig. 6. The results are listed in full detail in Table A.4, and illustrated for each galaxy in Figs. 7 and 8. In summary, we obtain a spectroscopic identification for 50% of our sample, with eight robust redshifts and four uncertain redshifts, and find a z_{phot} catastrophic failure rate of 8%, where the contaminants are $z \sim 2.5$ dusty galaxies. We quantify the accuracy of the photometric redshifts to a median $|z - z_{phot}|$ of 1.2%, which implies that even the galaxies without z_{spec} should be reliable. We describe these results in more detail in the following sub-sections.

3.3.1. Robust redshifts

In total, we obtain eight robust spectroscopic identifications, with z_{spec} ranging from 2.210 to 3.715. The highest measured redshift, $z_{spec} = 3.715$, is that of ZF-COS-20115, which was first reported in G17, and is based on the detection of Hβ, Hγ, and Hδ absorption. We note that this value is slightly lower than the redshift obtained in G17 ($z_{spec} = 3.717$); this results from the accumulation of more data, and a slightly different measurement method. The change, contained within the error bars, has no implication on the nature of the neighboring dusty source (Schreiber et al. 2018b). Balmer absorption lines are found in two other galaxies, 3D-EGS-18996 at $z_{spec} = 3.239$ and 3D-EGS-40032 at $z_{spec} = 3.219$ (see Fig. 7, top). These two galaxies being at slightly lower redshifts, the rest of the Balmer series appears at the red end of the H band. Although at this stage the continuum model was not yet fine-tuned to reproduce the

strength of the Balmer absorption lines (this is done later in Sect. 4.1), the quality of the fit is already excellent. This illustrates the good agreement between the photometric and spectroscopic age-dating, which was already pointed out in G17 and Schreiber et al. (2018b) when studying the case of ZF-COS-20115.

Beside these three galaxies, the rest of the redshifts were determined using emission lines. Two galaxies turn out to be redshift interlopers, ZF-COS-20032 at $z_{spec} = 2.474$ and 3D-UDS-27939 at $z_{spec} = 2.210$, for which we detected Hα and [N II]. These two galaxies are shown in Fig. 8 (bottom). ZF-COS-20032 is significantly extended in the F160W image and is detected by ALMA at $890\,\mu m$, which indicates it might be an obscured disk. 3D-UDS-27939 is also extended, and blended with another galaxy. In the 3DHST catalog, this blended system was split in two galaxies, one of which was our target with $z_{phot} = 3.22$, while the other was attributed a lower $z_{phot} = 2.24 \pm 0.02$. This value is in fact consistent with our measured z_{spec} for the quiescent candidate, which suggests the two objects are either a major merger, or two parts of the same galaxy with a strong attenuation gradient. Regardless, as can be seen in Figs. 7 and 8, the morphologies of both ZF-COS-20032 and 3D-UDS-27939 stands apart from that of the rest of the sample, where galaxies are typically more compact; this could be a natural consequence of the different mass-size relation for star-forming and quiescent galaxies (e.g., van der Wel et al. 2014a; Straatman et al. 2015).

The redshifts for the remaining three galaxies (ZF-COS-20133, 3D-EGS-26047, and ZF-UDS-8197) were obtained using a combination of the [O III] doublet, Hβ, and [O II]. ZF-COS-20133 and ZF-UDS-8197 are both found to have particularly bright [O III] emission and little to no Hβ and [O II], as shown in Fig. 8. Their line widths, however, are very different: the former has unresolved line profiles ($\sigma_v \leq 60\,km\,s^{-1}$) in both [O III] and Hβ, while the latter has extremely broad [O III] ($\sigma_v = 530 \pm 54\,km\,s^{-1}$). A more detailed description of the emission line properties of these galaxies is provided later in Sect. 3.5. Lastly, 3D-EGS-26047 has faint [O III] and Hβ lines of comparable fluxes, as a well as [O II]. As shown in Fig. 8, the lines of this galaxy are only marginally detected, and it is only by combining them in the redshift search that we could obtain a measure of the redshift (which is in excellent agreement with the z_{phot}).

The comparison of our spectroscopic redshifts against the photometric redshifts from EAzY is presented in Fig. 9. Excluding the two outliers, the agreement between the z_{spec} and z_{phot} is excellent: the largest $|z_{spec} - z_{phot}|/(1 + z)$ is 6.3% for 3D-18996, and the median is 1.1%.

3.3.2. Uncertain redshifts

A further four galaxies were attributed an uncertain redshift: ZF-COS-17779, ZF-COS-18842, ZF-COS-19589, and 3D-EGS-31322. We included in this list the galaxy ZF-COS-19589, whose redshift of $z_{spec} = 3.715$ (identified using Balmer absorption features, see Fig. 7) should have been rejected on the basis of its $p = 32\%$. Indeed, this redshift lies within $\Delta z < 0.01$ of ZF-COS-20115, which has a robust z_{spec} and is located only 23″ away. Based on the possibility of these two galaxies being physically associated, we gave extra credit to this z_{spec} and promoted it to the uncertain category. Otherwise, the constraints on the redshift form the spectrum alone are relatively poor, but the absence of a break in the K band rules out redshifts $z > 3.8$.

The galaxy 3D-EGS-31322 (shown in Fig. 7) has a well-detected continuum emission and a significant break at the red
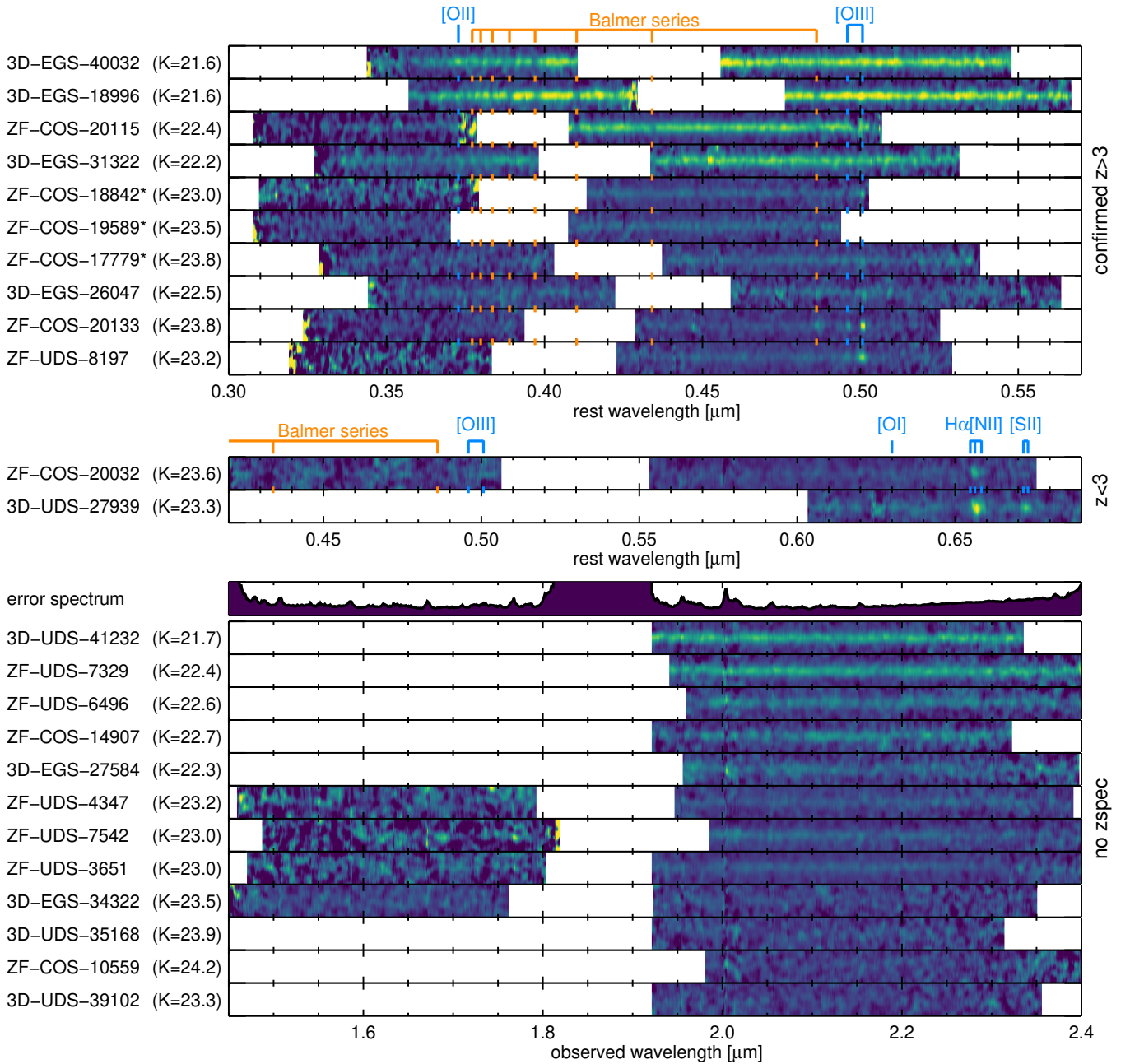
Fig. 6. 2D spectra of our sample, flux-calibrated and corrected for telluric absorption. For display purposes only, these spectra were smoothed with a 70 Å boxcar filter in wavelength and 0.7″ FWHM Gaussian along the slit. The galaxies are sorted by decreasing *K*-band continuum S/N. *Top*: galaxies with a spectroscopic redshift $z_{spec} > 3$, aligned on the same rest frame wavelength grid. The most prominent emission and absorption lines are labeled in blue and orange ticks, respectively. Galaxies with an uncertain redshift (see text) are marked with an asterisk. *Middle*: same as top but for $z_{spec} < 3$. *Bottom*: galaxies without spectroscopic redshift, aligned on the same observed wavelength grid. The average error spectrum is shown at the top, to illustrate the regions with the strongest atmospheric features (telluric absorption and OH lines).

end of the *H* band. This break is sufficient to confirm that the redshift is indeed $z \sim 3.5$, but a more precise redshift requires line identifications. Balmer absorption features may be identified, in particular Hε at the red edge of the *H* band and Hγ at the blue edge of the *K* band, along with weak [O II] emission. However the S/N is low enough that these identifications are ambiguous. In all cases however, Hβ absorption and [O III] emission are weak or non-existent.

The last two galaxies, ZF-COS-17779 and ZF-COS-18842 (shown in Fig. 7), are essentially identified using a single narrow emission line. While this emission is securely detected in both

cases (7.7 and 5.6σ, respectively), the identification of the corresponding emission line is partly degenerate. For both galaxies, the automated redshift search attributed this emission to the brightest line of the [O III] doublet, [O III]$_{5007}$. For ZF-COS-17779, this solution is also backed up by a plausible detection of Hβ (4.1σ) and tentative [O II] (1.7σ). For ZF-COS-18842 on the other hand, Hβ is detected at only 1.0σ, and since the line is located almost at the edge of the *K* band, the detected emission could also be attributed either to the fainter line of the [O III] doublet, [O III]$_{4959}$, or to Hβ. The only reason why these alternative solutions are disfavored is because they provided a poorer fit
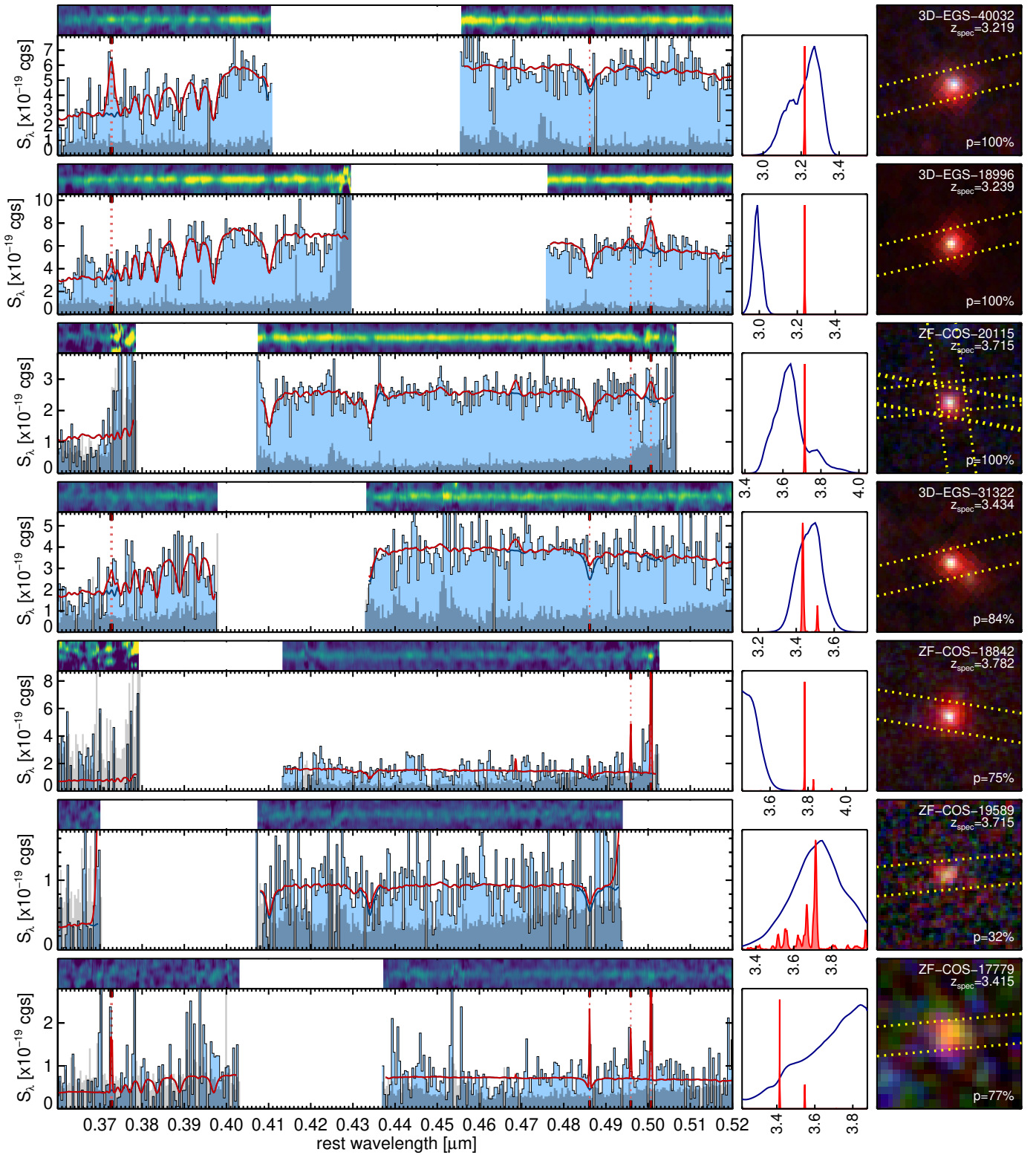
Fig. 7. From left to right: MOSFIRE spectrum, redshift probability distribution, and false-color image of the galaxies with a measured $z_{\rm spec}$. The galaxies are sorted as in Fig. 6. The spectrum on the left is displayed as a function of rest-frame wavelength. The observed spectrum is shown with a black solid line and blue shading, and the best-fit model obtained at the end of the redshift fitting procedure is shown in red. The uncertainty is shown as a dark shaded area at the bottom of each plot, and the 2D spectrum is displayed at the top, with smoothing to enhance the display. For the redshift probability distribution, the $p(z)$ from the spectra are shown in red, while the $p(z)$ from the photometry (EAzY) are shown in dark blue. Finally, the false-color images are composed of the WFC3-F125W (blue), WFC3-F160W (green) and $K_{\rm s}$ bands (red, either from ZFOURGE, HawK-I, or Ultra-VISTA), with linear scaling. Each image is $3.6'' \times 3.6''$ across. We also show the extents of the MOSFIRE slits as dotted yellow rectangles.
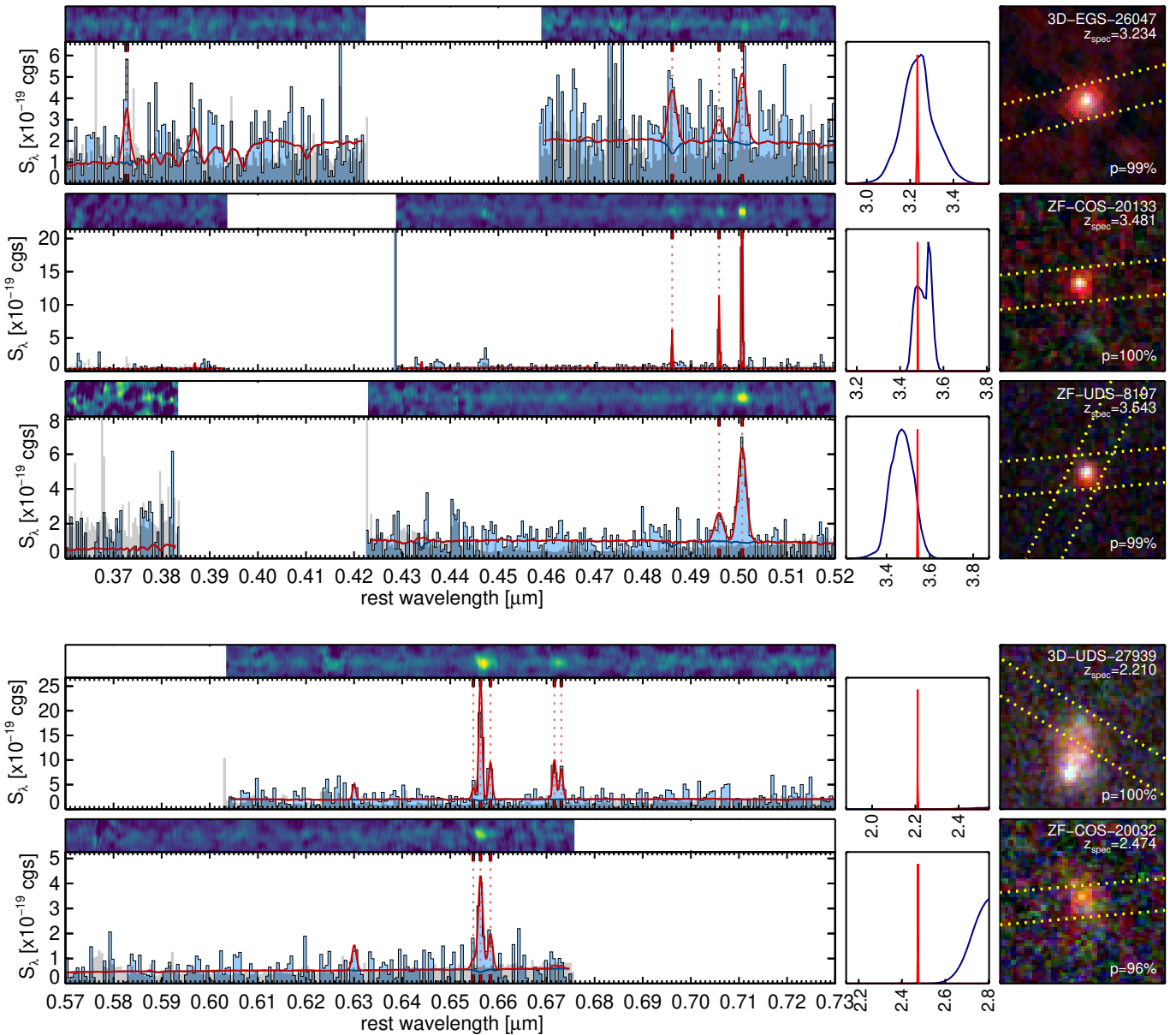
Fig. 8. Fig. 7 continued.

to the continuum emission, in particular regarding the presence of absorption features. Indeed, at these higher redshifts, the Hδ line enters the K band but does not correspond to any absorption feature in the observed spectrum, and thus would have created a tension of 2.0 and 2.9σ (if the detected emission line is $[O\,\textsc{iii}]_{4959}$ or Hβ, respectively). Likewise, the Hγ line is covered for all three solutions, and although there is no clear evidence that this absorption line is actually detected, the $[O\,\textsc{iii}]_{5007}$ solution provides the smallest tension (1.1σ, versus 2.4σ for the other two solutions). This evidence is however marginal, since an alternative possibility is that we overestimated the strength of the absorption lines in the continuum template, that is, if the galaxy is younger (or older) than its broadband photometry initially suggested.

Lastly, we manually rejected from the uncertain category the $z_{spec} = 4.194$ for the galaxy ZF-COS-14907 which had $p = 63\%$; its surprisingly high value (highest z_{spec} of all the sample), poor fit (reduced $\chi^2 = 1.2$, highest of all the sample), and blatant

inconsistency with the photometric redshift ($z_{phot} = 2.89^{+0.06}_{-0.06}$, ~20σ difference, again the highest of the sample) suggested an issue with the spectrum.

In the end, for the four galaxies in the uncertain category, the highest $|z_{spec} - z_{phot}|/(1 + z)$ is 10% for ZF-COS-17779, with a median of 3.9%. This is higher than for the galaxies of the robust sample, and could be expected since the uncertain galaxies are on average 0.7 mag fainter in K. Considering the combined robust and uncertain sample, the median $|z_{spec} - z_{phot}|/(1 + z)$ is 1.2%; we can therefore conclude that, save for the few galaxies with strong emission lines, the z_{phot} were highly accurate, confirming the results of Straatman et al. (2016) obtained with galaxy pairs.

3.3.3. Unconfirmed redshifts

We could not determine spectroscopic redshifts for the remaining 12 galaxies. As can be seen on Fig. 6, these
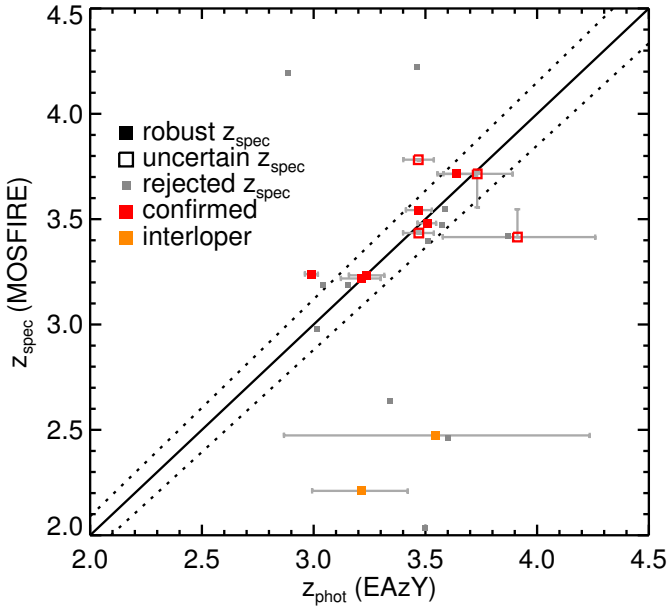
Fig. 9. Comparison between the photometric redshifts obtained with EAzY from the broadband photometry alone (z_{phot}) and the spectroscopic redshifts determined from the MOSFIRE spectra (z_{spec}). Robust redshifts ($p > 90\%$) are shown with large filled squares, uncertain redshifts ($50\% < p < 90\%$) are shown with open squares, and galaxies with rejected z_{spec} are shown with small gray squares. Galaxies confirmed to be at $z \geq 3$ are displayed in red, and interlopers are displayed in orange. The error bars show the 68% confidence intervals. The solid line shows the one-to-one match, while the dotted lines above and below indicate the typical z_{phot} uncertainty of 3%, as estimated for ZFOURGE at $3 < z < 4$ in Straatman et al. (2016). These values are listed in Table A.4.

are not particularly fainter, and a Kolmogorov–Smirnov test gives $p = 99\%$ of the two samples having the same K-band magnitude distribution. Likewise, their photometric redshift distribution is consistent with being the same as that of the spectroscopically confirmed galaxies (KS test: $p = 99\%$). However, the five brightest of these 12 galaxies have no H-band coverage from MOSFIRE. As demonstrated with 3D-EGS-40032 and 3D-EGS-18996, the H band can prove particularly useful in determining redshifts when the high-order lines from the Balmer series are observed, or simply to confirm the absence of continuum emission (e.g., ZF-COS-20115). For the bright but unconfirmed galaxies, a possible explanation for their lack of identification would be that they have weaker Balmer absorption owing to them having older or younger stellar populations. But, in general, it is also possible that we simply missed the emission lines because of sky lines. Indeed, based on the calculations in Sect. 2.7, we can statistically expect this to happen in two of these 12 galaxies.

Nevertheless, and statistically excluding two galaxies for which lines are not detectable because of sky lines, we could confirm that their K-band (and, for a few, also H-band) photometry is not significantly contaminated by emission lines (see Sect. 2.7). As per the above, this implies their photometric redshifts and derived UVJ colors should not suffer from systematic errors, hence that most of these unconfirmed galaxies should be reliable quiescent candidates. Consequently, ZF-COS-20032 and 3D-UDS-27939 should be the only two galaxies with catastrophic redshift failure, resulting in a failure rate of 8% (or 9% if we account for galaxies with potentially missed emission lines).

3.4. Stacked spectrum

We show in Fig. 10 the stack of all the eight $z > 3$ galaxies with robust or uncertain redshifts. This stack was obtained as the inverse-variance-weighted average flux, after each spectrum was re-normalized to a unit flux at rest-frame $0.48\,\mu m$. The galaxies that entered the stack have different rest-wavelength coverage (as shown in Fig. 6), such that only the wavelengths from 0.475 to $0.495\,\mu m$ (which includes Hβ) were covered for all galaxies. The [O III] and [O II] emission lines were covered in all but one galaxy, so their stacked amplitude should be representative, but Hγ and Hδ were only covered in half of the sample. We simultaneously stacked the best-fitting stellar continuum models of the galaxies (derived below in Sect. 4.1), using the same weighting.

In this stacked spectrum, the Balmer absorption series can be readily identified, with Hβ, Hγ, Hδ, Hε, Hζ, Hη and H10. We also identified the calcium H absorption feature (calcium K is blended with Hε), and tentatively the G-band and Mg I absorption. In emission, we only find [O II] and [O III]$_{5007}$ to be significantly detected in the residual spectrum.

3.5. Emission lines ratios and equivalent widths

The measured emission line properties for all galaxies in the robust and uncertain categories are listed in Table A.5. The most commonly detected line ($>2\sigma$) is the [O III] doublet, which was detected in five galaxies with EW$_{\mathrm{rest}}$ ranging from 14 to $282\,\text{Å}$ (median $49\,\text{Å}$), while Hβ emission was detected in three galaxies, with EW$_{\mathrm{rest}}$ ranging from 8 to $34\,\text{Å}$. For one galaxy, 3D-EGS-18996, we find [O III] in emission and Hβ in absorption, as shown in Fig. 7. We also formally detected [O II] in two galaxies, 3D-EGS-26047 and 3D-EGS-40032, with equivalent widths of 43 and $23\,\text{Å}$, respectively.

Among the galaxies with [O III] detections, the $\log_{10}($[O III]$/$H$\beta)$ ratio (corrected for Balmer absorption, see Sect. 3.1) ranges from $0.17^{+0.26}_{-0.30}$ (3D-EGS-26047) to $1.41^{+0.07}_{-0.26}$ (ZF-UDS-8197), with a median of 0.91. Using the stellar masses derived in the next section (or the ones initially derived at $z = z_{\mathrm{phot}}$), the mass-excitation diagram (Juneau et al. 2011) classifies all the [O III]-detected galaxies as "AGN", and this remains true even if we use the stricter criterion derived for $z > 2$ galaxies in Coil et al. (2015). Recent results suggest this criterion should be made even stricter, shifting the $z \sim 0$ criterion of Juneau et al. (2011) by 1 dex in mass (Strom et al. 2017); this would reduce the fraction of AGNs among our [O III] emitters to 40%, which remains substantial. The [O III] luminosity ranges from 1.5×10^8 to $1.8 \times 10^9\,L_\odot$ (0.7 to $7.8 \times 10^{42}\,\mathrm{erg\,s^{-1}}$). The line velocity profile are unresolved ($\sigma_v < 60\,\mathrm{km\,s^{-1}}$) for two galaxies, ZF-COS-20133 and ZF-COS-17779, and particularly broad for all other galaxies, with $\sigma_v = 530\text{--}582\,\mathrm{km\,s^{-1}}$. While the narrow [O III] in ZF-COS-20133 may be powered by an AGN, the broad [O III] of ZF-UDS-8197 should instead reflect shocked gas in the galaxy's gravitational potential, since [O III] is not produced in AGN broad line regions (e.g., Baldwin 1975). We defer further analysis of these line kinematics and links to AGN activity to a future paper.

Lastly, for the two redshift interlopers we detected the Hα line with an EW$_{\mathrm{rest}}$ of 84 ± 31 for ZF-COS-20032 and $109 \pm 9\,\text{Å}$ for 3D-UDS-27939. The [N II] doublet was weakly detected in the former, and more clearly in the latter; the resulting $\log_{10}($[N II]$/$H$\alpha)$ are $-0.33^{+0.27}_{-0.30}$ and $-0.40^{+0.06}_{-0.07}$, respectively, which are both inconclusive as they might correspond to any category in the Baldwin–Phillips–Terlevich (BPT; Baldwin et al. 1981) diagram (Kauffmann et al. 2003a). The [S II] doublet was
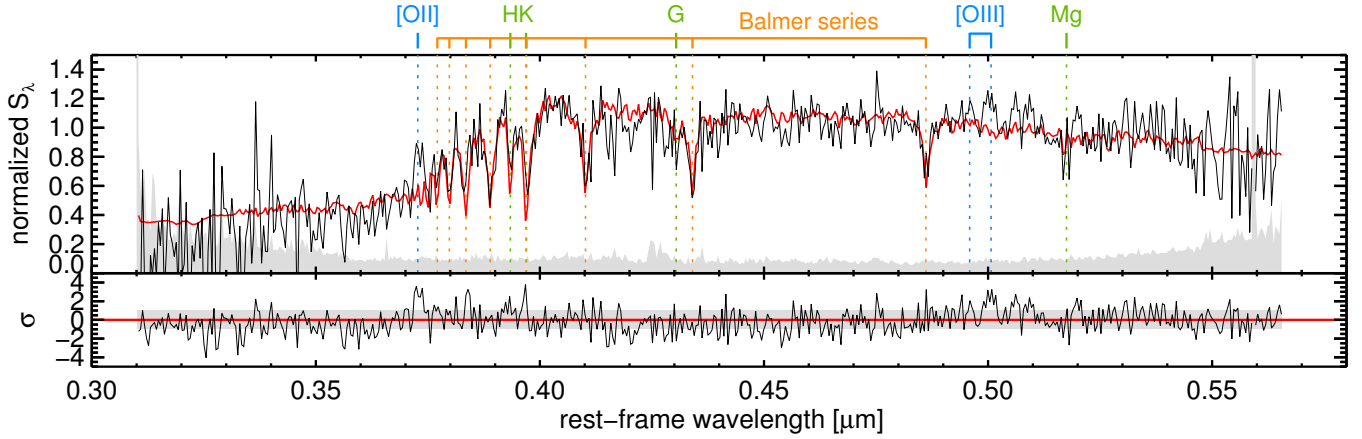
Fig. 10. Stacked rest-frame optical spectrum of the eight spectroscopically confirmed $z > 3$ galaxies with no strong emission lines (i.e., all the galaxies with robust or uncertain redshifts at $z > 3$, except ZF-COS-20133 and ZF-UDS-8197). The stacked MOSFIRE spectrum is shown in black, normalized to unit flux density at $\lambda \sim 0.48\,\mu m$, and the error spectrum is shown in shaded gray at the bottom. We overlay the stacked model spectrum of all the galaxies in red (as obtained in Sect. 4.1), and we indicate the main absorption and emission lines with colored lines (blue: emission, green: absorption, orange: Balmer series) with labels at the top of the figure. The residuals of the spectrum, after subtracting the stacked model and normalizing by the uncertainty, are displayed at the bottom of the figure.

also covered and only detected for 3D-UDS-27939, leading to $\log_{10}([S\,\textsc{ii}]/H\alpha) = -0.24^{+0.06}_{-0.07}$, which is similarly inconclusive.

Over the entire sample, "high-EW" emission line complexes with a summed $EW_{rest} > 100\,\text{Å}$ were observed in four galaxies. This implies that UVJ-selected samples are contaminated by high-EW lines at the rate of 17% (or 18% if we account for galaxies with potentially missed emission lines), half of these being redshift interlopers. Even if all of these high-EW galaxies happened to not be truly quiescent (which is a question we address later in Sects. 3.7 and 4), this would not affect the number densities (e.g., Straatman et al. 2014) in a significant way.

3.6. Subtracting emission lines from the broadband photometry

To go forward (see Sect. 4) we needed to analyze the continuum emission, using both the spectra and the broadband photometry. Since some of our galaxies displayed particularly high EW emission lines, we had to correct the H and K broadband photometry for this contamination. For each NIR broadband, we selected the lines with $EW_{obs}/\Delta EW_{obs} > 3$ and computed the corrected flux densities S^{cor}_{BB}. With a similar reasoning as in Sect. 2.7, we can derive

$$\frac{S_{BB}}{S^{cor}_{BB}} = 1 + \sum_{\ell} EW^{\ell}_{obs} \frac{R(\lambda_\ell)}{\int d\lambda\, R(\lambda)} \qquad (7)$$

where S_{BB} is the original flux density, $R(\lambda)$ is the response curve of the corresponding filter, and where EW^{ℓ}_{obs} and λ_ℓ are the observer-frame equivalent width and central wavelength of the line ℓ, respectively. The above equation assumes a constant continuum flux density within the filter, and a constant filter response over the spectral extent of the line. The flux uncertainties were updated to account for the uncertainty associated with this correction, using Monte Carlo simulations where the original flux and all the EWs were randomly perturbed within their respective uncertainties. Because it is based only on the measured equivalent widths, this correction is by construction not affected by systematic errors in flux calibration.

In the end, this had a significant impact only in the K_s band, and only for the galaxies ZF-COS-20133 (30% of the flux
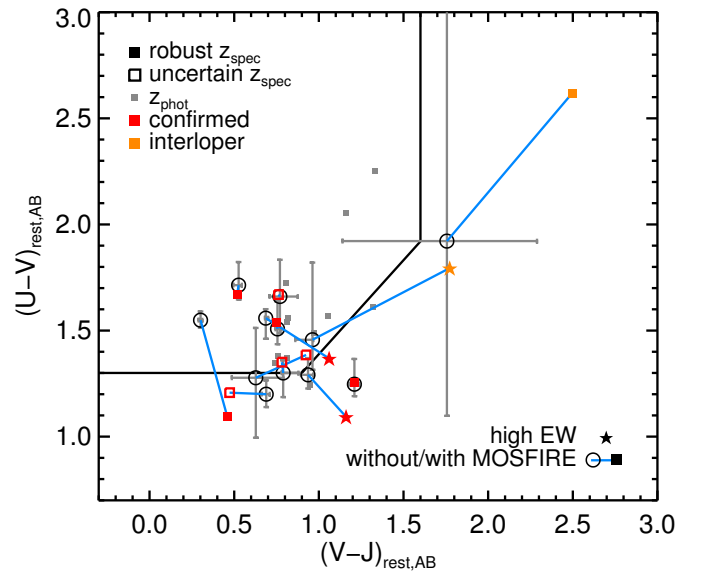
Fig. 11. UVJ colors of our candidate quiescent galaxies targeted with MOSFIRE. The legend is the same as in Fig. 9. The black solid line is the quiescent-or-star-forming dividing line used in S14. For galaxies with no spectroscopic redshift, shown in small gray symbols, the colors shown were computed at $z = z_{phot}$. For galaxies with a spectroscopic redshift, the colors shown were computed at $z = z_{spec}$; the colors initially computed at $z = z_{phot}$ and without the correction for emission lines are shown with empty squares, and blue lines connect the old (empty black circle) and new (colored star or square) colors of a given galaxy. These values are listed in Table 3.

removed), ZF-UDS-8197 (19%) and 3D-UDS-27939 (16%). The fluxes of the other galaxies were affected by less than 5%, and the corrected photometry is displayed in Fig. 3.

3.7. Updated rest-frame colors

Using the photometry corrected for emission lines, as described in the previous section, and assuming $z = z_{spec}$, we then recomputed the UVJ colors with EAzY. We show in Fig. 11 the

change in colors resulting from the knowledge of the spectroscopic redshifts and the line-subtracted photometry.

The most striking change naturally occurs for the two redshift interlopers, which are now clearly located in the "dusty star-forming" region of the UVJ diagram. The reason for this change is different for the two galaxies. For ZF-COS-20032, the observed colors are redder than presently allowed by the SED template set of EAzY, which may explain why its redshift was incorrect in the first place. Indeed we show later in Sect. 4.1 that this galaxy suffers exceptional obscuration by dust, with $A_V \sim 4$ mag, while the dustiest template provided with EAzY has $A_V = 2$ mag; this suggests that such redshift outliers could be avoided if redder templates were included in the z_{phot} determination, but demonstrating this goes beyond the scope of this paper. For 3D-UDS-27939, the rest-frame colors are still within range of the EAzY template set; the main reason for the z_{phot} failure was that the K_s-band flux was significantly contaminated by Hα, mimicking the Balmer break at $z \sim 3$.

Otherwise, the colors also changed significantly for the two confirmed $z > 3$ galaxies with high equivalent-width [O III] emission, namely ZF-COS-20133 and ZF-UDS-8197. These two galaxies are now outside the fiducial UVJ "quiescent" region, although ZF-COS-20133 still lies within 0.05 mag of the dividing line. The situation for these two objects is similar to that of 3D-UDS-27939, in that the apparent strength of the Balmer break was reduced once the emission line was subtracted from the photometry.

For the rest of the sample, the only significant change was for 3D-EGS-18996 which saw its $U-V$ color reduced by about 0.5 mag, moving it outside of the quiescent region. This change was caused by a revision of the redshift, as the z_{phot} was significantly underestimated. Interestingly, this galaxy nevertheless displays strong Balmer absorption and no Hβ or [O II] emission, which demonstrates the absence of current star formation. In fact, this region of the UVJ diagram was shown to be mainly populated by post-starburst galaxies (e.g., Whitaker et al. 2012; Wild et al. 2014). Albeit not satisfying the fiducial UVJ color cut owing to their too recent quenching, such galaxies still host little to no current star-formation activity (e.g, Merlin et al. 2018), and can thus still be considered quiescent.

In the end, out of the 24 galaxies that were observed with MOSFIRE, two turned out to be redshift interlopers, two had bright emission lines contaminating their rest V magnitude, and one saw its redshift sufficiently revised to change its UVJ colors and move it out of the quiescent region (albeit in the post-starburst area). Combined, this implies that 21% of the galaxies initially classified as UVJ-quiescent were spurious. We thus concluded that the initial UVJ colors of our galaxies were robust for the majority of the sample, but that the number of quiescent galaxies estimated at $z > 3$ from UVJ-selected samples is overestimated by 20% because of contaminants. In the following, we take this figure into account to correct the observed number densities.

4. Star formation rates and histories

In this section we take advantage of the knowledge derived from the MOSFIRE spectra, namely the redshifts and absence of strong emission lines, as well as the spectra themselves, to model the star formation histories of our quiescent galaxies candidates. The goal of this section is to investigate if these galaxies are truly quiescent, and if so, to provide the first constraints on their star formation histories.

4.1. Modeling

Using the updated photometry and redshifts, we re-ran FAST++ to update the stellar masses and other physical properties of the quiescent galaxies. In addition, for each galaxy with a measured z_{spec}, we used the MOSFIRE spectrum to constrain the fit further, masking emission lines detected at more than 2σ significance since FAST++ does not model them, and only using spectra for which the synthetic broadband flux (in the H or K passband) was detected at more than 5σ. In the fit, the spectrum was renormalized independently of the broadband photometry to account for mis-corrections of the slit losses and residual aperture systematics (AUTO_SCALE = 1). The spectrum was still included in the χ^2, but the independent rescaling ensured that only the shape of the spectrum constrained the fit (i.e., absorption lines and spectral breaks, or the absence thereof) and not its absolute normalization. Lastly, if a galaxy was detected by ALMA, we computed its L_{IR} using the dust templates from Schreiber et al. (2018a), assuming the average dust temperature at the redshift of each galaxy ($T_{dust}(z)$, Eq. (15) in Schreiber et al. 2018a; e.g, $T_{dust} \sim 40$ K at $z \sim 3.5$), and used this value to constrain the attenuation in FAST++. For ZF-COS-20115 we used the L_{IR} non-detection derived in Schreiber et al. (2018b).

Similarly to Schreiber et al. (2018b), we post-processed each model SFH generated by FAST++ and identified the two main phases of the galaxy's history, as illustrated in Fig. 1. Firstly, we located the time of peak SFR and determined the smallest contiguous time period surrounding it where 68% of the integrated SFR took place. We considered this as the "main" formation phase, and defined its length as T_{SF} and its mean SFR as $\langle SFR \rangle_{main}$. To locate this formation phase in time, we computed the time at which half of the mass had formed, t_{form}, ignoring mass loss. Having identified the main formation phase, we then looked for the longest contiguous time period, starting from the epoch of observation and running backward, where the SFR was less than 10% of $\langle SFR \rangle_{main}$. If such a time period existed, we considered it as the "quenched" phase, and defined its starting time as t_{quench}. Knowing the redshift of the galaxy, we then inferred the formation and quenching redshifts, z_{form} and z_{quench}, respectively.

For each galaxy and for all model parameters, we defined the best-fit values from the model with $\chi^2 = \min(\chi^2)$, and defined the range of allowed values as the range spanned by models with $\chi^2 - \min(\chi^2) < 2.71$. This corresponds to a 90% confidence interval (Avni 1976). The resulting galaxy properties for the entire sample can be found in Table 3. To study these two quantities in more detail, we also computed the probability distribution functions for t_{quench} and t_{form}. We defined this probability on a one-dimensional grid of values t_i with fixed step $\Delta t = 50$ Myr such that $p(t_i) \propto \exp(-\chi_i^2)$, where χ_i^2 is the minimum χ^2 of all models with $|t - t_i| < \Delta t/2$ (where t is either t_{quench} or t_{form}). This same approach was used in CIGALE (Noll et al. 2009; see in particular their Fig. 6 for an illustration).

To check that our modeling provided a good description of the data, we also computed the reduced χ^2 (χ^2_{red}) for each galaxy. For this exercise, we excluded the spectra from the χ^2 since we already showed they have $\chi^2_{red} \sim 1$ (see Table A.4). We find a median χ^2_{red} of 1.13, which indicates an overall good fit to the photometry, however three galaxies have values larger than 2.0 which deserved further inspection. The largest values is $\chi^2_{red} = 2.34$ for ZF-UDS-6496, and is mainly caused by a flux excess in the B band blueward of the Lyman limit (rest $860-1070$ Å). Excluding this band brings the χ^2_{red} down to 1.39. The second highest value is obtained for 3D-EGS-18996, with

Table 3. Final properties of the galaxies observed with MOSFIRE (90% confidence error bars, unless otherwise stated).

ID	z^a	$U-V^a$ rest, AB	$V-J^a$ rest, AB	M_* $10^{11}\,M_\odot$	A_V mag.	SFR_{10} $M_\odot\,\mathrm{yr}^{-1}$	t_{quench} Gyr	z_{form}	t_{SF} Gyr	$\langle\mathrm{SFR}\rangle_{\mathrm{main}}$ $\log(M_\odot\,\mathrm{yr}^{-1})$
Robust z_{spec}										
ZF-COS-20032	2.474	$2.62^{+0.71}_{-1.61}$	$2.50^{+0.05}_{-1.40}$	$1.12^{+0.19}_{-0.47}$	$3.7^{+0.3}_{-0.2}$	139^{+75}_{-139}	$0.28^{+0.17}_{-0.28}$	$2.9^{+0.3}_{-0.3}$	$0.06^{+0.56}_{-0.05}$	$3.26^{+0.58}_{-0.98}$
ZF-COS-20115	3.715	$1.67^{+0.16}_{-0.02}$	$0.52^{+0.03}_{-0.03}$	$1.15^{+0.16}_{-0.09}$	$0.3^{+0.1}_{-0.1}$	$0.0^{+0.7}_{-0.0}$	$0.51^{+0.19}_{-0.24}$	$6.1^{+0.8}_{-0.7}$	$0.15^{+0.58}_{-0.13}$	$2.93^{+0.83}_{-0.66}$
ZF-COS-20133	3.481	$1.37^{+0.32}_{-0.01}$	$1.06^{+0.03}_{-0.01}$	$0.33^{+0.01}_{-0.06}$	$0.1^{+0.1}_{-0.1}$	$4.6^{+0.1}_{-1.3}$	$1.43^{+0.00}_{-0.44}$	$32.2^{+0.0}_{-21.1}$	$0.14^{+0.28}_{-0.02}$	$2.43^{+0.00}_{-0.54}$
3D-EGS-18996	3.239	$1.10^{+0.47}_{-0.01}$	$0.46^{+0.01}_{-0.15}$	$0.98^{+0.04}_{-0.06}$	$0.0^{+0.1}_{-0.0}$	$1.0^{+1.0}_{-0.9}$	$0.33^{+0.09}_{-0.10}$	$4.3^{+0.3}_{-0.1}$	$0.27^{+0.14}_{-0.20}$	$2.58^{+0.53}_{-0.16}$
3D-EGS-26047	3.234	$1.25^{+0.11}_{-0.06}$	$1.21^{+0.03}_{-0.03}$	$0.99^{+0.21}_{-0.14}$	$0.2^{+0.2}_{-0.2}$	$0.1^{+1.9}_{-0.0}$	$0.52^{+0.66}_{-0.42}$	$7.5^{+7.2}_{-1.4}$	$0.42^{+0.53}_{-0.32}$	$2.42^{+0.64}_{-0.34}$
3D-EGS-40032	3.219	$1.54^{+0.04}_{-0.09}$	$0.75^{+0.02}_{-0.01}$	$2.03^{+0.16}_{-0.14}$	$0.4^{+0.1}_{-0.1}$	$6.1^{+3.7}_{-3.4}$	$0.51^{+0.14}_{-0.20}$	$5.0^{+1.3}_{-0.4}$	$0.23^{+0.76}_{-0.17}$	$2.99^{+0.60}_{-0.61}$
ZF-UDS-8197	3.543	$1.09^{+0.25}_{-0.01}$	$1.16^{+0.02}_{-0.18}$	$0.36^{+0.04}_{-0.04}$	$0.0^{+0.3}_{-0.0}$	$1.7^{+3.3}_{-0.7}$	$0.43^{+0.39}_{-0.40}$	$6.7^{+3.0}_{-1.5}$	$0.69^{+0.21}_{-0.64}$	$1.76^{+1.03}_{-0.11}$
3D-UDS-27939	2.210	$1.79^{+1.27}_{-1.20}$	$1.77^{+0.87}_{-0.81}$	$0.46^{+0.05}_{-0.17}$	$1.0^{+0.6}_{-0.2}$	$1.8^{+8.0}_{-1.2}$	$2.36^{+0.00}_{-2.36}$	$23.6^{+0.0}_{-20.1}$	$0.23^{+1.14}_{-0.10}$	$2.39^{+0.17}_{-0.90}$
Uncertain z_{spec}										
ZF-COS-17779	3.415	$1.39^{+0.13}_{-0.45}$	$0.92^{+0.01}_{-0.43}$	$0.36^{+0.12}_{-0.16}$	$1.4^{+0.6}_{-0.8}$	$0.7^{+100}_{-0.7}$	$0.03^{+0.29}_{-0.03}$	$4.4^{+5.0}_{-0.9}$	$0.85^{+0.66}_{-0.84}$	$1.65^{+1.76}_{-0.21}$
ZF-COS-18842	3.782	$1.21^{+0.19}_{-0.17}$	$0.47^{+0.26}_{-0.01}$	$0.45^{+0.06}_{-0.04}$	$0.0^{+0.2}_{-0.0}$	$2.5^{+2.4}_{-1.2}$	$0.33^{+0.16}_{-0.12}$	$5.3^{+1.1}_{-0.3}$	$0.27^{+0.50}_{-0.22}$	$2.24^{+0.79}_{-0.40}$
ZF-COS-19589	3.715	$1.67^{+0.16}_{-0.18}$	$0.76^{+0.12}_{-0.06}$	$0.62^{+0.10}_{-0.08}$	$0.9^{+0.4}_{-0.5}$	$0.0^{+4.4}_{-0.0}$	$0.38^{+0.39}_{-0.32}$	$5.1^{+2.9}_{-0.7}$	$0.09^{+1.12}_{-0.08}$	$2.87^{+0.86}_{-1.11}$
3D-EGS-31322	3.434	$1.35^{+0.02}_{-0.16}$	$0.78^{+0.10}_{-0.01}$	$0.98^{+0.12}_{-0.08}$	$0.3^{+0.2}_{-0.2}$	$0.0^{+2.3}_{-0.0}$	$0.28^{+0.25}_{-0.20}$	$4.9^{+1.6}_{-0.4}$	$0.28^{+0.87}_{-0.24}$	$2.57^{+0.83}_{-0.58}$
z_{phot}										
ZF-COS-10559	$3.34^{+0.30}_{-1.04}$	$1.57^{+0.75}_{-0.88}$	$1.05^{+0.89}_{-0.32}$	$0.23^{+0.08}_{-0.05}$	$0.0^{+0.5}_{-0.0}$	$0.9^{+2.2}_{-0.5}$	$1.06^{+0.46}_{-0.98}$	$9.0^{+23.3}_{-3.6}$	$0.16^{+0.74}_{-0.07}$	$2.22^{+0.29}_{-0.75}$
ZF-COS-14907	$2.89^{+0.06}_{-0.06}$	$1.38^{+0.08}_{-0.07}$	$0.76^{+0.04}_{-0.03}$	$0.49^{+0.06}_{-0.05}$	$0.2^{+0.3}_{-0.2}$	$2.4^{+3.2}_{-1.8}$	$0.36^{+0.24}_{-0.29}$	$4.3^{+2.0}_{-0.6}$	$0.68^{+0.55}_{-0.63}$	$1.90^{+1.13}_{-0.25}$
3D-EGS-27584	$3.60^{+0.18}_{-0.23}$	$2.25^{+0.29}_{-0.23}$	$1.33^{+0.16}_{-0.14}$	$4.59^{+0.82}_{-1.02}$	$1.3^{+0.7}_{-0.3}$	$11.5^{+29.7}_{-11.5}$	$0.82^{+0.61}_{-0.76}$	$7.1^{+30.7}_{-2.7}$	$0.09^{+1.23}_{-0.07}$	$3.73^{+0.60}_{-1.16}$
3D-EGS-34322	$3.59^{+0.33}_{-0.32}$	$1.25^{+0.51}_{-0.16}$	$0.95^{+0.05}_{-0.23}$	$0.42^{+0.15}_{-0.08}$	$0.9^{+0.7}_{-0.8}$	$2.6^{+25.0}_{-2.6}$	$0.21^{+0.35}_{-0.21}$	$4.3^{+8.5}_{-0.6}$	$0.09^{+1.28}_{-0.08}$	$2.67^{+0.92}_{-1.09}$
ZF-UDS-3651	$3.87^{+0.12}_{-0.12}$	$1.35^{+0.11}_{-0.25}$	$0.74^{+0.01}_{-0.01}$	$0.76^{+0.12}_{-0.09}$	$0.7^{+0.2}_{-0.5}$	$14.8^{+25.1}_{-14.7}$	$0.31^{+0.21}_{-0.31}$	$5.0^{+3.9}_{-0.4}$	$0.13^{+0.93}_{-0.09}$	$2.79^{+0.50}_{-0.89}$
ZF-UDS-4347	$3.58^{+0.04}_{-0.05}$	$1.54^{+0.06}_{-0.09}$	$0.81^{+0.08}_{-0.04}$	$0.32^{+0.06}_{-0.04}$	$0.4^{+0.2}_{-0.2}$	$1.3^{+3.7}_{-1.3}$	$0.30^{+0.12}_{-0.24}$	$4.4^{+0.8}_{-0.2}$	$0.06^{+0.75}_{-0.05}$	$2.75^{+0.71}_{-1.09}$
ZF-UDS-6496	$3.50^{+0.04}_{-0.04}$	$1.73^{+0.04}_{-0.02}$	$0.81^{+0.01}_{-0.02}$	$0.89^{+0.07}_{-0.10}$	$0.3^{+0.2}_{-0.1}$	$8.4^{+6.2}_{-2.7}$	$0.34^{+0.22}_{-0.34}$	$6.5^{+0.6}_{-1.5}$	$0.81^{+0.10}_{-0.73}$	$2.08^{+0.99}_{-0.08}$
ZF-UDS-7329	$3.04^{+0.17}_{-0.17}$	$2.05^{+0.28}_{-0.32}$	$1.16^{+0.13}_{-0.10}$	$1.82^{+0.14}_{-0.55}$	$0.5^{+0.3}_{-0.1}$	$0.9^{+3.3}_{-0.4}$	$1.51^{+0.22}_{-1.21}$	$15.0^{+17.2}_{-9.5}$	$0.27^{+0.54}_{-0.19}$	$2.90^{+0.36}_{-0.53}$
ZF-UDS-7542	$3.15^{+0.06}_{-0.06}$	$1.49^{+0.08}_{-0.05}$	$0.97^{+0.02}_{-0.01}$	$0.55^{+0.14}_{-0.04}$	$1.1^{+0.2}_{-0.5}$	$0.0^{+4.7}_{-0.0}$	$0.28^{+0.09}_{-0.23}$	$3.7^{+1.9}_{-0.1}$	$0.02^{+1.23}_{-0.01}$	$3.43^{+0.29}_{-1.69}$
3D-UDS-35168	$3.46^{+0.32}_{-0.29}$	$1.37^{+0.35}_{-0.28}$	$0.81^{+0.11}_{-0.12}$	$0.34^{+0.10}_{-0.11}$	$0.3^{+1.1}_{-0.3}$	$0.9^{+24.6}_{-0.9}$	$0.10^{+1.41}_{-0.10}$	$9.6^{+27.4}_{-5.7}$	$0.73^{+0.77}_{-0.71}$	$1.72^{+1.56}_{-0.37}$
3D-UDS-39102	$3.51^{+0.42}_{-0.36}$	$1.61^{+0.51}_{-0.50}$	$1.32^{+0.11}_{-0.24}$	$0.89^{+0.44}_{-0.22}$	$1.9^{+0.4}_{-0.8}$	$0.0^{+100}_{-0.0}$	$0.11^{+0.36}_{-0.11}$	$3.8^{+6.2}_{-0.2}$	$0.07^{+1.39}_{-0.06}$	$3.11^{+0.85}_{-1.20}$
3D-UDS-41232	$3.01^{+0.07}_{-0.08}$	$1.56^{+0.16}_{-0.12}$	$0.82^{+0.04}_{-0.03}$	$1.50^{+0.16}_{-0.24}$	$0.3^{+0.3}_{-0.2}$	$0.9^{+2.9}_{-0.9}$	$0.32^{+0.54}_{-0.23}$	$5.4^{+3.0}_{-1.4}$	$0.98^{+0.18}_{-0.92}$	$2.24^{+1.20}_{-0.08}$

Notes. This table is also available in electronic format at the CDS. [a] 68% confidence intervals.

$\chi^2_{\mathrm{red}} = 2.28$, and this is caused by the WirCAM J band, which is inconsistent with the well-measured HST F125W flux at the 2.4σ level. Without this band, we find $\chi^2_{\mathrm{red}} = 1.46$. The last case is ZF-COS-10559, the faintest galaxy of our sample, with $\chi^2_{\mathrm{red}} = 2.22$. There we could not find a single band causing the poor χ^2, rather a number of bands with inconsistent fluxes (see Fig. 3). The only trend we found was for the CFHT photometry to be lower than that from Subaru (the CFHT G band in particular is lower than the overlapping Subaru B and V bands at 3.8 and 4.2σ, respectively). This may indicate variability. All the other galaxies have $\chi^2_{\mathrm{red}} < 1.5$, indicating that our models were able to capture all the significant features of the observed photometry.

4.2. Impact of star-formation history parametrization

As discussed in Sect. 2.2, in this work we considered a more involved SFH parametrization than the traditional models (e.g., delayed exponentially declining, or constant truncated SFH). In this section we discuss the impact of adding an additional degree of freedom regarding the recent SFR Eq. (2). For this purpose, we ran FAST++ again with a simplified SFH where we forced $R_{\mathrm{SFR}} = 1$, and therefore where this freedom was removed.

Compared to the fits with $R_{\mathrm{SFR}} = 1$, the best-fit values obtained with the SFH of Eq. (2) are mostly the same except for one galaxy, ZF-COS-20133, which is discussed below. For the rest of the sample, the stellar masses have a median ratio of unity and a scatter of only 0.07 dex. The quiescence and formation times have a scatter of 115 and 175 Myr, respectively. The most important difference is found for the 10 Myr-averaged SFRs, which are increased on average by 35%. Still, for 80% of the sample these changes are contained within the error bars, and are thus not significant.

The case of ZF-COS-20133 clearly stands apart from the rest of the sample. As illustrated in Fig. 12, this galaxy is the only one for which setting R_{SFR} free provided a significant improvement of the fit ($\Delta\chi^2 = 39$). With $R_{\mathrm{SFR}} = 1$, the adopted models consisted of an SFR that slowly declined since the Big Bang ($t_{\mathrm{SF}} \sim 600\,\mathrm{Myr}$), with a high formation redshift ($z_{\mathrm{form}} = 11.1-12.7$) and a low current SFR ($2.0-2.4\,M_\odot\,\mathrm{yr}^{-1}$). With R_{SFR} free, the main formation episode of the galaxy was shortened ($t_{\mathrm{SF}} < 400\,\mathrm{Myr}$) and pushed to even higher redshifts, and a recent short burst was used to reproduce the current SFR ($3.2-4.7\,M_\odot\,\mathrm{yr}^{-1}$). The galaxy was thus modeled as a maximally old stellar population with a small rejuvenation event.
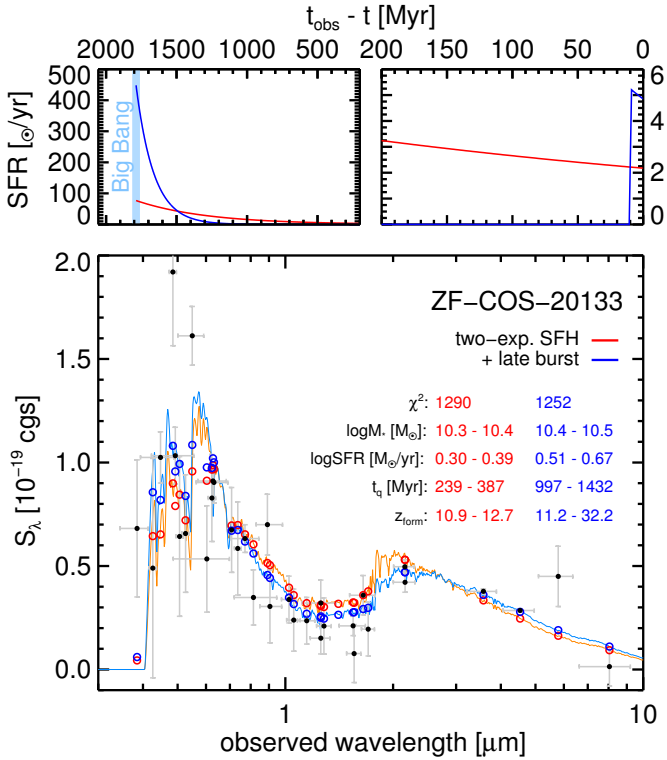
Fig. 12. Photometry and models of ZF-COS-20133. The observed photometry is shown as black circles with gray error bars. Our model spectrum and photometry using the fiducial SFH parametrization Eq. (2) is shown in blue, and the model without the additional late burst Eq. (1) is shown in red/orange. The fit parameters are given in inset. At the *top* of the figure, we show the star formation histories of both models. The time axis was split in half to better display the behavior during the last 200 Myr.

Such a large age would imply an extreme star-formation efficiency within the first few hundred million years after the Big Bang. However, this galaxy is one of the few for which we observed strong [O III] emission ($\log([O\,III]/H\beta) = 0.84$). This line is unresolved, which suggests it may originate from the narrow-line region of an AGN. Since it also has a remarkably high EW of ~300 Å, and since the [O III] EW is known to correlate with AGN obscuration (e.g., Caccianiga & Severgnini 2011), an alternative and more plausible scenario is that the photometry contains some continuum emission from an obscured AGN (see, e.g., Marsan et al. 2017). This would impact particularly the IRAC bands, increasing the flux there and thus faking the presence of an old population (the galaxy is not detected at 24 μm however, so this putative AGN cannot be very luminous). If this is true, then the SFH of ZF-COS-20133 cannot be constrained without accurately accounting for the presence of the AGN, which we cannot do here. In the following, we therefore do not use the inferred SFH for this galaxy.

4.3. Comparison of SFR estimates

While we have explored a large variety of star formation histories, our SFR estimates may still be biased low if we underestimated the attenuation by dust, since red colors can be produced both by old stellar populations and dust obscuration (e.g., Dunlop et al. 2007). The few galaxies in our sample covered by ALMA show no detection in the sub-millimeter (save for the
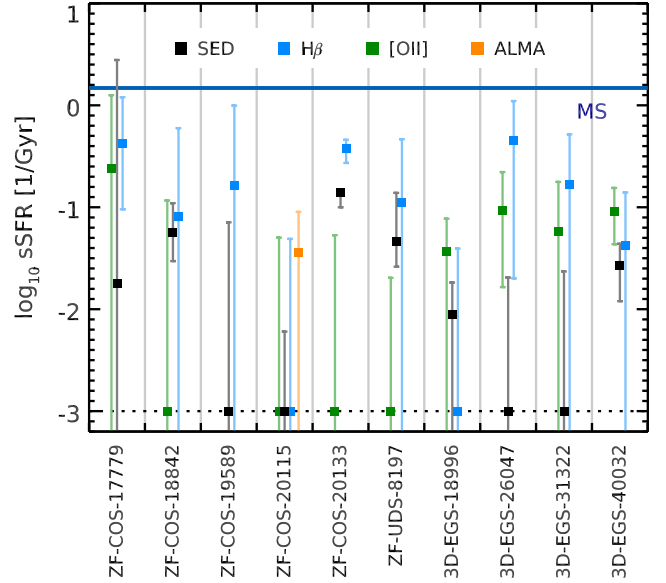

Fig. 13. Comparison of sSFR estimates from several sources: SED modeling with FAST++ (black), $H\beta$ luminosity (blue), [O II] luminosity (green), and, when available, L_{IR} from ALMA (orange). The error bars for all quantities indicate the 90% confidence interval. The line luminosities were corrected for dust attenuation using the A_V from the SED modeling. Each galaxy with a spectroscopic redshift at $z > 3$ is shown in a different column. For visualization purposes we limit the minimum sSFR to 10^{-3} Gyr^{-1}, which is indicated with the dotted line; galaxies on this line are actually at lower sSFR. The blue horizontal line gives the average sSFR of main-sequence galaxies (see Sect. 4.4).

two redshift outliers). Stacking the ALMA-derived L_{IR} for these galaxies leads to SFR $= 11 \pm 17\,M_\odot$ yr^{-1}, while the deep upper limit for ZF-COS-20115 is SFR $< 13\,M_\odot$ yr^{-1} (Schreiber et al. 2018b). The non-detections with ALMA therefore rules out strong starbursts, but still allows for moderate amounts of star formation.

To double check our SFRs, we therefore obtained alternative estimates using the measured emission line luminosities (Table A.5), empirical conversion factors from the literature, and the dust attenuation estimated by FAST++ from the broadband photometry. While it is usually assumed that emission lines suffer more attenuation than the stellar continuum (e.g., Calzetti et al. 2000), recent results suggest this reddening excess becomes negligible at high redshifts (e.g., Pannella et al. 2015; Reddy et al. 2015). Here we therefore assume the same A_V for the lines as for the continuum. Even though the assumed reddening is the same as for the SED-based SFRs, these line-based estimates are still independent. Indeed, the emission lines are located at visible rather than far-UV wavelengths, and are therefore affected by dust differently than the continuum flux of young OB stars (which drives the SED-based SFR estimates). The comparison can thus allow us to reveal systematic issues with the dust correction (with the caveat that part of the star-forming regions may be optically thick; this can only be tackled with deep FIR imaging). The results of this analysis are summarized in Table A.6 and Fig. 13.

Given the wavelength coverage of our MOSFIRE observations, the best emission line at our disposal for SFR estimates is $H\beta$, which is covered for all our spectroscopically-confirmed galaxies. We recall that our line flux measurement procedure automatically corrects for the underlying stellar absorption, see Sect. 3.1. To translate the $H\beta$ luminosities to SFR,

we assumed case-B recombination ($L_{H\alpha}/L_{H\beta} = 2.86$) and the Kennicutt (1998) conversion factor (converted to a Chabrier IMF):

$$\mathrm{SFR}_{H\beta}^{\mathrm{nodust}}/L_{H\beta} = 5.46 \times 10^{-8} \, M_\odot \, \mathrm{yr}^{-1} \, L_\odot^{-1}. \qquad (8)$$

Using this relation and without dust correction, the 1σ SFR sensitivity from the MOSFIRE spectra is of the order of $1-3 \, M_\odot \, \mathrm{yr}^{-1}$ (see Table A.6). We also considered SFRs estimated from the [O II] line luminosity, which is covered in all but two galaxies. We used the Kewley et al. (2004) calibration (converted to a Chabrier IMF):

$$\mathrm{SFR}_{[\mathrm{O\,II}]}^{\mathrm{nodust}}/L_{[\mathrm{O\,II}]} = 1.59 \times 10^{-8} \, M_\odot \, \mathrm{yr}^{-1} \, L_\odot^{-1}. \qquad (9)$$

Without correction for dust, the [O II] emission in our MOS-FIRE spectra provided comparable SFR sensitivity to Hβ. We considered the [O II]-based SFRs less reliable than that based on Hβ because the physical connection between star-formation and [O II] emission is less immediate. In fact, given that our galaxies have (a priori) low sSFR, both emission lines may be significantly contaminated by energy sources other than star-formation; an AGN, or low-ionization emission from old/intermediate-age stars (see, e.g., Cid Fernandes et al. 2010, 2011 for Balmer lines, and Yan et al. 2006 or Lemaux et al. 2010 for [O II]). This will tend to bias the line-based SFRs toward higher values.

From the broadband modeling (Table 3), we find only modest attenuation in the confirmed $z > 3$ galaxies, with $A_V < 0.5$ for all but two galaxies. At the wavelength of Hβ, and assuming the Calzetti et al. (2000) attenuation curve, this implies attenuation factors of at most 1.7 for Hβ, and 2.1 for [O II]. In all that follows, we considered the entire range of allowed A_V for each galaxy to correct the line SFRs for attenuation.

As illustrated in Fig. 13, in most cases (8/10) the Hβ-based SFRs are consistent with the range of allowed 10 Myr-averaged SFR derived using FAST++. Only one galaxy is clearly discrepant, with larger Hβ-based SFRs than expected from the SED: ZF-COS-20133. This is probably caused by it hosting an AGN (see Sect. 4.2), so we did not consider it further.

The [O II]-based SFRs are consistent with the FAST++ estimates in most cases (7/9), and the most striking inconsistency is again for ZF-COS-20133, which we thus ignored. The other inconsistent galaxy is for ZF-UDS-8197, for which the [O II]-based SFR is actually lower than that inferred from the SED modeling. In fact, overall we find the [O II]-based SFRs are systematically lower than that derived from Hβ. This could imply that we underestimated the attenuation by dust, and that star-forming regions are more attenuated than the older stars. Before applying any dust correction, the median $\mathrm{SFR}_{[\mathrm{O\,II}]}^{\mathrm{nodust}}/\mathrm{SFR}_{H\beta}^{\mathrm{nodust}}$ is 0.3, and reproducing this ratio with the Calzetti et al. (2000) attenuation would require an average $A_V \sim 3$ mag, which is substantially larger than the $A_V < 0.5$ we obtain from the SED modeling. However, with $A_V \sim 3$ the observed Hβ luminosities would then translate to a median SFR $= 100 \, M_\odot \, \mathrm{yr}^{-1}$, and violate the upper limit from ALMA ($< 62 \, M_\odot \, \mathrm{yr}^{-1}$ at 3σ) or *Herschel* ($< 85 \, M_\odot \, \mathrm{yr}^{-1}$ at 3σ, Straatman et al. 2014).

Therefore, either the $z \sim 0$ calibration of the $L_{[\mathrm{O\,II}]}$-to-SFR conversion factor does not apply to $z > 3$, or the luminosities of these lines is not related to star formation. Providing a definitive answer to this question would require observing the Hα line to infer the Balmer decrement, which will only be possible with the upcoming *James Webb* Space Telescope (JWST), although given the expected change in gas-phase metallicity and ionization parameters in high redshift galaxies (e.g.,

Erb et al. 2010; Steidel et al. 2016; Shapley et al. 2017), an evolution of the [O II] SFR calibration would not be surprising. In any case, the fact that the Hβ-based SFRs agree with our SED-based estimates suggests that our SED modeling is not affected by significant biases regarding the dust attenuation, and that the derived star-formation histories should thus be reliable.

4.4. Specific star-formation rates

We show in Fig. 14 the distribution of the specific SFR (sSFR) of our $z > 3$ quiescent galaxies, and compare it to that of "normal" main-sequence galaxies. Here and in all that follows, we excluded the two $z < 3$ interlopers and ZF-COS-20133 from the quiescent sample. We defined the locus of the main sequence as the average sSFR of the ZFOURGE galaxies at $3 < z_{\mathrm{phot}} < 4$ and $3 \times 10^9 < M_*/M_\odot < 10^{10}$, using the values derived in the present paper with the same SED modeling as for the quiescent galaxies (Sect. 2.2). We find $\mathrm{sSFR}_{\mathrm{MS}} = 1.5 \, \mathrm{Gyr}^{-1}$, which is comparable to dust-based estimates (e.g., Schreiber et al. 2017). We chose to use the same approach to derive SFRs for both the main-sequence and quiescent galaxies (i.e., rather than using stacked far-IR measurements to define the main sequence) in order to make the comparison between the two populations as straightforward as possible.

Looking only at the best-fitting models to begin with, we find our quiescent candidates are located more than a factor of ten below the $z \sim 3.5$ main sequence, with the exception of ZF-UDS-3651 which is a factor of seven below. When using the upper limits on the sSFR, only five galaxies (24%) can actually be located within a factor of three of the main sequence, and thus not be truly quiescent; these are the galaxies ZF-COS-17779, ZF-UDS-3651, 3D-UDS-35168, 3D-UDS-39102, and 3D-EGS-34322. ZF-COS-17779, in particular, is the only one which may be located on or above the main sequence. The vast majority of the MOSFIRE sample thus remains at least a factor of ten below the main sequence, within the uncertainties, which confirms that these galaxies must be genuinely quiescent.

In the absolute sense, although the recovered sSFR are low, the best-fit values still span a range up to $\mathrm{sSFR} = 0.1 \, \mathrm{Gyr}^{-1}$ (and ZF-UDS-3651 at $0.2 \, \mathrm{Gyr}^{-1}$). In particular, 57% of our galaxies are above the $\mathrm{sSFR} = 0.01 \, \mathrm{Gyr}^{-1}$ threshold used in other studies to isolate "red-and-dead" galaxies (shown in Fig. 14; see, e.g., Fontana et al. 2009; Merlin et al. 2018), and only one has an upper limit below this threshold (ZF-COS-20115). This shows that the UVJ selection does not only select galaxies with zero on-going star-formation, but can also include galaxies with low residual SFR (see also the discussion in Merlin et al. 2018). While here we consider this more of a feature than a fundamental flaw (these galaxies are still an order of magnitude below the main sequence), it is clearly an important distinction to keep in mind when comparing observed galaxy number counts to models, which we do in Sect. 5.

4.5. Inferred star-formation histories

In this section we discuss the inferred star-formation histories of the quiescent galaxies. In Fig. 15, we show the values for z_{form} – the point in time when half of the mass had formed, and z_{quench} – the point in time when the galaxy's SFR dropped below 10% of their respective past average (see Sect. 4.1 for the precise definitions of these quantities). The values are listed in Table 3. In addition, we show in Fig. 16 the probability distribution functions (PDFs) of t_{form} and t_{quench}. We show
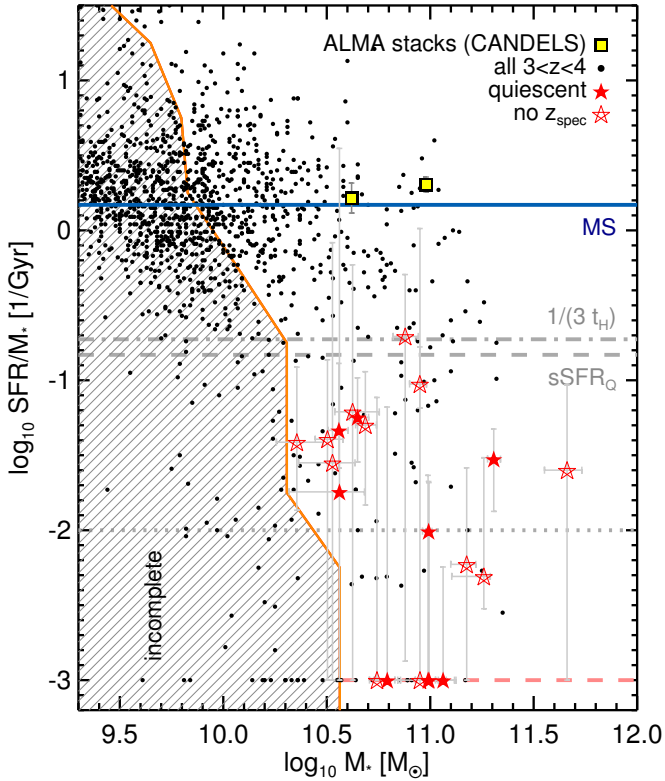
Fig. 14. Best fit specific SFR (sSFR) as a function of stellar mass for galaxies at $3 < z < 4$, as derived from their UV-to-NIR SEDs. The whole ZFOURGE sample (in COSMOS and UDS) is shown as small dots, and the $z > 3$ quiescent galaxies observed with MOSFIRE are shown as red stars with gray error bars. Filled and open symbols refer to galaxies with and without a z_{spec}, respectively (values are listed in Table 3). The locus of the main sequence for the ZFOURGE galaxies is shown with a blue line. Different definitions for "quiescence" are shown in gray: the one we adopted in this paper, sSFR $=$ sSFR$_Q = 0.15$ Gyr^{-1} (dashed), sSFR $= 1/(3 t_H)$ (dot-dashed, t_H being the age of the Universe), and sSFR $= 0.01$ Gyr^{-1} (dotted). The sSFR measured in stacks of ALMA imaging at $z \sim 4$ is shown with yellow squares (Schreiber et al. 2017), and was corrected down by a factor 1.3 to account for redshift evolution of the sSFR between $z = 3.5$ and $z = 4$. As in Fig. 13, for display purposes we place a limit on the minimum sSFR of 10^{-3} Gyr^{-1}, which is indicated with the dashed pink line. The hashed region on the bottom *left* shows the region of this parameter space where the ZFOURGE catalogs are incomplete because of the K_s magnitude limit.

the mean of all PDFs, to illustrate of how well the quantity is constrained for individual objects, as well as the sample's average value. We note that, because of outshining from the youngest stars, our modeling can tend to underestimate the mass of old stellar populations, and the actual age of formation and quenching (see, e.g., Papovich et al. 2001 and the discussion in the appendix of Schreiber et al. 2018b). In this context, our estimates of z_{quench} and z_{form} could be considered as lower limits. These values are later compared to models in Sect. 6.3.

4.5.1. When did star-formation stop?

All the galaxies in our sample have best-fit solutions which have quenched (as per our adopted definition for t_{quench}). Quantitatively, the 16 and 84th percentiles of t_{quench} are 210 and 510 Myr, 71% of the galaxies have an upper bound on t_{quench} larger than 500 Myr, and $t_{quench} = 0$ is ruled for 76% of the sample. Stack-
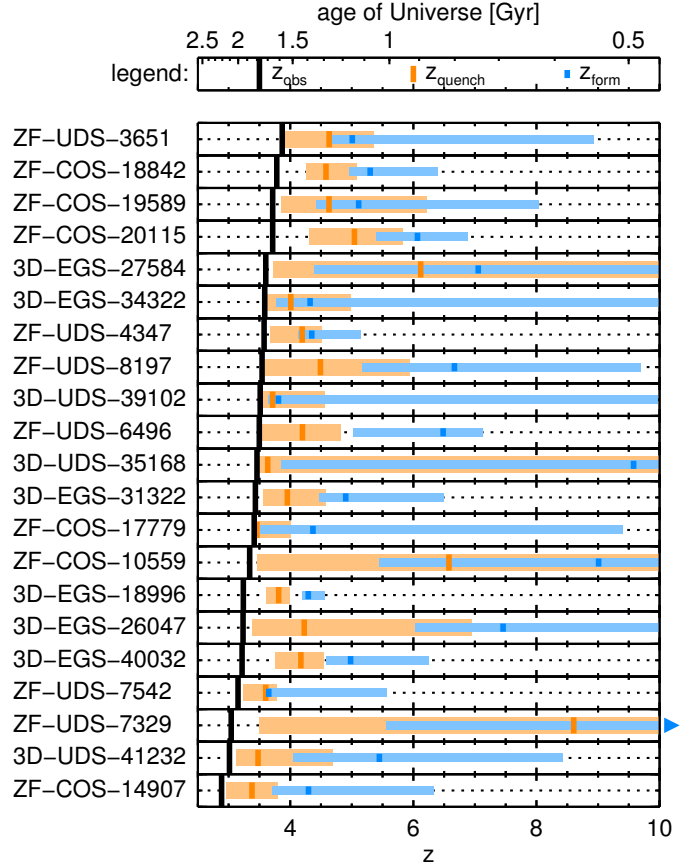

Fig. 15. Summarized star-formation histories of our $z > 3$ quiescent galaxies. Galaxies are sorted by descending redshift, and are each displayed on a separate line. The black vertical bar indicates the redshift at which the galaxy is observed. The orange and blue bands show the 90% confidence range for the quenching and formation redshifts, respectively. Bars of darker colors indicate the corresponding best fit values. These values are listed in Table 3.

ing the probability distributions, as shown in Fig. 16, we find that the average t_{quench} is 330 ± 30 Myr. In other words, 76% of our quiescent galaxies have quenched with certainty, on average ~ 330 Myr before being observed. For reference, at $z = 3, 3.5$, and 4, $t_{quench} = 330$ Myr implies a quenching at $z_{quench} = 3.5, 4.2$ and 4.9, respectively.

Looking at individual objects, $t_{quench} < 200$ Myr is excluded for five of our quiescent galaxies (25%). These are ZF-COS-18842, ZF-COS-20115, ZF-UDS-7329, 3D-EGS-18996, and 3D-EGS-40032 (all but ZF-UDS-7329 are spectroscopically confirmed). ZF-COS-20115 has the highest redshift of these five galaxies, and its case was already discussed at length in our previous works (Glazebrook et al. 2017; Schreiber et al. 2018b). Here we confirm once more that the galaxy must have quenched between 270 and 700 Myr prior to observation ($4.3 < z_{quench} < 5.8$).

Otherwise, ZF-UDS-7329, albeit not spectroscopically confirmed, is particularly interesting: its photometry requires a quenching as recently as 300 Myr prior to observation ($z_{quench} > 3.5$), but it also allows much older solutions up to 1.7 Gyr prior to observation, equivalent to $z_{quench} = 12$, while the other galaxies are limited to $z_{quench} < 6$. The reason why is apparent in Fig. 3, where it is clear that ZF-UDS-7329 has the SED least resembling that of an A star (as opposed, e.g., to ZF-COS-20115). An older stellar population may also explain why no significant absorption line was found in its deep K-band spectrum
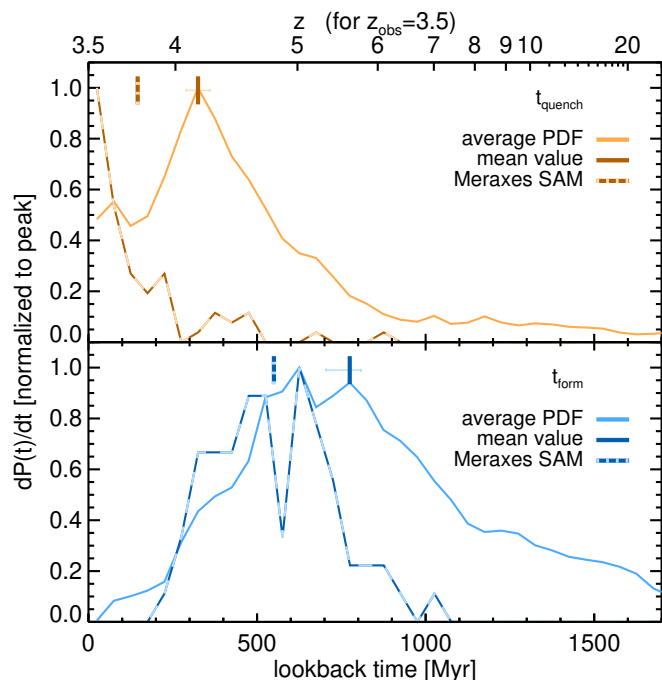
Fig. 16. Probability distributions functions (PDF) of the quenching (*top*) and formation (*bottom*) lookback times for our $z > 3$ quiescent galaxies observed with MOSFIRE. For each quantity we show the average of the individual PDFs with a solid line, the mean of the population with a darker vertical line (error bars are the error on the mean), as well as the distribution of values measured in a $z = 3.5$ snapshot of the MERAXES semi-analytic model (SAM) with a dashed line (the dashed vertical bar indicates the mean value for the MERAXES galaxies).

(at its z_{phot}, we would expect to see only Hβ). If it indeed stopped forming stars at such extremely high redshifts, this galaxy may prove difficult to explain in our current understanding of cosmology (e.g., Behroozi & Silk 2018). Follow up observations are under way to confirm its redshift and constrain further its star formation history.

Two other galaxies in our sample may have had such an early quenching, ZF-COS-10559 and 3D-EGS-27584. However, their photometry also allows a more recent quenching, as late as 70 Myr prior to observation. While ZF-COS-10559 is simply too faint to offer reliable constraints on the quenching, 3D-EGS-27584 has a bright and clean SED. The main source of uncertainty for this galaxy lies in dust obscuration, which could be as high as $A_V = 2$, in which case the quenching might be quite recent, while the earliest quenching scenarios correspond to the fits with $A_V < 1.2$. Deep sub-mm imaging could break this degeneracy.

On the other hand, five galaxies have a lower bound of $t_{quench} = 0$ Myr, that is, their photometry is also compatible with SFHs that have not fully quenched. These are ZF-COS-17779, ZF-UDS-3651, 3D-UDS-35168, 3D-UDS-39102, and 3D-EGS-34322. All of them still have best-fit solutions which have quenched however, and the possibility of non-quenched solutions can be partly explained by their simply being fainter (median K magnitude fainter than the rest of the sample by 0.5 mag), so their photometry is less constraining. This is apparent in Fig. 3. In fact, only one of these has a spectroscopic redshift. Based on the brighter galaxies, we expect in most cases that deeper NIR spectra or photometry would lift this ambiguity and eventually rule out these non-quenched solutions.

4.5.2. When and how did these galaxies form?

From Fig. 15, the formation epoch of these galaxies appears less well constrained, but this is partly an effect of the non-linearity of the redshift. The 16 and 84th percentiles of t_{form} are 360 and 1070 Myr, 71% of the galaxies have an upper bound on t_{form} larger than 1 Gyr, and values less than 400 Myr are excluded for 57% of the sample. From the stacked probability distributions, we find an average of $t_{form} = 780^{+30}_{-70}$ Myr. At $z = 3$, 3.5, and 4, $t_{form} = 780$ Myr implies a formation at $z_{form} = 4.4$, 5.6 and 7.1, respectively.

During this main formation epoch, the average SFR of the galaxies must have been high: the 16 and 84th percentiles of $\langle SFR \rangle_{main}$ are 80 and 850 $M_\odot$ yr^{-1}, and the stacked probability distributions lead to an average of 340^{+150}_{-30} $M_\odot$ yr^{-1}. We recall that this value is the average SFR during the formation phase; the peak SFR would be even higher. The duration of this phase, T_{SF}, has a 16 and 84th percentile range of 90−700 Myr, with an average of 280^{+40}_{-134} Myr. This implies that substantial star-formation must have occurred at $z \sim 5$ in brief and intense episodes in order to form the bulk of the $3 < z < 4$ quiescent population, and supports the results of G17.

Considering individual galaxies, $z_{form} < 5$ is excluded for six galaxies (28%), namely ZF-COS-10559, ZF-COS-20115, ZF-UDS-6496, ZF-UDS-8197, ZF-UDS-7329, and 3D-EGS-26047. The latter is the only galaxy with a lower limit of $z_{form} > 6$, while ZF-COS-20115 has the highest implied past SFR, with a lower limit of $\langle SFR \rangle_{main} > 190$ $M_\odot$ yr^{-1}.

5. Number density

In the previous sections, we have demonstrated that the UVJ colors can be used to select genuinely quiescent galaxies at $z > 3$ with a high purity (80%). Using this knowledge, we came back to the full ZFOURGE sample and revised the observed number density of quiescent galaxies. This is discussed in the following subsections.

By coming back to the entire ZFOURGE catalogs, rather than only working with those galaxies observed with MOSFIRE, we ensured that the derived number densities are unaffected by the non-trivial selection function of our masks (Sect. 2.3). Therefore the only relevant factor limiting our completeness is the K-band flux, which we already addressed in Sect. 2.2.

5.1. Number density of UVJ-quiescent galaxies

Combined, the three ZFOURGE fields contain 130 galaxies at $3 < z < 4$ and $M_* > 3 \times 10^{10}$ $M_\odot$, where our catalogs are complete (see Sect. 2.2). The UVJ colors classify 23 galaxies as quiescent, eight of which were observed with MOSFIRE and confirmed to be robust candidates from the absence of strong emission lines. For the quiescent galaxies without MOSFIRE coverage, following our results from Sect. 3.7 we assumed that 20% are not truly quiescent (i.e., they are either dusty interlopers or high equivalent-width [O III] emitters). This lead to a statistically corrected number of 20 galaxies, corresponding to a number density of $(1.4 \pm 0.3) \times 10^{-5}$ Mpc^{-3} and a stellar mass density of $(1.0 \pm 0.2) \times 10^6$ $M_\odot$ Mpc^{-3} (error bars are only Poisson noise). Following Moster et al. (2011), we estimate the amplitude of cosmic variance on these numbers to be 22%. This number density is lower by 30% compared to the value first reported in Straatman et al. (2014), even though their sample was built at slightly higher redshift and masses ($3.4 < z < 4.2$ and $M_* > 4 \times 10^{10}$ $M_\odot$). This is mostly the result of the
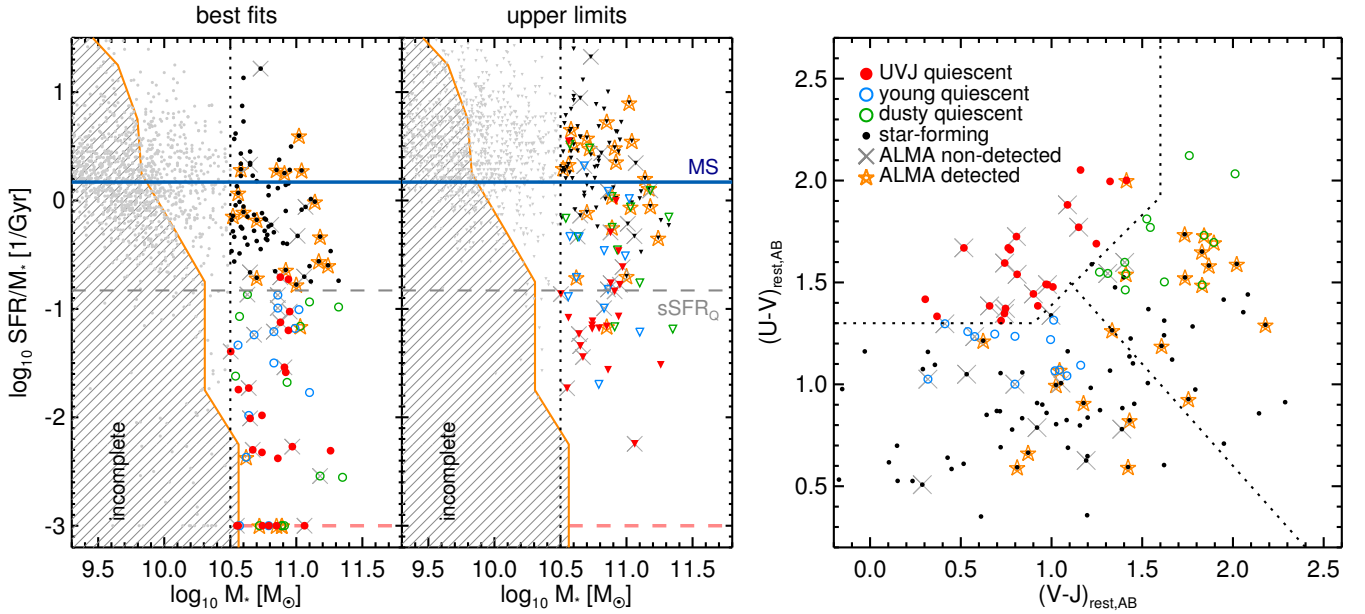
Fig. 17. Comparison of sSFR and *UVJ* classification for the $3 < z < 4$ galaxies with $M_* > 3 \times 10^{10} M_\odot$ in ZFOURGE (COSMOS and UDS). The sSFR are displayed on the *left* (best fits on the left, and upper limits on the right), while *UVJ* colors are shown on the *right*. *UVJ*-quiescent galaxies are highlighted in red. Galaxies outside of the quiescent region but with sSFR $<$ sSFR$_Q = 0.15$ Gyr^{-1} are labeled with green and blue open symbols, depending on whether they are in the "dusty" or "non-dusty" part of the *UVJ* diagram (respectively), as delimited with the diagonal dotted line ($U - J = 2.6$).

de-contamination from redshift interlopers, and strong [O III] emitters. Given the moderate number of objects at play, this change is nevertheless contained within the Poisson error bars.

5.2. Completeness of the UVJ selection and link to sSFR

While the MOSFIRE observations we present in this paper demonstrate that the *UVJ* selection has a high purity, its completeness still remains to be addressed. Indeed, the galaxy 3D-EGS-18996 studied in the present paper (see Sect. 3.7) and the analysis of Merlin et al. (2018) suggest that genuinely quiescent galaxies can be found outside of the fiducial *UVJ* boundaries. Merlin et al. showed that this can be the case either when a quiescent galaxy is both old and significantly obscured by dust (it then fails the $V - J$ cut), or if it is devoid of dust and quenched abruptly less than a few million years prior observation (in which case it fails the $U - V$ cut, as is the case for 3D-EGS-18996). The latter could be a common occurrence, especially at high redshifts where galaxies had little available time to evolve passively after quenching. If true, *UVJ*-selected sample may underestimate the actual number densities of quiescent galaxies.

We illustrate this issue in Fig. 17 by showing the sSFRs of all the ZFOURGE galaxies, in relation to their position on the *UVJ* diagram. As discussed in the previous section, *UVJ*-quiescent galaxies in our sample are predominantly located at sSFR $<$ sSFR$_{MS}$/10, which we adopt as a threshold for quiescence: sSFR$_Q = 0.15$ Gyr^{-1}. We note that this threshold is valid only for $z > 3$ galaxies, and corresponds to a "relative" quiescence; galaxies below this threshold at $z > 3$ are forming stars with rates an order of magnitude lower than the bulk of the star-forming population. Galaxies thus labeled "quiescent" may still form stars at low rates of up to $10 M_\odot$ yr^{-1}, which would be high for $z \sim 0$ galaxies of similar masses, but is nonetheless far enough from the average at $z > 3$ to require an abnormal event

(e.g., quenching) in the galaxy's history (see discussion in the introduction).

As shown in Fig. 17, while the vast majority of the *UVJ*-quiescent galaxies are found below sSFR $=$ sSFR$_Q$, 55% of the galaxies below this threshold are not classified as *UVJ*-quiescent. This would imply indeed that the *UVJ* selection is incomplete.

We find 26 such galaxies with a best-fit sSFR $<$ sSFR$_Q$ but *UVJ*-"star-forming" colors, and note that their colors remain nonetheless within 0.4 mag of the *UVJ* dividing line. Based on the above discussion, we split this sample in two parts according to their *UVJ* colors (see Fig. 17): "young-quiescent" galaxies with "blue" colors, $(U - J) < 2.6$, and "dusty-quiescent" galaxies with "red" colors, $(U - J) > 2.6$. Each sample contains 13 galaxies. In the following, we quantify what fraction of these are genuinely quiescent, that is, with sSFR $<$ sSFR$_Q$ and no dust-obscured star formation.

5.3. Young or dusty quiescent galaxies

In this section we analyze the available data for these "young" or "dusty-quiescent" galaxies to prune the sample, and determine which of them are most likely to be genuinely quiescent. We first use archival ALMA data to identify dust-obscured star-forming galaxies and, for those galaxies without ALMA coverage, to statistically infer the rate of contamination from dusty SFGs. We then use the probability distribution function of the sSFR as obtained from the SED modeling to estimate how many are truly at sSFR $<$ sSFR$_Q$. Using these data, we finally estimate their contribution to the number density of quiescent galaxies.

5.3.1. Removing dusty SFGs

We cross-matched all the ZFOURGE galaxies at $3 < z < 4$ and $M_* > 3 \times 10^{10} M_\odot$ with the ALMA archive, as in Sect. 2.8 for the

Table 4. ALMA coverage and detection rates.

Category[a]	Num.[b]	Cov.[c]	Det.[d]	Frac.[e]
All	130	46	22	48%
Star-forming	81	23	16	70%
... Dusty	24	8	8	100%
... Non-dusty	57	15	8	53%
Dusty-quiescent	13	6	4	67%
Young-quiescent	13	5	1	20%
UVJ-quiescent	23	12	1	8%

Notes. [a] Star-forming: UVJ star-forming and $sSFR > sSFR_Q$, dusty-quiescent: UVJ star-forming, $(U - J) > 2.6$, and $sSFR < sSFR_Q$, young-quiescent: UVJ star-forming, $(U - J) < 2.6$, and $sSFR < sSFR_Q$. For star-forming galaxies, "dusty" and "non-dusty" refer to the location of galaxies on the UVJ diagram as per the "dusty" and "young" quiescent samples. [b] Total number of galaxies in this category. [c] Number of galaxies located inside the full-width-half-power area covered by archival ALMA observations. [d] Number of detections at more than 3σ significance. [e] Fraction of detections among the covered galaxies.

MOSFIRE sample, and looked for detections[4]. Of the 130 massive galaxies in ZFOURGE, 46 were observed with ALMA, and 22 were detected at more than 3σ significance. The detection limit is typically 0.8 and 0.2 mJy in Band 7 and 6, respectively, which is equivalent to an L_{IR} detection limit of about $10^{12} L_\odot$ (Schreiber et al. 2018a), or $SFR \gtrsim 100\ M_\odot\ yr^{-1}$ (Kennicutt 1998). An ALMA detection can therefore be used to rule out quiescence (see however Schreiber et al. 2018b). The coverage and detection rates are summarized in Table 4. There, and in what follows, we only consider as star-forming the galaxies that have UVJ-star-forming colors and $sSFR > sSFR_Q$.

Among galaxies covered by ALMA, dusty-quiescent galaxies have a detection rate of 67% (4/6), while the young-quiescent have a detection rate of only 20% (1/5). Albeit to a lesser extent, a similar trend can be observed for star-forming galaxies, which have detection rates of 100% (8/8) for the dusty galaxies and 53% (8/15) for the non-dusty galaxies. We also note that the detection rate for UVJ-quiescent galaxies is particularly low (8%, 1/12).

From this we conclude that the vast majority of the dusty-quiescent galaxies are not robustly quiescent. The possibility of their dust emission being powered by old or intermediate age stars should be explored in more detail (e.g., Fumagalli et al. 2014; Schreiber et al. 2018b), but to be conservative we plainly discarded them as unreliable and focused exclusively on the young-quiescent galaxies.

Of the 13 young-quiescent galaxies, five were observed with ALMA, and only one is detected (ZF-GS-8706), which we excluded in the following. This implies that, although a priori non-dusty from their UVJ colors, a small fraction are nevertheless dusty contaminants, of order 20%. We statistically take this into account later for the eight galaxies lacking ALMA coverage. The SEDs of the 12 remaining galaxies, excluding ZF-GS-8706, are shown in Fig. 18. These are the galaxies we considered as genuine quiescent candidates complementing the UVJ-selected quiescent galaxies.

5.3.2. Spectroscopic redshifts and high AGN fraction

Four of the young-quiescent galaxies are X-ray detected, either by *Chandra* (ZF-GS-7932, 9332, and 13954) or *XMM-Newton* (ZF-UDS-8092), resulting in an X-ray detection rate of 33%, twice higher than for UVJ-quiescent galaxies (18%). This suggests that AGNs are particularly common in this population, similarly to "green-valley" galaxies at lower redshifts (e.g., Wang et al. 2017).

One galaxy, ZF-UDS-8197, had its redshift confirmed to $z_{spec} = 3.543$ in our MOSFIRE masks (see Sect. 3.3). Otherwise, four galaxies have spectroscopic redshifts from the literature, all in excellent agreement with the photometric redshifts: ZF-GS-7363 at $z_{spec} = 3.579$ (Tasca et al. 2017), ZF-GS-9332 at $z_{spec} = 3.700$ (Hernán-Caballero & Hatziminaoglou 2011), ZF-GS-13954 at $z_{spec} = 3.064$ (Silverman et al. 2010), and ZF-UDS-8092 at $z_{spec} = 3.222$ (Akiyama et al. 2015). This suggests that redshift outliers are not a major concern for these galaxies, especially given that three of these galaxies with confirmed redshifts are among the X-ray detected AGNs, for which we could expect the photometric redshifts to be the most uncertain. This can be explained by their characteristic SEDs (Fig. 18), in which both the Lyman and Balmer breaks are present and unambiguously constrain the redshift. We thus expect this population to not be significantly affected by redshift outliers, and assumed in the following that their z_{phot} are accurate.

However, a fraction of this sample could be affected by high equivalent-width emission lines, particularly those hosting an AGN. Two young-quiescent galaxies were observed in our MOSFIRE runs, ZF-UDS-8092 and ZF-UDS-8197 (the former was not included in our analysis so far because it failed to fulfill our UVJ color cuts, see Sect. 2.4). Both galaxies indeed show significant [O III] emission. ZF-UDS-8197 was already discussed in Sect. 3.6 as one of the "high-EW" galaxy, and its photometry was subsequently corrected for the [O III] flux. ZF-UDS-8092 has a much weaker equivalent-width (35 Å) such that its impact on the photometry is marginal. A more systematic near-IR spectroscopic follow-up would be required to definitively address this question, but this suggests the implied SFH for these galaxies should be interpreted with caution, as ages (or stellar masses) might be biased high if [O III] contributes significantly to the K band flux[5]. We show however in Sect. 5.3.4 that these galaxies have measurably younger stellar populations than the other quiescent galaxies, which suggests this should not be a major issue.

5.3.3. Contribution to the number density

While these young-quiescent galaxies have best-fit $sSFR$ lower than $sSFR_Q$, 67% actually have an upper limit on their $sSFR$ larger than $sSFR_Q$, compared to only 30% for the UVJ-quiescent galaxies. To account for this, we used the probability distribution of the $sSFR$ derived from the SED modeling and weighted each galaxy by the probability that its $sSFR$ is indeed lower than $sSFR_Q$. We find that 10.6 galaxies should be truly below this threshold. Assuming a 20% ALMA detection rate for the galaxies without ALMA coverage, this number was further reduced to a total of 9.2 genuinely quiescent galaxies. Applying the same weighting schemed to the UVJ-quiescent galaxies, we find 18.6 UVJ-quiescent galaxies below $sSFR_Q$.

---

[4] For this exercise we used data from the same programs listed in Sect. 2.8, adding Band 7 data from 2015.1.01495.S (PI: Wang), 2012.1.00307.S (PI: Hodge), 2012.1.00869.S (PI: Mullaney), as well as Band 6 observations from 2013.1.00151.S and 2015.1.00379.S (PI: Schinnerer), and 2013.1.00118.S (PI: Aravena).

[5] To some extent, [O II] could counter-balance this bias by increasing the flux in the H band, however none of the galaxies we observed with MOSFIRE here had [O II] contributing more than a few percent of the H-band photometry. We therefore only expect [O III] to be a source of concern in case of bright emission lines.
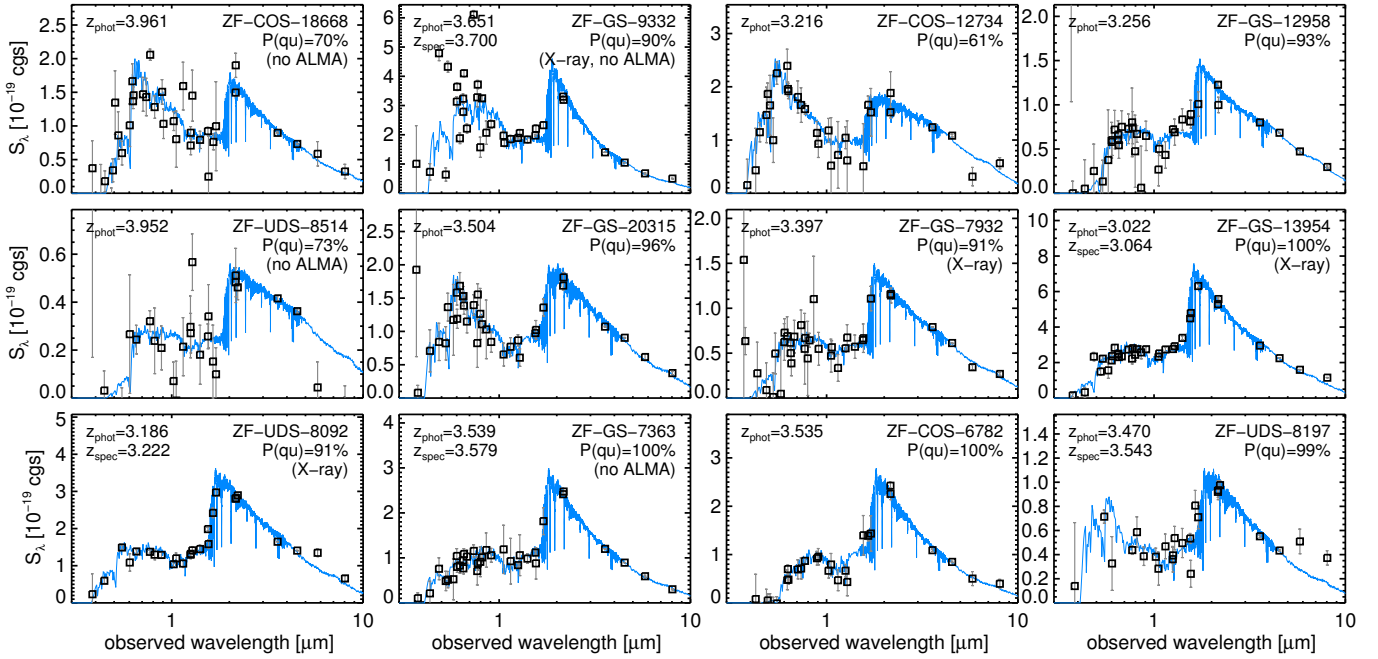
Fig. 18. Same as Fig. 3, but for the "young-quiescent" galaxies, namely galaxies in ZFOURGE that lie outside of the UVJ-quiescent region but are probably still quiescent. Beside the name of each galaxy, we show the probability that its sSFR is lower than sSFR$_Q$, our adopted threshold for quiescence, and we also indicate which galaxies have X-ray detections ("X-ray") and those which are covered by ALMA but not detected ("no ALMA").

Combined, the two samples produce a total of 27.8 galaxies with sSFR < sSFR$_Q$, or a space density of $(2.0 \pm 0.3) \times 10^{-5}$ Mpc^{-3}. This is 40% larger than the space density of UVJ-quiescent galaxies alone: young-quiescent galaxies therefore form a sizable fraction of the quiescent population at $z > 3$.

5.3.4. Star formation histories

To understand what sets these young-quiescent galaxies apart from the rest of the UVJ-quiescent population, we compared their estimated star-formation histories from the SED modeling, keeping in mind that high-EW [O III] emission might tend to artificially increase the strength of their Balmer break (see Sect. 5.3.2), hence drive models toward older ages.

As expected based on the analysis of Merlin et al. (2018), we find that young-quiescent galaxies have experienced a more recent quenching: the 16th and 84th percentiles of $t_{\rm quench}$ are 15 and 330 Myr, respectively, about 200 Myr later than the other quiescent galaxies. There is evidence that they also entered their main formation phase on average 300 Myr later. This would indicate at first order that their SFHs are not intrinsically different from that of other quiescent galaxies, but are simply shifted to later times.

The existence of such objects is natural if the quiescent population was assembled gradually, with new galaxies becoming quiescent at all epochs. In this context, their relative importance to the overall quiescent population can be expected to increase toward earlier times, to the point where they must be dominant at the epoch of formation of the very first quiescent galaxies. As showed in the previous section, UVJ-quiescent galaxies were still dominant at $3 < z < 4$ (66% of the population), but not by a large margin. This suggests that the UVJ selection may not be adequate in finding $z > 4$ quiescent galaxies, and that a revised color selection or an sSFR criterion will have to be used instead.

One way to achieve this would be to adjust the bottom edge of the UVJ selection. For example, at $3 < z < 4$, 6 out of 13 of the young-quiescent galaxies would be included in the UVJ selection if we were to remove the constraint $(U - V) > 1.3$. This constraint was initially introduced in Williams et al. (2009) to prevent blue galaxies from entering this region by random scatter, which is critical when blue galaxies dominate the parent sample. For galaxies as massive as those considered here, this is not so much an issue, both because the photometry is quite accurate so random scatter is limited, but also because relatively few massive galaxies are blue or un-obscured. One could therefore adopt this as a revised UVJ selection, fine-tuned for more massive samples. However, if we were to remove this constraint we would also include 7 galaxies with sSFR > sSFR$_Q$, which would reduce the purity of the sample. This could be alleviated by adding more colors, for example sensitive to more recent star-formation (far-UV).

6. Discussion

Our observations clearly demonstrate the existence of quiescent galaxies at $z > 3$, and confirm the high number density estimated from photometric samples. With these facts solidly established, the main question still remains open: do we understand how these objects formed? In this section we describe if and how the $z > 3$ quiescent population is reproduced in current galaxy formation models. Following G17, we looked in particular at how well the number densities are reproduced, and, when possible, how the simulated star formation histories compare to that inferred from our sample.

We based the comparison of number densities on the observed number of galaxies with low sSFR, since this selection is immediately applicable to models, and we therefore included the young-quiescent galaxies in this analysis (i.e., the galaxies with low sSFR located outside of the UVJ quiescent region, see

Sect. 5.3). The following discussion would not change significantly if we were to only consider the UVJ quiescent galaxies.

6.1. Number densities predicted by semi-analytic models

We first looked at semi-analytic models (SAMs). We downloaded the first $3.14 \deg^2$ light cone of the Henriques et al. (2015) SAM[6] and searched for galaxies in the same range of mass and redshift as ours. Looking at quiescent galaxies with $sSFR < sSFR_Q$ we find only four objects. The corresponding number density, $1.1 \times 10^{-7} \, \mathrm{Mpc}^{-3}$, is smaller than our observed value by a factor ~ 200. Simulating uncertainties in stellar masses as a log-normal error of $0.07 \, \mathrm{dex}$ (the median uncertainty of our galaxies) increased the predicted number density by 20%, which is insufficient to match the observed value. The overall number density of massive galaxies this model predicts, regardless of sSFR, is also too small by a factor of about four: $2.4 \times 10^{-5} \, \mathrm{Mpc}^{-3}$, compared to $9.3 \times 10^{-5} \, \mathrm{Mpc}^{-3}$ in the ZFOURGE catalogs. This suggests that both the formation of massive galaxies and their subsequent quenching happens too late in this model.

More recently, Rong et al. (2017) used an earlier version of this SAM (Guo et al. 2011) to study analogs of ZF-COS-20115, the $z = 3.7$ quiescent galaxy discussed in G17. While they were indeed able to find quiescent galaxies at these redshifts, their predicted space density was comparably low: $7.5 \times 10^{-8} \, \mathrm{Mpc}^{-3}$ for galaxies with $sSFR < sSFR_Q$ and $M_* > 10^{11} \, M_\odot$, a factor 30 lower than we observe for this same mass threshold (i.e., $(2.2 \pm 1.1) \times 10^{-6} \, \mathrm{Mpc}^{-3}$).

In both these SAMs, quenching of massive galaxies is explicitly caused by AGNs in the so-called radio mode, that is, when the black hole slowly accretes hot gas from the host galaxy. Energy is artificially deposited into the galaxy's halo, supposedly from the action of the AGN radio jets, and inhibits cooling. The more intense quasar mode, which happens only during galaxy mergers, does not generate explicit feedback but is accompanied by stellar feedback from the resulting starburst, which Guo et al. (2011) found is sufficient to remove all the gas from the galaxy. In the Henriques et al. SAM, the parameters of the quenching model were tuned to match additional observables at $0 < z < 3$, chiefly the fraction of color-selected (UVJ) quiescent galaxies as function of mass and redshift. The fact that, despite a calibration up to $z = 3$, this model still cannot reproduce the number densities of QGs at $3 < z < 4$ is surprising. We suspect this could be caused by two effects. First, an over-estimation of the stellar mass uncertainties; Henriques et al. assumed $0.36 \, \mathrm{dex}$ at $z \sim 3.5$, which is a factor two larger than reported in observations (e.g., Ilbert et al. 2013). Second, the mismatch of definition for "quiescent" galaxies between their model and observations. Indeed, the dividing line between quiescent and star-forming in their UVJ diagram is different than in observations, and is shifted toward dusty galaxies; their predicted quiescent sample used in the model calibration phase will therefore include a number of red but nonetheless star-forming contaminants. In the model calibration, this will make it possible to match the observed number of "red" galaxies while under-predicting the actual number of galaxies with low sSFR.

The MERAXES SAM (Qin et al. 2017a,b) was recently shown to predict more accurate number densities, both for the quiescent and the overall massive population. After correcting their masses from the Salpeter (1955) to the Chabrier (2003) IMF we used

here, their model predicts a number densities of quiescent galaxies of $1.1 \times 10^{-5} \, \mathrm{Mpc}^{-3}$ at $z = 3.5$ (Y. Qin, priv. comm.), which is only a factor two lower than our observed value. The prediction for the number density of massive galaxies, regardless of sSFR, is within 10% of the observed value ($8.5 \times 10^{-5} \, \mathrm{Mpc}^{-3}$). Therefore massive galaxies in this model are formed in the right numbers, but quench slightly too late (see below).

The ingredients for quenching in MERAXES are essentially the same as in the Henriques and Guo SAMs, but the calibration strategy and implementation differ. Perhaps the most immediate difference is that this model, introduced to study reionization, was calibrated to match the observed stellar mass functions at $0.6 < z < 7$ with an emphasis on $z > 5$. Most other models, like the Henriques SAM, are instead calibrated with data at $0 < z < 3$, usually with a special emphasis on the $z = 0$ mass function. This different focus may allow the MERAXES model to better reproduce the $3 < z < 4$ galaxy population, but possibly at the expanse of a poorer description of the $z \sim 0$ Universe (their merger trees stop at $z = 0.56$). MERAXES also includes a calibration of the black hole masses (anchored at $z \sim 0$), and an indirect calibration on SFRs through the matching of reionization. Although the latter should only constrain star formation in low-mass galaxies (which dominate reionization), this may impact the star-formation efficiency in the progenitors of massive $z \sim 3.5$ quiescent galaxies as well. Finally, a last notable difference is that this SAM was implemented on a dark matter merger tree generated with a finer time step (11 Myr at $z > 5$, 30 Myr at $z \sim 4$), which allows a better temporal sampling of star-formation and accretion histories.

The fact that this model manages to reproduce the number density of quiescent galaxies at $3 < z < 4$ is encouraging and suggests that, with adequate calibration, the AGN-based quenching model is indeed able to explain the existence of such galaxies. However, it remains to be shown that this agreement is not reached at the expense of other observable properties which were not the focus of the MERAXES model, in particular at lower redshifts.

6.2. Number densities predicted by cosmological hydrodynamic simulations

The strength of SAMs is their ability to be anchored to different observables while exploring a wide space of parameters and models. This freedom is also a weakness: since the calibration of these models plays such an important role, the predictive power of SAMs, albeit not null, is limited. Cosmological hydrodynamic simulations come closer to a "first-principle" approach, and although their limited spatial resolution still requires the use of empirical sub-grid recipes and regularization schemes, their predictive power is enhanced compared to SAMs. The downside is that the studied volume is usually smaller, but modern simulations are now executed on large enough volumes that a comparison to our results can be attempted.

In Illustris (Wellons et al. 2015, Fig. 6), only five galaxies have $M_* > 3 \times 10^{10} \, M_\odot$ and $sSFR < sSFR_Q$ at $z = 3$, and none exist at $z = 4$ (as first noticed in G17). Given the box size of Illustris, this translates into an average number density at $3 < z < 4$ of $2.1 \times 10^{-6} \, \mathrm{Mpc}^{-3}$, an order of magnitude too low. The MUFASA simulation (Davé et al. 2016) has a three times smaller volume, $(50 \, h^{-1})^3 \, \mathrm{Mpc}^3$, and will therefore offer more limited statistics, but a comparison is still possible. In this simulation, only two galaxies match the required properties at $z = 3$, and none at $z = 4$ (R. Davé, priv. comm.). The average number density at $3 < z < 4$ is therefore similar to that of Illustris, $2.5 \times 10^{-6} \, \mathrm{Mpc}^{-3}$.

---

[6] http://galformod.mpa-garching.mpg.de/public/LGalaxies/downloads.php

While Illustris implemented AGN feedback in a fashion similar to that discussed above for SAMs (i.e., including both quasar- and radio-mode feedback), MUFASA adopted a simpler phenomenological model. There, AGN feedback was not invoked explicitly; instead, a "halo quenching" mass was introduced above which all the gas which was not self-shielded was artificially heated, hence prevented from forming stars. The source of this heating is unspecified in the model, but AGNs would be obvious candidates. Despite its simplicity, this model was still able to offer a good match to the $z < 4$ stellar mass functions, but with a lack of massive galaxies at $z > 4$ (Davé et al. 2016).

These two simulations therefore fall short of the observed number density by about an order of magnitude despite implementing different quenching mechanisms. Given that the MERAXES SAM was able to match our observations, a side-to-side comparison of galaxy properties with these different models could reveal what feature might be currently missing in the hydrodynamic simulations, or if the feedback model implemented in MERAXES can be implemented at all in a more physically-motivated context.

6.3. Star formation histories in MERAXES

Since, among all the models we explored, MERAXES was the only one that came close to reproducing the observed number density of quiescent galaxies, we proceeded to investigate whether the star formation histories of the quiescent galaxies in this model match the ones we inferred in Sect. 4.5. For this analysis we used the 70 quiescent galaxies in the $z = 3.5$ MERAXES snapshot kindly provided by Y. Qin (priv. comm.). We applied the same procedure as for the observed galaxies to compute t_{quench} and t_{form}, using the tabulated SFHs produced by MERAXES. The results are illustrated on Fig. 16.

We find that, on average, quiescent galaxies in MERAXES have been quiescent for 150 Myr, a duration about twice shorter than that measured for the observed galaxies, 330 ± 30 Myr. The 16th and 84th percentiles of t_{quench} are 0 and 210 Myr in MERAXES, versus 210 and 510 Myr in observations. In this simulation, the quenching therefore happened at later times, which is consistent with the slight under-prediction of the number density. However we find that this time shift also affects the formation epoch: in the simulation, the quiescent galaxies had assembled half of their observed mass on average 550 Myr prior to observation (percentiles: 370 and 700 Myr), compared to 780^{+30}_{-70} Myr for the observed galaxies (percentiles: 360 and 1070 Myr). While these numbers are comparable at first order, the simulated SFHs are systematically shifted to later times by about 200 Myr (10% of the *Hubble* time at $z = 3.5$). This would be consistent with the predicted space density of $z \sim 3.5$ quiescent galaxies being lower than the observed value.

In addition, we find that the simulated SFHs are more extended in time than the observed ones. The average SFR of these galaxies during their formation phase was $160\,M_{\odot}\,\mathrm{yr}^{-1}$ in the simulation, while we inferred $340^{+150}_{-30}\,M_{\odot}\,\mathrm{yr}^{-1}$ for the observed galaxies. Consequently, the duration of the star-forming phase was longer in the simulation: 480 Myr versus 280^{+40}_{-134} Myr for our observed sample.

Our observations therefore suggest that the formation of $3 < z < 4$ quiescent galaxies happened earlier, in shorter, more intense bursts than what this model predicts. This inevitably requires a higher star formation efficiency, as already argued in G17, and a corresponding increase in quenching efficiency to avoid over-producing star-forming galaxies. The possibility of achieving this with an AGN-driven quenching model goes beyond the scope of this paper, but should definitely be explored.

We recall, however, that the galaxies studied in Sect. 4.5 are only those for which we obtained a MOSFIRE spectrum, and these are mostly UVJ-quiescent. Young-quiescent galaxies are under-represented in this sample, and including them would tend to shift the observed population toward overall younger ages (Sect. 5.3.4). While this would improve the agreement with MERAXES, we do not expect the discrepancy to disappear entirely. Indeed, we quantified the fraction of young-quiescent galaxies in MERAXES by producing synthetic UVJ colors with the Bruzual & Charlot (2003) model and assuming $A_{\mathrm{V}} = 0.3$ mag of attenuation (the median of our sample, see Table 3). In this model, young-quiescent galaxies make up 73% of the quiescent population, while they remain a minority in ZFOURGE (33%). It is therefore clear that quiescent galaxies in MERAXES are overall younger.

7. Conclusions

We have obtained NIR spectra for a sample of $3 < z < 4$ massive, UVJ-selected quiescent galaxies with MOSFIRE. These spectra allow us to measure a spectroscopic redshift for 40% of our targets, revealing that their photometric redshifts have an excellent accuracy of 1.2%, with a catastrophic failure rate of 10% coming from dusty galaxies at lower redshifts. An additional 10% are found at the correct redshift, but with high-equivalent-width [O III] emission contributing significantly to the K-band flux, hence for which the steepness of the Balmer break was overestimated. The rest of the sample shows no strong emission lines. Combined, this demonstrates that the UVJ selection of quiescent galaxies suffers from a contamination rate of only 20% at $3 < z < 4$.

Balmer absorption lines are detected in four galaxies (among the brightest of the sample), a clear signpost of a recent quenching, and in agreement with expectations from their broadband photometry. The star-formation histories inferred for the entire sample show that all but one of the galaxies have quenched with certainty, on average 330 Myr before being observed, sometime between $z = 3.5$ and $z = 5$. Half of their stars were formed by $z = 4.4$ to $z = 7.1$ in brief (280 Myr) star-formation episodes with SFRs of $80–850\,M_{\odot}\,\mathrm{yr}^{-1}$.

These results therefore confirm that the UVJ selection is efficient in selecting genuinely quiescent galaxies. Building on this result and our estimated contamination rate, we came back to the parent catalogs and updated the number density of UVJ-quiescent galaxies at $3 < z < 4$. At $M_* > 3 \times 10^{10}\,M_{\odot}$, we find $(1.4 \pm 0.3) \times 10^{-5}\,\mathrm{Mpc}^{-3}$. This number is slightly lower than previous estimates, but the overall picture remains unchanged: a substantial population of quiescent galaxies did exist at these early epochs.

To compare this number to models, we then investigated the completeness of the UVJ selection in terms of sSFR. We find a sizable population of galaxies with sSFR $< 0.15\,\mathrm{Gyr}^{-1}$ (a factor of ten below the main sequence) and no dust-obscured star-formation, some of which were missing from our initial UVJ-selected sample. We dubbed these "young-quiescent" galaxies. Their modeling suggest they quenched later than the other quiescent galaxies, such that their colors are not yet red enough to enter the UVJ selection. We find they have a number density comparable to that of the UVJ-quiescent galaxies. The combined number density of galaxies with low sSFR is therefore larger than that of UVJ-quiescent galaxies alone, namely, $(2.0 \pm 0.3) \times 10^{-5}\,\mathrm{Mpc}^{-3}$.

We finally compared our results to recent models of galaxy formation. All the models we explored predict number densities from one to two orders of magnitude too low. The only exception is the MERAXES semi-analytic model, which was tuned with a particular emphasis on high redshift observations, and which falls short by only a factor two. This shows that the AGN-based quenching model, adopted in MERAXES, can produce quiescent galaxies in almost the right numbers, albeit only with some specific tuning to high-redshift galaxies. Yet the predicted star formation histories in MERAXES do not match our observations. The formation and quenching of massive $z \sim 3.5$ galaxies in this model happens 200 Myr too late, and their star-forming phase is twice longer than observed, with past SFRs too low by a similar factor. This suggests that quenching of high-redshift galaxies is not yet a fully understood process.

From here, a number of follow-up studies can be undertaken. In a future work, we will attempt to predict the properties of the star-forming progenitors of our quiescent galaxies, and compare their inferred SFRs and space densities to known SFGs at high redshifts. A more detailed analysis of the ALMA-detected quiescent candidates is also in order, to investigate whether the sub-mm emission is always powered by obscured star-formation, or if it may originate from heating by older stars (e.g., Schreiber et al. 2018b). In addition, for those which are truly star-forming, it would be worthwhile to understand why the color classification failed in the first place.

Follow-up observations would also be beneficial to better understand this population. Deeper spectra would allow us to place tighter constraints on the SFHs, pin-pointing with greater accuracy the time of quenching and the duration of the star-forming episode. Perhaps more importantly, this would also allows us to start constraining the chemical enrichment histories, which can further constrain the SFH and the physics of stellar evolution in the early, metal-poor Universe. Velocity dispersion measurements would constrain dynamical masses and the stellar initial mass function, which may or may not be as universal as thought. There is only so much one can do from ground-based 8 m class telescopes however, and these are questions that will undoubtedly require the help of the next generation of instruments, chiefly the JWST and its on-board spectrograph, NIR-Spec.

The origin of the low level residual star-formation in these otherwise "dead" galaxies is also intriguing, and could be linked to different processes. New gas can be brought in from the generous infall expected at these high redshifts, or from gas recycling. Alternatively, this residual activity may just be the declining "tail" of a past starburst, if quenching is not an instantaneous event. These different scenarios could be investigated with higher resolution NIR imaging to locate these star-forming regions, and with deep spectroscopy to reveal the physical condition in these small reservoirs of star-forming gas (i.e., metallicity). JWST will again be a key instrument to answer these questions.

Finally, while it was thought for a long time that galaxies must fully deplete (or expel) their gas reservoirs before quenching, evidence is building that this is not always true and that sizable reservoirs exist in quenched or post-starburst galaxies. This question could be addressed by looking at those galaxies in our sample which have spent the smallest amount of time since quenching, and study their gas or dust content, for example through systematic and deep follow-ups with ALMA. At these high redshifts, ALMA can cover a number of interesting emission lines, chiefly [C II], [N II], and the CO ladder, while providing a reasonable sampling of the dust SED to derive infrared luminosities and dust (or gas) masses.

References

Akiyama, M., Ueda, Y., Watson, M. G., et al. 2015, PASJ, 67, 82
Alatalo, K., Appleton, P. N., Lisenfeld, U., et al. 2014, ApJ, 795, 159
Alatalo, K., Lacy, M., Lanz, L., et al. 2015, ApJ, 798, 31
Audouze, J., & Tinsley, B. M. 1976, ARA&A, 14, 43
Avni, Y. 1976, ApJ, 210, 642
Baldry, I. K., Glazebrook, K., Brinkmann, J., et al. 2004, ApJ, 600, 681
Baldwin, J. A. 1975, ApJ, 201, 26
Baldwin, J. A., Phillips, M. M., & Terlevich, R. 1981, PASP, 93, 5
Barro, G., Faber, S. M., Pérez-González, P. G., et al. 2013, ApJ, 765, 104
Barro, G., Kriek, M., Pérez-González, P. G., et al. 2016, ApJ, 827, L32
Behroozi, P., & Silk, J. 2018, MNRAS, 477, 5382
Belli, S., Newman, A. B., Ellis, R. S., & Konidaris, N. P. 2014, ApJ, 788, L29
Belli, S., Genzel, R., Förster Schreiber, N. M., et al. 2017a, ApJ, 841, L6
Belli, S., Newman, A. B., & Ellis, R. S. 2017b, ApJ, 834, 18
Benítez, N. 2000, ApJ, 536, 571
Benson, A. J., Bower, R. G., Frenk, C. S., et al. 2003, ApJ, 599, 38
Birnboim, Y., & Dekel, A. 2003, MNRAS, 345, 349
Blanton, E. L., Sarazin, C. L., McNamara, B. R., & Wise, M. W. 2001, ApJ, 558, L15
Bower, R. G., Benson, A. J., Malbon, R., et al. 2006, MNRAS, 370, 645
Brammer, G. B., van Dokkum, P. G., & Coppi, P. 2008, ApJ, 686, 1503
Brammer, G. B., Whitaker, K. E., van Dokkum, P. G., et al. 2011, ApJ, 739, 24
Bruzual, G., & Charlot, S. 2003, MNRAS, 344, 1000
Butcher, H., & Oemler, A. J. 1978, ApJ, 219, 18
Caccianiga, A., & Severgnini, P. 2011, MNRAS, 415, 1928
Calzetti, D., Armus, L., Bohlin, R. C., et al. 2000, ApJ, 533, 682
Cappellari, M., & Emsellem, E. 2004, PASP, 116, 138
Cattaneo, A., Faber, S. M., Binney, J., et al. 2009, Nature, 460, 213
Chabrier, G. 2003, PASP, 115, 763
Cid Fernandes, R., Stasińska, G., Schlickmann, M. S., et al. 2010, MNRAS, 403, 1036
Cid Fernandes, R., Stasińska, G., Mateus, A., & Vale Asari, N. 2011, MNRAS, 413, 1687
Ciesla, L., Boselli, A., Elbaz, D., et al. 2016, A&A, 585, A43
Ciesla, L., Elbaz, D., & Fensch, J. 2017, A&A, 608, A41
Cimatti, A., Pozzetti, L., Mignoli, M., et al. 2002, A&A, 391, L1
Cimatti, A., Daddi, E., Renzini, A., et al. 2004, Nature, 430, 184
Coil, A. L., Aird, J., Reddy, N., et al. 2015, ApJ, 801, 35
Croton, D. J., Springel, V., White, S. D. M., et al. 2006, MNRAS, 365, 11
Daddi, E., Cimatti, A., & Renzini, A. 2000, A&A, 362, L45
Daddi, E., Cimatti, A., Renzini, A., et al. 2004, ApJ, 617, 746
Davé, R., Thompson, R., & Hopkins, P. F. 2016, MNRAS, 462, 3265
Davé, R., Rafieferantsoa, M. H., & Thompson, R. J. 2017, MNRAS, 471, 1671
Davis, T. A., Young, L. M., Crocker, A. F., et al. 2014, MNRAS, 444, 3427
Dekel, A., & Birnboim, Y. 2008, MNRAS, 383, 119
de Vaucouleurs, G. 1948, Ann. Astrophys., 11, 247
Dunlop, J. S., Cirasuolo, M., & McLure, R. J. 2007, MNRAS, 376, 1054

Elbaz, D., Daddi, E., Le Borgne, D., et al. 2007, A&A, 468, 33
Elbaz, D., Leiton, R., Nagar, N., et al. 2018, A&A, 616, A110
Erb, D. K., Pettini, M., Shapley, A. E., et al. 2010, ApJ, 719, 1168
Faber, S. M., Willmer, C. N. A., Wolf, C., et al. 2007, ApJ, 665, 265
Fabian, A. C. 1994, ARA&A, 32, 277
Fontana, A., Santini, P., Grazian, A., et al. 2009, A&A, 501, 15
Franx, M., Labbé, I., Rudnick, G., et al. 2003, ApJ, 587, L79
French, K. D., Yang, Y., Zabludoff, A., et al. 2015, ApJ, 801, 1
Fumagalli, M., Labbé, I., Patel, S. G., et al. 2014, ApJ, 796, 35
Gabor, J. M., & Davé, R. 2012, MNRAS, 427, 1816
Gatz, D. F., & Smith, L. 1995, Atmos. Environ., 29, 1185
Glazebrook, K., Abraham, R. G., McCarthy, P. J., et al. 2004, Nature, 430, 181
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
Grogin, N. A., Kocevski, D. D., Faber, S. M., et al. 2011, ApJS, 197, 35
Guo, Q., White, S., Boylan-Kolchin, M., et al. 2011, MNRAS, 413, 101
Henriques, B. M. B., White, S. D. M., Thomas, P. A., et al. 2015, MNRAS, 451, 2663
Hernán-Caballero, A., & Hatziminaoglou, E. 2011, MNRAS, 414, 500
Hill, A. R., Muzzin, A., Franx, M., & van de Sande, J. 2016, ApJ, 819, 74
Hopkins, P. F., Hernquist, L., Cox, T. J., & Kereš, D. 2008, ApJS, 175, 356
Husser, T., Wende-von Berg, S., Dreizler, S., et al. 2013, A&A, 553, A6
Ilbert, O., Salvato, M., Le Floc'h, E., et al. 2010, ApJ, 709, 644
Ilbert, O., McCracken, H. J., Le Fèvre, O., et al. 2013, A&A, 556, 55
Inami, H., Bacon, R., Brinchmann, J., et al. 2017, A&A, 608, A2
Juneau, S., Dickinson, M., Alexander, D. M., & Salim, S. 2011, ApJ, 736, 104
Kado-Fong, E., Marchesini, D., Marsan, Z. C., et al. 2017, ApJ, 838, 57
Kauffmann, G., Heckman, T. M., Tremonti, C., et al. 2003a, MNRAS, 346, 1055
Kauffmann, G., Heckman, T. M., White, S. D. M., et al. 2003b, MNRAS, 341, 33
Kennicutt, R. C. Jr. 1998, ARA&A, 36, 189
Kewley, L. J., Geller, M. J., & Jansen, R. A. 2004, AJ, 127, 2002
Koekemoer, A. M., Faber, S. M., Ferguson, H. C., et al. 2011, ApJS, 197, 36
Kriek, M., van Dokkum, P. G., Franx, M., et al. 2006, ApJ, 645, 44
Kriek, M., van Dokkum, P. G., Labbé, I., et al. 2009, ApJ, 700, 221
Kriek, M., Shapley, A. E., Reddy, N. A., et al. 2015, ApJS, 218, 15
Kriek, M., Conroy, C., van Dokkum, P. G., et al. 2016, Nature, 540, 248
Labbé, I., Huang, J., Franx, M., et al. 2005, ApJ, 624, L81
Lacey, C., & Cole, S. 1993, MNRAS, 262, 627
Lemaux, B. C., Lubin, L. M., Shapley, A., et al. 2010, ApJ, 716, 970
Lin, L., Belfiore, F., Pan, H., et al. 2017, ApJ, 851, 18
Marsan, Z. C., Marchesini, D., Brammer, G. B., et al. 2017, ApJ, 842, 21
Martig, M., Bournaud, F., Teyssier, R., & Dekel, A. 2009, ApJ, 707, 250
Martis, N. S., Marchesini, D., Brammer, G. B., et al. 2016, ApJ, 827, L25
Mawatari, K., Yamada, T., Fazio, G. G., Huang, J., & Ashby, M. L. N. 2016, PASJ, 68, 46
McLean, I. S., Steidel, C. C., Epps, H. W., et al. 2012, Proc. SPIE, 8446, 84460J
McLure, R., Pentericci, L., & VANDELS Team 2017, The Messenger, 167, 31
Merlin, E., Fontana, A., Castellano, M., et al. 2018, MNRAS, 473, 2098
Moster, B. P., Somerville, R. S., Newman, J. A., & Rix, H. 2011, ApJ, 731, 113
Muzzin, A., Labbé, I., Franx, M., et al. 2012, ApJ, 761, 142
Muzzin, A., Marchesini, D., Stefanon, M., et al. 2013, ApJS, 206, 8
Nanayakkara, T., Glazebrook, K., Kacprzak, G. G., et al. 2016, ApJ, 828, 21
Newman, A. B., Ellis, R. S., Bundy, K., & Treu, T. 2012, ApJ, 746, 162

Noeske, K. G., Faber, S. M., Weiner, B. J., et al. 2007, ApJ, 660, L47
Noll, S., Burgarella, D., Giovannoli, E., et al. 2009, A&A, 507, 1793
Onodera, M., Carollo, C. M., Lilly, S., et al. 2016, ApJ, 822, 42
Pannella, M., Elbaz, D., Daddi, E., et al. 2015, ApJ, 807, 141
Papovich, C., Dickinson, M., & Ferguson, H. C. 2001, ApJ, 559, 620
Papovich, C., Moustakas, L. A., Dickinson, M., et al. 2006, ApJ, 640, 92
Papovich, C., Finkelstein, S. L., Ferguson, H. C., Lotz, J. M., & Giavalisco, M. 2011, MNRAS, 412, 1123
Press, W. H., & Schechter, P. 1974, ApJ, 187, 425
Qin, Y., Mutch, S. J., Duffy, A. R., et al. 2017a, MNRAS, 471, 4345
Qin, Y., Mutch, S. J., Poole, G. B., et al. 2017b, MNRAS, 472, 2009
Reddy, N. A., Kriek, M., Shapley, A. E., et al. 2015, ApJ, 806, 259
Rees, M. J., & Ostriker, J. P. 1977, MNRAS, 179, 541
Rong, Y., Jing, Y., Gao, L., et al. 2017, MNRAS, 471, L36
Salpeter, E. E. 1955, ApJ, 121, 161
Schreiber, C., Pannella, M., Elbaz, D., et al. 2015, A&A, 575, A74
Schreiber, C., Elbaz, D., Pannella, M., et al. 2016, A&A, 589, A35
Schreiber, C., Pannella, M., Leiton, R., et al. 2017, A&A, 599, A134
Schreiber, C., Elbaz, D., Pannella, M., et al. 2018a, A&A, 609, A30
Schreiber, C., Labbé, I., Glazebrook, K., et al. 2018b, A&A, 611, A22
Shapley, A. E., Sanders, R. L., Reddy, N. A., et al. 2017, ApJ, 846, L30
Silk, J., & Rees, M. J. 1998, A&A, 331, L1
Silverman, J. D., Mainieri, V., Salvato, M., et al. 2010, ApJS, 191, 124
Simpson, J. M., Smail, I., Swinbank, A. M., et al. 2017, ApJ, 839, 58
Skelton, R. E., Whitaker, K. E., Momcheva, I. G., et al. 2014, ApJS, 214, 24
Smit, R., Bouwens, R. J., Franx, M., et al. 2012, ApJ, 756, 14
Smit, R., Bouwens, R. J., Labbé, I., et al. 2016, ApJ, 833, 254
Sparre, M., Hayward, C. C., Springel, V., et al. 2015, MNRAS, 447, 3548
Spitler, L. R., Straatman, C. M. S., Labbé, I., et al. 2014, ApJ, 787, L36
Stefanon, M., Marchesini, D., Rudnick, G. H., Brammer, G. B., & Whitaker, K. E. 2013, ApJ, 768, 92
Steidel, C. C., Strom, A. L., Pettini, M., et al. 2016, ApJ, 826, 159
Straatman, C. M. S., Labbé, I., Spitler, L. R., et al. 2014, ApJ, 783, L14
Straatman, C. M. S., Labbé, I., Spitler, L. R., et al. 2015, ApJ, 808, L29
Straatman, C. M. S., Spitler, L. R., Quadri, R. F., et al. 2016, ApJ, 830, 51
Strom, A. L., Steidel, C. C., Rudie, G. C., et al. 2017, ApJ, 836, 164
Suess, K. A., Bezanson, R., Spilker, J. S., et al. 2017, ApJ, 846, L14
Tacconi, L. J., Genzel, R., Neri, R., et al. 2010, Nature, 463, 781
Tasca, L. A. M., Le Fèvre, O., Ribeiro, B., et al. 2017, A&A, 600, A110
Tomczak, A. R., Quadri, R. F., Tran, K. H., et al. 2014, ApJ, 783, 85
van de Sande, J., Kriek, M., Franx, M., et al. 2013, ApJ, 771, 85
van der Wel, A., Chang, Y., Bell, E. F., et al. 2014a, ApJ, 792, L6
van der Wel, A., Franx, M., van Dokkum, P. G., et al. 2014b, ApJ, 788, 28
van Dokkum, P. G., Franx, M., Kriek, M., et al. 2008, ApJ, 677, L5
van Dokkum, P. G., Whitaker, K. E., Brammer, G., et al. 2010, ApJ, 709, 1018
van Dokkum, P. G., Nelson, E. J., Franx, M., et al. 2015, ApJ, 813, 23
Wang, T., Elbaz, D., Daddi, E., et al. 2016, ApJ, 828, 56
Wang, W., Faber, S. M., Liu, F. S., et al. 2017, MNRAS, 469, 4063
Wellons, S., Torrey, P., Ma, C., et al. 2015, MNRAS, 449, 361
Whitaker, K. E., Kriek, M., van Dokkum, P. G., et al. 2012, ApJ, 745, 179
White, S. D. M., & Rees, M. J. 1978, MNRAS, 183, 341
Wild, V., Almaini, O., Cirasuolo, M., et al. 2014, MNRAS, 440, 1880
Williams, R. J., Quadri, R. F., Franx, M., van Dokkum, P., & Labbé, I. 2009, ApJ, 691, 1879
Wuyts, S., Förster Schreiber, N. M., van der Wel, A., et al. 2011, ApJ, 742, 96
Yan, R., Newman, J. A., Faber, S. M., et al. 2006, ApJ, 648, 281

Appendix A: Tables

Table A.1. Galaxies targeted with MOSFIRE, photometry.

ID[a]	RA deg.	Dec. deg.	Observed mag.[b] H	K	$870\,\mu$m mJy
ZF-COS-10559	150.07147	2.2911844	25.74 ± 0.18	24.15 ± 0.08	0.09 ± 0.17
ZF-COS-14907	150.12422	2.3374486	23.44 ± 0.02	22.65 ± 0.02	–
ZF-COS-17779	150.04651	2.3673911	24.94 ± 0.15	23.82 ± 0.06	–
ZF-COS-18842	150.08728	2.3960431	24.45 ± 0.08	23.03 ± 0.04	–
ZF-COS-19589	150.06671	2.3823645	25.41 ± 0.15	23.54 ± 0.06	–
ZF-COS-20032	150.11256	2.3765266	25.14 ± 0.13	23.62 ± 0.06	1.64 ± 0.28
ZF-COS-20115	150.06149	2.3787093	24.39 ± 0.03	22.43 ± 0.02	0.09 ± 0.06[c]
ZF-COS-20133	150.12173	2.3745940	25.75 ± 0.19	23.84 ± 0.06	–
3D-EGS-18996	214.89563	52.856556	22.75 ± 0.01	21.60 ± 0.02	–
3D-EGS-26047	214.90512	52.891621	23.83 ± 0.03	22.55 ± 0.05	–
3D-EGS-27584	214.85387	52.861393	24.62 ± 0.12	22.25 ± 0.07	–
3D-EGS-31322	214.86606	52.884312	23.72 ± 0.03	22.20 ± 0.04	–
3D-EGS-34322	214.81316	52.858986	25.02 ± 0.12	23.52 ± 0.21	–
3D-EGS-40032	214.76062	52.845383	22.92 ± 0.02	21.59 ± 0.03	–
ZF-UDS-3651	34.289452	−5.2698030	24.54 ± 0.05	22.95 ± 0.03	0.64 ± 0.42
ZF-UDS-4347	34.290428	−5.2620687	24.90 ± 0.07	23.17 ± 0.03	−0.44 ± 0.45
ZF-UDS-6496	34.340358	−5.2412550	24.27 ± 0.05	22.62 ± 0.02	−0.04 ± 0.24
ZF-UDS-7329	34.255872	−5.2338210	23.90 ± 0.04	22.36 ± 0.01	–
ZF-UDS-7542	34.258888	−5.2322803	24.16 ± 0.04	23.02 ± 0.02	–
ZF-UDS-8197	34.293755	−5.2269478	24.77 ± 0.07	23.22 ± 0.03	–
3D-UDS-27939	34.487273	−5.1821730	24.69 ± 0.09	23.26 ± 0.07	1.57 ± 0.57
3D-UDS-35168	34.485131	−5.1578340	25.40 ± 0.13	23.89 ± 0.09	–
3D-UDS-39102	34.526210	−5.1438100	25.10 ± 0.18	23.28 ± 0.15	0.11 ± 0.56
3D-UDS-41232	34.526589	−5.1360390	22.90 ± 0.02	21.74 ± 0.01	0.09 ± 0.26

Notes. [a] IDs starting with "ZF" are from the ZFOURGE DR1 catalogs, and IDs starting with "3D" are from the 3DHST v4.1 catalogs. [b] Observed AB magnitudes. [c] Obtained at $\lambda = 744\,\mu$m (Band 9).

Table A.2. Galaxies targeted with MOSFIRE, properties from photometric redshift.

ID	z_{phot}	M_* $10^{11}\,M_\odot$	$U-V$ rest, AB	$V-J$ rest, AB
ZF-COS-10559	$3.34^{+0.30}_{-1.04}$	0.23	$1.57^{+0.75}_{-0.88}$	$1.05^{+0.89}_{-0.32}$
ZF-COS-14907	$2.89^{+0.06}_{-0.06}$	0.47	$1.38^{+0.08}_{-0.07}$	$0.76^{+0.04}_{-0.03}$
ZF-COS-17779	$3.91^{+0.35}_{-0.33}$	0.43	$1.28^{+0.23}_{-0.28}$	$0.63^{+0.16}_{-0.14}$
ZF-COS-18842	$3.47^{+0.07}_{-0.07}$	0.40	$1.20^{+0.07}_{-0.06}$	$0.69^{+0.02}_{-0.03}$
ZF-COS-19589	$3.73^{+0.16}_{-0.15}$	0.64	$1.66^{+0.17}_{-0.17}$	$0.77^{+0.11}_{-0.06}$
ZF-COS-20032	$3.55^{+0.69}_{-0.68}$	2.54	$1.92^{+1.13}_{-0.82}$	$1.76^{+0.53}_{-0.62}$
ZF-COS-20115	$3.64^{+0.08}_{-0.09}$	1.24	$1.71^{+0.11}_{-0.07}$	$0.53^{+0.02}_{-0.03}$
ZF-COS-20133	$3.51^{+0.04}_{-0.04}$	0.24	$1.56^{+0.04}_{-0.10}$	$0.69^{+0.01}_{-0.01}$
3D-EGS-18996	$2.99^{+0.03}_{-0.03}$	0.86	$1.55^{+0.04}_{-0.04}$	$0.30^{+0.01}_{-0.01}$
3D-EGS-26047	$3.24^{+0.08}_{-0.08}$	1.03	$1.25^{+0.12}_{-0.06}$	$1.21^{+0.03}_{-0.03}$
3D-EGS-27584	$3.60^{+0.18}_{-0.23}$	4.54	$2.25^{+0.29}_{-0.23}$	$1.33^{+0.16}_{-0.14}$
3D-EGS-31322	$3.47^{+0.07}_{-0.07}$	0.98	$1.30^{+0.06}_{-0.11}$	$0.79^{+0.09}_{-0.01}$
3D-EGS-34322	$3.59^{+0.33}_{-0.32}$	0.43	$1.25^{+0.51}_{-0.16}$	$0.95^{+0.05}_{-0.23}$
3D-EGS-40032	$3.22^{+0.08}_{-0.09}$	1.98	$1.51^{+0.13}_{-0.07}$	$0.76^{+0.02}_{-0.01}$
ZF-UDS-3651	$3.87^{+0.12}_{-0.12}$	0.77	$1.35^{+0.11}_{-0.25}$	$0.74^{+0.01}_{-0.01}$
ZF-UDS-4347	$3.58^{+0.04}_{-0.05}$	0.31	$1.54^{+0.06}_{-0.09}$	$0.81^{+0.08}_{-0.04}$
ZF-UDS-6496	$3.50^{+0.04}_{-0.04}$	0.79	$1.73^{+0.04}_{-0.02}$	$0.81^{+0.01}_{-0.02}$
ZF-UDS-7329	$3.04^{+0.17}_{-0.17}$	1.67	$2.05^{+0.28}_{-0.32}$	$1.16^{+0.13}_{-0.10}$
ZF-UDS-7542	$3.15^{+0.06}_{-0.06}$	0.70	$1.49^{+0.08}_{-0.05}$	$0.97^{+0.02}_{-0.01}$
ZF-UDS-8197	$3.47^{+0.06}_{-0.06}$	0.36	$1.29^{+0.08}_{-0.06}$	$0.94^{+0.03}_{-0.04}$
3D-UDS-27939	$3.22^{+0.20}_{-0.22}$	0.54	$1.46^{+0.36}_{-0.14}$	$0.96^{+0.01}_{-0.10}$
3D-UDS-35168	$3.46^{+0.32}_{-0.29}$	0.32	$1.37^{+0.35}_{-0.28}$	$0.81^{+0.11}_{-0.12}$
3D-UDS-39102	$3.51^{+0.42}_{-0.36}$	0.86	$1.61^{+0.51}_{-0.50}$	$1.32^{+0.11}_{-0.24}$
3D-UDS-41232	$3.01^{+0.07}_{-0.08}$	1.52	$1.56^{+0.16}_{-0.12}$	$0.82^{+0.04}_{-0.03}$

Table A.3. MOSFIRE observations of our targets.

ID	Integration time h		Uncertainty[a,b] 10^{-19} erg s^{-1} cm^{-2} Å^{-1}		Median S/N^b		Maximum S/N^b		Masks
	H	K	H	K	H	K	H	K	
ZF-COS-10559	–	1.6	–	0.50	–	0.7	–	4.3	Z245
ZF-COS-14907	–	3.3	–	0.54	–	3.6	–	9.9	Y259-A
ZF-COS-17779	3.9	7.2	0.40	0.24	1.1	3.1	5.1	6.7	W182
ZF-COS-18842	0.3	3.6	0.89	0.36	0.7	3.9	3.5	8.7	U069
ZF-COS-19589	3.9	7.2	0.42	0.25	0.8	3.7	3.2	7.9	W182
ZF-COS-20032	3.9	7.2	0.84	0.31	1.5	1.5	3.4	6.8	W182
ZF-COS-20115	4.2	14.4	0.47	0.21	1.1	12.2	3.9	22.9	W182, Y259-B, U069, Z245
ZF-COS-20133	3.9	7.2	0.38	0.29	0.4	2.0	3.2	19.7	W182
3D-EGS-18996	0.8	4.8	0.72	0.54	6.7	10.1	11.3	21.1	W057
3D-EGS-26047	0.8	4.8	0.70	0.83	2.1	2.8	6.3	5.9	W057
3D-EGS-27584	0.8	4.8	–	0.86	–	3.5	–	6.0	W057
3D-EGS-31322	0.8	4.8	0.50	0.51	3.9	7.4	10.1	15.4	W057
3D-EGS-34322	0.8	4.8	0.63	0.72	1.1	1.4	3.8	5.7	W057
3D-EGS-40032	0.8	4.8	0.50	0.52	7.1	11.0	16.0	19.3	W057
ZF-UDS-3651	0.3	7.3	0.87	0.39	1.5	3.6	4.3	8.9	W182, Y259-A
ZF-UDS-4347	0.3	2.4	0.89	0.52	1.3	2.6	4.2	6.8	W182
ZF-UDS-6496	–	4.9	–	0.52	–	3.6	–	8.6	Y259-A
ZF-UDS-7329	–	9.6	–	0.35	–	7.2	–	17.0	Y259-A, U069
ZF-UDS-7542	0.3	7.3	0.93	0.34	1.4	4.2	3.8	9.1	W182, Y259-A
ZF-UDS-8197	0.3	7.3	0.90	0.38	0.7	2.4	3.6	16.0	W182, Y259-A
3D-UDS-27939	–	4.0	–	0.71	–	3.2	–	15.8	Y259-B
3D-UDS-35168	–	4.0	–	0.50	–	1.0	–	4.9	Y259-B
3D-UDS-39102	–	4.0	–	0.71	–	0.8	–	4.3	Y259-B
3D-UDS-41232	–	4.0	–	0.60	–	7.9	–	17.3	Y259-B

Notes. [a] Median uncertainty of the MOSFIRE spectrum (1σ), accounting for slit loss. [b] For a binning of 70 Å.

Table A.4. Spectroscopic identification of our targets.

ID	z_{phot}	z_{spec}	$\dfrac{z_{phot} - z_{spec}}{\Delta z}$	p^a	χ^2_{red}	Balmer abs.[b]	Emission lines[c]	$\sigma_v{}^d$ km s^{-1}
Robust z_{spec}								
ZF-COS-20032	$3.55^{+0.69}_{-0.68}$	$2.4736^{+0.0008}_{-0.0010}$	1.6	96%	1.0	–	[O I]$_{6300}$ Hα [N II]$_{\lambda\lambda}$	217 ± 161
ZF-COS-20115	$3.64^{+0.08}_{-0.09}$	$3.7145^{+0.0015}_{-0.0015}$	−0.9	100%	1.0	Yes	–	–
ZF-COS-20133	$3.51^{+0.04}_{-0.04}$	$3.4806^{+0.0002}_{-0.0002}$	0.6	100%	1.0	–	Hβ [O III]$_{\lambda\lambda}$	60 ± 26
3D-EGS-18996	$2.99^{+0.03}_{-0.03}$	$3.2390^{+0.0007}_{-0.0009}$	−8.3	100%	1.0	Yes	[O III]$_{\lambda\lambda}$	530 ± 199
3D-EGS-26047	$3.24^{+0.08}_{-0.08}$	$3.2337^{+0.0020}_{-0.0016}$	0.0	99%	1.0	–	[O II]$_{\lambda\lambda}$ Hβ [O III]$_{\lambda\lambda}$	530 ± 263
3D-EGS-40032	$3.22^{+0.08}_{-0.09}$	$3.2187^{+0.0011}_{-0.0013}$	−0.0	100%	1.0	Yes	[O II]$_{\lambda\lambda}$	582 ± 236
ZF-UDS-8197	$3.47^{+0.06}_{-0.06}$	$3.5431^{+0.0007}_{-0.0010}$	−1.2	99%	1.0	–	[O III]$_{\lambda\lambda}$	530 ± 53
3D-UDS-27939	$3.22^{+0.20}_{-0.22}$	$2.2104^{+0.0004}_{-0.0004}$	4.5	100%	1.0	–	[O I]$_{6300}$ Hα [N II]$_{\lambda\lambda}$ [S II]$_{\lambda\lambda}$	154 ± 7
Uncertain z_{spec}								
ZF-COS-17779	$3.91^{+0.35}_{-0.33}$	$3.4150^{+0.1320}_{-0.0003}$	1.4	77%	1.0	–	Hβ [O III]$_{\lambda\lambda}$	60 ± 26
ZF-COS-18842	$3.47^{+0.07}_{-0.07}$	$3.7823^{+0.0023}_{-0.0031}$	−4.5	75%	1.1	–	[O III]$_{\lambda\lambda}$	60 ± 279
ZF-COS-19589	$3.73^{+0.16}_{-0.15}$	$3.7152^{+0.0094}_{-0.1589}$	0.1	32%	0.9	–	–	–
3D-EGS-31322	$3.47^{+0.07}_{-0.07}$	$3.4344^{+0.0104}_{-0.0029}$	0.5	84%	1.0	Yes	–	–
Rejected z_{spec}								
ZF-COS-10559	$3.34^{+0.30}_{-1.04}$	$2.6376^{+1.6758}_{-0.2266}$	0.4	2%	0.9	–	–	–
ZF-COS-14907	$2.89^{+0.06}_{-0.06}$	$4.1935^{+0.1241}_{-0.0030}$	−23.1	63%	1.2	–	–	–
3D-EGS-27584	$3.60^{+0.18}_{-0.23}$	$2.4620^{+0.1355}_{-0.3406}$	4.3	9%	1.0	–	–	–
3D-EGS-34322	$3.59^{+0.33}_{-0.32}$	$3.5460^{+0.1150}_{-0.3692}$	0.1	18%	0.9	–	–	–
ZF-UDS-3651	$3.87^{+0.12}_{-0.12}$	$3.4214^{+0.1193}_{-0.0600}$	2.7	33%	1.1	–	–	–
ZF-UDS-4347	$3.58^{+0.04}_{-0.05}$	$3.4731^{+0.2220}_{-0.0186}$	0.5	39%	0.9	–	–	–
ZF-UDS-6496	$3.50^{+0.04}_{-0.04}$	$2.0332^{+1.1110}_{-0.0003}$	1.3	33%	1.0	–	–	–
ZF-UDS-7329	$3.04^{+0.17}_{-0.17}$	$3.1857^{+0.0188}_{-1.0035}$	−0.1	25%	1.0	–	–	–
ZF-UDS-7542	$3.15^{+0.06}_{-0.06}$	$3.1885^{+0.2709}_{-0.0029}$	−0.6	35%	1.0	–	–	–
3D-UDS-35168	$3.46^{+0.32}_{-0.29}$	$4.2240^{+0.2913}_{-1.5226}$	−0.5	4%	1.0	–	–	–
3D-UDS-39102	$3.51^{+0.42}_{-0.36}$	$3.3982^{+0.3368}_{-1.1912}$	0.2	4%	1.0	–	–	–
3D-UDS-41232	$3.01^{+0.07}_{-0.08}$	$2.9810^{+0.1448}_{-0.5831}$	0.2	25%	1.0	–	–	–

Notes. This table is also available in electronic format at the CDS. [a] Fraction of the redshift probability distribution enclosed within $z_{spec} \pm 0.01$. [b] Indicates galaxies for which the redshift could be determined with $p > 50\%$ from the continuum emission alone, using Balmer absorption lines. [c] List of emission lines detected with a significance of at least 2σ. Line doublets names are shortened as in Table A.5. [d] Best-fit velocity dispersion of the emission lines (assumed identical for all the lines in a spectrum).

Table A.5. Measured emission line properties.

ID	$[\text{O\,II}]_{\lambda\lambda}$	Hβ	$[\text{O\,III}]_{\lambda\lambda}$	$[\text{O\,I}]_{6300}$	Hα	$[\text{N\,II}]_{\lambda\lambda}$	$[\text{S\,II}]_{\lambda\lambda}$
Luminosities ($10^8\ L_\odot$)							
ZF-COS-17779	0.47 ± 0.27	0.62 ± 0.14	1.51 ± 0.21	–	–	–	–
ZF-COS-18842	–	0.66 ± 1.63	6.19 ± 4.14	–	–	–	–
ZF-COS-19589	–	0.70 ± 1.02	–	–	–	–	–
ZF-COS-20032	–	0.19 ± 0.50	2.06 ± 1.36	0.52 ± 0.23	2.05 ± 0.77	0.94 ± 0.45	0.12 ± 0.27
ZF-COS-20115	–	0.00 ± 0.33	1.86 ± 1.43	–	–	–	–
ZF-COS-20133	–	2.05 ± 0.21	14.31 ± 0.31	–	–	–	–
3D-EGS-18996	1.61 ± 0.88	0.00 ± 0.35	7.14 ± 1.46	–	–	–	–
3D-EGS-26047	3.04 ± 1.16	6.61 ± 3.12	8.44 ± 3.97	–	–	–	–
3D-EGS-31322	1.55 ± 0.99	2.18 ± 1.60	0.82 ± 1.14	–	–	–	–
3D-EGS-40032	4.59 ± 1.04	1.02 ± 0.99	1.13 ± 1.29	–	–	–	–
ZF-UDS-8197	–	0.74 ± 0.74	18.26 ± 1.49	–	–	–	–
3D-UDS-27939	–	–	–	0.84 ± 0.31	6.69 ± 0.46	2.68 ± 0.33	3.96 ± 0.47
Rest-frame equivalent widths (Å)							
ZF-COS-17779	13.4 ± 7.9	7.5 ± 1.8	21.0 ± 3.1	–	–	–	–
ZF-COS-18842	–	3.0 ± 7.6	32.8 ± 23.2	–	–	–	–
ZF-COS-19589	–	5.2 ± 7.6	–	–	–	–	–
ZF-COS-20032	–	12.9 ± 34.1	136.6 ± 87.3	22.3 ± 10.1	84.1 ± 31.2	40.2 ± 19.2	5.2 ± 11.3
ZF-COS-20115	–	0.0 ± 0.9	5.8 ± 4.5	–	–	–	–
ZF-COS-20133	–	34.3 ± 4.2	282.2 ± 19.4	–	–	–	–
3D-EGS-18996	6.7 ± 3.7	0.0 ± 0.6	14.2 ± 3.0	–	–	–	–
3D-EGS-26047	42.8 ± 16.5	33.9 ± 16.7	49.2 ± 24.2	–	–	–	–
3D-EGS-31322	9.6 ± 6.1	5.0 ± 3.7	2.2 ± 3.1	–	–	–	–
3D-EGS-40032	23.0 ± 5.3	1.8 ± 1.8	2.3 ± 2.7	–	–	–	–
ZF-UDS-8197	–	5.7 ± 5.8	165.7 ± 16.4	–	–	–	–
3D-UDS-27939	–	–	–	13.8 ± 5.3	108.5 ± 9.3	43.1 ± 6.0	62.2 ± 8.4

Notes. For doublets, luminosities and EW are the sum of the two lines in the doublet. $[\text{O\,II}]_{\lambda\lambda}$ is the sum of $[\text{O\,II}]_{3729}$ and $[\text{O\,II}]_{3726}$. $[\text{O\,III}]_{\lambda\lambda}$ is the sum of $[\text{O\,III}]_{5007}$ and $[\text{O\,II}]_{4959}$. $[\text{N\,II}]_{\lambda\lambda}$ is the sum of $[\text{N\,II}]_{6583}$ and $[\text{N\,II}]_{6548}$. $[\text{S\,II}]_{\lambda\lambda}$ is the sum of $[\text{S\,II}]_{6731}$ and $[\text{S\,II}]_{6716}$.

Table A.6. Summary of SFR estimates from SED modeling or emission lines.

ID	$\mathrm{SFR_{10}}^{a}$ $M_\odot\,\mathrm{yr}^{-1}$	Line	$\mathrm{SFR_{line}^{nodust}}^{b}$ $M_\odot\,\mathrm{yr}^{-1}$	$\mathrm{SFR_{line}^{corr}}^{c}$ $M_\odot\,\mathrm{yr}^{-1}$
Confirmed $z > 3$ galaxies				
ZF-COS-17779	$0.66^{+100.77}_{-0.66}$	Hβ	3.4 ± 0.8	$15.5^{+21.3}_{-10.6}$
...	...	[O II]	1.1 ± 0.6	$8.7^{+24.8}_{-7.6}$
ZF-COS-18842	$2.53^{+2.36}_{-1.21}$	Hβ	3.6 ± 8.9	$3.6^{+12.0}_{-8.9}$
...	...	[O II]	0.0 ± 1.9	$0.0^{+2.6}_{-1.9}$
ZF-COS-19589	$0.00^{+4.41}_{-0.00}$	Hβ	3.8 ± 5.6	$10.2^{+28.5}_{-12.9}$
ZF-COS-20115	$0.00^{+0.69}_{-0.00}$	Hβ	0.0 ± 1.8	$0.0^{+2.8}_{-2.5}$
...	...	[O II]	0.0 ± 1.6	$0.0^{+2.9}_{-2.5}$
ZF-COS-20133	$4.57^{+0.09}_{-1.30}$	Hβ	11.2 ± 1.1	$12.5^{+1.3}_{-2.4}$
...	...	[O II]	0.0 ± 0.8	$0.0^{+0.9}_{-0.8}$
3D-EGS-18996	$0.88^{+0.91}_{-0.86}$	Hβ	0.0 ± 1.9	$0.0^{+1.9}_{-1.9}$
...	...	[O II]	3.6 ± 2.0	$3.6^{+2.0}_{-2.0}$
3D-EGS-26047	$0.07^{+1.95}_{-0.04}$	Hβ	36.0 ± 17.0	$44.8^{+37.3}_{-25.8}$
...	...	[O II]	6.8 ± 2.6	$9.2^{+7.9}_{-5.0}$
3D-EGS-31322	$0.04^{+2.26}_{-0.04}$	Hβ	11.9 ± 8.8	$16.5^{+19.1}_{-13.0}$
...	...	[O II]	3.6 ± 2.3	$5.7^{+6.8}_{-4.1}$
3D-EGS-40032	$5.50^{+3.38}_{-3.07}$	Hβ	5.6 ± 5.4	$8.6^{+10.3}_{-8.4}$
...	...	[O II]	10.3 ± 2.3	$18.6^{+7.9}_{-6.2}$
ZF-UDS-8197	$1.68^{+3.33}_{-0.73}$	Hβ	4.0 ± 4.1	$4.0^{+7.2}_{-4.1}$
...	...	[O II]	0.0 ± 0.2	$0.0^{+0.4}_{-0.2}$
Interlopers at $z < 3$				
ZF-COS-20032	$139.80^{+75.78}_{-139.80}$	Hα	3.9 ± 1.5	$63.6^{+46.0}_{-29.4}$
...	...	Hβ	1.0 ± 2.7	$58.1^{+238.9}_{-136.8}$
3D-UDS-27939	$1.81^{+7.97}_{-1.16}$	Hα	12.8 ± 0.9	$27.1^{+18.4}_{-5.4}$

Notes. [a] SFR estimated from the SED modeling, averaged over the last 10 Myr preceeding observation. [b] SFR estimated from the observed emission line luminosity, without correcting for dust attenuation. [c] Same, but corrected for dust attenuation using the best-fit A_V from the SED modeling.

Appendix B: Reduction of MOSFIRE spectra

B.1. Transmission correction

As in Nanayakkara et al. (2016), a standard star was observed at the beginning and end of the observing runs to monitor telluric absorption. Since this star was not observed at the same time as the science targets, it cannot capture variations of the effective transmission within the run, and it is affected by different systematic errors (e.g., in the background subtraction).

Each of our masks contained a "slit star" of moderate brightness, which was used mostly to monitor the seeing. These are typically M or K stars, which are not optimal flux calibrators because their continuum emission contains a variety of features which need to be properly modeled. However these features are relatively weak in H and K, and we can model them using theoretical stellar spectra. We therefore used these "slit stars" to perform an independent telluric correction. Our approach, described below, generates an effective "transmission correction" which accounts for telluric absorption, filter transmission, slit losses for a point source, and absolute flux calibration (electron/s to cgs).

For each mask, we first fit the broadband photometry of the slit star with the PHOENIX theoretical star models (Husser et al. 2013) to estimate its intrinsic spectrum. We performed this fit only using the NIR photometry ($0.8\,\mu m < \lambda < 3\,\mu m$) to maximize the fidelity of the fit in the H and K bands. Because the stars are not extremely bright, this photometry is not saturated and can be used reliably. The best-fitting PHOENIX spectrum, normalized to fit the photometry, was used as the intrinsic spectrum of the star. We then used the MOSFIRE pipeline to reduce each pair of "AB" exposures into 2D spectra for all our targets, including the slit star (see next section for detail). At this stage, no transmission correction was applied, and the spectra were still in raw units. We collapsed the slit star's 2D spectrum along the wavelength axis to form the spatial profile of the star, which we fit with a Gaussian model to determine the location of the peak emission, as well as the seeing. We extracted the 1D spectrum of the star using this Gaussian model, normalizing it to the emission in the slit at each wavelength element while keeping the position and width of the Gaussian fixed. We finally estimated the transmission correction by computing the ratio of the intrinsic spectrum of the star to this 1D spectrum. Because the intrinsic spectrum was rescaled to match the broadband photometry, this correction also includes the conversion from raw units to physical flux, including slit loss correction for a point source.

A downside of using the slit star is that its S/N is not as high a that of the standard star. The derived transmission corrections are more noisy, and this can degrade the final S/N of the science targets. To mitigate this, we first applied a 3 pixel boxcar filter to the star spectra, then stacked the transmission correction of all exposures to build a "template" curve with much higher S/N. We then fit this template to the transmission correction of each exposure. To allow the strength of the atmospheric absorption features to vary from one exposure to the other, we decomposed the template into multiple wavelength "chunks" – each containing one major atmospheric feature – and let their amplitude vary independently over a fixed baseline. For each exposure, we adopted the best-fitting combination of these chunks as the final transmission correction. An example fit is shown in Fig. B.1, and more detail on this procedure are provided in Appendix B.5. We then extended each curve to cover wavelengths beyond that covered by the slit star's spectrum using the average atmospheric transmission at Mauna Kea[7], and thus built the final transmission correction $T_i(\lambda)$ of each exposure i.

We find that the baseline of these corrections (which reflects the overall transmission across the band) shows variations of order $10-20\%$ between exposures. Factoring out these global fluctuations, the residual wavelength-dependent transmission variations are smaller: of the order of 1% and 3% in H and K, respectively, and reaching a maximum of 4% and 10% at the position of strong telluric absorption or toward the edges of the spectrum (see insets in Fig. B.1).

B.2. Extracting spectra

Before extracting 1D spectra, we removed hot pixels and cosmic rays by flagging strong pixel outliers from the 2D spectra produced by the pipeline. While the MOSFIRE pipeline does include a cosmic ray rejection algorithm, this feature can only be enabled when reducing all the exposures at once. In our case, where

---

[7] https://www.gemini.edu/sciops/telescopes-and-sites/observing-condition-constraints/ir-transmission-spectra
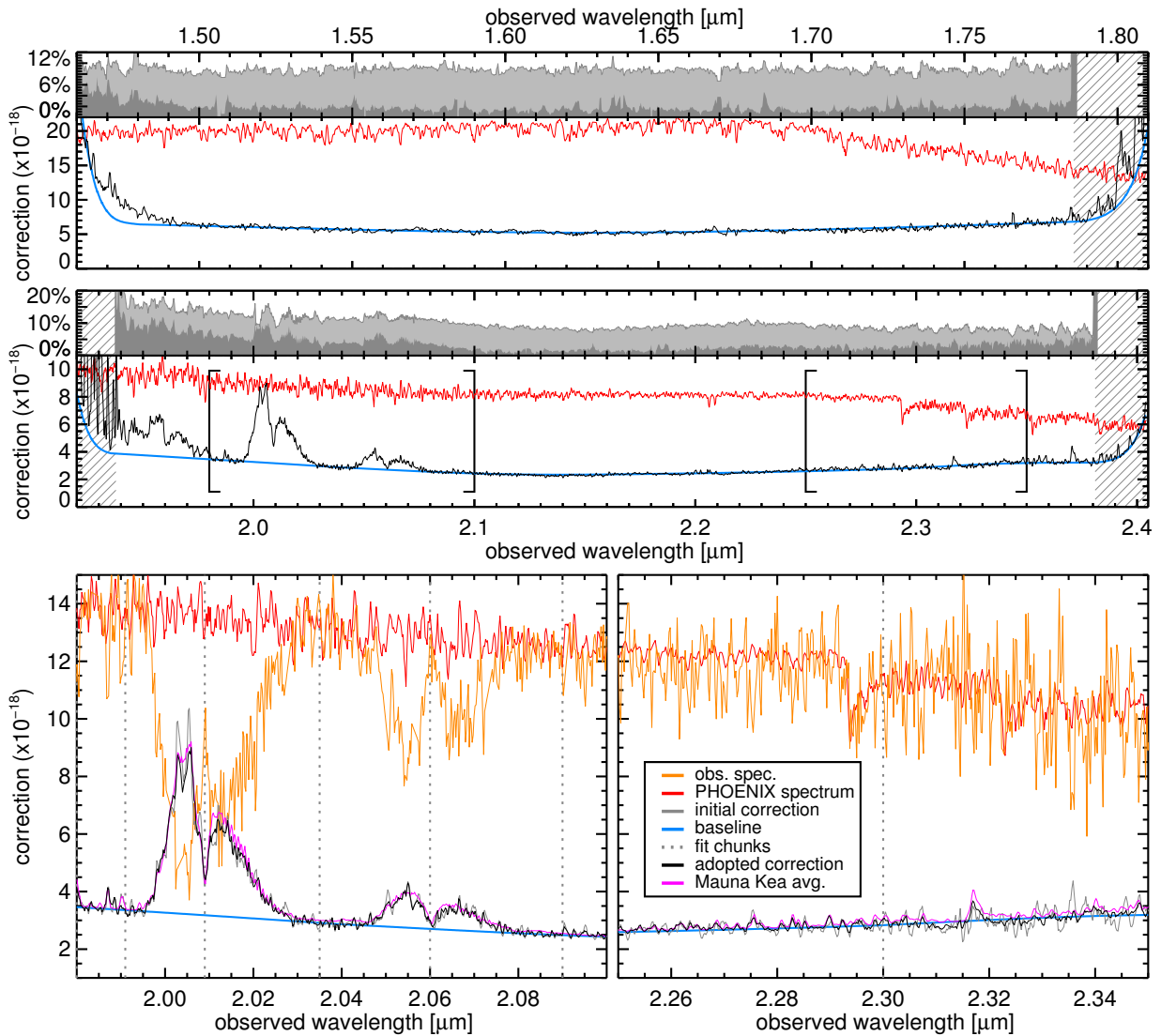
Fig. B.1. Example transmission correction curve for one exposure of the COSMOS run using the slit star (an M 5 star). As discussed in the text, this correction accounts for the telluric absorption, the filter transmission, slit loss (for a point source), and absolute calibration from electron/s to flux. *Top*: adopted correction curve in *H* (*top*) and *K* (*bottom*) for this exposure (black line), assumed intrinsic spectrum of the slit star (red line) and adopted transmission "baseline" (blue line, see text). The hashed regions indicate the portion of the spectrum for which we used the Mauna Kea average transmission. The inset at the top of each plot shows the relative variation of the transmission among all exposures, showing either the full variation between exposures (light gray) or what remains after factoring out variations of global transmission (dark gray). *Bottom*: zoom-ins on the *K*-band correction curve (areas bracketed in the plot above). In addition to the lines shown above, we show the raw observed spectrum multiplied by the baseline correction (orange line), the initial correction derived from this spectrum (light gray line), the individual wavelength chunks of the template curve (dotted vertical lines) and the expected correction based on the average Mauna Kea transmission (see text).

we reduced each "AB" exposure separately, we used a different approach which is described and illustrated in Appendix B.6. We also systematically masked strong OH residuals.

For each galaxy, we then stacked the 2D spectra with uniform weighting and used it to identify the position within the slit of the target, either from the continuum emission (as for the slit star) or using clearly detected emission lines. Compared to the position predicted by the pipeline, we find systematic offsets within each masks of up to two pixels (a pixel is 0.18″ wide), and residual per-target offsets of up to three pixels (the RMS ranges from 0.5 to 1.5 pixels, depending on the mask). For sources without detected continuum or lines, we used the average shift within their respective mask.

Using these positions, we extracted 1D spectra for our targets on each pair of "AB" exposures by fitting the amplitude of

a fixed Gaussian profile along the slit axis, for each resolution element. We set the width of this Gaussian to that measured for the slit star for each exposure, which is equivalent to assuming that the targets are point-like. This is a reasonable assumption given that our galaxies are both distant and intrinsically compact objects (Straatman et al. 2015), and potential flux loss resulting from this choice are accounted for at a later stage when the spectra are rescaled to match the broadband photometry.

For each target t and exposure i, this produced a counts spectrum $\mathrm{ADU}_{t,i}(\lambda)$. We then applied the transmission corrections (described in the previous section) to all these 1D spectra to obtain the total flux, $F_{t,i}(\lambda) = \mathrm{ADU}_{t,i}(\lambda) \times T_i(\lambda)$. We also derived the formal uncertainty $\sigma_{t,i}^{\mathrm{th}}(\lambda)$ on this flux by propagating the uncertainties from the 2D error spectrum produced by the pipeline (see next section for more detail on the uncertainty estimates).
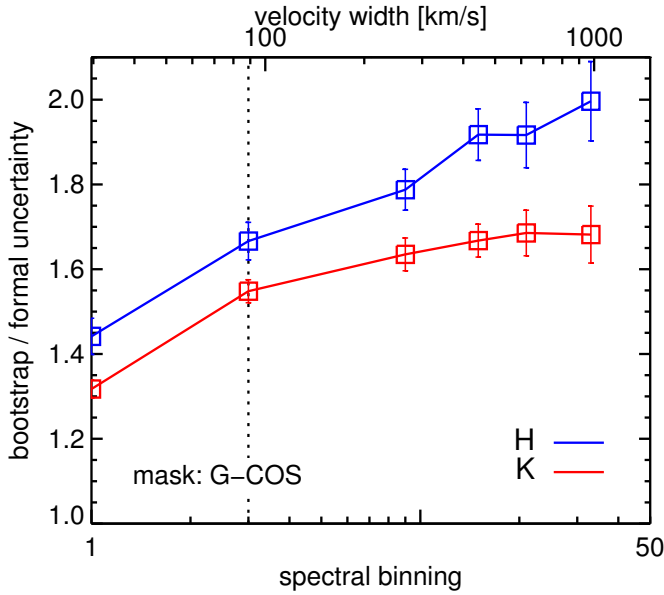
Fig. B.2. Ratio of uncertainties derived by bootstrapping the 1D spectra against that obtained by formally propagating the uncertainties from the 2D spectra and assuming uncorrelated Gaussian noise. The values shown are the median of the ratios among all the targets within the mask (COS-W182) and error bars show the error on the median. We show this ratio for the two bands (H in blue, K in red) as a function of spectral binning. The dashed line indicates the binning adopted in this work.

Then, for each target t the final spectrum $F_t(\lambda)$ was obtained by stacking the 1D spectra of all exposures using inverse variance weighting, $w_{t,i}(\lambda) = 1/\sigma_{t,i}^{th}(\lambda)^2$, and sigma clipping to remove any remaining strong outlier (exposures differing from the median by more than 5σ were given a weight of zero). This weighting scheme optimally penalizes exposures with poor observing conditions, in particular in case of bad seeing: the width of the Gaussian profile used for the flux extraction is larger, which results in larger formal uncertainties in the extraction. Furthermore, the transmission correction is also larger for these exposures because of larger slit losses, which contributes in weighting them down.

We also stacked the 2D spectra for visualization purposes, using the same weighting, flagging and telluric correction as for the 1D spectra. In case of small offsets along the slit between exposures, the spectra were shifted to a common grid before stacking.

B.3. Determining uncertainties

We determined the uncertainties on the final stacked spectra in two ways. A first "theoretical" value, $\sigma_{th,t}(\lambda)$, was obtained for each target t by propagating the formal uncertainties of each exposure in the stack, as derived originally from the pipeline's 2D error spectrum, namely, combining Poisson statistics on the measured electron counts with the detector read noise. As an alternative to this first method, we empirically estimated the uncertainties from the variance between exposures, $\sigma_{var,t}$, using:

$$\sigma_{var,t}(\lambda) = \frac{1}{\sum_i w_{t,i}(\lambda)} \sqrt{\frac{N_{exp}}{N_{exp}-1} \sum_i w_{t,i}(\lambda)^2 \left[F_{t,i}(\lambda) - F_t(\lambda)\right]^2},$$
(B.1)

where N_{exp} is the number of exposures; the other quantities were defined in the previous section. This formula produces results
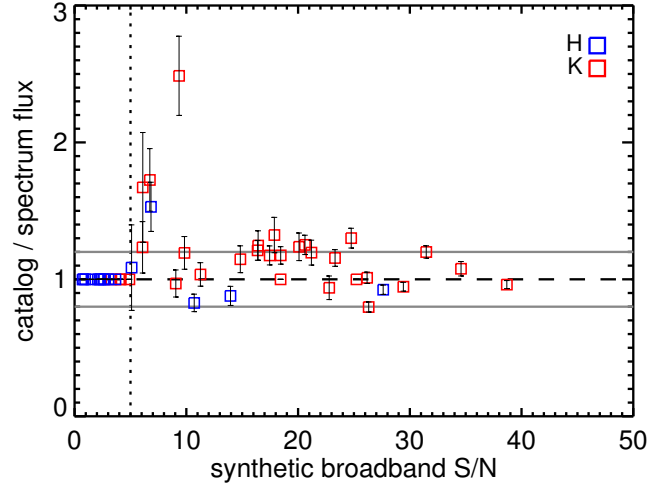

Fig. B.3. Ratio of the flux listed in the ZFOURGE or 3DHST catalog to the flux measured in the MOSFIRE spectrum for each of our targets, as a function of the S/N of the spectrum flux. The flux in the spectrum was integrated in the same passband as the catalog fluxes. Fluxes in H band are shown in blue, and fluxes in K band are shown in red. The vertical dotted line shows the S/N limit below which we did not perform the flux rescaling to avoid introducing noise.

identical to bootstrapping (as demonstrated in Gatz & Smith 1995, and we double checked that it was indeed the case on our data) but it requires much less computing time.

Using the native pixel scale of the spectra (i.e., without binning), we find that σ_{var} is systematically larger than σ_{th} by $30-45\%$, depending on the mask and band as shown in Fig. B.2 (left). This suggests the uncertainties produced by the pipeline are substantially underestimated. In fact, the pipeline version we used was indeed reported to employ an incorrect treatment of the error spectra, which could be the source of this discrepancy[8]. Using their own MOSFIRE pipeline (and a slightly different observing strategy), Kriek et al. (2015) found no such issue when performing a similar test on their spectra. However, their formula for σ_{var} (their Eqs. (5) and (6)) also differs from ours, making any direct comparison difficult. Using MC simulations of a weighted mean, we find that their formula actually overpredicts the uncertainty by $20-50\%$, depending on the choice of weights, while our formula is accurate to 1%.

We also find the discrepancy between σ_{var} and σ_{th} grows even larger for binned spectra: using a binning of three (resp. nine) wavelength elements, the bootstrapped uncertainties are an additional $12-18\%$ (resp. $18-25\%$) larger than the formal uncertainties. This suggests that, at the native pixel scale produced by the pipeline ($\lambda/\Delta\lambda = 9000-11\,000$), the noise is spectrally correlated. Since we did not require high spectral resolution for this work, we chose to avoid the native pixel scale in the following, and considered instead a binning of at least three spectral elements corresponding to $\lambda/\Delta\lambda \sim 3000$ (or resolution elements of $\geq 80\,km\,s^{-1}$), which matched the spectral resolution of MOSFIRE with 0.7″ slits, and ensured that our uncertainties were accurate within 20% at worst. Consequently, the line spread function was reduced to little more than a pixel, and was thus ignored.

While σ_{var} is larger than σ_{th} on average, it is not always the case for every spectral element of a spectrum. In particular

---

[8] See https://github.com/Keck-DataReductionPipelines/MosfireDRP/issues/40

for masks with few exposures, the noise fluctuations can make the data of each exposure coincide (by pure chance) to similar flux values, leading to an underestimated σ_{var}. To avoid this, we defined a "rescaled" formal uncertainty

$$\bar{\sigma}_{\mathrm{th},t}(\lambda) = \sigma_{\mathrm{th},t}(\lambda) \left\langle \frac{\sigma_{\mathrm{var},t}}{\sigma_{\mathrm{th},t}} \right\rangle_{\lambda}, \qquad (\mathrm{B.2})$$

where the second term is the median of $\sigma_{\mathrm{var}}/\sigma_{\mathrm{th}}$ across the spectrum of the target t. We then adopted as final uncertainty the largest of σ_{var} and $\bar{\sigma}_{\mathrm{th}}$.

As we have just shown, bootstrapping uncertainties are more conservative than the formal uncertainties in general. However, if some source of error affects all the exposures in a similar way, for example owing to an imperfect background subtraction, they will not be accounted for by bootstrapping by construction. To verify that our data were globally unaffected by such issues, we reduced 14 "sky" spectra in the COSMOS mask, extracted at offset positions in the slits, avoiding known sources. If our reduction procedure was unbiased, these spectra should only contain noise with a zero flux average. We therefore stacked these spectra, aligned on the same wavelength grid, and find reduced $\chi^2 = 1.2$ and 1.7 in H and K, respectively, for a binning of 15 resolution elements ($500\,\mathrm{km\,s^{-1}}$, the expected full line width for our objects). These residuals were not perfect, as would be indicated by a reduced χ^2 of unity, but they were still small: at this resolution the largest residuals were of the order of $3 \times 10^{-20}\,\mathrm{erg\,s^{-1}\,cm^{-2}\,\mathring{A}^{-1}}$, which is lower than the error bar of our deepest spectra. We therefore concluded that our spectra were not affected by background subtraction issues, and that the bootstrapping uncertainties derived above accounted for the main sources of error in the reduction.

B.4. Rescaling to total flux

Figure B.3 illustrates the final rescaling of the spectra using the catalog fluxes for each source.

B.5. Detail of telluric correction

Here we describe in more detail the procedure through which the telluric correction of each exposure was fit with a template curve to improve the S/N.

We first defined the "initial" observed correction (i.e., inverse transmission) for the ith exposure as

$$T_i^{\mathrm{init}}(\lambda) = \frac{S_\lambda(\lambda)\,[\mathrm{cgs}]}{\mathrm{ADU}(\lambda)\,[e^-/\mathrm{s}]}, \qquad (\mathrm{B.3})$$

where S_λ is the physical flux of the star (taken from the PHOENIX synthetic spectrum and renormalized to the broadband fluxes of the star) and ADU is the raw electron flux measured by the MOSFIRE detector. This ratio could be used directly to perform the telluric correction of the science targets for the corresponding exposure, but since the slit star is only of moderate brightness, it can be undesirably noisy.

We therefore attempted to reduce the noise. To do so, we assumed the above correction can be empirically decomposed as

$$T_i(\lambda) = B_i(\lambda) \times \left[1 + \sum_{k} \alpha_{ik} \times A_k(\lambda) \right], \qquad (\mathrm{B.4})$$

where B_i is a strictly positive and smooth baseline function that captures the overall instrument transmission, A_k is a function
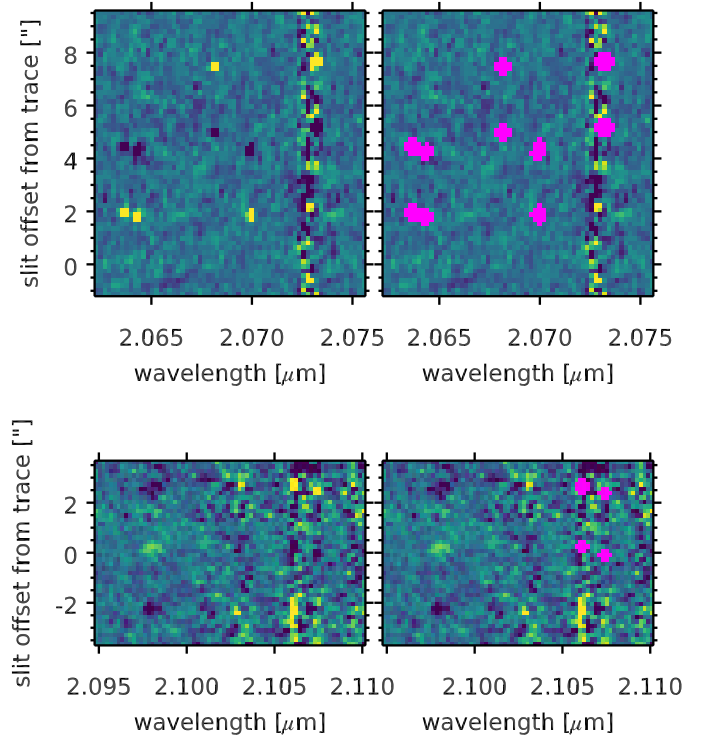
Fig. B.4. Examples of the output of the cosmic ray and hot pixel rejection algorithm. The original unmasked 2D spectra are shown in the *left*, and the masked spectra are shown on the *right* with masked regions displayed in pink. *Top*: excerpt from the brightest $z \sim 4$ quiescent galaxy; *bottom*: excerpt from a $z \sim 2$ filler SFG, which has an emission line at $\lambda = 2.098\,\mu\mathrm{m}$. Both spectra are taken from the same K-band exposure in the COSMOS mask, with excellent seeing ($0.42''$).

Table B.1. Wavelength chunks used to define the A_k functions in the telluric correction.

λ_k^{cut} (μm)	
H	K
1.450	1.900
1.565	1.951
1.585	1.961
1.615	1.980
1.660	1.991
1.670	2.009
1.850	2.035
	2.060
	2.090
	2.200
	2.300
	2.450

capturing the kth "chunk" of atmospheric absorption features, and α_{ik} is a strictly positive number that defines the strength of the kth set of features in the ith exposure. We obtained high S/N versions of the A_k functions by stacking all exposures, as described below.

For each exposure, we started by determining the baseline B_i. We first multiplied T_i^{init} by the response curve of the MOSFIRE filter to account for the sharp drop of transmission at the edges, and then isolated several disjoint wavelength regions within the band that are free of strong atmospheric feature. We computed

the average correction within each of these regions, and interpolated these values using a cubic spline with natural boundary conditions (vanishing second derivatives). We eventually defined the baseline as the product of this spline with the inverse of the filter response curve.

We then stacked T_i^{init}/B_i among all exposures to produce the high S/N template curve $\bar{T}$. We used the inverse of the Mauna Kea average transmission to fill the gaps in $\bar{T}$ which are not covered by the star's spectrum (at the edges of the H and K bands). We then used this template to define A_k as

$$A_{\mathrm{k}}(\lambda) = \begin{cases} \bar{T}(\lambda) - 1 & \text{for } \lambda_{\mathrm{k}}^{\mathrm{cut}} \leq \lambda < \lambda_{k+1}^{\mathrm{cut}} \\ 0 & \text{otherwise.} \end{cases} \tag{B.5}$$

The values of λ^{cut} in both H and K are given in Table B.1 and are illustrated in Fig. B.1 (bottom).

As a last step, for each exposure i we determined the values of α_{ik} by performing a linear fit of the A_k functions to $(T_i^{\mathrm{init}}/B_i - 1)$. If some of the α_{ik} came out negative from the fit, we considered that the S/N was too low and fixed them to $\alpha_{ik} = 1$ (i.e., assume the average transmission from the stack).

B.6. Cosmic ray and hot pixel rejection

In this section we describe the algorithm used for cosmic ray and hot pixel rejection in the 2D spectra produced by the MOSFIRE pipeline.

For each pixel p in the 2D spectrum, we built two lists of neighboring pixels: a first list containing the 12 neighboring pixels along the wavelength axis (corresponding to $\Delta\lambda \simeq 20\,\text{Å}$), and a second list containing all the neighboring pixels along the slit axis. We then computed the median value $\bar{p}_{\mathrm{wave}}$ among the first list, and computed the scatter of the values in the two lists σ_{wave} and σ_{slit} (respectively) using the median absolute deviation, multiplied by 1.48 to get a robust standard deviation.

The median $\bar{p}_{\mathrm{wave}}$ measures the emission in the spectrum on scales larger or equal to $20\,\text{Å}$ (i.e., continuum emission or relatively wide emission lines). The two scatter values, σ_{wave} and σ_{slit}, are used to estimate the expected noise amplitude for the pixel p without relying on the uncertainty spectrum produced by the pipeline (which, as we discuss in Sect. B.3, can be significantly underestimated and would cause many spurious hot pixel or cosmic ray rejections). The scatter σ_{slit} will capture increased noise at the wavelengths affected by OH lines, while σ_{wave} will be larger if the target shows detectable emission (which will result in increased Poisson noise). We therefore kept the highest value of the two as the best noise estimate, σ, and finally flagged out the pixel if $(p - \bar{p}_{\mathrm{wave}})/\sigma > 4$.

The result of this filtering is shown in Fig. B.4 for the K-band exposure in COSMOS with the best seeing (i.e., where the risk of filtering out features of the science target is the highest). We show two examples: the brightest $z \sim 4$ quiescent galaxy and a $z \sim 2$ low-mass filler with an emission line. In the first case, the algorithm correctly identified a number of hot pixels, even within the residuals of an OH line. The rest of the spectrum appears visually well-behaved. In the second case, a few hot pixels were also identified, but the emission line of the target galaxy at $\lambda = 2.098\,\mu\mathrm{m}$ was not inadvertently masked.