MNRAS **000**, 1–?? (2019) Preprint 10th January 2020 Compiled using MNRAS LATEX style file v3.0

The KLEVER Survey: Spatially resolved metallicity maps and gradients in a sample of $1.2 < z < 2.5$ lensed galaxies

Mirko Curti,[1,2] ⋆ Roberto Maiolino,[1,2] Michele Cirasuolo,[3] Filippo Mannucci,[4] Rebecca J. Williams,[5,1] Matt Auger,[6] Amata Mercurio,[7] Connor Hayden-Pawson,[1,2] Giovanni Cresci,[4] Alessandro Marconi,[8] Francesco Belfiore,[3] Michele Cappellari,[9] Claudia Cicone,[10,11] Fergus Cullen,[12] Massimo Meneghetti,[13] Kazuaki Ota,[14] Yingjie Peng,[15] Max Pettini,[6] Mark Swinbank,[16,17] and Paulina Troncoso[18]

[1] *Cavendish Laboratory, University of Cambridge, 19 J. J. Thomson Ave., Cambridge CB3 0HE, UK*
[2] *Kavli Institute for Cosmology, University of Cambridge, Madingley Road, Cambridge CB3 0HA, UK*
[3] *European Southern Observatory, Karl-Schwarzschild-Strasse 2, D-85748 Garching bei Muenchen, Germany*
[4] *INAF - Osservatorio Astrofisico di Arcetri, Largo E. Fermi 5, I-50125, Firenze, Italy*
[5] *Observatory Sciences Ltd, 1 New Road, St Ives, PE27 5BG, Cambridgeshire, UK*
[6] *Institute of Astronomy, University of Cambridge, Cambridge CB3 0HA, UK*
[7] *INAF - Osservatorio Astronomico di Capodimonte, Via Moiariello 16, I-80131 Napoli, Italy*
[8] *Dipartimento di Fisica e Astronomia, Università di Firenze, Via G. Sansone 1, I-50019, Sesto Fiorentino (Firenze), Italy*
[9] *Sub-department of Astrophysics, Department of Physics, University of Oxford, Denys Wilkinson Building, Keble Road, Oxford OX1 3RH, UK*
[10] *INAF - Osservatorio Astronomico di Brera, via Brera 28, 20121 Milan, Italy*
[11] *Institute of Theoretical Astrophysics, University of Oslo, P.O. Box 1029, Blindern, 0315 Oslo, Norway*
[12] *Institute for Astronomy, University of Edinburgh, Royal Observatory, Edinburgh EH9 3HJ, UK*
[13] *INAFâĂŤOsservatorio Astronomico di Bologna, Via Ranzani 1, I-40127 Bologna, Italy*
[14] *Kyoto University Research Administration Office, Yoshida-Honmachi, Sakyo-ku, Kyoto 606-8501 Japan*
[15] *Kavli Institute for Astronomy and Astrophysics, Peking University, Beijing 100871, China*
[16] *Institute for Computational Cosmology, Durham University, South Road, Durham DH1 3LE UK*
[17] *Center for Extra-galactic Astronomy, Durham University, South Road, Durham DH1 3LE UK*
[18] *Universidad Autónoma de Chile, Facultad de Ingeniería, Núcleo de Astroquímica & Astrofísica, Av. Pedro de Valdivia 425, Providencia, Santiago*

Accepted 2019 November 25. Received 2019 October 28; in original form 2019 June 5

ABSTRACT

We present near-infrared observations of 42 gravitationally lensed galaxies obtained in the framework of the KMOS Lensed Emission Lines and VElocity Review (KLEVER) Survey, a program aimed at investigating the spatially resolved properties of the ionized gas in $1.2 < z < 2.5$ galaxies by means of a full coverage of the YJ, H and K near-infrared bands. Detailed metallicity maps and gradients are derived for a subsample of 28 galaxies from reconstructed source plane emission line maps, exploiting the variety of different emission line diagnostics provided by the broad wavelength coverage of the survey. About 85% of these galaxies are characterized by metallicity gradients shallower than 0.05dex/kpc and 89% are consistent with flat slope within 3σ (67% within 1σ), suggesting a mild evolution with cosmic time. In the context of cosmological simulations and chemical evolution models, the presence of efficient feedback mechanisms and/or extended star formation profiles on top of the classical "inside-outâĂİ scenario of mass assembly is generally required to reproduce the observed flatness of the metallicity gradients beyond $z \sim 1$. Three galaxies with significantly ($> 3\sigma$) "inverted" gradients are also found, showing an anti-correlation between metallicity and star formation rate density on local scales, possibly suggesting recent episodes of pristine gas accretion or strong radial flows in place. Nevertheless, the individual metallicity maps are characterised by a variety of different morphologies, with flat radial gradients sometimes hiding non-axisymmetric variations on kpc scales which are washed out by azimuthal averages, especially in interacting systems or in those undergoing local episodes of recent star formation.

Key words: galaxies: high-redshift – galaxies: abundances – galaxies: evolution

1 INTRODUCTION

During the epoch characterized by the peak of the cosmic star formation history (i.e. 1.5 < z < 3) galaxies were experiencing dramatic transformations affecting their morphology and dynamics. The enhanced star formation activity, regulated by the interplay between cosmic gas accretion, merger events and gas outflows due to the stellar and AGN winds, was responsible for the the bulk of the cosmic evolution of galaxies (e.g. Somerville & Davé 2015). All these processes have also left a clear imprint on the content of heavy elements (i.e. metals) in the interstellar medium (ISM) and their spatial distribution across a galaxy (Davé et al. 2011). Measurements of the gas-phase metallicity (which we simply refer henceforth as metallicity) at these epochs therefore provides unique insights on the history of the baryonic cycling and its influence on the evolution of galaxies (see Maiolino & Mannucci 2019 for an extensive review).

The existence of global scaling relations involving metallicity and other intrinsic galaxy properties, like stellar mass (i.e. the mass-metallicty relation, MZR, Tremonti et al. 2004, Andrews & Martini 2013 and many others), has been assessed through the years thanks to the advent of large astronomical databases such as the Sloan Digital Sky Survey (SDSS, Yuan et al. 2011). These relationships have been further investigated at higher redshifts to search for clues of a possible cosmic evolution. Several lines of evidence supporting an overall decrease in metallicity at fixed stellar mass have been found from deep observational campaigns conducted in the near-infrared (e.g. Erb et al. 2006, Maiolino et al. 2008, Mannucci et al. 2009,Zahid et al. 2011, Cullen et al. 2014, Steidel et al. 2014,Sanders et al. 2015, Guo et al. 2016). Moreover, the scatter in the local MZR has been observed to exhibit a clear secondary dependence on the star formation rate (Ellison et al. 2008); Mannucci et al. (2010) first proposed the existence of a tight (i.e. ~ 0.05 dex dispersion) relation, followed by local galaxies, in the three-dimensional space defined by stellar mass ($M_\star$), metallicity (Z) and star formation rate (SFR), which is usually referred to as the Fundamental Metallicity Relation (FMR). This tight relation, later tested and revised by different authors (e.g. Yates et al. 2012; Salim et al. 2014; Curti et al. 2019), has been interpreted as a consequence of a long lasting equilibrium between gas accretion, mass growth, metal production and outflows of enriched material, and does not show any clear sign of evolution up to $z \sim 2.5$ (Mannucci et al. 2010; Cresci et al. 2018). In this picture, the observed MZR (at any cosmic time) just follows from the two-dimensional projection of the FMR on the $M_\star$ vs log(O/H) plane, while its cosmic evolution arises from sampling different regions of the FMR due to the increase in the average SFR density with redshift. Despite a variety of physically motivated theoretical frameworks supporting this scenario (e.g. Lilly et al. 2013; Dayal et al. 2013; Hunt et al. 2016; Davé et al. 2017), as well as several confirmations from observations of high-z galaxies (e.g. Richard et al. 2011; Belli et al. 2013; Wang et al. 2017), the existence and the possible evolution of the M-Z-SFR relation at high redshift is currently debated (Steidel et al. 2014; Cullen et al. 2014; Wuyts et al. 2014, 2016; Sanders et al. 2018, see the discussion in Cresci et al. 2018).

With the increased availability of multi-object and integral field spectrographs, many studies have also turned to investigating the spatial distribution of metals inside galaxies, assessing the presence of radial variations in the chemical enrichment levels. Tracing the evolution of these metallicity gradients across the cosmic epochs is a crucial benchmark for theoretical models aimed at describing the relative contributions that star formation, gas flows and feedback processes play in driving galaxy evolution. Although relatively well characterized in the local Universe, the main features of metallicity gradients still remain so far very poorly constrained at higher redshifts. Locally, the large majority of spiral galaxies exhibit negative metallicity gradients, with inner regions more chemically enriched with respect to the outskirts of galactic discs. This follows observationally from the spectroscopic analysis of HII regions in the Milky Way (e.g. Magrini et al. 2010; Stanghellini & Haywood 2010) and in nearby galaxies (e.g. Zaritsky et al. 1994; Kewley et al. 2010; Bresolin 2011; Berg et al. 2012, 2015), from large campaign based on integral field spectroscopy (e.g. Sánchez et al. 2014; Ho et al. 2015; Belfiore et al. 2017), as well as from abundance measurements from stellar spectroscopy of individual massive young stars (Kudritzki et al. 2015; Gazak et al. 2015; Bresolin et al. 2016) and spatially resolved analysis of stellar population properties (e.g Li et al. 2018, see Conroy 2013 for a review). Negative gradients are generally interpreted as indicative of the so called inside-out growth scenario of galaxy formation (Samland et al. 1997; Portinari & Chiosi 1999; Prantzos & Boissier 2000; Pilkington et al. 2012; Gibson et al. 2013a). Indications of flattening gradients beyond a certain radius may indicate the presence of radial mixing processes or (re)accretion of metal-enriched gas in the outer regions (Bresolin et al. 2012), whilst intense accretion af external pristine gas may also produce inverted gradients (as seen for instance in some dwarfs galaxies, Sánchez Almeida et al. 2018). Merger events could also play an important role in flattening the metallicity gradients (Kewley et al. 2010; Rupke et al. 2010b,a).

On the contrary, there is no general consensus yet in the literature about the behaviour of metallicity gradients at high redshift and their cosmic evolution. Planetary nebulae have been extensively used to investigate the time evolution of abundance gradients in the Milky Way and in nearby galaxies, as they probe the enrichment of the gas on different (i.e. older) timescales than HII regions (Maciel et al. 2003; Magrini et al. 2009; Henry et al. 2010; Stanghellini et al. 2010, 2014). In general, it is found that gradients inferred from tracers of long time-scales enrichment tend to be flatter than those inferred from HII regions (see e.g. Stanghellini et al. 2014). However, it is not trivial to account for the effects of radial stellar migration, which could potentially bias the results of stellar archaeology studies. The situation is perhaps even more uncertain for high redshift galaxies, where the investigations of metallicity gradients conducted so far have led to diverse and sometimes conflicting conclusions (e.g. Cresci et al. 2010; Swinbank et al. 2012; Queyrel et al. 2012; Stott et al. 2014; Jones et al. 2013; Wuyts et al. 2016; Leethochawalit et al. 2016; Wang et al. 2017). This could be mainly ascribed to the intrinsic challenges in obtaining reliable measurements of metallicity gradients at high-z, due to the poor angular resolution (Yuan et al. 2013) and to uncertainties affecting metallicity diagnostics. If the latter problem is currently still a source of large systemat-

ics, the former could be addressed and partly mitigated by exploiting new adaptive optics systems and/or the magnification provided by gravitational lensing.

The number of direct measurements of metallicity gradients at high redshift has been constantly increasing in recent years thanks to the advent of multi-IFU instrumentation like KMOS (Sharples et al. 2013) on the Very Large Telescope (VLT), with surveys targeting hundreds of galaxies at $z \sim 1-2$ (e.g. Stott et al. 2014; Wuyts et al. 2016). However, a typical seeing $\geqslant 0.6''$, which roughly corresponds to $\sim$ 5kpc at z $\sim$ 2, does not allow us to properly resolve the inner structure of high-redshift galaxies, especially for low mass objects. As previously stated, adaptive optics assisted observations could greatly enhance the spatial resolution down to $\sim$ 1kpc (Swinbank et al. 2012), which can be further improved by targeting gravitationally lensed sources (Jones et al. 2010a,b, 2013; Leethochawalit et al. 2016). Space-based grism spectroscopy from *HST* has also been recently used to infer sub-kpc resolution metallicity maps and gradients (Jones et al. 2015a; Wang et al. 2017), which are however affected by poor spectral resolution. Despite these efforts, high angular resolution measurements of metallicity gradients are still scarce in terms of statistical significance. Moreover, the large majority of the studies conducted so far rely only on a limited number of emission line detections, with metallicity estimates that could therefore suffer from potential biases due to the different physical properties of high redshift galaxies with respect to the local samples used to calibrate the abundance diagnostics.

In this paper we investigate the metallicity properties of a sample of 42 galaxies between 1.2 < z < 2.5, gravitationally lensed by either foreground galaxy clusters (in particular the MS2137 and RXJ2248 clusters) or individual galaxies. The analysis presented in this work is based on the first observations conducted in the framework of the KMOS LEnsed galaxies Velocity and Emission line Review (KLEVER) Large Programme (39 galaxies) and include also 3 strongly lensed sources observed with SINFONI in the context of different observational programs. We take advantage of multi-band observations conducted in the YJ, H and K bands, which allow the simultaneous coverage of many rest-frame optical emission lines, to derive spatially resolved metallicity maps using different strong-line diagnostics and assess the presence and the cosmic evolution of metallicity gradients. We manage to obtain at least marginally resolved gradients for 28 out of 42 the galaxies presented in this work: we refer to these galaxies throughout the paper as the *metallicity gradient sample*.

The paper is organized as follows: in Section 2 we give an overview on the KLEVER Programme, describing the target selection, the observing strategy and the data reduction. Section 3 describes the emission line fitting method, the lens modelling and Source Plane reconstruction for our galaxies and how we measured stellar masses, star formation rates and gas-phase metallicities. In Section 4 we present our results in terms of global sample properties and spatially resolved metallicity maps and gradients, which are then discussed in Section 5. Finally, our conclusions are reported in Section 6.

In Table 1, we report the notations used throughout the paper to indicate the line ratios adopted in our analysis. Throughout this work we assume a standard ΛCDM cosmology based on the results from Planck Collaboration et al. (2016), with $H_0 = 67.8$ km s^{-1} Mpc^{-1}, $\Omega_M = 0.308$, $\Omega_\Lambda = 0.69$.

Table 1. Definition of line ratios adopted throughout this paper.

Notation	Line Ratio
R_2	[O II]λ3727, 29/Hβ
R_3	[O III]λ5007/Hβ
N_2	[N II]λ6584/Hα
S_2	[S II]λ6717, 31/Hα
O_3O_2	[O III]λ5007/[O II]λ3727, 29
O_3N_2	[O III]λ5007/Hβ / [N II]λ6584/Hα

2 OBSERVATIONS

2.1 The KLEVER Survey: sample selection and observing strategy

KLEVER is an ESO Large Program (197.A-0717, PI: Michele Cirasuolo) conducted with the multi-object near-IR integral field spectrograph KMOS on the VLT (Sharples et al. 2013) and aimed at investigating spatially resolved kinematics, dynamics and properties of the ionised gas in a sample of $\sim$ 200 galaxies at $1.2 \lesssim z \lesssim 2.5$. The survey is designed to provide a full coverage of the near-infrared region of the spectrum by observing each galaxy in the YJ, H and K band (spanning respectively the 1.025 – 1.344 μm, 1.456 – 1.846 μm and 1.92 – 2.46 μm wavelength ranges), hence allowing us to detect and spatially map almost the entire set of the brightest optical rest-frame nebular lines. The full KLEVER sample comprises both gravitationally lensed galaxies within well studied cluster fields from the CLASH (Postman et al. 2012) and FRONTIER FIELDS (Lotz et al. 2017) programs, as well as un-lensed galaxies in the southern CANDELS fields UDS, COSMOS and GOODS-S.

The analysis presented in this paper is based on the first available observations in KLEVER, targeting 39 gravitationally lensed galaxies behind the clusters MS2137-2353 (hereafter MS2137) and RXJ2248.7-4431 (also know as AS1063, hereafter RXJ2248) and carried out in Service Mode during Periods 95-97 (from May 2015 to September 2016). During the creation of the KMOS mask, the targets within each pointing have been prioritised according to the observability of the emission lines of interest, in order to maximise the number of lines falling within the different NIR bands while minimising at the same time the sky contamination from OH lines (as identified from the catalogue provided by Rousselot et al. 2000). The targets have been primarily selected to fall within specific redshift ranges, either at z $\in$ [1.2 , 1.65] to have Hβ+[O III]λ5007 in the *YJ* band, Hα+[N II]λ6584+[S II]$\lambda\lambda$6717, 31 in the *H* band and [S III]$\lambda\lambda$9068, 9530 in the *K* band, or at z $\in$ [2 , 2.6] to have [O II]$\lambda\lambda$3727, 29 in the *YJ* band, Hβ+[O III]λ5007 in the *H* band and Hα+[N II]λ6584+[S II]$\lambda\lambda$6717, 31 in the *K* band. Spectroscopic redshifts used for target selection and prioritisation were provided as part of the CLASH-VLT survey (Rosati et al. 2014) conducted with VIMOS on the VLT. Three IFUs have been assigned, for each pointing, to bright continuum sources for alignment purposes (see Sect. 2.2),

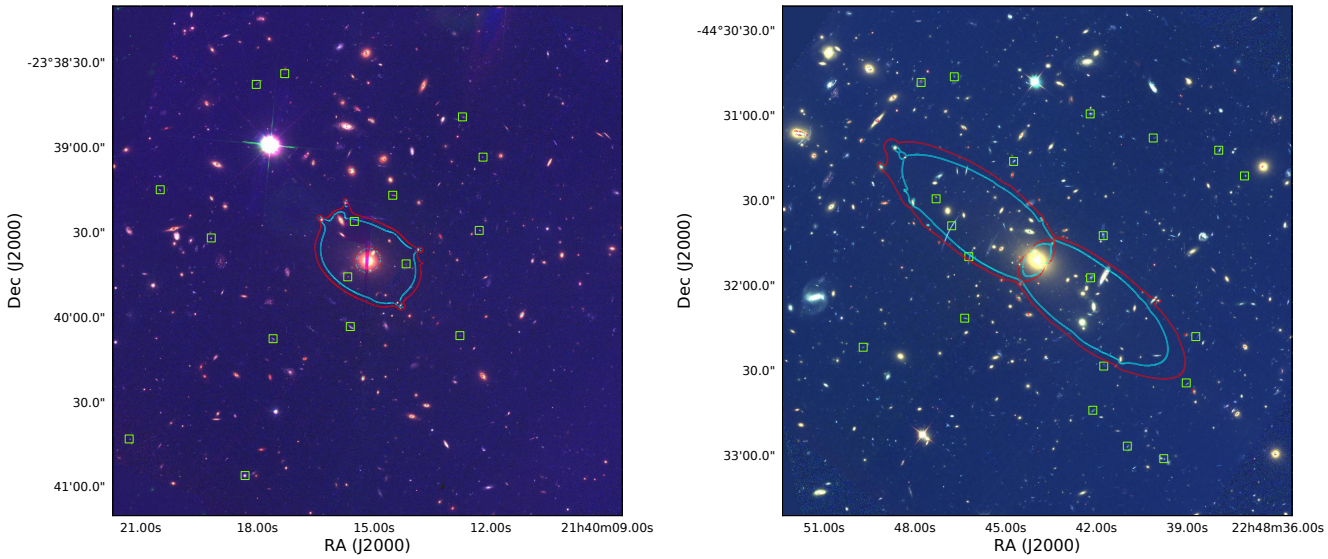

Figure 1. Composite RGB-color image of the MS2137 (*left panel*) and RXJ2248 (*right panel*) clusters. The red, green and blue channels are assigned to broad band F814W, F606W and F435W images obtained in the framework of the CLASH and Frontier Fields program respectively. The green squares superimposed mark the position of deployed KMOS IFUs targeting the high redshift sources analysed in this work, while cyan and red contours represent the critical lines of infinite magnifications predicted, for z = 1.4 and z = 2.2 source redshift respectively, by the lensing models adopted in this work, i.e. Zitrin et al. (2013).

while remaining spare IFUs were assigned to low priority targets or z> 3 sources. No prior screening to identify AGN contamination was performed for the galaxies within the clusters.

Each lensed galaxy belonging to RXJ2248 has been observed for a total exposure time on source of 11 hours (3h in the *YJ* band, 3h in *H* and 5h in *K* respectively), whereas galaxies from MS2137-2353 have been observed in total for 13.1 hours (4.2 in YJ, 4.5 in H and 4.4 in K band respectively). For KMOS observations, we adopted an A-B-A nodding (with dithering) strategy for sky sampling and subtraction. The average seeing of the observations (as inferred from the reference stars observed in three KMOS IFUs) ranged between 0.5″ and 0.6″. In Figure 1 the position of the KMOS IFUs, deployed on the target galaxies, are shown on top of color-composite RGB images of both cluster fields. The coloured curves in Fig. 1 denote the critical lines of infinite magnification within the two clusters for source redshift 1.4 (cyan) and 2.2 (red) respectively, as predicted by the adopted lensing models (see Section 3.2).

In addition, we have included in the analysis three strongly lensed galaxies at z> 2 observed with the integral field spectrograph SINFONI and part of galaxy-galaxy lensing systems or lensed by galaxy clusters. These galaxies fully match the criteria described above for the KMOS-sample in terms of emission lines detectability. Two out of the three galaxies were observed as part of some of our previous programs: SDSS J114833+193003 (also known as the *Horseshoe*) and a strongly lensed, arc-like shaped galaxy within the MACS J0451+0006 (MACS0451) cluster. Seeing-limited mode observations have been performed, with a PSF-FWHM ranging between 0.4″ to 0.9″, and the 0.25″ x 0.25″ pixel scale was adopted. The observations of the third additional galaxy (SDSS J0232-0323, also know as CSWA164 as part of the CASSOWARY Survey Belokurov et al. (2009)) were retrieved instead from the ESO archive. This brings the total observed sample to 42 galaxies. We note that, due to scheduling constraints, not all the galaxies observed with SINFONI have the desired full wavelength coverage (i.e. *J*, *H* and *K* band observations) as the KMOS galaxies.

The complete list of the targets analysed in this work is given in Table 2, for which we report coordinates, redshift and emission lines detected above the 3σ level within each band from the integrated spectra. Unfortunately, low signal-to-noise affected the detection of the faintest nebular lines (i.e Hβ, [N II]λ6584, [S II]λ6717, 31) in some of the KMOS datacubes, reducing the effective number of sources for which we could reliably derive a metallicity gradient to 28 (the *metallicty gradient sample* see Sect. 4.2 for details).

2.2 Data Reduction

The KMOS data were reduced using the pipeline provided by ESO (v.1.4.3). Within the pipeline environment, we implemented the advanced sky subtraction technique from Davies (2007) as well as the sky-stretch algorithm, which stretches the sky-cube involving a relatively high degree polynomial in order to align the sky lines with those in the object-cube. This slightly improved the residuals from the first correction. The final datacubes were then reconstructed onto a 0.1″ x 0.1″ pixel scale. To properly align and combine the individual exposures, both within a single observing block (OB) and between different OBs, three IFUs (i.e. one for each of the KMOS detectors) were devoted to observe bright stars; we then exploited the relative position of their centroids in each exposure to compute the shifts to apply to the scientific sources that were observed on the same detector of the corresponding reference star. This method has proven to provide more precise alignments than just relying only on the informations stored in the header, since it

Galaxy	R.A.	Dec.	z [Hα]	*J*-Band	*H*-Band	*K*-Band
SINFONI Galaxies						
Horseshoe	11:48:32.7	19:30:03.5	2.383	-	[O III]+Hβ	Hα+[N II]+[S II]
(Western arc)						
Horseshoe	11:48:32.7	19:30:03.5	2.383	-	-	Hα+[N II]+[S II]
(South+western arc)						
MACS0451 Arc	04:51:57.3	00:06:19.7	2.014	[O II]	[O III]+Hβ	Hα+[N II]+[S II]
CSWA164	02:32:49.8	-03:23:26.6	2.518	[O II]	[O III]+Hβ	Hα+[N II]+[S II]
KMOS Galaxies						
MS2137						
sp1	21:40:18.031	-23:38:37.87	1.393	[O III]+Hβ	Hα+[N II]+[S II]	-
sp2	21:40:12.730	-23:38:49.34	2.2425	[O II]	[O III]+Hβ	Hα+[N II]+[S II]
sp3	21:40:14.522	-23:39:17.03	1.2645	[O III]+Hβ	Hα+[N II]+[S II]	[S III]
sp5	21:40:12.298	-23:39:29.44	2.0146	[O II]	[O III]+Hβ	Hα+[N II]+[S II]
sp6	21:40:19.188	-23:39:32.11	2.4881	[O II]	[O III]+Hβ	Hα+[S II]
sp7	21:40:17.599	-23:40:07.70	1.6523	[O III]+Hβ	Hα+[N II]+[S II]	-
sp9*	21:40:18.319	-23:40:56.04	1.6497	Hβ +[O III]	Hα+[N II]	-
sp10	21:40:15.679	-23:39:45.84	3.0843	-	-	[O III]
sp13	21:40:15.510	-23:39:26.27	1.4951	[O III]+Hβ	Hα+[S II]	-
sp14	21:40:14.178	-23:39:41.26	1.4948	[O III]+Hβ	Hα+[S II]	-
sp15	21:40:15.614	-23:40:03.42	1.4967	[O III]+Hβ	Hα+[N II]+[S II]	-
ph6532	21:40:21.265	-23:40:43.11	2.071	[O II]	[O III]+Hβ	Hα
ph2594	21:40:17.300	-23:38:34.06	1.348	[O III]	Hα+[N II]+[S II]	-
ph3729	21:40:12.214	-23:39:3.561	1.4291	[O III]+Hβ	Hα+[N II]	-
ph3912	21:40:20.497	-23:39:15.06	1.3275	[O III]+Hβ	Hα+[S II]	-
ph5514	21:40:12.795	-23:40:6.609	1.5254	-	Hα+[N II]+[S II]	-
ph7727	21:40:12.228	-23:41:25.74	1.5789	[O III]	Hα	-
ph8073	21:40:10.607	-23:41:37.73	1.2386	[O III]+Hβ	Hα+[S II]	[S III]
RXJ2248 (AS1063)						
GLASS_00093-99-99	22:48:46.701	-44:30:46.27	1.4317	[O III]+Hβ	Hα+[N II]	-
R2248_LRb_p1_M3_Q4_58__2	22:48:42.213	-44:30:59.38	1.4199	[O III]+Hβ	Hα+[N II]+[S II]	-
MUSE_SW_462-99-99	22:48:41.781	-44:31:42.39	1.4286	[O III]+Hβ	Hα+[N II]+[S II]	-
GLASS_00333-99-99	22:48:40.130	-44:31:07.90	1.4285	[O III]+Hβ	Hα	-
R2248_LRb_p3_M4_Q3_93__1	22:48:37.966	-44:31:12.21	1.485	[O III]+Hβ	Hα+[N II]+[S II]	[S III]
R2248_MR_p1_M1_Q4_10__1	22:48:35.016	-44:30:30.07	1.4312	[O III]	Hα	-
R2248_LRb_p3_M4_Q3_94__1	22:48:37.116	-44:31:21.29	2.0665	[O II]	[O III]+Hβ	Hα+[N II]+[S II]
MUSE_SW_48-99-99	22:48:38.723	-44:32:18.05	2.5804	-	-	-
MUSE_SW_51-99-99	22:48:41.764	-44:32:28.47	3.2275	-	-	[O III]
R2248_MR_p3_M1_Q3_43__1	22:48:39.037	-44:32:34.41	3.2421	-	[O II]	[O III]+Hβ
GLASS_01891-99-99	22:48:39.780	-44:33:01.17	1.159	-	-	-
GLASS_01845-99-99	22:48:40.984	-44:32:56.75	2.3014	[O II]	[O III]+Hβ	Hα
MUSE_SW_45-99-99	22:48:42.125	-44:32:44.08	1.2694	[O III]+Hβ	Hα+[N II]+[S II]	[S III]
MUSE_SW_461-99-99	22:48:42.199	-44:31:57.29	1.4286	[O III]+Hβ	Hα+[S II]	-
MUSE_NE_111-99-99	22:48:46.358	-44:32:11.56	1.3975	[O III]+Hβ	Hα+[S II]	-
GLASS_01404-99-99	22:48:49.712	-44:32:21.81	1.4475	-	-	[S III]
R2248_MR_p1_M1_Q4_51__1	22:48:46.222	-44:31:49.81	1.2593	[O III]+Hβ	Hα	[S III]
GLASS_00800-99-99	22:48:46.787	-44:31:38.91	1.2282	[O III]+Hβ	Hα+[N II]+[S II]	[S III]
MUSE_NE_23-99-99	22:48:44.742	-44:31:16.18	1.2282	[O III]+Hβ	Hα+[S II]	-
R2248_MR_p1_M1_Q4_59__1	22:48:47.796	-44:30:48.28	1.4281	-	Hα+[N II]+[S II]	-
MUSE_NE_117-99-99	22:48:47.302	-44:31:29.37	3.4519	-	-	-

Table 2. The full sample of galaxies analysed in this work. The systemic redshift reported is derived from the Hα detection in the integrated spectra. In case of no Hα detection (e.g. for $z > 3$ galaxies), the redshift is computed from [O III]λ5007. We also report the main emission lines detected above 3σ in the integrated spectra in each targeted band. A '-' denotes that observations in that band were available, but we did not detect any emission line with at least 3σ significance, with the exception of the Horseshoe galaxy for which observations in the J-band were not available.
*: type-1 AGN

mitigates potential offsets introduced by the differential rotation of the instrument's IFUs at different times that are non tracked by the World Coordinate System (WCS). The final cubes were then created through a (three) sigma-clipped average.

The SINFONI data were reduced following the latest version of the ESO-SINFONI pipeline to perform the flat fielding, wavelength calibration and reconstruct a non-sky subtracted cube for each observation, after the removal of cosmic rays from raw data using the "L.A.Cosmic" procedure by van Dokkum (2001). The pixels in the datacubes were resampled to a symmetric angular size of 0.125″ x 0.125″. Then, we implemented the sky subtraction technique from Davies (2007) to perform a better removal of the residual

OH airglow emission lines from the data. We corrected for the atmospheric absorption and instrumental response using a telluric standard star, which also provides the flux calibration, before finally combining all the single OBs through a sigma-clipped average to produce the final science cubes.

3 ANALYSIS

3.1 Emission line fitting

On spatially resolved basis, we performed the emission line fitting on the fully reduced datacubes, which sample the image plane of each galaxy. We spatially smoothed the datacubes with a Gaussian kernel of 0.4″ FWHM, below the average PSF-FWHM of the observations ($\sim 0.5'' - 0.6''$, as measured from the three reference stars), to increase the signal-to-noise ratio (SNR) of the weakest emission lines that we aim to detect and map in our galaxies (e.g. [N II]λ6583, [S II]λ6717, 6731 and Hβ).

All the emission lines of interest were fitted with single Gaussian components at the location of each spaxel in the datacube and we included a linear component to account for continuum emission in each band. The Hα, [N II] and [S II] lines were jointly fit by linking their velocity and line width, and the same criteria was applied when fitting [O III] and Hβ. When available, the [O II]$\lambda\lambda$3726, 3729 and [S III]$\lambda\lambda$9068, 9532 line doublets are fitted with two Gaussian components which are linked in velocity and width as well. Moreover, the [O III]$\lambda\lambda$4959, 5007 and [S III]$\lambda\lambda$9068, 9532 doublets were fixed in flux ratios to their relative Einstein spontaneous emission coefficients (i.e. 3 and 2.47 respectively), and the relative intensity of the two lines of the [S II]λ6717, 31doublet was constrained to be within the physical range associated to the low- and high-density regimes (i.e. [S II]λ6717/[S II]λ6731 $\in$ [0.4, 1.45]). We choose to keep the fitting procedure within each band self-consistent: we do not link the line widths between different bands because of the different KMOS resolving powers (namely 3582, 4045 and 4227 in the centre of YJ, H and K band respectively). Velocities were not linked as well to avoid any systematic introduced by possible offsets in the wavelength calibration between the three bands. Each fit on spaxel-by-spaxel basis was inverse weighted by the corresponding noise spectrum extracted from the noise datacube provided by the pipeline, and the spectra were totally masked at the position of the brightest sky lines.

For each galaxy, the emission line maps from different bands were finally re-aligned exploiting the detection of the underlying stellar continuum or, in case of no continuum detection, aligning the peaks of the brightest emission lines detected in each band (e.g. [O III]λ5007 in YJ, Hα in H and [S III]λ9530 in K for a typical $z \sim 1.4$ galaxy). After the fitting procedure is completed, all the emission line maps are mapped back into the Source Plane of the galaxy after converting them to surface brightness units, which is the physical quantity conserved by gravitational lensing (we refer to the following sub-section for more details).

Representative integrated spectra for each source have also been extracted from circular pseudo-fibres of 0.6″-radius (equivalent to the average FWHM of the seeing-limited PSF) centred on the position of the peak of the emission in each band, and we fit the emission lines following the same prescriptions described above. The choice of the aperture width is arbitrary, but has proven to be an effective compromise to be representative of the bulk of the emission while encompassing at the same time a significant fraction of the total flux. An example of integrated YJ, H and K band spectra for a $z \sim 1.4$ and a z$\sim$ 2.2 galaxy are shown in Fig.2; the best-fit to continuum and nebular line emission is shown in red, while the blue shaded areas mark the regions masked-out in the fitting procedure.

The list of emission lines detected in each galaxy is reported in Table 2. We report a detection if the signal-to-noise ratio (SNR) on the integrated emission line flux is higher than 3. We achieved a 92% detection rate for integrated Hα and 83% for [O III]λ5007. The detection rate drops to 71% for Hβ, 56% for [N II]λ6584 and 53% for [S II]λ6717, 31 (i.e. where at least one of the lines in the doublet is detected with SNR> 3.). In 7 of the z$\sim$ 1.4 sources, the [S III]λ9530 line (the brightest of the [S III] doublet) is also detected in the integrated spectra: these represent some of the very first observations of this emission line at such redshifts, and provide key insights regarding the excitation state of the gas. Finally, 64% of the z> 2 galaxies present a detection of the [O II]λ3727, 29 doublet.

3.2 Lens modelling and deprojection

In order to reconstruct our galaxies in the source plane, we have to build lens models that describe how the foreground mass (either a single galaxy or a cluster) has re-distributed the emission from the background lensed galaxy into the image plane, i.e. the image that we see on the sky. In the following we discuss in more detail the lens modelling by distinguishing between the cases of lensing by individual galaxies and lensing by galaxy clusters. After the lens models have been constrained by the broadband imaging (retrieved from both *HST* and Keck archives), we then use deflection maps generated from these models to map the source properties (line intensities, velocities, etc.) back to the un-distorted source plane.

3.2.1 Galaxy lenses

Two of our targets are part of a galaxy-galaxy lensing system (i.e. the foreground mass is a single galaxy). For the Horseshoe galaxy, we use *F606W HST* observations to perform the modelling where we assume an elliptical power law with external shear lensing mass distribution. We use the adaptive-source-plane technique of Vegetti & Koopmans (2009) to create a galaxy model in the source plane, which we then lens into the image plane using a trial lens model before comparing the results with the *HST* image. We then vary the mass-model parameters until an optimal match to the *HST* data is obtained. This can be done particularly well for the Horseshoe as it forms an almost complete Einstein ring. The same procedure is employed to model the CSWA164 lensed galaxy (which also consists of a nearly complete Einstein ring) using data from the ESI imager on KECK II.

To de-lens the galaxies and reconstruct them back in the source plane we use these lens mass models to compute deflection maps. We first have to align the SINFONI

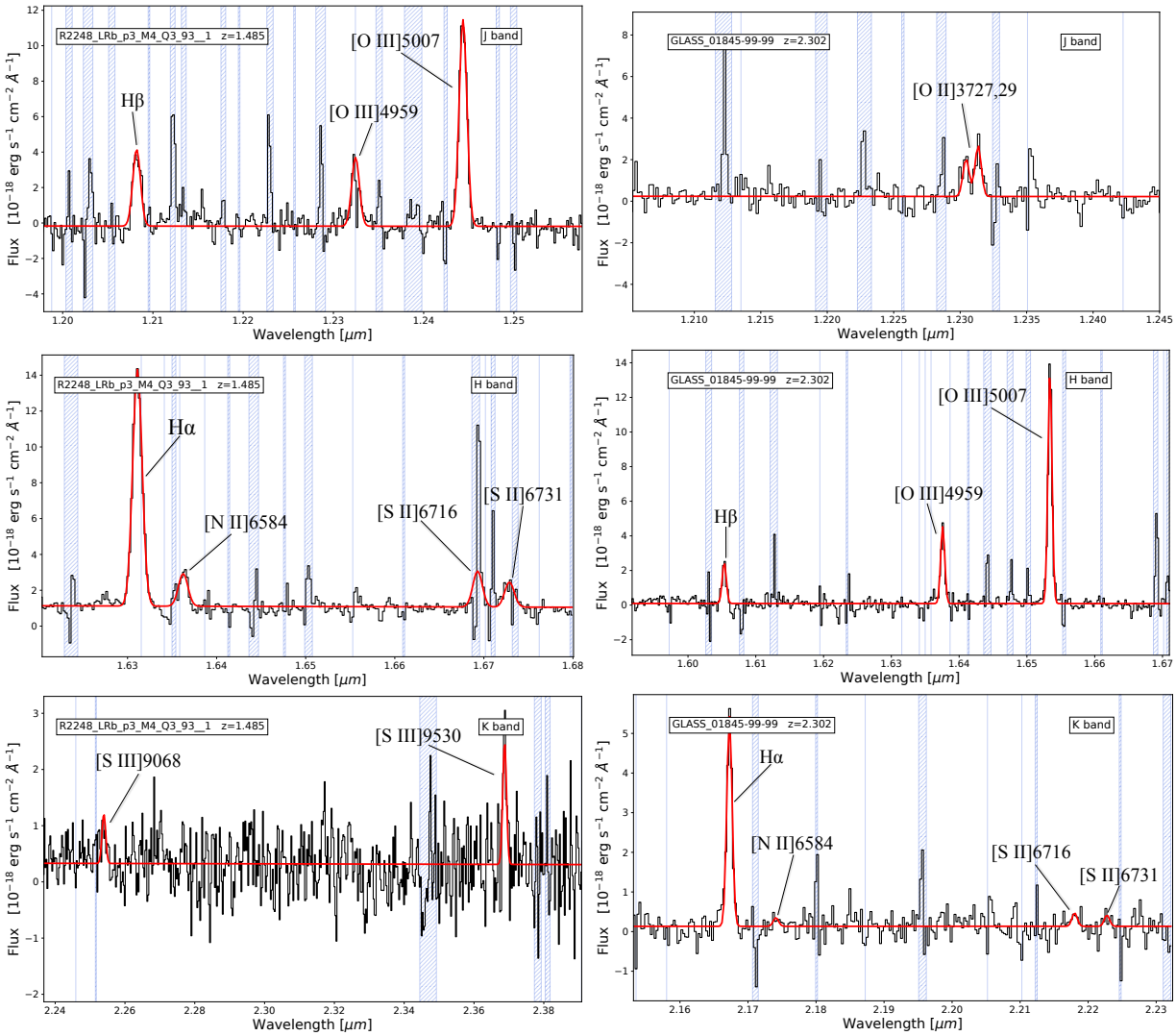

Figure 2. Integrated spectra for a z = 1.485 galaxy (R2248_LRb_p3_M4_Q3_93_1, left panels) and for a z = 2.302 galaxy (GLASS_01845-99-99, right panels) in the YJ, H and K band. The best-fit to the main emission lines targeted in this study (i.e. [O II]λ3727, 29, [O III]λ5007, Hβ, Hα, [N II]λ6584, [S II]λ6717, 31, [S III]λ9068, 9530) are shown by the red component, while the blue shaded areas mark the regions masked out during the fitting procedure.

datacubes with the broadband images, and we do so by determining the centre of the foreground galaxy (i.e. the lens) in both datasets and applying a scaling factor to the Einstein radius equal to the ratio of the different pixel scales of SINFONI and the broadband data. The mass model then defines a mapping for every pixel in the SINFONI cube back to its un-distorted location in the source plane.

3.2.2 Cluster lenses

In the case of the galaxy strongly lensed by the MACS0451 cluster (and part of the SINFONI sub-sample), we use a different technique. We exploit the multiple imaged system to constrain the model by focussing the images back to the source plane so that the multiple images all map back to the same position. We assume an elliptical power law with external shear for the cluster halo and also use singular isothermal sphere mass profiles to model the mass in each of the individual cluster galaxies. As explained in the Sect. 3.2.1, we then use this mass model to create deflection maps that allow us to deproject the SINFONI data and reconstruct the source properties in the un-distorted source plane. Interestingly, both the de-lensed HST image and Hα map of the strongly lensed galaxy in MACS0451 (Fig. 3) highlight what appears to be two separate peaks approximately $\sim$ 2 kpc apart. Further investigation conducted on the velocity field confirms the possibility that this is a merging system, as the two Hα peaks are apparently associated with different rotation patterns. Therefore, for the purpose of this work, this system is considered as constituted by a pair of galaxies and we thus compute two separate metallicity gradients, one for the *northern* and one for the *southern* region of each map.

The galaxies observed with KMOS are background galaxies of lensing clusters with publicly available mass mod-

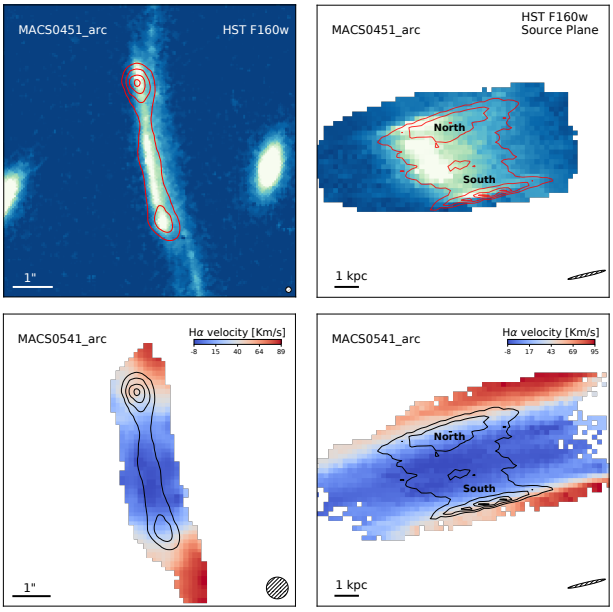

Figure 3. *Upper left panel:* HST F160W broad filter image of the galaxy lensed by the MACS0451 cluster (μ = 11.7). *Upper right panel:* Source plane reconstructed HST image of the source. The contours of the Hα emission from SINFONI observations are overlaid in red in both panels. *Lower panels:* Velocity field from Hα emission in the image plane (left) and reconstructed velocity field in the source plane (right); the source plane maps and velocity field reconstruction of this system suggest the presence of two distinct sources, which we indicate as the *northern* and *southern* galaxy. Hα contours are overlaid in black. The PSF in the image and source plane is shown in the bottom right corner of each panel.

els[1] provided by the CLASH and FRONTIER FIELDS collaborations. In particular, we exploited the mass models described in Zitrin et al. (2013, 2015), which are based on assuming a Pseudo Isothermal Elliptical Mass Distributions for the galaxies and an analytical elliptical-NFW dark matter halo, primarily centred on the BCG(s). We use the deflection maps generated from these models to map the lensed source properties back to the de-lensed source plane; these deflection maps are defined from *HST* images and we therefore re-bin our IFU data to the same pixel scaling as the *HST* observations (i.e. 0.065″/pixel) and then align our surface brightness maps with the *HST* images for each emission line. In the cases where we do not have a strong detection of the continuum from our IFU data, we are forced to align the peak of the *H*-band emission line (Hα or [O III] depending on the source redshift), with the peak in the *HST* *H*-band image, hence assuming that the latter is dominated by emission line flux, or that continuum and line emission are co-spatial.

3.2.3 Source plane reconstruction and the PSF

We reconstruct the de-lensed surface brightness maps by defining a regular grid in the source plane for each system and using the deflection maps described in the previous two sections to define on which source-plane pixel a given observed pixel will fall. The source-plane pixel is then evaluated as the average of all of the image-plane pixels that were mapped to it (the average is used because lensing conserves surface brightness). The pixel size of the grid was chosen to allow a proper sampling of the source plane PSF (see below). For KMOS data, maintaining the original HST pixel scale is enough for this purpose, while different resampling factors were applied to the reconstructed SINFONI data. The final pixel size of our reconstructed images on the source plane is therefore 0.065″/pixel for galaxies observed with KMOS, 0.062″/pixel for CSWA164 and 0.025″/pixel for both Horseshoe and the galaxy lensed by the MACS0451 cluster.

It is worth mentioning here what effect the de-lensing has on the PSF. Because we do not perform forward modelling of the moment maps, the reconstructed sources are significantly affected by the observational PSF. For example, what may be a circularly symmetric PSF in the image plane will be significantly skewed in the source plane as a result of removing the (preferentially tangential) lensing distortion. To investigate this effect, we take a very small (i.e. point-like) synthetic source in the source plane and use the lens models to produce mock image-plane observations. We simulate what would be observed by SINFONI or KMOS by convolving this with the PSF from each observation, as measured from the alignment stars observed in three dedicated IFUs. The PSF FWHM of our seeing-limited observations ranges between 0.5″ − 0.6″. We then de-lens this image back into the source plane to visualise how the observational PSF affects the morphology of the reconstructed source (i.e. to see the distortions imposed on a source-plane point-like object as a result of our reconstruction technique). We fit an ellipse to model the source plane PSF and give the major and minor FWHM in physical distance scales for each galaxy in Table 4. On average, the typical source plane resolution obtained is then of the order of ∼ 3 kpc. It is also worth recalling that, despite the apparent high spatial sampling obtained in the source plane provided by our procedure, the information encoded in individual pixels in the final reconstructed and interpolated images is not fully independent from that of the neighbouring spatial elements. However, taking into account the shape of the source plane PSF allows us to properly interpret the reconstructed emission line maps.

We estimate the total magnification factor μ for our sources in two different ways. In the first case, we compute $\mu_{\mathrm{H}\alpha}$ as the ratio of the total Hα flux in the image plane to the total Hα flux in the source plane, obtained by co-adding the flux of all the spaxels in the Hα map with signal-to-noise ratio > 3 before and after the reconstruction procedure. The uncertainty on $\mu_{\mathrm{H}\alpha}$ is estimated by propagating the relative uncertainties on the source plane and image plane Hα fluxes. In the second case, we exploit the magnification maps generated from the mass and shear maps provided by the lens models, once rescaled to the proper source redshift for each galaxy, to compute the magnification (μ_{model}) at the position of each galaxy. The statistical uncertainties in this latter case are estimated from the ranges of magnification values in 100 different realizations of each model, i.e. taking the 1σ interval of the resulting magnification distribution. Both values are reported in Table 3, with the two different estimates consistent within 1σ for almost all sources.

[1] downloadable from the Frontier Fields and CLASH models repository

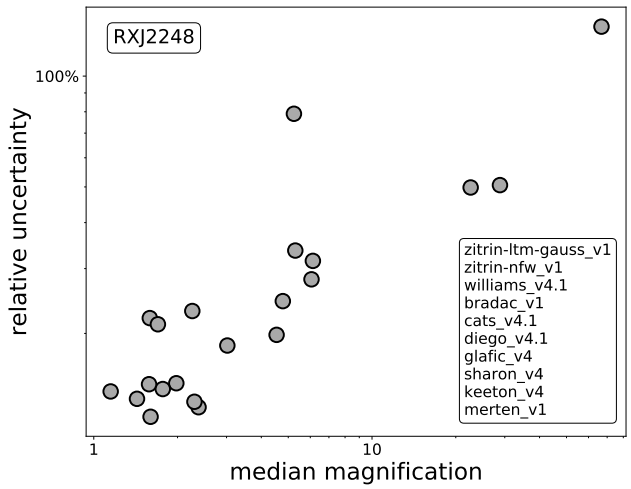

Figure 4. The (relative) systematic uncertainties on the magnification are plotted as a function of the magnifications themselves for the KLEVER galaxies within the RXJ2248 cluster. Adopting 10 different realisations of the lens model (listed in the bottom-right corner) we estimate the median and 1σ interval of the magnification distribution for each source, assuming then σ as the typical systematic uncertainty on the magnification factor associated to the choice of a particular lens model.

It is finally worth recalling that different approaches in the mass reconstruction algorithms of various lensing models could produce significant discrepancies in the estimated magnifications, especially in the proximity of the critical lines where they can reach factors of $\sim$ 30% (Meneghetti et al. 2017). For this reason, systematics uncertainties on the lens modelling are likely to dominate the error budget over the statistical uncertainties computed for each individual model. To test the impact of such effect on our study, we created magnification maps at each source redshift adopting the prescriptions of 10 different models available for the Frontier Field cluster RXJ2248 and computed the magnifications for each galaxy in our sample. We can then assume the 1σ interval of the resulting distributions as an estimate of the systematic uncertainty on the magnification associated with the choice of a particular lens model. In Fig. 4 we plot the relative uncertainties, i.e. 1σ divided by median of the magnification distribution, as a function of the median magnification itself for the KLEVER galaxies within RXJ2248. The typical systematic uncertainties for the majority of our galaxies are comprised between 10%−25%, and increase with increasing magnification. The list of lens models included in this test is reported in the bottom-right corner of the plot and each of them can be downloaded from the web page of the Frontier Fields project.

3.3 Stellar Mass and Star Formation Rate

Stellar masses and star formation rates (SFR) for our sample are listed in columns 2 and 3 of Table 3 and were derived as follows. For KMOS galaxies lensed by clusters an SED-fitting was performed on the photometric measurements extracted (with SExtractor) from broadband HST images provided by the CLASH and FRONTIER FIELDS (FF) programs. In particular, for galaxies observed in MS2137 we implemented the photometry from the 7 optical and near-infrared filters covered by CLASH (namely $F105W, F125W, F140W, F160W, F435W, F606W, F814W$), while for RXJ2248 we could exploit the deepest exposures in the same filters provided by the FF images. Stellar masses were obtained using the high-z extension of the MAGPHYS program (da Cunha et al. 2015), assuming a Chabrier (2003) IMF and exploting the Bruzual & Charlot (2003) stellar population synthesis models; in particular, MAGPHYS adopts the two-component model of Charlot & Fall (2000) to describe the attenuation of stellar emission at ultraviolet, optical, and near-infrared wavelengths. Stellar masses for a few objects within the clusters are not available because these galaxies are not covered by the HST imaging. Similarly, for the Horseshoe, we used the available HST images in multiple bands ($F110W, F160W, F475W, F660W, F814W$) to perform the SED-fitting and derive the stellar mass. The uncertainties on stellar masses are derived from the 1σ interval of the resulting likelihood distribution and include also the contribution from statistical uncertainties on the magnification, but does not account for systematic uncertainties on the lensing model.

For the remaining two SINFONI galaxies we adopt the stellar masses provided in the literature, after a proper scaling (when needed) to the same Chabrier (2003) IMF. In detail, the stellar mass for MACS0451 is taken from Richard et al. (2011) and has been derived from a SED fitting including HST and Spitzer/IRAC data in the $3.6\mu m$ and $4.5\mu m$ filters with a Calzetti et al. (2000) extinction law. As pointed out in Section 3.2.2, we are here considering this system as constituted by a pair of galaxies for which we can compute separate metallicity maps and gradients. Therefore, in addition to the global $M_\star$ value, we used the continuum flux from the de-lensed HST H-band images as a proxy for the stellar mass distribution in order to split the global value and provide an $M_\star$ estimate for both the *northern* and *southern* region of this system. The $M_\star$ for CSWA164 is taken instead from Kostrzewa-Rutkowska et al. (2014) and has been derived from SED fitting to the SDSS photometry. All values reported in Table 3 are corrected for the $\mu_{H\alpha}$ magnification factor (or alternatively μ_{maps} if the former is not available.)

The global star formation rate for our sources is calculated from the extinction corrected Hα luminosity, which is converted to SFR assuming the Kennicutt & Evans (2012) relation and applying a scaling factor of 1.06 to convert from Kroupa et al. (1993) to Chabrier (2003) IMF. The amount of nebular reddening has been deduced from the Balmer decrement and corrected assuming an intrinsic ratio of Hα/Hβ = 2.87, adopting the Cardelli et al. (1989) extinction law. We finally corrected for the magnification factor derived from the de-lensing procedure (see Section 3.2) and reported in Table 3 as $\mu_{H\alpha}$. Both individual measurements errors on $E(B\text{-}V)_{neb}$ and the magnification factors are included in the total uncertainties on SFR.

The total Hα flux required to compute the SFR was estimated by co-adding the flux from all individual spaxels with robust Hα detection (i.e. above 5σ). Indeed, for extensive quantities like the SFR, different choices of the aperture width adopted to extract the integrated spectra would provide different measurements of the Hα flux and thus change the final inferred SFR, whereas for physical quant-

ities derived from line ratios, like the gas-phase metallicity, changing the aperture width has negligible impact on the results. On average, the fraction of the total Hα flux collected within the 0.6″-wide aperture is ∼ 75%. The lowest fractions (i.e around 50%) occur in composite or interacting systems, whose spatial emission line profiles can be significantly smeared or even double-peaked (for instance MUSE_SW_461-99-99).

Additional care has to be taken when computing the SFR and magnification for the Horseshoe galaxy. Our IFU observations of this object only encompass the western arc, hence do not allow us to sample the entire galaxy when re-constructed in the source plane. Therefore, we use additional SINFONI observations of the Horseshoe, conducted in 2009, which sample both the western arc and the southern high surface brightness peak. However, the observations from 2009, covering the full arc, only provide K-band data; therefore, for the rest of the current analysis, which depends on detecting multiple emission lines in different bands, we will use only the latest SINFONI multi-band observations (from 2013-2014) of the western arc of this system.

3.4 Metallicity determination

We derive the gas-phase metallicity using different diagnostics, exploiting the coverage of multiple emission lines. In this work we exploit the strong-line calibrations presented in Curti et al. (2017, 2019) (hereinafter C17 and C19). These works redefined the diagnostics from Maiolino et al. (2008) to fully anchor them to the T_e-abundance scale defined in the local Universe by the SDSS galaxies. A combination of the R_3 ($\log([\mathrm{O\,III}]\lambda5007/\mathrm{H}\beta)$), N_2 ($\log([\mathrm{N\,II}]\lambda6484/\mathrm{H}\alpha)$), S_2 ($\log([\mathrm{S\,II}]\lambda6717,31/\mathrm{H}\alpha)$) and O_3O_2 ($\log([\mathrm{O\,III}]\lambda5007/[\mathrm{O\,II}]\lambda3727,29)$) diagnostics, depending on source redshift and detectability of each emission line at $\geqslant 3\sigma$, are jointly used to tightly constrain the metallicity. We ran a Monte Carlo Markov chain algorithm (MCMC, implemented through the *emcee* package in python) to sample the log(O/H) posterior distribution, minimizing at each step the chi-square defined as:

$$\chi^2 = \sum_i \frac{(R_i^{\mathrm{obs}} - R_i^{\mathrm{exp}})^2}{\sigma_{\mathrm{obs}}^2 + \sigma_{R_i}^2}\,, \qquad (1)$$

where R_i^{obs} is the observed line ratio while R_i^{exp} is the expected value, according to each calibration, at fixed metallicity. The median of the resulting distribution is then assumed as the *true* metallicity and its 1σ interval defines the associated uncertainties. Both the uncertainty on the observed line ratio σ_{obs} and the intrinsic dispersion of each calibrated indicator σ_{R_i} are taken into account in the procedure. For most of the targets, the metallicity determination is based solely on the R_3 and N_2 (and, in some cases, S_2, see Sect.4.2) indicators, hence we do not require any reddening correction since all these diagnostics are constituted by ratios of very nearby emission lines. For z > 2 galaxies, where $[\mathrm{O\,II}]\lambda3727,29$ is detected and included in our routine, line fluxes are extinction corrected exploiting the Balmer decrement by assuming an intrinsic Hα/Hβ ratio equal to 2.87 and the Cardelli et al. (1989) extinction law. We note that in these cases the extinction correction often represents the largest contribution to the uncertainty in the metallicity determination, given the relatively low SNR of the Hβ line. In fact, for some of our galaxies we could not produce a reliable E(B-V) map to correct the line fluxes on a spaxel by spaxel basis. In such cases, we inferred the global E(B-V) value from the total Hα and Hβ fluxes and applied it to all spaxels, assuming that the extinction does not change dramatically across the galaxy.

3.4.1 Metallicity diagnostics at high redshift

We briefly discuss here the robustness of the metallicity measurements performed on our galaxies by means of strong-line diagnostics. At the present time, any result involving gas-phase metallicity measurements from strong-line calibrations at high redshift should always be interpreted with full awareness of these potential caveats. Nonetheless we also stress that, since the main results presented in this paper are based on relative metallicity measurements, they can be considered more robust against biases in the metallicity calibrations than those relying on absolute measurements.

As a general remark, whether the locally calibrated metallicity diagnostics are applicable to high-redshift galaxies is still a matter of great debate. Diagnostics that are expected to be little affected by the ionization conditions of the gas (see e.g. Dopita et al. 2016) have been suggested to be valuable for high-redshift galaxies, where strong variations in ionization parameter and excitation conditions compared to local galaxies have been invoked to explain the observed evolution in the emission line ratios (as seen for example from the offset of high-z sources in the classical BPT diagrams with respect to the local sequence (Kewley et al. 2013; Nakajima et al. 2013; Steidel et al. 2014; Kashino et al. 2017; Strom et al. 2017). However, since such diagnostics usually involve the $[\mathrm{N\,II}]/[\mathrm{O\,II}]$ or the $[\mathrm{N\,II}]/[\mathrm{S\,II}]$ line ratios, they are strongly dependent on the assumed relation between the N/O ratio as a function of the oxygen abundance O/H, which is affected by a large scatter and whose evolution with cosmic time and/or dependence on galaxy mass is also indicated as a possible origin of the observed evolution of the emission line properties in high-z galaxies (Masters et al. 2014, 2016; Shapley et al. 2015). Therefore, strong-line indicators based only on alpha-elements (like, e.g., oxygen) have also been suggested as appropriate to high redshift studies, since galaxies at z∼ 1.5 − 2.5 seem to show no appreciable offset from local trends in oxygen based diagnostic diagrams (e.g. R_{23} vs O_{32}, Shapley et al. 2015). However, the location on the abovementioned diagram could even be sensitive to a variation in the hardening of the radiation field at fixed metallicity rather than a variation in abundances (Steidel et al. 2016; Strom et al. 2017). In any case, at redshifts ∼ 1.5 (where the majority of KLEVER galaxies considered in this work lie), the lack of the $[\mathrm{O\,II}]$ doublet in the NIR bands observable from KMOS prevents us from using purely oxygen diagnostics, thus forcing us to exploit the nitrogen-based ones. When the survey will be complete, we will investigate the spatially resolved behaviour of z∼ 2 galaxies on the R_{23} vs O_{32} diagram in a more statistically robust manner. Recently, Patrício et al. (2018) have shown that oxygen-based diagnostics z∼ 2 provide metallicities comparable to those inferred from the electron temperature method; unfortunately, just a handful of robust auroral line detections have been reported so far in high-z sources (e.g. Jones et al. 2015b;

Sanders et al. 2016, see also Patrício et al. 2018 and references therein), due to the intrinsic observational challenges in detecting the faint auroral lines with current instrumentation. Only the advent of new facilities like JWST or the MOONS spectrograph on the VLT will ultimately allow us to tackle this issue in the next few years, allowing to properly calibrate the metallicity diagnostics against fully Te-based abundances determination at high redshifts and providing the key to overcome all these potential discrepancies.

We finally recall here that the strong-line calibrations adopted throughout this work are valid only if the gas is photoionised by stellar continuum from young massive stars, with no contribution to ionisation due to AGN or shocks. The sp9 galaxy presents clear signatures of a Seyfert 1 galaxy (with emission lines as broad as $\gtrsim$ 1000 km s^{-1}) and has been thus removed form the analysis. We further checked the possible contamination from AGN or shocks in the other galaxies of our sample in two different ways, which will be described more in detail in a forthcoming paper (Curti et al., in prep.). First, we found no clear correlation between spaxels lying above the theoretical dividing line of Kewley et al. (2001) and their distance from the central regions of the galaxy, as it would have reasonably been if the ionisation was driven by an AGN. Moreover, we find no clear trend between the observed BPT-offsets and velocity dispersion, as one would expect in case of shock-driven line ratios. Finally, we stacked the spectra of these spaxels from both the [N II] and [S II]-based BPT diagrams, to look for clear spectral signatures of AGN or Wolf-Rayet contamination (like prominent He IIλ4686 emission), but we did not find any evidence for those either. Therefore, we conclude that the contribution from AGN or shocks is negligible in the majority of cases and that our metallicity determination based on emission line ratios is reliable for the purposes that we pursue in this work.

4 RESULTS AND DISCUSSION

4.1 Global properties

We briefly discuss here the global properties of our sample by examining line ratios and physical quantities inferred from integrated spectra.

In Fig 5 we show the position of our galaxies on the BPT diagrams, named after Baldwin, Phillips & Telervich (Baldwin et al. 1981) and defined as [N II]λ6584/Hα vs [O III]λ5007/Hβ (the N2-BPT) and [S II]λ6716, 31/Hα vs [O III]λ5007/Hβ (the S2-BPT). We indicate in blue galaxies at z $\sim$ 1.4, while in red those at z $\sim$ 2.2. The sources observed within the RXJ2248 cluster are represented by filled circles, while empty circles mark those observed in MS2137. Finally, empty squares are assigned to the sample of strongly lensed galaxies observed with SINFONI, where both the northern and southern component of the MACS041 Arc are shown.

Consistently with typical findings at these redshifts, a systematic offset from the tight sequence occupied by galaxies in the local Universe (encompassed by the grey contours represented by SDSS galaxies) is seen on the N2-BPT diagram, observed towards higher [O III]/Hβ and/or [N II]/Hα ratios. Nonetheless, the position of our galaxies is still consistent, within the uncertainties, with the star forming region predicted by theoretical classification schemes like those proposed by Kewley et al. (2001) and Kauffmann et al. (2003) (the solid and dashed black lines respectively). On the S2-BPT diagram the offset is less prominent than in the N2-BPT diagram, and the points are more scattered. For comparison, we also show in Fig. 5 the fits to the average position of high-z galaxies on BPT diagrams as inferred from previous large surveys in the near-infrared (i.e. Shapley et al. 2015 for MOSDEF, Kashino et al. 2017 for FMOS and Strom et al. 2017 for KBSS). Compared to these studies, KLEVER provides spatially resolved information which could be used to investigate radial trends and to what extent the observed evolution in line ratios is driven locally (and by which physical effect) or whether it is a global property of the galaxy as a whole. For a more in-depth, spatially resolved analysis and discussion on the physical drivers of the BPT-offset for the KLEVER sample, we refer to a forthcoming paper by the collaboration (Curti et al., in prep.).

In the left panel of Figure 6 we show instead the distribution of our sample on the M$_\star$vs log(SFR) plane; colours and symbols are the same as in Fig. 5. The plotted error bars include the statistical errors on the magnification, but do not take into account the systematic uncertainties on the lensing models (see Sect. 3.2). For comparison, the Star-Forming Main Sequence (SFMS) at z = 0.07, z = 1.4 and z = 2.2 (i.e. the local Universe and the two average redshift intervals probed by the KLEVER sample) from Speagle et al. (2014) are shown as shaded regions, assuming a nominal intrinsic scatter of 0.2 dex. This is based on a compilation of many different works in the literature, which are predominantly based on UV/IR SFR indicators. Although Speagle et al. (2014) assume a Kroupa et al. (1993) IMF, converting to a Chabrier (2003) IMF introduce negligible effects. In addition, the best-fit to the SFMS derived in the context of the FMOS Survey (Kashino et al. 2019, their power-law equation at z$\sim$ 1.4) and the MOSDEF Survey (Shivaei et al. 2015, z$\sim$ 2.3) are also shown with green and orange solid lines respectively, with an observational scatter of 0.3 and 0.36 dex as marked by the dashed lines. The latter relations are based on dust-corrected Hα-SFR measurements, hence they are more consistent with the methodology followed in this work.

The sample of lensed galaxies analysed in this work allows us to extend the probe of ionised gas properties in the SFR vs M$_\star$ plane towards lower masses (in particular below $10^{9.5}$M$_\odot$) compared to typical magnitude-limited surveys of field galaxies, populating both the region below and above the representative SFMS at z$\sim$ 1.5 – 2.5. In particular, 8 galaxies at z$\sim$ 1.4 lie within 1σ of the Kashino et al. (2019) relation, while 7 galaxies fall more than 1σ below and 7 galaxies more than 1σ above the best-fit SFMS. However, we note that assuming a different parametrisation of the SFMS, like the broken power-law proposed in Kashino et al. (2019) (or by Whitaker et al. (2014)), or even considering the extrapolation of the Speagle et al. (2014) relation, would result in a steepening of the slope of the low-mass end of the relation, making the bias towards high specific star formation rates (sSFR) of the KLEVER sample in that regime more prominent. At z$\sim$ 2.2, our sample comprises only 1 galaxy below the Shivaei et al. (2015) SFMS parametrisation, while the remaining 9 lie above (although 3 of them are still consistent within the observational scatter). Remarkably, a few

Galaxy	SFR(Hα) [$M_\odot yr^{-1}$]	log($M_\star/M_\odot$)	12 + log(O/H)	$\mu^{(1)}_{H\alpha}$	$\mu^{(2)}_{maps}$
SINFONI Galaxies					
Horseshoe	54.60*±12.02	10.18±0.16	$8.49^{+0.04}_{-0.05}$	12.11±1.56	–
MACS0451 Arc (Total)	93.04±11.08	9.71±0.10	$8.24^{+0.05}_{-0.06}$	11.85 ±0.59	–
MACS0451 Arc (North)	56.08±6.74	9.50±0.12	$8.24^{+0.05}_{-0.07}$	9.18±1.01	–
MACS0451 Arc (South)	36.04±4.20	9.28±0.10	$8.25^{+0.04}_{-0.06}$	14.51±1.45	–
CSWA164	108.50±20.30	9.40±0.19	$8.45^{+0.06}_{-0.05}$	12.62±2.25	–
KMOS Galaxies					
MS 2137					
sp1	40.48±8.81	8.95±0.18	$8.37^{+0.04}_{-0.04}$	1.12±0.16	$1.22^{+0.01}_{-0.01}$
sp2	70.24±13.02	10.19±0.18	$8.43^{+0.04}_{-0.04}$	1.23±0.18	$1.27^{+0.01}_{-0.02}$
sp3	92.89±25.71	9.90±0.10	$8.45^{+0.03}_{-0.04}$	2.26±0.19	$2.57^{+0.12}_{-0.12}$
sp5	30.67±4.81	9.39±0.15	$8.30^{+0.05}_{-0.05}$	1.73±0.15	$1.76^{+0.04}_{-0.04}$
sp6	27.45±11.43	9.47±0.14	$8.34^{+0.02}_{-0.02}$	1.43±0.4	$1.55^{+0.02}_{-0.02}$
sp7	67.9±28.69	9.83±0.17	$8.55^{+0.04}_{-0.03}$	1.37±0.23	$1.4^{+0.02}_{-0.02}$
sp9	–	–	–	–	$1.1^{+0.01}_{-0.01}$
sp10	–	–	–	–	$3.62^{+0.22}_{-0.19}$
sp13	8.53±8.1	8.84±0.21	$8.14^{+0.05}_{-0.06}$	13.48±0.71	$25.51^{+15.84}_{-7.13}$
sp14	3.54±2.01	8.86±0.26	$8.3^{+0.04}_{-0.04}$	4.93±0.54	$5.69^{+0.6}_{-0.53}$
sp15	134.14±19.86	9.66±0.10	$8.18^{+0.03}_{-0.04}$	1.97±0.1	$2.28^{+0.09}_{-0.1}$
ph6532	21.70±3.21	9.58±0.38	$8.25^{+0.04}_{-0.04}$	1.07±0.14	$1.08^{+0.0}_{-0.0}$
ph2594	4.13**±1.88	9.58±0.14	$8.67^{+0.03}_{-0.03}$	1.28±0.44	$1.28^{+0.01}_{-0.01}$
ph3729	30.37±9.67	9.61±0.30	$8.46^{+0.1}_{-0.1}$	1.33±0.21	$1.34^{+0.02}_{-0.02}$
ph3912	114.12±25.57	9.24±0.12	$8.23^{+0.06}_{-0.06}$	1.22±0.13	$1.27^{+0.01}_{-0.01}$
ph5514	4.01**±0.07	–	$8.79^{+0.04}_{-0.04}$	1.40±0.38	$1.71^{+0.02}_{-0.03}$
ph7727	–	–	–	–	$1.07^{+0.0}_{-0.0}$
ph8073	199.55±46.72	9.36±0.26	$8.14^{+0.13}_{-0.16}$	1.0±0.19	$1.05^{+0.0}_{-0.0}$
RXJ2248					
GLASS_00093-99-99	6.05±1.13	9.40±0.13	$8.37^{+0.12}_{-0.12}$	1.62±0.02	$1.93^{+0.17}_{-0.17}$
R2248_LRb_p1_M3_Q4_58__2	23.51±2.62	10.38±0.07	$8.47^{+0.04}_{-0.04}$	1.36±0.01	$1.55^{+0.13}_{-0.07}$
MUSE_SW_462-99-99	4.55±1.51	9.83±0.12	$8.51^{+0.06}_{-0.06}$	3.21±0.03	$4.35^{+1.15}_{-0.4}$
GLASS_00333-99-99	4.54±0.76	8.41±0.11	$8.25^{+0.10}_{-0.12}$	1.24±0.01	$1.44^{+0.13}_{-0.05}$
R2248_LRb_p3_M4_Q3_93__1	26.02±1.38	9.76±0.08	$8.45^{+0.01}_{-0.02}$	1.23±0.01	$1.35^{+0.1}_{-0.03}$
R2248_MR_p1_M1_Q4_10__1	1.35**±0.07	–	–	1.01±0.02	$1.14^{+0.05}_{-0.02}$
R2248_LRb_p3_M4_Q3_94__1	17.24±0.66	9.67±0.19	$8.36^{+0.04}_{-0.04}$	1.22±0.01	$1.53^{+0.11}_{-0.09}$
MUSE_SW_48-99-99	–	8.30±0.21	–	–	$8.28^{+2.78}_{-2.58}$
MUSE_SW_51-99-99	–	7.30±0.22	–	–	$26.15^{+50.92}_{-13.78}$
R2248_MR_p3_M1_Q3_43__1	–	9.58±0.07	–	–	$28.61^{+57.4}_{-16.91}$
GLASS_01891-99-99	–	9.82±0.06	–	–	$1.88^{+0.15}_{-0.16}$
GLASS_01845-99-99	6.85±0.24	9.83±0.04	$8.09^{+0.06}_{-0.07}$	1.76±0.02	$2.19^{+0.25}_{-0.25}$
MUSE_SW_45-99-99	11.61±2.03	9.01±0.20	$8.32^{+0.06}_{-0.05}$	1.69±0.02	$2.15^{+0.21}_{-0.22}$
MUSE_SW_461-99-99	6.72±2.02	10.11±0.15	$8.52^{+0.05}_{-0.04}$	2.81±0.03	$3.66^{+0.4}_{-0.55}$
MUSE_NE_111-99-99	9.18±1.13	8.58±0.02	$8.35^{+0.04}_{-0.04}$	1.9±0.01	$2.28^{+0.33}_{-0.16}$
GLASS_01404-99-99	–	9.45±0.13	–	–	$1.4^{+0.12}_{-0.04}$
R2248_MR_p1_M1_Q4_51__1	0.65±2.0	8.62±0.10	$8.36^{+0.12}_{-0.10}$	–	$20.65^{+16.55}_{-8.34}$
GLASS_00800-99-99	0.6±0.07	9.97±0.09	$8.58^{+0.04}_{-0.04}$	4.83±0.08	$10.52^{+7.45}_{-2.87}$
MUSE_NE_23-99-99	1.40±0.07	9.37±0.07	$8.42^{+0.06}_{-0.06}$	2.64±0.03	$3.64^{+0.41}_{-0.55}$
R2248_MR_p1_M1_Q4_59__1	2.89**±0.78	10.90±0.07	$8.81^{+0.06}_{-0.06}$	1.86±0.02	$2.34^{+0.2}_{-0.23}$
MUSE_NE_117-99-99	–	7.75±0.25	–	–	$3.05^{+0.73}_{-0.66}$

Table 3. Global properties of the analysed sample. The star formation rate, the stellar mass and the oxygen abundance are reported for each source. SFRs are derived from the total H_α flux in the image plane and corrected for magnification. Stellar masses are derived from SED fitting, while global metallicities are computed from the integrated spectra adopting the Curti et al. (2017, 2019) calibrations. Galaxies with no measured $M_\star$ were not covered by multi-band HST images (but sometimes they are still covered by the lensing maps), while galaxies with no measured SFR have no Hα detection.
In the last columns we report two different estimates of the magnification factor: $\mu_{H\alpha}$ is derived from the ratio between the Hα flux in the image plane and that in the source plane, while $\mu_{\rm maps}$ exploits the magnification maps computed, for each source redshift, from the output files of the lens models adopted in this work (Zitrin et al. 2013). The uncertainties on the latter value are purely statistical, as computed from the standard deviation of 100 different realisations of the magnification maps, and do not take into account the systematic uncertainties on the lens models, which are of the order of 10% – 20% (see Section 3.2 for more details).
Notes - * Total SFR derived from SINFONI observations of both the western and southern arc. The SFR computed for the western arc only, which is the only region covered also by observations in the J and H band, is 26.2 ± 4.3 $M_\odot yr^{-1}$.
**: Hα not corrected for reddening due to absence of Hβ detection.

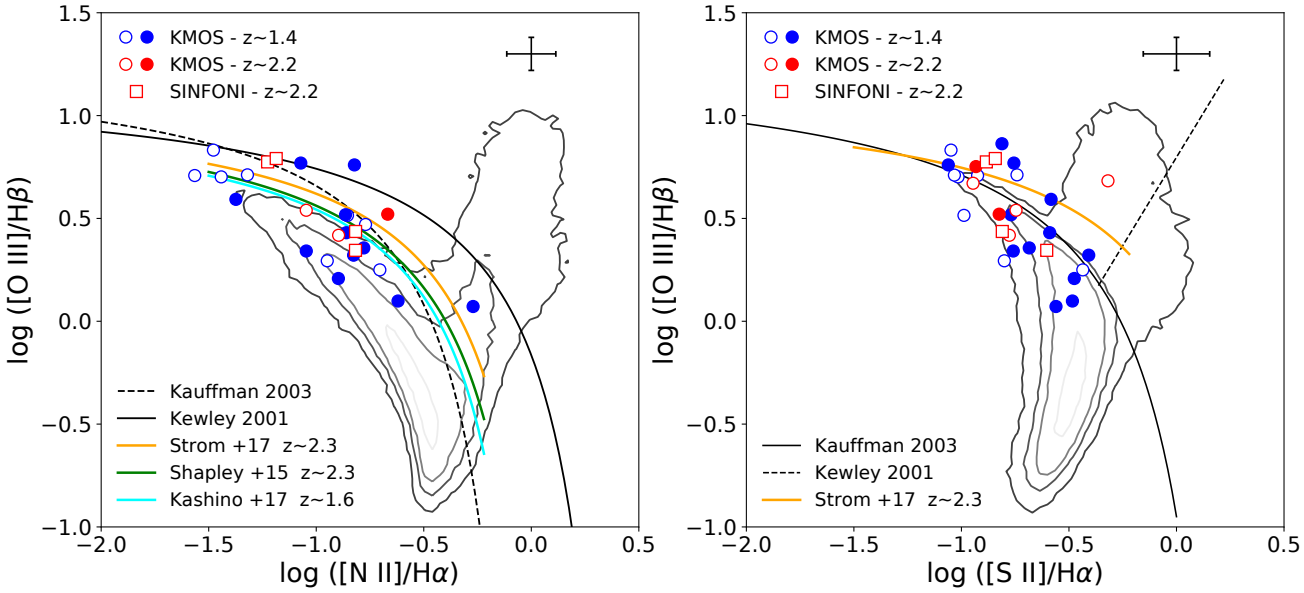

Figure 5. Global BPT diagrams for the KLEVER sample analysed in this work. In both panels the positions of the z ∈ [1.2, 1.6] sources are marked in blue, while that of the z ∈ [2, 2.4] ones are marked in red. Empty and filled circles denote galaxies belonging to the MS2137 and RXJ2248 cluster respectively, while empty squares represent the sample of strongly lensed galaxies observed with SINFONI. The theoretical demarcation line between the star forming locus and AGN/LINER galaxies from Kewley et al. 2001 and Kauffmann et al. 2003 are shown with solid and dashed black lines respectively, while the underlying grey contours encompass the 25, 75, 90 and 97.5% of the distribution of SDSS galaxies. On the N2-BPT in the *left panel* a fit to the location of high redshift galaxies from the FMOS survey (Kashino et al. 2017, z ∼ 1.6), MOSDEF survey (Shapley et al. 2015, z ∼ 2.3) and KBSS survey (Strom et al. 2017, z ∼ 2.3) is shown with cyan, green and orange lines respectively. The typical error on the emission line ratios is indicated in the upper right corner of each panel. Consistently with findings reported by the abovementioned surveys, many of our galaxies appear shifted from the position occupied by the bulk of the local galaxy distribution, especially in the N2-BPT.

galaxies in the sample are approximately ten times more star forming than the average population at those epochs. The impact that the distribution in sSFR of our sample has in the interpretation of the observed metallicity gradients is discussed in Sect. 4.2. The full KLEVER sample will comprise also a large number of more massive, unlensed sources from the CANDELS fields, which are complementary to this sample for a full exploration of the $M_\star$-SFR plane.

The right panel of Figure 6 shows instead the relation between stellar mass and oxygen abundance (the mass-metallicity relation, MZR) for our sample. The gas-phase metallicity is derived exploiting emission line diagnostics measured from the integrated spectra, following the scheme presented in Sect. 3.4 and the C17, C19 calibrations. The grey shaded area encloses the 68%, 84%, 95% and 99% of local SDSS galaxies and the relative best-fit MZR from Curti et al. (2019) is shown in green. In agreement with typical findings at these redshifts, our sample is characterized by lower metallicities, at fixed stellar mass, compared to the values assumed by local galaxies, indicative of the cosmic evolution of the MZR (see Maiolino & Mannucci 2019 and references therein). The mean offsets in metallicity, at fixed $M_\star$, from the local MZR are −0.25 and −0.36 dex for z ∼ 1.4 and z ∼ 2.2 galaxies respectively. In violet we also show the MZR at z∼ 2.3 from bins in stellar mass of individual KBSS galaxies (Steidel et al. 2014), derived exploiting the C17 calibration of the O3N2 indicator for consistency with the abundance scale adopted in this work; 8 out of 10 KLEVER galaxies at z∼ 2.2 with robust metallicity determination are consistent within the uncertainties with the Steidel et al. (2014) curve. Finally, we also plot for comparison the predicted mass-metallicity relations from the FIRE cosmological simulations at redshift 1.4 and 2.2, which are shown by the yellow and brown curves respectively, as presented in Ma et al. (2016). Despite the very different methodologies considered (T_e-based strong line methods versus prescriptions from zoom-in cosmological simulations), the predicted MZR are in reasonable agreement with our measurements for both redshift intervals considered, although showing a small systematic offset towards lower metallicities, possibly due to the large uncertainties affecting the stellar yields used in simulations (e.g. Wiersma et al. 2009).

4.2 Spatially resolved metallicity maps and gradients

We have derived the metallicity maps for our galaxy sample from the de-lensed source plane emission line maps, using the scheme presented in Sec. 3.4 and a combination of the diagnostics and calibrations presented in Curti et al. (2017, 2019). As general criteria we impose, on pixel-by-pixel basis, a SNR threshold of 5 on Hα and [O III]λ5007and of 2.5 on the other lines in order to include them in the metallicity calculation. To maintain a sufficient level of self-consistency, the same combination of diagnostics was used for all spatial elements across a given galaxy, in order to avoid possible systematics introduced by the differential dependence of each of the line ratios considered on the gas properties

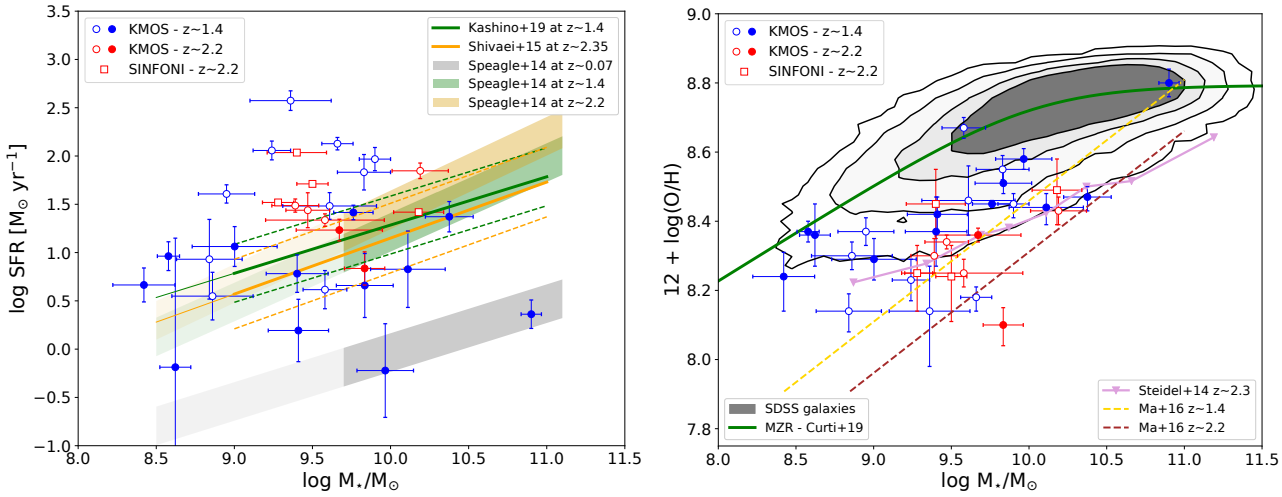

Figure 6. *Left Panel*: Distribution of our galaxy sample in the SFR vs $M_\star$plane. Symbols and colours are the same as in Fig. 5. Both SFR and $M_\star$measurements are corrected for lensing magnification and the error bars account for the uncertainties associated to the reddening correction and the statistical uncertainty on the magnification itself. The representative star-forming main sequence (SFMS) at redshift 0.07, 1.4 and 2.2 from Speagle et al. (2014) is represented by grey, green and yellow shaded regions respectively, with a nominal intrinsic scatter of 0.2 dex. The best-fit to SFMS (based on dust corrected Hα-based SFR) from Kashino et al. (2019) (FMOS Survey, z~ 1.4) and Shivaei et al. (2015) (MOSDEF Survey, z~ 2.3) are also plotted with green and golden solid lines, with an observational scatter of 0.3 and 0.36 dex respectively. All SFRs are reported to a common Chabrier (2003) IMF. *Right Panel*: Relationship between stellar mass and gas-phase metallicity for our sample, derived with the C17 and C19 calibrations as described in the text. Colors and symbols are the same as in the left panel. The region within grey contours encloses the 68%,84% 95% and 99% of the distribution of local SDSS galaxies in the $M_\star$ vs O/H plane, with abundances calculated from the same C17 and C19 calibrations, and the relative best-fit MZR to median points in bins of stellar mass from Curti et al. (2019) in shown in green. We also plot the MZR at z~ 2.3 from Steidel et al. (2014) in bins of stellar mass, shown in violet, as obtained from the O3N2 diagnostic rescaled to the C17 calibrations. All these metallicity measurements are consistent with the abundance scale defined by the T_e method. For comparison, we also show,with yellow and brown dashed curves respectively, the predicted MZR at z= 1.4 and z= 2.2 from high-resolution FIRE simulations, as presented in Ma et al. (2016). In agreement with the prescriptions of an evolving MZR, our galaxies are characterised by lower metallicities, at fixed $M_\star$, compared to local galaxies (0.24 and 0.36 dex lower on average than the local MZR for z~ 1.4 and z~ 2.2 sources respectively). However, C17-C19 calibrations provide higher metallicities compared to the abundance scale defined by both Ma et al. (2016) simulations at the reference redshift, possibly due to the uncertainties affecting the stellar yields adopted in the simulations.

of that particular source. However, the set of diagnostics involved in the metallicity calculation can vary from galaxy to galaxy, according to the availability and signal-to-noise of the different emission lines for each source. In general, the majority of our metallicity maps involves the simultaneous use of the R_3 and N_2 diagnostics. The S_2 line ratio is adopted instead of N_2 when a resolved gradient can not be obtained with N_2 or the latter is heavily contaminated by residuals of badly subtracted sky-lines. Moreover, for galaxies at z > 2 we could include the information from the O_3O_2 line ratio (once properly corrected for reddening) in the metallicity determination. The list of diagnostics adopted for each galaxy to derive the relative metallicity map and gradient is reported in the last column of Table 4.

To derive the metallicity gradients, we radially bin our data defining a series of elliptical apertures according to the shape of the source plane PSF. Increasing radii are taken in steps of a fraction of the linear PSF size, defined as the geometric mean of the model ellipse, i.e. $r = \sqrt{x \times y}$, where x and y represent the FWHM along the major and minor axis respectively (which are reported for each source in Table 4). This is aimed at partially overcoming the smearing effect introduced by the distorted source plane PSF, which would be particularly intense, especially at large distances from the centre, if we adopted simple circular apertures of increasing radius. The potential systematics associated to this choice are discussed in Appendix A. We then estimate the weighted average of the metallicity within each annulus; each value is weighted on the fraction of the relative pixel falling into the considered aperture. We note that the result does not change if we compute instead the average emission line ratios and re-compute the metallicity accordingly. The uncertainty associated with the average metallicity within each annulus takes into account both the 1σ dispersion of values of individual pixels inside the aperture and the error on each individual metallicity measurement (introduced by adopting our set of strong-line calibrations). The centre of our apertures is assumed to be the peak of the continuum emission in the H band or, alternatively, the peak of the Hα emission (hence the peak of the SFR surface density) for galaxies with no continuum detection. This choice does not affect the inferred gradients for the majority of the cases, with the exception of a few interacting systems with double-peaked (or strongly smeared) distribution of the Hα emission (see discussion in Sect. 4.5). Finally, the metallicity gradients are derived by performing a linear fit to the extracted metallicity values in each annulus and their uncertainties are evaluated by sampling the posterior distribution of the parameters through a Markov Chain Monte Carlo.

From the analysis of the source plane PSF we determine whether we are truly resolving a metallicity gradient or we are affected by poor angular resolution. We claim our

Galaxy	Metallicity Gradient [dex kpc^{-1}]	Resolved	PSF-FWHM (SP) [kpc × kpc]	Diagnostics
SINFONI Galaxies				
Horseshoe	0.012±0.014	Yes	1.82×0.37	R_3,N_2
MACS Arc (North)	-0.011±0.012	Yes	1.74×1.33	R_3,N_2,O_3O_2
MACS Arc (South)	-0.011±0.016	Yes	1.64×0.76	R_3,N_2,O_3O_2
CSWA164	-0.006±0.013	Yes	2.58×0.36	R_3,N_2,O_3O_2
KMOS Galaxies				
MS2137				
sp1	0.019±0.012	Yes	3.57×2.8	R_3,S_2
sp2	-0.007±0.015	Yes	3.48×2.65	R_3,O_3O_2
sp3	0.025±0.013	Yes	3.28×1.52	R_3,N_2
sp5	0.001±0.020	Yes	3.6×1.96	R_3,O_3O_2
sp6	0.063±0.012	Yes	3.39×2.18	R_3,O_3O_2
sp7	0.024±0.014	Yes	3.67×2.44	R_3,N_2
sp13	0.060±0.084	Yes	2.61×0.49	R_3,N_2
sp15	0.005±0.010	Yes	3.63×1.56	R_3,S_2
ph6532	0.036±0.012	Yes	3.65×3.01	R_3,O_3O_2
ph3729	0.002±0.048	Yes	3.62×2.56	R_3,N_2
ph3912	0.051±0.012	Yes	3.58×2.67	R_3,S_2
ph8073	0.006±0.105	Marg	3.59×3.15	R_3,S_2
RXJ2248				
GLASS_00093-99-99	0.011±0.039	Yes	4.14×2.61	R_3,N_2
R2248_LRb_p1_M3_Q4_58__2	0.002±0.010	Yes	4.46×3.01	R_3,N_2
MUSE_SW_462-99-99	-0.005±0.012	Yes	3.14×1.62	R_3,N_2
GLASS_00333-99-99	0.031±0.033	Marg	4.58×3.13	R_3,S_2
R2248_LRb_p3_M4_Q3_93__1	0.008±0.006	Yes	4.66×3.28	R_3,N_2
R2248_LRb_p3_M4_Q3_94__1	0.051±0.018	Yes	4.62×2.97	R_3,O_3O_2
GLASS_01845-99-99	0.029±0.023	Yes	4.1×2.27	R_3,O_3O_2
MUSE_SW_45-99-99	0.009±0.015	Yes	4.1×2.38	R_3,S_2
MUSE_SW_461-99-99	-0.019±0.018	Yes	2.56×2.52	R_3,N_2
MUSE_NE_111-99-99	0.012±0.038	Marg	3.8×2.42	R_3,N_2
GLASS_00800-99-99	-0.014±0.019	Yes	3.36×0.75	R_3,N_2
MUSE_NE_23-99-99	0.011±0.021	Yes	3.6×1.68	R_3,S_2

Table 4. Radial gradients for galaxies for which they could be at least marginally resolved (the *metallicity gradients sample*). We indicate to what degree each gradient has been resolved following the criteria described in Section 4.2. "Resolved" indicates galaxies where the gradient is extracted out to a radius larger than 2 times the linear size of the PSF-HWMH ($\sqrt{\mathrm{HWHM}_x \times \mathrm{HWHM}_y}$), while "Marginal" denotes that the gradient is extracted within a radial distance between 1 and 2 times this value. The third column report the major- and minor-axis FWHM (in kpc) of the ellipse used to model the PSF in the source plane, while in the last one we indicate the combination of strong-line diagnostics adopted to derive the metallicity map in the source plane for each galaxy.

gradients to be fully resolved if we can extract metallicities up to a radial distance at least twice the linear size of the PSF-HFHM (i.e. $\sqrt{\mathrm{HWHM}_x \times \mathrm{HWHM}_y}$), while we consider them marginally resolved if the gradient is evaluated between 1 and 2 times the linear PSF-HFHM; we finally classify them as unresolved otherwise. We show an example of a resolved and marginally resolved metallicity gradient in Fig. 7. From left to right we show respectively the source plane HST F160W image, the normalised source plane Hα map, the metallicity map and the extracted metallicity at increasing radii, with the linear fit to the points represented by the red line. The elliptical apertures adopted in the extraction of the radial gradient overlay with black contours the 2D metallicity map. The shape of the source plane PSF (third panel), as well as the linear size of the PSF-HWHM (in blue, fourth panel), are reported for each source to aid visualization.

Following the procedure described above, we manage to construct reliable metallicity maps and compute resolved (or, at least, marginally resolved) metallicity gradients for a total of 28 galaxies out of the 42 presented in Table 2. We refer to this subsample as the *metallicity gradients sample*. The derived values of the metallicity gradients for these sources are reported in Table 4, while the associated metallicity and line ratios maps are shown in Appendix B. Overall, we report 6 negative gradients and 22 positive gradients respectively, with gradients resolved for 25 out of 28 sources while only "marginally resolved" for the remaining 3. The 86% of the galaxies present gradients shallower than 0.05dex/kpc (and 72% shallower than 0.025dex/kpc). Moreover, we note that 25 out of 28 gradients (i.e. 89% of the sample) are totally consistent with a flat slope within their 3σ uncertainty (67% within 1σ), with only three galaxies significantly diverging from zero in showing inverted gradients.

4.3 Cosmic evolution of metallicity gradients

We discuss here the cosmic evolution of metallicity gradients, including the sample derived in this work in the context of the current observational framework and comparing the trends with those predicted by chemical evolution mod-

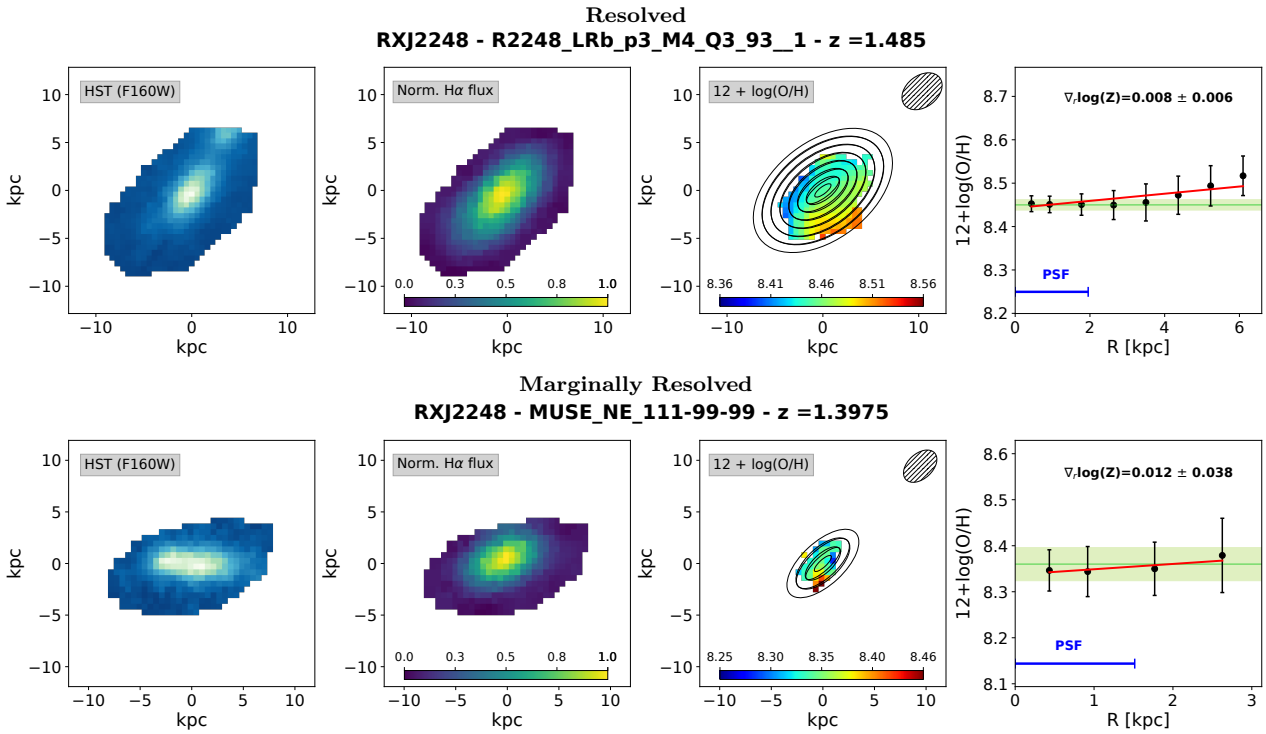

Figure 7. Two examples of metallicity maps and gradients for KLEVER galaxies, illustrating a resolved (*top* panel) and a marginal resolved (*bottom* panel) case. We show, from left to right respectively, the source plane HST reconstructed image in HST-F160w filter, the normalized source plane Hα flux maps, the full metallicity map and the extracted metallicities at increasing radii across each galaxy. The shaded black region in the *third* panel reproduce the shape and size of the PSF when mapped back into the source plane. The elliptical apertures used to derive the radial gradient overlay the 2D metallicity map. In the *fourth* panel, the linear fit to the metallicity gradient (in red) and the linear size of the PSF-HWHM ($\sqrt{\mathrm{HWHM}_x \times \mathrm{HWHM}_y}$, in blue) are also shown, while the green region encompass the uncertainty associated with the global metallicity of the galaxy (inferred from the integrated spectrum and marked in lime green by the horizontal line).

els and simulations. In Fig. 8 we plot the *metallicity gradients sample* as a function of redshift, together with a collection of radial abundance gradients (in units of dex/kpc) from previous works, both at high redshift (namely Swinbank et al. 2012; Queyrel et al. 2012; Jones et al. 2013, 2015a; Stott et al. 2014; Troncoso et al. 2014; Leethochawalit et al. 2016; Wuyts et al. 2016; Wang et al. 2017, 2019; Carton et al. 2018) and based on local galaxies (namely Rupke et al. 2010b; Magrini et al. 2009, 2010; Stanghellini & Haywood 2010; Stanghellini et al. 2014; Stanghellini & Haywood 2018). A brief description of the samples involved and the observational techniques adopted in the determination of the metallicity gradients collected from the literature is given in the caption of Fig. 8. The gradients derived in this work for the KLEVER galaxies are shown with dark blue filled circles (for resolved gradients) and light blue circles (for marginally resolved gradients, see Sect. 4.2). The results presented in this paper are generally consistent with previous investigations of metallicity gradients at high redshift: despite the variety of observational set up, resolution, sample selection, metallicity calibration etc.. the bulk of observed star forming galaxies does not exhibit, on average, strong radial trends (i.e. above 0.05dex/kpc) in the metal distributions between 1 <z< 2.5, with a large part of them characterised by gradients consistent with being flat within their 1σ uncertainty.

We can compare these results with the predictions from cosmological hydrodynamical simulations which trace the gas phase metallicity of disc galaxies from $z \sim 2$ to the present day under different feedback modes in the framework of the MUGS (Stinson et al. 2010) and MaGICC (Gibson et al. 2013b) schemes. We show the evolutionary track for the simulated galaxy g15784 under both *normal* and *enhanced* feedback conditions. The "normal" feedback scenario, in which 10 to 40% of the energy from each supernova is used to heat the surrounding ISM, predicts steeper metallicity gradients at earlier redshifts (i.e. ~ -0.2dex/kpc) which become flatter over cosmic time. The totality of our measurements and even the largest part of those reported in the literature are inconsistent with this scenario, with just a few example of observed gradients that match its predictions at redhifts higher than 1.5 (Jones et al. 2013; Wang et al. 2017). The enhanced feedback scheme instead, where outflows re-distribute energy and re-cycle ISM material over larger galactic scales, with preferential re-accretion into the outer regions (i.e. metal re-cycling), is characterised by a milder evolution with a progressive (minimal) flattening of the metallicity gradients with increasing redshift. Our results, together with the majority of previous indications from the literature, seem to point towards this latter regime, which suggests that at high-z, stronger star formation feedback is in place causing flatter gradients due to rapid and efficient recycling of enriched gas and re-distribution of the ISM, which occurs on timescales shorter that those related to star formation and metal enrichment. Within the

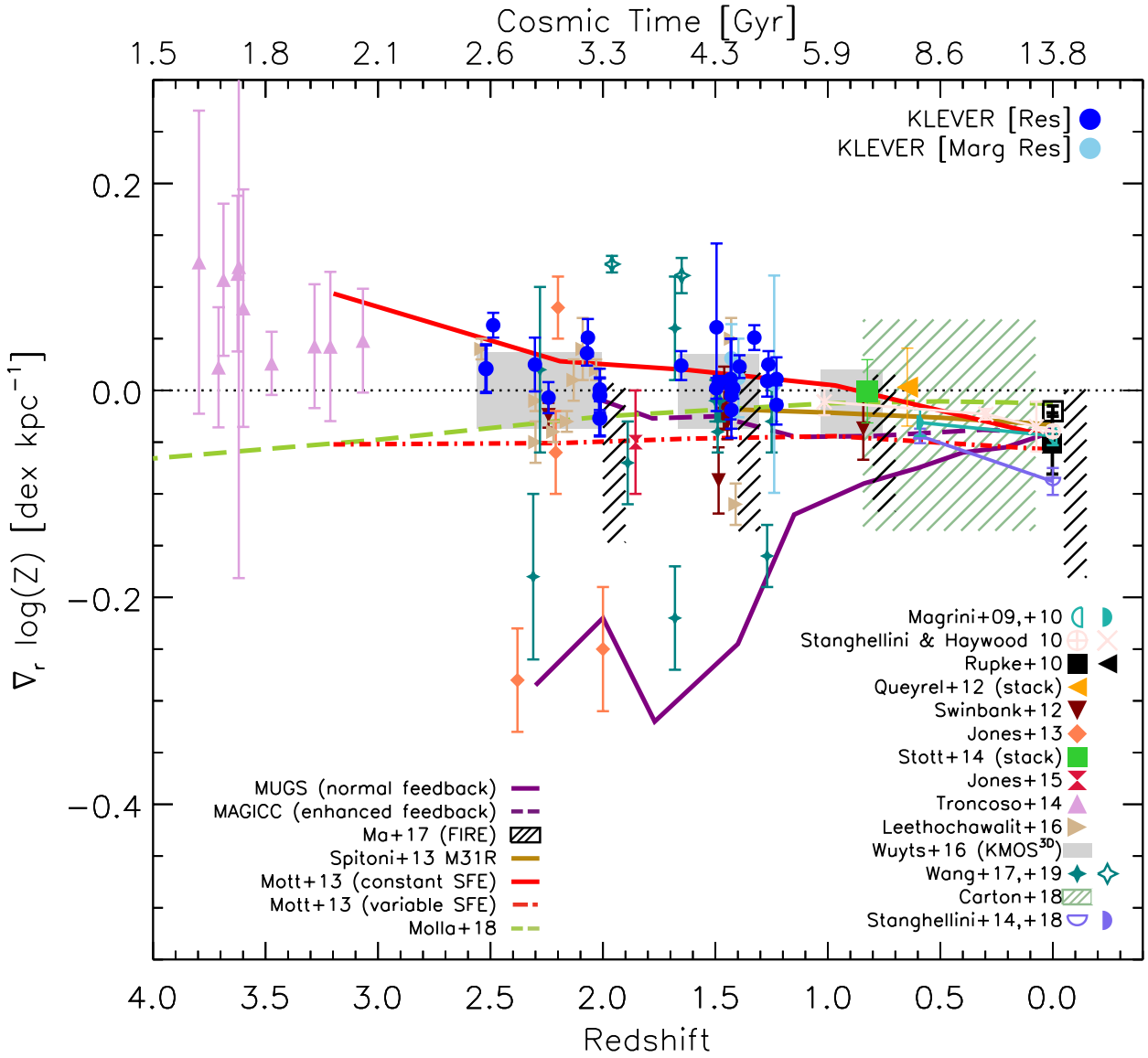

Figure 8. Compilation of metallicity gradients plotted as a function of redshift (or equivalently the age of the Universe). Blue points mark the results derived in this work, where we distinguish between resolved (blue circles) or marginally resolved (light blue circles) gradients. Additional measurements of metallicity gradients from previous works, derived with different techniques on both local and high-z galaxies, are included with different symbols/colors as explained in the legend. The metallicity gradients by Carton et al. (2018) are based on MUSE observations of $0.1 < z < 0.8$ galaxies, for which we report the spread in their measurements both in redshift and ∇Z. The sample from Queyrel et al. (2012) reports SINFONI observations of z$\sim$ 1.2 galaxies from the MASSIV survey, while Stott et al. (2014) reports KMOS observations of z$\sim$ 0.8 galaxies from the HiZELS survey. For these works, we report the average and 1σ dispersion for their metallicity gradients measurements at the mean redshift of each sample. The grey shaded areas, instead, represent the spread in the metallicity gradients measured in the framework of the KMOS^{3D} Survey within three different redshift intervals and presented by Wuyts et al. (2016). Jones et al. (2013) and Leethochawalit et al. (2016) report observations conducted with AO-assisted spectroscopy with OSIRIS at Keck on gravitationally lensed systems, while Jones et al. (2015a), Wang et al. (2017) and Wang et al. (2019) are based on HST-grism slitless spetroscopy of lensed galaxies behind the Frontier Fields clusters as part of the GLASS program (Treu et al. 2015). The metallicity gradients a z$>$ 3 from Troncoso et al. (2014) are based on SINFONI observations in the framework of the AMAZE and LSD programs (Maiolino et al. 2008; Mannucci et al. 2009). For what concern the metallicity gradients in the local Universe, we show the average of local gradients for isolated (filled square) and interacting (open square) systems from Rupke et al. (2010b). We also report the time evolution of metallicity gradients for the Milky Way, M33 and M81, as inferred from abundances computed for HII regions (open half-circles) and Planetary Nebulae (PNe, filled half-circles, associated to older stellar population progenitors) and presented in Magrini et al. (2009, 2010); Stanghellini & Haywood (2010); Stanghellini et al. (2014); Stanghellini & Haywood (2018). The predictions from MUGS and MaGICC cosmological simulations reported in Gibson et al. (2013a), which tracks the evolution of a typical disc galaxy implementing two different feedback modes, are shown with a dashed purple line for the enhanced feedback scenario and with a straight purple line for the normal feedback scenario respectively. The black hatched regions cover instead the predictions of metallicity gradients at four different epochs (i.e. z= 0, 0.8, 1.5, 2) from the FIRE simulations, as presented by Ma et al. (2017). The brown curve shows the predicted cosmic evolution of the metallicity gradient from the Spitoni et al. (2013) modelling of M31, while in light-green those from the chemical evolution models by Mollá et al. (2019). Finally, the predictions of chemical evolution models by Mott et al. (2013) assuming constant (solid line) and variable (dot-dashed line) star formation efficiencies are plotted in red. Our results are consistent with little or no evolution of the metallicity gradients with redshift, in better agreement with predictions by simulations and models including strong feedback mechanisms and/or gas radial flows in place.

context of the high resolution FIRE simulations, Ma et al. (2017) presented gas-phase metallicity maps and measured radial gradients within $0.25 - 1R_{90}$[2] for a sample of simulated galaxies up to z$\sim$ 2. They are capable of predicting a variety of metallicity distributions and gradients at each cosmic epoch (represented by the black hatched regions in Fig. 8), spanning from strongly negative ones in ordered discs to flatter gradients in more perturbed systems affected by mergers and significant gas flows, which mix the ISM on large galactic scales ($\lesssim$ 10kpc) on relatively short timescales ($\sim 10 - 50$Myr), nicely reproducing the scatter in the observational results.

In Fig 8 we also show the prediction for the metallicity gradient of M31 at different redshifts according to the Spitoni et al. (2013) chemical evolution model (solid brown line) which assumes no gas threshold on star formation, an inside-out formation of the disk, constant star formation efficiency along the disk and radial gas flows; this model is consistent with enhanced-feedback simulation schemes in prescribing a mild steepening of the gradient with time. Interestingly, no or little evolution in the predicted metallicity gradients can be achieved also by purely- or semi-analytical chemical evolution models which do not include any prescription about feedback processes mixing up the gas. The models from Mollá & Díaz (2005) (then revisited by Mollá et al. 2017, 2019, green dashed line), for example, assuming an inside-out disc growth, an ISM structured as a multiphase mix of diffuse gas and cold clouds and prescribing star formation efficiencies, stellar yields, gas infall rates and atomic to molecular hydrogen conversion factors, predict flat metallicity gradients at all redshifts within 2.5 effective radii in disc galaxies, with no assumption on radial gas mixing.

Among our galaxies, we report several examples of positive, 'inverted' metallicity gradients, despite only three determined with $> 3\sigma$ significance (i.e. sp6, ph3912 and ph6532, with R2248_LRb_p3_M4_Q3_94_1 at 2.8σ). Previous claims of such gradients have been reported both at the same redshifts (Wuyts et al. 2016; Wang et al. 2019) and at earlier epochs (z > 3), mostly in seeing-limited data of unlensed galaxies (Cresci et al. 2010; Troncoso et al. 2014); similarly to these findings, the central metal poor regions in our galaxies are co-spatial with the regions of highest star formation rate density (and hence higher gas fractions). These observations have been interpreted as evidence of cold flows of pristine gas into the central regions of such galaxies, which both fuels new star formation episodes and dilutes the metallicity (Dekel et al. 2009; Cresci et al. 2010). An alternative explanation consist in considering these galaxies affected by strong feedback processes, in the form of outflows that rain material from the inner regions back down in the outskirts of the galaxy (the so called 'galactic fountain' effect, Fraternali & Binney 2008; Werk et al. 2011). This process, whose primary effect results into an average flattening of the abundance gradients, can contribute in some cases to invert the gradient itself, as the highest mass loss rate is observed in the central regions (Wang et al. 2019). However, the impact of these outflows does not seem to be powerful enough to represent the dominant contribution to such a flattening for the ensemble population, since the measured mass outflow rates are often too small (Cresci et al. 2010; Troncoso et al. 2014). In the framework of chemical evolution models, Mott et al. (2013) managed to reproduce inverted gradients at z> 1 assuming the inside-out formation of the disc, a threshold in the gas density for star formation, a variable star formation efficiency along the disc and the presence of radial flows with varying speed. In particular, they find the velocity pattern for radial gas flows to be a crucial parameter in shaping the chemical evolution of the disc and that a constant star formation efficiency along the disc is needed to reproduce the observed inverse gradients (solid red line in Fig. 8).

Finally, It is worth noting here that the large majority of currently available results rely on a variety of different indicators for metallicity determination: many of them are derived only from nitrogen based diagnostics ([N II]/Hα in particular, being one of the most easily accessible at high-z), some from purely oxygen based ones. Moreover, beside the fact that the metallicity gradients shown in Fig. 8 are obtained with different diagnostics, the latter are sometimes even based on different calibration methods (e.g. T_e versus photoionisation modelling). This means not only that the absolute metallicity estimates could disagree between various methods, but also that the dynamic range of possible values spanned by the calibration itself could be different, which might translate in a change of the slope of the calibrated relation. In general, theoretical calibrations provide a wider range of possible metallicities with respect to T_e-based methods, even when considering the same set of strong-line ratios, as they often provide higher normalisations in the high-metallicity regime. Therefore, the use of different diagnostics/calibrations could introduce systematics in the slope of the inferred gradients. However, given the generally small range of abundances spanned by individual regions across a galaxy (very strong gradients are rare), and considering that relative metallicity measurements are always much more robust than absolute ones, this effect is likely to be negligible in the majority of cases, making the comparison of samples derived with different techniques more fair.

4.4 Correlation with stellar mass and SFR

We discuss here whether the metallicity gradients correlate with global galaxy properties. As previously mentioned, simulations of the inside-out growth of galaxies predicts initial steep negative gradients which flatten at later times, whereas observed positive or flat gradients are predicted by simulations and models as due to efficient re-accretion and gas mixing processes (e.g. Gibson et al. 2013b; Mott et al. 2013). In particular, the specific star formation rate (sSFR=SFR/$M_\star$) is a measured quantity that holds information about galaxy star formation history: galaxies exhibiting high sSFR are experiencing a starbursting phase and might be in an early stage of their cosmic evolution, possibly fuelled by inflowing gas into their central regions and affected by strong feedback episodes. One can therefore expect that signatures of this process might also be observed in the metallicity gradients, in the sense that galaxies with higher sSFR exhibit on average flatter (or more positive) gradients than those characterised by low sSFR.

In the upper panel of Fig. 9 we plot the *metallicity gradi-*

[2] the radius enclosing 90% of the star formation within 10 kpc

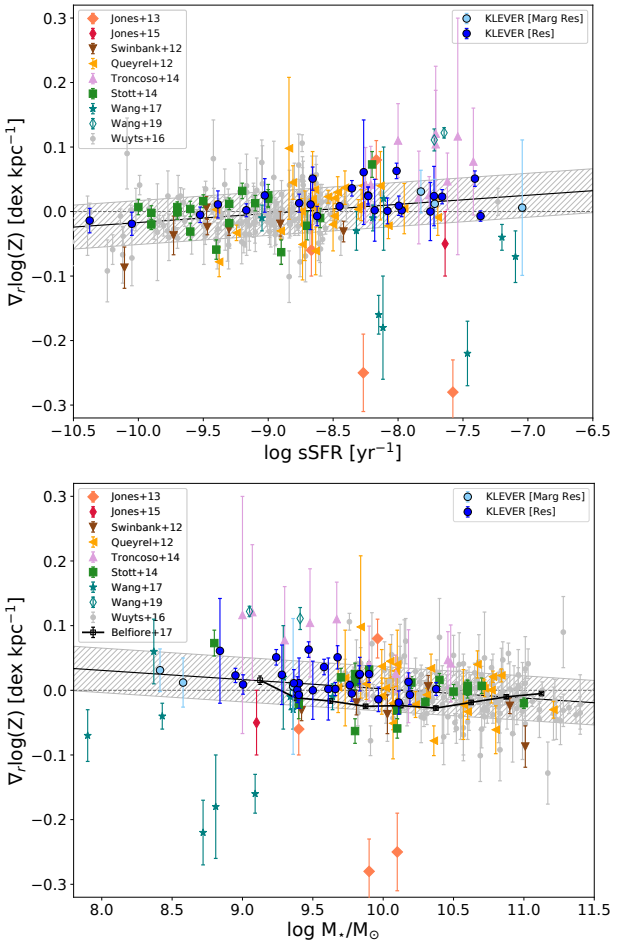

Figure 9. Metallicity gradients as a function of global specific star formation rate (sSFR, *Upper Panel*) and stellar mass (*Bottom Panel*). The sample includes both objects from KLEVER and additional gradients from previous works on z$\gtrsim$ 0.8 sources; symbols and colors are the same as in Fig. 8. A robust linear regression fit to the data is shown in black in both panels, with the RMS region marked in grey. A mild correlation is present between metallicity gradients and both quantities, in the sense of a positive trend with increasing sSFR and a negative trend with increasing M$_\star$, significant at 3.2σ and 2.7σ and with a Spearman correlation coefficient $\rho = 0.29$ and $\rho = -0.25$ respectively. In the bottom panel, the trend between gradients and stellar mass in the local Universe (inferred from a sample of galaxies in MANGA, Belfiore et al. (2017)) is traced by the black line.

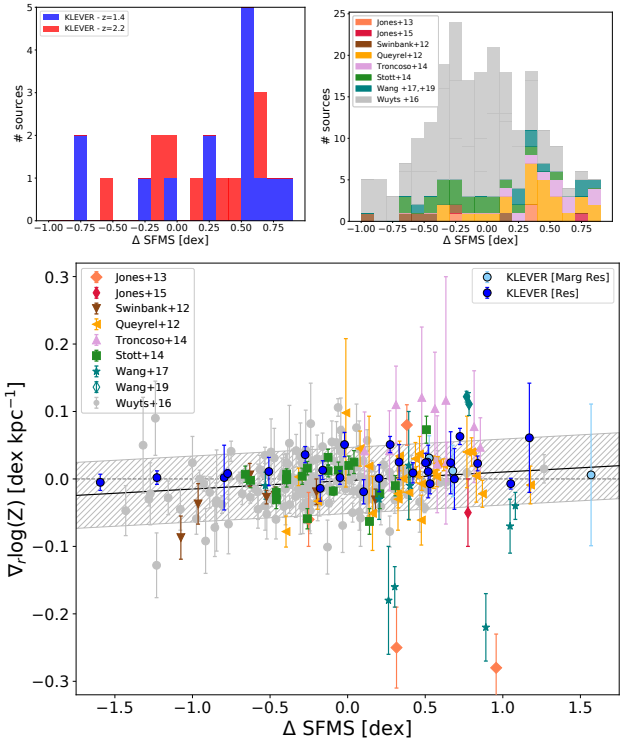

Figure 10. *Upper Panels:* Stacked histograms of the distribution in Δ SFMS, defined as the distance from the star-forming main sequence of Speagle et al. (2014) at the mean redshift of each reference sample, for the KLEVER galaxies hereby analysed (left panel) and the compiled sample from the literature at z$\gtrsim$ 0.8 (right panel). *Bottom Panel:* Metallicity gradients as a function of Δ SFMS. Compared to the trends shown in Fig. 9, the significance is reduced to $\sim 2\sigma$, with Spearman coefficient $\rho = 0.20$.

ents sample as a function of sSFR, along with gradients from previous works at $z \gtrsim 0.8$ (i.e. Swinbank et al. 2012; Queyrel et al. 2012; Jones et al. 2013; Troncoso et al. 2014; Stott et al. 2014; Jones et al. 2015a; Wuyts et al. 2016; Wang et al. 2017, 2019)). We exploited the *sklearn* python package to perform a robust linear regression and quantify the level of correlation: the result is shown with the black line in Fig. 9, with the shaded region indicating the r.m.s. of the fit. The slope of the relation is equal to 0.013 ± 0.004 (i.e. 3.2σ from being flat), with an r.m.s. of 0.05dex, suggesting that a correlation between metallicity gradients and sSFR might be present in the sense of most active galaxies (i.e. higher sSFR) being characterized by flatter or positive gradients. This is consistent with the trend originally found by Stott et al. (2014) and the predictions of the FIRE simulations in Ma et al. (2017). Nevertheless, the level of correlation is very mild, as a Spearman test gives only $\rho = 0.29$ with a 99.9% level of confidence.

We investigate also whether our metallicity gradients correlate with total stellar mass in the bottom panel of Fig. 9. We perform the same linear regression and find a slightly negative slope (-0.013 ± 0.005), significant at the 2.6σ level. A Spearman correlation test gives $\rho = -0.25$ with the 99.9% of confidence. Again, this is consistent with previous observational assessments (Stott et al. 2014; Wuyts et al. 2016) and simulations (Ma et al. 2017). In the local Universe, a clear trend of metallicity gradients with stellar mass is observed, as recently shown for example by Belfiore et al. (2017) for a large sample of galaxies within the MANGA survey: less massive systems are characterized by flatter metallicity gradients, steepening (i.e. becoming negative) with increasing stellar mass up to $10^{10.5}$ M$_\odot$, before flattening again at higher masses (especially in their central regions). In the framework of the classical inside-out growth scenario, and assuming a radially decreasing star formation efficiency, feedback and/or gas recycling mechanisms are required to reproduce the observed flattening of the gradients at low masses (see also Belfiore et al. 2019).

As also seen in Sect. 3.3, the sample analysed in the present paper is distributed across a wide region of the M$_\star$-SFR plane, although being preferentially constituted by

higher sSFR galaxies than representative samples at the corresponding redshifts, especially at z > 2. Fig. 10 shows how the *metallicity gradient sample* analysed in this paper and the compiled sample from the literature are distributed as a function of Δ SFMS, defined as the distance from the star forming main sequence computed at each source redshift, adopting the redshift-dependent parametrisation of Speagle et al. (2014). In the bottom panel of Fig. 10 we investigate the correlation between metallicity gradients and Δ SFMS. When normalised to the main sequence at each redshift, the correlation is found to be weaker compared to the trends previously discussed, with $\rho = 0.20$ and a slope significantly different from being flat only at the 2σ level.

In light of the trends shown and discussed above, we note that it might be therefore more likely for KLEVER galaxies to show flatter gradients compared to the rest of the complied sample, although flat gradients are found at all $M_\star$and SFR within our sample. Nonetheless, the full high-z sample shown in Fig. 8 is almost uniformly distributed in Δ SFMS, hence the overall bias introduced in the interpretation of the cosmic evolution of the gradients is supposed to be minimal. The only exception is the sample of z> 3 galaxies from Troncoso et al. (2014), which appears to be constituted by galaxies lying systematically above the SFMS at that redshift. Therefore, it is possible that the incidence of inverted gradients in the population, as inferred from that particular sample, is overestimated, modulo the uncertainties associated with the parametrisation of the SFMS at high redshift which should always taken into account when performing this type of analysis.

Interestingly, Wang et al. (2017) report a few examples of steep, negative gradients (< −0.1 dex/kpc) at low masses (i.e. $M_\star < 10^9 M_\odot$) from their HST-grism spectroscopic survey of lensed galaxies, which deviate both from the average trend followed by the rest of the high-z sample and from the picture outlined in the local Universe. According to the authors, they could be interpreted as systems with a rapid and highly centrally concentrated star formation history and limited feedback mechanisms and are, to date, the only reported systems that fit the "purely inside-out growth' scenario of mass assembly. However, a larger number of robust measurements of metallicity gradients in low mass galaxies is required in order to carefully estimate the occurrence rate of such systems and whether they might represent a different galaxy population. In this sense, the full KLEVER sample will strongly contribute to increase the statistics in the low-mass regime.

4.5 Are radial gradients always meaningful ?

A first look to the derived metallicity gradients reveals how the vast majority of the sample is characterised by radial gradients consistent with being flat within their uncertainties. However, we note that some of the metallicity maps are characterised by the presence of non-axisymmetric patterns, revealing a more complex distribution in the level of chemical enrichment. This means that a formally flat radial metallicity gradient may be either the consequence, on one hand, of a fairly homogeneous metallicity map, resulting from the mixing of the ISM on large scales or from a flat, extended star formation profile, while on the other hand can hide the presence of non-radial variations (even on ~ kpc scales) which are washed out (producing a flat gradient) when azimuthally averaged. On average, the maximum metallicity variation between different regions in most of our galaxies is around ~ 0.1 − 0.2 dex; however, when this regions are not symmetrically distributed around the galaxy centre, they might be hidden by the azimuthal averages and the resulting gradient would result artificially flatter. Interestingly, some of those galaxies showing irregularities in their emission line maps and/or velocity fields, present clear signatures of disturbed morphology in the HST images, which could reveal ongoing interactions or suggest clumpy structures in their disks. A few examples are discussed in the following and shown in Fig. 11, where both image plane and source plane HST images, velocity fields and metallicity maps are reported for four different systems.

The complex kinematics seen in the velocity fields of sp15 suggests a primary galaxy (associated with the Hα peak and with a clear velocity gradient) interacting with a fainter companion responsible for the redshifted Hα emission towards the bottom right part of the image. The Hα emission is elongated towards the north-est, co-spatial with the "tail" visible in the HST images. The metallicity map is strongly irregular, with an apparent spatial anti-correlation between the intensity of the Hα and the metallicity. However, the azimuthal asymmetry is partially washed out when averaging, with the final gradient equal to -0.005 ± 0.010. A similar behaviour can be seen for the galaxy sp1, which shows the presence of multiple components, clearly visible in the F814W image, with the centroid of the extended Hα emission misaligned from the peak of the F160W continuum, more representative of the stellar mass density distribution. The velocity map is nevertheless quite regular and does not show any clear deviation from an ordered disk-like rotation pattern (making its interpretation often non trivial in seeing-limited data, see e.g. Simons et al. 2019). A clear linear gradient in the metallicity map can be seen from the most active, metal poor region in the south-est area to the more metal rich region towards the north-west direction, which is associated with the main disk component (as marked by the peak of H-band continuum image); however, the gradient is lost if radially averaging around the centroid of the emission in the system. MUSE_SW_461-99-99 is characterized by the presence of a distinct Hα blob in the north-west region of the image plane map (south-west in the source plane), which is not detected in the KMOS H-band continuum map nor can be clearly associated with any visible structure in the broad band HST-F160w image (nor in the de-lensed velocity field). However, the HST-F814w image reveals a more complex morphology, with an offset component (likely a companion galaxy) aligned with the secondary Hα peak. It is worth noting how in this case the metallicity of the region co-spatial with the secondary Hα blob is significantly lower than that of the central galaxy. Finally, R2248_LRb_p1_M3_Q4_58__2 presents a bright, elongated knot westward (~ 1.5 kpc far) of the centre in the F814W image which is not detected in H band continuum. However, the centroid of the Hα emission is shifted in the same direction, and the region is characterised by a slightly enhanced level of enrichment.

As we have seen, these configurations may be either related to the presence of close interacting companions or even suggest that different processes, occurring on shorter timescales than secular processes and mainly associated to

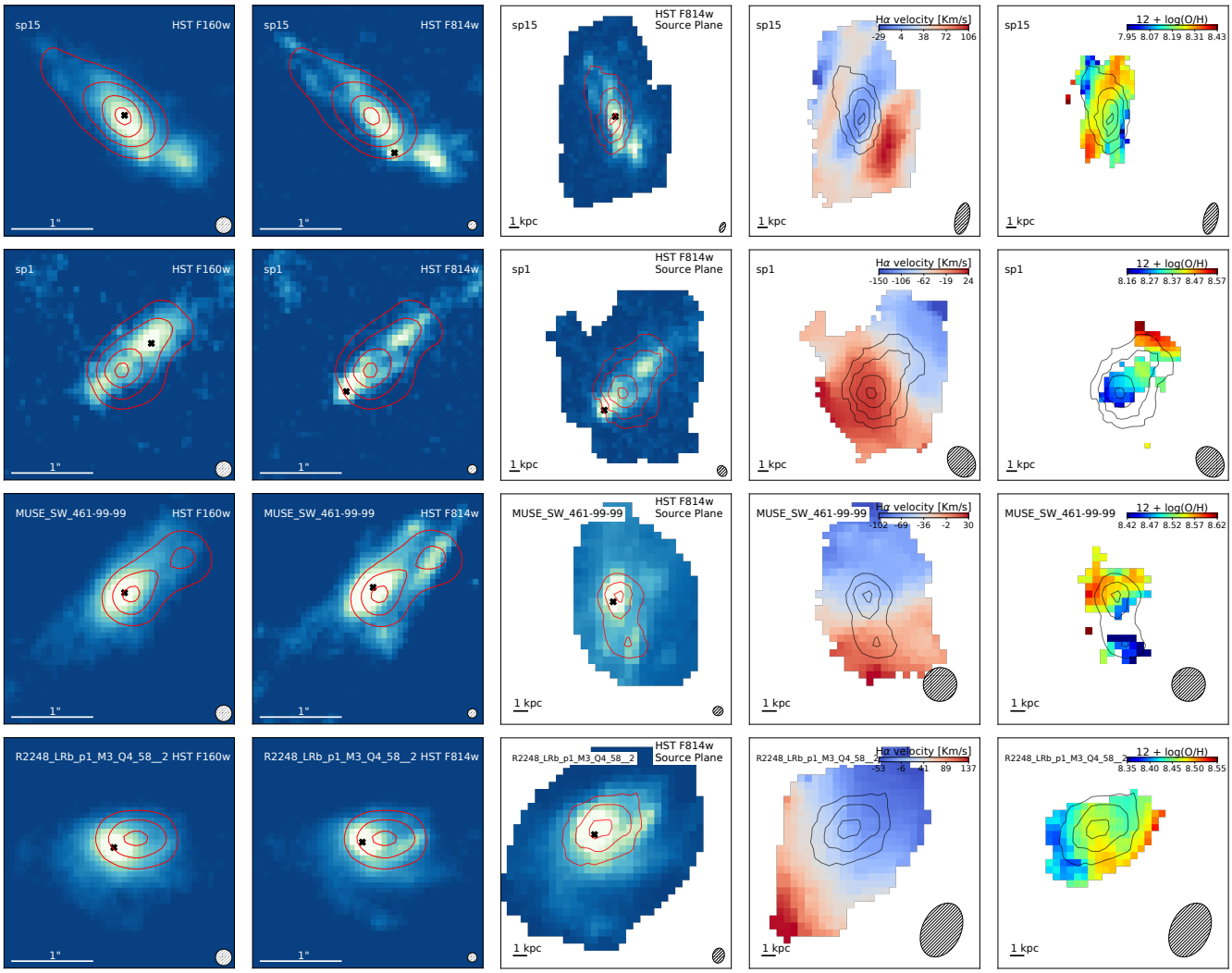

Figure 11. Examples of systems showing disturbed morphologies, kinematics and metallicity maps. From *left* to *right*: HST F160W image, HST F814W image, reconstructed HST F814W source plane image, Hα velocity field and metallicity map are shown for four galaxies analysed in this work. These galaxies present extended Hα emission or secondary peaks misaligned from the bulk of the underlying H-band continuum. These regions of intense off-nuclear Hα emission may be either associated with episodes of recent star formation occurring in compact clumps which form in the turbulent gas-rich disk or the results of merger events which, however, can not always be obviously confirmed from the velocity fields. These regions are often characterised by a different level of chemical enrichment compared to the bulk of the system. In each panel, the overlaid contours encompass the 50%, 68%, 85% and 97% of the Hα emission, while black crosses mark the position of the peak of the underlying HST continuum image.

gas flows, could contribute to the complexity seen in some of the 2D metallicity maps. High resolution simulations of disc galaxies (Milky Way-like at redshift zero) have shown in fact that strong fluctuations in their star formation history and gas outflow rate occurring on relatively short timescales (i.e. $\lesssim$ 1Gyr) can induce significant fluctuations in the gas-phase metallicity maps and hence in the derived radial gradients (Ma et al. 2017). In general, at the peak of the cosmic star formation history (z$\sim$ 1.5 – 2.5, the epoch probed by KLEVER), galaxies are often characterized by irregular morphologies and their emission lines luminosity might be dominated by the presence of young, massive star forming clumps (Elmegreen 2009; Förster Schreiber et al. 2011; Genzel et al. 2011). Recent cosmological simulations by Ceverino et al. (2016) also show that in 50% of cases at redshifts lower than 4, pristine gas accretion from the cosmic web can give rise to the formation of local star-forming clumps and subsequent drop in metallicity ($\sim$ 0.3 dex) compared to the surrounding ISM. They further assert that the accretion should be rapid enough to sustain such a metallicity drop, as the turbulence mixing would dissolve these features within a few disc dynamical timescales. Another possibility is that these clumps may form within their gas-rich disks due to gravitational instabilities (Bournaud et al. 2007; Elmegreen et al. 2008; Genzel et al. 2008); in this case, they might leave a different imprint on the metallicity maps, showing similar level of chemical enrichment to those of the material from which they have formed. The presence of such regions of different star formation and chemical enrichment histories within the same galaxy may reflect itself in the azimuthal asymmetries seen in some of the metallicity maps, despite the impossibility to resolve them spatially given the intrinsic

limitations of seeing-limited observations in reaching the required resolution (~ 100pc, Tamburello et al. 2017). Nonetheless, recent observations conducted with SINFONI AO-assisted spectroscopy have also reported similar asymmetries in the distribution of the [N II]λ6584/Hα line ratio in z ~ 2 galaxies (Förster Schreiber et al. 2018). Therefore, although azimuthally averaged gradients are still useful and commonly adopted as benchmark for theoretical predictions, this suggests that only the full characterization of the spatial distribution of the heavy elements within a galaxy (i.e. the 2D metallicity maps) can ultimately provide the key to resolve the interplay between physical processes occurring on local scales (gas accretion, gas flows, turbulent mixing, etc...) and global, secular processes, constraining their role in shaping the formation of galactic disks at these epochs.

5 SUMMARY AND CONCLUSIONS

We have analysed a sample of gravitationally lensed galaxies between redshift $1.2 < z < 2.5$, observed with KMOS and SINFONI as a first part of the KLEVER Large Programme. The observations conducted in the YJ, H and K bands provided spatially resolved mapping of many rest-frame optical nebular emission lines, allowing us to assess different ISM properties in these objects on both global and local scales. In particular, in this work we have constrained gas-phase metallicities by simultaneously adopting different strong-line diagnostics and calibrations tied to the T_e abundance scale. The sample covers a range of $10^{8.5} - 10^{10.5}$ $M_\odot$ in stellar mass and $\sim 1 - 200$ $M_\odot yr^{-1}$ in SFR, sampling galaxies on, above and below the star-forming main sequence at these redshifts, with a slight bias towards high specific star formation rates especially at z > 2. On global scales (i.e. from the analysis of their integrated spectra) the galaxies analysed are consistent with the observed evolution in the BPT diagrams and in the mass-metallicity relation reported in previous studies (Fig. 5 and 6).

In this paper, we mainly focused on the analysis of metallicity gradients. We have exploited robust lens modelling of foreground sources (mainly galaxy clusters) to obtain source plane emission line maps with a typical resolution of ~ 3 kpc, from which we have derived metallicity maps and extracted radial metallicity gradients by means of azimuthal averages within elliptical apertures defined according to the shape of the source plane PSF (e.g. Fig. 7). The main conclusions drawn in this paper are summarized as follows.

- The 86% of our galaxy sample present radial metallicity gradients shallower than 0.05dex/kpc (72% within 0.025dex/kpc), with 89% of the gradients consistent with being flat within 3σ uncertainty (67% within 1σ). This is in agreement with the majority of previous results reported in the literature which investigated metallicity gradients at z ~ 1 – 2.5, despite the diversity of observational techniques and diagnostics adopted (Fig. 8). Comparison with cosmological simulations that explore the effect of different feedback modes suggest a scenario where efficient mixing processes, that redistribute a significant amount of gas over large scales, are in place at these epochs (Gibson et al. 2013b). However, predictions from analytical models which assumes fairly constant star formation profiles and no prescriptions about radial gas mixing are also broadly consistent with the observed distribution of metallicity gradients observed in the galaxy population at $1 \lesssim z \lesssim 2.5$ (Mollá et al. 2019). Galaxies showing a relatively homogeneous metal distribution across large spatial scales might also be consistent with a scenario of uniform disk mass assembly.
- Three sources in the sample are characterised by significantly (i.e. above 3σ) positive ("inverted") metallicity gradients. Similarly with previous findings in the literature (Cresci et al. 2010; Troncoso et al. 2014), in these galaxies the central metal poor regions are associated with the highest level of Hα emission, hence with the highest star formation rate surface density, suggesting recent accretion of pristine gas. Alternatively, chemical evolution models assuming a constant SFE along the disc and allowing for large variations in the velocity of radial gas flows have also shown to be able to predict an "inversion" of the metallicity gradients (Mott et al. 2013).
- We explored the correlation between the sSFR and metallicity gradients at z > 0.8, extending the relation towards the highest sSFRs probed by our sample (Fig. 9, upper panel). We find a mild trend in which galaxies with a higher sSFR exhibit flatter or 'more positive' gradients, with a significance of 3.2σ, in agreement with previous findings (Stott et al. 2014; Wuyts et al. 2016). When normalising the SFR to that expected from the star forming main sequence at each redshift, the significance of the trend reduces to 2σ (Fig. 10). We have found a milder, negative correlation between metallicity gradients and stellar mass (significant at the 2.7σ level, lower panel of Fig. 9), consistent though with the trend followed in the local Universe (i.e. gradients steepening with increasing stellar mass, Belfiore et al. 2017).
- Despite the apparent radial invariance, some of our galaxies exhibit complex patterns in the metallicity maps, with variations in log(O/H) of the order of 0.1 – 0.2 dex on kpc scales which are, however, not symmetrically distributed around the galaxy centre (Fig. 11). In these cases, flat gradients might artificially arise from azimuthal averages, washing out and therefore hiding the presence of regions with different level of chemical enrichment. Strong fluctuations in the star formation history and outflow rates or local episodes of gas accretion occurring on short timescales could explain the observed irregular morphologies. This warns against using radial gradients as the only constrains for galaxy evolution models. Extracting more information from the full 2D metallicity maps is key to discriminate between the contribution that different local processes (like gas recycling via stellar feedback, pristine gas accretion, turbulent mixing and merger events) have in shaping the observed morphologies.

In conclusion, mutual interactions between star formation and different processes involving efficient radial mixing driven by intense galactic feedback or merger events could strongly affect the complex morphology of the maps and the average radial flattening of the gradients that we see at these epochs. A more comprehensive and self-consistent view of spatially resolved metallicity properties in high-z galaxies will be possible as soon as the KLEVER Programme is completed, comprising a larger sample spanning a wider range in stellar masses and star formation rates. The complexity revealed by current generation IFU instruments is also paving the way for observations with next generation facilities like NIRSPEC on JWST and ERIS (the AO-assisted Enhanced

Resolution Imager and Spectrograph, aimed at substituting SINFONI on the VLT from 2020, Davies et al. 2018), that will offer in the near future the opportunity to probe the metal distribution in high redshfit galaxies with unprecedented detail, helping to discriminate between different theoretical predictions and setting new and powerful constraint for cosmological simulations and chemical evolution models.

ACKNOWLEDGEMENTS

Based on observations made with ESO Telescopes at the La Silla Paranal Observatory under programme ID 083.A-0918(A), 083.B-0108(A), 090.B-0313(A), 092.A-0426(A), 092.B-0677(A), 094.A-0279(A) , 095.B-0480(A) and 197.A-0717(A). The mass models for the lensed galaxies in galaxy clusters were developed by Zitrin et al. (2009, 2013), and obtained through the Hubble Space Telescope Archive, as a high-end science product of the CLASH (Postman et al. 2012) and Frontier Fields (Lotz et al. 2017) programs.

MC and RM acknowledge support by the Science and Technology Facilities Council (STFC) and from ERC Advanced Grant 695671 "QUENCH". GC acknowledges the support by INAF/Frontiera through the "Progetti Premiali" funding scheme of the Italian Ministry of Education, University, and Research; GC has been also supported by the INAF PRIN-SKA 2017 program 1.05.01.88.04. MC acknowledges support from a Royal Society University Research Fellowship. CC acknowledges funding from the European UnionâĂŹs Horizon 2020 research and innovation programme under the Marie Sklodowska-Curie grant agreement no. 664931. Y. P. acknowledges support from the National Key Program for Science and Technology Research and Development under grant number 2016YFA0400702, and the NSFC grant no. 11773001, 11721303.

This work utilizes gravitational lensing models produced by PIs Bradac, Natarajan & Kneib (CATS), Merten & Zitrin, Sharon, Williams, Keeton, Bernstein and Diego, and the GLAFIC group. This lens modeling was partially funded by the HST Frontier Fields program conducted by STScI. STScI is operated by the Association of Universities for Research in Astronomy, Inc. under NASA contract NAS 5-26555. The lens models were obtained from the Mikulski Archive for Space Telescopes (MAST).

References

Andrews B. H., Martini P., 2013, ApJ, 765, 140
Baldwin J. A., Phillips M. M., Terlevich R., 1981, PASP, 93, 5
Belfiore F., et al., 2017, MNRAS, 469, 151
Belfiore F., Vincenzo F., Maiolino R., Matteucci F., 2019, MNRAS, 487, 456
Belli S., Jones T., Ellis R. S., Richard J., 2013, ApJ, 772, 141
Belokurov V., Evans N. W., Hewett P. C., Moiseev A., McMahon R. G., Sanchez S. F., King L. J., 2009, MNRAS, 392, 104
Berg D. A., et al., 2012, ApJ, 754, 98
Berg D. A., Skillman E. D., Croxall K. V., Pogge R. W., Moustakas J., Johnson-Groh M., 2015, ApJ, 806, 16
Bournaud F., Elmegreen B. G., Elmegreen D. M., 2007, ApJ, 670, 237
Bresolin F., 2011, ApJ, 730, 129
Bresolin F., Kennicutt R. C., Ryan-Weber E., 2012, ApJ, 750, 122
Bresolin F., Kudritzki R.-P., Urbaneja M. A., Gieren W., Ho I.-T., Pietrzyński G., 2016, ApJ, 830, 64
Bruzual G., Charlot S., 2003, MNRAS, 344, 1000
Calzetti D., Armus L., Bohlin R. C., Kinney A. L., Koornneef J., Storchi-Bergmann T., 2000, ApJ, 533, 682
Cardelli J. A., Clayton G. C., Mathis J. S., 1989, ApJ, 345, 245
Carton D., et al., 2018, MNRAS, 478, 4293
Ceverino D., Sánchez Almeida J., Muñoz Tuñón C., Dekel A., Elmegreen B. G., Elmegreen D. M., Primack J., 2016, MNRAS, 457, 2605
Chabrier G., 2003, PASP, 115, 763
Charlot S., Fall S. M., 2000, ApJ, 539, 718
Conroy C., 2013, ARA&A, 51, 393
Cresci G., Mannucci F., Maiolino R., Marconi A., Gnerucci A., Magrini L., 2010, Nature, 467, 811
Cresci G., Mannucci F., Curti M., 2018, arXiv e-prints,
Cullen F., Cirasuolo M., McLure R. J., Dunlop J. S., Bowler R. A. A., 2014, MNRAS, 440, 2300
Curti M., Cresci G., Mannucci F., Marconi A., Maiolino R., Esposito S., 2017, MNRAS, 465, 1384
Curti M., Mannucci F., Cresci G., Maiolino R., 2019, arXiv e-prints,
Davé R., Finlator K., Oppenheimer B. D., 2011, MNRAS, 416, 1354
Davé R., Rafieferantsoa M. H., Thompson R. J., Hopkins P. F., 2017, MNRAS, 467, 115
Davies R. I., 2007, MNRAS, 375, 1099
Davies R., et al., 2018, in Ground-based and Airborne Instrumentation for Astronomy VII. p. 1070209 (arXiv:1807.05089), doi:10.1117/12.2311480
Dayal P., Ferrara A., Dunlop J. S., 2013, MNRAS, 430, 2891
Dekel A., et al., 2009, Nature, 457, 451
Dopita M. A., Kewley L. J., Sutherland R. S., Nicholls D. C., 2016, Ap&SS, 361, 61
Ellison S. L., Patton D. R., Simard L., McConnachie A. W., 2008, ApJ, 672, L107
Elmegreen B. G., 2009, in Jogee S., Marinova I., Hao L., Blanc G. A., eds, Astronomical Society of the Pacific Conference Series Vol. 419, Galaxy Evolution: Emerging Insights and Future Challenges. p. 23 (arXiv:0903.1937)
Elmegreen B. G., Bournaud F., Elmegreen D. M., 2008, ApJ, 688, 67
Erb D. K., Shapley A. E., Pettini M., Steidel C. C., Reddy N. A., Adelberger K. L., 2006, ApJ, 644, 813
Förster Schreiber N. M., et al., 2011, ApJ, 739, 45
Förster Schreiber N. M., et al., 2018, ApJS, 238, 21
Fraternali F., Binney J. J., 2008, MNRAS, 386, 935
Gazak J. Z., et al., 2015, ApJ, 805, 182
Genzel R., et al., 2008, ApJ, 687, 59
Genzel R., et al., 2011, ApJ, 733, 101
Gibson B. K., Courty S., Cunnama D., Mollá M., 2013a, Asociacion Argentina de Astronomia La Plata Argentina Book Series, 4, 57
Gibson B. K., Pilkington K., Brook C. B., Stinson G. S., Bailin J., 2013b, A&A, 554, A47
Guo Y., et al., 2016, ApJ, 822, 103
Henry R. B. C., Kwitter K. B., Jaskot A. E., Balick B., Morrison M. A., Milingo J. B., 2010, ApJ, 724, 748
Ho I.-T., Kudritzki R.-P., Kewley L. J., Zahid H. J., Dopita M. A., Bresolin F., Rupke D. S. N., 2015, MNRAS, 448, 2030
Hunt L., Dayal P., Magrini L., Ferrara A., 2016, MNRAS, 463, 2020
Jones T. A., Swinbank A. M., Ellis R. S., Richard J., Stark D. P., 2010a, MNRAS, 404, 1247
Jones T., Ellis R., Jullo E., Richard J., 2010b, ApJ, 725, L176
Jones T., Ellis R. S., Richard J., Jullo E., 2013, ApJ, 765, 48
Jones T., et al., 2015a, AJ, 149, 107
Jones T., Martin C., Cooper M. C., 2015b, ApJ, 813, 126

Kashino D., et al., 2017, ApJ, 835, 88
Kashino D., et al., 2019, ApJS, 241, 10
Kauffmann G., et al., 2003, MNRAS, 346, 1055
Kennicutt R. C., Evans N. J., 2012, ARA&A, 50, 531
Kewley L. J., Dopita M. A., Sutherland R. S., Heisler C. A., Trevena J., 2001, ApJ, 556, 121
Kewley L. J., Rupke D., Zahid H. J., Geller M. J., Barton E. J., 2010, ApJ, 721, L48
Kewley L. J., Maier C., Yabe K., Ohta K., Akiyama M., Dopita M. A., Yuan T., 2013, ApJ, 774, L10
Kostrzewa-Rutkowska Z., Wyrzykowski Ł., Auger M. W., Collett T. E., Belokurov V., 2014, MNRAS, 441, 3238
Kroupa P., Tout C. A., Gilmore G., 1993, MNRAS, 262, 545
Kudritzki R.-P., Ho I.-T., Schruba A., Burkert A., Zahid H. J., Bresolin F., Dima G. I., 2015, MNRAS, 450, 342
Leethochawalit N., Jones T. A., Ellis R. S., Stark D. P., Richard J., Zitrin A., Auger M., 2016, ApJ, 820, 84
Li H., et al., 2018, MNRAS, 476, 1765
Lilly S. J., Carollo C. M., Pipino A., Renzini A., Peng Y., 2013, ApJ, 772, 119
Lotz J. M., et al., 2017, ApJ, 837, 97
Ma X., Hopkins P. F., Faucher-Giguère C.-A., Zolman N., Muratov A. L., Kereš D., Quataert E., 2016, MNRAS, 456, 2140
Ma X., Hopkins P. F., Feldmann R., Torrey P., Faucher-Giguère C.-A., Kereš D., 2017, MNRAS, 466, 4780
Maciel W. J., Costa R. D. D., Uchida M. M. M., 2003, A&A, 397, 667
Magrini L., Stanghellini L., Villaver E., 2009, ApJ, 696, 729
Magrini L., Stanghellini L., Corbelli E., Galli D., Villaver E., 2010, A&A, 512, A63
Maiolino R., Mannucci F., 2019, A&ARv, 27, 3
Maiolino R., et al., 2008, A&A, 488, 463
Mannucci F., et al., 2009, MNRAS, 398, 1915
Mannucci F., Cresci G., Maiolino R., Marconi A., Gnerucci A., 2010, MNRAS, 408, 2115
Masters D., et al., 2014, ApJ, 785, 153
Masters D., Faisst A., Capak P., 2016, ApJ, 828, 18
Meneghetti M., et al., 2017, MNRAS, 472, 3177
Mollá M., Díaz A. I., 2005, MNRAS, 358, 521
Mollá M., Díaz Á. I., Ascasibar Y., Gibson B. K., 2017, MNRAS, 468, 305
Mollá M., Díaz Á. I., Cavichia O., Gibson B. K., Maciel W. J., Costa R. D. D., Ascasibar Y., Few C. G., 2018, preprint, (arXiv:1810.09182)
Mollá M., Díaz Á. I., Cavichia O., Gibson B. K., Maciel W. J., Costa R. D. D., Ascasibar Y., Few C. G., 2019, MNRAS, 482, 3071
Mott A., Spitoni E., Matteucci F., 2013, MNRAS, 435, 2918
Nakajima K., Ouchi M., Shimasaku K., Hashimoto T., Ono Y., Lee J. C., 2013, ApJ, 769, 3
Patrício V., Christensen L., Rhodin H., Cañameras R., Lara-López M. A., 2018, MNRAS, 481, 3520
Pilkington K., et al., 2012, MNRAS, 425, 969
Planck Collaboration et al., 2016, A&A, 594, A13
Portinari L., Chiosi C., 1999, A&A, 350, 827
Postman M., et al., 2012, ApJS, 199, 25
Prantzos N., Boissier S., 2000, MNRAS, 313, 338
Queyrel J., et al., 2012, A&A, 539, A93
Richard J., Jones T., Ellis R., Stark D. P., Livermore R., Swinbank M., 2011, MNRAS, 413, 643
Rosati P., et al., 2014, The Messenger, 158, 48
Rousselot P., Lidman C., Cuby J.-G., Moreels G., Monnet G., 2000, A&A, 354, 1134
Rupke D., Kewley L., Barnes J., 2010a, in Smith B., Higdon J., Higdon S., Bastian N., eds, Astronomical Society of the Pacific Conference Series Vol. 423, Galaxy Wars: Stellar Populations and Star Formation in Interacting Galaxies. p. 355
Rupke D. S. N., Kewley L. J., Chien L.-H., 2010b, ApJ, 723, 1255
Salim S., Lee J. C., Ly C., Brinchmann J., Davé R., Dickinson M., Salzer J. J., Charlot S., 2014, ApJ, 797, 126
Samland M., Hensler G., Theis C., 1997, ApJ, 476, 544
Sánchez Almeida J., Caon N., Muñoz-Tuñón C., Filho M., Cerviño M., 2018, MNRAS, 476, 4765
Sánchez S. F., et al., 2014, A&A, 563, A49
Sanders R. L., et al., 2015, ApJ, 799, 138
Sanders R. L., et al., 2016, ApJ, 825, L23
Sanders R. L., et al., 2018, ApJ, 858, 99
Shapley A. E., et al., 2015, ApJ, 801, 88
Sharples R., et al., 2013, The Messenger, 151, 21
Shivaei I., et al., 2015, ApJ, 815, 98
Simons R. C., et al., 2019, ApJ, 874, 59
Somerville R. S., Davé R., 2015, ARA&A, 53, 51
Speagle J. S., Steinhardt C. L., Capak P. L., Silverman J. D., 2014, ApJS, 214, 15
Spitoni E., Matteucci F., Marcon-Uchida M. M., 2013, A&A, 551, A123
Stanghellini L., Haywood M., 2010, ApJ, 714, 1096
Stanghellini L., Haywood M., 2018, preprint, (arXiv:1806.02276)
Stanghellini L., Magrini L., Villaver E., Galli D., 2010, A&A, 521, A3
Stanghellini L., Magrini L., Casasola V., Villaver E., 2014, A&A, 567, A88
Steidel C. C., et al., 2014, ApJ, 795, 165
Steidel C. C., Strom A. L., Pettini M., Rudie G. C., Reddy N. A., Trainor R. F., 2016, ApJ, 826, 159
Stinson G. S., Bailin J., Couchman H., Wadsley J., Shen S., Nickerson S., Brook C., Quinn T., 2010, MNRAS, 408, 812
Stott J. P., et al., 2014, MNRAS, 443, 2695
Strom A. L., Steidel C. C., Rudie G. C., Trainor R. F., Pettini M., Reddy N. A., 2017, ApJ, 836, 164
Swinbank A. M., Sobral D., Smail I., Geach J. E., Best P. N., McCarthy I. G., Crain R. A., Theuns T., 2012, MNRAS, 426, 935
Tamburello V., Rahmati A., Mayer L., Cava A., Dessauges-Zavadsky M., Schaerer D., 2017, MNRAS, 468, 4792
Tremonti C. A., et al., 2004, ApJ, 613, 898
Treu T., et al., 2015, ApJ, 812, 114
Troncoso P., et al., 2014, A&A, 563, A58
Vegetti S., Koopmans L. V. E., 2009, MNRAS, 392, 945
Wang X., et al., 2017, ApJ, 837, 89
Wang X., et al., 2019, ApJ, 882, 94
Werk J. K., Putman M. E., Meurer G. R., Santiago-Figueroa N., 2011, ApJ, 735, 71
Whitaker K. E., et al., 2014, ApJ, 795, 104
Wiersma R. P. C., Schaye J., Theuns T., Dalla Vecchia C., Tornatore L., 2009, MNRAS, 399, 574
Wuyts E., et al., 2014, ApJ, 789, L40
Wuyts E., et al., 2016, ApJ, 827, 74
Yates R. M., Kauffmann G., Guo Q., 2012, MNRAS, 422, 215
Yuan T.-T., Kewley L. J., Swinbank A. M., Richard J., Livermore R. C., 2011, ApJ, 732, L14
Yuan T.-T., Kewley L. J., Rich J., 2013, ApJ, 767, 106
Zahid H. J., Kewley L. J., Bresolin F., 2011, ApJ, 730, 137
Zaritsky D., Kennicutt Jr. R. C., Huchra J. P., 1994, ApJ, 420, 87
Zitrin A., Broadhurst T., Rephaeli Y., Sadeh S., 2009, ApJ, 707, L102
Zitrin A., et al., 2013, ApJ, 762, L30
Zitrin A., et al., 2015, ApJ, 801, 44
da Cunha E., et al., 2015, ApJ, 806, 110
van Dokkum P. G., 2001, PASP, 113, 1420

APPENDIX A: SYSTEMATICS IN THE CHOICE OF THE AVERAGING APERTURES

We discuss here the potential biases associated to the radial binning scheme. As already discussed in Sect. 4.2, radial averaging within elliptical apertures is required to account for the smearing effect introduced by the distorted source plane PSF. However, this choice is not without its issues. In fact, the information from different radii is combined within the same aperture when considering the directions perpendicular and parallel to the PSF major axis, and this effect can be particularly severe in those cases where the source plane PSF is extremely stretched (as in the case of strongly lensed sources), possibly introducing systematics in the determination of the gradients. Therefore, to test the impact of this issue we compute the radial gradients adopting purely circular apertures as well: overall, we find that the choice of the radial binning scheme does not affect the main results and the interpretation presented in this paper, as statistically significant differences are limited to few individual cases. For the 83% of the sample, the two estimates are consistent within the uncertainties, while for 5 sources they diverge more than 1σ; however, only in two of the these cases there is also an inversion in the sign of the gradient (from negative to positive). The results are reported in Table A1.

In Fig. A1 the metallicity gradients adopting elliptical apertures are plotted against the gradients computed within circular apertures, and the points are colour coded according to the axis ratio of the source plane PSF. Some of the points scatter around the 1:1 line and no clear systematic trends are highlighted. The most deviating points are, not surprisingly, associated with galaxies whose source plane PSF is extremely elongated. For instance, the two different configurations in the case of CSWA164 are compared in Fig. A1: for this source, the circularly-averaged gradient seems to be more representative of the metallicity distribution of individual spaxels across the map, although in both cases the slope remains consistent with zero within its uncertainty.

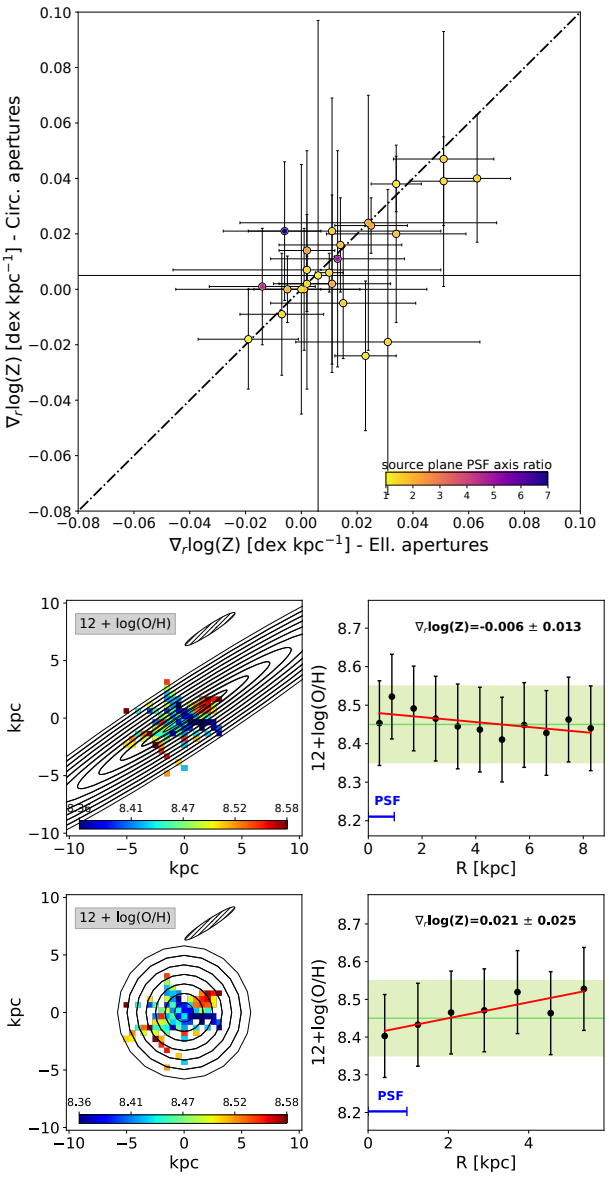

Figure A1. *Upper Panel:* Metallicity gradients computed by averaging within elliptical apertures in the source plane are plotted against those derived assuming purely circular apertures. The points are colour coded by the corresponding value of the axis ratio of the PSF in the source plane. The two estimates are consistent for more than 80% of the sources, with only two showing significant inversion of the slope. *Bottom Panels:* The difference between elliptically-averaged (top panel) and circularly-averaged (bottom panel) metallicity gradients is shown for one of the most extreme cases in our sample (i.e. the CSWA164 galaxy, $\mu \approx 13$) . Here, the significant elongation of the PSF causes the information from very different radii to be combined within the same annulus, potentially biasing the determination of the gradient. In this case, the circularly-averaged gradient seems to be more representative of the metallicity distribution of individual spaxels across the map.

APPENDIX B: RECONSTRUCTED EMISSION LINE RATIO AND METALLICITY MAPS

In the upper panels of Figures B1- B10 we show, for each galaxy in the *metallicity gradient sample* and from left to right respectively, the source plane HST F160w image, the normalized source plane Hα flux maps, the metallicity map and the metallicities extracted at increasing radii across each galaxy. In the last panel the linear fit to the metallicity gradient (in red) and the linear size of the PSF-HWHM (in blue) are also shown, together with the global metallicity value and its associated uncertainty (in green). The shaded black region in the *third* panel reproduce the shape and size of the PSF when mapped back into the source plane. The elliptical apertures used to derive the radial gradient overlay the 2D metallicity map.

In the bottom panels, we show instead some of the emission line ratios maps in the source plane. In particular, from left to right, we show [N II]/Hα, [O III]/Hβ, [S II]/Hα and [S III]/[S II]. For z> 2 galaxies, given the different wavelength range covered by our observations, we show [O III]/[O II] instead of [S III]/[S II].

Galaxy	Gradient (Circ. apertures) [dex kpc^{-1}]	Resolved	Diagnostics
SINFONI Galaxies			
Horseshoe	0.011±0.039	Yes	R_3,N_2
MACS Arc (North)	-0.021±0.033	Yes	R_3, N2, O_3O_2
MACS Arc (South)	0.002±0.038	Yes	R_3, N2, O_3O_2
CSWA164	0.021±0.025	Yes	R_3, N2, O_3O_2
KMOS Galaxies			
MS2137			
sp1	0.023±0.018	Yes	R_3,S_2
sp2	-0.009±0.022	Yes	R_3,O_3O_2
sp3	0.023±0.01	Yes	R_3,N_2
sp5	-0.0±0.022	Yes	R_3,O_3O_2
sp6	0.04±0.023	Yes	R_3,O_3O_2
sp7	0.251±0.008	Yes	R_3,N_2
sp13	0.282±0.02	Yes	R_3,N_2
sp15	0.014±0.013	Yes	R_3,S_2
ph6532	0.035±0.013	Yes	R_3,O_3O_2
ph3729	0.007±0.043	Yes	R_3,N_2
ph3912	0.039±0.016	Yes	R_3,S_2
ph8073	0.005±0.092	Marg	R_3,S_2
RXJ2248			
GLASS_00093-99-99	0.021±0.048	Yes	R_3,N_2
R2248_LRb_p1_M3_Q4_58__2	0.002±0.01	Yes	R_3,N_2
MUSE_SW_462-99-99	-0.0±0.012	Yes	R_3,N_2
GLASS_00333-99-99	-0.019±0.055	Yes	R_3,S_2
R2248_LRb_p3_M4_Q3_93__1	0.006±0.007	Yes	R_3,N_2
R2248_LRb_p3_M4_Q3_94__1	0.047±0.046	Yes	R_3,O_3O_2
GLASS_01845-99-99	0.016±0.017	Yes	R_3,O_3O_2
MUSE_SW_45-99-99	0.02±0.032	Yes	R_3,S_2
MUSE_SW_461-99-99	-0.018±0.018	Yes	R_3,N_2
MUSE_NE_111-99-99	-0.005±0.02	Marg	R_3,N_2
GLASS_00800-99-99	0.001±0.021	Yes	R_3,N_2
MUSE_NE_23-99-99	0.002±0.032	Yes	R_3,S_2

Table A1. Radial metallicity gradients computed adopting circular apertures in the source plane. For more details see the caption of Table 4.

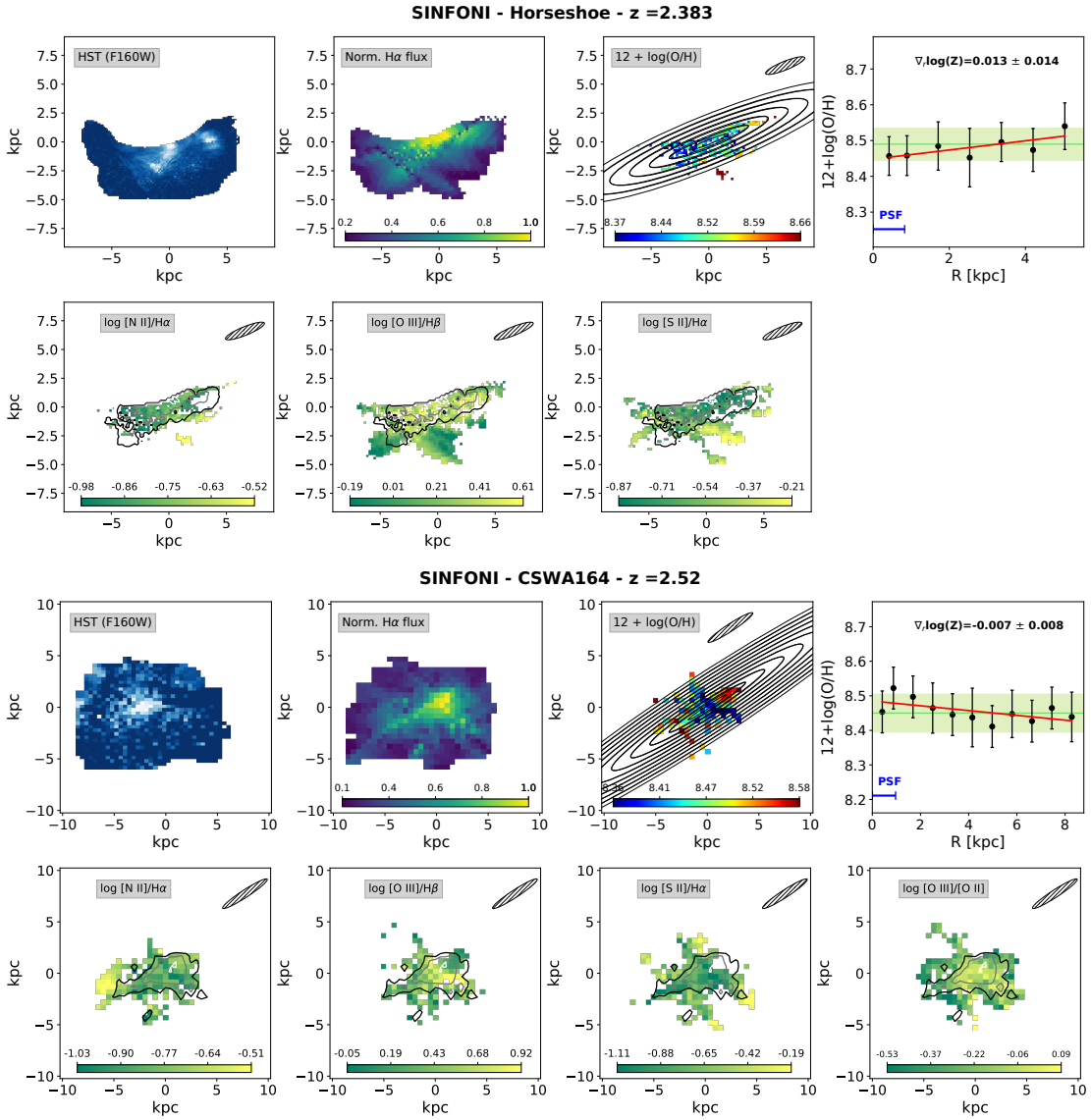

Figure B1. Source plane HST F160w image, normalised Hα flux map, metallicity map, line ratios maps and metallcity gradients for our *metallicity gradient sample*. For details see the text of Appendix B or the caption of Figure 7.

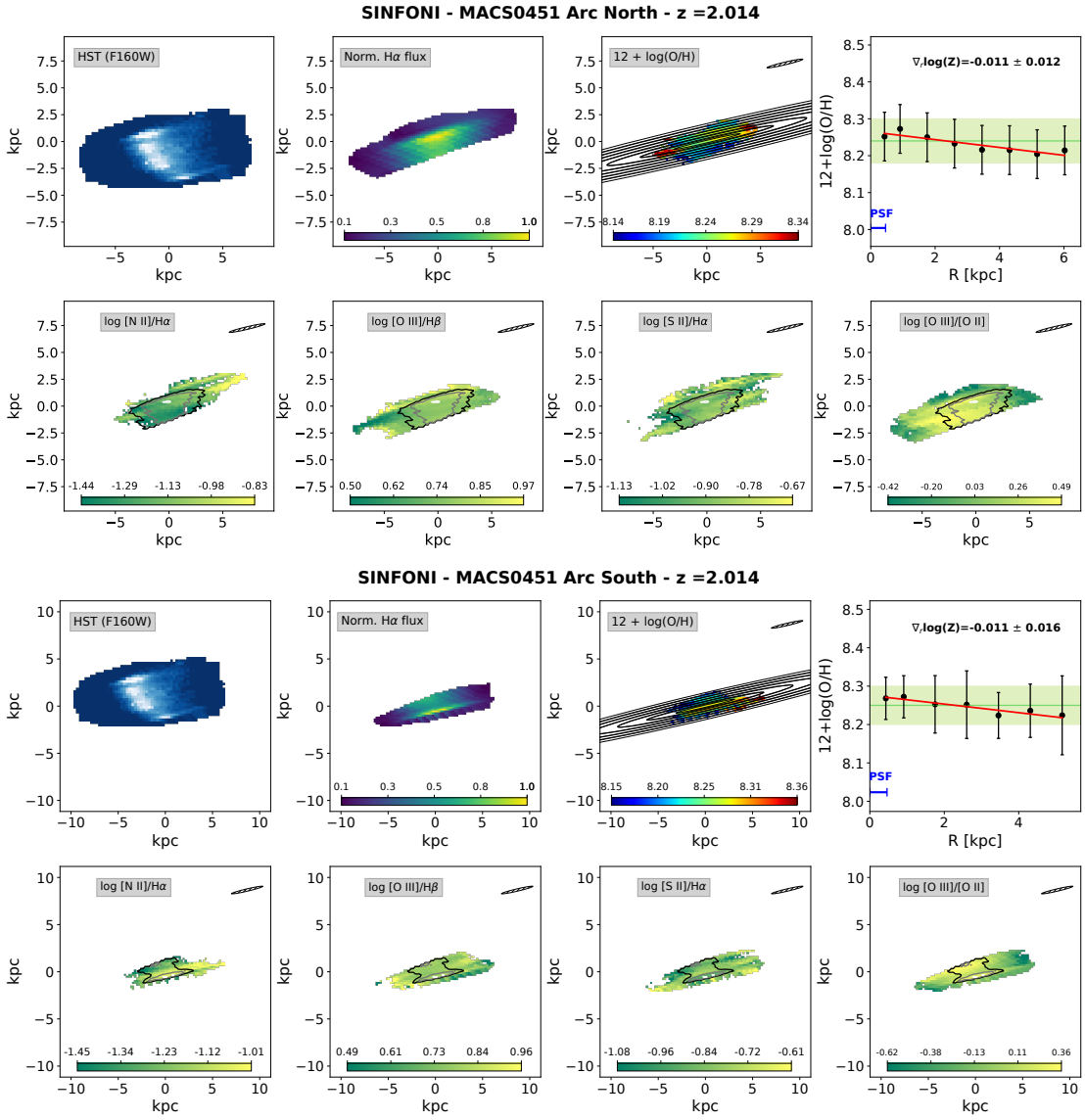

Figure B2. Same as Fig. B1.

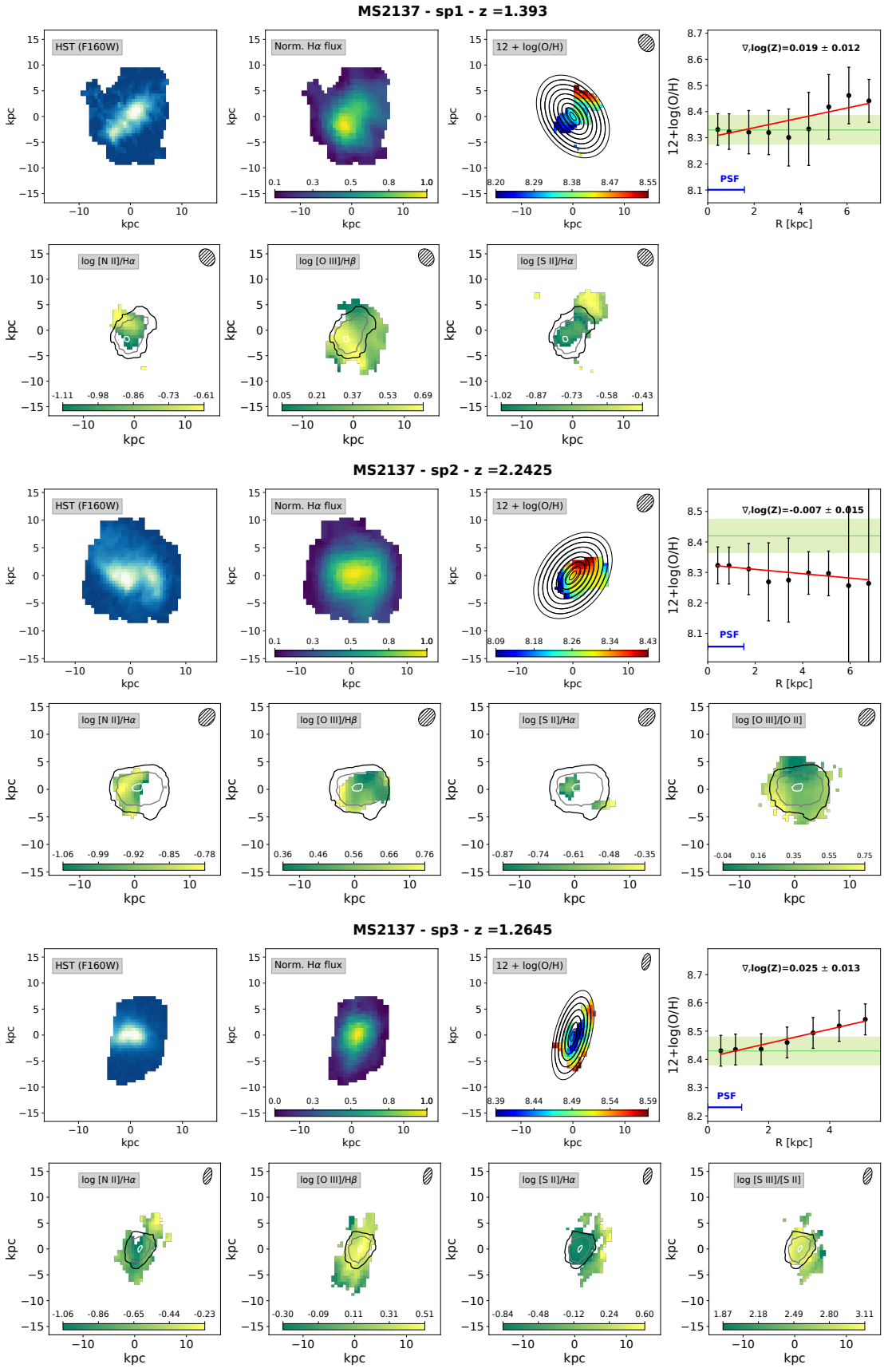

Figure B3. Same as Fig. B1.

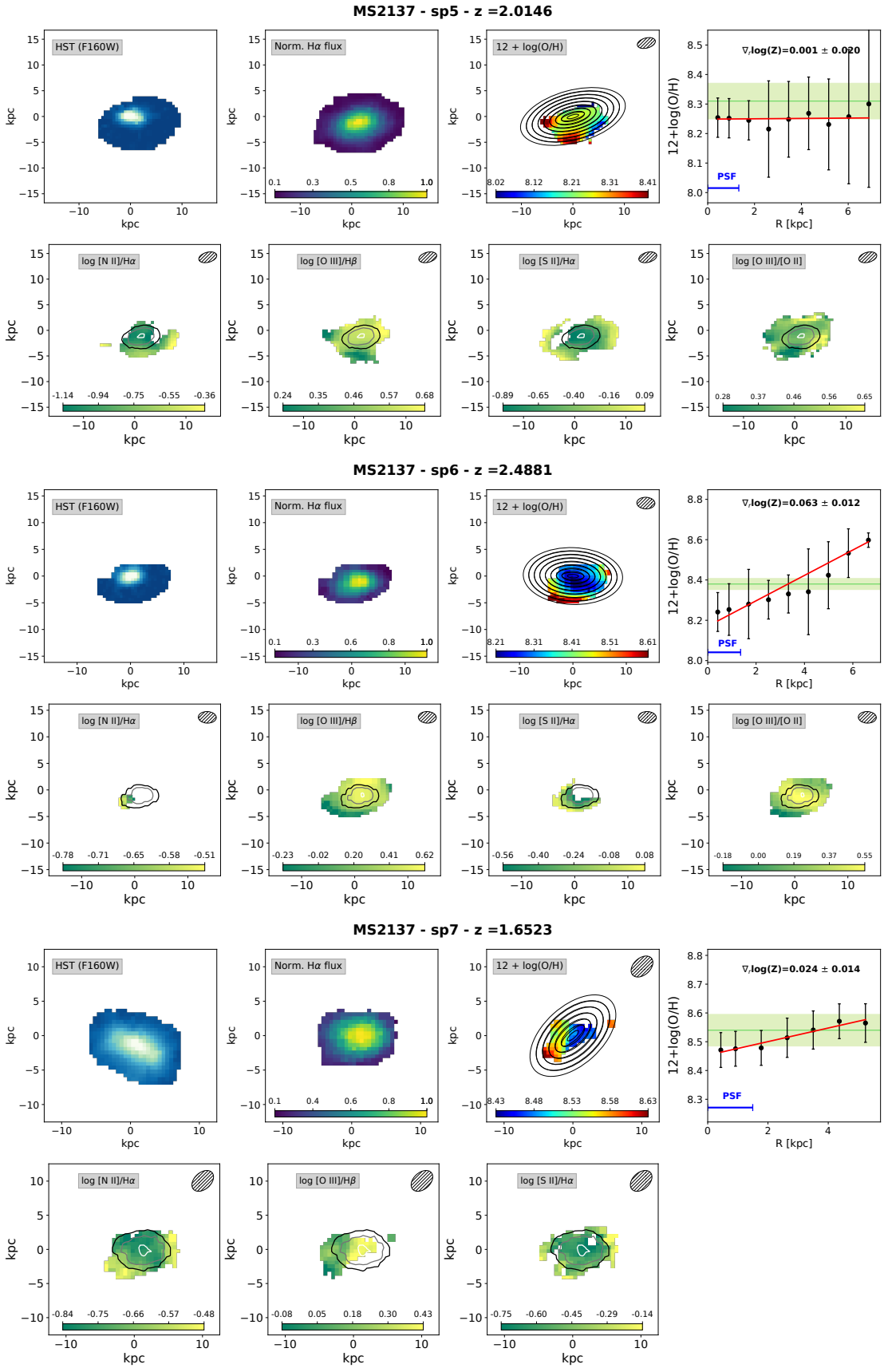

Figure B4. Same as Fig. B1.

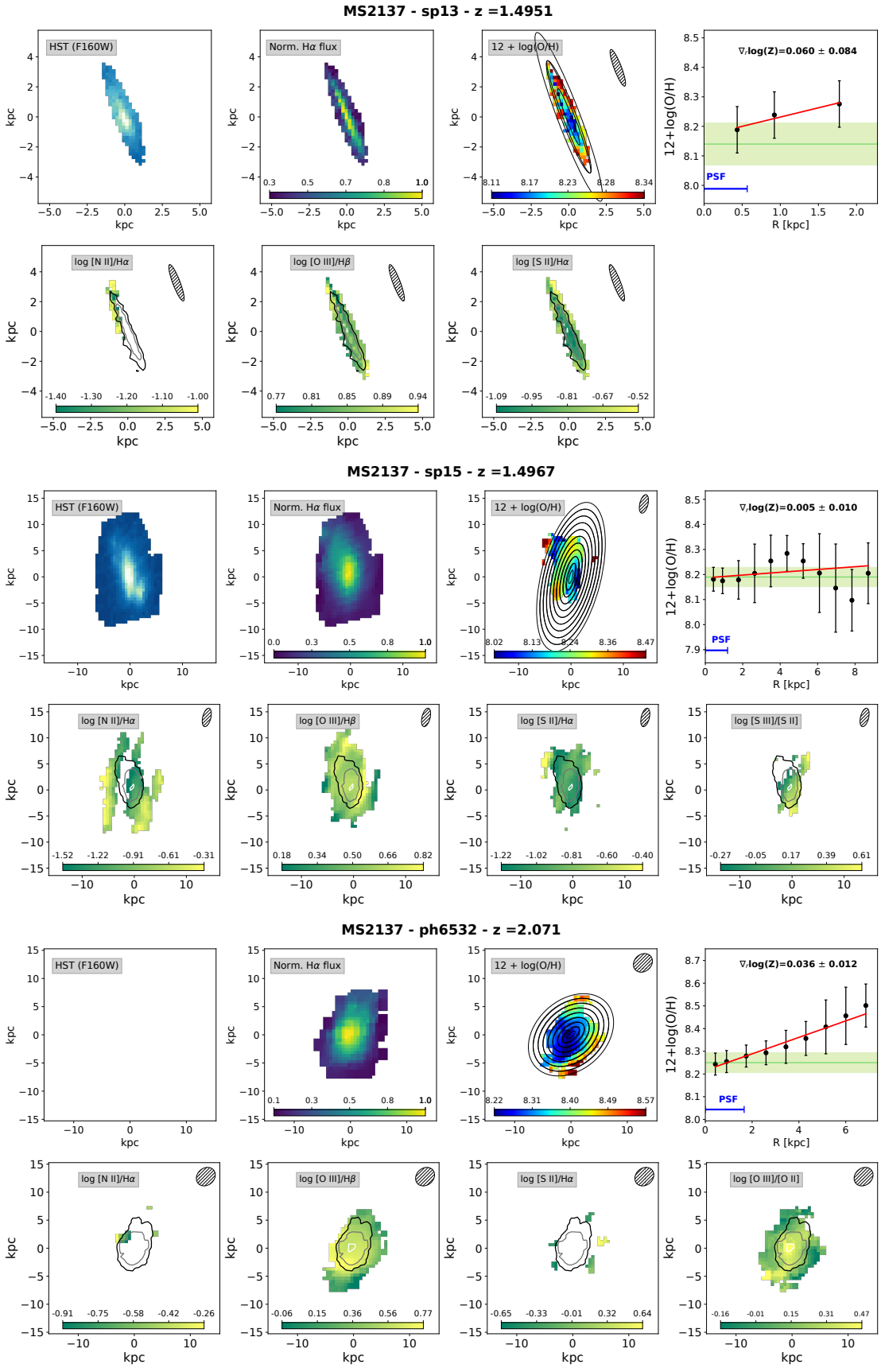

Figure B5. Same as Fig. B1.

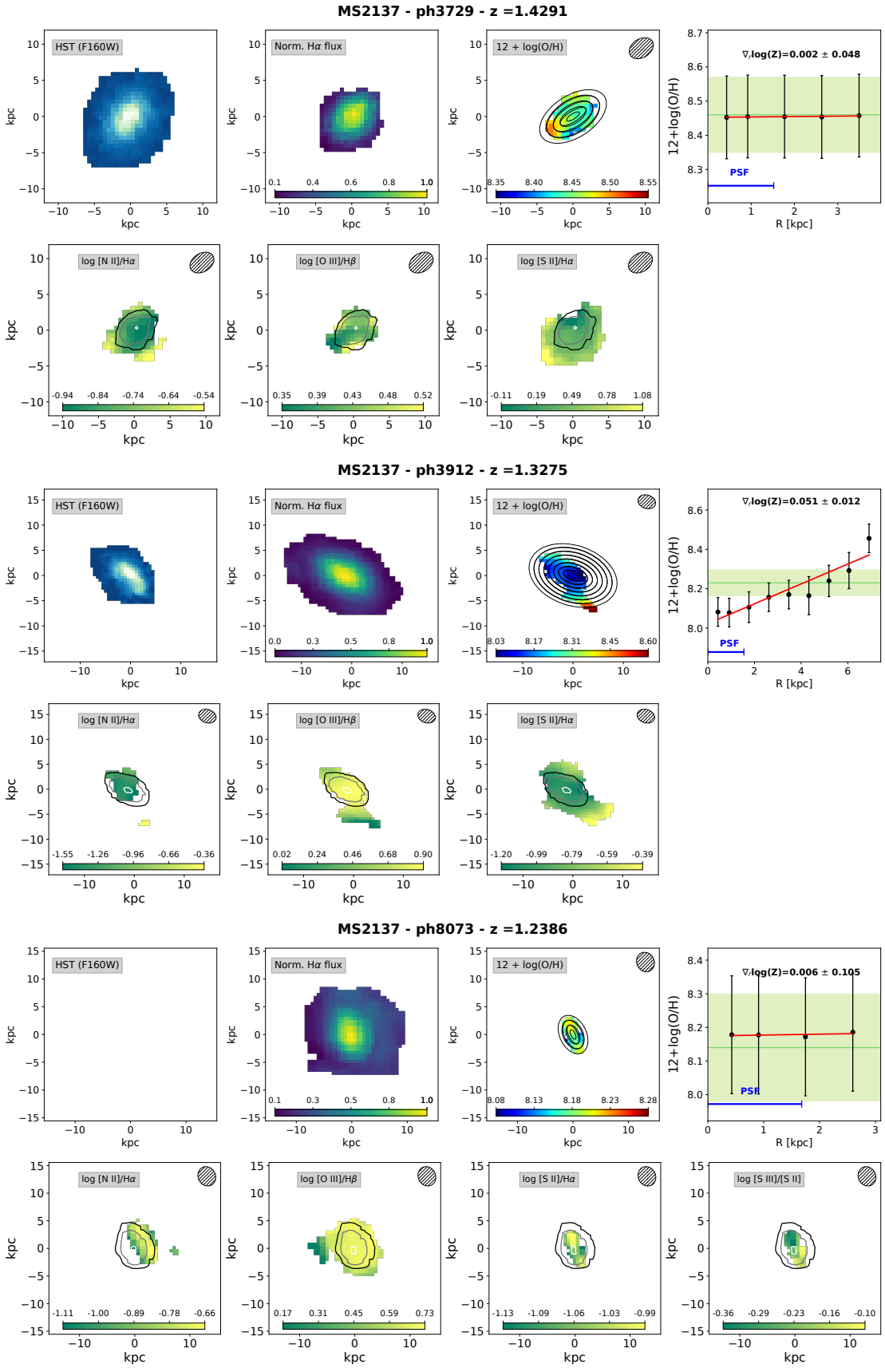

Figure B6. Same as Fig. B1.

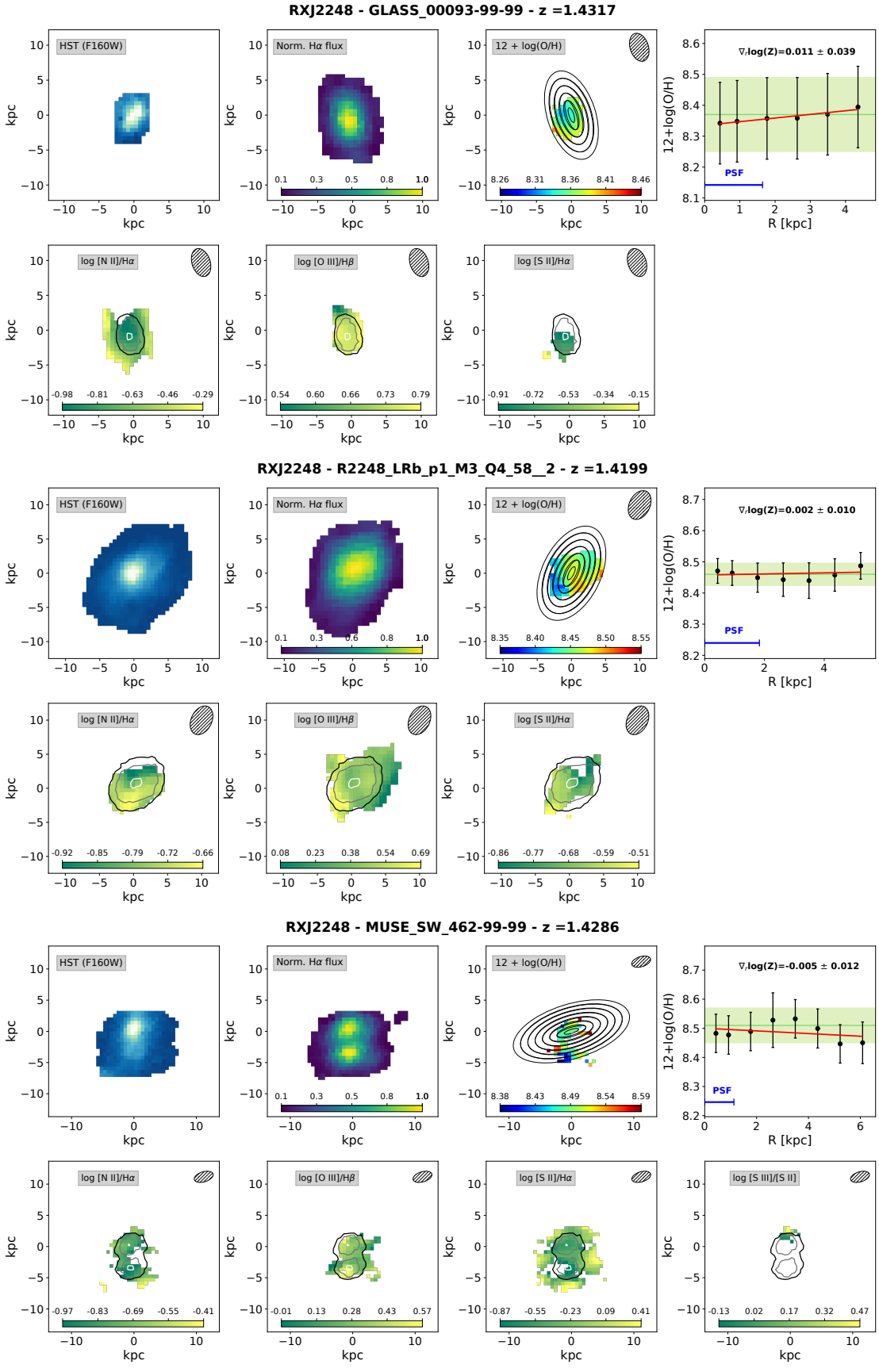

Figure B7. Same as Fig. B1.

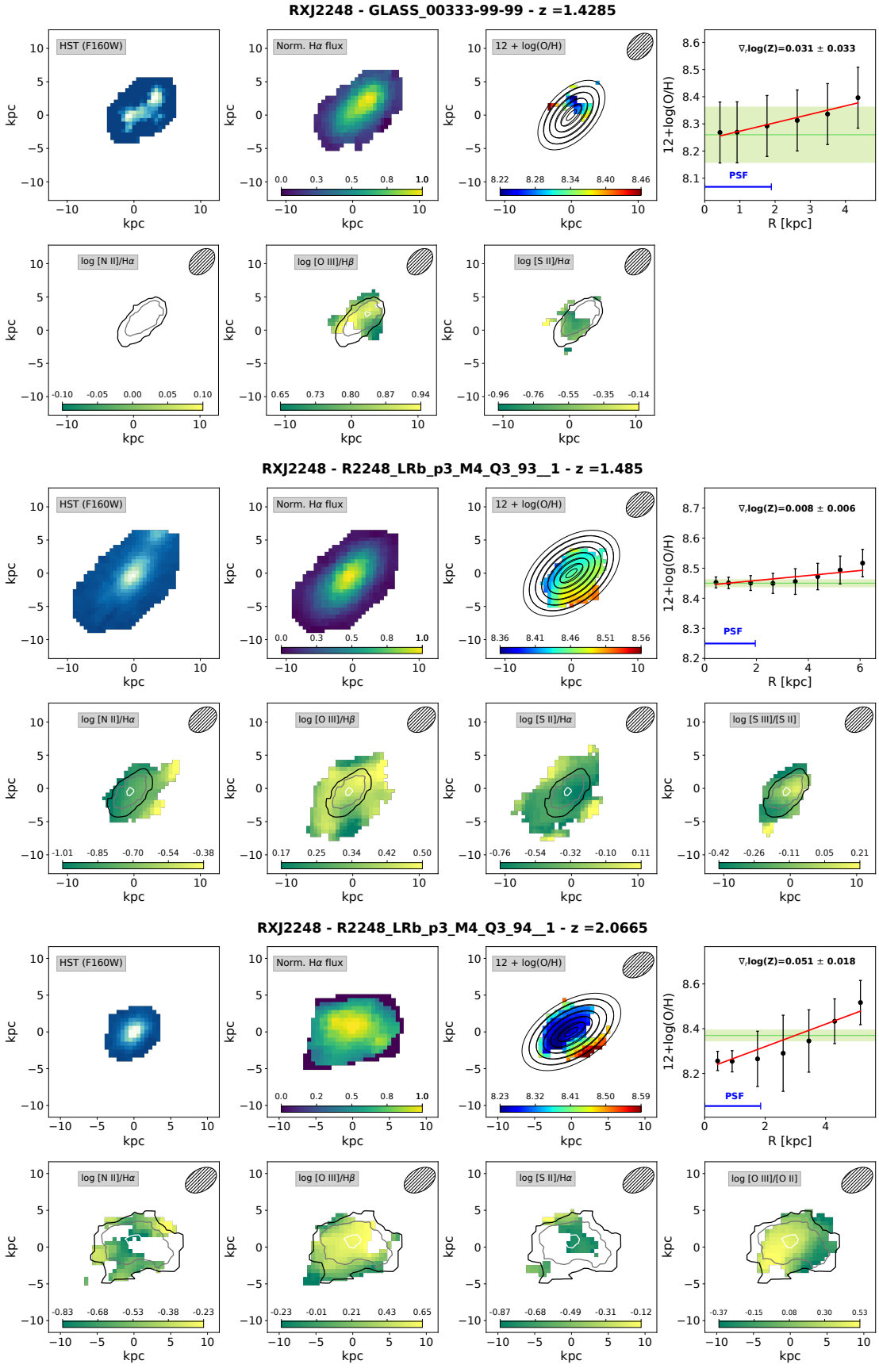

Figure B8. Same as Fig. B1.

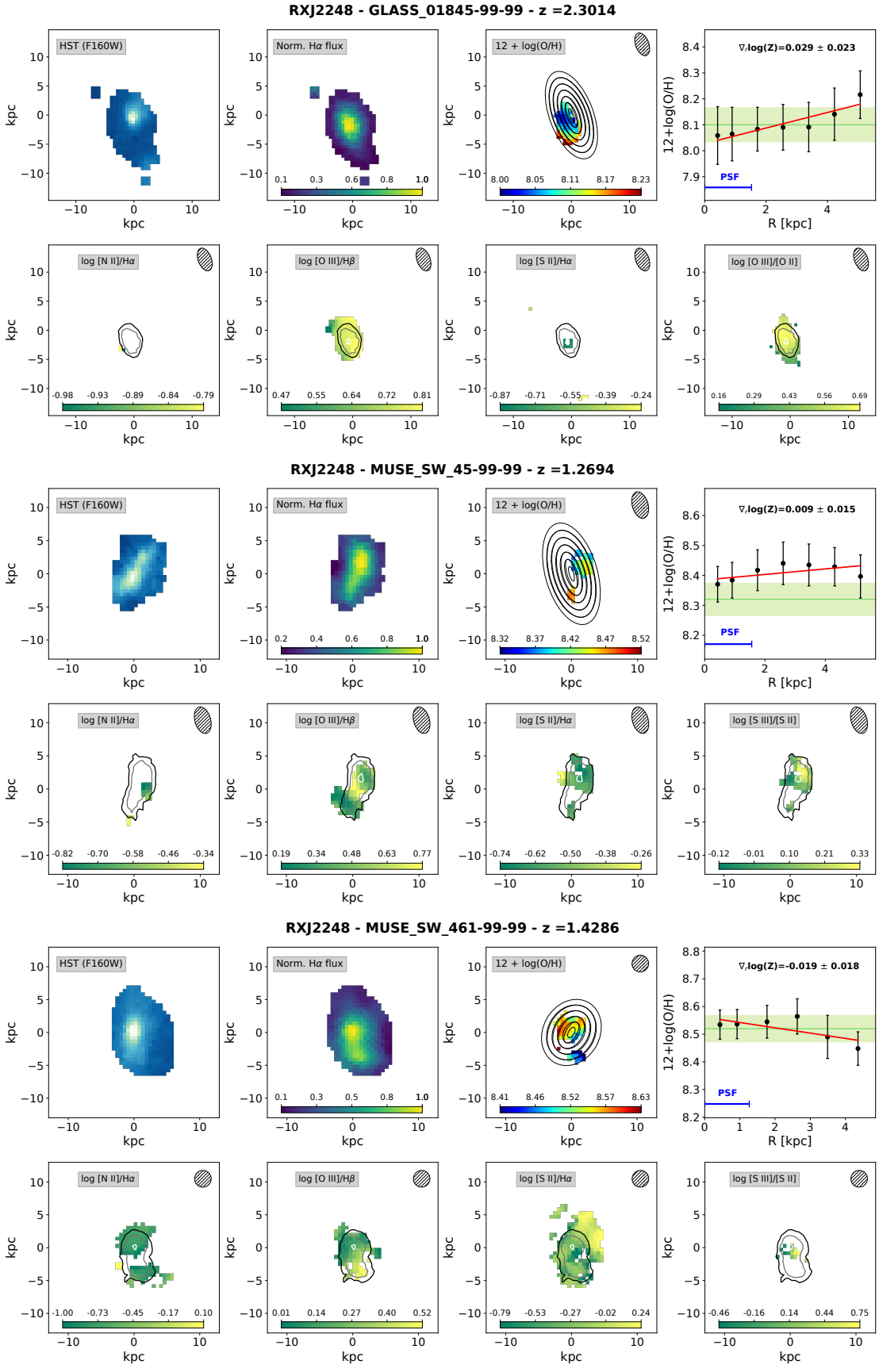

Figure B9. Same as Fig. B1.

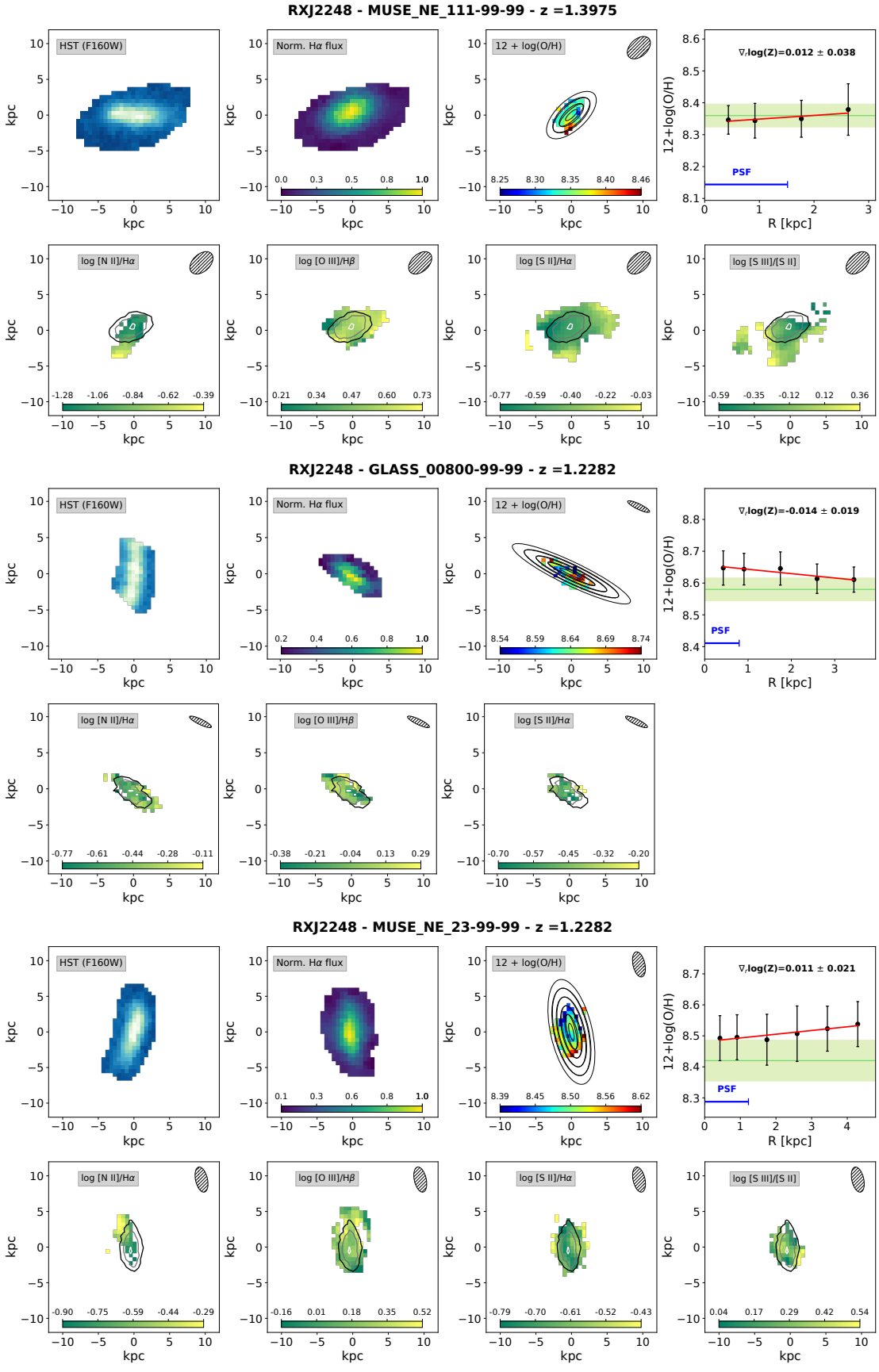

Figure B10. Same as Fig. B1.