The MUSE Hubble Ultra Deep Field surveys: Data release II

Roland Bacon, Jarle Brinchmann, Simon Conseil, Michael Maseda, Themiya Nanayakkara, Martin Wendt, Raphael Bacher, David Mary, Peter M. Weilbacher, Davor Krajnovic, et al.

A&A 670, A4 (2023)
https://doi.org/10.1051/0004-6361/202244187

Astronomy & Astrophysics

The MUSE *Hubble* Ultra Deep Field surveys: Data release II★,★★

Roland Bacon[1], Jarle Brinchmann[2], Simon Conseil[3], Michael Maseda[4], Themiya Nanayakkara[5], Martin Wendt[6,7], Raphael Bacher[1,8], David Mary[9], Peter M. Weilbacher[6], Davor Krajnović[6], Leindert Boogaard[16], Nicolas Bouché[1], Thierry Contini[14], Benoît Epinat[3,18], Anna Feltre[11], Yucheng Guo[1], Christian Herenz[15], Wolfram Kollatschny[17], Haruka Kusakabe[12], Floriane Leclercq[10], Léo Michel-Dansac[1], Roser Pello[3], Johan Richard[1], Martin Roth[6], Gregory Salvignol[1], Joop Schaye[13], Matthias Steinmetz[6], Laurence Tresse[3], Tanya Urrutia[6], Anne Verhamme[12], Eloise Vitte[12], Lutz Wisotzki[6], and Sebastiaan L. Zoutendijk[13]

(Affiliations can be found after the references)

Received 4 June 2022 / Accepted 17 October 2022

ABSTRACT

We present the second data release of the MUSE *Hubble* Ultra-Deep Field surveys, which includes the deepest spectroscopic survey ever performed. The MUSE data, with their 3D content, amazing depth, wide spectral range, and excellent spatial and medium spectral resolution, are rich in information. Their location in the *Hubble* ultra-deep field area, which benefits from an exquisite collection of ancillary panchromatic information, is a major asset. This update of the first release incorporates a new 141-h adaptive-optics-assisted MUSE eXtremely Deep Field (MXDF; 1 arcmin diameter field of view) in addition to the reprocessed 10-h mosaic (3×3 arcmin^2) and the single 31-h deep field (1×1 arcmin^2). All three data sets were processed and analyzed homogeneously using advanced data reduction and analysis methods. The 3σ point-source flux limit of an unresolved emission line reaches 3.1×10^{-19} and 6.3×10^{-20} erg s^{-1} cm^{-2} at 10- and 141-h depths, respectively. We have securely identified and measured the redshift of 2221 sources, an increase of 41% compared to the first release. With the exception of eight stars, the collected sample consists of 25 nearby galaxies ($z < 0.25$), 677 [O II] emitters ($z = 0.25$–1.5), 201 galaxies in the MUSE redshift desert range ($z = 1.5$–2.8), and 1308 Lyα emitters ($z = 2.8$–6.7). This represents an order of magnitude more redshifts than the collection of all spectroscopic redshifts obtained before MUSE in the *Hubble* ultra-deep field area (i.e., 2221 versus 292). At high redshift ($z > 3$), the difference is even more striking, with a factor of 65 increase (1308 versus 20). We compared the measured redshifts against three published photometric redshift catalogs and find the photo-z accuracy to be lower than the constraints provided by photo-z fitting codes. Eighty percent of the galaxies in our final catalog have an HST counterpart. These galaxies are on average faint, with a median AB $F775W$ magnitude of 25.7 and 28.7 for the [O II] and Lyα emitters, respectively. Fits of their spectral energy distribution show that these galaxies tend to be low-mass star-forming galaxies, with a median stellar mass of $6.2 \times 10^8\ M_\odot$ and a median star-formation rate of $0.4\ M_\odot$ yr^{-1}. We measured the completeness of our catalog with respect to HST and found that, in the deepest 141-h area, 50% completeness is achieved for an AB magnitude of 27.6 and 28.7 ($F775W$) at $z = 0.8$–1.6 and $z = 3.2$–4.5, respectively. Twenty percent of our catalog, or 424 galaxies, have no HST counterpart. The vast majority of these new sources are high equivalent-width $z > 2.8$ Lyα emitters that are detected by MUSE thanks to their bright and asymmetric broad Lyα line. We release advanced data products, specific software, and a web interface to select and download data sets.

Key words. galaxies: high-redshift – galaxies: distances and redshifts – techniques: imaging spectroscopy – catalogs

1. Introduction

In 2017 we published the first MUSE spectroscopic survey data release (Bacon et al. 2017; Inami et al. 2017, hereafter DR1) in the *Hubble* ultra-deep field (Beckwith et al. 2006) area. These two papers were published together with eight other papers, providing a first glimpse into the richness of the data and their scientific impact. In that series of papers we investigated the photometric redshift properties of the sample (Brinchmann et al. 2017), the properties of C III] emitters (Maseda et al. 2017), the spatially resolved stellar kinematics of galaxies at $z = 0.2$–0.8 (Guérou et al. 2017), the faint end of the Lyα luminosity function (LF; Drake et al. 2017), the properties of Fe II* emission (Finley et al. 2017), the extended Lyα halos (Leclercq et al. 2017), the evolution of the galaxy merger fraction (Ventou et al. 2017), and the equivalent-width properties of Lyα emitters (Hashimoto et al. 2017).

These initial studies did not exhaust the scientific content of the data set and were quickly followed by many others: the study of Mg II emission and absorption in star-forming galaxies (Feltre et al. 2018), the recovery of systemic redshifts from Lyα line profiles (Verhamme et al. 2018), the covering fraction of Lyα emission (Wisotzki et al. 2018), the low-mass end of the star formation sequence (Boogaard et al. 2018), the spatially resolved properties of Lyα halos (Leclercq et al. 2020), the evolution of the Lyα emitter fraction (Kusakabe et al. 2020), the molecular gas properties of high redshift galaxies (Boogaard et al. 2019, 2020, 2021; Inami et al. 2020), the study of high equivalent-width Lyα emitters (Maseda et al. 2018, 2020), the study of He II emission line properties (Nanayakkara et al. 2019), the rest-UV properties of Lyα emitters (Feltre et al. 2020), and the angular momentum of low-mass star-forming galaxies (Bouché et al. 2021).

Shortly after the completion of the MUSE *Hubble* Ultra-Deep Field (hereafter HUDF) survey, MUSE was successfully coupled to the ground-layer module of the VLT adaptive optics

★ Catalogs are only available at the CDS via anonymous ftp to cdsarc.cds.unistra.fr (130.79.128.5) or via https://cdsarc.cds.unistra.fr/viz-bin/cat/J/A+A/670/A4

★★ Based on observations made with ESO telescopes at the La Silla Paranal Observatory under the programs 094.A-0289(B), 095.A-0010(A), 096.A-0045(A), 096.A-0045(B) and 1101.A-0127.

facility (AOF; Leibundgut et al. 2017; Madec et al. 2018). The new system offers improved spatial resolution as well as faster survey times by allowing a wider range of atmospheric conditions (i.e., seeing and airmass) for operations.

The DR1 was based on two data sets: a 3×3 arcmin^2 mosaic of nine MUSE fields at a 10-h depth (hereafter MOSAIC) and a single 1×1 arcmin^2 31-h depth field (hereafter UDF-10). The improved depth of UDF-10 proved to be essential, especially for the study of the faint end of the Lyα emitter population and the study of diffuse emission. We then decided to push forward in depth and start a new adaptive-optics-assisted GTO survey in the same area with the goal of reaching a depth of over 100 h. This new survey, called the MUSE eXtremly Deep Field (MXDF), was completed in January 2019. With an achieved depth of 141 h, it is a key addition to the existing MOSAIC and UDF-10 data sets and is the deepest spectroscopic survey ever performed.

In addition to these new observations, we significantly improved the data-reduction and data-analysis system with respect to DR1 (Sect. 6.1.2). The MOSAIC and UDF-10 data sets were therefore reprocessed with the same tools and methodology used for the MXDF processing. This second data release (DR2), which incorporates these three data sets, aims to provide a comprehensive and homogeneous deep spectroscopic survey in the HUDF iconic field.

The paper is organized as follows: the MXDF observations are reported in Sect. 2 and the improved data reduction in Sect. 3 and Appendix B. Section 4 presents the data properties. The processes of source detection and classification are described in Sect. 5, and the resulting catalogs and sample properties are presented in Sect. 6. A summary and conclusions are given in Sect. 7. Finally, the released advanced data products and software are detailed in Appendix A.

Data reduction and catalog building for deep integral field unit (IFU) data is not yet common. Therefore, we aim to provide sufficient details on the methodology used in all our data processing and data analysis steps, to enable future users to perform similar deep fields. Readers not interested in these technical details can skip Sects. 3–5.

2. Observations

In this section we report only recent observations of MXDF (GTO Large Program 1101.A-0127, PI R. Bacon). For previous observations related to the UDF-10 and MOSAIC fields, we refer to Bacon et al. (2017). The observing campaign started in August 2018 and lasted until January 2019 for a total of 6 runs made during new moon periods.

All observations were made with the VLT's AOF and GALACSI, its dedicated ground-layer adaptive optics (GLAO) system (Kolb et al. 2016; Madec et al. 2018). With respect to non-AOF observations, the only change in the MUSE instrumental configuration is the notch filter that blocks bright light due to the four sodium laser guide stars in the 5800–5966 Å wavelength range (Fig. 1). The AOF ran smoothly and achieved robust performance during all runs.

The location of the MXDF field (Fig. 2) was chosen (i) to be in the HUDF extremely deep field region with WFC3 deep imaging (Illingworth et al. 2013) (ii) to match the AOF tip/tilt natural star requirement (i.e., $V > 18.5$ in a 3.5 arcmin field of view and outside the MUSE field of view), and (iii) to have a usable slow-guiding star in the outer circle to compensate for the derotator's wobble and possible misalignment of the Nasmyth platform with respect to the telescope's focal plane.

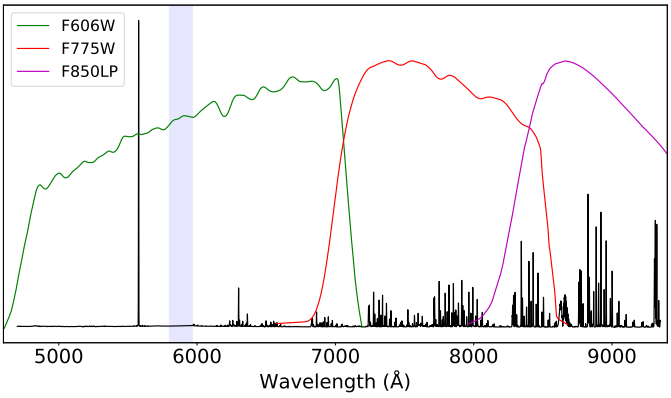


Fig. 1. MUSE spectral range. The typical sky spectrum (shown in black) is observed in one 25 min observing block. The blue shaded wavelength region shows the location of the sodium notch filter used in MUSE GLAO mode. The response curves of the HST ACS filters *F606W*, *F775W*, and F850LP are also indicated.

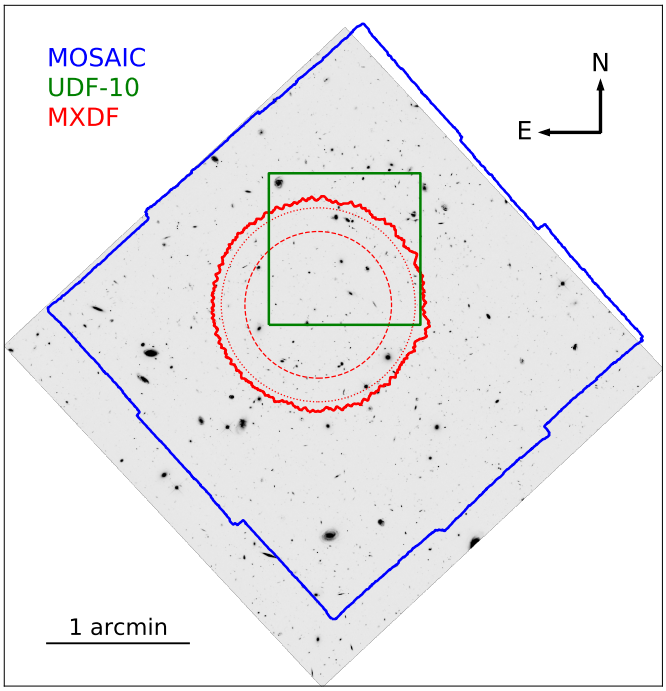


Fig. 2. Location of the three deep fields used in this paper: MXDF (141-h depth), MOSAIC (10-h depth), and UDF-10 (31-h depth) overlaid on the HST *F775W* UDF image. The dotted and dashed red circles show the MXDF 10- and 100-h exposure time isocontours, respectively.

As shown in Table 1, we observed the field over a wide range of atmospheric conditions, including poor seeing (1.2 arcsec) and up to high air mass (1.4). The ground layer fraction, a key parameter for GLAO observations – the larger the better – averaged 65%, and rarely was less than 50%. We observed some correlation between seeing and ground layer fraction: on many occasions, poor seeing is related to increased ground layer turbulence. This behavior of the Paranal atmospheric turbulence is fortunate because it makes the GLAO mode very effective.

It is instructive to compare the observing conditions of the UDF-10 and MOSAIC campaigns with the MXDF campaign (Table 1). The first campaign, conducted before the AOF was commissioned, was limited to the natural seeing observing mode of MUSE. To maximize the final spatial resolution, we therefore

Table 1. Summary of MXDF, UDF-10, and MOSAIC observations.

Field	Runs [a]	Integ [b]	Airmass [c]			Seeing [c], [d] (″)			GL frac [c], [e]		
			q50	q10	q90	q50	q10	q90	q50	q10	q90
MXDF	08/18–01/19 (6)	155	1.08	1.01	1.37	0.76	0.54	1.11	0.65	0.46	0.83
UDF-10 & MOSAIC	09/14–02/16 (8)	116	1.05	1.00	1.19	0.75	0.60	0.86			

Notes. [a]Starting and end dates. The number of conducted runs is indicated in parentheses. [b]Total open shutter time in hours. [c]50% (q50, median), 10% (q10) and 90% (q90) percentile. [d]Seeing is measured in the pointing direction with the AO telemetry for the MXDF observations and with the telescope guiding camera for the MOSAIC and UDF-10 non-AO observations. [e]Fraction of 1 km ground-layer turbulence.

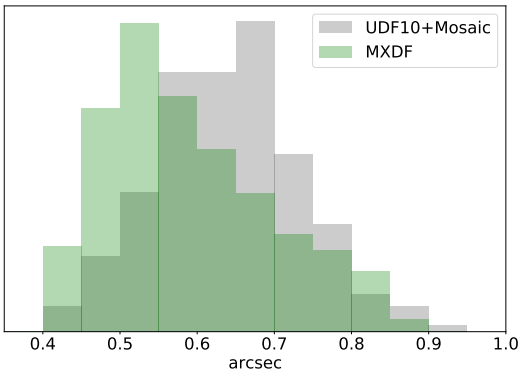


Fig. 3. Spatial resolution of individual MXDF observations compared to the UDF-10 plus MOSAIC observations. The FWHMs (in arcsec) are derived from the Moffat PSF model at 7000 Å. The normalized FWHM histograms of the MXDF and UDF-10 plus MOSAIC observations are shown in green and gray, respectively. The FWHM MXDF values have been corrected for a +0.06 arcsec offset (see text).

decided to observe the UDF field only under good seeing conditions (seeing less than 0.8 arcsec) and at low air mass (less than 1.2). The consequence is that it took 2.5 yr and 8 runs to accumulate the required telescope time. In contrast, the MXDF campaign took only 6 months and 6 consecutive runs to reach 155 h of integration time. As we will see in Sect. 4.2.1 (Fig. 3), thanks to the performance of the AOF/GALACSI GLAO, the spatial resolution achieved by MXDF is better than that of UDF-10 and MOSAIC, despite its relaxed atmospheric observing conditions.

The MXDF observing strategy also differs from the 90-degree rotation plus small offset dithering patterns commonly used in MUSE observations. Each observing block is similar to the scheme used in the first campaign: that is, a set of four 25 min exposures with successive 90° instrument rotations. But in order to further reduce the systematics and to break the horizontal and vertical patterns introduced by the instrument field splitter and slicer geometry, we systematically rotated the field of view by a few degrees between each observing block. Consequently, the final combined field of view is approximately circular (Fig. 2) with a radius of 41″ and 31″ for respectively 10+ and 100+ h of depth. The field center celestial coordinates are 53.16467 deg, −27.78537 deg (J2000 FK5). After the rejection of a few bad exposures due to satellite track contamination, poor final spatial resolution, or cloud absorption, the achieved final maximum depth was 141 h.

3. Data reduction

After DR1, described in Bacon et al. (2017), we continued to work on the data reduction process, which led to several major improvements. The overall process is similar to DR1, with important changes in the self-calibration algorithm, in the sky-subtraction with the Zurich Atmospheric Purge (ZAP) software (Soto et al. 2016, Appendix B.3), and the use of a "superflat." This process was first applied to the UDF-10 and MOSAIC (278 exposures) and then to the MXDF (373 exposures). The resulting depth and color images for the three data sets are shown in Figs. 4 and 5, respectively. A detailed description of this improved data reduction process is given in Appendix B.

4. Data properties

4.1. Astrometry and photometry

As in DR1, the datacube's world coordinate system has been matched to the *Hubble* ACS astrometry (Appendix B). By construction, the astrometry of the datacube should therefore be consistent with the HUDF published photometric catalogs: UVUDF (Rafelski et al. 2015), 3D-HST (Skelton et al. 2014), and CANDELS v2 (Whitaker et al. 2019). We note that Dunlop et al. (2017) and Franco et al. (2018) have shown that the *Hubble* Space Telescope (HST) astrometry is offset from the *Gaia* DR2 catalogs (Gaia Collaboration 2016, 2018). The offset, as measured by Whitaker et al. (2019), amounts to ΔRA = +0.094 ± 0.042 arcsec and ΔDec = −0.26 ± 0.10 arcsec. In this paper, we have ignored this offset in order to maintain the same astrometric reference as the HST catalogs.

We measured the astrometric accuracy achieved by comparing source centroids in the HST ACS *F775W* high resolution image with the corresponding values measured in MUSE reconstructed image. We proceeded as follows: For each field, we derived the equivalent *F775W* broadband image by computing at each spaxel the transmission-weighted average flux of the corresponding *F775W* ACS filter (Fig. 1). This reconstructed image is then convolved by the HST ACS point spread function (PSF). We approximated this PSF using a Moffat function[1] with a 0.085 arcsec full width at half maximum (FWHM) and β = 1.60. The HST ACS *F775W* image is convolved with the MUSE PSF value at the *F775W* reference wavelength (7750 Å), that is, 0.45 arcsec FWHM and β = 1.89, 0.60 arcsec FWHM and β = 2.80, and 0.63 arcsec FWHM and β = 2.80 for the MXDF, UDF-10, and MOSAIC data sets, respectively. Source detection is performed on both convolved images with SExtractor (Bertin & Arnouts 1996). We tune SExtractor's detection parameters to obtain similar segmentation maps between HST and MUSE convolved images. The two catalogs are crossmatched and the mean offset and its standard deviation are derived using 3σ sigma-clip statistics on the matched catalog. The astrometric

[1] The circular 2D Moffat profile (Moffat 1969) has two parameters: the FWHM in arcsec and the β shape parameter, β = 1 and $\beta \gg 1$ for respectively a Lorentzian or a Gaussian distribution.

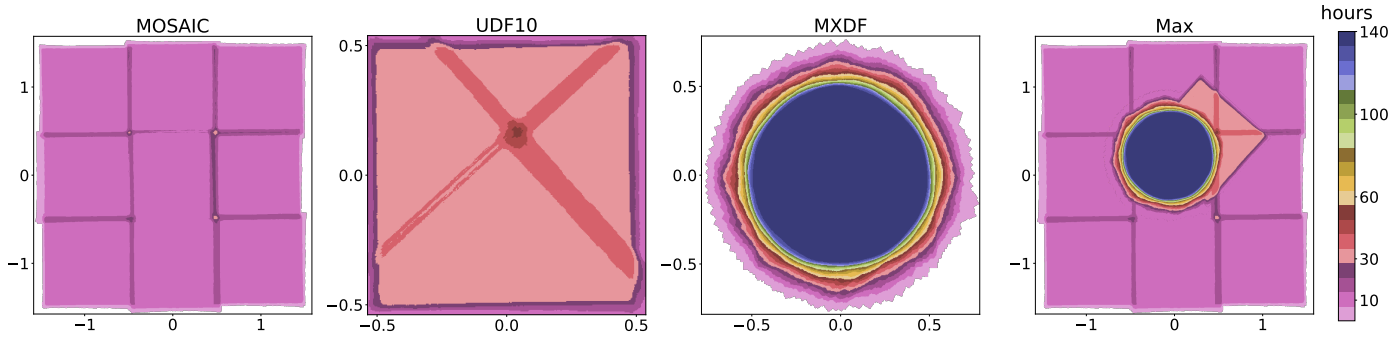


Fig. 4. Achieved depth in hours. The MOSAIC, UDF-10, and MXDF exposure maps are shown in the first three panels, starting from the left. The Max exposure map, computed as the maximum depth for each spaxel, is shown in the *right panel*. UDF-10 and MXDF are to the north (*top*) and the east (*left*), respectively, while the MOSAIC and Max exposure maps are rotated by 42° (see Fig. 2). Axis labels are in arcmin.

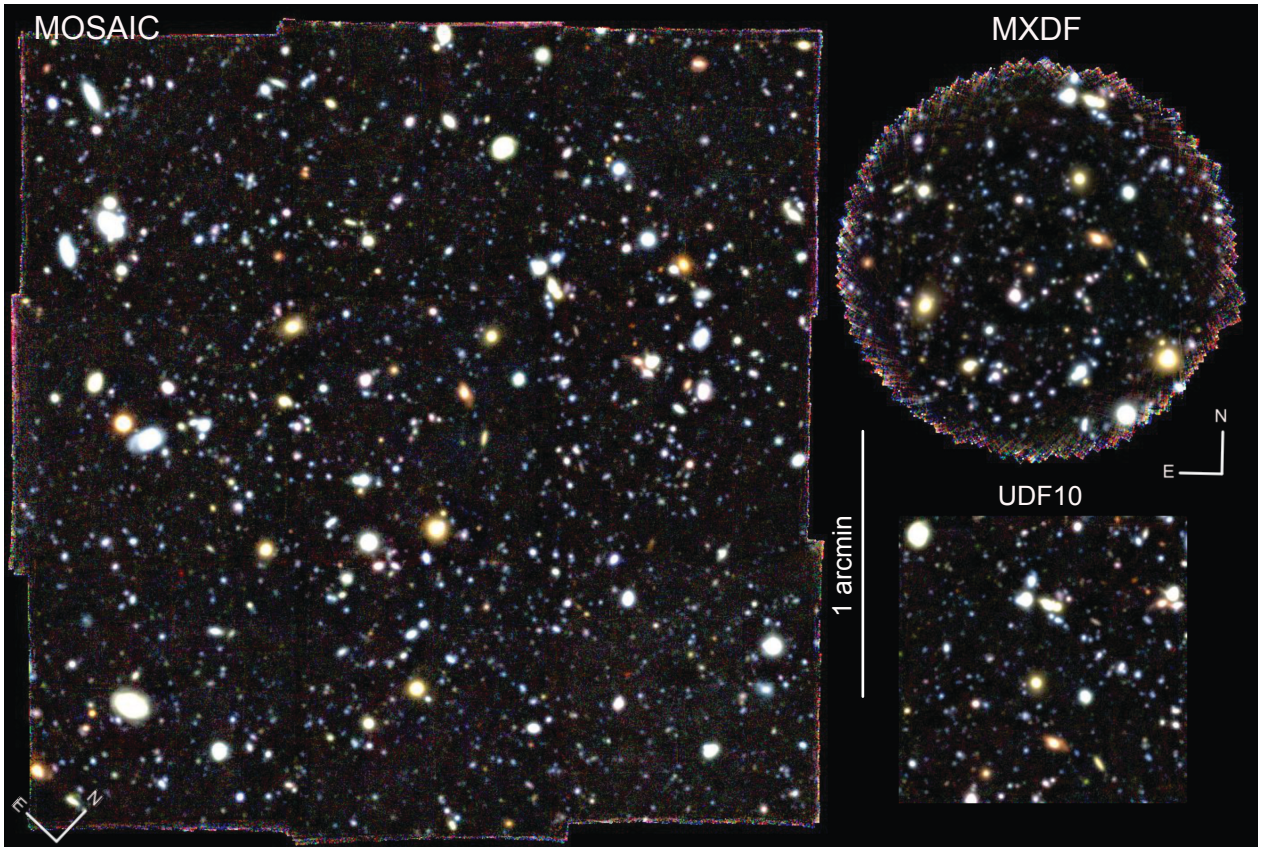


Fig. 5. Reconstructed pseudo-color images of the MOSAIC, MXDF, and UDF-10 data sets.

differences between the MUSE datacubes and HST ACS F775W image are given in the upper part of Table 2. If we exclude the faint sources with AB > 27, the mean astrometric relative error is about 0.1 arcsec, which is half of a spaxel or about 1/6 of the spatial resolution. We note that astrometric errors are larger for MXDF than for the MOSAIC data set. Such an increase with depth is expected given the increase of source confusion.

We used the same crossmatched catalogs to evaluate the achieved accuracy in source AB magnitude for MUSE data sets with respect to HST photometry. The results, reported in the lower part of Table 2, are based on SExtractor computed automatic aperture magnitude[2]. We note that the magnitude difference is below 0.1 for MXDF sources with AB < 27. As expected, the magnitude differences decrease with the source brightness and increase with depth.

4.2. Spatial and spectral resolution

4.2.1. MXDF spatial PSF

Spatial resolution is an important ingredient for most analyses. The spatial PSF was first estimated for each individual raw exposure to reject bad exposures before the final datacube was produced. Then, the final PSF was estimated on the combined datacube.

[2] SExtractor's automatic aperture photometry routine is derived from Kron's algorithm (Kron 1980); see `https://sextractor.readthedocs.io/en/latest/Photom.html` for more detailed information.

Table 2. Estimated astrometric and photometric accuracy for the three data sets when compared to the HST ACS *F775W* image.

Data set	AB < 25	25 < AB < 27	27 < AB
	Average astrometric offsets (arcsec)		
MXDF	0.12 ± 0.04	0.13 ± 0.05	0.13 ± 0.06
UDF10	0.08 ± 0.02	0.10 ± 0.05	0.15 ± 0.09
MOSAIC	0.05 ± 0.02	0.08 ± 0.05	0.16 ± 0.10
	Average AB magnitude differences		
MXDF	0.08 ± 0.04	0.09 ± 0.07	0.25 ± 0.19
UDF10	0.07 ± 0.04	0.23 ± 0.13	0.35 ± 0.24
MOSAIC	0.10 ± 0.08	0.24 ± 0.16	0.32 ± 0.23

Notes. Mean and standard deviation are given for three AB magnitude intervals. The mean AB magnitude error reported by SExtractor for the HST convolved image amounts to 0.002, 0.011 and 0.030 in the AB < 25, 25 < AB < 27 and AB > 27 intervals, respectively.

Unfortunately, there is no point source bright enough in the MXDF field of view to estimate the PSF in a single exposure of 25 min integration. Moreover, the seeing value reported by the Paranal seeing monitor is not sufficient by itself to estimate the PSF given the adaptive optics (AO) correction provided by the AOF GLAO system. We therefore used the tool muse-psfr developed by Fusco et al. (2020) to derive an estimate of the PSF using AO telemetry. During an exposure, atmospheric parameters such as seeing, ground-layer turbulence fraction, turbulence outer scale, are derived from wavefront sensor telemetry and regularly recorded in a table in the raw FITS exposure. The algorithm uses this information to produce an estimate of the spatial PSF in the form of a circular 2D Moffat profile. Fusco et al. (2020) measure a standard deviation of 0.06 arcsec for the derived FWHM. We note that some systematic offsets were found at high spatial resolution (Fig. 17 of Fusco et al. 2020). We also found a similar offset when we compared the derived value of muse-psfr with a direct fit to the combined datacube (Table 3).

We show in Fig. 3 the distribution of the computed MXDF FWHMs[3] at 7000 Å for the 373 individual exposures. The benefit of the GLAO AOF system is evident when the histogram of MXDF FWHMs is compared to the corresponding MOSAIC plus UDF-10 values. Due to the specific rotation scheme used for the MXDF observations, the outer ring of the field results from the combination of different exposures and may thus have a potentially different PSF than the central part. We show in Fig. 6 that this is indeed the case with a variation of 0.10 arcsec FWHM. We note, however, that within the area with depth >100 h there is no detectable spatial variation in the PSF.

Although there is no bright star in the field of view, there is a fainter M star (ID 5102 in R15 catalog, *F775W* AB magnitude 24.7) located in the 100-h depth area that can be used to measure the PSF. To take into account the evolution of the PSF with wavelength, we divided the final combined datacube into 20 wavelength slices of 232 Å each and fit a circular Moffat function to each image. As shown in Fig. 7 (left panels), the circular Moffat function represents the data well. The figure also shows the evolution of the shape parameters FWHM and β as a function of wavelength. Except at the blue end of the MUSE wavelength range, where the M star becomes too faint, the shape parameter β is approximately constant. We adopt $\beta = 2.123$ and perform the Moffat fit again, leaving the FWHM as a variable. The adopted third-order polynomial fitted to the evolution of the FWHM as a function of wavelength is shown in the upper right panel of Fig. 7. The final spatial PSF is then given by

$$\mathrm{PSF}(r,\lambda) = (\beta-1)/(\pi\alpha(\lambda)^2) \times (1 + r^2/\alpha^2)^{-\beta} \quad (1)$$

$$\alpha(\lambda) = FWHM(\lambda) \times \left[2\sqrt{2^{1/\beta}-1}\right]^{-1}$$

$$FWHM(\lambda) = a_3\lambda'^3 + a_2\lambda'^2 + a_1\lambda' + a_0$$

$$\lambda' = (\lambda - 4850)/(9350 - 4850) - 0.5.$$

The wavelength (λ) and FWHM units are Å and arcsec, respectively. The β value and the FWHM polynomial coefficients (a_n) are given in Table 4.

We compared this PSF estimate with the values derived by convolution of the HST images for the *F606W*, *F775W*, *F814W*, and F850LP broadband filters. The method is identical to that used in the paper I (Bacon et al. 2017, Sect. 5.1) for the MOSAIC and UDF-10 datacubes. For ease of comparison, a

[3] A value of 0.06 arcsec was added to the values of MXDF muse-psfr to account for the measured offset between muse-psfr and the direct Moffat fit.

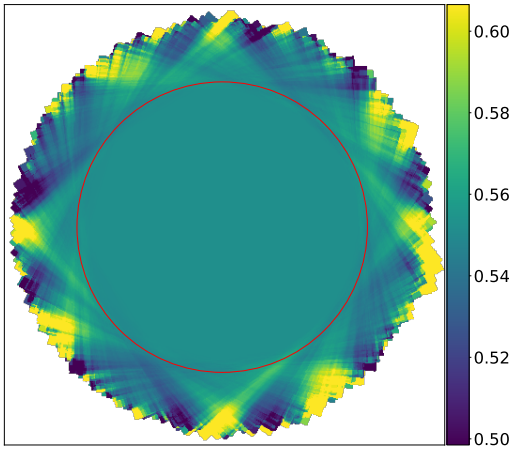


Fig. 6. Field evolution of the MXDF spatial resolution. The colors indicate the mean value (in arcsec) of all individual Moffat FWHMs at 7000 Å at each spaxel location as measured by the PSF AO reconstruction algorithm and corrected for an offset of +0.060 arcsec. The red circle displays the 100-h depth contour.

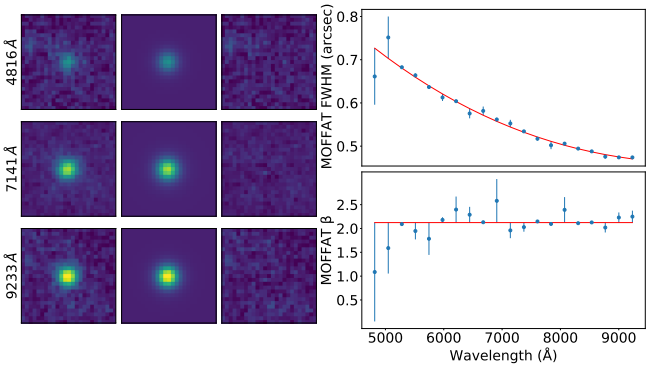


Fig. 7. MXDF Moffat fit of the M star RID-5102. The fit is performed on 20 narrowband images of 232 Å width, equally distributed along the MXDF cube wavelength axis. Examples of the first, central, and last images (each box side is 5 arcsec in length) are displayed in the first column. The corresponding Moffat fit and residuals are shown in the two other columns. The FWHM and β Moffat fitted values are shown in the upper and bottom right panels, together with their polynomial approximation.

Table 3. Comparison of different spatial MXDF PSF estimates for four HST reconstructed broadband images.

Filter	Moffat FWHM			Moffat β			Eq. Gaussian FWHM		
	Star	HST	AO tel	Star	HST	AO tel	Star	HST	AO tel
F606W	0.60	0.64	0.54	2.10	2.41	2.01	0.72	0.75	0.65
F775W	0.52	0.56	0.45	2.17	2.25	1.88	0.61	0.66	0.55
F814W	0.50	0.55	0.43	2.19	2.35	1.84	0.59	0.64	0.53
F850LP	0.48	0.51	0.41	2.25	2.25	1.80	0.57	0.60	0.50

Notes. FWHM (arcsec) and β values are given for the direct Moffat fit of the RID-5102 M star (Star columns), the comparison with HST (HST columns) and the mean of the PSF reconstruction values (AO tel columns). The Gaussian equivalent FWHMs are given in the last three columns.

Table 4. Circular Moffat PSF models for the three MUSE data sets.

Data set	β	a_0	a_1	a_2	a_3	F_{4850}	F_{7000}	F_{9350}
MXDF	2.123	0.5465	−0.2474	0.1975	−0.0295	0.72	0.55	0.47
UDF-10	2.800	0.6179	−0.1353			0.69	0.62	0.55
MOSAIC	2.800	0.6378	−0.1515			0.71	0.64	0.56

Notes. β is the Moffat shape parameter and a_n are the coefficients of the polynomial approximation to FWHM(λ) defined in Eq. (1). In the case of the MOSAIC data set, we provide the median PSF for the nine fields. FWHM values in arcsec (e.g., F_{4850}) are given for the blue (4850 Å), central (7000 Å) and red (9350 Å) wavelengths.

direct Moffat fit was performed on the M star for the four reconstructed HST broadband MUSE images. An additional comparison was performed by fitting a Moffat distribution to the average of all individual muse-psfr PSFs. The Moffat FWHM and β shape parameter for the three different estimates are shown in the Table 3. Since the FWHM and β parameters are not independent, we also provide the FWHM obtained by fitting a Gaussian profile to judge the difference in FWHM. However, we recall that the MUSE PSF is not well represented by a Gaussian function (Husser et al. 2016; Weilbacher et al. 2020). As shown in Table 3, all methods give similar results with the expected decrease in FWHM with increasing wavelengths. However, there is a systematic offset between the methods: the HST convolution method overestimates the FWHM by ≈+0.05 arcsec while the muse-psfr method underestimates it by ≈–0.06 arcsec.

4.2.2. Line spread function

As in DR1, we measure the line spread function (LSF) on a combined cube of data where the sky has not been subtracted. A Gaussian fit was performed on the bright sky lines taking into account the contribution of adjacent fainter lines. As shown in Fig. 8, the MXDF LSF shows the expected instrumental evolution with wavelength. A good model of the median LSF FWHM is given by the UDF-10 model: LSF(λ) = $5.866 \times 10^{-8}\lambda^2 - 9.187 \times 10^{-4}\lambda + 6.040$, with LSF and λ in Å. The LSF is approximately constant in the field of view, except in the outer rings, which show larger variation. This is mainly due to the limited number of exposures that are combined together at the edges, which leaves the intrinsic variation between the 24 IFUs more visible.

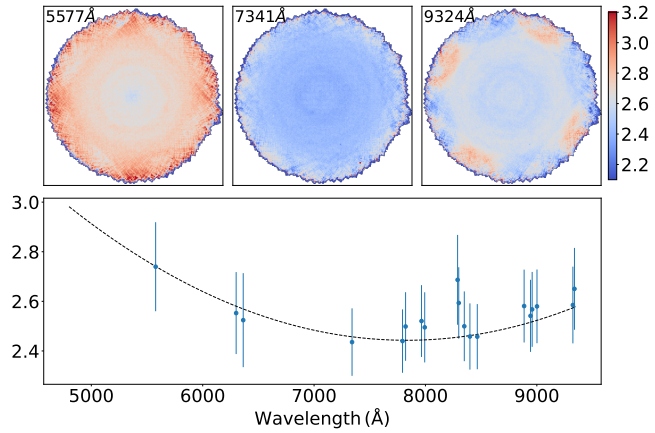


Fig. 8. Measurements of the MXDF LSF in the combined datacube (not sky subtracted). The bottom panel displays the fitted Gaussian FWHM for a few bright sky lines. Variations with the field of view are given by the error bars (standard deviation). The DR1 UDF-10 polynomial approximation is shown as a dashed line. Examples of FWHM spatial maps are given for three sky lines in the top panels. FWHM units is Å.

4.3. Noise property and limiting flux

To compare the noise level of the three data sets, we selected 214 one-arcsecond-diameter apertures located in empty locations and common to all data sets. The location of the aperture was chosen to have a similar depth for each data set: 140.8 ± 0.06 h, 31.2 ± 0.2 h, and 10.0 ± 0.3 h for the MXDF, UDF-10, and MOSAIC data sets, respectively. The corresponding noise level as a function of wavelength is calculated as the median of the 214 standard deviations. It should scale with exposure time (t) as $\sqrt{t}$. In Fig. 9 we display the UDF-10 and MXDF noise level scaled by the square root of the depth and divided by the MOSAIC scaled value. The figure shows that the deviation from the theoretical expectation is indeed very small (i.e., 10–20%). The slight improvement with depth results from a better control of the systematics due to the observation strategy (dithering and rotation; see Sect. 2).

We estimated the 1σ surface brightness limit for an unresolved lines ($SB_{\rm line}$) with the following formula:

$$SB_{\rm line}(\lambda) = \frac{\sqrt{V_{\rm line}(\lambda)}\,\Delta\lambda}{\sqrt{25}\,\Delta s^2} \text{with} V_{\rm line}(\lambda) = \sum_{\lambda'=\lambda_0-k\sigma(\lambda)}^{\lambda_0+k\sigma(\lambda)} V_s(\lambda'), \qquad (2)$$

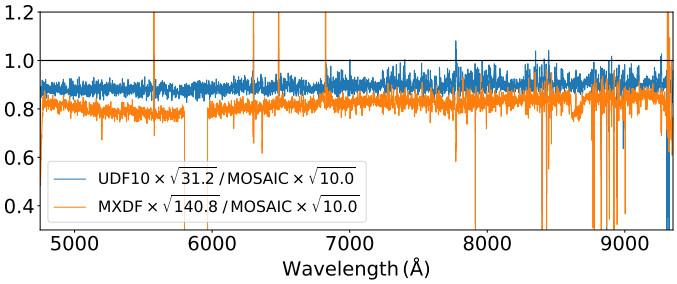


Fig. 9. Scaled raw noise comparison between the three data sets. The plot displays the UDF-10 and MXDF to MOSAIC scaled noise ratio. The scaled noise is computed as the median noise level for each data set in a common set of 1″ empty apertures, after scaling by the square root of the depth in hours. We note that the noise level has not been corrected for correlations caused by the interpolation.

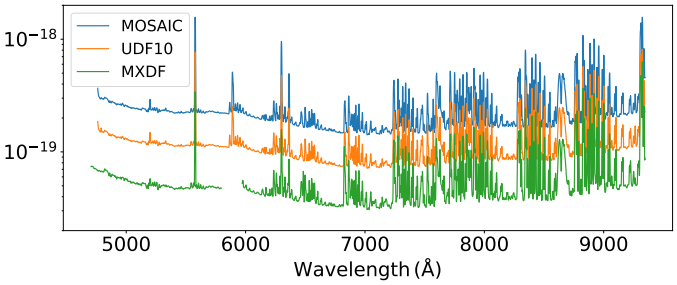


Fig. 10. Unresolved emission line median 1σ surface brightness limit for the three data sets in units of erg s^{-1} cm^{-2} arcsec^{-2}.

where $V_s(\lambda)$ is the variance by spaxel, $\Delta\lambda$ the wavelength bin size (1.25 Å), Δs the spaxel bin size (0.2 arcsec), 25 the number of spaxels in 1 arcsec^2 and $\sigma(\lambda)$ is derived from the LSF approximation given in Sect. 4.2.2 with $\sigma(\lambda) = \mathrm{LSF}(\lambda)/2.355$. The wavelength interval ($k = 1.29$) was chosen to capture 90% of the line flux. The variance by spaxel was computed on the central MXDF area with a depth larger than 120 h after masking the bright sources and taking the median of the datacube variance. We note that the variance is already corrected for the correlated noise as described in Appendix B.6.

The same computation was performed for the UDF-10 data set, restricted to the area with depths greater than 30 h, and the MOSAIC data set. Results are shown in Fig. 10. The best sensitivity is achieved in the MXDF deep area and in the 6800–7800 Å wavelength range (outside sky lines) where an unresolved emission line surface brightness of 1.0×10^{-19} erg s^{-1} cm^{-2} arcsec^{-2} will be detected at 3σ.

We perform a similar estimation for a point-like source with an unresolved emission line, using the spatial PSF estimated in Sect. 4.2.1 with the following formula:

$$F_{pl}(\lambda) = \Delta\lambda \sqrt{\frac{\pi r^2}{\Delta s^2} V_{\mathrm{line}}(\lambda)}, \quad (3)$$

where V_{line} is defined in Eq. (2) and r is the 80% enclosed flux circular radius for the corresponding Moffat PSF model at 7000 Å. The best point-source sensitivity (Fig. 11) is achieved in the MXDF central area and in the 6700–9300 Å wavelength range (sky lines excluded) with a 3σ detection limit of 6.3×10^{-20} erg s^{-1} cm^{-2} for an unresolved emission line.

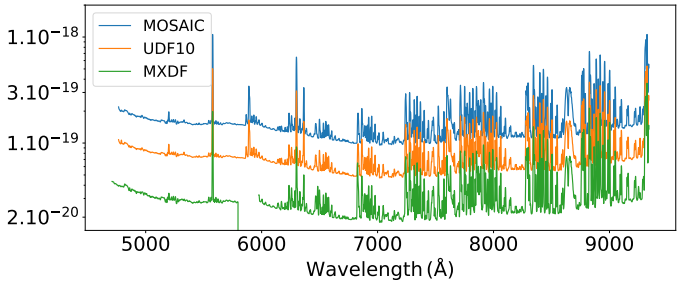


Fig. 11. Unresolved emission line median 1σ point-like source limit for the three data sets in units of erg s^{-1} cm^{-2}.

5. Source detection and classification

The next stage after the production of the final datacubes for the three data sets (MOSAIC, UDF-10 and MXDF) is the detection and classification of sources. This is a delicate task, involving many steps to ensure the data quality and homogeneity of the final product. An overview, followed by detailed descriptions of each of the step, is presented in the following subsections.

5.1. Overview

A schematic of the process is shown in Fig. 12. The first phase is the production of inspection data sets (process chart one in Fig. 12). Candidate sources are produced in two parallel branches: (i) from blind detection using ORIGIN software (Sect. 5.2) and (ii) from de-blended extraction using the ODHIN algorithm (Sect. 5.3) based on the R15 HST catalog and segmentation map. For each candidate source, the first five redshift solutions obtained with the pyMarZ software are recorded (Sect. 5.4). For each redshift solution, the corresponding emission and absorption lines are computed with pyPlatefit (Sect. 5.5) and narrow bands for the highest S/N lines are derived (Sect. 5.6). The data resulting from the different processes are organized into a database with a set of candidate source catalogs, one for each input datacube and each detection branch (i.e., ORIGIN and ODHIN).

The second phase (process chart two in Fig. 12, Sect. 5.7) consists of the visual inspection of a set of candidate sources by a group of experts who select the appropriate redshift solution, assigns confidence level and crossmatch the source with the HST R15 and DR1 catalogs. Conflicts between the experts' solutions are identified and resolved in reconciliation meetings. Duplicates between ORIGIN and ODHIN sources are sorted out and the retained solutions are attributed a unique MUSE identifier[4].

In the last step (process chart three in Fig. 12, Sect. 5.8), we selected the most appropriate extraction and refined the redshift solution, emission, and absorption line fits and narrow bands for each source. The results are inspected and corrective actions are identified and executed. This process is iterated until the results are deemed satisfactory. At the end of this last step, the final data products are produced, that is, a set of catalogs (Appendix A.2) and data organized in the MUSE Python Data Analysis Framework (MPDAF) source format (Appendix A.3).

[4] We use the following naming convention: MID-nnnn for MUSE identifiers and RID-nnnn for R15 identifiers.

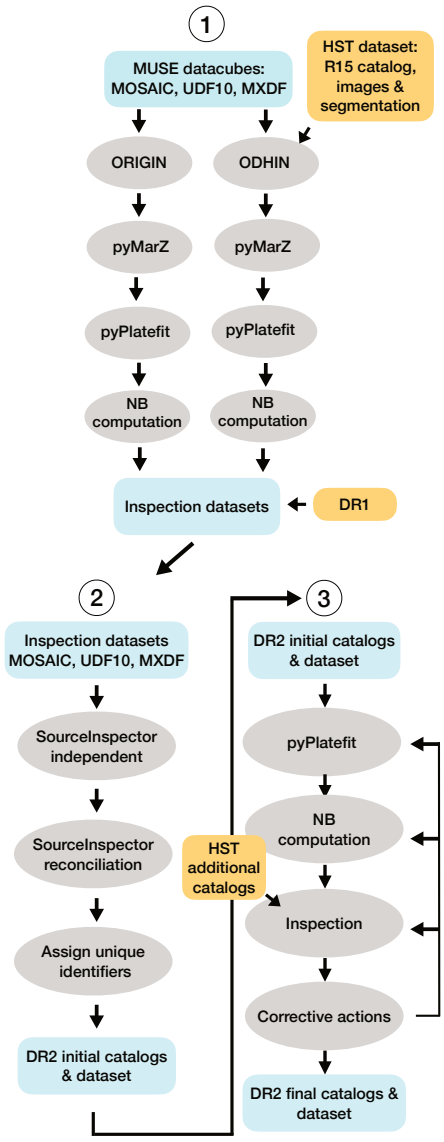


Fig. 12. Schematic of the processes involved in the data release production.

5.2. ORIGIN blind source detection

The ORIGIN software (Mary et al. 2020) has been developed to automatically detect faint line emitters in MUSE datacubes. It is optimized for the detection of compact sources with faint spatial-spectral emission signatures and provides an automated and reliable estimate of the purity (i.e., one minus the fraction of false positives, Benjamini & Hochberg 1995; Mary & Roquain 2022). All three data sets were processed with the same input parameters. Below we present the example of the MXDF data set.

The preprocessing steps (Fig. 13) perform a first pass continuum removal using a discrete cosine transformation of order ten along the datacube wavelength axis. The resulting white light continuum image is then segmented to mask the bright continuum sources. A second segmentation is also performed on the S/N residual image (i.e., the reconstructed continuum-free image of the datacube divided by the square root of the variance). The two segmentation masks are merged. The unmasked region defines the "faint region" that will be used in the following steps to avoid disturbance due to bright or noisy sources.

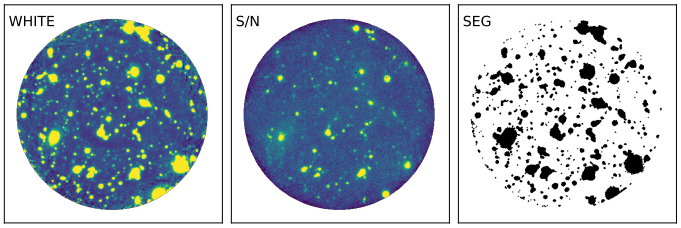


Fig. 13. ORIGIN preprocessing steps. From left to right, the MXDF MUSE white-light image, the S/N image and the segmentation map used to identify continuum sources and regions of low S/N.

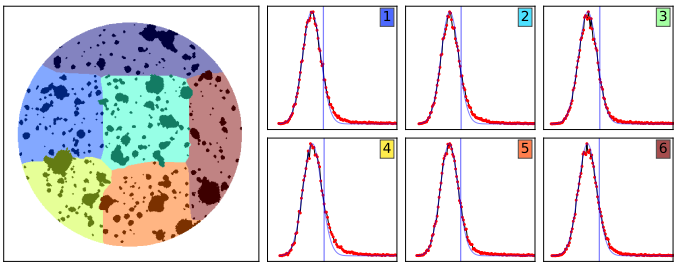


Fig. 14. ORIGIN PCA thresholds. The left image displays the field segmentation in 6 zones. The 6 plots show the corresponding histograms of the PCA test statistics for each zone. The red curves exhibit the Gaussian fit of the distribution. The computed thresholds are displayed as blue vertical lines. They identify the minimum value above which the spectrum triggers the PCA cleaning process. This value corresponds to the 10% upper quantile of the empirical distribution.

Although most of the continuum signal was removed in the preprocessing steps, there are still residuals in the resulting datacubes that will create too many false positives in the detection step, forcing the use of a higher detection threshold and preventing the detection of faint emission line sources. The continuum residuals, as well as the systematics present in the datacube, are subsequently removed with the iterative Principal Component Analysis (PCA) process described in Sect. 3.1.3 of Mary et al. (2020). The MXDF field of view is segmented into six zones (Fig. 14 left panel; see Sect. 3.1.2 of Mary et al. 2020). In each segment, a threshold is calculated on the spaxel distribution to evaluate the deviation from the normal distribution[5]. This threshold is then used by the iterative PCA process to clean the signal from nuisance sources. In the left panel of Fig. 15, we show the number of cleaning iterations at each spaxel. We note that while in most of the spaxels only a few iterations were needed to clean the data, many more were required at specific locations (e.g., at the edge of the field or in the presence of complex bright sources residuals).

The datacube is then filtered using a library of possible 3D signatures, constructed as spectral line profiles spread spatially by the PSF and a generalized likelihood ratio (GLR) is computed for each voxel, as described in Sect. 3.2.1 of Mary et al. (2020). The image of the maxima of the resulting GLR datacube is shown in Fig. 15, right panel. A comparison of the local maxima counts in the GLR datacube, restricted to the faint area, with the corresponding local minima counts is used to derive the purity as a function of the threshold applied to the cube of local maxima (Fig. 16, left panel). Using a purity of 0.8 for the GLR datacube, we identify 984 sources after spatial grouping of

[5] The noise distribution in the MXDF is well represented by a normal distribution, as are the UDF-10 and MOSAIC noise statistics; see Fig. 18 of Bacon et al. (2017).

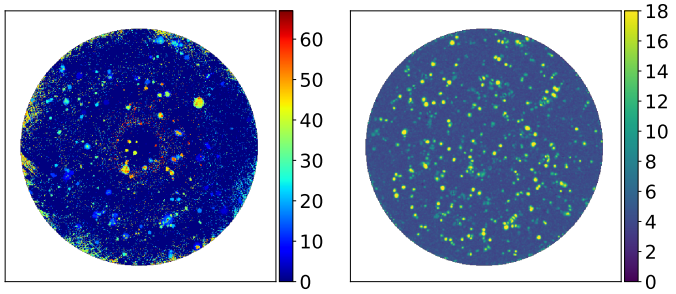


Fig. 15. Number of iterations where each spaxel was involved in the iterative PCA is shown in the left panel. The ORIGIN maxmap, computed as the maximum of the General Likehood Ratio test datacube over the wavelength axis, is displayed in the right panel.

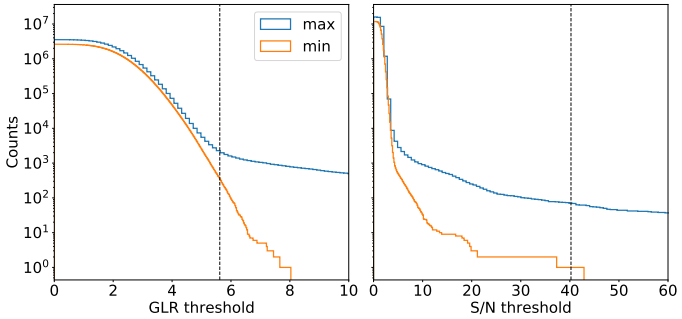


Fig. 16. Counts of positive and negative local maxima in the GLR datacube as a function of the threshold are shown in the left panel in blue and orange, respectively. The GLR threshold value computed to achieve the target purity value of 0.8 is displayed as a dotted vertical line. A similar plot for the S/N datacube and a target purity of 0.90 is shown on the right panel.

the detected emission lines. A similar operation is performed in the S/N datacube, prior to PCA subtraction, to recover possible bright sources affected by the PCA process. In that case we use a higher purity of 0.9 to search for these additional bright sources. This yields 18 more sources.

For each source, an optimal extraction is performed on the MUSE datacube with the Horne (1986) algorithm and using as a weighting and segmentation image, the image obtained by summing the GLR datacube over a wavelength window centered on the detection peak wavelength and with a width of $\pm 2 \times$FWHM, where FWHM is the width of the spectral template that provides the highest correlation peak (see Sect. 3.5 of Mary et al. 2020). We emphasize that the extraction is always performed on the original MUSE datacube, the ORIGIN pseudo narrowband image being used only as weighting image.

The process was similar for the UDF-10 and MOSAIC data sets. We compared our results with another blind source detection software program, LSDCat, developed by Herenz & Wisotzki (2017) in the context of the MUSE-Wide survey (Herenz et al. 2017; Urrutia et al. 2019; Schmidt et al. 2021). LSDCat is based on a 3D matched filter, which is, to a first approximation, similar to the mathematical basis of the ORIGIN algorithm. However, there are some important differences (e.g., the continuum subtraction or the use of GLR) and we should then be able to assess the robustness of our results by comparing the two detection catalogs. We use the same MOSAIC datacube to perform this comparison. Excluding the UDF-10 footprint to have a homogeneous depth, the LSDCat catalog is composed of 1190 sources. This must be compared to the 1287 ORIGIN sources with a redshift confidence ZCONF $>$ 0 (see Sect. 5.7.3) using the same selection. 81 LSDCat sources (6%) are unmatched in the ORIGIN catalog. A detailed examination of these sources shows that most of them are low purity sources that were detected by ORIGIN but not confirmed in the inspection process. Of the ORIGIN sources, 246 (19%) are not matched to the LSDCat catalog. Half of these sources are low confidence (ZCONF = 1) sources that fall below the LSDCat detection threshold. Overall, the vast majority of high confidence sources (93%) are detected by the two methods, giving confidence in the robustness of the results.

5.3. ODHIN HST-prior spectra extraction

For deep exposures, source blending increases, to a point where source confusion begins to be a problem. Source blending is already present in the 10-h depth MOSAIC field and of course has an increased impact in the ten times deeper MXDF data. Fortunately, the 3D information content of MUSE helps to identify source mixing. For example, the location of the emission line can be measured to an accuracy of about half a spaxel (0.″1) in reconstructed narrow bands when the S/N is sufficiently high. For blended sources with different emission lines it is then possible to separate the different contributions. Although this does not work with the continuum, we have another very valuable piece of information: the high spatial resolution broadband images from HST.

The combination of data at different spatial and spectral resolutions is known as pansharpening in signal processing of hyperspectral data (see the review in Loncan et al. 2016). The aim of these methods is to reconstruct a datacube with high spatial and spectral resolution. However, they are not directly applicable to our problem due to the very different spatial and spectral resolution in MUSE and HST[6] and the presence of noise. We have therefore developed a new and more robust method, nicknamed ODHIN, based on linear source demixing.

The method assumes that at HST spatial resolution a source can be modeled as a single spectrum. This is obviously not valid for nearby galaxies that exhibit spatially resolved spectral variations. This is also not strictly valid at high redshift where some line emission may be more extended than the broadband morphology (e.g., Leclercq et al. 2017). ODHIN is similar to the TDOSE method (Schmidt et al. 2019) developed for the MUSE-Wide survey (Urrutia et al. 2019; Schmidt et al. 2021) but it differs in three aspects: it is nonparametric, it uses multiple broadband HST images, and it implements a regularization process to avoid noise amplification for very close sources. The details of the algorithm are presented in Appendix C.

The inputs to the algorithm are the 3 HST broadband images ($F606W$, $F775W$, and F850LP) that fall within the MUSE spectral range (Fig. 1), the HST source catalog and its segmentation map and the MUSE datacube. We ran ODHIN for three HST catalogs: our primary catalog R15 (Rafelski et al. 2015) but also the CANDELS v2 (Whitaker et al. 2019) and 3D-HST (Skelton et al. 2014) catalogs. Although these extractions use the same HST broadband images (XDF, Illingworth et al. 2013), they are based on different segmentation maps and may thus provide different results.

An example of source de-blending is given in Fig. 17. Two close sources can be identified in the HST image (upper left panel) separated by 0.4 arcsec, but they are blended at the MUSE

[6] There is a factor of ≈5 in spatial resolution and ≈1000 in spectral resolution between MUSE and HST.

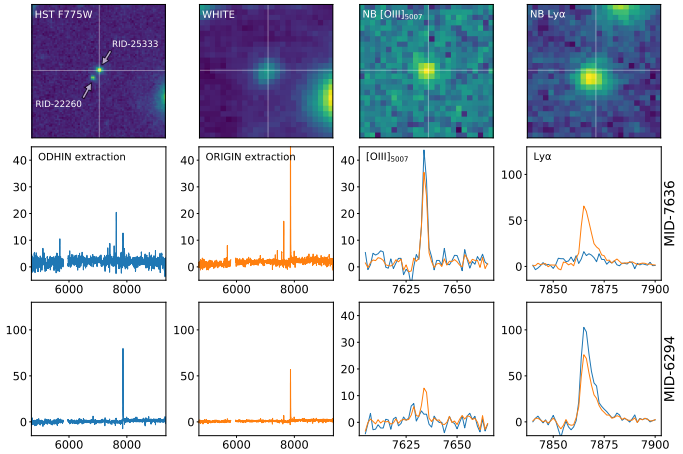


Fig. 17. Example of ODHIN de-blending of two close sources MID-7636 (RID-25333), a $z = 0.5$ star forming galaxy, and MID-6294 (RID-22260), a $z = 5.5$ Lyα emitter. The separation of the sources is 0.4 arcsec. The top row of panels shows, from left to right: the HST $F775W$, MUSE white-light, [OIII]5007 Å and Lyα narrow band images (5×5 arcsec^2 size). The central and bottom panels display the spectra derived from ODHIN (in blue) and ORIGIN extraction (in orange) for each source. Fluxes and observed wavelengths are expressed in units of 10^{-20} erg s^{-1} cm^{-2} Å^{-1} and Å, respectively.

spatial resolution as shown in the MUSE white light image (second upper panel). The brightest source ($F775W$ AB 27.2) MID-7636 (RID-25333) is a nearby star-forming galaxy at $z = 0.5$, with multiple emission lines: for example, [O II]$\lambda\lambda$3726,3729, [O III]$\lambda\lambda$4959, Hβ and [O III]$\lambda\lambda$5007(central row of Fig. 17). The second source ($F775W$ AB 27.8) MID-6294 (RID-22260) is a $z = 5.5$ Lyα emitter with a strongly asymmetric Lyα line (last row of Fig. 17). The fact that the narrow-band image of the Lyα is slightly offset from the [O III]$\lambda\lambda$5007 narrow-band image indicates that the Lyα emission is indeed related to RID-22260 and not RID-25333. However, the overlap is too large to de-blend it from the narrowband segmentation at the spatial resolution of MUSE. Therefore, for each source, the spectra provided by the ORIGIN extraction (shown in orange in Fig. 17) show strong contamination due to the nearby source. In contrast, the ODHIN extraction (shown in blue) is able to de-blend the contributions from each source.

One can see that, despite its intrinsic limitations[7], ODHIN is capable of de-blending closed sources with good accuracy. We note, however, that in the presence of emission that extends much beyond the HST broadband, the method will miss part of the flux (see Fig. 18 for an example). Furthermore, the method is by construction blind to any source undetected by HST.

5.4. Redshift determination with pyMarZ

The software pyMarZ is a Python implementation of the redshift fitting software MarZ (Hinton et al. 2016) originally developed for multi-object fiber spectroscopy with OzDES (Lidman et al. 2020). MarZ itself is based on a modified version of the cross-correlation algorithm AUTOZ (Baldry et al. 2014). It is written in JavaScript with a dedicated user interface. The development of the Python version was motivated by the need to integrate it more easily into the full data analysis chain where all parts are written in Python. We have also developed our own interface SourceInspector (Sect. 5.7.2), which is run separately from pyMarZ.

Compared to the version used in DR1 (Inami et al. 2017), the main change, apart from the Python language, is the update of the spectral templates based on a set of higher S/N spectra of selected sources in UDF-10 and MOSAIC datacubes. We tested it on a selection of UDF-10 high confidence isolated sources and found that the first ranked solution returned by pyMarZ is correct 84% of the time. We note that the probability of finding the correct solution among the first two pyMarZ solutions is 95%. In the previous version used for DR1, the probability of finding the correct solution was only 71% and 87% for the first and first two returned solutions, respectively. Although the improvement is significant, the selection of the correct redshift solution requires additional information and cannot be performed automatically based only on the top ranked pyMarZ solution. For this reason, we retain the top five pyMarZ solutions for the manual inspections presented in Sect. 5.7. The software, its documentation and the templates used for this work are made publicly available (Appendix A.5).

5.5. Line flux measurements with pyPlatefit

In the past years, the increasing number of galaxy surveys with wide field or highly multiplexed IFUs (e.g., SAMI, MANGA, and MUSE) has been accompanied by the development of software tools to efficiently fit emission and/or absorption lines in spectra or datacubes (e.g., pPXF Cappellari 2017; Sarzi et al. 2017, ALFA Wesson 2016, LZIFU Ho et al. 2016, pipe3D Sánchez et al. 2021). Most of these tools have been developed for the spatially resolved study of nearby galaxies and are not optimized for the population of medium to high redshift galaxies present in our observations. We have therefore developed a specific fitting code pyPlatefit, inspired by the Platefit IDL code developed for the SDSS survey (Tremonti et al. 2004; Brinchmann et al. 2004).

pyPlatefit is a python module to perform emission line fitting of astronomical spectra. Although it was specifically developed for this data release[8], it should work for spectra delivered by other instruments, provided that the input spectrum is in MPDAF format (Piqueras et al. 2019). The program takes as input a spectrum and an input redshift and returns a list of emission line parameters: for example, flux, velocity dispersion, equivalent width and their errors. The program performs a continuum fit, using a simple stellar population model (Bruzual & Charlot 2003, see Brinchmann et al. 2013), and then fits the emission lines after subtracting the continuum. In addition, the program can also perform a basic fit of the absorption lines. Lines are grouped into families, each of which is fitted independently. All lines belonging to the same family share the same velocity (or redshift) and velocity dispersion. The default lines table (Appendix D) has 3 different families: the Balmer series, the non-Balmer emission lines and the ISM absorption lines. In addition, some major resonant lines like Lyα or Mg II are fitted separately as distinct families.

All emission and absorption lines are modeled as Gaussian with 3 parameters: central wavelength (λ_0), wavelength dispersion (σ) and flux (F_0). The only exception is the Lyα line, which is fitted as a skewed Gaussian (Birnbaum 1950), with an additional unit-less parameter (γ) measuring the asymmetry of the

[7] We assume that the broadband light distribution is representative of any light emission and do not allow for spectral variation within a source.

[8] We used the Platefit IDL code for the DR1 release.

line:

$$F(\lambda) = F_0 \left[1 + \mathrm{erf}\left(\gamma \frac{\lambda - \lambda_0}{\sqrt{2}\sigma}\right)\right] \exp\left(-\frac{(\lambda - \lambda_0)^2}{2\sigma^2}\right). \quad (4)$$

Optionally, a double peaked Lyα line profile can be fitted as the sum of two skewed Gaussians. The program uses the front-end lmfit minimization package[9] (Newville et al. 2014) based on scipy optimize nonlinear least-squares minimization (Virtanen et al. 2020). The default minimization routine is the trust region reflective least squares algorithm, which is more robust and gives a better error estimate than the classical Levenberg–Marquardt algorithm. Other minimization algorithms can be selected (see the lmfit documentation for details). The Markov chain Monte Carlo emcee package (Foreman-Mackey et al. 2013) is also used to improve error computation. Information about the public released version of the software and its documentation is given in Appendices A.5 and D.

5.6. Narrowband measurements

For each redshift solution, we compute the narrowband images associated with each detected emission or absorption line, as provided by pyPlatefit. These narrowband images are important to confirm the detection of the line and to verify its origin. For example, spectral pollution due to nearby sources will be more easily spotted by looking at the location of the corresponding narrowband image (e.g., Fig. 17). Another example is the case of a low S/N spectrum when a noise pattern due to sky subtraction residuals is fitted by pyPlatefit because its spectral location matches the expected wavelength of a line. In this case, the narrowband image should not present a coherent peak at the source location, thus helping discard the line measurement.

The narrowband process selects each emission and absorption line with $(S/N)_{\rm line} > 2$ plus all major[10] emission lines, regardless of their S/N. The continuum is then subtracted from the datacube using a median filter size of 101 Å total width (81 pixels). A narrowband image is finally obtained using a simple summation of the continuum-subtracted datacube over the wavelength range defined by the fitted center of the line and a spectral width of $\pm 2 \times FWHM$, where FWHM is the fitted line width at half maximum. For emission line doublets like [O II]$\lambda\lambda$3726,3729, the wavelength range is extended to cover both lines. For each narrowband image with $(S/N)_{\rm line} > 2$, we perform a SExtractor segmentation and derive its center, flux and signal-to-noise ratio $(S/N)_{nb}$.

We also derived composite emission and absorption narrowband images obtained by optimally adding all emission or absorption narrow bands. The process starts with the highest $(S/N)_{nb}$ of the emission or absorption narrow bands and adds additional narrowband images until the $(S/N)_{nb}$ of the composite narrowband stops increasing.

5.7. Visual inspections and reconciliations

The material produced in the previous steps, that is, for each candidate source, the five redshift solutions provided by pyMarZ as well as their emission and absorption lines fits and their associated narrow bands, must be organized and evaluated by a group of experts. Once the conflicts have been resolved during the reconciliation meetings, one can proceed to the creation of the final catalog by assigning a unique identifier to the detected sources. The details of the process are presented in the following sections. The figures and statistics quoted are given for the MOSAIC largest data set, but the process was similar for the UDF-10 and MXDF data sets.

5.7.1. Inspection sample

The MOSAIC parent sample is composed of 10 450 sources divided into 7977 ODHIN sources derived from the R15 catalog (Sect. 5.3) and 2473 ORIGIN sources (Sect. 5.2). Previous investigation has shown that all DR1 ORIGIN sources to which a secure redshift could be assigned have a purity greater than 0.9. We therefore restrict the ORIGIN sample to the 1946 sources with purity greater than 0.9.

The ORIGIN subsample contains all line sources that have at least one detectable emission line, regardless of their continuum flux. The remaining sources of interest are those whose continuum flux is bright enough to identify other spectral features (e.g., absorption lines, breaks). After some trial, we adopted a continuum S/N cut[11] of 0.8×10^{-20} erg s^{-1} cm^{-2} Å^{-1} per spectral pixel, which corresponds roughly to an AB F775W magnitude of 27.6. This reduces the ODHIN sample to 782 sources[12].

We further reduced the MOSAIC inspection sample by removing all sources located in the UDF-10 area, taking a margin of 1 arcsec. The final MOSAIC inspection sample is composed of 2412 sources: 1727 ORIGIN and 685 ODHIN sources. This sample was randomly distributed to four groups of three experts. In addition, a randomly selected common subsample of 50 sources was given to the four groups. This control sample will be used to assess the homogeneity of the classification between the groups. We proceeded likewise for the UDF-10 and MXDF data sets and produced final (i.e., ODHIN, ORIGIN, and control) inspection samples of 1234 and 1599 sources, respectively.

5.7.2. SourceInspector

The inspection package provided to each expert consists of a small database, a set of interactive html files and the SourceInspector software. SourceInspector is a PyQt tool that allows the expert to display each candidate source, to select the redshift solution or to provide a new one, to give a confidence to the redshift assignment and to match the source to other sources (e.g., ORIGIN to ODHIN) and to HST catalogs. All operations are performed locally. An example of the SourceInspector interface is shown in Fig. 19. The inspection results are exported at the end of the process to be used later for the reconciliation step.

5.7.3. Redshift confidence

The expert must assign a confidence to the redshift solution. This redshift confidence (called ZCONF in the tables) is defined as follows.

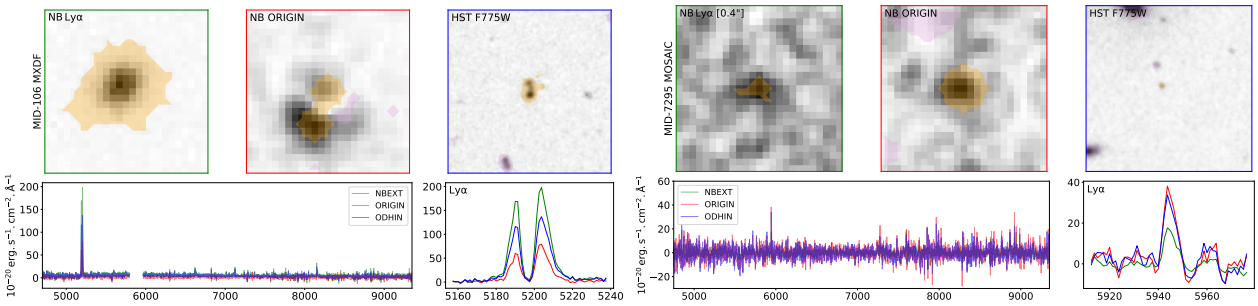
A confidence level of 0 means no redshift solution can be found. It should be noted that this is decoupled from the detection status. An emission line can be well detected with high

[9] lmfit.github.io/lmfit-py

[10] E.g., Lyα or [O II]$\lambda\lambda$3726,3729; see pyPlatefit input line table in Appendix D.1.

[11] The continuum S/N is computed as the maximum of the S/N spectrum after running a median filter of 100 pixels width to remove the impact of emission lines.

[12] Our experience has taught us that galaxies with fainter continuum cannot be attributed an aborption line redshift and that galaxies with emission lines are already detected with ORIGIN.



(a) (b)

Fig. 18. Example of sources where spectral extraction gives different results. Three different extraction schemes are shown for the Lyα emitters MID-106 (*a*) and MID-7295 (*b*). For each subfigure we show the full spectrum (*lower left panel*) and a zoomed region around the Lyα line (*lower right panel*). In green we display the spectrum derived from the weighted extraction using the narrowband Lyα weighted map (NBEXT) and its segmentation (*top left panel*), in red the ORIGIN weighted extraction (*top central panel*) and in blue the ODHIN extraction based on the R15 segmentation map (*upper right panel*). The colored area in the left and central top panels indicate the source (orange) and masked objects (magenta) segmentation maps. In the case of the low S/N source MID-7295 (b), the spectra have been smoothed with a 5 pixels width kernel and a Gaussian filter of 0.4 arcsec FWHM has been applied to the Lyα narrowband image (*top left panel* of subfigure b). Observed wavelengths are in Å units.

purity by ORIGIN, even though no redshift solution can be found.

A confidence level of 1 indicates a low confidence. A tentative redshift solution has been found. Reasons for assigning ZCONF = 1 include low S/N of the lines (e.g., $S/N_{\rm line} < 3$), poor fit, existence of other valid redshift solutions, noisy associated narrow bands, additional lines present in the spectrum with reasonable S/N but unexplained by the proposed redshift solution.

A confidence level of 2 indicates a good confidence. For non Lyα emitters (i.e., $z < 2.9$) sources, we assign this confidence level to sources with multiple lines detected with good S/N ($S/N_{\rm line} > 5$). For example, a resolved [O II]$\lambda\lambda$3726,3729 doublet with a clear narrowband would be sufficient to assign ZCONF = 2. For Lyα emitters, the requirement is to have a Lyα line with good S/N (i.e., $S/N_{\rm line} > 5$) and a width and asymmetry compatible with Lyα line shapes.

Finally, a confidence level of 3 indicates a high confidence. For non-Lyα emitters, the criteria are similar to the ZCONF = 2 requirements, but we expect more lines and higher S/N, as well as high S/N narrowband images. For Lyα emitters, if there is no other line than Lyα, we require the Lyα line to achieve high S/N ($S/N_{\rm line} > 7$) with the expected line shape: that is a pronounced red asymmetrical line profile[13] ($\gamma > 2$) and/or a blue bump or double peaked line profile[14].

In the case of ORIGIN detections and if the source can be matched to an HST counterpart, it adds confidence to the detection. If, in addition, the R15 photometric redshift is reliable and matches well with the MUSE redshift (see Eq. (5)), this adds confidence to the redshift assignment and could lead to a higher ZCONF than in the case where the ORIGIN source is not matched with an HST source. On the other hand, photometric redshifts for faint sources are often not very reliable and thus disagreement between the MUSE and photometric redshifts is not considered a strong negative constraint.

We note that in the case of low S/N spectra with a single emission line, we are not always able to determine the nature of that line with a high degree of certainty. In many cases, the combination of line shape and photometric redshift argues in favor of Lyα, but it is desirable to have a handle on the amount of misclassifications in this subsample. We focus on [O II] and Lyα emitters here as other emission lines are much less likely because of the small accessible volume.

In order to estimate the expected fraction of Lyα and [O II] emitters as a function of magnitude, we need LFs and rest equivalent-width distribution functions (EWDFs) for both. We take the Lyα LF from Herenz et al. (2019) and the Lyα EWDF from Hashimoto et al. (2017). For [O II] we adopted the evolving LF from Saito et al. (2020), which agrees fairly well with our UDF data. We also need an EWDF for [O II], which we prefer to derive from our high quality (ZCONF $\geq$ 2) data. To do that we adopt the point-source selection function from Herenz et al. (2019), which we derived for Lyα emitters but which reproduces the minimum [O II] flux as a function of wavelength fairly well for the present sample when the selection function is adjusted for difference in exposure time by $\sqrt{t}$. We then use a simple Vmax estimator for the EWDF after verifying that the [O II] line luminosity and equivalent width are approximately independent, in good agreement with previous work (Blanton & Lin 2000; Ciardullo et al. 2013). The distribution is found to be well fitted by a Gaussian in log EW(Å) with mean = 1.61 and sigma = 0.17.

Taking these together with the assumption that the spectrum of our galaxies is approximately flat in flux density (f_ν) we can calculate the expected number of [O II] and Lyα emitters with observed magnitude $F775W$ fainter than a particular value. When we do that, we find that at magnitudes $F775W > 28.5$ the number of Lyα emitters is expected to be 10–100 times that of the [O II] emitters at all redshifts of relevance down to a flux limit of 10^{-18} erg s^{-1} cm^{-2} Å^{-1}. Thus, we conclude that the likelihood of misclassified lines is less than 1/10 for galaxies fainter than 28.5.

5.7.4. Cross identifications

For ORIGIN sources, the expert must also identify the possible HST counterparts and, likewise, must find the possible ORIGIN counterparts for ODHIN sources. Crossmatching ORIGIN and HST sources can be difficult when the field is overcrowded or if

[13] γ is the asymmetric Gaussian parameter used by pyPlatefit (Sect. 5.5).
[14] Double peaked Lyα profiles are easily differentiated from [O II]$\lambda\lambda$3726,3729 or C III]$\lambda\lambda$1907,1909 using the line spacings and widths.

the ORIGIN location is offset with respect to HST (see an example in Fig. 20). The use of the photometric redshift can help with the HST association, but not always. At a later stage we perform an evaluation of the HST crossmatch confidence (Sect. 5.8.2).

Occasionally a source is visually identified in one of the HST images at the ORIGIN location, but is not present in the HST R15 catalog. This might occur when the source has been incorrectly merged with a nearby source by SExtractor, but in other cases the source is simply too faint and escapes detection. We identify these cases by setting "missed, faint, or undetect" in the corresponding HST crossmatch flag (see Sect. 6.2).

The ORIGIN process identifies individual emission lines and merges them into a single source when they are close (see Sect. 3.5 in Mary et al. 2020). However, sometimes the process fails to merge emission lines when the source is very extended, or, on the contrary, it merges lines that belong to two different sources (i.e., superpositions). The expert is then responsible for evaluating if the ORIGIN sources should be split or merged with another ORIGIN source. Such actions are then performed in the final step (Sect. 5.8).

Cross-identification with the DR1 catalog is also performed during this evaluation process. When a source is found at the location of a DR1 source but is assigned a different redshift, we mark it as a DR1 unmatched source. On the other hand, if only the redshift confidence is different, we mark the source as DR1 matched.

5.7.5. Reconciliation

When all experts in the group have completed their evaluation, the group leader performs the reconciliation process. All sources that disagree on any of the assigned items (i.e., redshift value, redshift confidence, HST, DR1, or ORIGIN crossmatch) are flagged for the reconciliation meeting. In addition, any source flagged with "revisit" by an expert is added to the review list. The only exception to this process is when the only discrepancy is due to redshift confidences between two and three, in which case the source is not flagged for reconciliation and its confidence is set to the average confidence rounded to the nearest integer.

An example of the statistics returned by SourceInspector in automatic checking mode is given in Table 5. Disagreements are then discussed and resolved in a face-to-face reconciliation meeting with the help of the reconciliation mode of SourceInspector. On average, between 30% and 40% of the original sample had to be revisited.

Most of the sources to be re-examined are weak, single-line sources for which experts have a different assessment of the reality of the measured redshift. For example, some less experienced experts would be inclined to identify a Lyα emitter for a low S/N line, even if there is no additional information to support this assumption. In most cases, however, when reconciling, the redshift solution will be ruled out.

5.7.6. Control sample

A control sample of 50 randomly selected MOSAIC sources was given to the same four experts groups. The same experiment was done for the MXDF evaluation. After reconciliation, the control sample was analyzed for remaining disagreements. The 50 sources consisted of 36 ORIGIN and 14 ODHIN sources. All groups agreed on the redshift assessments. Eight sources did not obtain a redshift solution. Of the 42 sources with a redshift solution, nine sources were credited with different redshift confidence: five sources were ranked with low and good/high confidence and four sources were ranked with good and high confidences. Four sources were also crossmatched to different HST sources.

Table 5. Example of automatic reconciliation.

MOSAIC group 1 reconciliation	
Total number of sources	790
Sources autochecked with ZCONF = 0	57
Sources autochecked with ZCONF > 0	430
Sources to be manually reconcilied	303
With revisit flag	24
With discrepant redshift	185
With discrepant redshift confidence	45
With discrepant matching information	36
With discrepant split or merge info	13

Notes. "autochecked" means that all experts agree.

The results for this control sample demonstrate that we achieved overall excellent agreement among all experts, with no discrepancies in redshift solutions. Redshift confidence is more subjective and we do identify 20% of disagreements between groups. We note that we do not consider the difference between confidence two and three to be critical. Nevertheless, we have 10% of sources that were considered low confidence by some groups and good or high confidence by others.

Similar results were obtained on a control sample of 49 MXDF sources given to three different groups. All groups agreed on the redshifts, but there were ten sources with different confidence assignments. In addition, some groups assigned a low confidence level to 10 sources, while the others classified them as having no redshift (confidence 0).

As we discuss in the next section, a review of all sources with redshifts is performed as a final step by a single expert. This helps to homogenize the results, but given the amount and complexity of information that goes into the confidence evaluation (Sect. 5.7.3), it currently seems difficult to have a more objective criterion for redshift confidence assignment.

5.8. Creation of final data products

In this final step, we create the final catalog by giving a unique MUSE identifier to all sources with an assigned redshift. We start by splitting the catalog in ORIGIN only, ODHIN only and crossmatched ORIGIN-ODHIN sources. All sources already matched to the DR1 catalog keep the same MUSE identifier, all others are given a unique new identifier.

Split and merge of ORIGIN sources defined during the inspection step was implemented. A final review of all sources with redshift was then completed. During this review the reference extraction (Sect. 5.8.1) and the reference center (Sect. 5.6) were chosen. Emission or absorption line fits were fine-tuned with pyPlatefit when needed and double asymmetric Lyα fits implemented for double peaked profiles. The computed narrowband S/N_{nb} was used as an additional guide for the redshift confidence (Fig. 21), which was updated when requested.

5.8.1. Selection of reference extraction

For each source, we derive the spectra from the MUSE datacube using different extraction schemes (e.g., ORIGIN or ODHIN).

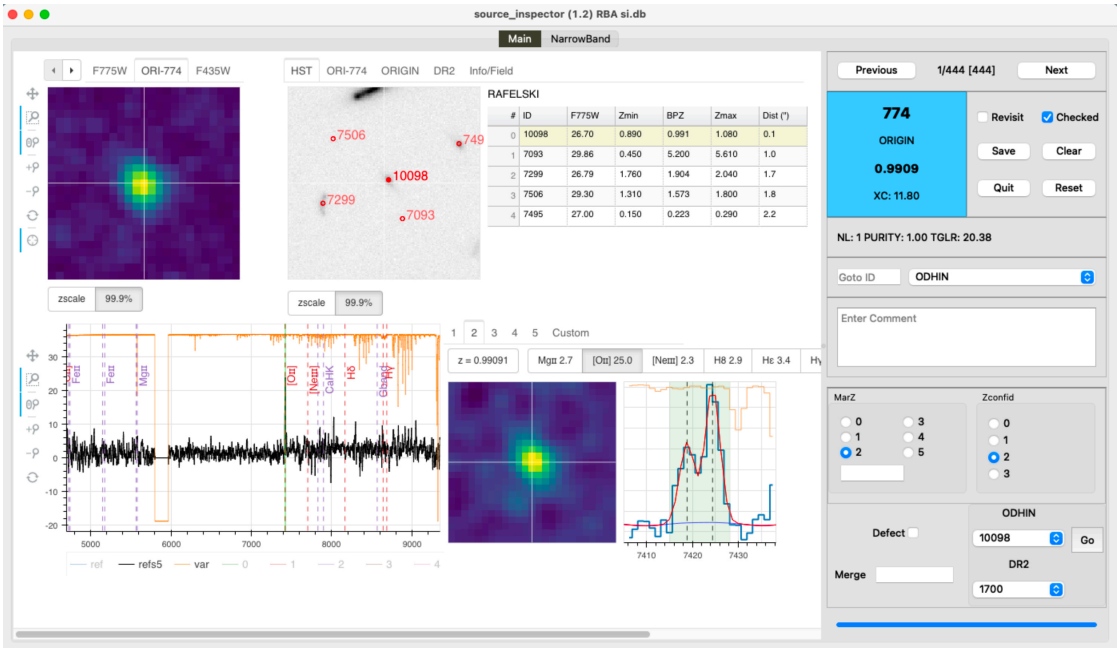


Fig. 19. SourceInspector interface in evaluation mode. The left side of the interface displays the interactive html file for the current source. It shows the five redshift solutions proposed by pyMarZ plus their corresponding pyPlatefit line fits and associated narrow bands in the bottom panels. The html upper panels display MUSE and HST images, and the locations of neighboring sources from HST catalogs. The right side of the interface allows the user to navigate in the source list, select the redshift solution or assign a new redshift, assign confidence and provide crossmatch information.

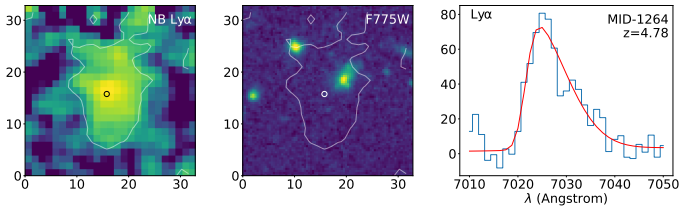


Fig. 20. Example of a Lyα emitter with extended Lyα emission, offseted with respect to its potential HST counterpart. The two first panels display the Lyα narrowband and HST $F775W$ images. The units of the image axes are physical kpc. The right panel shows the Lyα profile and its fit with a skewed Gaussian. Observed wavelength and flux units are respectively Å and 10^{-20} erg s^{-1} cm^{-2} Å^{-1}.

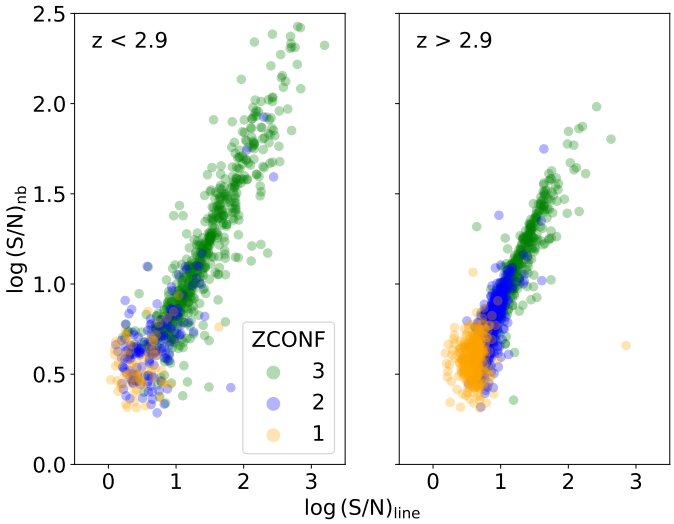


Fig. 21. S/N estimates as a function of redshift confidence. The log of the S/N emission or absorption composite narrowband is shown as a function of the log of the leading line S/N derived from pyPlatefit. Low (1), good (2) and high (3) redshift confidences are shown in respectively orange, blue and green colors, while low-z ($z < 2.9$) and high-z ($z > 2.9$) sources are displayed in the left and right panels, respectively.

There is no single way to extract the signal: for example, the stellar continuum is not necessarily superimposed on the location of the flux peak of the emission lines, and its spatial extend is often different. The most obvious cases are those of Lyα emitters, which show extended Lyα emission that is on average ten times larger than their continuum counterparts (e.g., Wisotzki et al. 2016; Leclercq et al. 2017). In this case, a spectrum based on the continuum surface brightness will miss a large faction of the Lyα flux, and vice versa, a spectrum optimally extracted for the Lyα emission will be suboptimal for the continuum S/N.

In the ORIGIN extraction scheme (Sect. 5.2), the pseudo-narrowband image is derived from an improved continuum subtracted datacube using an iterative principal component (PCA) scheme and thus has the advantage of being mostly free from continuum residuals. It also has a high S/N, as it results from a process similar to optimal filtering (see Sect. 3.2 of Mary et al. 2020). However, in the case of a very bright emission line, the PCA process may produce artifacts by removing part of emission line. In this case, we perform an optimal extraction based on the narrowband composite image and its segmentation (Sect. 5.6). This narrowband extraction method is called NBEXT in the rest of the document.

We illustrate the difference between the two methods with two examples in Fig. 18. Figure 18a shows the results of the ORIGIN and NBEXT weighted extractions for a bright Lyα emitter in which this effect is particularly strong. It can be seen that in the case of the ORIGIN pseudo narrowband image (upper central panel), the bright central Lyα emission has been partially

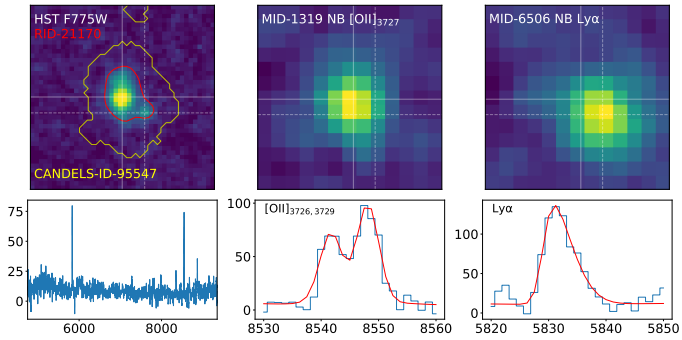


Fig. 22. Example of a MUSE source missed in HST catalogs. The upper left panel display the HST $F775W$ image ($2\farcs5 \times 2\farcs5$) at the location of RID-21170. The red line shows the R15 segmentation contour, while the yellow line displays the CANDELS v2 segmentation contour of the corresponding source (ID 95547) in Whitaker et al. (2019). Although all HST catalogs identify a single source, MUSE unambiguously identifies two sources: MID-1319, an [O II] emitter at $z = 1.292$ (bottom central panel) and MID-6506 (bottom right panel), a Lyα emitter at $z = 3.798$. The MUSE reference spectrum is shown in the bottom left panel. From the corresponding [O II] and Lyα narrowband images, one can see that the MID-1319 [O II] emitter matches the brighter part of the HST image, while the MID-6506 Lyα emitter coincides with the southeast secondary HST peak. Observed wavelength and flux units are respectively Å and 10^{-20} erg s^{-1} cm^{-2} Å^{-1}.

removed by the PCA, leaving only the outskirts. As a result, only a small portion of the Lyα flux is recovered by the weighted extraction (shown in red in the lower panels). We note that the ODHIN extraction (Sect. 5.3) based on the R15 segmentation map (blue line in bottom panels) is more efficient but still misses some Lyα flux compared to the narrow band Lyα extraction (green line in bottom panels). The second example shown in Fig. 18b presents the same extraction methods applied to a faint Lyα emitter. We observe that the narrowband image is too noisy to capture the Lyα flux and that the ORIGIN weighted extraction performs better. We stress that in most cases, however, the different extractions methods give very similar results.

For each source, we assigned a reference extraction. In general, we favored ORIGIN or NBEXT for Lyα emitters and ODHIN for low-z sources when the continuum was not too faint. In some cases where the contamination is strong, ODHIN was preferred for Lyα emitters, even if it misses some Lyα flux. The selection was made on a case-by-case basis. However, the reference extraction may not be the best for a given science case and so we provide all extractions in the delivered data products (Appendix A.3).

5.8.2. Crossmatch with HST source catalogs

During the inspection step (Sect. 5.7) the ORIGIN sources were matched to the HST R15 catalog. However, since the publication of Rafelski et al. (2015), new HST source catalogs in the CANDELS area have been published. These catalogs are based on an improved data reduction of HST photometry (Whitaker et al. 2019), additional HST GRISM information as in 3D-HST (Skelton et al. 2014), or additional ground-based photometric bands (Merlin et al. 2021). In addition, the detection image[15] and the segmentation input parameters were different, resulting in different segmentation maps. The biggest difference is for closely packed faint galaxies, which can be merged in a single source in some catalogs, whereas they are split into subcomponents in others. In our investigation, we found many cases where the HST source should be split into two different sources (e.g., Fig. 22) or, on the contrary, where two close HST sources should be considered as a single one.

In this final step, we therefore crossmatched our results with the CANDELS v2 (Whitaker et al. 2019), 3D-HST (Skelton et al. 2014) and ASTRODEEP (Merlin et al. 2021) catalogs. These catalogs and their corresponding segmentation maps were also used to produce the corresponding ODHIN extraction (Sect. 5.3).

For all sources, we calculated an HST environment index to quantify whether the source can be considered spatially isolated or has close or distant neighbors. This index is the percentage of the area covered by neighboring sources measured in the R15 segmentation map[16] in a given aperture centered on the MUSE source location. We used aperture radii of 0.5 and 1.0 arcsec to define the close and distant environment index, respectively. At the MUSE spatial resolution, an object with a nonzero close environment index will overlap with its neighbors, at least partially, while an object with zero close environment index and nonzero distant environment index will be resolved by MUSE. A source is considered isolated if its distant environment index is zero.

The presence or absence of close or distant neighbors is coded in the IFLAG column of the main catalog (Appendix A.2). It is defined as follows: (i) Sources with a close HST environment index of zero and a distant environment index of less than 5% are considered isolated with respect to HST. If, in addition, there is no other non-HST DR2 sources within 1 arcsec radius, they are defined as fully isolated. Their isolation indicator ICONF is set to three. (ii) Sources with ICONF = 2 have only distant neighbors, either from HST or from DR2 non-HST sources. (iii) All sources with at least one close HST or non-HST neighbor have ICONF set to one.

To assess the quality of the provided matching information, we assigned a matching confidence indicator (hereafter MCONF) to each source. Its values are given in Table 6 and range from 0 (ambiguous) to 3 (high confidence). MCONF combines the neighborhood information as defined above with additional redshift and narrowband offset information.

An HST source is matched in redshift with a MUSE source when its photometric redshift is considered reliable and the MUSE redshift is within the 95% confidence interval, that is,

$$\frac{z_{\rm max} - z_{\rm min}}{1 + z_{\rm best}} < 0.22 \qquad z_{\rm min} < z_{\rm MUSE} < z_{\rm max}, \tag{5}$$

where $z_{\rm best}, z_{\rm min}, z_{\rm max}$ are respectively the photometric redshift solution and its bounds given by the 95% error probability. In the case of 3D-HST spectroscopic redshifts we use $|z_{\rm MUSE} - z_{\rm spec}|/(1 + z_{\rm MUSE}) < 0.01$.

The last criterion is the offset between the MUSE absorption or emission composite narrowband (Sect. 5.6) from the center of the matched HST source. This offset is compared to the size of the HST source and is considered acceptable when it satisfies the following condition: offset $<$ max $(2 \times$ size$, 0.4'')$ where size is the radius of the circular HST aperture (in arcsec) enclosing half of the total flux, as given in the catalog.

[15] The Rafelski et al. (2015) detection image is the average of the four optical ACS images with the four WFC3 images. Whitaker et al. (2019) perform a noise-equalized combination of one ACS (F850LP) with three WFC3 bands ($F125W$, $F140W$, $F160W$). The Merlin et al. (2021) catalog is based on the Guo et al. (2013) catalog, which uses the $F160W$ as detection image.

[16] We use the CANDELS v2 segmentation map when the HST source is not present in R15.

Table 6. Isolation flag (IFLAG) and HST crossmatch confidence (MCONF) definition and counts.

IFLAG	Description	Counts	MCONF	Description	Count
			0	Ambiguous	8
1	Close neighbors	679	1	Low confidence	61
2	Distant neighbors	154	2	Good confidence	644
3	Isolated	1388	3	High confidence	1508

All isolated sources that do not have an HST counterpart in any catalog are assigned a high matching confidence (MCONF = 3). There are 99 sources in this category. All isolated sources matched to a single HST counterpart are also assigned a MCONF value of three if all reliable photometric redshifts of the two used HST catalogs can be matched to the MUSE redshift and if there is no significant offset between the HST source and the narrowband location. If the offset is too large and/or there is at least one reliable photo-z discrepant with respect to MUSE, MCONF is set to two.

When sources cannot be considered isolated, in addition to the previous criteria, we search if any neighbor, not already assigned to any MUSE sources, has a reliable photo-z that could be matched to the MUSE redshift. If so, MCONF is lowered to two or one, depending on if the neighboring source is distant or close to the MUSE source. Finally, sources with multiple HST counterparts assigned during the manual evaluation are assigned a zero value of MCONF. There are only eight sources in this category.

The counts of the matching confidence are given in Table 6. Most MUSE sources have a high HST matching confidence (68%), either because the MUSE source can be assigned to a unique identifier in the HST catalogs (53%) or because they are, instead, isolated and have no HST counterpart in the catalog (15%).

5.8.3. Systemic redshift estimation for Lyα emitters

For most Lyα emitters the measured redshift is based on the peak of the Lyα line. However, due to the resonant scattering properties of Lyα photons in the interstellar medium, the Lyα redshift is systematically different from the systemic redshift (e.g., Shapley et al. 2003; McLinden et al. 2011; Rakic et al. 2011; Song et al. 2014). Typical velocity offsets are $\approx$200 km s^{-1} for Lyα emitters with larger values ($\approx$500 km s^{-1}) for Lyman-break galaxies (Shibuya et al. 2014; Muzahid et al. 2020).

We used the empirical recipes provided by Verhamme et al. (2018) to estimate the velocity offset between the Lyα and systemic redshift. In their paper, Verhamme et al. (2018) supply two types of corrections. The first correction is based on the separation between the two peaks of the Lyα line (i.e., the blue bump separation). The center of the two peaks is a good estimate of the systemic redshift with a rest-frame scatter of 53 km s^{-1}. The asymmetric double peaked line profile used in pyPlatefit (Sect. 5.5) provides this value directly. We simply add the 53 km s^{-1} standard deviation derived by Verhamme et al. (2018) to the redshift error budget in quadrature.

The second type of correction is based on the FWHM of the Lyα line profile. The rest-frame velocity offset (ΔV in km s^{-1}) is inferred from the following linear relation with the rest frame measured FWHM in km s^{-1} (uncorrected for the MUSE LSF): $\Delta V = 0.9 \times FWHM - 34$ km s^{-1}. It has a larger scatter of 72 km s^{-1} and therefore we use the first correction when a double Lyα peak was detected and successfully fitted.

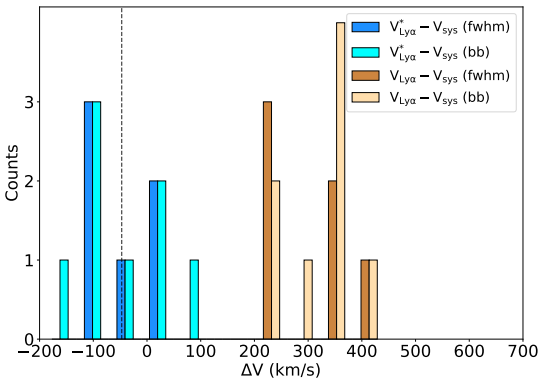


Fig. 23. Check of systemic redshift correction for a subsample of DR2 Lyα emitters with systemic redshifts measurements. The histogram of the rest-frame velocity difference between the measured Lyα red peak location ($V_{Ly\alpha}$) and the systemic velocity (V_{sys}) is shown in orange. The same histogram after systemic velocity correction ($V^*_{Ly\alpha}$) is shown in blue. The type of correction, that is, blue bump separation (bb) or Lyα line FWHM (fwhm), are identified in respectively darker and lighter colors. The dashed vertical line displays the median value of the velocity offset distribution after correction (-47 km s^{-1}).

We checked the accuracy of the correction by selecting the few Lyα emitters with high redshift confidence, high S/N Lyα line and additional high S/N nonresonant emission or absorption lines. This selection resulted in a sample of 14 Lyα emitters: six with single Lyα line profiles and eight with double peaked profiles. Systemic velocities were measured from forbidden emission lines (mostly C III]$\lambda\lambda$1907,1909) or strong absorption lines (e.g., Boogaard et al. 2021). The comparison of the velocity offsets before and after correction is presented in Fig. 23. We can see that the systematic velocity offsets of about 300 km s^{-1} are reduced to a few tens of km s^{-1}. The measured scatter (77 km s^{-1} RMS) is comparable to the values measured by Verhamme et al. (2018).

5.8.4. Final catalogs

In Table 7 we display the source count by extraction method and data set. We find that the majority of sources (67%, 1499/2221) detected with ORIGIN have also an HST counterpart, but we also observe that the fraction of ORIGIN sources without HST counterparts increases with depth, from 13% (202/1505) in the MOSAIC to 29% (203/691) in the MXDF data sets.

When a source is detected in multiple data sets (e.g., UDF-10 and MXDF), we retain the data set with the greatest depth at the source location. The only exception is MID-171, where we selected the UDF-10 instead of MXDF data set, despite its lower depth (31 versus 136 h). The reason for this is the Lyα line that falls into the MXDF masked wavelength range because of the sodium laser guide star notch filter (Fig. 1). When available, the

Table 7. Source counts with respect to detection method.

Data set	ORI [(a)]	ODH [(b)]	ORI & ODH	Total
MOSAIC	202	220	1083	1505
UDF-10	78	48	225	351
MXDF	203	78	410	691
DR2	428	294	1499	2221

Notes. [(a)]ORIGIN, [(b)]ODHIN.

other data sets are also recorded in the delivered data product (Appendix A.3).

The final catalog contains 2221 sources with redshifts. The statistics and properties of these sources are discussed in the following sections.

6. Results

6.1. Redshifts

One of the most visible results of the survey is the number of galaxies for which we were able to assign high quality redshifts. A detailed presentation is given in the following subsections.

6.1.1. Redshift distribution and confidence

The distribution of redshifts is shown in Fig. 24. Table 8 presents the breakdown of the 2221 redshifts in five categories. Given the small area (8.3 arcmin^2) and the Galactic pole location, we identified only eight stars, mostly M types. Likewise, the number of nearby galaxies ($z < 0.25$) is limited (25) in this small volume (280 cMpc^3). For this latter category, dedicated wide field spectroscopic surveys (e.g., SDSS Blanton et al. 2017 and MANGA Bundy et al. 2015) are much more efficient. Nevertheless, for these few bright galaxies (median AB $F775W$ of 23.3), we obtain spatially resolved spectroscopy with exquisite S/N up to large galactocentric distance. An example is given in Fig. 25 row a.

In the redshift range 0.25–1.5, the [O II]$\lambda\lambda$3726,3729 doublet falls in the MUSE wavelength range. At the MUSE spectral resolution ($R \approx 3000$), the [O II] doublet is resolved in most cases, making line identification easier. With 677 galaxies, this population is the second most populated category, after the Lyα emitters. These galaxies cover a broad range in $F775W$ AB magnitude with 23.3, 25.7 and 27.6 for the 10%, 50% and 90% percentiles, respectively. Examples are given in Fig. 25, rows b and c. We note that a small fraction (5%) of the galaxies in this wavelength range are not star forming [O II] emitters, but passive galaxies detected through their absorption lines.

The redshift range 1.5–2.8 corresponds to the MUSE redshift desert, a region where [O II] is redshifted out of the MUSE wavelength range and Lyα is not yet observable. The number of galaxies drops to 201. In this redshift range, galaxies are detected either through their UV absorption lines (50%) or by fainter emission lines, mainly the C III]$\lambda\lambda$1907,1909 or Mg II $\lambda\lambda$2797,2803 doublets (see examples in Fig. 25, rows d and e).

The last category comprises the Lyα emitters, covering the redshift range 2.8–6.7. With 1308 Lyα emitters or 60% of the complete sample, it is the most numerous and also the faintest population (29.4 median AB $F775W$). The vast majority of these galaxies (98%) are identified by their high equivalent width asymmetric Lyα emission line but a few galaxies (3%) exhibit

Table 8. Distribution of redshifts by confidence and for the different classes of sources.

Class	ZCONF				
	1	2	3	1–3	
Stars	0	0	8	8	0.4%
Nearby ($z < 0.25$)	4	0	21	25	1.1%
[O II] ($0.25 < z < 1.5$)	55	122	500	677	30.5%
Desert ($1.5 < z < 2.8$)	55	57	89	201	9.0%
Lyα ($z > 2.8$)	396	538	374	1308	58.9%
All	510	717	994	2221	100.0%

Notes. ZCONF range from 1 (less secure redshift) to 3 (most secure redshift); see Sect. 5.7.3.

in addition fainter emission lines such as C IV $\lambda\lambda$1548,1550 or C III]$\lambda\lambda$1907,1909. A small percentage (2%) are Lyα absorbers with broad absorption and little or no detectable Lyα emission. We note that a significant fraction of the identified Lyα emitters have no HST counterpart (see Sect 6.2). A few examples of Lyα emitters are displayed in Fig. 25, rows f to h.

It can be seen in Table 8 and in the right panel of Fig. 24 that the fraction of good to high redshift confidence (ZCONF > 1) is high (87%) for the low and intermediate redshift populations (i.e., $z < 2.8$). This ratio is slightly lower (70%) for the Lyα emitters for the reason that in most cases the Lyα line is the only line identified.

In Sect. 5.7.3 we presented the redshift confidence definition and assessment. Although it is a complex process based on multiple parameters and data sets, including the expert's judgement, emission line S/N is one of the primary factors in the decision process. This can be seen in Fig. 21 where we show the confidence values as a function of the main emission line S/N derived from the extracted spectra and the corresponding narrowband S/N.

We point out that sources with low redshift confidence (ZCONF = 1) have nevertheless a clear detection signal, but in general the S/N is too low to assign a redshift with high confidence. An example is given in Fig. 26 where the same source is shown for the UDF-10 and MXDF data sets with 34- and 128-h depths, respectively. In the UDF-10 data set (upper left panel in Fig. 26), the emission line is detected with an S/N of 5.3 and does not appear to be very asymmetric ($\gamma = 1.5 \pm 2.6$). The corresponding narrowband (lower left panel) shows a clear peak, although with a low S/N of 3.5. The measured redshift is 3.17, which is not far from the R15 photometric redshift of 2.74 estimated for the matched HST source. We note, however, that the photometric redshift (zp) measurement is not very reliable with $(zp_{\rm max} - zp_{\rm min})/(1 + zp) = 0.27$. Taking all these points into consideration, the expert gave a low redshift confidence to the source. The same source with the deepest MXDF observations (right column in Fig. 26) obtains a higher S/N of 10.7 and 7.8 for the emission line and the corresponding narrowband, respectively. The line profile is now clearly asymmetric with $\gamma = 3.6 \pm 2.6$. Given these measurements, a high redshift confidence (ZCONF = 3) was given to the source.

6.1.2. Comparison with DR1

We recall that the DR1 catalog (Inami et al. 2017) was based on an early data reduction of the UDF-10 and MOSAIC data sets and did not include the new MXDF data set. Furthermore, although the overall detection and inspection process

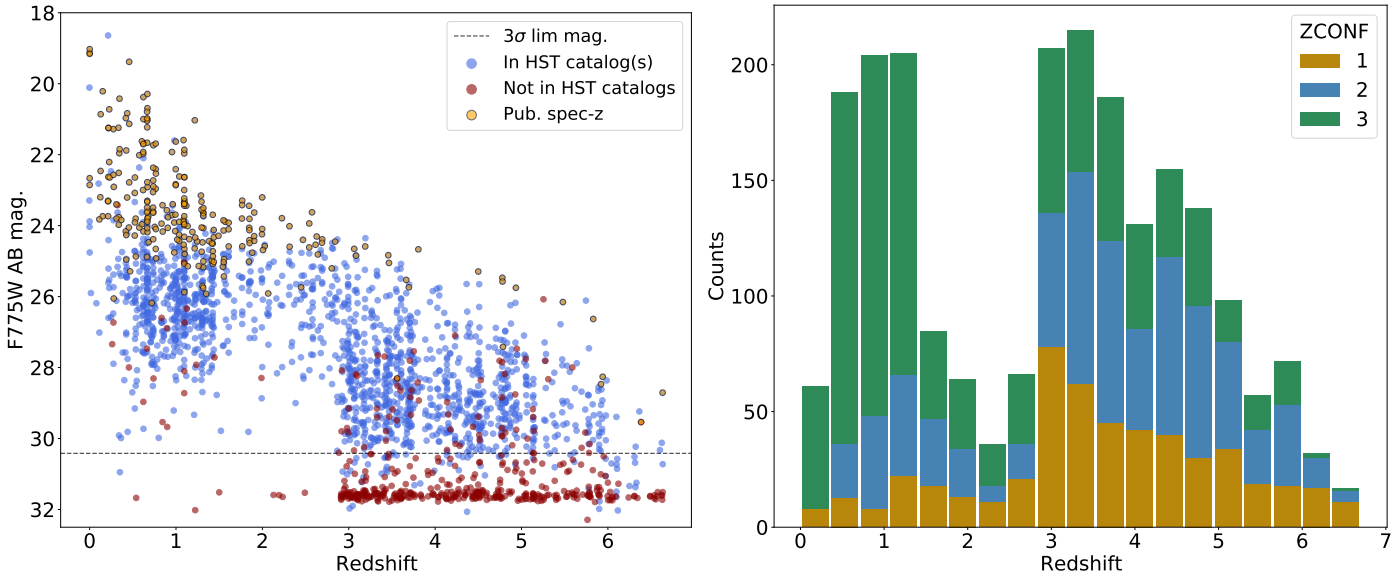


Fig. 24. Redshift distribution of the MUSE data release II. The magnitude – redshift scatter plot is displayed on the left panel. All MUSE sources with an HST counterpart in at least one of the HST catalogs are shown as blue symbols. We note that objects with low S/N in $F775W$ could still be in HST catalogs if they are detected in another filter (e.g., NIR filters for high-z objects). The matched spectroscopic redshifts previously published (Sect. 6.1.3) are shown as yellow circles. For sources without catalog HST counterparts, the brown circles show the 0.″4 aperture magnitude as measured on the HST F775 image (cf. Sect. 6.2). The horizontal dashed line indicates the median $F775W$ 3σ limiting magnitude. The redshift histogram, colored by redshift confidence, is presented in the right panel.

was similar, there are a few important differences. First, the data reduction was significantly improved with the use of super-flatfield, improved sky-subtraction and self-calibration (Appendix B). Second, the HST basic aperture extraction was replaced by the de-blending ODHIN procedure (Sect. 5.3). Third, we use new templates for pyMarZ (Sect. 5.4) and the improved line fitting software pyPlatefit (Sect. 5.5). So we expect some differences between DR1 and the new catalog, even in the area outside the MXDF footprint. A summary of the differences is given in Tables 9a and b.

The DR1 catalog contains 1574 sources. While most of the DR1 sources are present in the new catalog, a small fraction (9%) were not confirmed. The reasons for the failure are given in Table 9a. The rejected sample is mainly composed of low-confidence DR1 sources that were rejected or assigned a different redshift in the new deeper (MXDF) or improved (MOSAIC, UDF-10) data sets. For traceability purposes, new MUSE identifiers were given to DR1 sources that had been assigned a different redshift.

In Fig. 27 we display the quantitative improvement between DR1 and DR2. As shown in Table 9b, the majority of the 792 new DR2 sources (53%) are provided by the MXDF data set. This was expected given the improved depth, that is, a factor of 12 and 4 with respect to MOSAIC and UDF-10, respectively. However, there is still a significant gain of about 25% for the UDF-10 and MOSAIC data sets. This demonstrates that the effort spent on improving the entire process (i.e., data reduction, source detection and classification) has paid off.

6.1.3. Comparison with published spectroscopic redshifts

We also compared our redshifts with the compilation of all published spectroscopic redshifts in the CDFS given by Merlin et al. (2021) prior to 2021[17]. The compilation has 4951 entries, of which 292 are within the area of MUSE HUDF, which is 14 times smaller than the CDFS area.

A large majority (255 or 87%) of the published spectroscopic redshifts were successfully matched to MUSE redshifts, with the exception of 26 sources (9%) that were assigned a different redshift. Excluding the three galaxies with a low MUSE redshift confidence (MID-7951, MID-8025 and MID-8038) for which the proposed published solution is a viable alternative, we are confident that the MUSE assigned redshift is the correct solution. In addition, 11 galaxies (4%) of the published sample did not get any redshift in MUSE. We verify the proposed redshift solution and confirm the absence of clear features in the MUSE wavelength range. For these galaxies, the main feature (e.g., Lyα emission or absorption, [O III]$\lambda\lambda$4959,5007 emission) lies outside the MUSE wavelength range.

The scaled deviation of the published redshift with respect to MUSE, that is, $\Delta z = (z_{\rm MUSE} - z_{\rm Pub})/(1 + z_{\rm MUSE})$, has a standard deviation of $\Delta z = 0.014$ after clipping 25 sources at 5σ. We note

Table 9. Comparison of DR2 and DR1 catalogs.

(a) DR1 sources not matched in DR2						
Reason	ZCONF = 1		ZCONF > 1		Total	
New Z	25	1.6%	5	0.3%	30	1.9%
No Z	88	5.6%	17	1.1%	105	6.7%
Duplicate	1	0.1%	5	0.3%	6	0.4%
New Match	1	0.1%	0	0.0%	1	0.1%
All	115	7.3%	27	1.7%	142	9.0%
(b) New DR2 sources with respect to DR1						
Data set	ZCONF = 1		ZCONF > 1		Total	
MOSAIC	214	15.6%	110	8.0%	324	23.6%
UDF-10	35	19%	17	9.2%	52	28.3%
MXDF	159	23.9%	257	38.7%	416	62.7%
All	408	18.4%	384	17.3%	792	35.7%

[17] We note that the MUSE DR1 redshifts published in Inami et al. (2017) were not included in the Merlin et al. (2021) compilation.

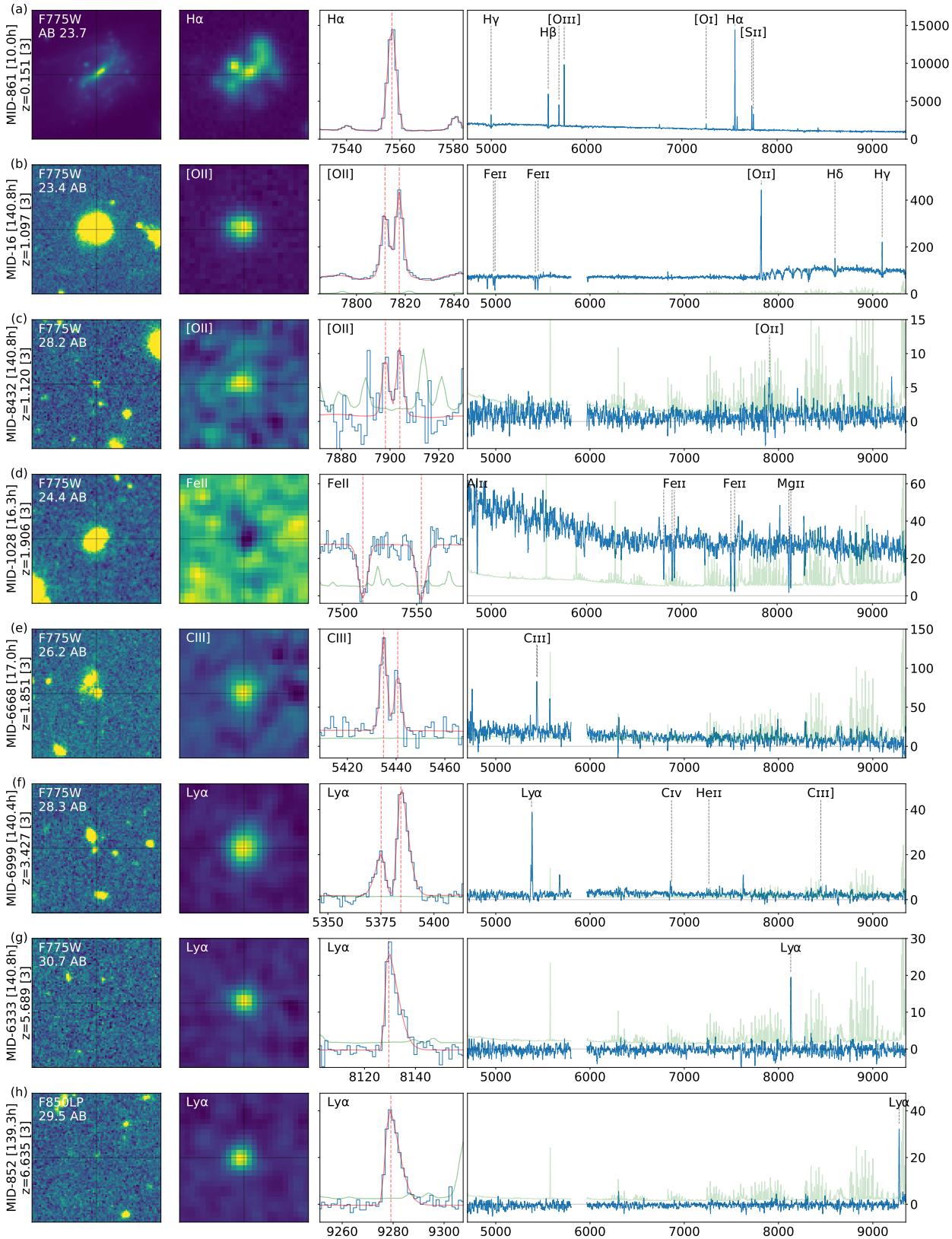


Fig. 25. Examples of sources. From left to right the columns display (1) the HST image in the *F775W* or F850LP filter (2) the emission or absorption narrowband for the main line (3) the spectrum zoomed-in on the main line (in blue) and its fit (in red) (4) the full spectrum (in blue). The green curves show the noise standard deviation. Except for the first two sources, the narrowband images and the full spectra have been filtered with respectively a Gaussian of 0.5″ FWHM and a box filter of 5 pixels. Image sizes are 5″ × 5″. Observed wavelengths are in Å and the flux unit is 10^{-20} erg s^{-1} cm^{-2} Å^{-1}. For each source the MUSE identifier, depth in hours, redshift value and confidence, and AB HST magnitude are indicated.

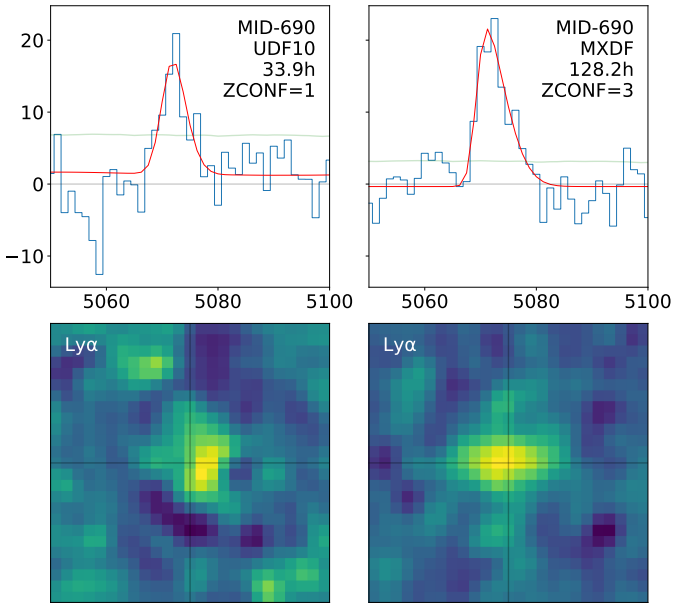


Fig. 26. Impact of depth on the redshift confidence assignment. The Lyα line (data in blue color, fit in red color) and the corresponding narrowband image of the Lyα emitter MID-690 are shown for the UDF-10 (*left column*) and the MXDF (*right column*) data sets. The green curves display the noise standard deviation. The narrow bands have been smoothed with a Gaussian of 0.5″ FWHM. Image sizes are 5″ × 5″. Observed wavelength and flux units are respectively Å and 10^{-20} erg s^{-1} cm^{-2} Å^{-1}.

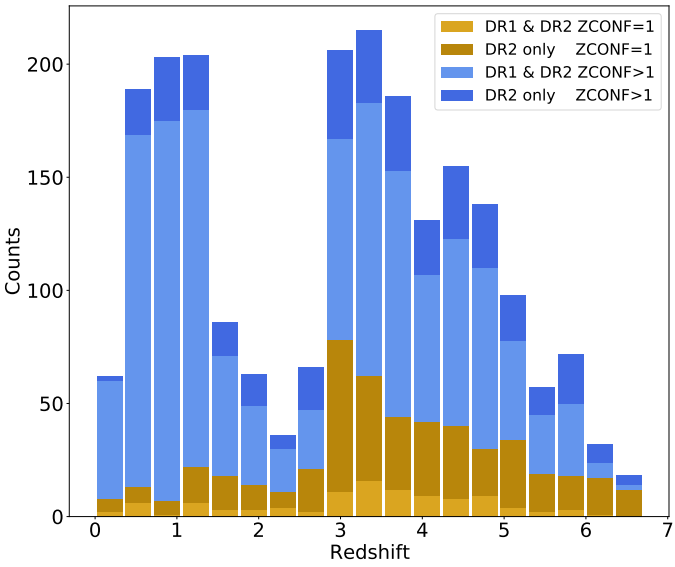


Fig. 27. Comparison of redshift statistics between the DR1 and DR2 data releases. Light and dark colors are for low (ZCONF = 1) and high (ZCONF > 1) redshift confidence, repectively.

that the redshift errors reported in the literature are often greatly underestimated: on average, they are 6 times smaller than the estimated error based on the MUSE reference redshifts.

Just looking at the raw numbers (i.e., 2221 versus 292 spectroscopic redshifts), one can appreciate the dramatic progress achieved by MUSE. As shown in Fig. 28, the difference is even more striking at high redshift ($z > 2.9$) where only 20 sources out of 1291 have previously been identified. The same holds for faint magnitudes: most of the published spectroscopic redshifts

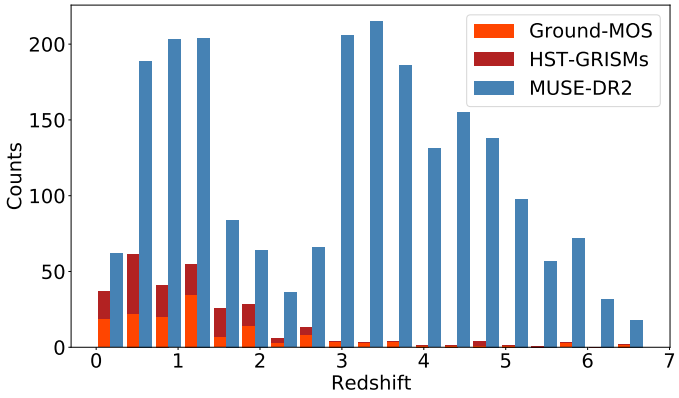


Fig. 28. Comparison of the spectroscopic redshift distributions of ground-based MOS and HST slitless GRISM published spectroscopic surveys in the HUDF area and the MUSE DR2 redshifts.

are from galaxies brighter than AB 24 in $F775W$, while the magnitudes of MUSE sources extend to AB 30 and beyond (Fig. 24 left panel).

6.1.4. Comparison with photometric redshifts

A comprehensive comparison of photometric redshifts with MUSE DR1 redshifts was performed by Brinchmann et al. (2017). Photometric redshifts are essential ingredients for on-going and upcoming cosmological surveys (e.g., KiDS, Hildebrandt et al. 2017; LSST, Ivezić et al. 2019; *Euclid*, Laureijs et al. 2011; Roman, Spergel et al. 2015). Some surveys have very stringent requirement in photo-z accuracy; for example, for weak lensing, the mean redshift in each redshift bin must be constrained to better than $2 \times 10^{-3}(1+z)$ (Newman et al. 2015). The goal of the Brinchmann et al. (2017) study was to test the performance of photometric redshifts for galaxies in the *Hubble* Ultra Deep Field down to the 30th magnitude.

One of the main conclusions of this study was the existence of systematic biases at intermediate (z = 0.4–1.5) and high redshift ($z > 3$), and a ≈10% proportion of outliers. We do not intend to repeat such a detailed analysis here, but we are interested in whether these conclusions based on DR1 are still valid for the more complete DR2 catalog.

For this comparison we use three different photometric redshifts catalogs. The first catalog is from Rafelski et al. (2015; hereafter BPZ-R15), which was already used in Brinchmann et al. (2017). We use the Bayesian photometric redshift values derived with the BPZ code (Benítez 2000). The second catalog contains the 3D-HST redshifts from Momcheva et al. (2016). These redshifts were computed from a simultaneous fit of multiband HST photometry and 2D spectra extracted from the HST WFC3 G141 GRISM. Although it is not a pure photometric redshift catalog, we included it in the comparison given that the additional GRISM data should provide improved redshifts with respect to purely photometric redshifts. We note, however, this is only true in the wavelength range covered by G141 and only for specific redshifts for which spectral features (especially breaks) are sampled by the G141 wavelength range. It is also restricted to object bright enough (i.e., with F140 JH magnitudes brighter than 26). The last one is the recent Merlin et al. (2021) ASTRODEEP-GS43 spectro-photometric catalog. It is based on HST CANDELS data complemented with additional ground-based photometric data, reaching a total of 43 different photometric wide and narrow

bands. Their photometric redshifts are derived with three different codes (LePhare, Ilbert et al. 2006; EAzY, Brammer et al. 2008; z-phot, Fontana et al. 2000), using the median of the three derived values.

The input catalog is composed of 1711 sources with good and high confidence redshifts (ZCONF > 1). The BPZ-R15, 3D-HST, and ASTRODEEP catalogs have respectively 1404, 1088, and 1009 matched sources. As in Brinchmann et al. (2017), we compute the normalized redshift error as

$$\Delta z = \frac{z_{\rm MUSE} - z_{\rm phot}}{1 + z_{\rm MUSE}}. \quad (6)$$

We evaluated the evolution of Δz with redshift and magnitude by computing the median error and the outlier fraction in bins of about 100 elements. We note that the distributions of redshift and magnitude are different for the three catalogs. For example, BPZ-R15 has more sources than the other two catalogs, it extends to fainter magnitudes and has a larger number of sources at high-z. Consequently, the redshift and magnitude bins differ between catalogs. In each bin, the outlier fraction is computed as the fraction of sources with $\Delta z > n\,\sigma_{\rm MAD}$ where $\sigma_{\rm MAD}$ is the median absolute deviation around the median, adjusted to match a standard deviation for a Gaussian distribution (Eq. (2) of Brinchmann et al. 2017). A value of $n = 3$ is used for simple outliers and $n = 10$ for catastrophic outliers.

The results are shown in Fig. 29. The BPZ-R15 curve (blue curve in top left panel) displays the same behavior with redshift as observed in DR1 (Fig. 5 left panel of Brinchmann et al. 2017), with a systematic negative offset ($\Delta z \approx -0.03$) at $z \approx 1.2$ and a positive offset ($\Delta z \approx +0.05$) at $z \approx 3$. We note, however, that the increased offset at $z > 4$ observed in DR1 is not observed in our larger sample. The median standard deviation measured in BPZ-R15 is 0.04, only slightly higher than the DR1 estimation. The evolution as a function of magnitude (blue curve and shaded area in the top right panel) displays a similar trend as for DR1: the standard deviation increases from 0.03 at AB < 25 to 0.06 at AB > 28. The observed fraction of outliers is 9.4% and 2.8% for the catastrophic outliers. These figures are consistent with the DR1 measurements.

The comparison with the ASTRODEEP photometric redshifts is shown in green in Fig. 29. The additional photometric bands have improved the accuracy of the photometric redshifts. There are almost no detectable systematic offset with either redshift or magnitude. The scatter is also much reduced compared to BPZ-R15. It is less than 0.02 at AB 26 and then increases to 0.04 at AB 28.5. We observe, however, that the average fraction of outliers increases, with 16% and 9% for the 3σ and 10σ outliers, respectively. We note that these numbers (scatter and outlier fraction) are higher than those quoted by Merlin et al. (2021). Their underestimation probably results from the limited spectroscopic sample used (see Sect. 6.1.3) in their tests.

Surprisingly, the comparison with the 3D-HST catalog (orange curve and shaded area in Fig. 29) reveals a net degradation of performance at AB > 26. We note that the Momcheva et al. (2016) catalog only includes grism information for sources with JH F140 magnitudes brighter than 26. The increased scatter must therefore be due to the photometric redshift estimate.

6.2. Sources without HST counterparts

During the inspection process (Sect. 5.7.4) we identified 424 (20%) MUSE sources that cannot be matched to any of the four HST catalogs: R15 UVUDF (Rafelski et al. 2015), CANDELS v2 (Whitaker et al. 2019), 3D-HST (Skelton et al. 2014; Momcheva et al. 2016), ASTRODEEP (Merlin et al. 2021). In a some cases (86), the source can still be clearly seen in the HST image. An example of such "missed" sources is shown in Fig. 22. This demonstrates the limitations of source de-blending based solely on morphology and broadband colors.

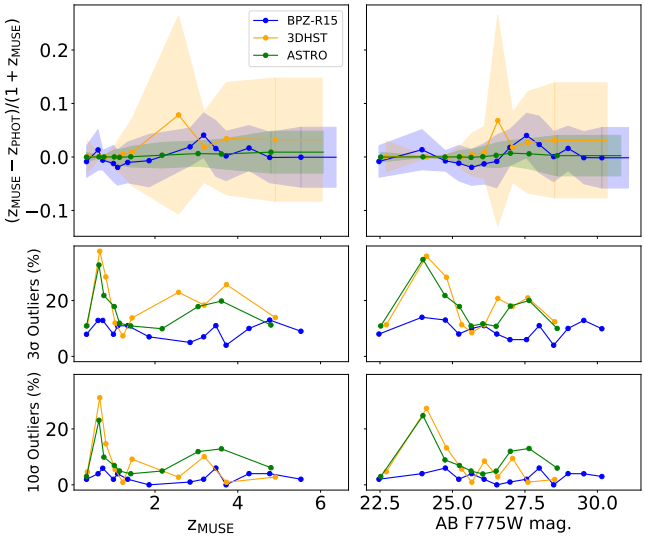


Fig. 29. Redshift off-set between spectroscopic redshifts from MUSE and the photometric redshifts from Rafelski et al. (2015) BPZ-R15 (blue), Momcheva et al. (2016) 3D-HST (orange) and Merlin et al. (2021) ASTRODEEP (green) catalogs, all normalized by $1+z_{\rm MUSE}$. The top left panel show the median (solid line) as a function of redshift and the scatter on the median is shown by the shaded areas. The middle and bottom left panels show the fraction of outliers at respectively 3σ and 10σ. The right panels display the same quantities as a function of HST $F775W$ AB magnitude.

The vast majority of sources not found in the HST catalogs were classified as "faint" or "undetect" by the experts. In order to give more quantitative ground for this classification, we performed our own HST photometry at the locations of MUSE sources unmatched to the HST catalogs.

As in Maseda et al. (2018), we measure the S/N ratio for objects using the HST/ACS mosaics from the XDF survey (Illingworth et al. 2013). We determine the fluxes in apertures of 0.4 arcsec diameter centered on the DR2 catalog positions. This aperture corresponds to a physical size of 3.1 kpc (2.2 kpc) at $z = 2.9$ (6.6). The local background level is calculated by measuring the standard deviation of the fluxes in 250 identical apertures spread randomly within a $10'' \times 10''$ cutout centered on the object, with other objects masked according to the R15 segmentation maps. If the aperture flux is greater than three times the local background level in an HST band, then we consider the object "detected" in that band. Otherwise we use the 3-σ upper-limit to the flux in that band.

The computed magnitudes of these sources are shown in the right panel of Fig. 30. Using the 3σ S/N detection limit, we have 189 sources that fall below the detection limit in all HST bands. We note that the majority of this sample (90% or 380 sources) have AB magnitude greater than 29. In practice, as shown in Sect. 6.1.4, no reliable photometric redshifts can be obtained for these faint sources, even at the UDF HST depth (Fig. 29).

We observe that almost all sources without HST counterparts are Lyα emitters (left panel of Fig. 30). The nature of these sources has been discussed by Maseda et al. (2018, 2020). The authors conclude that they constitute the high equivalent-width

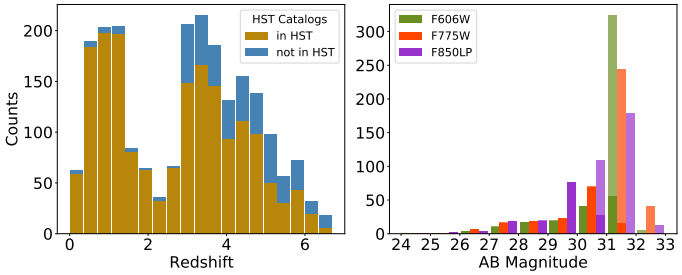


Fig. 30. Statistics of MUSE sources without HST counterpart. *Left panel*: comparison of the redshift distributions of MUSE sources with (brown color) and without (blue color) HST counterparts. *Right panel*: estimated AB aperture magnitude for MUSE sources without HST counterparts. The histogram is colored according to the HST filters. Successful photometric measurements ($S/N > 3$) are displayed in dark colors, while measurements with $S/N < 3$ are displayed with light colors. In the latter case the reported magnitude is the measured noise RMS standard deviation.

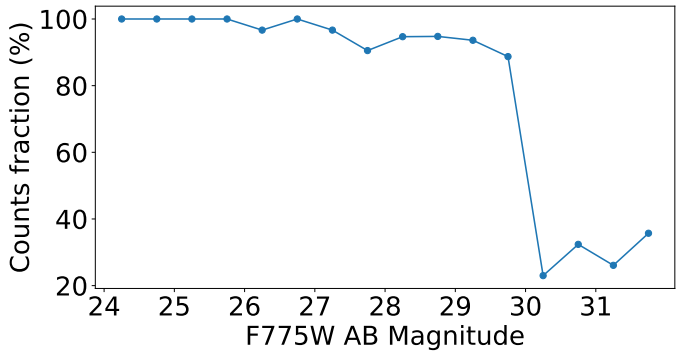


Fig. 31. Fraction of MUSE Lyα emitters with HST counterparts as a function of AB $F775W$ magnitude in bins of 0.5 magnitude.

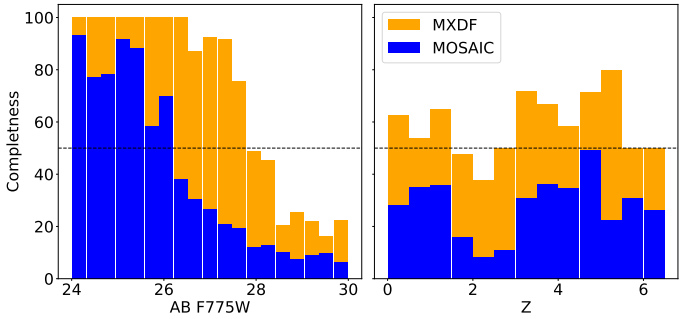


Fig. 32. DR2 completeness with respect to HST deep imaging sample. Completeness is computed as the number of galaxies with a MUSE redshift in a 0.31 HST AB $F775W$ magnitude (*left panel*) or 0.5 photometric redshift interval (*right panel*). The measurements have been performed for the deep MXDF (+100-h depth, orange bars) and MOSAIC (+10-h depth, blue bars) data sets and corresponding areas.

tail of the Lyα emitter population and that they have on average a high ionizing photon production efficiency. Although their study is based on 103 faint sources only and does not include the deepest part of the DR2 catalog related to MXDF observations, these conclusions are not invalidated by the present data set.

6.3. Completeness

The completeness measures how complete our sample is with respect to one of its observational or physical properties (e.g., source magnitude or line flux). It is an essential measurement to infer the LF of a given galaxy population. To estimate the completeness, we have to evaluate the number of missed sources at a given luminosity and therefore capture all the biases introduced by the whole data chain (i.e., sample selection, observation, data reduction, and analysis).

In this paper, the objective is to maximize the number of reliable redshifts in the HUDF area. We therefore have used multiple detection methods: blind detection with ORIGIN and HST prior extraction with ODHIN. We also used additional available information (e.g., photometric redshifts) to choose between various redshift solutions. Each detection method has its own bias. While HST prior ODHIN detection is based on continuum detection and will therefore be biased toward galaxies with stellar continuum, ORIGIN blind detection will preferentially select high equivalent-width emission line galaxies. Inferring the completeness of our catalog is therefore a challenging task and beyond the scope of this paper.

A useful estimate of completeness can nevertheless be made by comparing our catalog with the HST spectro-photometric catalogs. Given the great depth reached by Hubble in the UDF (i.e., 5σ 29.5 AB $F775W$ magnitude), the HST catalogs should be quite complete and could be used as an unbiased parent sample of broadband-selected galaxies. A rough estimate of HST Lyα completeness can be obtained for the Lyα emitter population by counting the fraction of MUSE sources with HST counterparts with respect to the full Lyα emitters sample. As shown in Fig. 31, the HST catalogs can be considered as complete (90%) up to AB 29.5. At fainter magnitudes, the HST catalogs have a significant drop.

We performed the comparison between MUSE and the R15 spectro-photometric catalog in two areas: the MXDF region limited to the 100+ h of depth and the MOSAIC area with 10+ h of depth, after removing the MXDF area. The former covers an area of 0.85 arcmin^2 and has 787 sources identified[18] in the R15 catalog, the latter covers 7.48 arcmin^2 and has 3619 sources identified in R15. The average depth achieved with MUSE in these two areas is 140.8 and 10.8 h, respectively. In each sample, we measure the fraction of HST sources that have a MUSE redshift in a given magnitude or redshift interval. The completeness evolution with respect to magnitude and photometric redshift is given in Fig. 32.

For a given magnitude, the evolution of completeness with redshift is not flat: first, we see a clear drop in the $z = 2-3$ range, corresponding to the redshift range of the MUSE desert, then we measure a higher completeness at $z > 3$, indicating that the fraction of Lyα emitters is high in this redshift range. In the MOSAIC sample at a 10-h depth, the 50% completeness is achieved at a magnitude of 26.6 ($F775W$) in the redshift range $z = 0.8–1.6$ and 27.5 at $z \approx 4$. In the deepest MXDF observations, the corresponding magnitude for the 50% completeness are 27.6 and 28.7 for $z = 0.8–1.6$ and $z = 3.2–4.5$, respectively.

It is instructive to compare these figures with those obtained for the published spectroscopic redshifts sample (Merlin et al. 2021 compilation). The 50% completeness rate is achieved at a magnitude of 25.2 ($F775W$) in the $z < 3$ redshift range. At higher redshifts ($z > 3$) the completeness drops very rapidly with increasing magnitude: 20% for AB 26, 5% for AB 27.

[18] We exclude all sources without measurable $F775W$ magnitude.

6.4. Physical properties

In this section we derive the stellar masses and star-formation rates (SFRs) for our galaxy sample using spectral energy distribution (SED) fitting. Since the wavelength range of MUSE does not extend far enough into the infrared to provide sufficient constraints for the SED fit, we selected the subsample of 1664 galaxies (75%) with ZCONF > 0 that have been identified in the HST R15 catalog. The R15 catalog has 11 photometric bands[19] ranging from the NUV (0.21 µm) to the WFC3/IR (1.5 µm).

We used the high-z extension of the code Magphys (da Cunha et al. 2008, 2015) with a minimum stellar mass of $10^6 M_\odot$ (see Sect. 3.2 of Maseda et al. 2017). The redshift of each object is fixed to the MUSE spectroscopic redshift. In cases where the fitting results in un-physically small errors in stellar mass and/or SFR (i.e., when the models cannot accurately reproduce the photometric data points), we adopted a fixed error of ±0.3 dex (the typical uncertainty in determining stellar masses from fits to broadband photometry; e.g., Conroy 2013; Maseda et al. 2014).

In addition to Magphys, we further utilized the Prospector SED fitting code (Johnson et al. 2021) to derive stellar masses and SFRs for the same R15 photometric catalog. Through Prospector we are able to consider the contribution from emission lines to galaxy photometry. Given that a large fraction of our sources have strong emission lines, constraining the contribution of the emission lines in deriving galaxy properties is important.

In Prospector we used FSPS models (Conroy et al. 2009) with the MILES spectral library and MIST isochrones to generate synthetic spectra to match with the observations. These models include contributions from nebular continuum and emission as described by Byler et al. (2017). We used a nonparametric approach to fit the observed photometry using the "continuity_sfh" template in Prospector. This computes the stellar mass produced in fixed time bins with a smoothness prior (Leja et al. 2019). We defined 7 time bins to fit for stellar mass. The first two bins were fixed at 0-30 and 30–100 Myr in lookback time. The furthest bin was fixed to be between 85%–100% in lookback time defined by the MUSE spectroscopic redshift. The remaining four bins were evenly distributed between the second and final bin in logarithmic space. We used a Calzetti et al. (2000) dust attenuation law and allowed the V-Band dust optical depth to vary as a free parameter between 0 and 2.0. Gas-phase metallicity and stellar metallicity were tied together and were allowed to vary between 0.01 and ≈1.5 times the solar metallicity. We let the gas ionization parameter vary between $\log(U) = -4$ and -1. In total we had 10 free parameters (6 SFH bins + stellar mass, dust, metallicity, and ionization parameter). Parameters were sampled using the Dynamic Nested Sampling method presented by Speagle (2020) using 200 live points and batches and 50 nested walks.

The stellar masses and the SFR averaged over the last 100 Myr are given in the main catalog (Appendix A.2). The error for the masses and SFR is based on the 1-sigma percentiles of the nested sampling chains. The stellar mass output by Prospector is the "total formed mass". In order to convert this to the observed mass, we use the mass correction factor from the maximum a posteriori solution (the posterior sample with the highest posterior possibility) to derive an approximate correction factor for the marginalized stellar mass values.

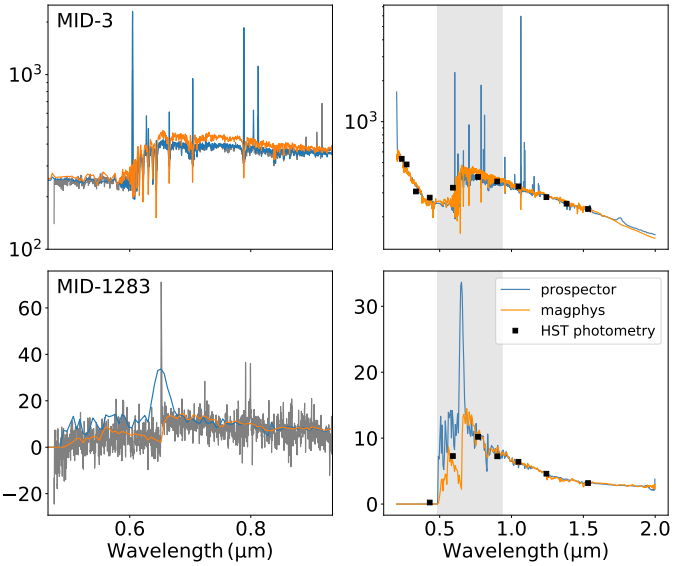


Fig. 33. Two example SED fits: the $z = 0.62$ MID-3 [O II] emitter (top panels) and the $z = 4.36$ MID-1283 Lyα emitter (bottom panels). The left column shows the MUSE spectrum (in black) and the Prospector (blue line) and Magphys (orange line) SED fits. The right column displays the SED fits and the HST R15 photometry (black symbols). The shaded area in the right column indicates the MUSE wavelength range. Flux and wavelength units are 10^{-20} erg s^{-1} cm^{-2} Å^{-1} and µm, respectively.

Two examples of the SED fit are given in Fig. 33. In the case of MID-3 (upper left panel), it can be seen that the SED fit of Magphys, unlike Prospector, overestimates the continuum due to the presence of strong emission lines[20]. In Fig. 34, we compare the Magphys and Prospector stellar mass and SFR results for the full sample. We note that Prospector tends to derive higher masses than Magphys with a median offset of 0.25 dex. This is a known feature of Prospector that is documented in Leja et al. (2019, 2020). The stellar mass and SFR distributions are presented in the first two panels of Fig. 35. As expected, the MUSE catalog probes low mass star forming galaxies with a median mass of $6.2 \times 10^8\, M_\odot$ and a median SFR of $0.4\, M_\odot \mathrm{yr}^{-1}$. The last panel of Fig. 35 displays the traditional star forming main sequence. We note that our main sequence deviates significantly from the best-fit star forming main sequence at $z = 1$ derived from a compilation of 25 studies by Speagle et al. (2014; dashed blue line). Such an offset between Prospector SED results and previous studies has been identified and discussed by Leja et al. (2022). As noted by the authors there are many reasons for the difference in inferred stellar mass and SFR between the Prospector's nonparametric star-formation histories and the classical approaches (see their Sect. 6.2 for a detailed discussion of the origin of this difference). We also compared our results at $z = 1$ with those of Boogaard et al. (2018; Fig. 35 solid blue line). The Boogaard et al. (2018) study was based on DR1, still using HST photometry SED to estimate stellar mass but with a different derivation of SFR based on the Hα and Hβ recombination line flux. As shown in the plot, there is still a small offset between our estimate and the results from Boogaard et al. (2018), but it is much smaller compared to the values of Speagle et al. (2014). This residual offset is most likely due to the different estimate

[19] WFC3/UVIS $F225W$, $F275W$, and $F336W$; ACS/WFC $F435W$, $F606W$, $F775W$, and F850LP and WFC3/IR $F105W$, $F125W$, $F140W$, and $F160W$.

[20] There are also some small differences in the photometry between MUSE and HST, as discussed in Sect. 4.1, that may account for a small offset between the two sets of measurements.

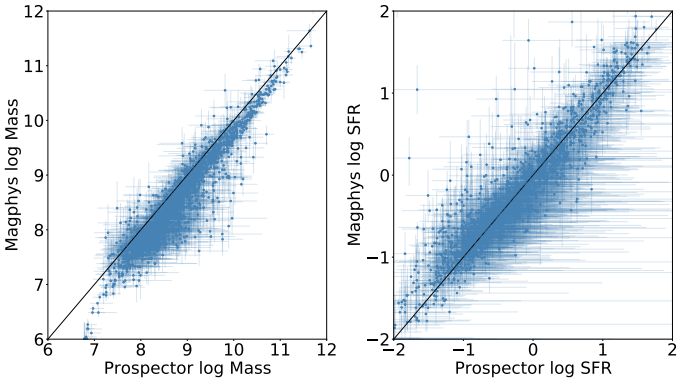


Fig. 34. Comparison of the Prospector and Magphys stellar mass and SFR results for the SED fits of all MUSE galaxies identified in the R15 catalog.

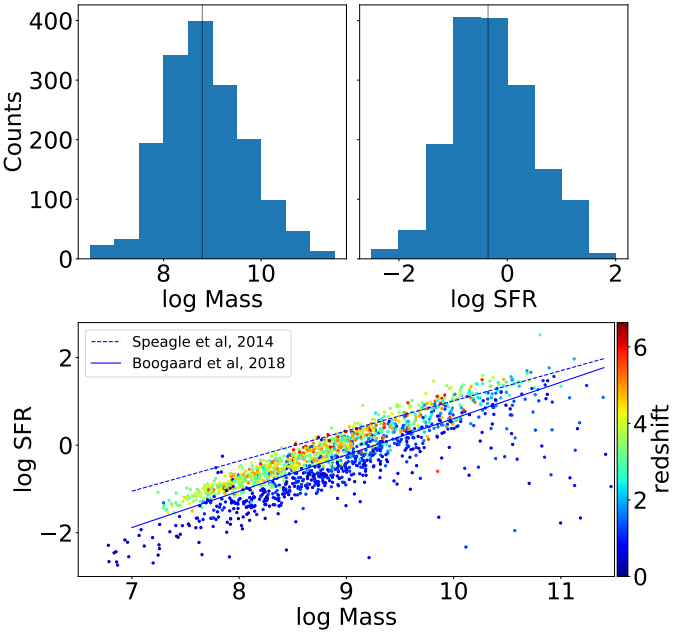


Fig. 35. Stellar mass (*top left panel*) and SFR (*top right panel*) distributions for the MUSE galaxies identified in the R15 catalog. Values are given for the Prospector SED fit in the ≈100 Myr lookback time. The stellar mass and the SFR are in log $M_\odot$ and log $M_\odot$ yr^{-1} units, respectively. The vertical black lines show the median values of the distributions. The *bottom panel* displays the star forming galaxies main sequence log SFR(log M). Symbols are colored according to the redshift. The $z = 1$ main sequence derived by Boogaard et al. (2018) and Speagle et al. (2014) are shown as blue solid and dashed lines, respectively. See the main text for a discussion of the offset from our inferred values.

of the SFR. Further investigations will be needed, for example into the completeness and SED degeneracy, to interpret this main sequence, but such an analysis is beyond the scope of this paper.

Finally, we recall that this SED analysis does not take into account the population of 577 (25%) galaxies without an HST counterpart in the R15 catalog. The majority of these sources are high equivalent-width Lyα emitters that are too faint in the HST broadband deep images to get meaningful SED measurements. However, as demonstrated by Maseda et al. (2018, 2020), this selection yields on average faint ($M_{\rm UV} \approx -15$), blue ($\beta \approx -2.5$) star-forming galaxies. Although we cannot directly determine the stellar masses from the UV continuum alone, an extrapolation to the Duncan et al. (2014) $M_{\rm UV} - M_\odot$ relation to $M_{\rm UV} = -15$ implies that these galaxies should have stellar masses below $10^7 M_\odot$ at these redshifts.

6.5. Peculiar objects

Among this rich set of data, we give in what follows some examples of interesting objects. The first category is the Lyα emitters with very extended Lyα emission. These objects are often called Lyman alpha blobs (LABs) in the literature. They are rare and mainly found in overdense regions like the two LABs discovered in the SSA22 supercluster (Steidel et al. 2000). While most Lyα emitters and rest-frame UV-selected galaxies exhibit diffuse Lyα emission (Wisotzki et al. 2016; Leclercq et al. 2017; Kusakabe et al. 2022), LABs differ in terms of their high luminosity (typically brighter than 10^{43}erg s^{-1}) and their large size (typically larger than 50 kpc; e.g., Herenz et al. 2020). A recent review is provided in Kimock et al. (2021). We show in Fig. 36, three examples of Lyα emitters with such extended Lyα emission. The first object (MID-208 and MID-6329) is a $z = 3.32$ source detected in the UDF-10 at a 29-h depth. With a Lyα luminosity of 8×10^{42} erg s^{-1} and a size of ≈60 kpc, it enters the LAB category. This object was already reported by Vanzella et al. (2017) using 20 h of the same GTO observations. The second example (MID-1056 and MID-3621) was discovered in the MOSAIC at an 11-h depth. This $z = 3.06$ LAB is also bright (1.4×10^{43} erg s^{-1} Lyα luminosity) and extended (≈60 kpc). This object coincides with a luminous type II active galactic nucleus identified in the *Chandra* 7Ms catalog (Luo et al. 2017). The extended Lyα emission was already reported by den Brok et al. (2020) using the same data set. It is part of a 1.2 Mpc long Lyα filament (Bacon et al. 2021). The third example, the $z = 3.2$ MID-1530 Lyα emitter, is less extended (30 kpc) and luminous (1.4×10^{42} erg s^{-1} Lyα luminosity) than the other two and would not be formally categorized as a LAB. It is representative of the class of the large Lyα halos found around Lyα emitters (e.g., Leclercq et al. 2020).

In addition to the MID-1056/3621 type II active galactic nucleus, we have two type I QSOs in the sample (Fig. 37). The first is MID-977, a $z = 1.2$ object with broad Mg II emission and some absorbers detected along the line of sight. The second is MID-1051, a $z = 3.2$ QSO with damped Lyα absorption and C IV, He II, and C III] emission.

As previously stated, Lyα emitters are typically surrounded by a Lyα halo that is detected out to approximatively ten times the continuum size (e.g., Leclercq et al. 2017). Moreover, the peak of the Lyα emission can be offset with respect to the UV continuum peak (e.g., Ribeiro et al. 2020; Claeyssens et al. 2022). We present an extreme example of this in Fig. 20, where the $z = 4.8$ Lyα emitter (MID-1264) has a 4.1 kpc (0.6 arcsec) offset between the UV and Lyα narrowband centers. We note that such an offset is much larger than the 0.6±0.05 kpc average offset observed by Ribeiro et al. (2020) in their deep VIMOS slit data. Nevertheless, these offset measurements rely on a correct match between the UV continuum source (in our case the deep HST image) and the Lyα emitter. This is not always an easy task, and MID-1264 is a good example. The Lyα emitter was matched to RID-6956. This object is bright enough (25.6 AB $F775W$ magnitude) to have a reliable photometric redshift. All HST catalogs give similar estimates, with for example $z_{\rm phot} = 4.77 \pm 0.07$ for the ASTRODEEP catalog. Our measurement of $z = 4.78$ fits well within the photo-z error bar, giving confidence that the Lyα emitter is matched with the HST source. In addition, all other

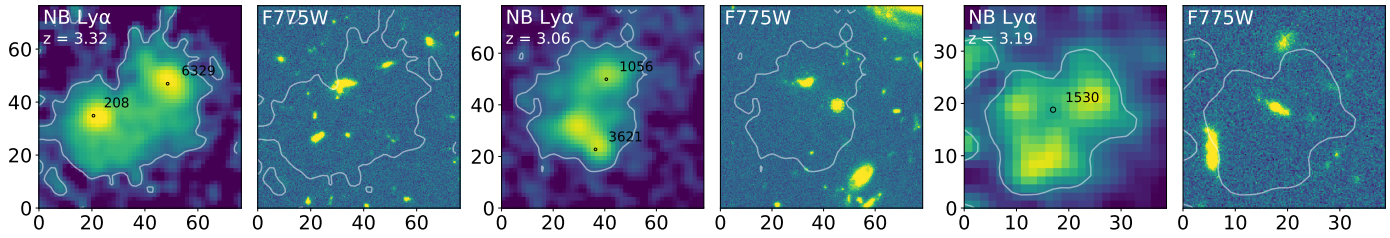


Fig. 36. Examples of sources with very extended Lyα emission. For each source we show the Lyα narrowband emission (NB Lyα) and the HST *F775W* image. Corresponding catalog source locations and MIDs are indicated in black. The 2.0, 7.0 and 2.8×10^{-20} erg s^{-1} cm^{-2} Lyα flux contours are overlaid in each image *from left to right*. The units of the image axes are physical kpc.

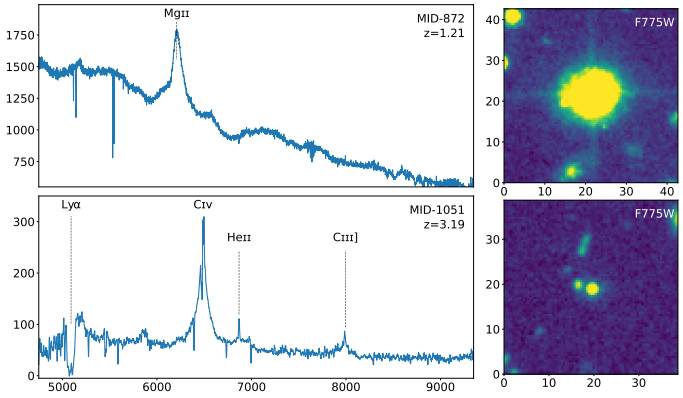


Fig. 37. Spectra of the two QSOs and their corresponding HST *F775W* images. The units of the image axes are physical kpc. Observed wavelength and flux units are respectively Å and 10^{-20} erg s^{-1} cm^{-2} Å^{-1}.

nearby HST sources are much farther from the location of the Lyα emission peak and their MUSE or photo-z redshifts indicate that they are foreground objects. However, the high S/N of the Lyα line profile shows evidence for a small dip at 7029 Å that can be interpreted as the superposition of two Lyα emitters at $z = 4.7806$ and $z = 4.7885$ respectively. In this case, the fainter Lyα line would be associated with RID-6956, while the brighter Lyα line would be a Lyα emitter without an HST counterparts, satellite to the main galaxy.

There are many examples where the HST sources have no reliable photo-z, are located in a crowded environment or have a complex morphology. In these cases, matching Lyα emission to HST imaging is a challenge. In general, we tend to favor HST-undetected Lyα emitter when the offset is large enough. This is in agreement with a recent study that used a large sample of lensed Lyα emitters observed with MUSE (Claeyssens et al. 2022); it shows that the average Lyα offsets are small, 0.66 ± 0.14 kpc, and that the large offsets are likely due to satellite or merging galaxies.

7. Summary and conclusions

This second data release of the MUSE *Hubble* Ultra Deep Field surveys is a major update of the DR1 published in Bacon et al. (2017) and Inami et al. (2017). Like DR1, it is based on deep MUSE GTO observations of the HST UDF area. It covers a field of view of 3×3 arcmin^2 at a 10-h depth (MOSAIC) and 1×1 arcmin^2 at a 31-h depth (UDF-10), with much improved data reduction and data analysis processes compared with DR1. A new data set, the MUSE extremely deep field (MXDF), also located in the UDF XDF area and covering an approximatively circular area of 1 arcmin in diameter, has been pushed to a depth of 141 h. This data set benefits from an improved spatial resolution of 0.55 arcsec at 7000 Å, thanks to the GLAO AOF system.

The achieved 3σ emission line surface brightness sensitivity at 7000 Å is 4.5, 2.3, and 1.0×10^{-19} erg s^{-1} cm^{-2} arcsec^{-2}for the 10-h depth MOSAIC, 31-h depth UDF-10, and 141-h depth MXDF datacubes, respectively. The corresponding 3σ point-source limiting flux for an unresolved emission line is 3.1, 1.5, and 0.6×10^{-19} erg s^{-1} cm^{-2}.

Advanced source detection and extraction using the ORIGIN blind detection (Mary et al. 2020) and the ODHIN HST-based source de-blending (Sect. 5.3) software resulted in the secure identification of 2221 sources, a 41% increase over DR1. Apart from eight stars, the collected sample is composed of 25 nearby galaxies ($z < 0.25$), 677 [O II] emitters ($z = 0.25$–1.5), 201 galaxies in the MUSE desert redshift range ($z = 1.5$–2.8), and 1308 Lyα emitters ($z = 2.8$–6.7).

We measure almost an order of magnitude more redshifts than the collections of all spectroscopic redshifts obtained before DR1 in the UDF (i.e., 2221 versus 292). At high redshift ($z > 3$), the difference is even more striking: 1308 versus 20. Since 2004, extensive campaigns of spectroscopic observations have been carried out with most large ground-based telescopes: for example, VIMOS/VLT (Le Fèvre et al. 2013; Garilli et al. 2021), FORS2/VLT (Mignoli et al. 2005; Vanzella et al. 2008), and DEIMOS/KECK (Doherty et al. 2005). It is surprising that, despite this significant effort in telescope time, so few spectroscopic redshifts have been obtained in the iconic HUDF field. The reason is technical: ground-based spectroscopy has been obtained with multi-object spectrographs (e.g., VIMOS; Le Fevre et al. 1998), which are most efficient when performing spectroscopy of galaxies at the bright end of the LF. At the faint end, the density of galaxies increases to a level where it becomes very inefficient to build masks or to pack enough fibers to get spectra of sources in such crowded environments. The other reason is inherent to the multi-object spectroscopy (MOS) concept, where a preselection is needed prior to the observations. Objects not or barely detected in deep broadband imaging will not be selected for the spectroscopic observations. This effect is most pronounced at high redshifts and explains why the large population of high equivalent-width Lyα emitters uncovered by MUSE is almost completely absent from the MOS observations.

The other sources of published redshifts are the slitless spectroscopic observations performed with HST (e.g., Brammer et al. 2012). Like in integral field spectroscopy, slitless spectroscopy does not require imaging preselection. There are, however, some limitations in crowded areas due to overlapping spectra. This is mitigated by performing multiple observations at different rotations, but it remains difficult to use for faint sources in crowded environments. The main limitation is, however, the small HST 2.4 m telescope aperture and the low

spectral resolution ($R \approx 50-100$), which limit its use to relatively bright compact sources.

Eighty percent of the galaxies in our final catalog have an HST counterpart identified in at least one of the published HST spectro-photometric catalogs. These galaxies are on average faint: a median AB $F775W$ magnitude of 25.7 and 28.7 for the [O II] and Lyα emitters, respectively. We measured the completeness of our catalog with respect to HST and found that at a 10-h depth, 50% of galaxies of magnitude $F775W$ AB 26.6 ($z = 0.8-1.6$) and 27.5 ($z \approx 4$) are present in DR2. This number increases in the deepest 140-h depth MXDF area with 50% completeness for $F775W$ AB 27.6 and 28.7 at $z = 0.8-1.6$ and $z = 3.2-4.5$, respectively.

The comparison with three published photometric redshifts catalogs (Rafelski et al. 2015; Momcheva et al. 2016; Merlin et al. 2021) confirms the results presented in Brinchmann et al. (2017): the scatter and outlier fraction increase with magnitude and exceed the errors claimed in the corresponding studies. Among the three catalogs tested, the recent ASTRODEEP catalog (Merlin et al. 2021) gives the smallest scatter and the smallest systematic, with $\Delta z = 0.02-0.04$ at $F775W$ AB < 26 and AB > 28.5, respectively.

We performed SED fitting based on the MUSE redshifts and the 11 HST broadband photometry measurements from the Rafelski et al. (2015) catalog. We used two different codes, Magphys (da Cunha et al. 2008, 2015) and Prospector (Johnson et al. 2021), to derive the stellar mass and SFR. Prospector uses the MUSE emission line flux as an additional fitting constraint. Both codes give comparable results, although with some systematic differences and a large scatter for faint galaxies. The MUSE sample probes on average low-mass star-forming galaxies with a median mass of $6.2 \times 10^8\,M_\odot$ and a median SFR of $0.4\,M_\odot \mathrm{yr}^{-1}$.

Twenty percent of our catalog, or 424 galaxies, have no counterpart in the HST source catalogs. While a few of these galaxies are clearly visible in the HST images and were just missed by the SExtractor source de-blending, the vast majority are too faint with respect to the HST limiting magnitude, which is between 29 and 30 AB magnitude depending on the filter. These new sources are high equivalent-width Lyα emitters that are detected by MUSE thanks to their bright and asymmetric broad Lyα line. As expected, the fraction of HST-undetected Lyα emitters is a function of depth, with a sharp increase from 18% in the 10-h depth MOSAIC to 49% in the 140-h depth MXDF. These figures confirm the existence of a large population of faint star-forming Lyα emitters revealed in MUSE deep fields (e.g., Bacon et al. 2015, 2017). As shown by Maseda et al. (2018, 2020), these galaxies are on average very faint ($M_{\mathrm{UV}} \approx -15$) star-forming galaxies with a high ionizing photon production efficiency.

The MXDF is the deepest spectroscopic survey ever performed. We believe it has great legacy value. The MUSE data, with their 3D content, amazing depth, wide spectral range, and excellent spatial and medium spectral resolution, are rich in information. Their location in the UDF area, which benefits from an exquisite collection of ancillary panchromatic information, is a major asset. We have made public the advanced data products, specific software programs, and a web interface to select and download data sets.

Acknowledgements. We warmly thank ESO Paranal staff for their great professional support during all GTO observing runs. We thank the anonymous referee for a careful reading of the manuscript and many suggestion for improving its readability. We thank Marc Rafelski, Katherine Whitekaer, Emiliano Merlin and Adriano Fontana for providing information on their published spectrophotometric catalogs and associated segmented maps. RB, YG acknowledge support from the ANR L-INTENSE (ANR-20-CE92-0015). NB acknowledges the support of the ANR 3DGasFlows (ANR-17-CE31-0017). SLZ acknowledges support by The Netherlands Organisation for Scientific Research (NWO) through a TOP Grant Module 1 under project number 614.001.652.

References

Bacher, R. 2017, Ph.D. Thesis, Université Grenoble Alpes, France
Bacon, R., Brinchmann, J., Richard, J., et al. 2015, A&A, 575, A75
Bacon, R., Conseil, S., Mary, D., et al. 2017, A&A, 608, A1
Bacon, R., Mary, D., Garel, T., et al. 2021, A&A, 647, A107
Bacon, R., Piqueras, L., Conseil, S., Richard, J., & Shepherd, M. 2016, Astrophysics Source Code Library [record ascl:1611.003]
Baldry, I. K., Alpaslan, M., Bauer, A. E., et al. 2014, MNRAS, 441, 2440
Beckwith, S. V. W., Stiavelli, M., Koekemoer, A. M., et al. 2006, AJ, 132, 1729
Benítez, N. 2000, ApJ, 536, 571
Benjamini, Y., & Hochberg, Y. 1995, J. R. Stat. Soc. Ser. B (Methodol.), 57, 289
Bertin, E., & Arnouts, S. 1996, A&AS, 117, 393
Birnbaum, Z. W. 1950, Ann. Math. Stat., 21, 272
Blanton, M., & Lin, H. 2000, ApJ, 543, L125
Blanton, M. R., Bershady, M. A., Abolfathi, B., et al. 2017, AJ, 154, 28
Boogaard, L. A., Brinchmann, J., Bouché, N., et al. 2018, A&A, 619, A27
Boogaard, L. A., Decarli, R., González-López, J., et al. 2019, ApJ, 882, 140
Boogaard, L. A., van der Werf, P., Weiss, A., et al. 2020, ApJ, 902, 109
Boogaard, L. A., Bouwens, R. J., Riechers, D., et al. 2021, ApJ, 916, 12
Bouché, N. F., Genel, S., Pellissier, A., et al. 2021, A&A, 654, A49
Brammer, G. B., van Dokkum, P. G., & Coppi, P. 2008, ApJ, 686, 1503
Brammer, G. B., van Dokkum, P. G., Franx, M., et al. 2012, ApJS, 200, 13
Brinchmann, J., Charlot, S., White, S. D. M., et al. 2004, MNRAS, 351, 1151
Brinchmann, J., Charlot, S., Kauffmann, G., et al. 2013, MNRAS, 432, 2112
Brinchmann, J., Inami, H., Bacon, R., et al. 2017, A&A, 608, A3
Bruzual, G., & Charlot, S. 2003, MNRAS, 344, 1000
Bundy, K., Bershady, M. A., Law, D. R., et al. 2015, ApJ, 798, 7
Byler, N., Dalcanton, J. J., Conroy, C., & Johnson, B. D. 2017, ApJ, 840, 44
Calzetti, D., Armus, L., Bohlin, R. C., et al. 2000, ApJ, 533, 682
Cantalupo, S., Pezzulli, G., Lilly, S. J., et al. 2019, MNRAS, 483, 5188
Cappellari, M. 2017, MNRAS, 466, 798
Carfantan, H. 2014, Modèles, estimateurs et algorithmes pour quelques problèmes inverses de traitement du signal et d'images en sciences de l'univers
Ciardullo, R., Gronwall, C., Adams, J. J., et al. 2013, ApJ, 769, 83
Claeyssens, A., Richard, J., Blaizot, J., et al. 2022, A&A, 666, A78
Conroy, C. 2013, ARA&A, 51, 393
Conroy, C., Gunn, J. E., & White, M. 2009, ApJ, 699, 486
da Cunha, E., Charlot, S., & Elbaz, D. 2008, MNRAS, 388, 1595
da Cunha, E., Walter, F., Smail, I. R., et al. 2015, ApJ, 806, 110
den Brok, J. S., Cantalupo, S., Mackenzie, R., et al. 2020, MNRAS, 495, 1874
Doherty, M., Bunker, A. J., Ellis, R. S., & McCarthy, P. J. 2005, MNRAS, 361, 525
Drake, A. B., Garel, T., Wisotzki, L., et al. 2017, A&A, 608, A6
Duncan, K., Conselice, C. J., Mortlock, A., et al. 2014, MNRAS, 444, 2960
Dunlop, J. S., McLure, R. J., Biggs, A. D., et al. 2017, MNRAS, 466, 861
Feltre, A., Bacon, R., Tresse, L., et al. 2018, A&A, 617, A62
Feltre, A., Maseda, M. V., Bacon, R., et al. 2020, A&A, 641, A118
Finley, H., Bouché, N., Contini, T., et al. 2017, A&A, 608, A7
Fontana, A., D'Odorico, S., Poli, F., et al. 2000, AJ, 120, 2206
Foreman-Mackey, D., Hogg, D. W., Lang, D., & Goodman, J. 2013, PASP, 125, 306
Franco, M., Elbaz, D., Béthermin, M., et al. 2018, A&A, 620, A152
Fusco, T., Bacon, R., Kamann, S., et al. 2020, A&A, 635, A208
Gaia Collaboration (Prusti, T., et al.) 2016, A&A, 595, A1
Gaia Collaboration (Brown, A. G. A., et al.) 2018, A&A, 616, A1
Garilli, B., McLure, R., Pentericci, L., et al. 2021, A&A, 647, A150
Golub, G. H., Heath, M., & Wahba, G. 1979, Technometrics, 21, 215
Guérou, A., Krajnović, D., Epinat, B., et al. 2017, A&A, 608, A5
Guo, Y., Ferguson, H. C., Giavalisco, M., et al. 2013, ApJS, 207, 24
Hashimoto, T., Garel, T., Guiderdoni, B., et al. 2017, A&A, 608, A10
Herenz, E. C., & Wisotzki, L. 2017, A&A, 602, A111
Herenz, E. C., Urrutia, T., Wisotzki, L., et al. 2017, A&A, 606, A12
Herenz, E. C., Wisotzki, L., Saust, R., et al. 2019, A&A, 621, A107
Herenz, E. C., Hayes, M., & Scarlata, C. 2020, A&A, 642, A55
Hildebrandt, H., Viola, M., Heymans, C., et al. 2017, MNRAS, 465, 1454
Hinton, S. R., Davis, T. M., Lidman, C., Glazebrook, K., & Lewis, G. F. 2016, Astron. Comput., 15, 61
Ho, I. T., Medling, A. M., Groves, B., et al. 2016, Ap&SS, 361, 280
Hoerl, A. E., & Kennard, R. W. 1970, Technometrics, 12, 55

Horne, K. 1986, PASP, 98, 609
Husser, T.-O., Kamann, S., Dreizler, S., et al. 2016, A&A, 588, A148
Ilbert, O., Arnouts, S., McCracken, H. J., et al. 2006, A&A, 457, 841
Illingworth, G. D., Magee, D., Oesch, P. A., et al. 2013, ApJS, 209, 6
Inami, H., Bacon, R., Brinchmann, J., et al. 2017, A&A, 608, A2
Inami, H., Decarli, R., Walter, F., et al. 2020, ApJ, 902, 113
Ivezić, Ž., Kahn, S. M., Tyson, J. A., et al. 2019, ApJ, 873, 111
Johnson, B. D., Leja, J., Conroy, C., & Speagle, J. S. 2021, ApJS, 254, 22
Kimock, B., Narayanan, D., Smith, A., et al. 2021, ApJ, 909, 119
Kolb, J., Madec, P. Y., Arsenault, R., et al. 2016, in Adaptive Optics Systems V, eds. E. Marchetti, L. M. Close, & J. P. Véran, SPIE Conf. Ser., 9909, 99092S
Kron, R. G. 1980, ApJS, 43, 305
Kusakabe, H., Blaizot, J., Garel, T., et al. 2020, A&A, 638, A12
Kusakabe, H., Verhamme, A., Blaizot, J., et al. 2022, A&A, 660, A44
Laureijs, R., Amiaux, J., Arduini, S., et al. 2011, ArXiv e-prints [arXiv:1110.3193]
Le Fèvre, O., Cassata, P., Cucciati, O., et al. 2013, A&A, 559, A14
Le Fevre, O., Vettolani, G. P., Maccagni, D., et al. 1998, in Optical Astronomical Instrumentation, ed. S. D'Odorico, SPIE Conf. Ser., 3355, 8
Leclercq, F., Bacon, R., Wisotzki, L., et al. 2017, A&A, 608, A8
Leclercq, F., Bacon, R., Verhamme, A., et al. 2020, A&A, 635, A82
Leibundgut, B., Bacon, R., Jaffé, Y. L., et al. 2017, The Messenger, 170, 20
Leja, J., Carnall, A. C., Johnson, B. D., Conroy, C., & Speagle, J. S. 2019, ApJ, 876, 3
Leja, J., Speagle, J. S., Johnson, B. D., et al. 2020, ApJ, 893, 111
Leja, J., Speagle, J. S., Ting, Y.-S., et al. 2022, ApJ, 936, 165
Lidman, C., Tucker, B. E., Davis, T. M., et al. 2020, MNRAS, 496, 19
Loncan, L., Almeida, L., Bioucas-Dias, J. M., et al. 2016, IEEE Geosci. Remote Sens. Maga., 3, 27
Luo, B., Brandt, W. N., Xue, Y. Q., et al. 2017, ApJS, 228, 2
Madec, P. Y., Arsenault, R., Kuntschner, H., et al. 2018, in Adaptive Optics Systems VI, eds. L. M. Close, L. Schreiber, & D. Schmidt, SPIE Conf. Ser., 10703, 1070302
Mary, D., & Roquain, E. 2022, Electron. J. Statist., 16, 4926
Mary, D., Bacon, R., Conseil, S., Piqueras, L., & Schutz, A. 2020, A&A, 635, A194
Maseda, M. V., van der Wel, A., Rix, H.-W., et al. 2014, ApJ, 791, 17
Maseda, M. V., Brinchmann, J., Franx, M., et al. 2017, A&A, 608, A4
Maseda, M. V., Bacon, R., Franx, M., et al. 2018, ApJ, 865, L1
Maseda, M. V., Bacon, R., Lam, D., et al. 2020, MNRAS, 493, 5120
McLinden, E. M., Finkelstein, S. L., Rhoads, J. E., et al. 2011, ApJ, 730, 136
Merlin, E., Castellano, M., Santini, P., et al. 2021, A&A, 649, A22
Mignoli, M., Cimatti, A., Zamorani, G., et al. 2005, A&A, 437, 883
Moffat, A. F. J. 1969, A&A, 3, 455
Momcheva, I. G., Brammer, G. B., van Dokkum, P. G., et al. 2016, ApJS, 225, 27
Muzahid, S., Schaye, J., Marino, R. A., et al. 2020, MNRAS, 496, 1013
Nanayakkara, T., Brinchmann, J., Boogaard, L., et al. 2019, A&A, 624, A89
Newman, J. A., Abate, A., Abdalla, F. B., et al. 2015, Astropart. Phys., 63, 81
Newville, M., Stensitzki, T., Allen, D. B., & Ingargiola, A. 2014, https://doi.org/10.5281/zenodo.11813
Pedregosa, F., Varoquaux, G., Gramfort, A., et al. 2011, J. Mach. Learn. Res., 12, 2825
Piqueras, L., Conseil, S., Shepherd, M., et al. 2019, in Astronomical Data Analysis Software and Systems XXVI, eds. M. Molinaro, K. Shortridge, & F. Pasian, ASP Conf. Ser., 521, 545
Rafelski, M., Teplitz, H. I., Gardner, J. P., et al. 2015, AJ, 150, 31
Rakic, O., Schaye, J., Steidel, C. C., & Rudie, G. C. 2011, MNRAS, 414, 3265
Ribeiro, B., Le Fèvre, O., Paulino-Afonso, A., et al. 2020, A&A, submitted [arXiv:2007.01322]
Saito, S., de la Torre, S., Ilbert, O., et al. 2020, MNRAS, 494, 199
Sánchez, S. F., Lacerda, E., Mejía-Narvaez, A., et al. 2021, Extragalactic Spectroscopic Surveys: Past, 29
Sarzi, M., Falcón-Barroso, J., Davies, R. L., et al. 2017, Astrophysics Source Code Library [record ascl:1708.012]
Schmidt, K. B., Wisotzki, L., Urrutia, T., et al. 2019, A&A, 628, A91
Schmidt, K. B., Kerutt, J., Wisotzki, L., et al. 2021, A&A, 654, A80
Schwarz, G. 1978, Ann. Stat., 6, 461
Shapley, A. E., Steidel, C. C., Pettini, M., & Adelberger, K. L. 2003, ApJ, 588, 65
Shibuya, T., Ouchi, M., Nakajima, K., et al. 2014, ApJ, 788, 74
Skelton, R. E., Whitaker, K. E., Momcheva, I. G., et al. 2014, ApJS, 214, 24
Song, M., Finkelstein, S. L., Gebhardt, K., et al. 2014, ApJ, 791, 3
Soto, K. T., Lilly, S. J., Bacon, R., Richard, J., & Conseil, S. 2016, MNRAS, 458, 3210
Speagle, J. S. 2020, MNRAS, 493, 3132
Speagle, J. S., Steinhardt, C. L., Capak, P. L., & Silverman, J. D. 2014, ApJS, 214, 15
Spergel, D., Gehrels, N., Baltay, C., et al. 2015, ArXiv e-prints [arXiv:1503.03757]
Steidel, C. C., Adelberger, K. L., Shapley, A. E., et al. 2000, ApJ, 532, 170
Tibshirani, R. 1996, J. R. Stat. Soc. Ser. B (Methodological), 58, 267
Tremonti, C. A., Heckman, T. M., Kauffmann, G., et al. 2004, ApJ, 613, 898
Urrutia, T., Wisotzki, L., Kerutt, J., et al. 2019, A&A, 624, A141
Vanzella, E., Cristiani, S., Dickinson, M., et al. 2008, A&A, 478, 83
Vanzella, E., Balestra, I., Gronke, M., et al. 2017, MNRAS, 465, 3803
Ventou, E., Contini, T., Bouché, N., et al. 2017, A&A, 608, A9
Verhamme, A., Garel, T., Ventou, E., et al. 2018, MNRAS, 478, L60
Villeneuve, E., Carfantan, H., & Serre, D. 2011, in 3rd Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS) (IEEE), 1
Virtanen, P., Gommers, R., Oliphant, T. E., et al. 2020, Nat. Methods, 17, 261
Weilbacher, P. M., Palsa, R., Streicher, O., et al. 2020, A&A, 641, A28
Wells, D. C., Greisen, E. W., & Harten, R. H. 1981, A&AS, 44, 363
Wesson, R. 2016, MNRAS, 456, 3774
Whitaker, K. E., Ashas, M., Illingworth, G., et al. 2019, ApJS, 244, 16
Wisotzki, L., Bacon, R., Blaizot, J., et al. 2016, A&A, 587, A98
Wisotzki, L., Bacon, R., Brinchmann, J., et al. 2018, Nature, 562, 229

[1] Univ Lyon, Univ Lyon1, ENS de Lyon, CNRS, Centre de Recherche Astrophysique de Lyon UMR5574, 9 avenue Charles André, 69230 Saint-Genis-Laval, France
e-mail: roland.bacon@univ-lyon1.fr
[2] Instituto de Astrofísica e Ciências do Espaço, Universidade do Porto, CAUP, Rua das Estrelas, 4150-762 Porto, Portugal
[3] Aix-Marseille Université, CNRS, CNES, LAM (Laboratoire d'Astrophysique de Marseille) UMR 7326UMR7326, 38 rue Frédéric Joliot-Curie, 13388 Marseille, France
[4] Department of Astronomy, University of Wisconsin, 475 N. Charter Street, Madison, WI 53706, USA
[5] Centre for Astrophysics and Supercomputing, Swinburne University of Technology, Melbourne, VIC 3122, Australia
[6] Leibniz-Institut für Astrophysik Potsdam (AIP), An der Sternwarte 16, 14482 Potsdam, Germany
[7] Institut für Physik und Astronomie, Universität Potsdam, Karl-Liebknecht-Str. 24/25, 14476 Golm, Germany
[8] Univ. Grenoble Alpes, CNRS, Grenoble INP, GIPSA-lab, 11 rue des Mathématiques, Grenoble Campus BP46, 38402 Saint-Martin-d'Hères Cedex, France
[9] Laboratoire Lagrange, CNRS, Université Côte d'Azur, Boulevard de l'Observatoire, CS34229, 06304 Nice, France
[10] Department of Astronomy, University of Texas at Austin, 2515 Speedway, Austin, TX 78712, USA
[11] INAF – Osservatorio di Astrofisica e Scienza dello Spazio di Bologna, Via P. Gobetti 93/3, 40129 Bologna, Italy
[12] Observatoire de Genève, Université de Genève, 51 Ch. des Maillettes, 1290 Versoix, Switzerland
[13] Leiden Observatory, Leiden University, PO Box 9513, 2300 RA Leiden, The Netherlands
[14] IRAP, Institut de Recherche en Astrophysique et Planétologie, CNRS, Université de Toulouse, 14, avenue Edouard Belin, 31400 Toulouse, France
[15] European Southern Observatory, Av. Alonso de Córdova 3107, 763 0355 Vitacura, Santiago, Chile
[16] Max Planck Institute for Astronomy, Königstuhl 17, 69117 Heidelberg, Germany
[17] Institut für Astrophysik, Universität Göttingen, Friedrich-Hund-Platz 1, 37077 Göttingen, Germany
[18] Canada-France-Hawaii Telescope, CNRS, 96743 Kamuela, Hawaii, USA

Appendix A: Data products and software releases

In this section we give practical information on the released data sets, namely images and datacubes, catalogs and source files. We also describe the features of the AMUSED web interface. Finally, we give the list of all specific software developed or updated in the context of this survey.

A.1. Images and datacubes

For each data set (i.e., MXDF, UDF-10, and MOSAIC) we make the final datacubes and associated images public. These data are available for download through the AMUSED web interface and are described in Table A.1. Images and datacubes are in the multi-extension FITS format, with a data and optionally a variance extension. Masked data have NaN values. The primary header of each file contain a WCS extension and a number of keywords, including the spatial PSF model polynomial coefficients (Sect. 4.2.1). The file format is described in detail in the corresponding MPDAF documentation[21]. These files can be used directly with any FITS reading tool (e.g., fitsio, ds9), but we recommend using MPDAF, which contains an easier and advanced manipulation of these files.

In addition we also provide a fits table DR2_SENSITIVITY.fits with the median surface brightness and point-source limiting flux (Sect. 4.3) as a function of wavelength. Column descriptions are given in Table A.2.

A.2. Catalogs

The main catalog gives a summary of source properties. It contains a subset of the main measured properties (i.e., the main emission and absorption line flux) and can be used standalone. For the full set of measurements three additional tables are given: the redshifts table, the line table and the narrow-band table. Catalogs are publicly available via the AMUSED web interface (Sect. A.4).

The main catalog is a table with 2221 entries, one line for each source. Sources have a unique MUSE identifier (MID). We note that sources already detected in DR1 have kept their identifier, except when the redshift or matching information was updated. In that case a new identifier was given. The catalog description is given in Table A.4.

The redshift and line tables that result from the fitting performed on the reference spectrum by pyPlatefit are described in Tables A.5 and A.6, respectively. The redshift table gives the common parameters (i.e., the redshift and the velocity dispersion) fitted for each line set. Some statistical (e.g., maximum S/N) and fitting (e.g., χ^2) information is also given. The line table contains detailed fitting information for each line.

The narrow-band table is described in Table A.7 and contains information for all narrow bands derived from emission or absorption, single or combined lines.

The list of DR1 sources that have been renamed in DR2 because of a new redshift assignment (Sect. 6.1.2) is part of the released data products. Its column description is given in Table A.8.

A.3. Sources

Each source identified in the DR2 catalog is available as a specific *Source* file, a multi-extension FITS file that gathers all source information in a single file (Wells et al. 1981). The *Source* format is part of the MUSE MPDAF (Bacon et al. 2016; Piqueras et al. 2019). A DR2 *Source* file contains generic information related to the source (e.g., its identifier and celestial coordinates), small images (e.g., white-light and narrow bands) and datacubes centered at the source location, spectra for various extraction scheme and tables (e.g., emission and absorption lines information).

The "Source" Python class is described in the MPDAF online documentation. A tutorial specific to the DR2 sources is also available online[22]. We give in Tables A.9,A.10,A.11, and A.12 the full description of the source content.

The DR2 source files, in addition of providing direct access to the datacube, spectrum, narrowband and white-light images contents for a given source, give access to a number of important ancillary information. For example, the sources contains spectra derived from different extractions. We note that all parameters provided in the catalogs have been computed with a reference extraction method, given in the main catalog shown in Table A.4, column `FROM`. However, this default extraction may not be optimal for a given science case as already discussed in Sect. 5.8.1, and thus we provide all other available extractions in the source file. In the following we give our prescription for the extraction method to use in some typical conditions.

ODHIN extraction should be preferred for sources with detectable continuum, or when the continuum and absorption line information is important (e.g., passive galaxies). It is also the best method when the source blending is substantial and all neighboring sources are present in the HST segmentation map.

ORIGIN extraction is obviously recommended for galaxies undetected in HST. It is also best for faint line emitters as it provides generally the highest S/N. It also works better when the emission flux is spatially extended (e.g., Lyα emitters).

NBEXT extraction is an alternative to ORIGIN extraction. Although it provides generally lower S/N than the ORIGIN extraction, it should be used in the few cases where the PCA continuum subtraction impacts the pseudo-narrow band. It should also be used in some case of close pairs of line emitters, when the ORIGIN resulting pseudo-narrowband, which implies a PSF convolution (Mary et al. 2020), is not able to distinguish between the two sources.

When available, we also give the spectrum derived from alternative data sets (i.e., MOSAIC extraction for a source with UDF-10 or MXDF reference data sets). pyPlatefit information is given in the form of tables (PL_LINES and PL_Z) and spectra (named with the "PL_" prefix).

Some additional information for the ORIGIN detected sources is given in the ORI_LINES Table A.15. The SED fitting information (Sect. 6.4) is optionally available in the SED Table A.19.

The sources also contain catalog information on surrounding HST objects derived from the 3 HST catalogs (R15, CANDELS v2 and 3D-HST), ORIGIN detected and DR2 neigbouring sources (named with "_CAT" suffix). If available, spectra of neighboring sources derived from ODHIN R15 extraction, are also given.

A.4. The AMUSED web interface

AMUSED is a public web interface for inspection and retrieval of MUSE data products. The data are organized in a Post-

[21] https://mpdaf.readthedocs.io/en/latest/obj.html

[22] https://amused.univ-lyon1.fr/project/UDF/HUDF/help

Table A.1. Released datacubes and images.

File name	Description	data set	Type	Size	Version
DATACUBE_MXDF.fits	Main datacube	MXDF	cube	6.9 Gb	3.0
IMAGE_MXDF.fits	White-light image	MXDF	image	2.0 Mb	3.0
EXPMAP_MXDF.fits	Exposure map datacube	MXDF	cube	3.1 Gb	3.0
EXPMAP-IMAGE_MXDF.fits	Exposure map image	MXDF	image	0.9 Mb	3.0
DATACUBE_UDF-10.fits	Main datacube	UDF-10	cube	2.9 Gb	1.1
IMAGE_UDF-10.fits	White-light image	UDF-10	image	0.9 Mb	1.1
EXPMAP_UDF-10.fit	Exposure map datacube	UDF-10	cube	1.5 Gb	1.1
EXPMAP-IMAGE_UDF-10.fits	Exposure map image	UDF-10	image	0.4 Mb	1.1
DATACUBE_MOSAIC.fits	Main datacube	MOSAIC	cube	25 Gb	1.1
IMAGE_MOSAIC.fits	White-light image	MOSAIC	image	6.9 Mb	1.1
EXPMAP_MOSAIC.fit	Exposure map datacube	MOSAIC	cube	13 Gb	1.1
EXPMAP-IMAGE_MOSAIC.fits	Exposure map image	MOSAIC	image	3.5 Mb	1.1
FIELDMAP_UDF-MOSAIC.fits	Field map image	MOSAIC	image	6.9 Mb	1.1

Notes. The flux unit of datacubes and white light images is 10^{-20} erg s^{-1} cm^{-2} Å^{-1}. Variance is given in the square of the flux unit. Units of exposure map images and datacubes are exposure counts, each of 25 mn duration. The exposure map image is obtained by taking the median over the wavelength axis. The MOSAIC field map image codes the sub-fields identifier (1-9) used in the combination at each spaxel location (see MPDAF fieldmap documentation (https://mpdaf.readthedocs.io/en/latest/muse.html#muse-mosaic-field-map)). The version refer to the version of the data reduction process.

Table A.2. Description of the columns of the sensitivity table.

Col. name	Description
WAVE	Wavelength in Å.
SB_MOSAIC	Median surface brightness 1σ limiting emission line flux for the MOSAIC data set (10-hour depth).
SB_UDF10	Median surface brightness 1σ limiting emission line flux for the UDF-10 data set (31-hour depth).
SB_MXDF	Median surface brightness 1σ limiting emission line flux for the MXDF data set (141-hour depth).
PS_MOSAIC	Median point-source 1σ limiting emission line flux for the MOSAIC data set (10-hour depth).
PS_UDF10	Median point-source 1σ limiting emission line flux for the UDF-10 data set (31-hour depth).
PS_MXDF	Median point-source 1σ limiting emission line flux for the MXDF data set (141-hour depth).

Notes. Computations are performed for an unresolved emission line. Surface brightness and point-source 1σ limiting flux are given in erg s^{-1} cm^{-2} arcsec^{-2} and erg s^{-1} cm^{-2} units, respectively.

Table A.3. Released tables.

File name	Description	Size	Rows	Version	Columns desc.
dr2_main_10.fits	Main source catalog	3.1 Mb	2221	0.9	Table. A.4
dr2_z_10.fits	Redshift catalog	0.9 Mb	6722	0.9	Table. A.5
dr2_lines_10.fits	Emission and absorption lines catalog	13 Mb	56968	0.9	Table. A.6
dr2_nb_10.fits	Narrowband catalog	4.2 Mb	11938	0.9	Table. A.7
dr2_noise_10.fits	Surface brightness and point source limiting emission flux	0.2 Mb	3705	0.9	Table. A.2
dr1_dr2_10.fits	DR1 sources that have been renamed	12 Kb	39	0.9	Table. A.8

These tables are only available in electronic form at the CDS via anonymous ftp to cdsarc.u-strasbg.fr (130.79.128.5) or via http://cdsweb.u-strasbg.fr/cgi-bin/qcat?J/A+A/ or via the AMUSED interface (Sect. A.4). We also provide another format for the fits table, named "*_astropy_*." These tables are astropy fits tables with masked values (https://docs.astropy.org/en/stable/api/astropy.table.MaskedColumn.html).

gresSQL relational database[23]. User can perform advanced source selection: for example, selecting all MUSE sources with Lyα flux higher than 8×10^{-19} erg s^{-1} cm^{-2} and S/N > 7, redshift in the $z = 3 - 4$ range, and without HST counterparts (Fig. A.1). The resulting list can be visually inspected by looking to spectra, emission and absorption lines fit, reconstructed broadband and narrowband images and ancillary information (e.g., HST images). This process allows users to refine the selection, possibly taking notes for sources of interest, and then download the selected data. Exported data can be tables (Sect. A.2) in CSV or FITS format, interactive html visualization files and sources files in the MPDAF multi-fits format (Sect. A.3).

The user can also download the full data set without the need to inspect the sources. The final reduced datacubes for MXDF, UDF-10, and MOSAIC are also make available (Sect. A.1). AMUSED is accessible at https://amused.univ-lyon1.fr.

[23] www.postgresql.org

Table A.4. Description of the columns of the main table.

Col. name	Description
ID	MUSE source identifier (int)
DATASET	MUSE data set[a]
DEPTH	exposure depth at source location in hours
FROM	spectrum extraction type[b] (Sect. 5.8.1)
ZCONF	redshift confidence: 1 (low) – 3 (high) (Sect. 5.7.3)
MCONF	matching confidence: 0–3, (Sect. 5.7.4)
IFLAG	isolation flag: 1–3, (Sect. 5.7.4)
ZSYS	systemic redshift in vacuum [c] (Sect.5.8.3)
ZSYS_ERR	error in systemic redshift (Sect.5.8.3)
REFZ	reference redshift line set [d] (Sect.5.5
Z	reference redshift value in vacuum
Z_ERR	error in reference redshift
DLYAFIT	flag to indicate double Lyα fit
DV_ttt	velocity offset with respect to reference redshift for redshift type ttt (km s^{-1})
DV_ERR_ttt	velocity offset with respect to reference redshift for redshift type ttt (km s^{-1})
RA	right ascension (J2000 degree), see astrometry Sect. 4.1
DEC	declination (J2000 degree) see astrometry Sect. 4.1
CENTER	reference center [e] (Sect.5.6)
IN_HST	HST matching flag [f]
IN_ORI	ORIGIN matching boolean flag
IN_DR1	DR1 matching boolean flag
IN_MXDF	source is located in MXDF footprint (bool)
IN_UDF10	source is located in UDF-10 footprint (bool)
RAF_ID	R15 catalog unique matched ID (int)
RAF_MIDS	R15 catalog multiple matched IDs [g]
CANDELS_ID	CANDELS v2 catalog unique matched ID (int)
CANDELS_MIDS	CANDELS v2 catalog multiple matched IDs [g]
C3DHST_ID	3D-HST catalog unique matched ID (int)
C3DHST_MIDS	3D-HST catalog multiple matched IDs [g]
ASTRO_ID	ASTRODEEP catalog unique matched ID (int)
ASTRO_MIDS	ASTRODEEP catalog multiple matched IDs [g]
MAG_SRC	source of magnitude [h]
MAG_FLAG	contamination flag (APER magnitude only) [i]
MAG_xxx	broadband AB magnitude in xxx HST filter [j]
MAGERR_xxx	AB magnitude error in xxx HST filter [k]

Notes. [a]MXDF,UDF10 or MOSAIC [b]ORIGIN,ODHIN or NBEXT [c]$Z_{sys} \equiv Z$, except for for simple peak Lyα emitters [d]BALMER,FORBIDDEN,LYALPHA,ABS,CIV1548 or MGII2796 [e]3DHST,CANDELS,CUSTOM,NB_EMI,ORIGIN or RAFELSKI [f]Ambiguous, Detected, Faint, Missed or Undetect [h]Coded as text with comma separator (e.g., 23,567) [h]APER,3DHST,CANDELS or RAF. Aperture photometry (APER) is used for undetected HST source (Sect. 6.2) [i]If true indicate source contamination [j]Filters are F606W, F775W or F850LP [k]If MAGERR_xxx < 0, then MAG_xxx ≡ noise stdev

Table A.4. Continued.

Col. name	Description
MASS_ff	log $M/M_\odot$ where M is the stellar mass derived from the ff[a] SED fit (Sect.6.4).
LERR_MASS_ff	Lower 1σ percentile of log $M/M_\odot$
HERR_MASS_ff	Upper 1σ percentile of log $M/M_\odot$
SFR_ff	log SFR/$M_\odot$yr^{-1} where SFR is the star formation rate at 100 Myr look-back time as derived from the ff[a] SED fit (Sect.6.4).
LERR_SFR_ff	Lower 1σ percentile of log SFR/$M_\odot$yr^{-1}
HERR_SFR_ff	Upper 1σ percentile of log SFR/$M_\odot$yr^{-1}
LINE_SNR_MAX	name of emission or absorption line with max S/N [b]
SNR_MAX	max S/N
FLUX_MAX	flux of the line with max S/N (10^{-20}erg s^{-1} cm^{-2})
lll_EMI_FLUX	flux of the lll[c] emission line (10^{-20}erg s^{-1} cm^{-2})
lll_EMI_SNR	S/N of the lll emission line
lll_EMI_EQW	Rest frame equivalent width of the lll emission line (Å)
lll_EMI_VD	Rest frame velocity dispersion[d] of the lll emission line (km s^{-1})
lll_ABS_FLUX	flux of the lll[i] absorption line (10^{-20}erg s^{-1} cm^{-2})
lll_ABS_SNR	S/N of the lll absorption line
lll_ABS_EQW	Rest frame equivalent width of the lll absorption line (Å)
lll_ABS_VD	Rest frame velocity dispersion[c] of the lll absorption line (km s^{-1})

Notes. [a]ff is PRO for Prospector and MAG for Magphys. [b]b at the end of line name indicate a blend (e.g., OII3727b is the sum of the [O II]$\lambda\lambda$3726,3729 doublet) [c]see table D.1 for line names [d]corrected for instrumental velocity dispersion

Table A.5. Description of the columns of the redshift table.

Col. name	Description
ID	MUSE source identifier
DATASET	MUSE data set[a]
LINESET	line set[b]
Z	redshift in vacuum
Z_ERR	error in redshift
VEL	velocity offset with respect to reference redshift (km s^{-1})
VEL_ERR	velocity offset error (km s^{-1})
VDISP	rest-frame velocity dispersion (km s^{-1})
VDISP_ERR	error in velocity dispersion (km s^{-1})
LINE	name of line with the highest S/N
SNRMAX	maximum S/N for the fitted lines
SNRSUM	total S/N for all lines
SNRSUM_CLIPPED	total S/N for lines with S/N > 3
NL	number of fitted lines
NL_CLIPPED	number of fitted lines with S/N > 3
RCHI2	returned reduced χ^2 by the minimization routine
STATUS	returned status of the fitting function[c]

Notes. [a]MXDF,UDF10 or MOSAIC [b]BALMER,FORBIDDEN,LYALPHA, ABS,CIV548 or MGII2796 [c]LMFIT

Table A.6. Description of the columns of the line table.

Col. name	Description
ID	MUSE source identifier
DATASET	MUSE data set[a]
LINESET	line set[b]
LINE	line identifier (e.g., LYALPHA)
LBDA_REST	rest wavelength (Å)
DNAME	display name for the line (e.g., Lyα), set to None for close doublets.
FLUX	total line flux (10^{-20}erg s^{-1} cm^{-2})
FLUX_ERR	line flux error (10^{-20}erg s^{-1} cm^{-2})
SNR	line S/N
Z	redshift in vacuum
Z_ERR	error in redshift
VDISP	rest-frame velocity dispersion (km s^{-1})
VDISP_ERR	error in velocity dispersion (km s^{-1})
SKEW	skewness parameter [c] (Eq. 4)
SKEW_ERR	skewness error [c]
SEP	separation between the two peaks [d] (km s^{-1})
SEP_ERR	error in peak separation [d] (km s^{-1})
VDINST	instrumental velocity dispersion (km s^{-1})
LBDA_OBS	fitted position of the line peak in observed frame (Å)
PEAK_OBS	maximum flux of the line peak in observed frame (10^{-20}erg s^{-1} cm^{-2} Å^{-1})
LBDA_LEFT	observed wavelength at the left of the peak with half peak value (Å)
LBDA_RIGHT	observed wavelength at the right of the peak with half peak value (Å)
FWHM_OBS	full width at half maximum of the line in the observed frame (Å)
EQW	restframe line equivalent width (Å)
EQW_ERR	restframe line equivalent width error (Å)
CONT_OBS	continuum mean value in observed frame (10^{-20}erg s^{-1} cm^{-2} Å^{-1})
CONT	continuum mean value in rest frame (10^{-20}erg s^{-1} cm^{-2} Å^{-1})
CONT_ERR	error in continuum mean value in rest frame (10^{-20}erg s^{-1} cm^{-2} Å^{-1})
NTSD	log$_{10}$ of the line fit relative error

Notes. [a]MXDF,UDF10 or MOSAIC [b]balmer, forbidden, lya, abs, civ1548 or mgii2796 [c]restricted to Lyα line [d]restricted to double peaked Lyα line fit

A.5. Software

In Table A.20 we give the list of software developed in the context of this data release. We make those who are stable enough and at least partially documented available on the musevlt github directory[24]. All software programs are python 3 modules.

The released software should work for any other MUSE data sets. Some may require some minor adaptation for non-deep-field observations (e.g., ORIGIN). Although some of these soft-

[24] https://github.com/musevlt

Table A.7. Description of the columns of the narrowband table.

Col. name	Description
ID	MUSE source identifier
DATASET	MUSE data set[a]
LINE	narrowband name[b] (e.g., NB_EMI_OII3727)
COMBINED	boolean to indicate if the image is a combination of multiple narrow bands
LINES	list of emission lines combined together (e.g., OII3726,OII3729)
DNAME	display name for the line (e.g., Lyα)
SNR_LINES	lines total S/N
FLUX[c]	narrowband flux measured over the segmented area (10^{-20}erg s^{-1} cm^{-2})
FLUX_ERR	narrowband flux error using datacube variance (10^{-20}erg s^{-1} cm^{-2})
FLUX_ISO_ERR	narrowband flux error from SExtractor "isocontour" mode (10^{-20}erg s^{-1} cm^{-2})
SNR_NB[c]	narrowband S/N derived from FLUX/FLUX_ERR
FLUX_AUTO	narrowband flux from SExtractor "automatic" mode (10^{-20}erg s^{-1} cm^{-2})
FLUX_AUTO_ERR	narrowband flux error from SExtractor "automatic" mode (10^{-20}erg s^{-1} cm^{-2})
LBDA_INF	lower limits[d] in wavelength used for narrowband computation (Å)
LBDA_SUP	upper limits[d] in wavelength used for narrowband computation (Å)
RA	right ascension of narrowband barycenter (J2000 degree)
DEC	declination of narrowband barycenter (J2000 degree)
OFFSET	offset (arcsec) with respect to source reference center
AREA	segmented area (arcsec^2)
PA	principal axis (°)
MAJAX	major axis (arcsec)
MINAX	minor axis (arcsec)
ELL	ellipticity
KRON	KRON radius (arcsec)
NSEG	number of segments

Notes. [a]MXDF,UDF10 or MOSAIC [b]The names NB_EMI_COMBINED or NB_ABS_COMBINED are used for the optimal combination of emission or absorption lines (see Sect. 5.6) [c]These values are obtained over the narrowband-segmented area, they are generally different from the values reported in the lines table. [d]This is a list if more than one interval was used for the narrowband computation

Table A.8. Description of the columns of the renamed DR1 sources table.

Col. name	Description
DR1_ID	Previous DR1 source identifier
DR1_Z	Previous DR1 redshift
DR1_ZCONF	Previous DR1 redshift confidence
DR2_ID	New DR2 source identifier
DR2_Z	New DR2 redshift
DR2_ZCONF	New DR2 redshift confidence
DR2_COM	Comment

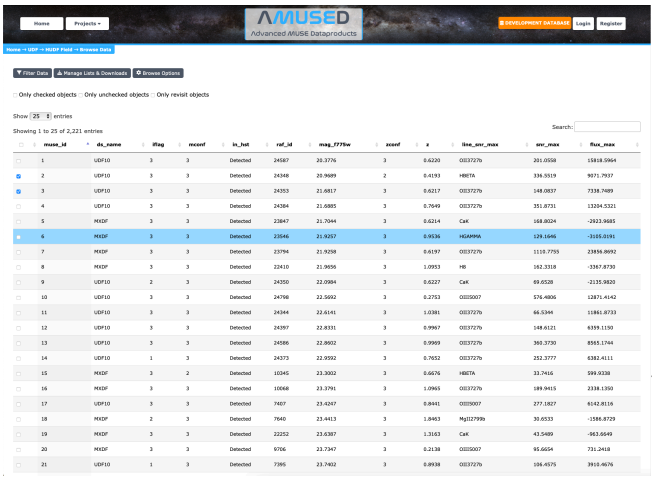


Fig. A.1. AMUSED web interface source selection window.

Table A.9. Source content: header.

Name	Description	Req.
ID	MUSE identifier	y
DATASET	data set (MXDF, UDF10 or MOSAIC)	y
RA	source right ascension (degree)	y
DEC	source declination (degree)	y
FROM	name of software used in source creation	y
FROM_V	version of software used in source creation	y
CUBE	MUSE datacube name	y
CUBE_V	version of MUSE datacube	y
SIZE	source square size (arcsec)	y
CATALOG	name of input catalog	y
ZCONF	redshift confidence (1-3)	y
DEPTH	average depth in hours at source location	y
EXPMEAN	average number of exposures at source location	y
EXPMIN	minimum number of exposures at source location	y
EXPMAX	maximum number of exposures at source location	y
FSFMODE	FSF (spatial PSF) model id (=2)	y
FSFLB1	FSF blue normalization wavelength	y
FSFLB2	FSF red normalization wavelength	y
FSF00FNC	FSF FWHM number of polynomial coef (=4)	y
FSF00Fxx	FSF FWHM polynomial coef value xx=00..03	y
FSF00BNC	FSF β number of polynomial coef (=1)	y
FSF00Bxx	FSF β polynomial coef value xx=00	y
REFSPEC	name of reference spectrum	y
REFZ	reference redshift	y
REFCENTER	identifier of reference center	y
ccc_ID	matched source ID for catalog ccc	n
ccc_RA	matched source RA for catalog ccc	n
ccc_DEC	matched source DEC for catalog ccc	n
ccc_OFF	matched source offset (arcsec) with respect to source center	n

Notes. ccc: catalog ORI (origin) RAF (Rafelski), CANDELS, C3DHST (3D-HST) Req: required y (yes) or n (no)

ware programs are very specific to MUSE (e.g., muse-psfr), most of them should also work with non-MUSE IFU data, provided that the MPDAF format is used.

Table A.10. Source content: spectra.

Name	Description	Req.
ORI_ds_id	ORIGIN extraction for ds data set, id is the ORIGIN ID	n
RAF_ds_id	ODHIN extraction based on Rafelski segmentation map for ds data set, id is the Rafelski ID	n
CAN_ds_id	ODHIN extraction based on CANDELS segmentation map for ds data set, id is the CANDELS ID	n
C3D_ds_id	ODHIN extraction based on 3D-HST segmentation map for ds data set, id is the 3D-HST ID	n
EMI_ds_id	NBEXT extraction based on narrowband segmentation map for ds data set, id is the MUSE ID	n
PL_FIT	pyPlatefit reference spectrum full fit (line emission + model continuum)	y
PL_CONT	pyPlatefit reference spectrum model continuum fit	y
PL_LINE	pyPlatefit continuum subtracted reference spectrum	y
PL_LINEFIT	pyPlatefit reference spectrum continuum subtracted fit	y
PL_FITP	pyPlatefit reference spectrum full fit (line emission and absorption + polynomial continuum)	n
PL_ABSINIT	pyPlatefit reference spectrum after subtraction of fitted emission lines	n
PL_ABSCONT	pyPlatefit reference spectrum polynomial continuum fit	n
PL_ABSLINE	pyPlatefit continuum and emission line subtracted reference spectrum	n
PL_ABSFIT	pyPlatefit reference spectrum absorption lines fit	n

Notes. ds: MXDF, UDF10 or MOSAIC. id: catalog identifier. Req: required y (yes) or n (no)

Table A.11. Source content: images.

Name	Description	Req.
MUSE_WHITE	reconstructed white-light source image	y
MUSE_EXPMAP	exposure map source image (in number of exposure)	y
ORI_CORR_REF	reference correlation image for ORIGIN	n
ORI_CORR_SEG	reference segmentation image for ORIGIN	n
ORI_CORR_lid	correlation image for ORIGIN line lid	n
ORI_MAXMAP	Maximum of correlation image for all ORIGIN sources	n
HST_ff	HST image in filter ff centered on source	n
HST_SEGRAF	HST segmentation image for Rafelski catalog and centered on source	n
HST_SEGCAN	HST segmentation image for CANDELS catalog and centered on source	n
HST_SEG3D	HST segmentation image for 3D-HST catalog and centered on source	n
NB_EMI_lll	narrowband source image for lll emission line	y
NB_ABS_lll	narrowband source image for lll absorption line	n
SEG_EMI_lll	narrowband source segmentation image for lll emission line	y
SEG_ABS_lll	narrowband source image for lll absorption line	n
NB_EMI_COMBINED	combined emission lines narrowband image	n
NB_ABS_COMBINED	combined absorption lines narrowband image	n
SEG_EMI_COMBINED	combined emission lines narrowband segmentation image	n
SEG_ABS_COMBINED	combined absorption lines narrowband segmentation image	n

Notes. ff: F435W, F606W, F775W, F850LP or F160W lll: narrowband line name (see table D.1). Req: required y (yes) or n (no)

Table A.12. Source content: tables.

Name	Description	Req[a]
PL_Z	table of fitted redshifts and related information for the source (columns description[b] in Table 8)	y
PL_LINES	table of fitted lines parameters for the source (columns description[b] in Table A.6)	y
NB_PAR	table of narrow bands for the source (columns description[b] in Table A.7)	y
ORI_LINES	table of ORIGIN line detections for the source (columns description in Table A.13)	n
DR2_CAT	table of all DR2 sources in the source field of view (columns description in Table A.14)	y
ORIG_CAT	table of all ORIGIN sources in the source field of view (columns description in Table A.15)	n
HST_CAT	table of all HST RAFELSKI sources in the source field of view (columns description in Table A.16)	n
CANDELS_CAT	table of all HST CANDELS sources in the source field of view (columns description in Table A.17)	n
HST3D_CAT	table of all HST 3D-HST sources in the source field of view (columns description in Table A.18)	n
SED	table with Magphys and Prospector SED fit (columns description in Table A.19)	n

Notes. [a]Required: y (yes) or n (no). [b]Note the omission of the two columns ID and data set in the source tables with respect to the main tables.

Table A.13. Column description of source table ORI_LINES

Col. name	Description
ID	ORIGIN source identifier
num_line	ORIGIN line identifier
ra	right ascension (°)
dec	declination (°)
lbda	wavelength (Å)
comp	matched continuum segment
T_GLR	correlation peaked value
STD	S/N peaked value
purity	purity estimate

Table A.14. Column description of source table DR2_CAT

Col. name	Description
ID	MUSE source identifier
DATASET	MUSE data set[a]
Z	redshift in vacuum
ZCONF	redshift confidence (1-3)
RA	right ascension (°)
DEC	declination (°)
ORI_ID	matched ORIGIN identifier
RAF_ID	matched RAFELSKI identifier
CANDELS_ID	matched CANDELS identifier
C3DHST_ID	matched HST-3D identifier
DIST	offset from source center (arcsec)

Notes. [a]MXDF, UDF10, or MOSAIC

Table A.15. Column description of source table ORIG_CAT.

Col. name	Description
ID	ORIGIN source identifier
ra	right ascension (°)
dec	declination (°)
waves	list of detected wavelengths (Å)
T_GLR	correlation peaked value
STD	S/N peaked value
purity	purity estimate
DIST	offset from source center (arcsec)

Table A.16. Column description of source table HST_CAT.

Col. name	Description
ID	RAFELSKI source identifier
RA	right ascension (°)
DEC	declination (°)
MAG_F775W	AB F775W magnitude
Z_BPZ	Bayesian photometric redshift
ZMIN_BPZ	lower limit for Bayesian photometric redshift
ZMAX_BPZ	upper limit for Bayesian photometric redshift
DIST	offset from source center (arcsec)

Notes. Columns description are described in Rafelski et al. (2015)

Table A.17. Column description of source table CANDELS_CAT.

Col. name	Description
ID	CANDELS source identifier
RA	right ascension (°)
DEC	declination (°)
MAG_F435W	AB F435W magnitude
MAG_F606W	AB F606W magnitude
MAG_F775W	AB F775W magnitude
MAG_F850LP	AB F850LP magnitude
MAG_F160W	AB F160W magnitude
DIST	offset from source center (arcsec)

Notes. Columns description are described in Whitaker et al. (2019)

Table A.18. Column description of source table HST3D_CAT.

Col. name	Description
ID	3D-HST source identifier
RA	right ascension (°)
DEC	declination (°)
MAG_F775W	AB F775W magnitude
z_best_s	redshift type
z_best	redshift
z_best_l95	lower limit for redshift
z_best_u95	upper limit for redshift
DIST	offset from source center (arcsec)

Notes. Columns description are described in Skelton et al. (2014)

Table A.19. Column description of source table SED.

Col. name	Description
wavelength	wavelength in Å
flux_prospector	Prospector SED flux
flux_magphys	Magphys SED flux
hst_wave	HST filter wavelength in Å
flux_hst_obs	HST observed flux

Notes. Flux unit are 10^{-20} erg s^{-1} cm^{-2} Å^{-1}

Table A.20. List and status of specific software.

Name	Description	Status	Availability	Version	Documentation	Reference
MPDAF	MUSE python data analysis framework	stable	pip, github	3.6	mpdaf.readthedocs.io	Bacon et al. (2016), Piqueras et al. (2019)
ORIGIN	Blind emission line source detection	stable	github	3.2	muse-origin.readthedocs.io	Mary et al. (2020)
ODHIN	De-blending of MUSE source using HST images	stable	github	1.0	odhin.readthedocs.io	this paper and Bacher (2017)
pyPlatefit	Emission and absorption line fitting	stable	github	0.7	pyplatefit.readthedocs.io	this paper
pyMarZ	Redshift estimation	dev		0.2		
SourceInspector	Interactive source evaluation and inspection	dev		1.3		
musered	Data reduction management	stable	github	0.3	musered.readthedocs.io	this paper
musex	Database management	dev		0.3		
imphot	MUSE PSF estimation based on HST photometry	stable	github	0.2	imphot.readthedocs.io	Bacon et al. (2017)
muse-psfr	MUSE PSF estimation based on AO telemetry	stable	github	1.1	muse-psfr.readthedocs.io	Fusco et al. (2020)

Notes. Software in development ("dev" in Status column) will be released later.

Appendix B: Advanced data reduction

B.1. Data reduction pipeline

We first run the raw science data through the standard MUSE pipeline (Weilbacher et al. 2020), using the development version[25]. Individual exposures are processed by the *scibasic* recipe with corresponding daily calibrations (flat fields, bias, arc lamps, twilight exposures) and geometry table (one per observing run) to produce the pixel tables (hereafter referred to as *pixtable*).

For the MXDF, we applied two changes related to the dark subtraction and the overscan parameters. To improve the bias subtraction, the overscan parameter was set to *vpoly:15,1.00001,1.00001* for all the recipes using that parameter (Weilbacher et al. 2020 Sect. 3.1). A sequence of long darks from June to August 2018 was used to produce a master dark, and then to derive a smooth model for each CCD as described in Weilbacher et al. (2020) Sect. 3.2.

The pipeline recipe *scipost* is then used to perform astrometric and flux calibrations on each pixable. For the MXDF, we computed a median standard response for each GTO run, using the standard exposures observed during that run and excluding those taken under non-photometric conditions.

The *scipost* recipe is run a first time in a "fast" mode (no sky subtraction, no Raman contamination correction) to produce images that are used to compute the centering offsets. Those offsets are calculated with the PSF fitting algorithm described in Bacon et al. (2017), Sect. 5.1, relative to the HST ACS images from the XDF data release (Illingworth et al. 2013).

The *scipost* recipe is then run a second time with sky subtraction, Raman correction and self-calibration but without applying the astrometry (*–astrometry=false*) to produce pixtables that we can use for the superflat (see Sect. B.4). The same recipe is run again to produce datacubes from these pixtables, using the astrometry and the offsets computed previously. In this case *scipost* will detect that the other steps have been done in the previous run so it will only need to produce the datacubes. The goal of this process is to minimize the computation time since we have to process hundreds of exposures.

For the different steps that require it (self-calibration, sky-subtraction and post-processing sky residual subtraction ZAP software) we use a source mask that we computed on the combined datacube from a previous version of the reduction.

B.2. New version of the self-calibration

We modified the recipe *scipost* to include the self-calibration algorithm. This algorithm was previously developed in the Python package MPDAF (Piqueras et al. 2019). Since this algorithm uses the average sky level as a reference to calculate the correction, we had to interrupt the *scipost* recipe before the sky subtraction, save the pixtables, reload them into Python and run the self-calibration, save again and resume processing with the MUSE pipeline. The overhead was huge, especially because the pixtables are very large files (8.6 GB) and saving and loading hundreds of them to remote storage takes time. Since the algorithm was already implemented as a C extension, moving it into the MUSE pipeline was an obvious choice to optimize processing, and it was also a great opportunity to make it available to the community.

Before porting the code to the MUSE pipeline, we developed a new version of the algorithm. When comparing the results of the first version used for DR1 with the flat field correction method *CubeFix* developed by Cantalupo (in preparation; see Cantalupo et al. 2019 for a description), it was clear that we could do better. The new algorithm is inspired by both *CubeFix* and the old version, and is described in more detail in Weilbacher et al. (2020). While *CubeFix* operates on datacubes, self-calibration operates directly on pixtables and could therefore be integrated into the *scipost* recipe.

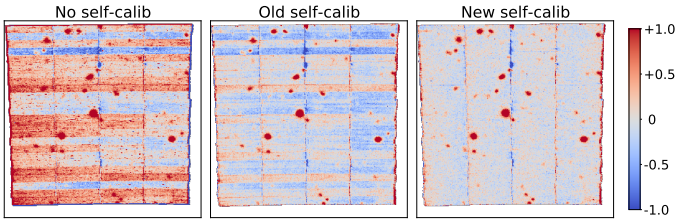


Fig. B.1. Illustration of the new self-calibration. Left panel: background subtracted white-light image for one exposure after the *scipost* recipe. Center panel: the same processed with the old version of the self-calibration. Right panel: with the new version. Flux units are 10^{-20} erg s^{-1} cm^{-2} Å^{-1}.

B.3. New version of ZAP (2.1)

The ZAP software (Soto et al. 2016) is a PCA-based tool developed to remove sky residuals in MUSE empty fields left by the imperfect sky subtraction pipeline process.

With the new version of self-calibration giving better results, it became clear that some of the artifacts we were observing in the reduced datacubes were due to ZAP. It introduced spatial variations for some pixels, which were visible in the white light images, and was also responsible for wiggles in the red part of the spectra. We corrected these problems with a new version of ZAP (version 2.1), where we made several changes: First, for the spatial variations we replaced the custom implementation of the PCA with the implementation from scikit-learn (Pedregosa et al. 2011), which is also much faster. Second, for the wiggles in the red part, we used the median continuum filter instead of the weighted one (which is a median filter weighted by the median sky level) and we increased the width of the filter window to 300 pixels. The default values were changed as well in version 2.0 for the window size and 2.1 for the filter type.

Third, the new version uses only one sky segment by default, which means that the cube is no longer split along the wavelength axis. Originally ZAP used 11 segments, whose purpose was to have coherent groups of sky emission lines, with a smaller number of eigenvalues per segment. It also allowed the computation to be parallelized. But the segments were also responsible for continuum oscillations, and made the choice of the number of eigenvalues per segment very difficult and very sensitive. With only one segment the performance of the sky subtraction is much better, thanks to the higher correlation between sky lines on the whole wavelength range. It also easier to control the power of the sky subtraction since the number of eigenvalues used for the sky reconstruction must be chosen or determined automatically for each segment.

Fourth, another issue was that sometimes ZAP suppressed signal for very bright lines, such as bright [O II] emitters. With the new version this is less likely to happen, thanks to the use of only one sky segment that provides a better estimation of the sky signal.

[25] Since v2.8.3 all development features used here are part of the public version.

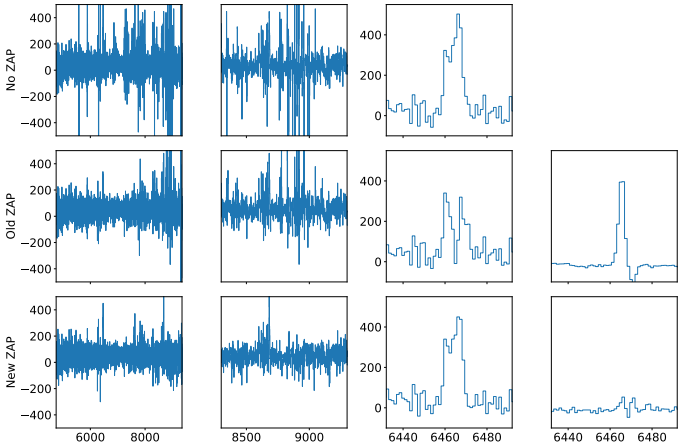


Fig. B.2. Example of improved performance of the new ZAP version. The top, middle and bottom rows display respectively the non-zapped, old zapped, and new zapped spectra of an [O II] emitter in the MXDF. From left to right, the columns display: the full spectral range, the red part of the spectra, the [O II]$\lambda\lambda$3726,3729 emission lines and the difference with the no ZAP spectrum. Flux units are 10^{-20} erg s^{-1} cm^{-2} Å^{-1}.

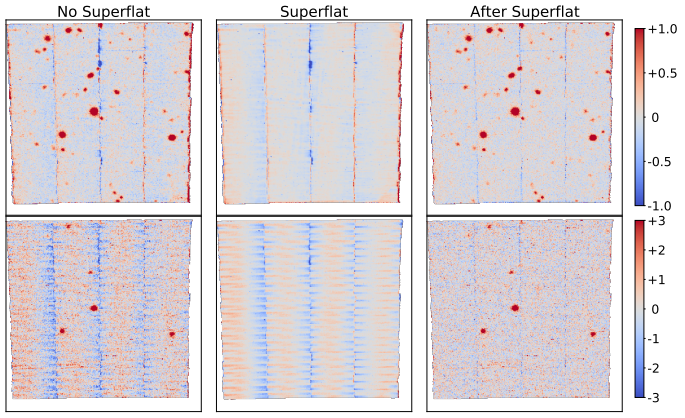


Fig. B.3. Example of superflat applied to a single exposure. The top and bottom rows display respectively the white light and red part (9000-9350Å) images. The left and right columns show the correction brought by the superflat (central column). Flux units are 10^{-20} erg s^{-1} cm^{-2} Å^{-1}.

As shown in Fig. B.2 this new version of ZAP has significantly improved the accuracy of sky subtraction and reduced the risk of capturing signal from bright emission lines. However, it revealed another issue: the flux at some spaxels are now sometimes rising or falling in the far red (above 8800Å; see the bottom left panel of Fig. B.3). This appears to be an instrumental effect, likely related to 2nd-order effects in the twilight-skyflat exposures used in the first step of the data reduction, and the old version of ZAP was removing this signal with its two last segment, instead of the sky emission in this wavelength range. Fortunately, this problem is now corrected by the superflat process described in the next subsection (see the bottom right panel of Fig. B.3).

Another change with respect to DR1 is that previously we were running ZAP on the individual exposures to get the best sky subtraction. With the superflat correction done on the individual datacubes (see the next section) this appeared not anymore necessary and we could run ZAP only once on the combined cube, which saves some computation time.

B.4. Superflat

In DR1, we masked the inter-stack holes in each individual exposure before combining them (Bacon et al. 2017, Sect.3.1.3). However, in addition to producing lower S/N region in the final datacube, the masking was never perfect and left some artifacts. To improve the process, we build a *superflat* for each datacube, by combining many exposures where we cancel the rotation and dithering. This is done by adding a `MUSE_SUPERFLAT_POS` environment variable in the MUSE pipeline, which allows the RA, DEC, and DROT keywords to be overridden by the values of the exposure for which the superflat is built. This way the "instrumental grid" is the same for all the datacubes and the sources move because of the rotation and dithering of the field. As we have sparse fields, combining those exposures with sigma-clipping removes the astrophysical signal and produces the superflat, where only the instrumental residuals remain.

For each exposure a superflat is built with 36 exposures from the same observing run, possibly supplemented with additional exposures from the previous or next run when there were not enough exposures in a single run. The superflat is subtracted from the science exposure[26]. To minimize the computation time, we build the datacubes tailored to each exposure from the pixtables that have been computed previously and where the self-calibration, sky-subtraction and Raman correction have already been done. Therefore, it is sufficient to produce the datacubes for each exposure after applying the astrometry with the modified RA, DEC and DROT. Even with this shortcut, applying the superflat correction to hundreds of exposures is computationally intensive.

As shown in the top panels of Fig. B.3 the superflat gives excellent results on the white light image. Most of the residuals and holes in the inter-stacks are corrected, which means that we no longer need to mask these areas. We also note that red effect in the 9000 – 9350Å band (see the previous section) is also corrected very well (Fig. B.3 bottom panels).

B.5. `musered`

To deal with the complexity of this data reduction, given the number of exposures and the number of reduction steps, we developed a Python package called `musered` on top of the MUSE pipeline. This package uses `python-cpl`[27] to run the pipeline recipes, and it uses a database to gather information about the raw files and to keep track of everything that has been done, for each file, each recipe, and each version of the reduction. `musered` takes care of file associations when running a recipe, to find the correct calibration files, and provides a convenient command line interface to run the various recipes, either from the pipeline or from custom Python recipes. It also makes it possible to inspect what was done for each exposure, when it was processed, and with which parameters and which calibration files.

The MXDF data set consists of 69 different nights and 373 exposures. It was then very time consuming to carefully inspect each exposure and each calibration frame. We therefore relied on several measurements and plots to identify problematic expo-

[26] After a few experiments, we found that subtracting the superflat from the science exposure produced better results than conventional division as it is typically done in imaging.
[27] https://github.com/olebole/python-cpl

sures, which sometimes led to the exclusion of a few problematic calibration sequences[28]. In this case, the sequence is flagged and another sequence from the previous or next night is used. The entire process costs 397 min of computing time per exposure on an 80-core workstation, of which 90% was due to the superflat computation. It took 103 days to completely reduce the MXDF data set.

B.6. Post-processing

Before combining the datacubes of all individual exposures, we identify those that should be discarded due to poor quality. The spatial PSF of each exposure is estimated with the muse-psfr tool, as described in Sect. 4.2.1. Twenty-two exposures with a PSF at 7000 Å greater than 0.″8 were rejected. The remaining 338 datacubes were combined using a 5σ sigma clipping scheme. The residuals left by sky subtraction were then removed using the new version of ZAP, as described in Sect. B.3. An additional check was performed on the resulting datacube to confirm that no offset was left in the background level at each wavelength plane.

As shown in Weilbacher et al. 2020, Sect. 4.6, the pipeline propagated variance of the datacube does not take into account the additional correlated noise due to the interpolation process. In DR1, we replaced the propagated variance of each individual datacube with a noise estimate derived from the datacube after masking the light sources, which is then rescaled to account for the impact of the correlated noise (Bacon et al. 2017, Sect. 3.5.1). The limitation is that this noise estimate is not valid for bright sources that are not sky-dominated. Another inaccuracy is that this noise estimate is by construction constant across the field of view and thus does not account for possible flat field variation. Thus, to address these limitations, we implemented a different scheme for MXDF: we computed the median ratio between the propagated variance and the variance estimated on the datacube itself after adjusting for correlated noise. This ratio has a slight evolution with the wavelength from 2.45 in the blue to 2.55 in the red. The propagated variance is then simply multiplied by this factor.

[28] 1 BIAS, 1 FLAT, 1 WAVE and 1 TWILIGHT exposures were discarded.

Appendix C: ODHIN

For more details of this whole procedure please refer to chapter 3 of Bacher (2017).

C.1. Model

C.1.1. Notations

Let $\boldsymbol{M}$ be a matrix composed of the elements $m_{i,j}$ and $\boldsymbol{v}$ be a vector. Let diag(s_j) be a diagonal matrix with diagonal elements s_j. $\|\boldsymbol{M}\|_2$ is the Frobenius norm of matrix $\boldsymbol{M}$. Tr$(\boldsymbol{M})$ is the trace of matrix $\boldsymbol{M}$. $\boldsymbol{M}^+$ is the pseudo-inverse of matrix $\boldsymbol{M}$. Hadamard (element-wise) product between a matrix $\boldsymbol{A}$ and a matrix $\boldsymbol{B}$ is noted $\boldsymbol{A} \circ \boldsymbol{B}$.

Let $\boldsymbol{Y}$ be the data matrix from MUSE of size $n \times \lambda$ where $n = pq$ is the number of spaxels (p,q are the two spatial axis) and λ is the number of spectral bands. Let $\boldsymbol{Z}$ be the data matrix from HST of size $N \times \Lambda$ with N the number of image pixels, Λ the number of filters (≈ 4). $\Lambda \ll \lambda$ and $n \ll N$.

Let $\boldsymbol{X}$ be the unknown field with a high spatial and spectral resolution of size $N \times \lambda$.

C.1.2. Model

Data observed by MUSE is assumed to be the addition of spatially degraded $\boldsymbol{X}$ data field and measurement noise $\mathcal{N}$. Spatial degradation is the multiplication of a convolution operation (by the FSF) and a subsampling operation. The spatial PSF (FSF) varies with the wavelength. We then have the following model:

$$\text{For} 1 \leq l \leq \lambda, \quad \underset{n\times 1}{\boldsymbol{Y}_l} = \underset{n\times N}{\boldsymbol{B}_l} \times \underset{N\times 1}{\boldsymbol{X}_l} + \underset{n\times 1}{\mathcal{N}}, \tag{C.1}$$

with $\boldsymbol{B}_l$ the spatial degradation matrix (subsampling and convolution), which can be written as $\boldsymbol{B}_l = \boldsymbol{S} \times \boldsymbol{C}_l$ where $\boldsymbol{C}_l$ is the convolution matrix by FSF at wavelength l and $\boldsymbol{S}$ is the subsampling matrix.

The HST observation, composed of images from different filters, is considered the result of a spectral degradation of $\boldsymbol{X}$:

$$\underset{N\times\Lambda}{\boldsymbol{Z}} = \underset{N\times\lambda}{\boldsymbol{X}} \times \underset{\lambda\times\Lambda}{\boldsymbol{A}}, \tag{C.2}$$

with $\boldsymbol{A}$ the matrix of the HST filter responses.

Let $\boldsymbol{H} \in \mathbb{R}^{N\times k}$ be the segmentation matrix at HST resolution. This segmentation is obtained using (Rafelski et al. 2015). Elements of $\boldsymbol{H}$ are defined as

$$h_{ij} = \begin{cases} 1 \text{ if the source } j \text{ is present on pixel } i, \\ 0 \text{ otherwise.} \end{cases}$$

It should be noted that we assume that MUSE and HST data have already been well aligned spatially. We also note that we do not take the LSF into account as it is negligible here.

C.1.3. Assumptions

We take the following assumptions and approximation:
A 1 (Spatial separability). All sources are spatially separated on HST images.
A 2 (Spatial invariability). The spatial shape of a source does not vary with the wavelength
A 3 (Spectral invariability). Sources are modeled with a unique spectrum.
A 4 (FSF invariability). The transfer function HST-MUSE is known, spatially invariant, and can be approximated as constant over a given spectral band.

Assumption A1 is valid for almost all sources in the considered fields. When it is not valid, the proposed method cannot do further de-blending. Assumptions A2 and A3 are also approximations that allow the sources to be separated. Note however that some sources, like close-by galaxies that are spatially resolved on MUSE, have velocity fields that will deform the observed spectra, breaking the A3 hypothesis. Other sources have emissions at certain wavelengths that differ greatly from the continuum, or even have no continuum, making the HST information irrelevant in these cases. Assumption A4 is in practice quite accurate. The FSF of HST is now well known and MUSE FSF has also been well studied (Villeneuve et al. 2011; Carfantan 2014). The latter is modeled by a 2D circular Moffat 2D function. The Moffat scale parameter varies slowly with the wavelength such that FSF can be considered constant on tens or even hundreds of MUSE spectral bands.

Assumption A2 allows the model to be placed within the framework of a linear mixing model:

$$\boldsymbol{X} \approx \underset{N\times k}{\boldsymbol{U}} \times \underset{k\times\lambda}{\boldsymbol{D}},$$

where k is the number of sources in the data ($k \ll N$), $\boldsymbol{D} \in \mathbb{R}^{k\times\lambda}$ is the sources spectra matrix , and $\boldsymbol{U} \in \mathbb{R}^{n\times k}$ the intensity matrix. We note that here we are talking of intensities, and not abundances like in other hyperspectral contexts (remote sensing), so we cannot use the classical sum-to-one constraint on the rows of this matrix.

All of these assumptions allow us to deduce an intensity matrix $\boldsymbol{U}$ using HST images. From the MUSE data and this matrix $\boldsymbol{U}$, we therefore try to directly estimate the matrix of the $\boldsymbol{D}$ spectra, without explicitly reconstructing $\boldsymbol{X}$.

C.2. Method

The strategy is as follows:

Algorithm 1

1: **for** each (vectorized) image $\boldsymbol{Z}_i$ of the HST **do**
2: Get intensity matrix $\boldsymbol{U}_i$ from $\boldsymbol{Z}_i$.
3: **for** each spectral block l **do**
4: Degrade $\boldsymbol{U}_i$ to MUSE spatial resolution to get $\widetilde{\boldsymbol{U}}_{i,l}$
5: Reverse the system $\boldsymbol{Y}_b l = \widetilde{\boldsymbol{U}}_{i,l}\boldsymbol{D}_{i,l}$ to find $\widehat{\boldsymbol{D}}_{i,l}$
6: Combine results: $\widehat{\boldsymbol{D}}_l = \sum_i a_{i,l}\widehat{\boldsymbol{D}}_{i,l}$, $1 \leq l \leq \lambda$ where $a_{i,l}$ is the spectral response of the filter i at the wavelength l.

Steps of lines (4) and (5) are costly to compute on all 3600 wavelength MUSE. As FSF varies slowly along wavelengths, we work on blocks (200 or 300 spectral bands) where we estimate a mean FSF over the block. This considerably reduces the computational cost since it is sufficient to pseudo-invert a single matrix $\widetilde{\boldsymbol{U}}_{i,l}$ for each block l. We note that we nevertheless obtain a full spectrum estimation, only the intensity matrix is assumed constant on the spectral bloc (A2). Along this process of linear operations we propagates the variance given with the MUSE data. Thus, we get at the end a variance for each estimated spectrum.

C.2.1. Intensity matrix estimation

First step (line 2) is done using the segmentation map of HST sources. From this map we build a $N \times k$ matrix $\boldsymbol{H}$, each column j of $\boldsymbol{H}$ corresponds to the presence/absence binary mask of the

source j in the vectorized HST field. We then apply element-wise product between $\boldsymbol{H}$ and HST image $\boldsymbol{Z}_i$ to get the intensity matrix $\boldsymbol{U}_i$:

$$\boldsymbol{U}_i = \boldsymbol{H} \circ \boldsymbol{Z}_i. \tag{C.3}$$

In step (4), we first perform a multiplication by the convolution transfer matrix between HST and MUSE. The convolution kernel from HST to MUSE K_{HM} is obtained from the convolution kernels of HST K_H and MUSE K_M, assumed to be known. Then the convolution result is resampled to the resolution of MUSE, using a linear interpolation.

We then have the intensity matrix at MUSE resolution $\widetilde{\boldsymbol{U}}_{i,l} = \boldsymbol{B}_l \boldsymbol{U}_i$.

Then for each object defined by the segmentation map, we get its own intensity map (even if the objects are not separable on the MUSE data).

C.2.2. Spectra estimation

Step (c) could be done by simply solving a linear mixing model in the least squares sense. But when the intensity matrix is poorly conditioned, that is to say, when objects are spatially very close, solving the problem in the least squares sense may over-fit the noise and create spectral artifacts (see Sect. C.3).

In order to cope with this phenomenon, a classical approach is to add constraints to regularize the problem. There are a large number of possible regularizations in the literature, among which we can notably mention the penalization regularizations of the ridge type (Hoerl & Kennard 1970) or LASSO (Tibshirani 1996) and the regularizations by informational criterion like the Bayesian information criterion (BIC: Schwarz 1978). To answer to MUSE data specificities, we chose to exploit both ridge and sparse regularizations, as spectra are composed of a smooth spectral continuum and a small set of sharp emission/absorption lines.

We thus separately process lines and continuum. For this, the procedure is as follows: (i) obtaining the line cube $\boldsymbol{Y}^r$ by subtracting the continuum cube estimated by a robust filtering (e.g., a median filter); (ii) reconstruction of the $\widehat{\boldsymbol{D}}^r$ lines associated with the objects from the line cube obtained previously: we will see that a parsimonious regularization approach with model selection has been chosen; (iii) subtraction of the estimated contribution of the lines to obtain the remaining data $\boldsymbol{Y}^c = \boldsymbol{Y} - \widetilde{\boldsymbol{U}}\widehat{\boldsymbol{D}}^r$; (iv) estimation of the spectral continua $\widehat{\boldsymbol{D}}^c$ using a ridge regularization on the data $\boldsymbol{Y}^c$; and (v) combination of the two estimates $\widehat{\boldsymbol{D}} = \widehat{\boldsymbol{D}}^r + \widehat{\boldsymbol{D}}^c$.

Spectral line estimations

Sparse regularization of spectral lines avoids over-fitting the noise and favors line attribution to a minimal number of spectra, which is the most physically probable. Spectral lines are assumed to extend over several spectral slices. We can thus exploit this assumption to ensure a good lines reconstruction and avoid spectral discontinuities (where part of a line is associated with one object and the other part to another). Thus, we seek to do the model selection jointly over all the spectral support of the spectral line. To do so, it is thus necessary to estimate this spectral support. Strategy is thus as follows: (i) fast detection of all potential spectral supports of lines and (ii) for each segment, selection of objects with nonzero spectrum, then estimation of spectra using only selected objects.

The lines support detection is based on local extrema detection and is detailed in (Bacher 2017). It allows all the spectral supports of the potential lines in the studied area to be obtained quickly. We can then estimate the spectra of objects with regularization by selecting models on each of these supports.

The sparse regularization is done using BIC selection. BIC is written as $\mathrm{BIC} = K\log(n) - 2\log(\hat{L})$ where $\hat{L}$ is the model likelihood maximum, K is the number of free parameters, n is the number of samples (here pixels in the area). In our case we assume that the noise is Gaussian. On a spectral band, for a model $\mathcal{M}$ with k objects with nonzero spectrum, we have

$$\mathrm{BIC}(\mathcal{M}) = K\log(n) + \log(\widehat{\sigma_{\mathcal{M}}}^2), \tag{C.4}$$

where $\widehat{\sigma_{\mathcal{M}}}^2$ is the empirical variance of the residuals obtained using the $\mathcal{M}$ model chosen. We have a nonzero spectrum free parameter plus one for the noise variance, and hence $K = k + 1$. To minimize the criterion BIC we see that there is a compromise to be found between a model that best explains the data ($\widehat{\sigma}^2$ minimal) and a simplest possible model (k minimal) . The selection of the best model is then reduced to a combinatorial problem where all the possible combinations of nonzero spectra are tested and the one that minimizes the BIC is retained. For computational reasons, instead of exploring all combinations we use a greedy variant for the BIC.

For each line l, we select the model $\mathcal{M}_l$ over the spectral support of l using

$$\mathcal{M}_l = \underset{\mathcal{M}}{\operatorname{argmin}}\, \mathrm{BIC}(\mathcal{M}), \tag{C.5}$$

where $\mathrm{BIC}(\mathcal{M})$ comes from Eq. (C.4).

If we then note $\widetilde{\boldsymbol{U}}_{\mathcal{M}}$ the intensity matrix composed only of the sources selected in $\mathcal{M}$, we can compute the associated spectra $\widehat{\boldsymbol{D}}_l^r$ for line l with

$$\widehat{\boldsymbol{D}}_l^r = \underset{\boldsymbol{D}}{\operatorname{argmin}} \|\boldsymbol{Y}_l^r - \widetilde{\boldsymbol{U}}_{\mathcal{M}_l}\boldsymbol{D}\|_2^2. \tag{C.6}$$

Continuum regularization

After subtracting the estimated contribution of the lines, we now estimate the spectral continua of the objects using ridge regularization. For a given block of sheets l and an HST image i we note the column vector $\boldsymbol{y} = \underbrace{\boldsymbol{Y}_l}_{n\times 1}$, $\widetilde{\boldsymbol{U}} = \widetilde{\boldsymbol{U}}_{i,l}$ and $\boldsymbol{d} = \underbrace{\boldsymbol{D}_{i,l}}_{k\times 1}$. Ridge regularized least squares admits an analytical solution of the form

$$\widehat{\boldsymbol{d}} = \left(\widetilde{\boldsymbol{U}}^T\widetilde{\boldsymbol{U}} + \alpha\boldsymbol{I}_k\right)^{-1}\widetilde{\boldsymbol{U}}^T\boldsymbol{y}, \tag{C.7}$$

with $\boldsymbol{I}_k$ the identity matrix of $\mathbb{R}^{k\times k}$ and where $\alpha \geq 0$ is the regularization parameter to be fixed.

The optimal value (in the Mean Square Error sense) of regularization parameter α is estimated using generalized cross validation (GCV; Golub et al. 1979). In order to increase the robustness of the cross-validation, only one regularization parameter is searched per block of a few tens of spectral sheets. This amounts to considering that the signal-to-noise ratio of the data (deprived of the lines) evolves weakly according to the wavelength. This hypothesis is valid if we consider that the sky's lines have been sufficiently well subtracted.

Flux correction

A natural drawback of ridge regularization is that the estimated solution is biased toward zero, reflected in our case by a loss of flux on the estimated spectra, which is of course not desired. To overcome this drawback, we take advantage of the redundancy of penalized estimators along the spectrum: by looking for a multiplicative factor per spectral block of sufficiently large size (on the order of a hundred sheets), we can make sure that the average flow (on the block) is preserved, while preserving the benefit of regularization. This correction factor is obtained in the following way: one looks for a diagonal matrix $\boldsymbol{F} = \mathrm{diag}(\{f_j\}_{1\leq j\leq k})$ such as one has for continuum $\boldsymbol{Y}^c = \widetilde{\boldsymbol{U}}\boldsymbol{F}\widehat{\boldsymbol{D}}^c$. The matrix $\boldsymbol{F}$ being diagonal this amounts to defining

$$f_j = \frac{\left\langle \left(\boldsymbol{Y}^c \left(\widehat{\boldsymbol{D}}^c\right)^+\right)_j, \widetilde{\boldsymbol{U}}_j \right\rangle}{\|\widetilde{\boldsymbol{U}}_j\|_2^2}, \text{ for } 1 \leq j \leq k. \tag{C.8}$$

We note that if this least-squares correction was done wavelength by wavelength it would in fact cancel the ridge regularization.

The regularized approach of spectra estimation for an HST image i within a FSF-constant spectral block j is summarized in the algorithm 2. We note that the whole method stays unsupervised as regularizations are without external parameters (for lines estimation) or with auto-estimated parameter (for continuum estimation).

Algorithm 2 Regularized procedure of spectra estimation

Input: data $\boldsymbol{Y} \leftarrow \boldsymbol{Y}_j$, intensity matrix $\widetilde{\boldsymbol{U}} \leftarrow \widetilde{\boldsymbol{U}}_{i,j}$

Estimation of $\boldsymbol{Y}^c$ ▷ Continuum estimation (robust filtering)
Lines estimation $\boldsymbol{Y}^r = \boldsymbol{Y} - \boldsymbol{Y}^c$ ▷ Continuum subtraction
Line detections
for line l **do**
    Computation of $\mathcal{M}_l$ using eq. (C.5) ▷
    Computation of $\widehat{\boldsymbol{D}}_l^r$ using eq. (C.6)
Construction of $\widehat{\boldsymbol{D}}^r$ ▷ Spectral concatenation
Computation of $\boldsymbol{Y}^c = \boldsymbol{Y} - \boldsymbol{Y}^r$
for bloc spectral b **do**
    Computation of α_{m+1s} using GCV
    Computation of $\widehat{\boldsymbol{D}}^r$ ▷ Continuum estimation using ridge
Construction of $\widehat{\boldsymbol{D}}^c$ ▷ Spectral concatenation
Computation of $\boldsymbol{F}$ using (C.8) ▷ Flux correction factors
Computation of $\widehat{\boldsymbol{D}} = \widehat{\boldsymbol{D}}^r + \boldsymbol{F}\widehat{\boldsymbol{D}}^c$ ▷ Lines + Continuum

Output: $\widehat{\boldsymbol{D}}$ ▷ Sources spectra (for image HST i and FSF-constant spectral band j)

Finally, in order to take into account the residuals of the atmosphere spectrum subtraction, we add throughout the unmixing procedure, the estimation of a spectrum spatially constant in the considered area, corresponding to the sky background spectrum. This spectrum can be seen as the intercept of the studied regression problem.

C.3. Validation on simulated data

C.3.1. Simulation settings

In order to evaluate the validity of this method, we built a series of data with sources that are increasingly close to each other spatially. To do this, we built a datacube (160×160) in flux units with high resolution (called HST data by abuse of language), which was then degraded following the MUSE model (same subsampling and core of convolution) to obtain data at low spatial resolution (which will be called MUSE data by abuse of language). An additive Gaussian noise is then applied, slightly correlated spatially by the application of a 3 by 3 spatial core.

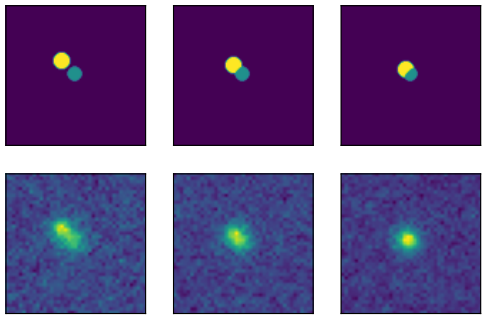


Fig. C.1. Simulation example. First row: "HST" segmentation map. Second row: "MUSE" white image. Distance of sources is characterized by the unit-less conditioning number c, increasing here from 1.2 (left column) to 3.7 (right column).

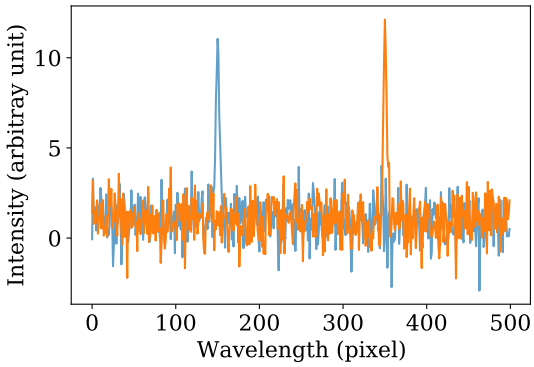


Fig. C.2. Ground truth spectra of the two simulated sources.

We can see in Fig. C.1 the simulated cube white image at MUSE and HST resolution. Simulated spectra are composed of an emission line for each object and a spectral continuum (see fig. C.2).

De-blending difficulty is measured using the conditioning number c of intensity matrix $\widehat{\boldsymbol{U}}$ (at MUSE spatial resolution):

$$c = \frac{s_{max}(\widehat{\boldsymbol{U}})}{s_{min}(\widehat{\boldsymbol{U}})}, \tag{C.9}$$

where $s_{max}(\widehat{\boldsymbol{U}})$ and $s_{min}(\widehat{\boldsymbol{U}})$ are the minimum and maximum singular values of the intensity matrix, and c is a bound of estimation error of $\widehat{\boldsymbol{D}}$ relative to a perturbation on data $\boldsymbol{Y}$. The closer c is to 1, the easier the inversion problem is. On the contrary, high values of c will imply instability and over-fitting of the noise.

We then measured the fidelity to the ground truth (the mean of the intercorrelations between each estimated spectrum, $\widehat{\boldsymbol{d}}_j$, and its ground truth, $\boldsymbol{d}_j$),

$$f = \frac{1}{k}\sum_{j=1}^{k} \frac{\langle \widehat{\boldsymbol{d}}_j, \boldsymbol{d}_j \rangle}{\|\widehat{\boldsymbol{d}}_j\|.\|\boldsymbol{d}_j\|},$$

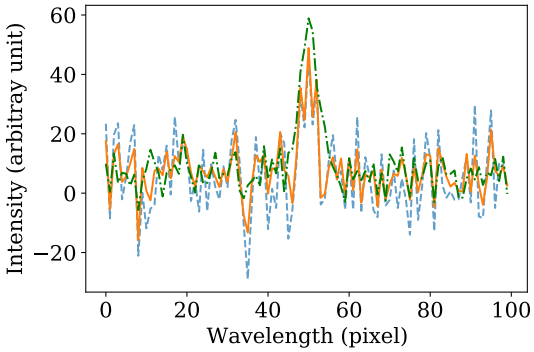


Fig. C.3. Ground truth (green dashed-dot) versus estimated spectra (proposed method in orange, least squares in dashed blue) for a difficult case (condition number equals 6)

and the inter-correlation,

$$ic = \frac{\langle \widehat{\boldsymbol{d}_0}, \widehat{\boldsymbol{d}_1} \rangle}{\|\widehat{\boldsymbol{d}_0}\|.\|\widehat{\boldsymbol{d}_1}\|},$$

between the two estimated spectra.

C.3.2. Results

We compared the proposed regularization method to a simpler least squares approach that has a closed-form solution given by

$$\widehat{\boldsymbol{D}}_{i,l} = (\widetilde{\boldsymbol{U}}_{i,l}^T \widetilde{\boldsymbol{U}}_{i,l})^{-1} \widetilde{\boldsymbol{U}}_{i,l}^T \boldsymbol{Y}_l. \tag{C.10}$$

Figure C.3 shows one of the two estimated spectra, around its spectral emission line, and when the sources are very close spatially ($c \approx 6$). We can see that both the proposed method and a non-regularized least squares approach estimate accurately the emission line but the proposed method is better at avoiding increased variance around the peak.

The results shown in Fig. C.5 describe the evolution of performance as the conditioning degrades. Performances of the proposed method are clearly more robust to ill-conditioning than a simple least squares approach. In particular, variance of the estimated spectra residuals increases and the estimated spectra become strongly negatively correlated. This anticorrelation is explained by the positivity of the $\widetilde{\boldsymbol{U}}$ intensity matrix. This phenomenon is also illustrated by Figs. C.3 and C.4, where the unmixed spectra when the sources are very close spatially ($c \approx 6$) are shown. The robustness to anticorrelated patterns in the proposed method allows it to keep a good fidelity to the sources spectra (fig C.5c). .

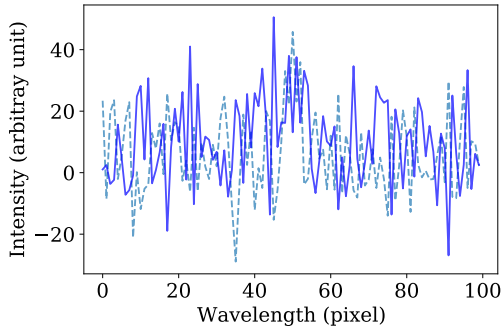
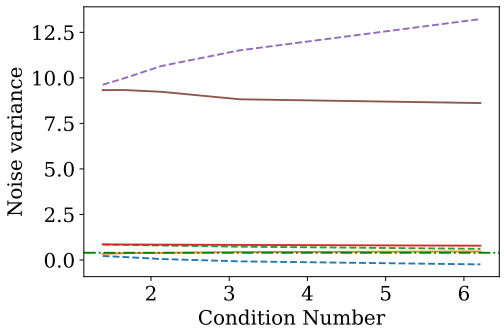
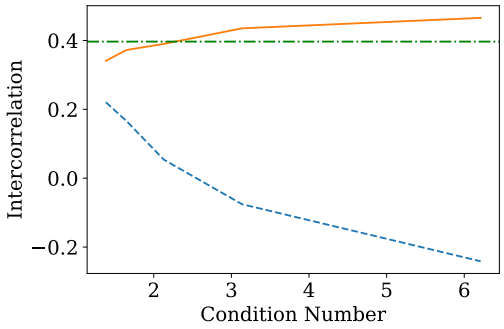


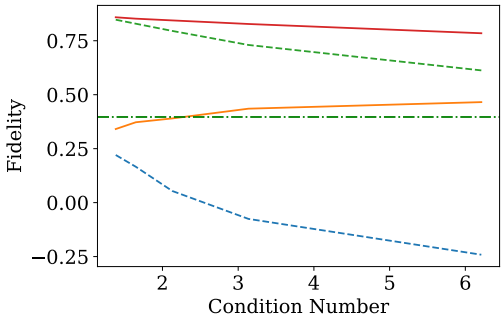
Fig. C.4. Zoom over a part of the non-regularized least squares estimation of both sources spectra, for a difficult case (condition number equals 6). Some anticorrelated artifacts are visible.



(a) Variance of spectra residuals



(b) Intercorrelation between objects spectra (dashed-dot green line is the ground truth)



(c) Fidelity to ground truth

Fig. C.5. Performance evolution function of conditioning. The dashed blue line indicates no regularization and the orange line with regularization. Results are averaged over ten Monte Carlo runs.

Appendix D: Emission and absorption lines

The lines input table used in pyPlatefit is given in Tab. D.1. We note that the default line table can also be retrieved directly from the pyPlatefit python interface[29].

The lines are grouped into 3 different families (column *Set*): the Balmer series (*bal*), non-balmer emission lines (*for*) and ISM absorption lines (*ism*). The vacuum rest wavelength in Å is given in the column *Wave*. The multiplet (*Mult*) column is used to group lines together when performing the fit. *Mult* is an integer value following the adopted notation in the literature (e.g., [OII]3727 for the [O II]$\lambda\lambda$3726,3729 doublet). The *Main* boolean column can be used to filter main and secondary lines. The two boolean columns, *Emi* and *Abs*, indicate if the line is in emission and absorption, respectively. The *Res* column is used to perform an independent fit for some major resonant lines. We note that some emission lines like OVI and NV, although resonant, are not fitted separately because they were too faint in our galaxy sample.

Table D.1. pyPlatefit lines definition.

Name	Set	Wave	Mult	Main	Emi	Abs	Res
OVI1032	for	1031.91	1033	F	T	F	F
OVI1038	for	1037.61	1033	F	T	F	F
LYALPHA	bal	1215.67	0	T	T	F	T
NV1238	for	1238.82	1240	F	T	F	F
NV1242	for	1242.80	1240	F	T	F	F
SiII1260	ism	1260.42	0	F	F	T	F
OI1302	ism	1302.17	1303	F	F	T	F
SiII1304	ism	1304.37	1303	F	F	T	F
CII1334	ism	1334.53	0	F	F	T	F
SiIV1394	for	1393.76	1403	F	T	T	F
OIV1397	for	1397.23	1403	F	T	F	F
OIV1400	for	1399.78	1403	F	T	F	F
SiIV1403	for	1402.77	1403	F	F	T	F
NIV1486	for	1486.50	0	F	T	F	F
SiII1527	ism	1526.71	0	F	F	T	F
CIV1548	for	1548.20	1549	T	T	T	T
CIV1550	for	1550.77	1549	T	T	T	T
FeII1608	ism	1608.45	1610	F	F	T	F
FeII1611	ism	1611.20	1610	F	F	T	F
HeII1640	for	1640.42	0	F	T	F	F
OIII1660	for	1660.81	1663	F	T	F	F
OIII1666	for	1666.15	1663	F	T	F	F
AlII1671	ism	1670.79	0	F	F	T	F
NIII1750	for	1749.67	0	F	T	F	F
AlIII1854	ism	1854.10	0	F	F	T	F
AlIII1862	ism	1862.17	0	F	F	T	F
SiIII1883	for	1882.71	1886	F	T	F	F
SiIII1892	for	1892.03	1886	F	T	F	F
CIII1907	for	1906.68	1909	T	T	F	F
CIII1909	for	1908.73	1909	T	T	F	F
CII2324	for	2325.40	2326	F	T	F	F
CII2325	for	2326.11	2326	F	T	F	F
CII2326	for	2327.64	2326	F	T	F	F
CII2328	for	2328.84	2326	F	T	F	F
FeII2344	ism	2344.21	0	F	F	T	F
FeII2374	ism	2374.46	0	F	F	T	F
FeII2382	ism	2382.76	0	F	F	T	F
NeIV2422	for	2421.83	2424	F	T	F	F
NeIV2424	for	2424.42	2424	F	T	F	F
OII2470	for	2471.02	0	F	T	F	F
FeII2586	ism	2586.65	0	F	F	T	F
FeII2600	ism	2600.17	0	F	F	T	F
MgII2796	for	2796.35	2799	F	T	T	T
MgII2803	for	2803.53	2799	F	T	T	T
MgI2853	ism	2852.97	0	F	F	T	F
NeV3426	for	3426.85	0	F	T	F	F
OII3726	for	3727.09	3727	T	T	F	F
OII3729	for	3729.88	3727	T	T	F	F
H11	bal	3771.70	0	F	T	T	F
H10	bal	3798.98	0	F	T	T	F
H9	bal	3836.47	0	F	T	T	F
NeIII3869	for	3870.16	0	T	T	F	F
HeI3889	for	3889.73	0	F	T	F	F
H8	bal	3890.15	0	F	T	T	F
CaK	ism	3933.66	0	F	F	T	F
CaH	ism	3968.45	0	F	F	T	F
NeIII3967	for	3968.91	0	F	T	F	F
HEPSILON	bal	3971.20	0	F	T	T	F
HDELTA	bal	4102.89	0	T	T	T	F
CaG	ism	4304.57	0	F	F	T	F
HGAMMA	bal	4341.68	0	T	T	T	F
OIII4363	for	4364.44	0	F	T	F	F
HBETA	bal	4862.68	0	T	T	T	F
OIII4959	for	4960.30	0	T	T	F	F
OIII5007	for	5008.24	0	T	T	F	F
MgB	ism	5175.44	0	F	F	T	F
HeI5876	for	5877.25	0	F	T	F	F
NaD	ism	5891.94	0	F	F	T	F
OI6300	for	6302.05	0	F	T	F	F
NII6548	for	6549.85	0	F	T	F	F
HALPHA	bal	6564.61	0	T	T	T	F
NII6584	for	6585.28	0	T	T	F	F
SII6717	for	6718.29	0	T	T	F	F
SII6731	for	6732.67	0	T	T	F	F
ARIII7135	for	7137.80	0	F	T	F	F

[29] pyplatefit.readthedocs.io/en/latest/tutorial.html#Selecting-and-updating-emission/absorption-lines