# Searching for Narrowband excess Ly$\alpha$ emitters at $z > 8$ in the COSMOS field

Using the new COSMOS2020 catalogs to find high redshift LBGs and locating emission line galaxies using the new ultra deep $NB$118 narrowband filter from UVISTA.

Written by *André Endrup Hartwigsen*
May 20, 2021

Supervised by
Johan P. U. Fynbo & Bo Milvang-Jensen

# University of Copenhagen

NAME OF INSTITUTE:   Niels Bohr Institute

NAME OF DEPART-   Faculty of Science
MENT:

AUTHOR(S):   André Endrup Hartwigsen

EMAIL:   andrehartwigsen@gmail.com

TITLE AND SUBTITLE:   Searching for Narrowband excess Ly$\alpha$ emitters at $z > 8$ in the COSMOS field
- Using the new COSMOS2020 catalogs to find high redshift LBGs and locating emission line galaxies using the new ultra deep $NB$118 narrowband filter from UVISTA.

SUPERVISOR(S):   Johan P. U. Fynbo & Bo Milvang-Jensen

HANDED IN:   20.05.2021

DEFENDED:   09.06.2021

# Abstract

Observing and classifying galaxies at high redshift is paramount to understanding the formation and evolution of the universe. This thesis aims to find and characterize Lyman break galaxies at $z > 7$, probing the early universe at the *Epoch of Reionization.* This is accomplished with preliminary access to the COSMOS2020 catalog, using new observations from the VISTA telescope.

UltraVISTA is a very deep imaging survey, totalling observations spanning over 6 years, containing 4 broadband filters ($Y$, $J$, $H$ and $K_s$) and one narrowband filter $NB118$ at $1.18\,\mu$m. COSMOS2020 (Weaver et al., subm) combines these in a multi-wavelength photometric and photo-z catalogue, including numerous of the deepest surveys of the $2\,\mathrm{deg}^2$ COSMOS field to date.

The narrowband filter $NB118$ is placed within the FWHM of broadband filter $J$, sources with excess flux (emission lines) can be identified in the narrowband filter. This enables the search for strong emitters such as [OII] at redshift 2.2, [OIII] and H$\beta$ around redshift 1.4, H$\alpha$ at redshift 0.8 and Ly$\alpha$ at redshift 8.8. In this work the focus is on searching for Lyman break galaxies with Ly$\alpha$ narrowband excess at $z \sim 8.75$. Candidates are selected by requiring flux blueward of the Ly$\alpha$ line to be consistent with zero, and inferring strict criteria for narrowband excess. No sources were found with satisfactory $NB118$ significance. A galaxy with substantial potential (J$_3$) at $z_{\mathrm{phot}} = 8.8 \pm 0.6$ was found, currently with narrowband excess indicating a Ly$\alpha$ line, but additional observation time will be required to be statistically significant.

COSMOS2020 uses a wide range of deep broadband filters to create an $izYJHK_s$ chi mean detection image. This does not include $NB118$, and it is therefore inherently interesting to use $NB118$ as detection image in contrast. This yielded 114 potential new sources not previously in the COSMOS2020 catalogue. With a handful of these presenting very good candidates to be real emission line galaxies.

Finally a search for lensed-, blended- and critically failed photometric redshift fit-galaxies was completed. Here sources with a photometric redshift within $2\sigma$ of the a major emission line in the $NB118$ filter are excluded, and the remaining galaxies are selected for narrowband excess and significance. This resulted in the detection of emission lines outside the search range and blended sources but no obvious lensed object. A bright $r < 21.5\,\mathrm{mag}$ galaxy with complex morphology was selected and sent as observation proposal to the NOT telescope on La Palma. Spectroscopy of H$\beta$, H$\gamma$, H$\alpha$, and two [OIII] lines visible in the otherwise flat spectrum, revealed a likely low metallicity irregular galaxy at redshift $z = 0.101\,21 \pm 0.000\,02$ corresponding to a He I emission line, causing the observed narrowband excess at $10\,830\,\text{Å}$.

# Contents

# 1   Introduction

The study of ultra high redshift galaxies has proven to be particularly difficult for many years. They potentially provide crucial knowledge into the formation and evolution of the early universe, and build the foundation for the galaxies around us today. Galaxies less than a billion years old, give insight into star formation, as well as the creation of supermassive black holes and quasars. The process of reionization and dust chemistry however also play an important role in the evolution of the universe, and it is therefore very important to develop methods to efficiently find and characterize these high redshift galaxies.

Star formation is a crucial piece in understanding the early universe, due to the most massive stars also producing by far the most output. Realizing the fact that galaxies are just ensembles of stars, with added dust, gas and dark matter, useful information about the star formation rate (SFR) and the distribution of stellar masses can be found in the spectral energy distribution (SED) of galaxies. Much of the radiation from these massive stars is emitted in the UV, and is therefore ionizing. However at ultra high $z$, these wavelengths subside in the visual due to redshift, making these features visible to earth based observation. Positively identifying Lyman break galaxies (LBGs), and extracting accurate photometric data in the visible to near infrared (NIR), is a key goal to better describe how the universe was re-ionized at $z \sim 7$.

The Epoch of Reionization is a physical observational limit to what we can observe, because the ionized medium is transparent to radiation. Characterizing galaxy evolution at $z > 8$ took place, is paramount to understanding how fast and efficient star formation was needed in this early time of the universe in order to ionize the neutral hydrogen. Some works indicate a rapid decline in star formation density in the $z \sim 8 - 10$ range (Oesch et al., 2018; McLeod et al., 2016), so a key question becomes: Could the faint galaxies emit enough ionizing radiation to re-ionize the universe?

The most prominent feature for all these high redshift galaxies, is the so-called Lyman break feature. At wavelengths shorter than 1215 Å, effectively all photons are scattered and or re-emitted at longer wavelengths (see figure 1.1). This starts from the Ly$\alpha$ line, where most the flux is absorbed, and for $\lambda < 912$ Å, effectively all flux is absorbed in the Lyman limit. These LBGs can therefore be identified by having very little to no flux after the Ly$\alpha$ line (Laursen et al., 2019).

The goal of this thesis is to use the new ultra-deep observations from the COS-MOS2020 survey (Weaver et al., subm) of the $2 \deg^2$ COSMOS field, to find and identify LBGs with an emission line right at the flux cut-off. This will be achieved by searching for emission line galaxies, in the $1.18\,\mu m$ $NB$118 (Milvang-Jensen et al., 2013) narrowband filter. Specifically comparing the narrowband magnitude to the $J$ broadband filter magnitude. Sources can be isolated with emission lines, in the $J$-filter regime, by searching for narrowband excess. This is possible because the entire width of the $NB$118 filter, subsides within the width broad $J$ filter. Two catalogues are paired, `Classic` and `The Farmer` extracted by Weaver et al. (subm), to find objects at ultra high redshifts. For each catalogue, the photometric redshift codes `LePhare` (Arnouts et al., 2002; Ilbert et al., 2006) and `EAZY` (Brammer et al., 2008) are applied, yielding a total of 4 photometric redshifts for each source detected in both catalogs.

The primary purpose is to find and characterize $z \simeq 8.8$ Ly$\alpha$ emitters in the $NB$118 narrowband filter. This will be done searching for $Y$ band dropouts, and selecting ones with emission lines in the narrowband filter. A search for sources not located in the COS-
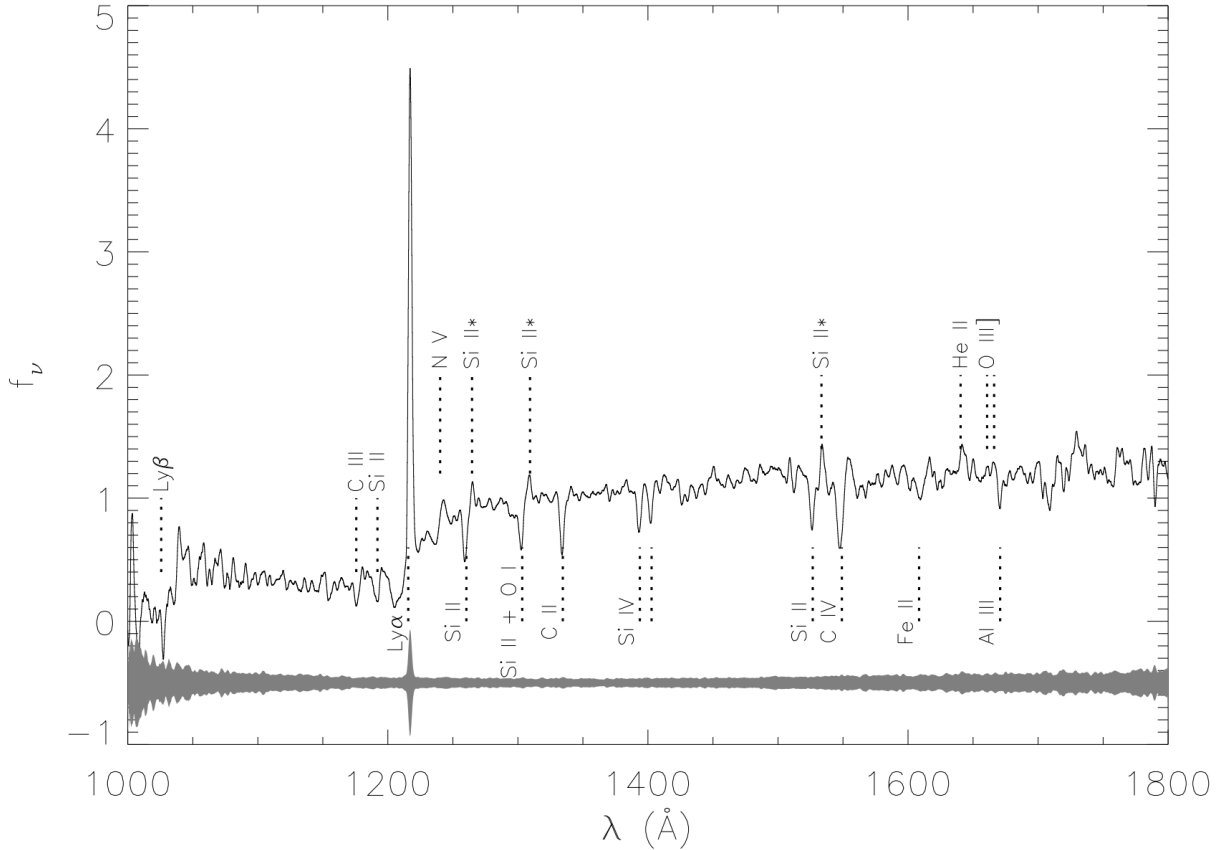
Figure 1.1: Example of a Lyman break galaxy spectrum, made from the composite spectrum of 81 galaxies between redshift 3.5 and 4.5 from Jones et al. (2012). The grey filled region represents the $\pm 1\sigma$, created by the scatter of the various spectra.

MOS2020 data will also be completed, using $NB118$ as basis for the detection image. In contrast to the deep `Classic` and `The Farmer` chi-mean $izYJHK_s$ composite image for source extraction. As a final results, a survey using narrowband excess and photometric redshift is conducted to search for possibly lensed objects and incorrectly determined photometric redshift. Where a source was picked out and spectroscopic redshift determined using the Nordic Optical Telescope (NOT).

This thesis has been given preliminary access to unfinished catalogs of the COS-MOS2020 data, making up the most filters, and deepest integration ever achieved. The two catalogues based on the `CHI_MEAN` combined $izYJHK_s$ image setting of `SWarp`, and explore two different methods of PSF homogenizing the photometry. The `Classic` catalog uses the classic `SExtractor` for source detection, and extracts flux in standard circular apertured of automatic, $2''$ and $3''$. `The Farmer` catalogue is focused on template fitting, varying from a simple point spread function composite galaxy models. `The Farmer` catalog does also not yet feature $NB118$ and some other broader filters like SuprimeCam channels B, V, $r$, $i^+$, $z^{++}$ and IRAC ch3 and ch4. Future releases of the `The Farmer` catalog will however have this filters included. Because Weaver et al. (subm) is in the submitted phase, we will go through how the data used in this work was extracted.

# 2 Theory

## 2.1 Epoch of Reionization

Since this work aim to look as far back as redshift $z \sim 8$, it is important to understand the epoch of reionization. Neutral hydrogen is the main cause for the Lyman break, so we aim to archive a better understanding of how the distribution of neutral hydrogen in the universe evolves with time.

After the the expansion and subsequent cooling of the universe following the Big Bang, stars began to form as gas became neutral at approximately $\sim$400,000 years or 3000 K surrounding the tiny overdensities in the universe. This also made the universe transparent as electrons and ions combined, with photons and baryon subsequently uncoupled, this leads to immense amounts of light being created, resulting in what we now know as the cosmic microwave background (CMB). Star formation leads to ionizing radiation and clustering of stars which eventually become galaxies. This radiation creates bubbles of high energy gas around them, and helps to clump matter further through radiation pressure and shocks. After sufficient amount of time had passed, galaxies formed and the entire universe eventually becomes re-ionized. This transition is referred to as the EoR (the Epoch of Reionization). The specific time period is not a solid point in time,
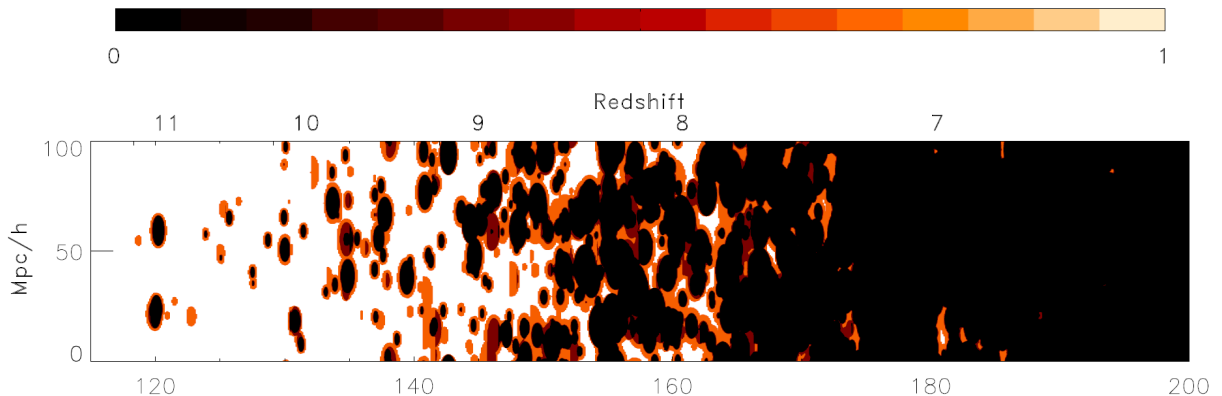


Figure 2.1: A slice through redshift of the 21 cm radiation in which the reionization process progresses through the volume of a cosmological simulation with radiative transfer, taken from: Thomas et al. (2009). Darkness represents the ionization level.

but rather, as would be expected, constrained to be within a redshift range (Zaroubi, 2012). Today we expect the EoR to be somewhere between $z \sim [6 - 15]$, corresponding to 12.8 Gyr and 13.4 Gyr ago. Looking back as far as $z \sim 8$, means a significant fraction of distance that light has to travel, will be through neutral hydrogen. Lyman alpha radiation at 1216 Å or shorter, will undergo significant scattering in the dense gas, including absorption and re-emission. Vast amounts of UV radiation from galaxies will therefore be spread out, and dim the galaxy in UV. The gas can however also contribute to adding luminosity to the Ly$\alpha$ line. Most of the light at the Ly$\alpha$ wavelength and shorter, will be absorbed and re-emitted as Ly$\alpha$ in the dense early universe. This can cause up to 10% of a galaxy's total bolometric luminosity to be in the Ly$\alpha$ peak alone (Laursen et al., 2019), with a high equivalent width due to Doppler shifts.

There are a few indicators assisting in determining the redshift of the EoR, including the 21 cm neutral hydrogen line, Thomson scattering of the CMB and the Lyman alpha forest (sec: 2.6).

## 2.2   Star formation

Fundamentally star formation is about the collapse of gas under its own gravity. This requires the gravitational force be greater than both thermal gas pressure, and radiation pressure. Using the ideal gas equation and hydrostatic equilibrium, we can derive the Jeans length ($\lambda_J$) and mass ($M_J$) (Shultz, 2012, p. 116). The critical mass at which a cloud of gas becomes unstable to its own gravity.

$$\lambda_J = \sqrt{\frac{\pi c_s^2}{G\rho}} = 0.2\sqrt{\frac{T}{10\,\text{K}}}\left(\frac{n}{10^4\,\text{cm}^{-3}}\right)^{-1/2} \quad M_J = \frac{4\pi}{3}\rho\lambda_J^3 = 1.6\left(\frac{T}{10\,\text{K}}\right)^{3/2}\left(\frac{n}{10^4\,\text{cm}^{-3}}\right)^{-1/2}$$

(1)

If a clump of hydrogen gas with density $\rho$, temperature $T$, at critical mass $M_J$ and extent $\lambda_J$, has any additional external pressure added, it will collapse. In most cases the temperature will play a very important role during this collapse, as the critical mass depends more strongly on this parameter. The contraction of a cloud converts potential gravitational energy to heat, and the mechanisms of cooling dominate the speed and size of these collapsing cores. In the early universe, the distribution was mostly primordial (75% hydrogen and 24% helium). Without any metallicity cooling effects such as CO-collisional de-excitation will not help energy escape the molecular cloud, as it is heated by gravitational potential energy during collapse. This means that these early stars, also referred to as population III stars, became much more massive before collapse compared to higher metallicity ones. Implying a very top heavy IMF in the early universe (Zaroubi, 2012), together with a lot of UV radiation abundance due to the stellar sizes. During this epoch, Type II supernovae (exploding stars) were the primary process of metallicity enrichment, for the transition to type II stars, with low metallicity. A typical example of type II stars would be halo stars of spirals or ellipticals. Type II stars are the precursor for type I stars, like the sun, we have today. The decreasing size of type II however means that the metallicity enriching process relies more on type I supernovae (exploding white dwarfs), due to fewer massive stars enriching the universe. The increase in metallicity inherently makes stars collapse faster due to more efficient cooling, which in turn makes them radiate earlier which leads to less time for collecting more gas and dust during collapse.

## 2.3   Initial mass function

For a newly born population of stars, the Initial Mass Function (IMF) describes the distribution of stellar masses entering the main sequence. As stars age and explode, the distribution will change into the Present Day Mass Function (PDMF). Depending on the stellar population age, this difference is mostly in high mass as these run out of hydrogen to fuse first. This specific change also makes it possible to estimate the IMF from the PDMF (Chabrier, 2003), which inherently means the current SED of a galaxy can tell us something about its IMF. We can describe the mass spectrum as the number density distribution per mass interval $\left(\frac{\mathrm{d}n}{\mathrm{d}m}\right)$ with:

$$\xi(m) = \frac{\mathrm{d}n}{\mathrm{d}m} = \frac{1}{m(\ln 10)}\xi(\log m) \qquad \rightarrow \qquad \xi(\log m) \propto m^{-x} \quad \text{and} \quad \xi(m) \propto m^{-\alpha}, \quad (2)$$

Here $n = N/V$ is the number of star per volume, $m$ is the mass in solar masses and $x = \alpha - 1$. In the original Salpeter (1955) paper, a value of $\alpha = 2.35$ was determined, which is believed to relatively accurately describe stars at $m \geq 1M_\odot$. For light from high redshift galaxies, $\alpha = 2.3$ however seems to fit better.

The power law is accurate for high mass stars, but the lower masses are much more abundant, so these also need to be properly described. A typical way is by using a Gaussian distribution, which Chabrier (2003) did in combination with the power law for higher mass using milky way data described as:



Figure 2.2: IMF described by the red striped line from equation (3). The black bars are a normalized histogram from randomly generated stars using the IMF.

$$\xi(m) = \begin{cases} A_1 \exp\left(-\frac{(\log m - \log m_c)^2}{2\sigma^2}\right) & m \leq 1M_\odot \\ A_2 m^{-x} & m > 1M_\odot \end{cases}, \quad (3)$$

Here $A_1 = 0.158^{+0.051}_{-0.064}$, $m_c = 0.097^{-0.016}_{+0.032}$, $\sigma = 0.69^{-0.01}_{+0.05}$, $A_2 = 4.43 \times 10^{-2}$ and $x = 1.3 \pm 0.3$. The normalization of the $A$-constants is $(\log M_\odot)^{-1}\,\mathrm{pc}^{-3}$. The original fit applied to the PDMF, and then converted to the IMF. Major differences only appear between the two distribution at masses equal to or greater than $1M_\odot$. It is for this exact reason that they went with the Salpeter (1955) slope of $x = 1.3 \pm 0.3$, since it fits fairly accurate for high mass stars.

With the definition of the IMF created in this way, the variable $\xi(m)$ is directly proportional to the probability density function $p(m) = \xi(m)/n_{tot}$.

$$\int_{m_{inf}}^{m_{sep}} p(m)\,\mathrm{d}m = 1 \quad \text{and} \quad P(m) = \int_{m_{inf}}^{m} p(x)\,\mathrm{d}x = \frac{1}{n_{tot}}\int_{m_{inf}}^{m} \xi(x)\,\mathrm{d}x\,. \quad (4)$$

5

The lower mass section of IMF's are always the hardest to model, because all the light is emitted by big bright stars, and not the much more abundant small ones. Modelling of IMF's is still a heavily discussed subject today, and is still very open for discussion in low mass stars, and ever evolving with redshift, as the primordial mix is enriched with metals through time.

## 2.4   Ensembles of stars

The basis for what we can see when observing a typical galaxy, is an ensemble of stars interacting with dust and gas, represented by a single observable spectrum. We know exactly how stars form, how long they live and what their spectrum is at any given point in their life cycle looks like using the equations for stellar structure. Applying a given IMF, it is possible to create an artificial galaxy, and evolve it in time to understand how the spectrum changes. This will vary depending on the type of IMF, as well as the efficiency and length of star formation. The simplest example is a galaxy that has undergone a brief starburst, which is then evolved in time. As the large stars die first, and have the most UV radiation, we expect the blue end of the spectrum to drop as time evolves. For observations this might not be a trivial feature to capture at high redshift, due to dense clouds of neutral hydrogen absorbing or scattering a significant fraction of the UV photons.
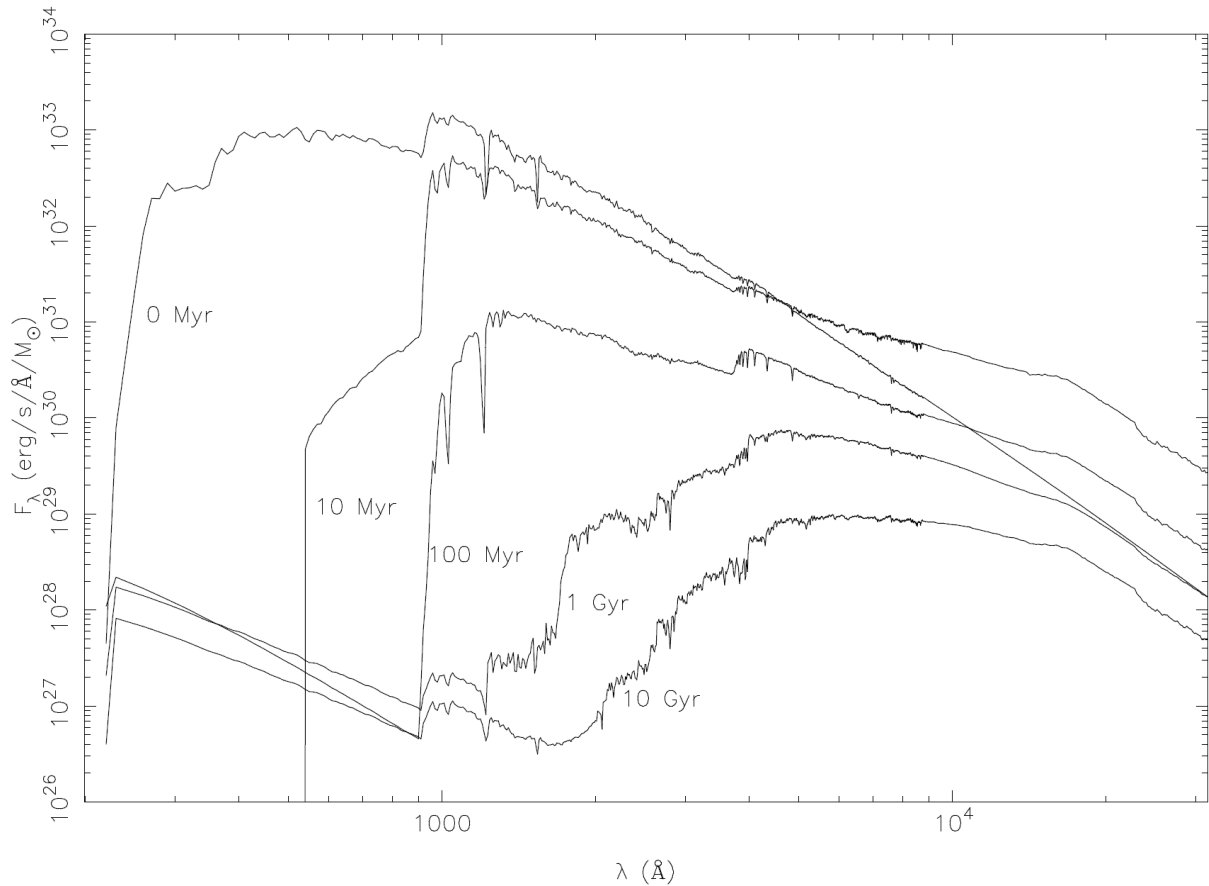


Figure 2.3: Evolution of a galaxy with brief starburst and no gas, taken from Fioc, M and Rocca-Volmerange, B (1997).

As the O and B stars die out the 4000 Å break will become more visible, A and F

stars with blanket metal absorption in their atmospheres, become the dominant sources of light.

In figure 2.3 we find a theoretical evolution of a galaxy containing only stars. The initial conditions are a Salpeter (1955) IMF and a brief starburst, which means no new stars are born after the initial surge. The stars are then evolved along their evolutionary tracks, taking into account all the final branches like the giant asymptotic branches, red, and blue giants. As expected the UV end of the spectrum dies of quickly, and the 4000 Å break becomes visible, as well as the 912 Å break. We don't find the 912 Å from the birth of the galaxy, because the massive stars keep the leftover hydrogen in their atmospheres ionized. The increase in UV flux as we pass 100 Myr, is due to a significant fraction of stars evolving into the asymptotic giant branch phases.

Generally the methodology used in figure 2.3 can be described as integrating over time, mass and star formation:

$$F_\lambda(t) = \int_0^t \int_{m_1}^{m_u} \tau(t - \theta)\phi(m)f_\lambda(m, \theta) \, \mathrm{d}m \, \mathrm{d}\theta \tag{5}$$

Here $\tau(t - \theta)$ is the star formation at time $t - \theta$ per unit time and mass, $\phi(m)$ is the IMF between $[m_1, m_u]$ normalized to $1 M_\odot$ and $f_\lambda(m, \theta)$ is the monochromatic flux of a star at wavelength $\lambda$ with mass $m$. The flux for any star that exceeds its lifetime is set to zero. This integration will work in general for any galaxy, not just the simple short starburst example presented here. The spectrum for a spiral galaxy with moderate but constant star formation rate (SFR), would be much more flat, but would still feature a weak 4000 Å break because large stars are much less abundant with a low SFR. The UV slope depends strongly on the fraction of big stars, with the addition of them dying relatively fast, means the two breaks (1215 Å and 4000 Å) can be used as an estimator for the SFR.

## 2.5   Line emitters

Only specific types of galaxies will have prominent emission lines. Sources with significant amounts of gas are required, since this is where all the emission lines originate. Galaxies with very active star formation, pump large amounts of energy into their surroundings, which causes the gas to be exited. The same goes for Active Galactic Nuclei (AGNs), where the gas is heated by the accreting black hole. The most typical galaxies with emission lines are spirals, with plenty of star formation and gas to excite. Most emission lines will be hydrogen just due to sheer abundance, but other elements become much stronger with increasing metallicity. For galaxies with moderate to high SFR, the hydrogen lines dominate together with forbidden oxygen lines. Gas is excited by the stellar light, and is then emitted from the clouds of gas at discrete wavelengths. Forbidden lines appear because the density of gas is very low, so atoms often stay in a exited state for much longer than the half life of their forbidden transitions, causing them to de-excite. Other strong lines include the 21 cm radio line for cold neutral hydrogen gas ($\sim 10$ K), undergoing a forbidden spin flip in the ground state. Both this feature and CO rotational states play

| Line | $\lambda$ |
|------|-----------|
| Ly$\alpha$ | 1215.670 Å |
| [OII] | 3726.032 Å |
| H$\delta$ | 4101.742 Å |
| H$\gamma$ | 4340.471 Å |
| H$\beta$ | 4861.333 Å |
| [OIII] | 4958.911 Å |
| [OIII] | 5006.843 Å |
| [NII] | 6548.050 Å |
| H$\alpha$ | 6562.819 Å |
| [NII] | 6583.460 Å |
| [SII] | 6716.440 Å |

Table 1:   Common strong galaxy emission lines.

important roles in cooling of cold dense gas, and the very long wavelengths makes them easily penetrate the clouds. There are hundreds of emission lines, which all depend on the metallicity, SFR, and geometry and orientation of the galaxy in question. Dust and gas chemistry play very important roles in the evolution and formation of pre-stellar cores (Doty and Neufeld, 1997).

## 2.6   The Lyα forest

The Lyman alpha forest is a very unique spectral feature for high redshift galaxies. Light emitted at the Lyα-wavelength of 1215 Å or lower, will undergo significant absorption and scattering. As the light travels towards us it will encounter large quantities of intergalactic neutral hydrogen. Since most of this gas is in the ground state, any light at the Lyα wavelength or bluer will be absorbed. These kinds of galaxies are referred to as Lyman Break Galaxies (LBGs).

A consequence of this, will be a lower average flux for wavelengths lower than the Lyα emitted line. Previously this feature has been used to estimate the amount of gas in the intergalactic medium, by looking at high redshift galaxies. Songaila (2004) used 50 quasars between redshifts 2 and 6.3 in order to estimate when the reionization of the universe took place. This was done by measuring the average transmission of the Lyα forest. The transmitted fractions showed a relatively smooth evolution over the entire range of redshifts, which can be modelled with a smoothly decreasing ionization rate.

This feature is exactly the goal of this work, because it will help put a constraint on the flux of the low wavelength filters such as $z$, $y$, $Y$, B and V, compared to the $NB118$ filter. For sources close to redshift 8.79, Lyα-emission line is expected to be in the filter. Since this is one of the largest peaks for many galaxies, we can comfortably require the flux in the filters at lower wavelengths than 1.18 μm, to be equivalent or close to zero.

## 2.7   Extinction

Extinctions is the wavelength dependent absorption and scattering of light, because of interactions with particles (dust). If a photon with $\lambda \ll a$, meets a dust grain of size $a$, then it will most likely be scattered or absorbed. As the wavelength increases, the likelihood of interaction will decrease, giving rise to an extinction law. The strength of interaction will vary with the column density of the dust, which a given galaxy is observed through. The transmission also varies depending on the composition of the dust, but the overall shape is very well defined.
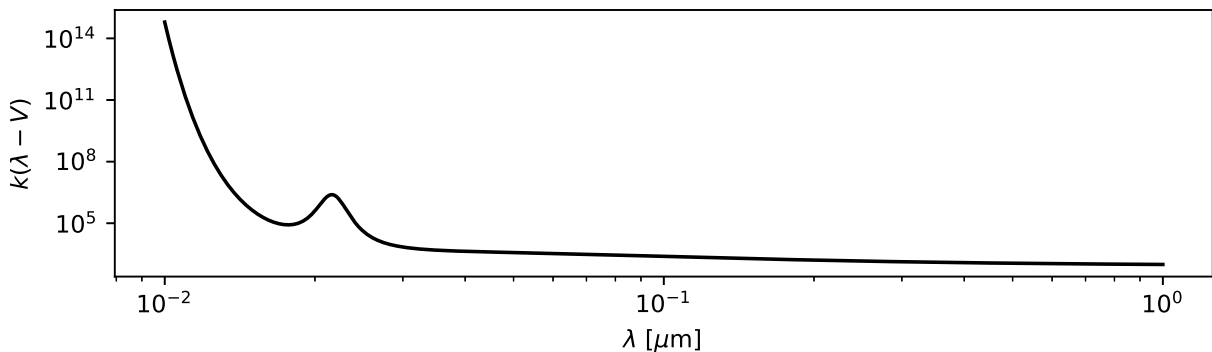


Figure 2.4: Extinction curve $k(\lambda - V)$, using the average LMC values from table 5 in Fitzpatrick and Massa (1990).

For most observations, the extinction curve will feature a prominent peak around 2175 Å seen in figure 2.4. This is believed to be caused by graphite grains in the ISM, but is still discussed as being a unique feature of the milky way (Pitman et al., 2000). Figure 2.4 was created using the symbolic function from Fitzpatrick and Massa (1990), fitted by observing the difference between close stars (with no extinction), and similar stars in a cluster within the Milky Way.

$$k(x - V) = c_1 + c_2 x + c_3 D(x; \gamma, x_0) + c_4 F(x),\tag{6}$$

Here the function $D(x)$ and $F(x)$ are defined as:

$$D(x) = \frac{x^2}{(x^2 - x_0^2)^2 + x^2 \gamma^2} \quad \text{and} \quad F(x) = \begin{cases} a_1(x - 5.9)^2 + a_2(x - 5.9)^3 & x \geq 5.9\,\mu m^{-1} \\ 0 & x < 5.9\,\mu m^{-1} \end{cases},\tag{7}$$

Where $x \equiv 1/\lambda$, $a_1 = 0.5392$ and $a_2 = 0.05644$, and the fit values for (6) are the average fit values for a cluster of stars in the LMC found in table 5 from Fitzpatrick and Massa (1990).

In 2018 a numerical study was conducted considering the likelihood of finding a Ly$\alpha$ emitter in the $NB$118 filter (Laursen et al., 2019). The probability of finding none, one or more emitters were calculated to be 90%, 10% and 1% respectively, whereas previous models predicted between 3 and 20 detections (Nilsson et al., 2007), using a simple light escape fraction between 0.1 and 0.01. In the improved study, theoretical models for halo mass (expected number of galaxy haloes per volume), were combined with high resolution cosmological hydro simulations. This was then post processed with radiative transfer, in order to predict the spectral shapes. This is a big step forward from a simple escape fraction models, and is more realistic. Interestingly, much of the UV radiation to the blue side of the 1216 Å, is converted to Ly$\alpha$
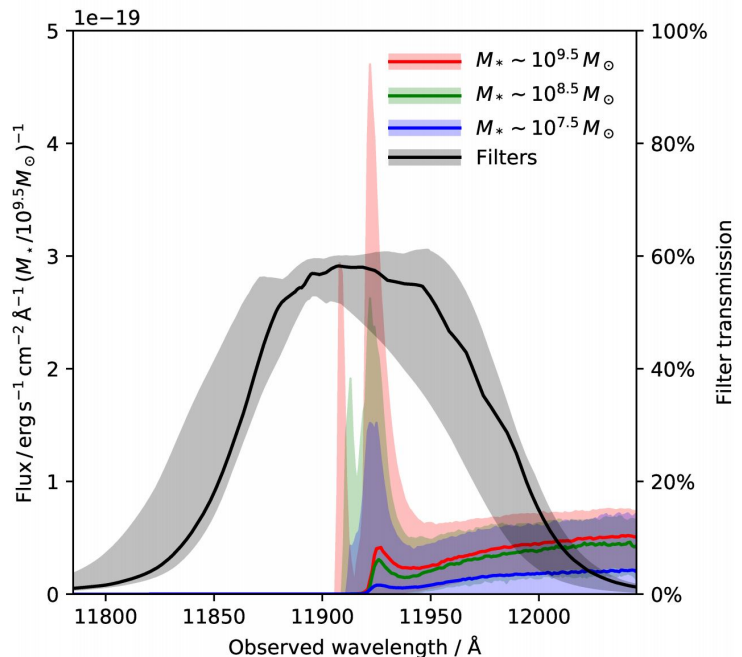


Figure 2.5: Median spectrum of galaxies after correction for radiative transfer from the IGM. Shaded areas indicate where 68% of the lines fall, and the black line represents the median NB118 filter (one filter for each of the 16 CCDs), where the shaded shows the 68%. Figure taken from Laursen et al. (2019).

due to the densities present at the epoch of reionization. While the neutral hydrogen will cut the blue end of the Lyman peak, also the red part can be suppressed due to in-falling gas. This is why the median spectre in fig (2.5) only rarely become very prominent. This ultimately amounts to very little Ly$\alpha$ luminosity escaping the early universe, making the $NB$118 narrowband excess detection much less probable.

## 2.8   Astronomical filters

The main focus on this thesis is the new Ultra deep VISTA data, and therefore also the specific filters. The idea is to only let certain wavelengths of light through, and detect it using a CCD. When multiple filters are used at varying wavelengths ranges, a somewhat crude photometric spectrum can be created. Compared to having a spectrum for each source, this method offers very little data. The broadband filters only supply a single integrated flux across their transmission wavelength range. The major advantage however is the large amount of data that can be collected with just a few filters, over a large area of the sky. This makes data collection both fast and efficient, but crude. Proper observation throughout many wavelength ranges, quickly covers a much broader range than many spectrometers provide from just a single telescope.
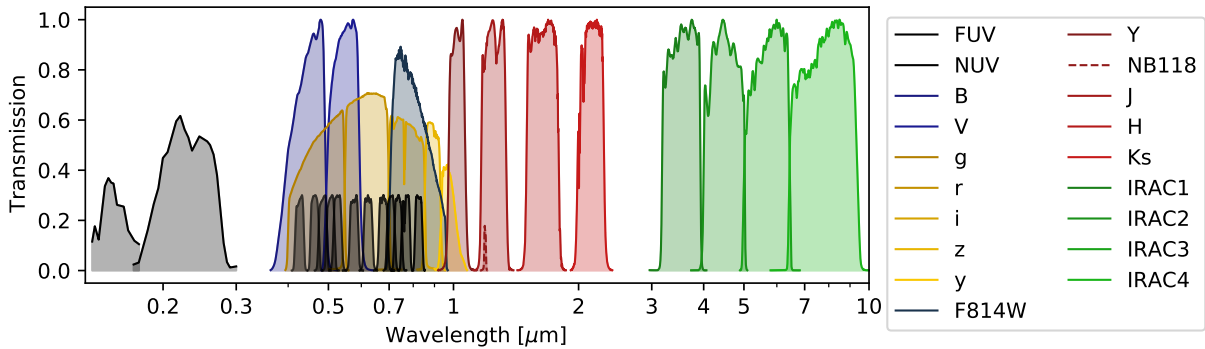


Figure 2.6: Filter curves for most of the important filters used in this study. Excluded here are Suprime-Cam filters ($g^+$, $r^+$, $i^+$, $z^+$ and $z^{++}$) due to overlap.

Each filter can be characterized by a central wavelength, and width defined through its shape. A typical broadband filter will let ~80% of the light in its range through, which then can be collected by a sensor mos suited for that wavelength, typically a CCD or CMOS. Narrowband filters are, like the name suggests, defined at a specific wavelength. The narrow width also makes it more difficult to archive a good transmission of light due to physical limitations. The width is however also its strength. Any emission line that happens to be in the filter, will be very obvious after processing. The wider filters will have much larger normalizations due to their widths, a very bright emission line might not be obvious in a broad filter, especially at low equivalent width. Generally, for any filter $f$ with transmission curve $T_f(\lambda)$, the average flux across the filter and mean value of filter are given by:

$$F_f = \frac{\int T_f(\lambda) F_\lambda(\lambda) \, d\lambda}{\int T_f(\lambda) \, d\lambda} \qquad\qquad \overline{T}_f = \frac{\int \lambda T_f(\lambda) \, d\lambda}{\int T_f(\lambda) \, d\lambda}, \qquad\qquad (8)$$

Here $F_\lambda(\lambda)$ represents the real flux as function of wavelength. Because the average flux across each filter is normalized by the integral of the filter, a sensitivity depending on the filter-curve arises. An emission line in a broad filter will not change the average flux by much, because it accounts for a small area compared to the width of the filter. A narrow filter is much more sensitive to such peaks in the flux, because of the smaller normalization. A good example of this is the $J$-filter, which has the NB118 filter within its width. This enables the estimation of average flux across the width of the $J$-filter, and if the average flux is much brighter for $NB$118, then there is likely a emission line placed exactly within the narrowband range. The far left placement of the narrowband filter

10

with the $J$-filter is also of advantage when looking for high redshift Ly$\alpha$ emitters. Flux will be low for wavelengths lower than the Lyman peak, but since most of the $J$-filter is at longer wavelengths, it can be expected that the average flux across the filter will be non-zero, due to much of the galaxy's spectrum at longer wavelengths than the Lyman peak, being within the filter.

## 2.9   Filters in this study

This work contains a wide selection of filters, ranging from the far UV all that way to the FIR and Radio. Most are shown in table 2, with the exception of filters centred at longer wavelengths than IRAC, due to both depth and relevance to this search. The faint nature of the high redshift survey, means the deepest filters will be in focus. Many of the potential high redshift targets will not be detected in the very high wavelength regime.

Available in the COSMOS2020 data are three different apertures. Firstly `AUTO`, which is produced by `SExtractor` for a specific band, the pseudo flux and flux error are very unreliable and often underestimated. The `AUTO` flux error is however not useless, as it can be used to estimate the pseudo signal to noise ratio $o$, and uncertainty for the $2''$ and $3''$ apertures.

$$o = \frac{1}{\sum_i w_i} \sum_i \left( \frac{\sigma_{\text{APER}}}{f_{\text{APER}}} \right)_i w_i \quad \text{with weight} \quad w_i = \frac{1}{\left( \frac{\sigma_{\text{AUTO}}}{f_{\text{AUTO}}} \right)^2 + \left( \frac{\sigma_{\text{APER}}}{f_{\text{APER}}} \right)^2} \tag{9}$$

The automatic flux extraction is done using the Kron radius (Kron, 1980). This step is necessary to keep the optical and near infrared colours unaffected. The usage of set aperture sizes is necessary to probe the same region for each source, ensuring this colour accuracy.

Depths are calculated using empty apertures clipped at $3\sigma$ in PSF homogenized images (`PSFEx`), and are done separately for the two main apertures ($2''$ and $3''$). Generally the extracted errors are still underestimated, and are corrected depending on each band. The correction factors are the ratio between the standard deviations of the fluxes measured in empty apertures and the median flux errors in the source catalogue, similarly to Laigle et al. (2016).

| Indtrument /Telescope | Band | Central $\lambda$ [Å] | Width [Å] | Depth (2″/3″) ±0.1 | Error Fact. (2″/3″) ±0.1 |
|---|---|---|---|---|---|
| GALEX | FUV | 1526 | 224 | … | … |
|  | NUV | 2307 | 791 | 25.5 | … |
| MegaCam /CFHT | $u$ | 3709 | 518 | 27.8/27.2 | 1.7/2.0 |
|  | $u*$ | 3858 | 598 | 27.7/27.1 | 1.4/1.6 |
| ACS/HST | F814W | 7890 | 1373 | 26.5 | … |
| HSC /Subaru HSC-SSP PDR2 | $g$ | 4847 | 1383 | 28.1/27.5 | 1.4/1.8 |
|  | $r$ | 6219 | 1547 | 27.8/27.2 | 1.4/1.7 |
|  | $i$ | 7699 | 1471 | 27.6/27.0 | 1.5/1.9 |
|  | $z$ | 8894 | 766 | 27.2/26.6 | 1.4/1.7 |
|  | $y$ | 9761 | 786 | 26.5/25.9 | 1.4/1.7 |
| Suprime-Cam /Subaru | B | 4488 | 892 | 27.8/27.1 | 1.5/1.8 |
|  | $g^+$ | 4804 | 1265 | 26.1/25.6 | 5.5/5.8 |
|  | V | 5487 | 954 | 26.8/26.2 | 2.1/2.3 |
|  | $r^+$ | 6305 | 1376 | 27.1/26.5 | 1.6/1.9 |
|  | $i^+$ | 7693 | 1497 | 26.7/26.1 | 1.5/1.8 |
|  | $z^+$ | 8978 | 847 | 25.7/25.1 | 1.5/1.7 |
|  | $z^{++}$ | 9063 | 1335 | 26.3/25.7 | 2.3/2.6 |
|  | $IB427$ | 4266 | 207 | 26.1/25.6 | 2.3/2.6 |
|  | $IB464$ | 4635 | 218 | 25.6/25.1 | 2.3/2.6 |
|  | $IA484$ | 4851 | 229 | 26.5/25.9 | 1.5/1.7 |
|  | $IB505$ | 5064 | 231 | 26.1/25.6 | 1.6/1.8 |
|  | $IA527$ | 5261 | 243 | 26.4/25.8 | 1.7/2.0 |
|  | $IB574$ | 5766 | 273 | 25.8/25.3 | 2.4/2.5 |
|  | $IA624$ | 6232 | 300 | 26.4/25.7 | 1.4/1.7 |
|  | $IA679$ | 6780 | 336 | 25.6/25.1 | 2.5/2.7 |
|  | $IB709$ | 7073 | 316 | 25.9/25.4 | 2.2/2.3 |
|  | $IA738$ | 7361 | 324 | 26.1/25.5 | 1.5/1.7 |
|  | $IA767$ | 7694 | 365 | 25.6/25.1 | 2.1/2.2 |
|  | $IB827$ | 8243 | 343 | 25.6/25.1 | 2.4/2.6 |
|  | $NB711$ | 7121 | 72 | 25.5/24.9 | 1.2/1.4 |
|  | $NB816$ | 8150 | 120 | 25.6/25.1 | 2.3/2.5 |
| VIRCAM /VISTA UltraVISTA DR4 | $Y$ | 10216 | 923 | 26.6/26.1 | 2.8/3.1 |
|  | $Y^{Deep}$ |  |  | 25.3/24.8 | 2.7/2.8 |
|  | $J$ | 12525 | 1718 | 26.4/25.9 | 2.7/2.9 |
|  | $J^{Deep}$ |  |  | 25.2/24.7 | 2.5/2.7 |
|  | $H$ | 16466 | 2905 | 26.1/25.5 | 2.6/2.9 |
|  | $H^{Deep}$ |  |  | 24.9/24.4 | 2.4/2.6 |
|  | $K_s$ | 21557 | 3074 | 25.7/25.2 | 2.4/2.6 |
|  | $K_s^{Deep}$ |  |  | 25.3/24.8 | 2.4/2.6 |
|  | $NB118$ | 11908 | 112 | 24.8/24.3 | 2.8/2.9 |
| IRAC /*Spitzer* | ch1 | 35686 | 7443 | 26.4/25.7 | … |
|  | ch2 | 45067 | 10119 | 26.3/25.6 | … |
|  | ch3 | 57788 | 14082 | 23.2/22.6 | … |
|  | ch4 | 79958 | 28796 | 23.1/22.5 | … |

Table 2: Instruments, telescopes/surveys and filters used in this study, taken from Weaver et al. (subm).

# 3   Telescopes and data collection

In this section the various telescopes and surveys will be described, as well as masking of bright stars and other artefacts present in the COSMOS2020 catalogue.

## 3.1   Masking and filter magnitudes

Filter depths vary from 24.8 on $NB$118 all the way down to 28.1 on HSC $g$, spread over 7 major telescopes and surveys. The various filters, instruments, detectors and observation locations all require various amounts of software to accurately process the information. Some telescopes use numerous CCD's at the same time, which requires complex data pipelines to process. Not all surveys have perfect overlap over the same region, data in some areas is still not complete or missing. Figure 3.1a presents illustrations of how all the surveys overlap, and figure 3.1b which depth each filter reaches.
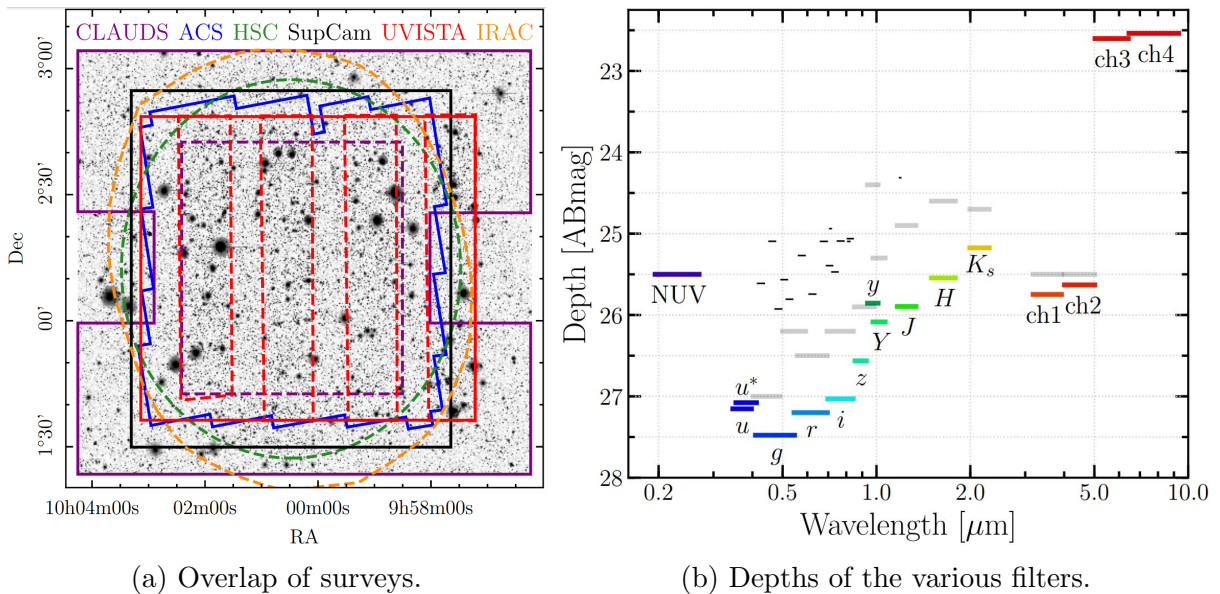


(a) Overlap of surveys.                    (b) Depths of the various filters.

Figure 3.1: (a) The coverage for most of the major surveys included in COSMOS2020 taken from Weaver et al. (subm). The CLAUDS solid line represents the $u$ filter, and the the striped $u^*$. The same applies for $NB$118, where the striped line represents the UVISTA ultra deep strips. Not all data overlaps, and there are areas where certain filters will not be available. The background image is the $izYJHK_s$ detection image. (b) Filters depths at $3\sigma$ in $3''$ apertures included in the `Classic` catalogue. This figure is from Weaver et al. (subm), where the UVISTA depth is in the ultra deep strips, the black bars represent the intermediate- and narrow-band filters. The with of each bar shows the FWHM, and the grey bars are the depths from Laigle et al. (2016). A substantial improvement has been made since COSMOS2015.

With bright stars and other artefacts in the images, the masking process is a very important part of extracting exact and reliable photometry. The masks are made from the HSC data (Coupon et al., 2017), and used to flag sources accordingly. The selection process for masked areas uses GAIA DR2 as a reference catalogue, where stars brighter than $G = 18$ mag are masked. The result is approximately 18% of the sources provided in the COSMOS2020 catalogue, are masked by bright stars and artefacts.

## 3.2   Observational data, surveys and telescopes

Source extraction and photometry done by Weaver et al. (subm) is based on data from a wide range of telescopes and surveys.

The COSMOS field is one of the most observed areas of the sky, with the widest wavelength range and number of filters. Not every telescope can observe every wavelength, so many different ones are used in combination to create a complete set of data. Earth based observation for example, can only be archived from the NIR and visual, including also some wide range of radio frequencies. UV, most of the IR and long range radio all have to be observed from space. CCD detectors additionally have a wavelength range at which they are most efficient, introducing a further constraint to the range of frequencies a single setup can archive. In the next section the various telescopes and surveys will be briefly introduced.

### 3.2.1   ULTRA-VISTA (NIR)

UVISTA DR4 was conducted with the 4.1 m Visible and Infrared Survey Telescope for Astronomy (VISTA), at the Paranal observatory in Chile. The main focus is on the near infrared, with the 1.65° wide VIRCAM instrument. The detector itself consists of 16 (2048×2048) CCD's arranged in a 4x4 pawprint. The pixel scale for the VISTA telescope is about 0.34″. The UltraVISTA survey used in this work uses five filters as characterized in Mc-Cracken et al. (2012), the four main filters ( $Y$, $J$, $H$ and $K_s$) at (0.88 μm, 1.02 μm, 1.25 μm, 1.65 μm and 2.15 μm) and the narrowband $NB$118 filter at 1.18 μm. This survey is very special because it is one of the deepest NIR surveys ever, focusing especially on the $NB$118 narrowband filter as described in Milvang-Jensen et al. (2013). When observing through the $NB$118 filter,
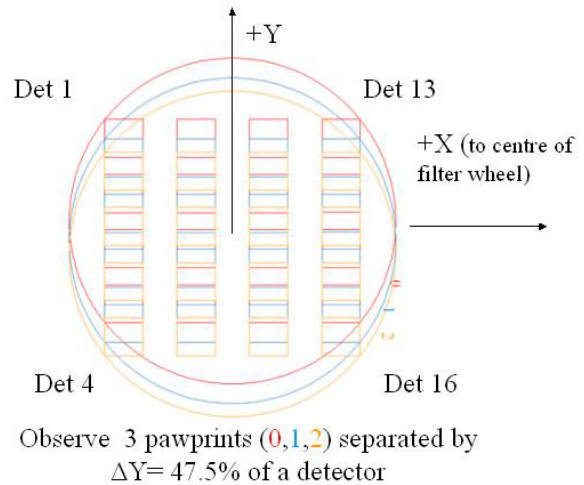


Figure 3.2: At least three pawprint exposures are required to create strips, and at least 6 to complete a fully connected image. Figure taken from J.P.Emerson (2019).

a filter is applied to each sensor, instead of one large for the entire pawprint. This means that there are 16 $NB$118 filters, all with slightly varying filter curves. The median and spread of the curves can also be observed in figure 2.5, and are characterized in detail individually in (Milvang-Jensen et al., 2013).

The telescope uses a detector artefact removal at Garching, that returns astrometrically and photometrically calibrated single pawprints. The data are then passed to UK-based VDFS (VISTA Data Flow System), where the pawprints are combined to a finished image as illustrated in figure 3.2. Currently the latest available data is DR4, which are about 0.8 mag deeper than DR2 used in McCracken et al. (2012).

### 3.2.2 Subaru (NIR/Visual/UV)

The Subaru telescope is a Japanese operated 8.3 m telescope, situated on the summit of mount Mauna Kea on Hawaii. The telescope contributes to the COSMOS survey with several different instruments ranging over a wide wavelengths spectrum. Firstly the 5 Hyper Suprime-Cam (HSC) filters ($g$, $r$, $i$, $z$ and $y$) (Miyazaki et al., 2017) as well as 21 filters from the Suprime-Cam (SC) (Miyazaki et al., 2002). These are split into broad band filters, (B, V, $g^+$, $r^+$, $i^+$, $z^+$ and $z^{++}$), intermediate band filters (IB427, IB464, IA484, IB505, IA527, IB574, IA624, IA679, IB709, IA738, IA767 and IB827) and two narrowband filters (NB711 and NB816).

The new Hyper Suprime-Cam uses 104 (2048x4096) CCD's covering a FOV diameter of 1.5°, corresponding to a pixel scale of 0.17″, with an additional four CCD's for auto guiding and eight for focus monitoring. The older Suprime-Cam, which was decommissioned in 2017, used ten (2048x4096) CCD's, which covered 34′×27′ FOV, with a pixel scale of 0.20″. Currently the latest public data release is included (PDR2) (Aihara et al., 2019) in the catalog.

### 3.2.3 CFHT (NUV)

The Canada-France-Hawaii telescope (CFHT), is a 3.58 m telescope located on Mauna Kea on Hawaii. The MegaCam instrument contributes to the COSMOS data with the CFHT Large Area $U$-band Deep Surver (CLAUDS) (Sawicki et al., 2019), with the filter $u$ and $u^*$, at wavelengths 3709 Å and 3858 Å respectively. Here the $u^*$ filter used in Laigle et al. (2016), now replaced with a new bluer $u$ filter, both are however featured in the COSMOS2020 release. Collected data was reprocessed using GAIA data, and stored in the Canadian Astronomy Data centre after being put through the *Elixir* pipeline (Magnier and Cuillandre, 2004). The wide-field imager, MegaCam, consists of 40 (2048x4612) pixel CCDs, covering a full 1x1 square degree field-of-view with a resolution of 0.187″.

The depth of the two $u$ bands are very deep reaching all the way to 27.8 mag, especially considering the UV being so important for determining crucial factors like the SFR.

### 3.2.4 Spitzer (MIR)

The Spitzer space telescope was a 85 cm infrared telescope operating in the 3.6 μm to 160 μm range. The telescope did several surveys using the IRAC (Infrared Array Camera) instrument, consisting of the 4 channels (ch1, ch2, ch3 and ch4) at respectively (3.6 μm, 4.5 μm, 5.8 μm and 8.0 μm). COSMOS2020 also includes the 3 MIPS (Multi-band Imaging Photometer for Spitzer) channels at (24 μm, 70 μm and 160 μm,).

There has been four large studies observing the COSMOS region, including the 2900 h SPLASH COSMOS survey from Laigle et al. (2016), the COSMOS-IRS survey, the SEDS survey and the S-COSMOS legacy survey done by Sanders et al. (2007).

Gathering data in the infrared band is very important for dust emission, greatly improving photometric redshift fitting and stellar mass estimates.

### 3.2.5 GALEX (FUV/NUV)

GALEX was an ultraviolet focused 50 cm telescope, focusing on near to far UV of the COSMOS field. The telescope has done one comprehensive 100 hour study, to a limiting magnitude of 25.5. It consists of photometry from the two detectors, one being FUV

at 152.8 nm and NUV at 227.1 nm. Due to the difference in the detectors, they have varying resolution, with the NUV being 5.3″ and the FUV being degraded by about 20% to 4.2″. The low resolution also means that for high redshift ($z > 0.7$), galaxies will fill up an entire pixel (Zamojski et al., 2007), making this data somewhat uncertain, due to blending introduced for these small sources.

# 4   Source detection in COSMOS2020

Source detection in COSMOS2020 was done using `SExtractor` or `SEP`, run on an $izYJHK_s$ detection image compiled using `SWarp`. To archive accurate photometry, each catalog PSF homogenizes the data in two different ways using `PSFEx`. Much of the photometry from Laigle et al. (2016) has been reprocessed, and has been years in the making. This thesis has been given exclusive access to preliminary data, which is subject to major changes before release.

## 4.1   The $izYJHK_s$-detection image

The detection image is one of the major differences introduced in COSMOS2020 compared to COSMOS2015. HSC filter $i$ is added for significantly improved detection of bluer objects. Using `SWarp` described in the appendix sec. 9.1, the science images are combined to a single map for source extraction in the two included catalogues. The COSMOS2020 image uses a new setting not described in the appendix called `CHI_MEAN`, introduced by Szalay et al. (1999). This method converts the science images and weight maps to create a chi squared distribution. The filters are then averaged and combined to one final image with accompanying weight map. This chi squared image aims to identify flux and eliminate probably sky pixels, and produces one of the most accurate detection images.

## 4.2   The `Classic` catalog

In this section we will describe the process of how the `Classic` catalog, used for this work has been extracted.

Initial source extraction on this image is done using `SExtractor`, with `DETECT_THRESH` of 1.5 and `DETECT_MINAREA` set to 5. This results in a total of 1 720 700 sources, from which 970 597 are within the HSC and Ultra VISTA overlapping regions.

With the sources extracted, the internal flags created by `SExtractor` can be used to identify stars and masked areas. The catalogue contains two types of flags, firstly each filter gets an two bit flag (`###_FALGS`, where the `###` represents each filter, for example: `UVISTA_J_FLAGS`), these can be found in table 8 in the appendix. Secondly masking flags ranging values between 0 and 1, where 0 is for sources within a certain survey and clear of bright stars. Here there are flags for each of the three deepest surveys `FLAG_HSC`, `FLAG_SUPCAM` and `FLAG_UVISTA` with a combined `FLAG_COMBINED`. Additionally a flag for the ultra deep areas of Ultra VISTA `FLAG_UDEEP`. Requiring `FLAG_COMBINED`=0 would select sources located within the HSC data, SUPCAM data, UVISTA data and not near any bright stars.

The next step is to PSF homogenize the images, applying `PSFEx` (section 9.2.5) using polar shapelet basis functions. Saturated stars have already been removed by `SExtractor` during the initial detection, and now just the artefacts left over from these remain. The target PSF is set to a Moffat profile, which describes how a point source spreads out. The parameters used are $(\theta, \beta) = (0.8'', 2.5)$, where $\theta$ is the FWHM. `PSFEx` then returns the PSF models, which `SExtractor` can apply to the images during the main source extraction. Some of the Suprime-Cam medium band channels have issues when the spatial variability in the PSF across the field is ignored. This was negated by running `PSFEx` on each individual exposure, to avoid the resulting large errors on the aperture photometry, which could be as high as 0.1 mag.

With the new extracted PSF models, the final source extraction is performed using `SExtractor`. This is done several times, using the automatic aperture mode built in, as well as manual 2″ and 3″ circular apertures. The errors extracted are underestimated if the noise in an image is correlated, so the magnitudes and fluxes are scaled with band-dependent corrections. The correction factors are the ratio between the depths and the median flux errors in the source catalog and can be found in table 2. All of these operations are done separately for all apertures, and also for the VISTA deep and ultra-deep strips.

In order to get the IRAC data implemented, software called `IRACLEAN` (Hsieh et al., 2012) was used. The PSF for IRAC images is larger than the optical data, and has significant confusion noise (noise related to the astronomical background). To tackle this issue, `IRACLEAN` uses a high-resolution image (and its segmentation map) as a prior to identify the centroid and the boundaries of the source, and iteratively subtract a fraction of its flux ('cleaning') until it reaches some convergence criteria specified by the user. The general assumption is that any source is the PSF convolved with a Dirac delta function. A target $S/N$ of 2.5 was set for this process, and will usually reveal new sources as a result of the cleaning. `IRACLEAN` uses the high resolution detection image as well as a segmentation map, the image is split into tiles in order to parallel the processing of the images.

## 4.3   The Farmer catalog

`SEP` is used for source extraction, which is a python port of `SExtractor` (Barbary, 2016). Detection is only performed on the Ultra VISTA area to minimize the processing in bright HSC masked areas, in order to avoid inhomogeneities in source modelling. Settings are set to be the same as `Classic`. A count comparison reaches agreement to within 0.01% at 893793 total, for the same UltraVISTA area and without regions containing HSC masked star halos.

After source detection and masking, the data is further processed for blended sources to improve flux measurements. `The Farmer` looks for crowded regions, and fits all the sources in that area at the same time, by dilating the source segmentation map to avoid double counting of pixels of the blended sources.

The major difference between the two catalogs is the PSF homogenization, where `The Farmer` convolves the models with the PSF in each band. A spatially constant PSF is created for each $u$, $u^*$, all HSC and UltraVISTA bands again using `PSFEx`. For the IRAC data, a similar procedure to `IRACLEAN` called `PRFMap` is used. PSF's are rescaled to match the 0.15″ pixel scale of the mosaics. For all sources, both resolved and unresolved, a best fitting model is determined:

- **PointSource** models are taken directly from the PSF, appropriate for unresolved sources.
- **SimpleGalaxy** round, exponential light profile with 0.45″ effective radius. This model represents the middle ground between point source, and very small sources.
- **ExpGalaxy** full on exponential fit with angle, axis ration, effective radius and flux.
- **DevGalaxy** similar to the previous model, but with a fixed de Vaucouleurs light profile.
- **CompositeGalaxy** combination of *ExpGalaxy* and *DevGalaxy*. These models are concentric and share centroid, added here is the bonus of having both a total flux

parameter and a fraction of total flux. The two component models have their own effective radius, axis ratio and position angles.

Models are fitted to give the most accurate shape parameters and the most likely shape of each source. The fit models draw constraints from each of the bands in the $izYJHK_s$ detection image, to avoid that the selection function is not adversely affected by choice of a particular band. Best model determination is done using a decision tree, starting with the two most simple ones. Then moves on to the three more complex ones using the same initial conditions. To avoid any bias from this usage of the same initial conditions, refitting is performed on the best model.

Flux is extracted in these finished models, in each desired band separately. For each band the model centroids are allowed to vary with a strict 0.3pix Gaussian prior to avoid catastrophic failures. This is done to mitigate the subtle astrometric offsets between different image bands. `The Farmer` does not contain apertures like `Classic`, but only the best fit model for each source.

## 4.4   Photometric redshift

Photometric redshifts are computed for both the catalogues extracted in Weaver et al. (subm). Fluxes are firstly corrected at each object position using the Schlafly and Finkbeiner (2011) dust map. A total of four redshift catalogs are thus created, two for each extraction using `LePhare` and `EAZY`.

### 4.4.1   `LePhare`

`LePhare` (Arnouts et al., 2002; Ilbert et al., 2006) is a template fitting code, that uses multiple interpolated spectroscopic redshifts to build a higher resolution spectrum. Included are star forming galaxies, ellipticals, spirals and two exponentially declining SFR quiescent galaxies. Parameters like the milky way reddening ($E(B-V) \leq 0.5$) and the $2175\,\text{Å}$ bump are also included (Fitzpatrick and Massa, 1986). The code is also sensitive to emission lines, which makes it very well suited for the $NB118$ data which is only included in the `Classic LePhare` photometric redshift catalog. Template fitting for stars, white and brown dwarfs are also performed, to the limitations presented by Saumon and Marley (2008).

Fitting is performed using the least squares method, where efficiency can be improved through indicators like HSC $i$-magnitude and half light radius to select which templates to apply for fitting.

$$\chi^2(z, T, A) = \sum_{f=1}^{N_f} \left( \frac{F_{\text{obs}}^f - A \times F_{\text{pred}}^f(z, T)}{\sigma_{\text{obs}}^f} \right)^2 \tag{10}$$

Here $F_{\text{obs}}^f$ and $\sigma_{\text{obs}}^f$ are the observed flux and the associated error for filter $f$, $F_{\text{pred}}^f(z, T)$ is the predicted flux for any template $T$ convoluted with filter $f$ redshifted to $z$ and $A$ is a normalization factor. Each filter is redshifts with a z-grid of step-size 0.01, to a maximum of $z = 10$. To make up for uncertainties in the colour modelling, $0.02\,\text{mag}$ are added in quadrature to the photometric errors of the data in the optical, $0.05\,\text{mag}$ for $J$, $H$, $K_s$, ch1, and the three narrow bands, and $0.1\,\text{mag}$ for ch2. For the fit itself, flux is used as to not introduce upper limits. Offsets are calculated for each filter to increase the precision using galaxies where the spectroscopic redshift is known, to more accurately

predict correct fluxes. This is done iteratively until convergence is reached, by minimizing the difference between observed and predicted fluxes. These offsets are located in table 2 in Weaver et al. (subm), but do not play a huge role for the means of this work. Lastly with the half light radius as an estimator, any template that fits better as star ($\chi_s^2 < \chi_g^2$), and is detected within $3\sigma$ in $K_s$ or IRAC ch1, will be marked as such instead of galaxy.

A very important output from `LePhare` is the likelihood function $\mathcal{L}(\text{data}|z)$, which yields the likelihood that a galaxy lies at a given redshift.

### 4.4.2   EAZY

`EAZY` is the second photometric redshift code,using the updated version running in python (Brammer et al., 2008). The underlying process is the same as `LePhare`, the templates are convoluted with each filter, redshifted over a pre-defined grid, and fitted at each step. The best fit is found using least squares similarly to equation (10), the templates are however not pre-defined for `EAZY`. Often the uncertainty in the redshift can be chalked down to a single template not fitting perfectly with the measured SED. `EAZY` aims to mitigate this issue by creating linear combination of several templates as described by:

$$\chi_{z,i}^2 = \sum_{f=1}^{N_f} \left( \frac{T_{z,i,f} - F_{\text{obs}}^f}{\sigma_{\text{obs}}^f} \right)^2 \quad \text{with linear combination} \quad T_z = \sum_{i=1}^{N_{\text{temp}}} \alpha_i T_{z,i} \qquad (11)$$

Here $T_{z,i,f}$ is the synthetic flux in filter $f$ at redshift $z$ for template $i$, $\alpha_i$ are the coefficients used to create the linear combination, that is $T_{z,i,f}$, from the templates $T_{z,i}$. `EAZY` contains 17 derived stellar population synthesis models (Conroy et al., 2009; Conroy and Gunn, 2010), that span broadly over a wide range of attenuations, ages and star formation histories.

Similarly to `LePhare`, a likelihood function is also produced, $\mathcal{L}(\text{data}|z)$, it is however corrected using a prior assumption. Many galaxies fit equally well at $z = 0$ and $z = 3$, because the templates are unable to distinguish blue colours redward of the Balmer and Lyman breaks respectively. To combat this a Bayesian approach is implemented, a prior assumption is made about the magnitude of the source, and its likely redshift based upon it, $p(z|m_v)$.

# 5   Selection and detection

## 5.1   Searching for $Y$-band dropouts

We will now try to describe the process of selecting $Y$-band dropouts in the search for Ly$\alpha$ emitters.

This will be a two part section, where the primary goal is to identify LBGs with narrowband excess Ly$\alpha$ emittance at redshift 8.8 and secondly a broader search for $Y$-dropouts.

The general criteria for the two are the same, because the $NB$118 filter is within the width of the $J$ filter. This will make them both surveys $Y$-dropouts, where we require the flux at bluer wavelengths to be consistent with zero within some uncertainty requirement. This could for example be a harsh requirement for all bluer filters than $J$ to be consistent with zero within $3\sigma$:

$$f_i - 3\sigma_{f_i} \leq 0 \quad \text{and} \quad f_J \geq 3\sigma_J. \tag{12}$$

Here the denotion $i$ represents any filter with a mean wavelength value lower than $J$, excluding $NB$118. We will however develop slightly on this approach, by instead using a weighted mean, but also having some requirements for flags and placement.

Certainty that the source exists is key, so avoiding masked areas, or borders of photometry due to inadequate S/N ratio is important. Therefore we require that `FLAG_SUPCAM`=0 and `FLAG_HSC`=0, but not `FLAG_UVISTA`, because the masks are cut too aggressively, and the interest here is in maximizing the amount of potential LBGs. This will have the additional repercussion of yielding some very noisy border sources in the $NB$118 strips, which will be visually easy to identify later. Sources with measured spectroscopic redshift are also ignored, for obvious reasons. This is done by using the `z_spec` column within the `EAZY` catalog, which we require to be less than 0 (not determined). These spectroscopic redshifts are from multiple articles, but compiled by Muzzin et al. (2013) currently at v4.1. As a final step, the detected flux must be free of errors such as nearby stars, saturated pixels or neighbour bias, therefore `UVISTA_J_FLAGS`=0.

To make any errors and variations in the various filters less important, the weighted mean is used to find the $Y$-dropouts. All the filters blue-ward of $J$ are combined this way, whereafter the resulting weighted mean $(\overline{f})$ with associated uncertainty $\sigma_{\overline{f}}$, must be consistent with zero within $2\sigma$:

$$\overline{f} - 2\sigma_{\overline{f}} \leq 0 \quad \text{where} \quad \overline{f} = \frac{\sum_i w_i f_i}{\sum_i w_i} \quad \text{and} \quad w_i = \frac{1}{\sigma_{f_i}^2}. \tag{13}$$

The weighted mean yields a more forgiving approximation of the flux at shorter wavelengths, and the occasional single filter that might be detected within the 3 sigma threshold proposed in (12) will not necessarily exclude a source from selection.

### 5.1.1   $NB$118 emitters

The narrowband magnitudes excess of $J$ and $NB$118 are not trivial, because the centres do not perfectly align. The $Y$ flux can be used to assist in gauging the narrowband flux for continuum SED's, because $NB$118 is on the blue side of the $J$ filter. This is done by observing the average trend of galaxies, as described in Milvang-Jensen et al. (2013). This is more difficult for the red side, as there is no clear-cut way. So here the brute force method is used, where a lot of galaxies with emission lines are redshifted over the filters,

and best fit values are determined to most accurately describe the expected ($J$-$NB$118) excess to the actual one. This can be boiled down to the symbolic expression in equation (14).

$$(J - NB118)_{\text{corr}} = \begin{cases} (J - NB118) + 0.34(Y - J) & \text{if } (Y - J) \leq 0.45 \\ (J - NB118) + 0.153 & \text{if } (Y - J) > 0.45 \\ (J - NB118) + 0.07 & \text{if } Y \text{ Not detected} \end{cases} \qquad (14)$$

This equation derived for line emitters below redshift 2.2, but this has no influence on our results for the LBG search, and helps describe the narrowband excess more accurately for the lower redshift sources. Determination of narrowband excess of an object is same as Milvang-Jensen et al. (2013).

$$(J - NB118)_{\text{corr}} \geq 0.2 \,\text{mag} \quad \text{and} \quad (J - NB118)_{\text{corr}} \geq 2.5\sigma_{(J-NB118)_{\text{corr}}} \qquad (15)$$

The error for equation (14) is the standard error propagation on $Y$, $J$ and $NB$118. The minimum value of 0.2 mag is the least required to be sure of narrowband excess.

Firstly all these criteria will by applied for the Ly$\alpha$ emitter search with $NB$118 excess. Then another search without any requirements on $NB$118, in the search for $z \sim 8$ $Y$-band dropouts.

## 5.2   Using $NB$118 to find new sources

Another application for the narrowband data is using it as a base for source extraction. Very faint galaxies with strong emission lines, are not easily detected in the broadband filters, because the emission line makes up a small portion of the total bandpass. Adding to this, filters are also normalized with their area in the transmission wavelength space, which further decreases what little flux is detected. With the narrowband data, and a strong emission line located in the filter range, it is theoretically possible for such a faint galaxy to be visible. It therefore makes sense to use the deep $NB$118 strips as the detection image.

The extraction is done using the python package `SEP` (Barbary, 2016), using both the science image and the weight map of $NB$118. Due to ram limitations source extraction is done in smaller squares, more specifically six, along each of the four strips totalling 24. The squares are sized 3000x3000 pixels, which leads to overlapping of a few hundred pixels. Double detected sources are eliminated, where the source with the least distance to the centre of a square is kept. This ensures that a source close to the border, is not picked over one that is fully in frame in another detection square.

A detection threshold of 2.5, and minimal area of 5 pixels and a de-blending contrast ratio of $10^{-5}$ are used as extraction settings. While only the narrowband is necessary, flux is extracted for all VISTA filters, $Y$, $J$, $NB$118, $H$ and $K_s$. Firstly the background is estimated using the `SEP background` function, and the `KRON_RADIUS` is calculated. Extraction is done using circular apertures with a radius 2.5 times the `KRON_RADIUS`. In order to imitate the `Classic` catalog closely, a minimal radius of 3.5 pixels is set for each aperture. This corresponds to a approximately $1''$ aperture diameter at the $0.15''$ pixel scale of the science images. Choosing the same aperture as Weaver et al. (subm), also means that the error factor is the same, so a factor of 2.8 is applied to the extracted flux error. The UVISTA data has a AB magnitude zero point of 30, so the extracted pixel count can readily be converted using:

$$m(F_p) = 30 - 2.5 \log_{10}(F_p) \quad \text{and} \quad \sigma_m(F_p, \sigma_{F_p}) = \frac{25}{4\log(10)} \frac{\sigma_{F_p}}{F_p}. \qquad (16)$$

Almost all sources will overlap with the Weaver et al. (subm) `Classic` catalog, so a 20 mag, $3\sigma$ selection criteria is implemented on the $NB$118 flux to save on calculations. A source is considered to be detected in `Classic`, if there is a source centre from `Classic` within 5 `KRON_RADIUS` of the narrowband extraction. Finally before visually inspecting the sources, filtering of artefacts and masked areas is applied. This is done by checking flags of the surrounding `Classic` sources, and removing the masked areas.

## 5.3   Identify lensed objects with $NB$118 narrowband excess

It is known that the COSMOS field contains many lensed objects (Faure et al., 2008), and searching for them is something that can be done using narrowband excess. By looking for galaxies with significant narrowband excess, which do not poses a photometric redshift, corresponding to a significant emission line in the $NB$118 filter. The idea is that the photometric redshift fits the lensing galaxy, while the lensed galaxy produces a narrowband excess, nowhere near the redshift range of the host galaxy.

$$
\begin{aligned}
(z_{\text{phot}} - 2\sigma_{z_{\text{phot}}})\lambda_{\text{line}} &< \lambda_{NB118} - \sigma_{\lambda_{NB118}} \\
(z_{\text{phot}} + 2\sigma_{z_{\text{phot}}})\lambda_{\text{line}} &> \lambda_{NB118} + \sigma_{\lambda_{NB118}}
\end{aligned}
\tag{17}
$$

This is going to yield a handful of different kinds of sources; lensed galaxies, non-Deblended galaxies and critically failed photometric redshift fitted galaxies. The procedure is completed by taking the lines from table 1, figuring at at which redshift range they will have emission in $NB$118, and excluding any sources that lie within these ranges. Then the $5\sigma$ $NB$118 and $3\sigma$ $J$ criteria is applied, also including the narrowband excess from equation (15) requirement to identify emitters. The used photometric redshift is the zPDF from `Classic LePhare`.

# 6   Results and discussion

To initially understand this very large amount of data better, we recreated a histogram of $K_s$ brightness selected galaxies from Weaver et al. (subm) (chapter 9.3). The histogram compares the number of galaxies with $K_s$ magnitude without stars. For the COSMOS2020 data this was trivial, because both `LePhare` and `EAZY` supply best fit models, and flag stars. We selected stars in a $YJHK_s$ colour-colour diagram, and reached a very good fit. The details of all this are in the appendix, and are not paramount to the main results of this work.

In the next few sections we will now present and characterize the three main results of this thesis.

## 6.1   $Y$-dropout $NB$118 emitters

131635 sources in the `CLASSIC` catalogue are detected with positive flux within $3\sigma$ in $J$ and $5\sigma$ in $NB$118, and out of these 5065 sources with significant narrowband excess in accordance with equation (14). As predicted with low likelihood by the Laursen et al.
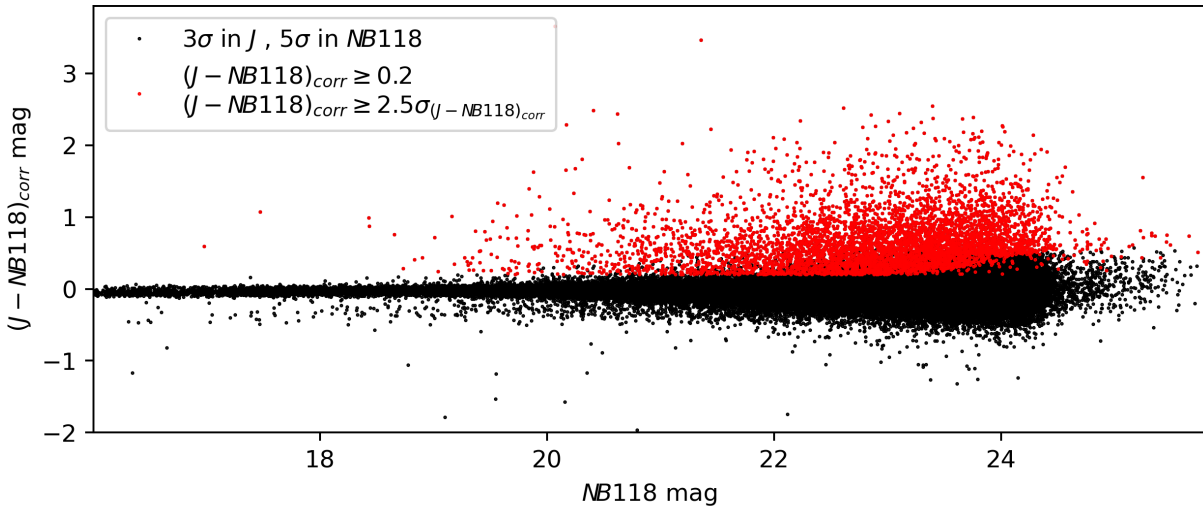


Figure 6.1: This figure represents the corrected narrowband excess magnitude versus $NB$118 magnitude. Black dots are sources detected at $3\sigma$ in $J$ and $5\sigma$ in $NB$118, red dots are the narrowband excess criteria of equation (15).

(2019) numerical study, the search for $Y$-dropouts with narrowband excess detection was not successful. The strict requirement for narrowband excess and significance did not include a clear single clear LBG at sufficient redshift. When calculating the weighted mean from equation (13), none of the galaxies selected in figure 6.1 reached a flux below zero, even when subtracting $3\sigma_{\overline{f}}$. This also implies that the requirement for all filters blueward of $NB$118 should have been consistent with zero within $3\sigma$ (equation (12)), would have yielded no results. Because if the weighted mean of all the bluer filters is too significant, then there will for certain also be at least one filter detected at equal significance.
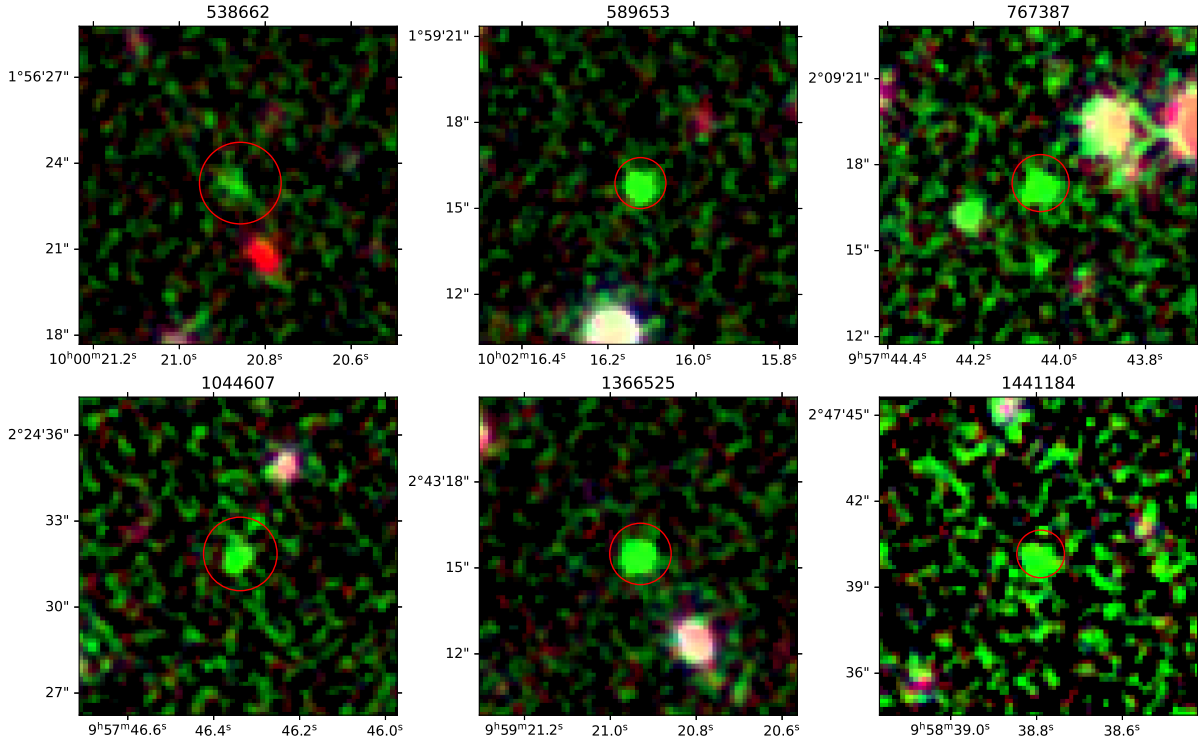
Figure 6.3: $K_s$, $NB$118 and $Y$ RGB images of the six sources, with their respective IDs in the `Classic` catalogue. Each of the sources is a high EW emitter, and very clearly spotted as a real source.

So the selection criteria was altered to yield SEDs with significant narrowband excess instead, but also low flux blueward of $NB$118. Instead of searching for fluxes at bluer wavelengths to be consistent with zero, we require that they are a factor of 8 lower than the $NB$118-flux. The value of 8 is somewhat arbitrarily chosen, and yields 6 sources, with some objectively interesting emitters. A seventh source was technically in this set, but it is located in the noisy edge of the narrowband stripes, and was therefore excluded.

Combining the large required ratio between the $NB$118 flux and the shorter wavelength filters, yields a sample of high-EW line emitters. The magnitude plots



Figure 6.2: Narrowband excess vs NB118 magnitude for the 6 line emitters.

will be represented for all detected filters for both `The Farmer` and `Classic`. If a filter is detected within 2 sigma or more it will be marked green and displayed with $1\sigma$ significance. For filters detected at less than $2\sigma$, a mark is set at the $2\sigma$ upper limit, and an arrow pointing downward to indicate that the source has a magnitude less than the marker within 98% confidence (lower than $2\sigma$ upper limit). In figure 6.2 we see the narrowband excess for the 6 sources. All of them are around 2 mag, making the emission
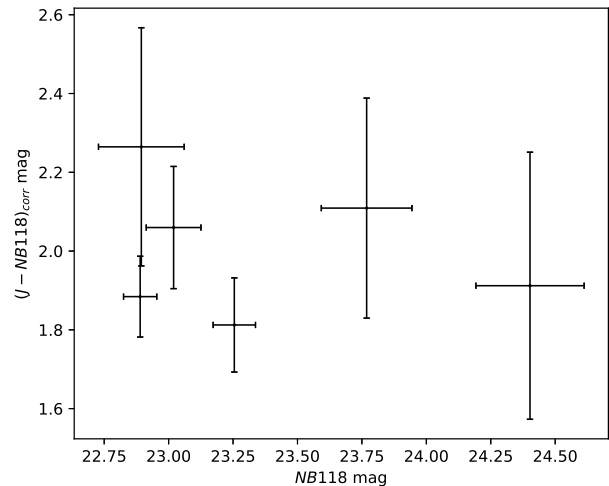
25

lines about 6 times brighter as the average flux of the galaxy in that wavelength region. Table 3 provides a quick overview of coordinates and photometric redshifts. The sources will be briefly characterized, and the job of the photometric redshift codes discussed.

|  |  |  | Classic | | The Farmer | |
|---|---|---|---|---|---|---|
| id | Ra , Dec (J2000) | Classic ID | LePhare | EAZY | LePhare | EAZY |
| $S_1$ | 10:0:20.9 +1:56:23:17 | 538662 | $1.75^{0.4}_{0.3}$ | $1.59^{0.05}_{0.06}$ | $1.86^{0.09}_{0.1}$ | $1.48^{0.1}_{0.05}$ |
| $S_2$ | 10:2:16.1 +1:59:15:67 | 589653 | $0.814^{0.008}_{0.009}$ | $0.847^{0.008}_{0.03}$ | $0.821^{0.02}_{0.03}$ | $0.319^{0.5}_{0.01}$ |
| $S_3$ | 9:57:44.1 +2:09:17:21 | 767387 | $0.817^{0.5}_{0.01}$ | $1.38^{0.05}_{1}$ | $0.334^{0.02}_{0.02}$ | $1.42^{0.04}_{0.03}$ |
| $S_4$ | 9:57:46.3 +2:24:31:73 | 1044607 | $1.64^{0.7}_{1}$ | $1.89^{0.2}_{0.1}$ | $0.03^{0.04}_{0.02}$ | $0.274^{0.03}_{0.03}$ |
| $S_5$ | 9:59:20.9 +2:43:15:37 | 1366525 | $0.815^{0.009}_{0.009}$ | $1.46^{0.02}_{0.03}$ | $1.37^{0.05}_{0.1}$ | $1.48^{0.004}_{0.01}$ |
| $S_6$ | 9:58:38.8 +2:47:40:00 | 1441184 | $1.29^{0.1}_{0.1}$ | $1.4^{0.04}_{0.05}$ | $1.3^{0.05}_{0.2}$ | $1.4^{0.05}_{0.07}$ |

Table 3: The six flux ratio selected sources. The table gives own identifier, J2000 coordinates, the `Classic` ID, and photometric redshift fit values for both catalogs.

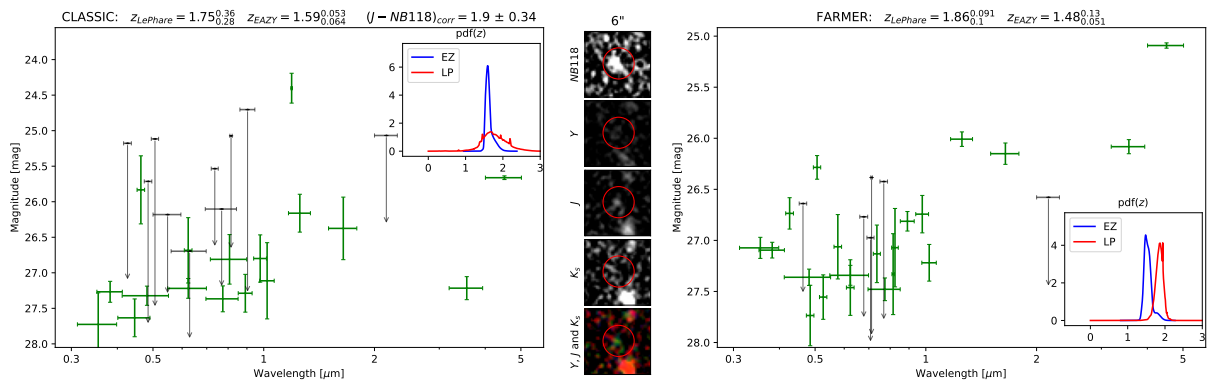Enlarged SEDs are in the appendix figures: 9.7, 9.8, 9.9, 9.10, 9.11 and 9.12

### 6.1.1   $S_1$



Figure 6.4: $S_1$ Classic ID: 538662, Ra,Dec (J2000)=(10h 0m 20.87s, 1° 56′ 23.17″).

This galaxy features a reasonably strong 4000 Å break indicating either an elliptical galaxy or a gas poor spiral. The strong emission line does however indicate a galaxy with gas. Initially this might seem like a high redshift galaxy, but sufficiently many are detected blue ward of the drop around 1 μm, which indicates a low SFR galaxy. This is especially visible in the `The Farmer` SED, where a good handful of intermediate filters are detected, which would not be expect for very high $z > 8$ galaxies. `LePhare` implies that the emission line is approximately around 4300 Å corresponding to a H$\gamma$ emission line. The `LePhare` also shows some small narrow peaks, where the emission lines likely are trying to fit, but the overall shape of the SED seems to have had a bigger importance here. The H$\gamma$ emission line is not very strong, which is probably why the pdz does not feature a visible additional peak at the $z = 1.75$ peak. The `The Farmer` extraction is much less in agreement. `EAZY` indicates a lower redshift, but has a small tail overlapping with `LePhare`. The small peaks in the `LePhare` pdz is likely from the intermediate band filters, as they are rising above the SED of the galaxy.

### 6.1.2   S$_2$
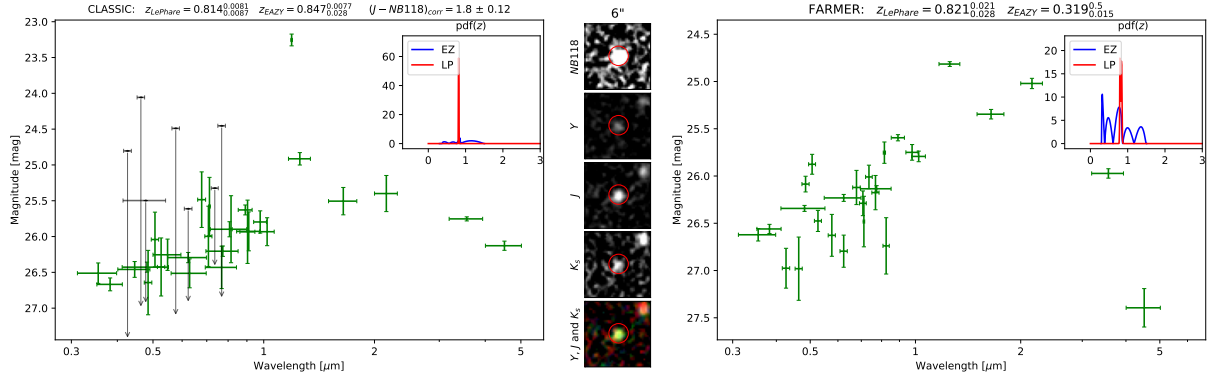


Figure 6.5: S$_2$ `Classic` ID: 589653, Ra,Dec (J2000)=(10h 2m 16.13s, 1° 59′ 15.67″).

A very bright high EW H$\alpha$ emitter at redshift 0.8. It becomes very apparent for a galaxy like this, that the addition of the narrowband $NB$118 gives a strong result for the photometric redshift. With the width of the $J$-filter being as wide as it is, the SII line will also be within this filter, which might be another part of the reason that also the flux of $J$ is so elevated. The two emission lines are only 279 Å apart at this redshift, and the FWHM of the $J$ filter is 1718 Å. For `Classic` both redshift extractions work fine, even `EAZY` without the $NB$118 does a fine job of fitting the clearly elevated flux of $J$, corresponding to the emission lines. The extra peak for `EAZY` at $z > 1$ presumably originates from the small decline in overall flux to the blue side of 1 µm. The code is likely trying to fit a SED break for low SFR galaxies, meaning a 4000 Å break, but the emission line barely won out. **The Farmer LePhare** did a good job at also finding the H$\alpha$ line, however `EAZY` had major trouble. The code presumable saw the SED as mostly flat, and tried to fit multiple emission lines to the data. The SED itself also seems to be a bit more uncertain here, with some of the intermediate bands indicating possible lines.

### 6.1.3   S$_3$
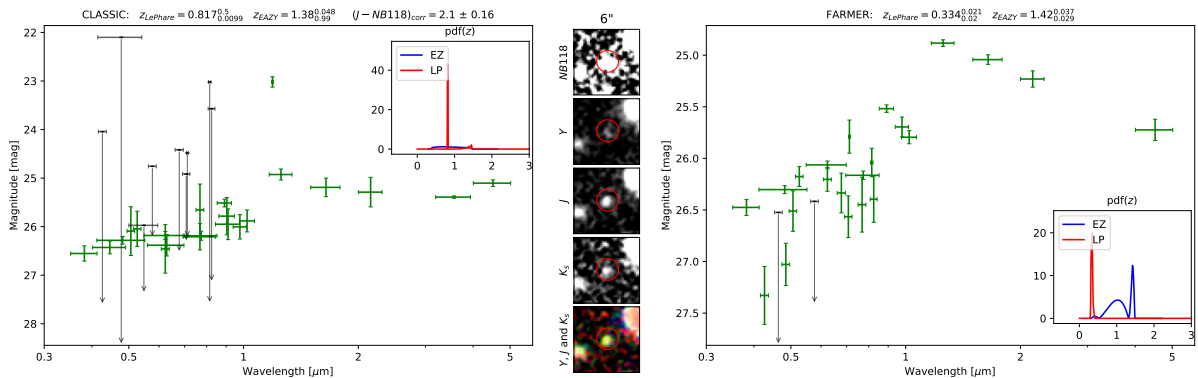


Figure 6.6: S$_3$ `Classic` ID: 767387, Ra,Dec (J2000)=(9h 57m 44.05s, 2° 9′ 17.21″).

This detection was very difficult for the codes to figure out. `Classic LePhare` with the strong emission line approximates this source to be a H$\alpha$ emitter at redshift 0.8, with very strong confidence. The SED is generally relatively flat, and contain a sufficient

amount of low uncertainty filters to be certain of this shape. With obvious features like the narrowband, we here see the codes without $NB118$ struggling to find a good fit. $\texttt{EAZY}$ aims for higher redshift for both catalogues, but only achieves a high EW [OIII] fit at $z = 1.4$ for $\texttt{The Farmer}$. The lower redshift ($z < 1$) seems more probable as many galaxies have a flat SED at lower redshift in this wavelength range.
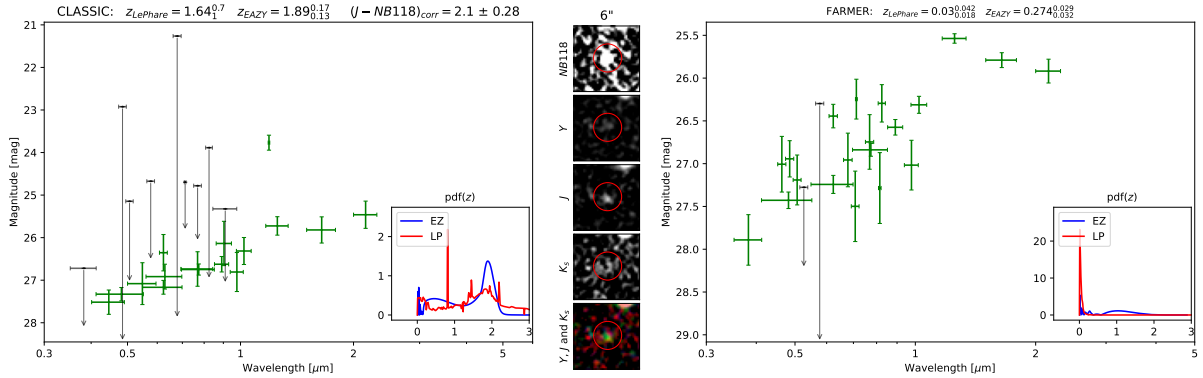
### 6.1.4   $S_4$



Figure 6.7: $S_4$ $\texttt{Classic}$ ID: 1044607, Ra,Dec (J2000)=(9h 57m 46.34s, 2° 24′ 31.73″).

A significantly fainter source than previous, which possible indicated a bit higher redshift. The problem is again the same, where many template shapes have this vague arching shape, and features like emission lines play a huge role in determining the best fit model. This is a really great example that shows the inner workings of these codes when a clear emission line in visible. The overall shape of all filters makes the pdz rise and fall slowly as they start to fit together, but the strong emission line gives peaks in the goodness of fit. For $\texttt{Classic}$ both codes show the same overall bump at just before redshift 2. Both codes would have shown about the same fit value at around 1.89. The lack of detected IR data (ch1 and ch2) adds to the uncertainty in these fits. The sharp peak at $\sim 0.8$ for $\texttt{LePhare}$ does not seem to matter to the best fit galaxy template. The value that minimizes the chi squared distribution of the pdz is however at 0.8 (H$\alpha$). Peaks around redshift 1.4 and 2.2 are also visible corresponding to [OIII] and [OII].
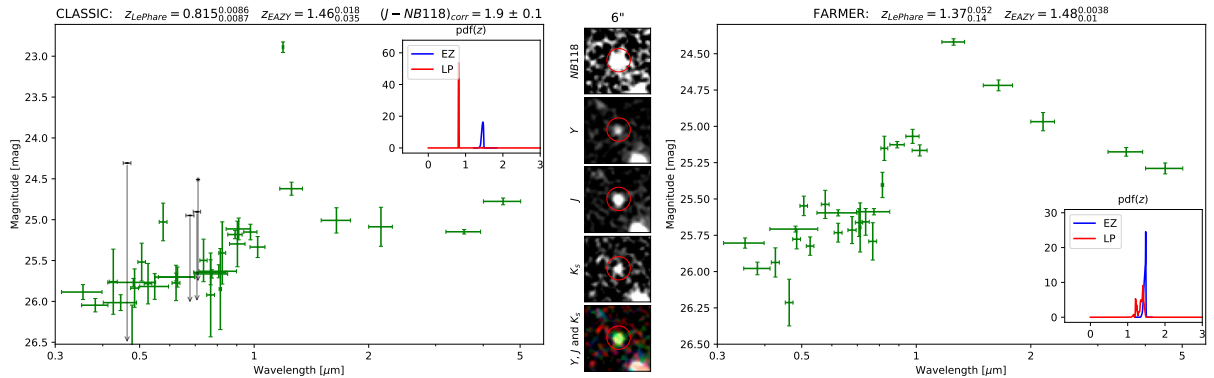
### 6.1.5   $S_5$



Figure 6.8: $S_5$ $\texttt{Classic}$ ID: 1366525, Ra,Dec (J2000)=(9h 59m 20.94s, 2° 43′ 15.37″).

This is a very high EW emitter with significant narrowband excess, with large disagreement between `Classic LePhare` and the three other fits. For both catalogues a redshift around 1.4, is found, except for `LePhare Classic`, where the extreme emission line becomes dominant in the fit due to the low uncertainty on the $NB118$ filter. At $z = 0.815$ with very low uncertainty, it seems `Lephare` has found a very well fitting H$\alpha$ line, and might be using dust obscuration to make up for the drop in flux at bluer wavelengths. In this case the redshift is more likely around 1.4, because of the strong [OIII] and H$\beta$ lines, together with the 4000 Å break, which also explains the subtle drop on the left side.
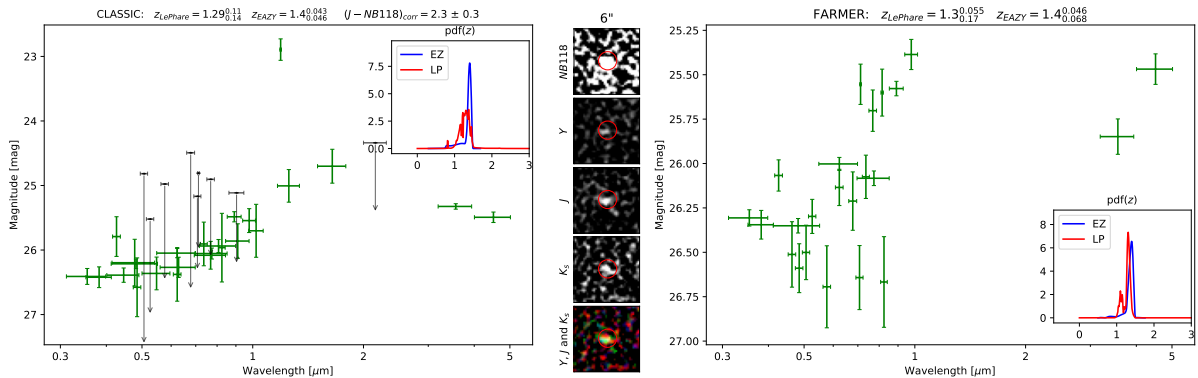
### 6.1.6   $S_6$



Figure 6.9: $S_6$ `Classic` ID: 1441184, Ra,Dec (J2000)=(9h 58m 38.8s, 2° 47′ 40″).

Another galaxy around redshift 1.4, where `LePhare` has issues because of the strong emission line for the `Classic` catalogue. The agreement between the other codes seems to be very good, the lack of detection of any VISTA filters gives a big gap in the 1-3 µm range. With all codes fitting to around 1.4, we see that the overall dropping SED left of 1 µm probably is the 4000 Å drop, for galaxies with low SFR. `The Farmer` for this area has a mistake in PR2 used in this study, while the source is obviously visible in all filters shown in the thumbnails (figure 6.9), the top and bottom few arcminutes are missing from the data in `The Farmer`, and flux could therefore not be extracted for VISTA.

## 6.2   $Y$-band dropout LBGs

While the search for Ly$\alpha$ emitters did not pan out for the sources detected at $5\sigma$ in $NB118$, the broader search of $Y$-band dropouts in the UVISTA filters revealed potential. Some LBGs in this chapter are located within the ultra deep $NB118$ strips, and they can therefore give an indication the Lyman limit likely subsides within the $J$ filter. The pdz in this section will have some added bars, to indicate at what redshift the Lyman break would lie for each relevant filter. Table 4 provides a quick overview, following discussion and characterization of the potential LGBs.

|    |    |    | Classic | | The Farmer | |
| --- | --- | --- | --- | --- | --- | --- |
| id | Ra , Dec (J2000) | Classic ID | LePhare | EAZY | LePhare | EAZY |
| $J_1$ | 10:01:47:49 +2:10:15:44 | 784810 | $8.37^{0.5}_{0.6}$ | $8.26^{0.4}_{0.3}$ | $8.67^{0.2}_{0.3}$ | $8.46^{0.4}_{0.4}$ |
| $J_2$ | 10:00:01:87 +2:16:14:77 | 895880 | $5.09^{3}_{2}$ | $3.13^{0.9}_{0.6}$ | $7.64^{0.1}_{0.1}$ | $2^{6}_{0.2}$ |
| $J_3$ | 09:58:59:78 +2:30:44:46 | 1151107 | $8.85^{0.6}_{0.6}$ | $8.51^{0.4}_{0.5}$ | $8.62^{0.2}_{0.3}$ | $8.46^{0.3}_{0.4}$ |
| $J_4$ | 10:00:15:97 +2:43:32:92 | 1371152 | $8.33^{0.8}_{2}$ | $8.4^{0.4}_{0.4}$ | $9.29^{0.2}_{0.2}$ | $9.36^{0.2}_{0.2}$ |
| $J_5$ | 09:57:43:02 +2:49:02:44 | 1464572 | $8.24^{0.5}_{0.5}$ | $7.9^{0.7}_{0.1}$ | $7.18^{0.7}_{3}$ | $4.03^{6}_{2}$ |

Table 4: Table over the five selected $J$ LBGs. The table gives own identifier, J2000 coordinates, the `Classic` ID, and photoetric redshift fit values for both catalogues and both codes.

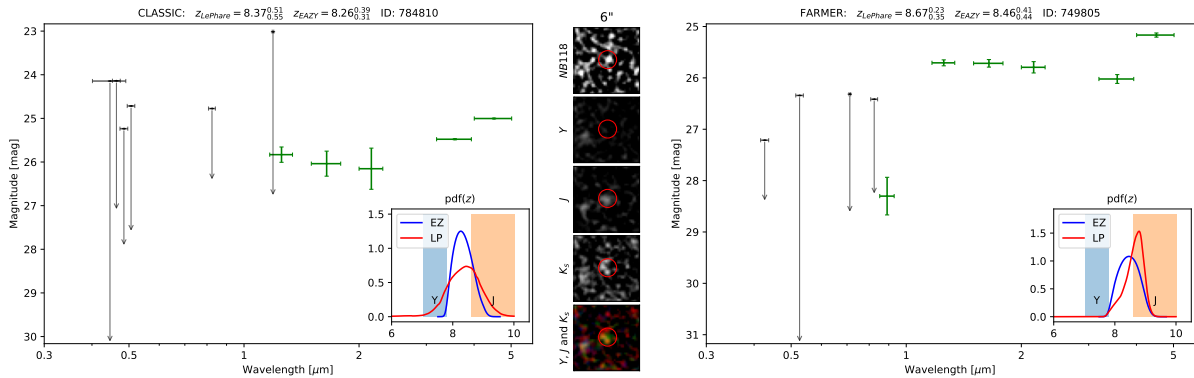Enlarged SEDs are in the appendix figures: 9.13, 9.14, 9.15, 9.16 and 9.17

### 6.2.1   $J_1$



Figure 6.10: $J_1$ `Classic` ID: 784810, Ra,Dec (J2000)=(10h 1m 47.49s, 2° 10′ 15.44″) This source is also known as UVISTA-Y10 and is part of the brightest $z \geq 8$ galaxies in Ultra-VISTA (Stefanon et al., 2019) and also UVISTA-598 from Bowler et al. (2020).

This is the first of three sources that lie within the ultra deep strips. With a only a $0.6\sigma$ detection of $NB118$, we here give the filter an arrow like any other filter detected at less than $2\sigma$. In `Classic` the $Y$ band is not detected at all for this source, and it is clear what that means if we look at the pdz. `EAZY` does not have the narrowband, so it bases the Lyman break between the detected filter $J$, and the non-existent $Y$ flux. Approximately the same behaviour is seen for `LePhare`, with a slightly higher weight towards the blue side of the $J$ filter. Even with the high uncertainty of $NB118$, an indication that the

SED might already be dropping within the FWHM of the $J$ filter can be extrapolated, because the value of the $NB118$ filter lies below that of $J$, if very uncertain. There is no way to be completely confident that all flux below the 1215 Å line is absorbed through the universe, which is probably what can be see in the `The Farmer` extraction. The single filter detected here is the HSC $z$ at $2.5\sigma$. Which is not a problem as long as $z < 9.6$ where the 912 Å Lyman break completely lies within this filter, where absolutely zero flux is expected.

### 6.2.2   $J_2$



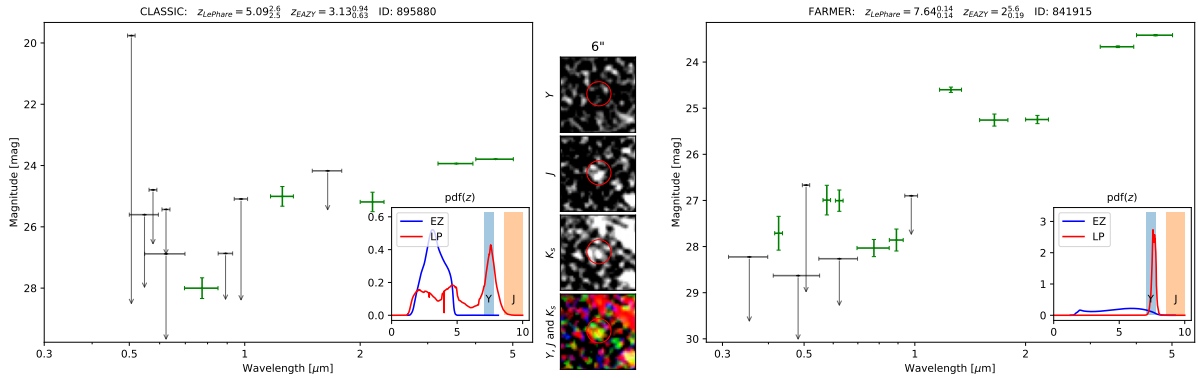Figure 6.11: $J_2$ `Classic` ID: 895880, Ra,Dec (J2000)=(10h 0m 1.874s, 2° 16′ 14.77″).

This is the most irregular SED out of the five, and is very likely not a $z \simeq 8$ emitter. This source is an example of the method being too soft, and allowing a few detected filters to still give a positive dropout. The cut value $(m - 2\sigma_m)$ was about $-0.001\,\mu\text{Jy}$, meaning just below the upper limit of 0, while the other 3 were less than $-0.01\,\mu\text{Jy}$. It was considered to remove this source, but the SED is worth discussing even if it is not a candidate. `Classic LePhare` places this source at $z \sim 5$ which corresponds to a Lyman break at less than $0.74\,\mu\text{m}$, which corresponds to the Lyman peak placement in the single HSC $i$ filter detected blueward of $J$. According to the pdz however, the most likely redshift seems to be around 7-8, placing the pdz peak for `LePahre` of both catalogs around this point. `EAZY` is does not seem to find a good fit for any of the two catalogs. The source seems very large in the thumbnails, which is another indication of a lower redshift source.
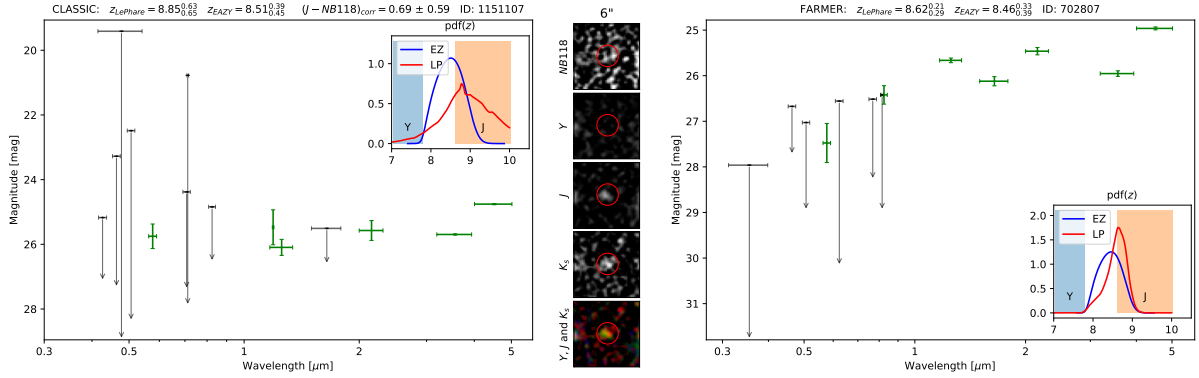
### 6.2.3   $J_3$



Figure 6.12: $J_3$ `Classic` ID: 1151107, Ra,Dec (J2000)=(9h 58m 59.78s, 2° 30′ 44.46″).

This is the source containing most potential to be a $z \simeq 8.8$ emitter. $NB$118 is detected at just $2.01\sigma$, just barely avoiding the selection process for a $2\sigma$ arrow, but the SED looks extremely promising with $J$ detected at $4.4\sigma$. While unlikely true, the current value of $NB$118 could be due to a Ly$\alpha$ emission line at narrowband excess 0.69. This could potentially be exactly what we set out to find in this work, even if lacking in exposure time to be statistically significant. The single filter marked green for `Classic` is $IB$574 detected within $2.86\sigma$ which is cause for some concern due to what should be expected from a LBG. With a filter detected to this degree within almost 3 sigma, is something unexpected considering the Lyman limit lies at $0.87\,\mu\text{m}$ at redshift 8.5. Now while a single seemingly statistically significant filter detected beyond the Lyman limit seems like an issue, because absolutely no flux should be expected beyond that point, the zphot codes still predict a high redshift. All 4 classify this as a LGB, even with the intermediate band detections in both images, `The Farmer` also getting a $4.9\sigma$ detection in $IB$827. The uncertainty on these fits is relatively high, because its difficult to gauge where the dropout is located, they do however all predict this source to be above redshift 8 or higher with 93% certainty, apart from `Classic LePhare` at 88%. We can see in the pdz for this fit that there is a tail to the lower side of $z \simeq 8.5$, this is likely because the code is using the additional data from $NB$118 to to more accurately fit a Ly$\alpha$ emission line. The other codes have to make due with the broad filters to find a redshift. This is exactly why we see broader peaks for the three other pdz's.
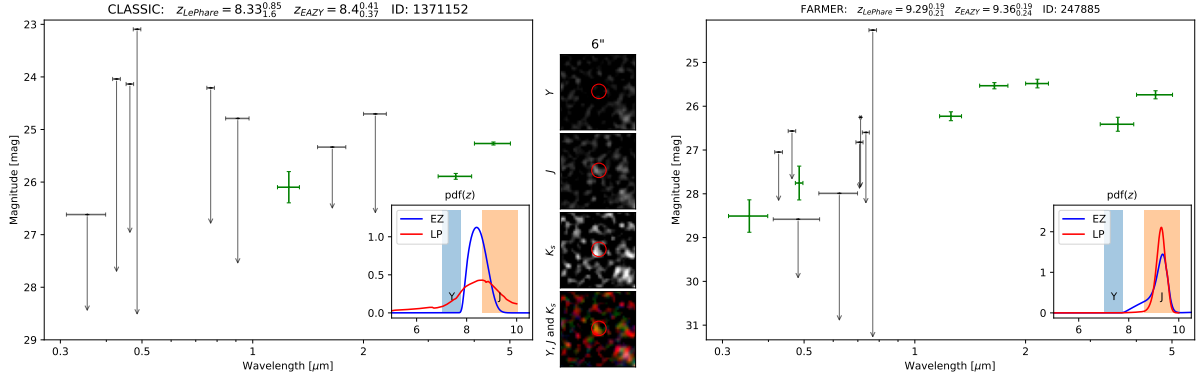
### 6.2.4  $J_4$



Figure 6.13: $J_4$ `Classic` ID: 1371152, Ra,Dec (J2000)=(10h 0m 15.97s, 2° 43′ 32.92″) Also known as UVISTA-Y12 and is part of the brightest $z \geq 8$ galaxies in Ultra-VISTA from Stefanon et al. (2019).

A good example of what we set out to find in the `Classic` SED, while not detected within $2\sigma$ for $H$ and $K_s$, it still has the characteristic $J$ detection at $3.7\sigma$ with no filters detected within 2 sigma blueward. Remembering that the arrows point towards the magnitude value they are extracted at, notice how the SED is majorly different between `Classic` and `The Farmer`. $J$ and ch1 (ch1 and ch2 being the right-most green markers in in both catalogs) are approximately at the same magnitude, however `Classic` detects $H$ and $K_s$ 1 mag lower for both. This causes the overall fits to be very different between the two. For `Classic`, the flux increases from $H$ to $J$, and so the fit will attempt to get some of the Ly$\alpha$ emission captured in the $J$ filter to make up for the increasing flux, hence the pdz predict the redshift to be on the left side of the $J$ filter range. `The Farmer` has a decreasing $J$ flux compared to $H$, so here the models predict that the Lyman drop lies within the filter FWHM. Agreement between `LePhare` and `EAZY` is very good between to two SEDs. There is is no doubt this is a $z > 8$ emitter.
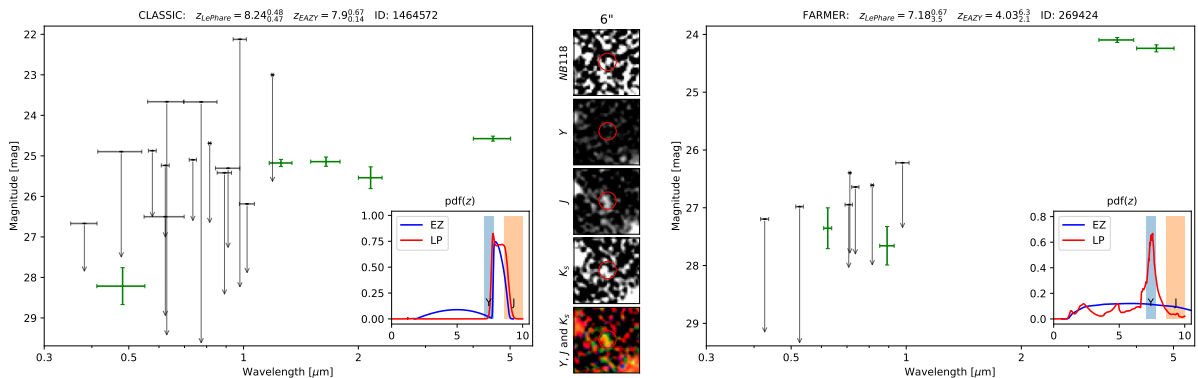
### 6.2.5  $J_5$



Figure 6.14: $J_5$ `Classic` ID: 1464572, Ra,Dec (J2000)=(9h 57m 43.02ss, 2° 49′ 2.439″).

The `Classic` SED looks very good, even with this many filters detected. The single detected filter here is HSC $g$ (at $2.4\sigma$), which is also the deepest filter for COSMOS2020.

Additionally the pictures of the source itself appear to contain a questionable source. It is obvious especially from $J$ and $K_s$ that the detection is between two larger sources, and it could almost reassemble outflows from one of them. It is however separated enough to be considered isolated in the end. For `Classic` there is a detection for the $Y$ filter, it is just detected at less than $2\sigma$. Obvious from the pdz of both codes, they attempt to place the redshift to be just at the red side of $Y$. This is to simulate that the $Y$ filter lies just to the blue side of the Lyman peak at $1215\,\text{Å}$, with just the blue tail being captured within the filter. The additional tail in `LePhare` reaching just within the $J$ filter is likely $NB118$ implying that the Lyman drop is located closer to the $J$ filter. In the end, $Y$ is only detected at $1.2\sigma$ and $NB118$ at $.8\sigma$, which makes both of them very uncertain. While this is very likely a high redshift galaxy, the SED for especially `The Farmer` leaves a lot to be desired. This specific source also suffers from the boundary mistake of $S_6$, so it unfortunately has no flux extracted for the VISTA data. There is not much point speculating over the `The Farmer` SED, because it simply lacks all the important filters for a source at this redshift.

## 6.3   Using $NB118$ as extraction image to find new galaxies

A total of 147962 sources were detected in the $NB118$ image. From these, 61104 are detected within $3\sigma$ and are fainter than 20 mag. Out of these, 10810 were duplicates, yielding a total of 50294 detected sources.
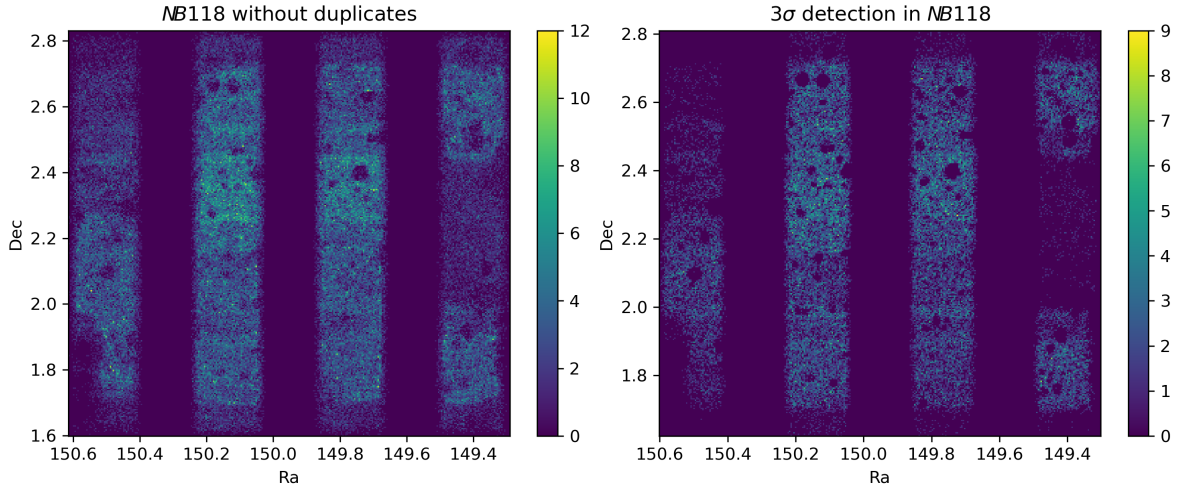


Figure 6.15: 2D histograms of the $NB118$ extracted source coordinates. The left side represents all sources detected by `SEP`. The right side is all the sources detected at minimum of $3\sigma$ in $NB118$ and fainter than 20 mag. Figure 9.6 in the appendix also contains the sources corresponding to the right side histogram, plotted over a $YJK_s$ RGB image, which shows stars and star halos.

The number of total detected sources is much lower than `Classic` and `The Farmer` in the same region. This is what should be expected, because the chi mean detection image of COSMOS2020 uses 6 broad filters, which are all on the order of 1 mag deeper. Figure 6.15 shows where the sources are extracted across the entire $NB118$ field. The $3\sigma$ lower limit for the sources create areas where more depth becomes easily spotted, the density of sources is much higher in the central area where the sky brightness is not so significant. The overlap among selected narrowband extracted sources and `Classic` is 99.4%, with the remaining .6% still containing artefacts. The artefacts in the images are not included in the `Classic` catalog for obvious reasons, but were detected for certain by `SExtractor`. After removal of these, 114 potential new sources remain. These are detected in ranges between $3\sigma$ up to $5.3\sigma$. The 36 most statistically significant sources, are located in the appendix in figure 9.4.

After the detection process is completed, artefacts, outliers and failed detections also need to be identified. Artefacts are easily removed, due to high luminosity in all VISTA filters. Some detected sources are clearly noise, and even with background subtraction, detection threshold and weight maps, some very questionable sources end up in the finished catalog. In figure 6.16 are 5 manually picked sources, specifically selected for 2 reasons, statistical significance, and overall visual appearance. Especially $NB_1$ and $NB_2$ are visually very believable, and represent the best candidates for real sources not detected in `Classic`. $NB_1$ is somewhat integrated into the surrounding noise, or is possibly an extended object. Its overall bright appearance and circular central shape, does however make it very good candidate. This was also the most statistically significant source, which was neither masked or an artefact. $NB_3$ is, while still a good detection, a good example of how many of the detected sources can be characterized. It has a few pixels just above the 2.5 detection threshold, and is encased in an area with otherwise low signal. This causes the source to visually seem more significant than it might be in reality, and yields a more significant detection from the flux extraction.

Out of the five $NB$118 extracted sources, the least robust one is $NB_4$. The shape is somewhat circular, but has a lot of additional connected signal to the left, it either seems extended, or is possibly just a noise object. It barely reaches above the noise of the surrounding area, but has a central point at the detection threshold, making it a very questionable object.

While the composite RGB using the $J$ band is not supposed to show anything for these sources, because that would imply detection in the `Classic` catalogue, a little dot can be observed for $NB_2$ in the $Y$, $J$ and $K_s$ image composite. It is clearly very small, but is located just within the detected source aperture. The photometry however tells a different story, where none of the filters, except
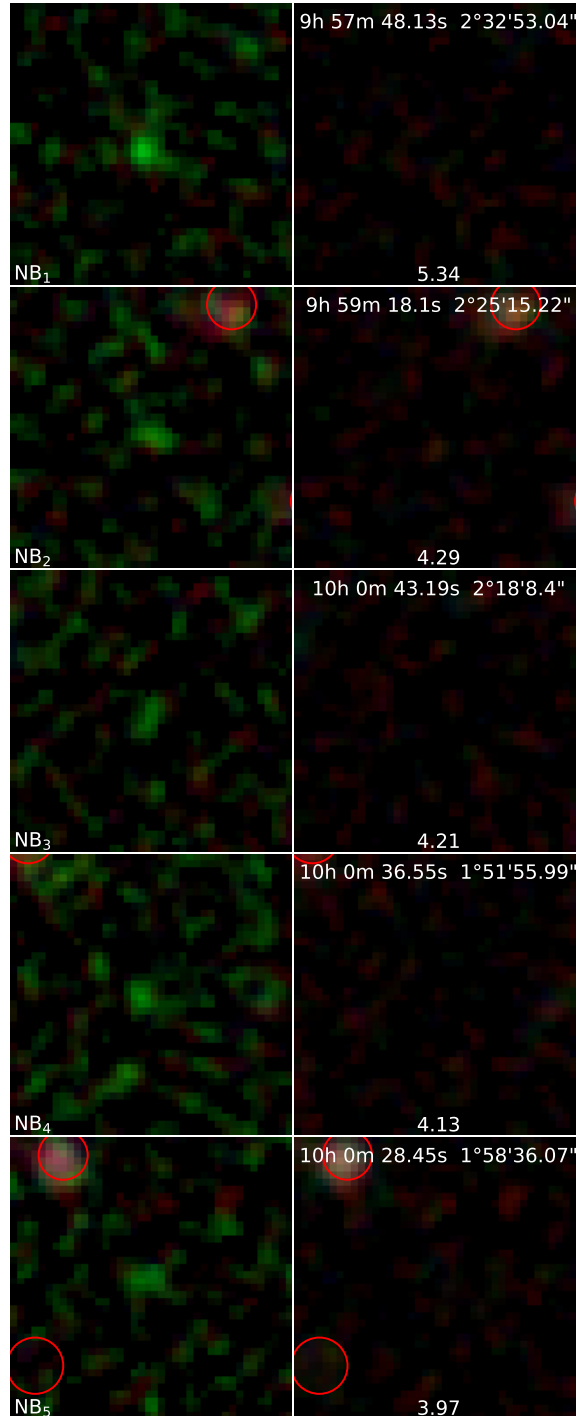


Figure 6.16: Hand-picked NB sources. Top text shows the coordinates for each set, the bottom is statistical detection significance. Left is $Y$, $NB$118, and $K_s$ while right is $Y$, $J$, and $K_s$, meaning the same source is represented with each horizontal column. Red circles represent the `Classic` detected sources, each source extracted in this figure is exactly at the centre of the thumbnails.

36

$NB$118, are detected with a positive flux after background subtraction. This is one way of realizing that visual inspection can be deceiving, but the undetected flux is also what should be expected with this method. Most objects will be at the border of detection in the narrowband image, and are therefore almost always questionable. The only two detections with flux detected are $NB_1$ and $NB_5$.

These five sources represent very good candidates, that almost certainly will end up being actual bright emission line galaxies. There is no way to be absolutely certain of their existence, without more integration time, or dedicated observations of these detections. An overview of them is summed up with significance of flux extraction for each VISTA filters in table 5.

| id | Ra Dec (J2000) | $\sigma Y$ | $\sigma NB118$ | $\sigma J$ | $\sigma H$ | $\sigma K_s$ |
|-----|----------------|-----------|---------------|-----------|-----------|-------------|
| $NB_1$ | 09:57:48:13 +2:32:53:04 | 1.33413 | 5.34352 | 6.20934 | 1.68679 | nan |
| $NB_2$ | 09:59:18:10 +2:25:15:22 | nan | 4.13434 | nan | nan | nan |
| $NB_3$ | 10:00:43:19 +2:18:08:40 | nan | 3.96798 | nan | nan | nan |
| $NB_4$ | 10:00:36:55 +1:51:55:99 | nan | 3.81721 | nan | nan | nan |
| $NB_5$ | 10:00:28:45 +1:58:36:07 | 0.934267 | 3.76416 | 6.7795 | 2.21932 | nan |

Table 5: The first two rows represent the identifier from this work and the J2000 coordinates. The last four are the detected significance of each filter using the same aperture. nan values represent failed detections.

During the source extraction, the Kron radii were tweaked and changed heavily depending on the detection limits and photometry extraction apertures. It therefore makes sense to compare the size of sources detected in this work, with the `Classic` detection sizes, and see if anything can be gained from the differences in detection images.
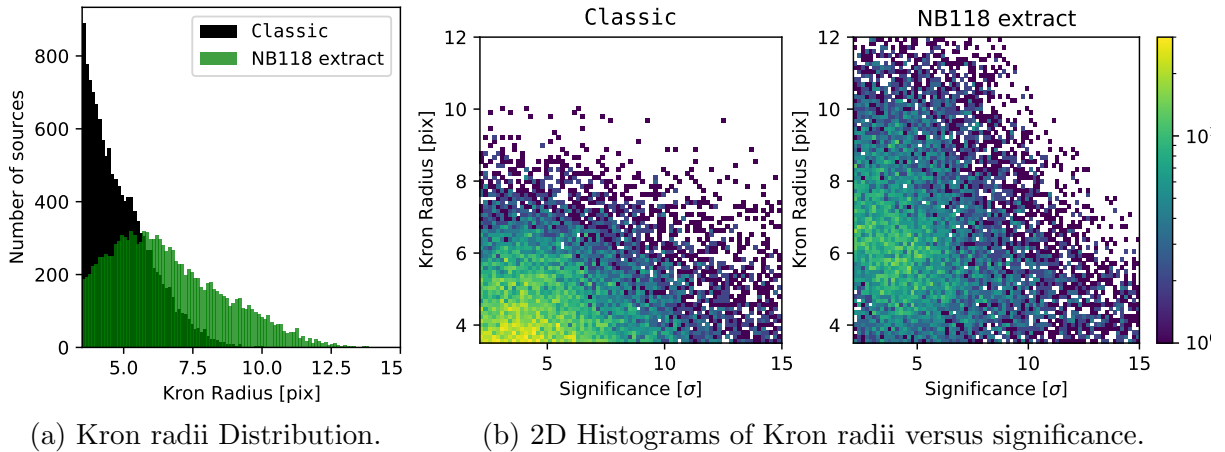


(a) Kron radii Distribution.           (b) 2D Histograms of Kron radii versus significance.

Figure 6.17: Histograms comparing the `KRON_RADIUS`, using the $NB$118 sources detected at $3\sigma$, are fainter than 20 mag, and that also have a corresponding source detected in the `Classic` catalog. This overlap makes up 49996 sources, where 13105 have a radius above 3.5 pix. (a) Histogram depicting the Kron radius for overlapping `Classic` sources and ones extracted in this work. (b) Histogram over Kron radii and significance of the $NB$118 detection. Both are normalized to the same bin range.

There are two size estimators available in the `Classic` catalogue, the `KRON_RADIUS` and the `FLUX_RADIUS`. The flux radius requires an additional extraction to be done in `SEP`,

so here the comparison is simply done for the KRON_RADIUS. As discussed in sec 5.2, the minimal radius for the automatic flux extraction performed by SExtractor is set at 3.5 pix. The same was done for the $NB$118 extraction (mimicking the FLUX_AUTO aperture), as a result, the requirement for the histograms in figure 6.17 is that KRON_RADIUS $> 3.5$ pix. Classic, just like the $NB$118 SEP source extraction, cannot find the radius for all sources, these are in both cases returned as zero, and therefore also not part of the histograms.

Figure 6.17a shows that the average source extracted by SEP in the narrowband is comparatively larger than Classic. This could imply a too sensitive detection threshold, or some major underlying difference in the detection image. The latter is inherently very true between the simple $NB$118, and the $izYJHK_s$ chi mean detection image. Figure 6.17b gives some more revealing insight into the issue. This time inspecting not just the Kron radius, but also the significance of detection for each detected source in $NB$118. The low significance sources have an approximately 50% larger Kron radius. This infers that the sources are much more stretched out at lower significance, ie. low flux. There are two likely reasons for this results, firstly the sources are already detected as closely possible to the noise, secondly the data is nowhere near as deep the $izYJHK_s$ image.

For more significant sources, the trend tends towards the same value around 4-5 pix. This is likely because the sources become bright enough the be clearly distinguished from the noise, yielding lower radius to more accurately predict the boundary of the source. Generally for many overlapping sources, especially the faint $NB$118 ones, the physical stretch will seem larger due to noise. The lack of depth for the sources where the overlap between the two detection images exists, is the main factor here. If the narrowband data was just as deep for each source with overlap, as it is in the Classic detection image, the comparison would likely have been much more uniform.
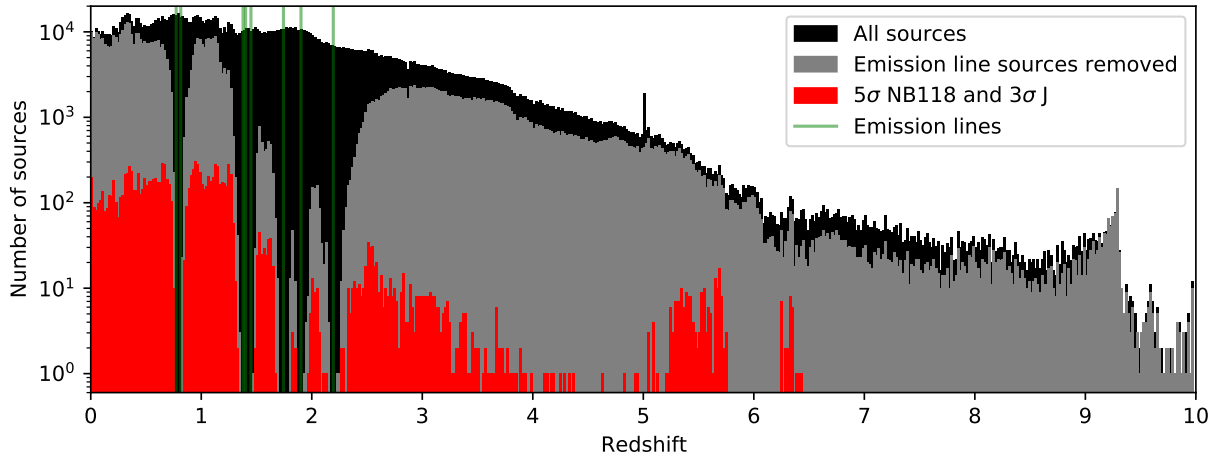
## 6.4   Possibly lensed object and spectroscopy



Figure 6.18: Histogram showing the progressive selection of sources using the criteria from equation (17). Black bars show all sources, grey the ones with photometric redshifts outside of major emission lines and the red, represent the final significance selection.

Out of the all the sources in classic with a `zPDF` determined, 605382 lie outside $2\sigma$ of the emission lines from table 6. Ly$\alpha$ was chosen not to be included, because it is potentially rare and likely weak. Within these, 11276 are detected at $3\sigma$ and $5\sigma$ in $J$ and $NB$118 respectively. Finally using equation (15) for narrowband excess, a total of 94 potentially lensed or otherwise interesting sources are found.

The search did not reveal any obvious lensed galaxies by visual inspection, but a reasonable amount of blended sources, with very apparently different colours. A few examples of these can be seen in figure 6.20. Instead the opportunity for spectroscopy turned up, and instead the search was lightened up on the redshift criteria to $1\sigma$ instead of the original 2 in equation (17). The target was $r < 21.5$ mag galaxies, due to observations constraints, ie. the size of the NOT telescope on La Palma, as well as available time for spectrum capture and a low position of the COSMOS region on the sky. The manually selected galaxy is in figure 6.19, and was unfortunately not in `The Farmer` due to HSC masks. It was specifically picked due to its shape. Featuring are strong emission line in the narrowband on the extremes of the galaxy, making it appear like a lensed object. Apart from the visual impression, the source also features heavily disagreeing photometric redshift fits, and a light drop in flux both on the red and blue side of $NB$118. There is an almost 2 mag drop in flux for ch1 and ch2, which is likely a normal low redshift IR drop-off. It should also be noted that this is HSC masked source, meaning it near a bright star or otherwise problematic.

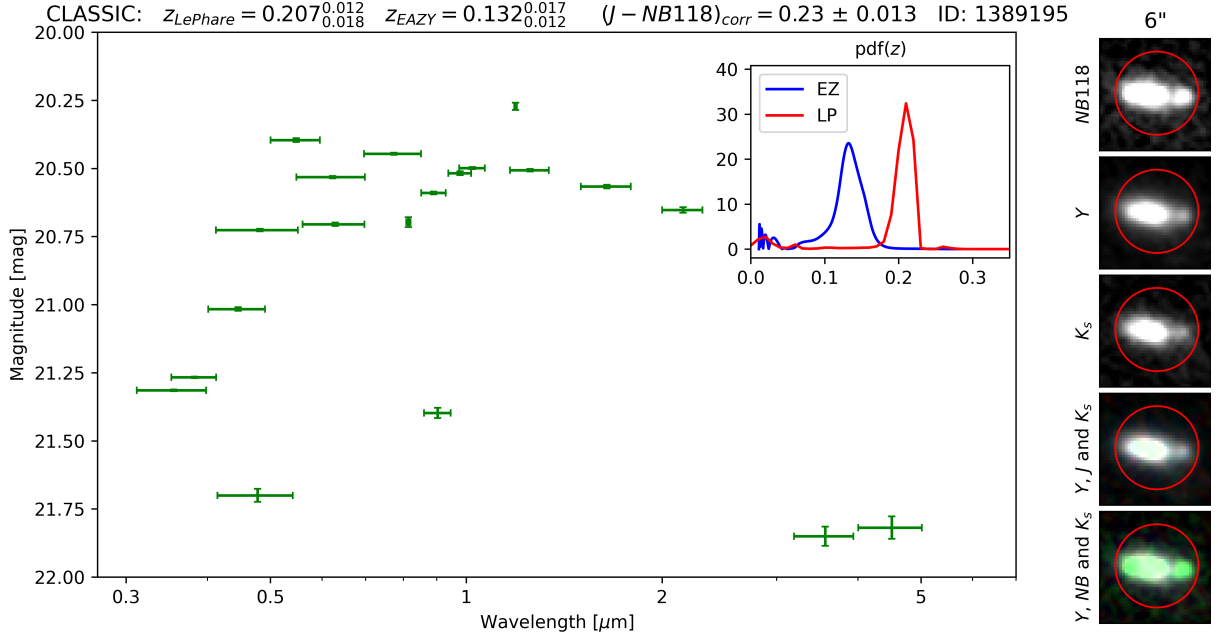| Line | $\lambda$ |
|------|-----------|
| [OII] | 3726.032 Å |
| H$\delta$ | 4101.742 Å |
| H$\gamma$ | 4340.471 Å |
| H$\beta$ | 4861.333 Å |
| [OIII] | 4958.911 Å |
| [OIII] | 5006.843 Å |
| H$\alpha$ | 6562.819 Å |
| [SII] | 6716.440 Å |

Table 6: Selected emission lines.

Figure 6.19: Bright possibly lensed galaxy. Ra,Dec (J2000)=(9h 57m 21.43s, 2° 44′ 18.31″)

The `LePhare` photometric redshift of 0.2, corresponds to [CI] line at $9850\,\text{Å}$, which is likely the peak on the pdf($z$). `EAZY` however is very certain that the source is lower than 0.75. The large disagreement between the two models is likely due to proximity to bright stars, especially when considering it is in a HSC masked area.
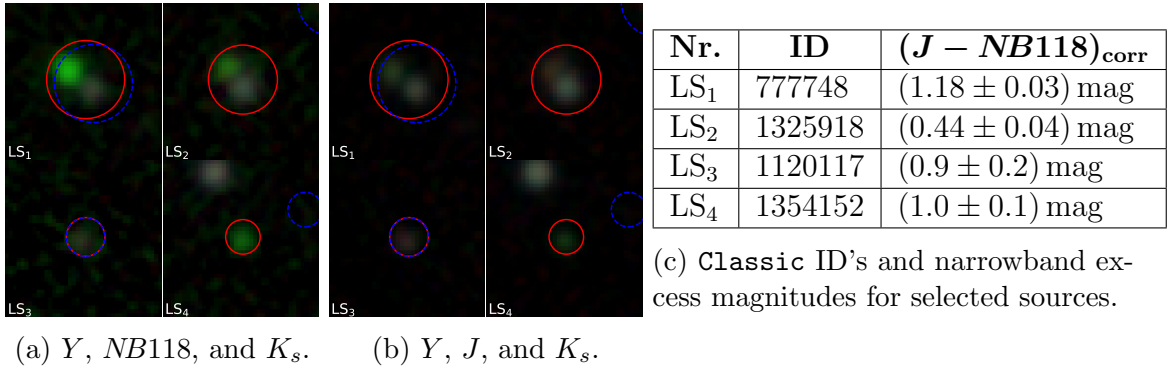


| Nr. | ID | $(J - NB118)_{\text{corr}}$ |
|---|---|---|
| $LS_1$ | 777748 | $(1.18 \pm 0.03)\,\text{mag}$ |
| $LS_2$ | 1325918 | $(0.44 \pm 0.04)\,\text{mag}$ |
| $LS_3$ | 1120117 | $(0.9 \pm 0.2)\,\text{mag}$ |
| $LS_4$ | 1354152 | $(1.0 \pm 0.1)\,\text{mag}$ |

(c) `Classic` ID's and narrowband excess magnitudes for selected sources.

(a) $Y$, $NB118$, and $K_s$.          (b) $Y$, $J$, and $K_s$.

Figure 6.20: Manually selected, potentially lensed object search in $6''\text{x}6''$ sized thumbnails. The red ring represents the `Classic` position, while the blue one represents the closest `The Farmer` detection (only $LS_3$ is detected for certain in both catalogs). $LS_1$ and $LS_2$ are obviously two sources, its difficult to judge if they are lensed, especially because both sources in both images are approximately equal size and colour when looking at 6.20b. $LS_3$ and $LS_4$ are examples of galaxies with emission lines outside the selection range. SED's, coordinates and photometric redshifts can be found in the appendix figure 9.5.
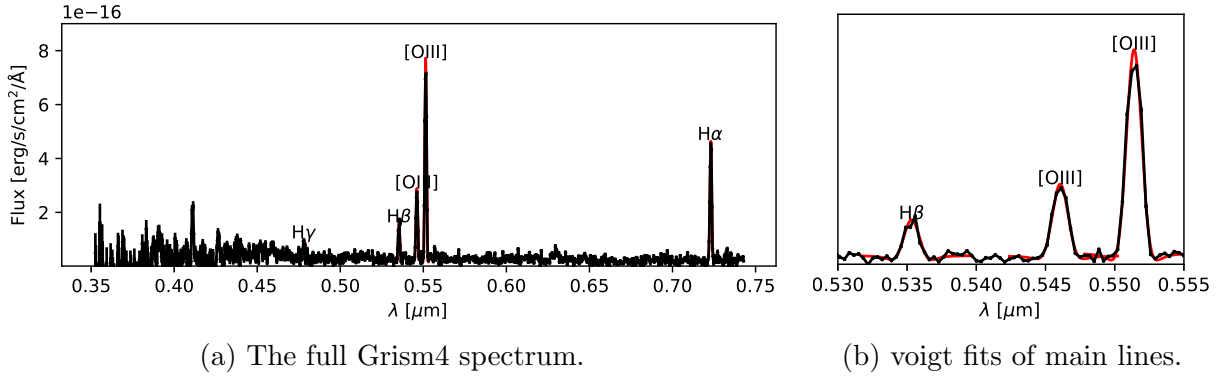
(a) The full Grism4 spectrum.

(b) voigt fits of main lines.

Figure 6.21: Grism4 spectrum taken by the NOT telescope on LaPalma May 13th 2021, using a $1'$ slit and instrument ALFOSC_FASU. The total spectrum is cut somewhat short, because of wavelength calibrations being inaccurate due to a partially broken HeNe calibration lamp. Voigt fits are used over Gaussian ones, to most accurately reconstruct any line broadening and line shapes. (a) Represents the full Grism4 spectrum, with all the fitted profiles. (b) Is a enhancement of (a), on the two [OIII] lines and H$\beta$ line.

| Line | $\lambda_{\text{vac}}$ | $\lambda_{\text{obs}}$ | Redshift |
|------|------------------------|------------------------|----------|
| H$\gamma$ | $4341.684\,\text{Å}$ | $(4781 \pm 1)\,\text{Å}$ | $0.1011 \pm 0.0002$ |
| H$\beta$ | $4862.683\,\text{Å}$ | $(5353.2 \pm 0.4)\,\text{Å}$ | $0.100\,88 \pm 0.000\,09$ |
| [OIII] | $4960.295\,\text{Å}$ | $(5461.6 \pm 0.2)\,\text{Å}$ | $0.100\,86 \pm 0.000\,09$ |
| [OIII] | $5008.24\,\text{Å}$ | $(5514.0 \pm 0.2)\,\text{Å}$ | $0.100\,99 \pm 0.000\,03$ |
| H$\alpha$ | $6564.614\,\text{Å}$ | $(7231.7 \pm 0.2)\,\text{Å}$ | $0.101\,46 \pm 0.000\,03$ |

Table 7: Selected emission lines, their vacuum wavelengths, redshifted voigt profile fit values and calculated redshift.

Initially inspecting the spectrum[1], shows that the source is not a lensed object. There are no obvious double emission lines, which should be visible within the spectrum. The limit on the spectrum is about $1\,\mu\text{m}$, meaning [OIII] lines shoul be visible all the way to $z \sim 1$. The outer clumps visible in the thumbnails in figure 6.19 are clearly visible in each filter, which indicates a sufficient brightness to know that emission lines would show up in the spectrum. This means that the source is just a split up, but the same object receding at possibly same redshift. The extra strong emission seen in the $NB118$ filter vs the $J$ filter, will likely be because the outer parts of the source are heated more effectively than the centre, indicating stronger star formation or other methods of gas heating. Taking a weighted mean from the fitted emission lines in table 7, a redshift of $0.101\,21 \pm 0.000\,02$ is found. The narrowband excess seen in the `Classic` SED should then be in the range $(10\,814 \pm 51)\,\text{Å}$, which coincides with a He I line at $10\,830\,\text{Å}$.

Gathering what was observed, this is likely a irregular galaxy, going through partial star formation. The few clear emission lines in the spectrum, indicate a probably metal poor source. Somewhat similar to dwarf galaxy I Zwicky 18 (Izotov and Thuan, 2004). The low metallicity indicates a late starting star formation, especially at a redshift this low. This will make the He I emission line in the narrowband filter stronger, because the lack of other elements also means a lack of cooling mechanism.

---

[1]The data presented here were obtained [in part] with ALFOSC, which is provided by the Instituto de Astrofisica de Andalucia (IAA) under a joint agreement with the University of Copenhagen and NOT.

# 7   Conclusion

## 7.1   $Y$-band dropouts and Ly$\alpha$ emitters

Using the `Classic` and `The Farmer` catalogs from Weaver et al. (subm), constrains on the flux at shorter wavelengths than the VISTA $J$ filter were created to identify Lyman break galaxies. Then using significance constraints of $3\sigma$ in $J$ and $5\sigma$ in $NB$118, also including a requirement for an emission line in the narrowband filter using equations (14) and (15).

The search for high redshift Ly$\alpha$ emitters at $z \sim 8.75$, with significant narrowband excess, did not directly yield a source. Even with loosened original constrains on the flux of filters blueward of the Lyman break, no good candidates were found. The broader search for general $Y$-dropouts in the range 7.8 to 8.6, did however yield 5 good LBGs. The most promising galaxy among these, was $J_3$, which also happens to be detected in the narrowband $NB$118 strips. This source currently shows narrowband excess of $(0.69 \pm 0.59)\,$mag, which carries great potential, with additional observation, to be exactly what was set out to discover in this thesis.

## 7.2   Using $NB$118 for source extraction

Using `SEP` , the python port of `SExtractor`, the narrowband strips of $NB$118 were used for source extraction in order to find new sources not included in the `Classic` catalog. Using a detection threshold of 2.5 and a minimum area of 5 pixels, a total of 147962 sources were extracted. Very bright sources have overlap with the `Classic` catalog, and therefore a 20mag upper limit and $3\sigma$ detection limit were set to lower computation. These remaining 50294 were then cross referenced for overlap and removed, which filtered off 99.4% after also removing sources in HSC masked areas, leaving 114 potential new detections.

The sources with both high significance but also visual impression were then picked out for discussion in figure 6.16. Where especially $NB_1$ and $NB_2$ are very convincing candidates, and almost certainly real sources with very bright emission lines. Most of the extracted objects are likely noise, due to a fast and simple extraction method.

## 7.3   Possibly lensed objects

Using the photometric redshift from `Classic LePhare`, galaxies with strong emission lines in the narrowband regime are removed within $2\sigma$ in $z$. These selected lines are show in table 6. With the most common line emitters gone, the remaining sources will either be line emitters outside the selected lines, or have incorrectly determined photometric redshifts. Again using equations (14) and (15), line emitters are picked out, and visually inspected. This resulted in 94 selected sources, varying between non-deblended sources, rarer emission lines and critically failed photometric redshift determinations in `LePhare`.

A bright $r < 21.5\,$mag possibly lensed galaxy was found (figure 6.19), with a central bright source, and two clouds on both sides, with clear narrowband excess. The photometric redshift for both sources was very out of agreement, so a observation proposal was sent to the NOT telescope. The proposal was accepted, and the detection turned out to be a He I emitter at $z = 0.1011 \pm 0.0005$, with no sign of other redshifts in the lines. It is likely a irregular galaxy, with low metallicity because of the low amount of lines in the spectrum.

## Acknowledgements

# 8   Bibliography

Aihara, H., AlSayyad, Y., Ando, M., Armstrong, R., Bosch, J., Egami, E., Furusawa, H., Furusawa, J., Goulding, A., Harikane, Y., and et al. (2019). Second data release of the hyper suprime-cam subaru strategic program. *Publications of the Astronomical Society of Japan*, 71(6).

Arnouts, S., Moscardini, L., Vanzella, E., Colombi, S., Cristiani, S., Fontana, A., Giallongo, E., Matarrese, S., and Saracco, P. (2002). Measuring the redshift evolution of clustering: the hubble deep field south. *Monthly Notices of the Royal Astronomical Society*, 329(2):355–366.

Barbary, K. (2016). Sep: Source extractor as a library. *Journal of Open Source Software*, 1(6):58.

Bertin, E. and Arnouts, S. (1996). SExtractor: Software for source extraction. *Astronomy and Astrophysics*, 117:393–404.

Bielby, R. M., Finoguenov, A., Tanaka, M., McCracken, H. J., Daddi, E., Hudelot, P., Ilbert, O., Kneib, J. P., Le Fèvre, O., Mellier, Y., and et al. (2010). The wircam deep infrared cluster survey. *Astronomy and Astrophysics*, 523:A66.

Bowler, R. A. A., Jarvis, M. J., Dunlop, J. S., McLure, R. J., McLeod, D. J., Adams, N. J., Milvang-Jensen, B., and McCracken, H. J. (2020). A lack of evolution in the very bright end of the galaxy luminosity function from z $\simeq$ 8 to 10. *Monthly Notices of the Royal Astronomical Society*, 493(2):2059–2084.

Brammer, G. B., van Dokkum, P. G., and Coppi, P. (2008). EAZY: A Fast, Public Photometric Redshift Code. *The Astrophysical Journal*, 686(2):1503–1513.

Chabrier, G. (2003). Galactic stellar and substellar initial mass function. *Publications of the Astronomical Society of the Pacific*, 115(809):763–795.

Conroy, C. and Gunn, J. E. (2010). The propagation of uncertainties in stellar population synthesis modeling. iii. model calibration, comparison, and evaluation. *The Astrophysical Journal*, 712(2):833–857.

Conroy, C., Gunn, J. E., and White, M. (2009). The propagation of uncertainties in stellar population synthesis modeling. i. the relevance of uncertain aspects of stellar evolution and the initial mass function to the derived physical properties of galaxies. *The Astrophysical Journal*, 699(1):486–506.

Coupon, J., Czakon, N., Bosch, J., Komiyama, Y., Medezinski, E., Miyazaki, S., and Oguri, M. (2017). The bright-star masks for the hsc-ssp survey. *Publications of the Astronomical Society of Japan*, 70(SP1).

Doty, S. D. and Neufeld, D. A. (1997). Models for dense molecular cloud cores. *The Astrophysical Journal*, 489(1):122–142.

Faure, C., Kneib, J., Covone, G., Tasca, L., Leauthaud, A., Capak, P., Jahnke, K., Smolcic, V., de la Torre, S., Ellis, R., and et al. (2008). First catalog of strong lens candidates in the cosmos field. *The Astrophysical Journal Supplement Series*, 176(1):19–38.

Fioc, M and Rocca-Volmerange, B (1997). PEGASE: a UV to NIR spectral evolution model of galaxies: Application to the calibration of bright galaxy counts. (astro-ph/9707017).

Fitzpatrick, E. L. and Massa, D. (1986). An Analysis of the Shapes of Ultraviolet Extinction Curves. I. The 2175 Angstrom Bump. *The Astrophysical Journal*, 307:286.

Fitzpatrick, E. L. and Massa, D. (1990). An Analysis of the Shapes of Ultraviolet Extinction Curves. III. an Atlas of Ultraviolet Extinction Curves. *Astrophysical Journal Supplement*, 72:163.

Franx, M. (2019). Ultra vista telescope website. `http://ultravista.org/`. Online; accessed 16 May 2021.

Hsieh, B.-C., Wang, W.-H., Hsieh, C.-C., Lin, L., Yan, H., Lim, J., and Ho, P. T. P. (2012). The taiwan ecdfs near-infrared survey: Ultra-deep $j$ and $k_s$ imaging in the extended chandra deep field-south. *The Astrophysical Journal Supplement Series*, 203(2):23.

Ilbert, O., Arnouts, S., McCracken, H. J., Bolzonella, M., Bertin, E., Le Fèvre, O., Mellier, Y., Zamorani, G., Pellò, R., Iovino, A., and et al. (2006). Accurate photometric redshifts for the cfht legacy survey calibrated using the vimos vlt deep survey. *Astronomy and Astrophysics*, 457(3):841–856.

Izotov, Y. I. and Thuan, T. X. (2004). Deephubble space telescopeacs observations of i zw 18: a young galaxy in formation. *The Astrophysical Journal*, 616(2):768–782.

Jones, T., Stark, D. P., and Ellis, R. S. (2012). Keck spectroscopy of faint $3 < z < 7$ lyman break galaxies. iii. the mean ultraviolet spectrum at $z \simeq 4$. *The Astrophysical Journal*, 751(1):51.

J.P.Emerson (2019). Vista telescope technical specifications page. `https://vista.maths.qmul.ac.uk/`. Online; accessed 19 May 2021.

Kron, R. G. (1980). Photometry of a complete sample of faint galaxies. *American Journal of Political Science*, 43:305–325.

Laigle, C., McCracken, H. J., Ilbert, O., Hsieh, B. C., Davidzon, I., Capak, P., Hasinger, G., Silverman, J. D., Pichon, C., Coupon, J., and et al. (2016). The cosmos2015 catalog: Exploring the $1 < z < 6$ universe with half a million galaxies. *The Astrophysical Journal Supplement Series*, 224(2):24.

Laursen, P., Sommer-Larsen, J., Milvang-Jensen, B., Fynbo, J. P. U., and Razoumov, A. O. (2019). Lyman $\alpha$-emitting galaxies in the epoch of reionization. *Astronomy and Astrophysics*, 627:A84.

Magnier, E. A. and Cuillandre, J. C. (2004). The Elixir System: Data Characterization and Calibration at the Canada-France-Hawaii Telescope. *TThe Publications of the Astronomical Society of the Pacific*, 116(819):449–464.

McCracken, H. J., Capak, P., Salvato, M., Aussel, H., Thompson, D., Daddi, E., Sanders, D. B., Kneib, J. P., Willott, C. J., Mancini, C., Renzini, A., Cook, R., Le Fèvre, O., Ilbert, O., Kartaltepe, J., Koekemoer, A. M., Mellier, Y., Murayama, T., Scoville, N. Z., Shioya, Y., and Tanaguchi, Y. (2010). The COSMOS-WIRCam Near-Infrared Imaging Survey. I. BzK-Selected Passive and Star-Forming Galaxy Candidates at z gsim 1.4. *The Astrophysical Journal*, 708(1):202–217.

McCracken, H. J., Milvang-Jensen, B., Dunlop, J., Franx, M., Fynbo, J. P. U., Le Fèvre, O., Holt, J., Caputi, K. I., Goranova, Y., Buitrago, F., and et al. (2012). Ultravista: a new ultra-deep near-infrared survey in cosmos. *Astronomy and Astrophysics*, 544:A156.

McLeod, D. J., McLure, R. J., and Dunlop, J. S. (2016). The $z = 9-10$ galaxy population in the hubble frontier fields and clash surveys: the $z = 9$ luminosity function and further evidence for a smooth decline in ultraviolet luminosity density at $z \geq 8$. *Monthly Notices of the Royal Astronomical Society*, 459(4):3812–3824.

Milvang-Jensen, B., Freudling, W., Zabl, J., Fynbo, J. P. U., Møller, P., Nilsson, K. K., McCracken, H. J., Hjorth, J., Le Fèvre, O., Tasca, L., and et al. (2013). On-sky characterisation of the vista nb118 narrow-band filters at $1.19\mu$m. *Astronomy and Astrophysics*, 560:A94.

Miyazaki, S., Komiyama, Y., Kawanomoto, S., Doi, Y., Furusawa, H., Hamana, T., Hayashi, Y., Ikeda, H., Kamata, Y., Karoji, H., Koike, M., Kurakami, T., Miyama, S., Morokuma, T., Nakata, F., Namikawa, K., Nakaya, H., Nariai, K., Obuchi, Y., Oishi, Y., Okada, N., Okura, Y., Tait, P., Takata, T., Tanaka, Y., Tanaka, M., Terai, T., Tomono, D., Uraguchi, F., Usuda, T., Utsumi, Y., Yamada, Y., Yamanoi, H., Aihara, H., Fujimori, H., Mineo, S., Miyatake, H., Oguri, M., Uchida, T., Tanaka, M. M., Yasuda, N., Takada, M., Murayama, H., Nishizawa, A. J., Sugiyama, N., Chiba, M., Futamase, T., Wang, S.-Y., Chen, H.-Y., Ho, P. T. P., Liaw, E. J. Y., Chiu, C.-F., Ho, C.-L., Lai, T.-C., Lee, Y.-C., Jeng, D.-Z., Iwamura, S., Armstrong, R., Bickerton, S., Bosch, J., Gunn, J. E., Lupton, R. H., Loomis, C., Price, P., Smith, S., Strauss, M. A., Turner, E. L., Suzuki, H., Miyazaki, Y., Muramatsu, M., Yamamoto, K., Endo, M., Ezaki, Y., Ito, N., Kawaguchi, N., Sofuku, S., Taniike, T., Akutsu, K., Dojo, N., Kasumi, K., Matsuda, T., Imoto, K., Miwa, Y., Suzuki, M., Takeshi, K., and Yokota, H. (2017). Hyper Suprime-Cam: System design and verification of image quality. *Publications of the Astronomical Society of Japan*, 70(SP1). S1.

Miyazaki, S., Komiyama, Y., Sekiguchi, M., Okamura, S., Doi, M., Furusawa, H., Hamabe, M., Imi, K., Kimura, M., Nakata, F., and et al. (2002). Subaru prime focus camera — suprime-cam. *Publications of the Astronomical Society of Japan*, 54(6):833–853.

Muzzin, A., Marchesini, D., Stefanon, M., Franx, M., Milvang-Jensen, B., Dunlop, J. S., Fynbo, J. P. U., Brammer, G., Labbé, I., and van Dokkum, P. (2013). A public ks-selected catalog in the cosmos/ultravista field: Photometry, photometric redshifts, and stellar population parameters. *The Astrophysical Journal Supplement Series*, 206(1):8.

Nilsson, K. K., Orsi, A., Lacey, C. G., Baugh, C. M., and Thommes, E. (2007). Narrow-band surveys for very high redshift lyman-$\alpha$ emitters. *Astronomy and Astrophysics*, 474(2):385–392.

Oesch, P. A., Bouwens, R. J., Illingworth, G. D., Labbé, I., and Stefanon, M. (2018). The dearth of $z \sim 10$ galaxies in all hst legacy fields—the rapid evolution of the galaxy population in the first 500 myr. *The Astrophysical Journal*, 855(2):105.

Pitman, K. M., Clayton, G. C., and Gordon, K. D. (2000). The 2175 Å Extinction Bump in QSO Spectra. *TThe Publications of the Astronomical Society of the Pacific*, 112(770):537–541.

Quadri, R., Marchesini, D., van Dokkum, P., Gawiser, E., Franx, M., Lira, P., Rudnick, G., Urry, C. M., Maza, J., Kriek, M., and et al. (2007). The multiwavelength survey by yale-chile (musyc): Deep near-infrared imaging and the selection of distant galaxies. *The Astronomical Journal*, 134(3):1103–1117.

Salpeter, E. E. (1955). The Luminosity Function and Stellar Evolution. *Astrophysical Journal*, 121:161.

Sanders, D. B., Salvato, M., Aussel, H., Ilbert, O., Scoville, N., Surace, J. A., Frayer, D. T., Sheth, K., Helou, G., Brooke, T., and et al. (2007). S-cosmos: The spitzer legacy survey of the hubble space telescope acs 2 deg 2 cosmos field i: Survey strategy and first analysis. *The Astrophysical Journal Supplement Series*, 172(1):86–98.

Saumon, D. and Marley, M. S. (2008). The Evolution of L and T Dwarfs in Color-Magnitude Diagrams. *The Astrophysical Journal*, 689(2):1327–1344.

Sawicki, M., Arnouts, S., Huang, J., Coupon, J., Golob, A., Gwyn, S., Foucaud, S., Moutard, T., Iwata, I., Liu, C., and et al. (2019). The cfht large area u-band deep survey (clauds). *Monthly Notices of the Royal Astronomical Society*.

Schlafly, E. F. and Finkbeiner, D. P. (2011). Measuring reddening with sloan digital sky survey stellar spectra and recalibrating sfd. *The Astrophysical Journal*, 737(2):103.

Shultz, N. S. (2012). *The formation and early evolution of stars*. Number 2. Springer, Berlin, Heidelberg.

Songaila, A. (2004). The evolution of the intergalactic medium transmission to redshift 6. *The Astronomical Journal*, 127(5):2598–2603.

Stefanon, M., Labbé, I., Bouwens, R. J., Oesch, P., Ashby, M. L. N., Caputi, K. I., Franx, M., Fynbo, J. P. U., Illingworth, G. D., Fèvre, O. L., and et al. (2019). The brightest $z \geq 8$ galaxies over the cosmos ultravista field. *The Astrophysical Journal*, 883(1):99.

Szalay, A. S., Connolly, A. J., and Szokoly, G. P. (1999). Simultaneous multicolor detection of faint galaxies in the hubble deep field. *The Astronomical Journal*, 117(1):68–74.

Thomas, R. M., Zaroubi, S., Ciardi, B., Pawlik, A. H., Labropoulos, P., Jelić, V., Bernardi, G., Brentjens, M. A., de Bruyn, A. G., Harker, G. J. A., and et al. (2009). Fast large-scale reionization simulations. *Monthly Notices of the Royal Astronomical Society*, 393(1):32–48.

Weaver, J. R., Kauffmann, O., Shuntov, M., Davidzon, I., Ilbert, O., Brammer, G., Hsieh, B., Capak, P., Moneti, A., McCracken, H. J., Toft, S., and The Cosmos Team (2021. subm). COSMOS2020: A next-generation catalog to explore the $1 < z < 8$ universe. volume 53, page 215.06. Submitted.

Zamojski, M. A., Schiminovich, D., Rich, R. M., Mobasher, B., Koekemoer, A. M., Capak, P., Taniguchi, Y., Sasaki, S. S., McCracken, H. J., Mellier, Y., and et al. (2007). Deep galex imaging of the cosmos hst field: A first look at the morphology of $z > 0.7$ star-forming galaxies. *The Astrophysical Journal Supplement Series*, 172(1):468–493.

Zaroubi, S. (2012). The epoch of reionization. *Astrophysics and Space Science Library*, page 45–101.

Here included is most of the major code used for this thesis on my GitHub page created just for this purpose. github.com/AndreHartwigsen/Masters-Thesis.

# 9   Appendix

## 9.1   SWarp

`SWarp` combines multiple FITS images together to one, and offers a lot of options for resampling, background subtraction and projection. The software is, on a basic level `SExtractor` without the source extraction. `SWarp` takes FITS files and their corresponding weight, uncertainty or variance-maps. The output is a single FITS file and weight-map.

Firstly, if necessary, the images will be resampled for sources and pixels to align. This process will both interpolate and filter the inputs. The program offers many interpolation options: `NEAREST`, `BILINEAR`, `LANCZOS2`, `LANCZOS3` and `LANCZOS4`. All these have an increasingly aggressive interpolation from surrounding pixels. The last option here provides the best resampling, but comes at a cost of computing power. An averaging of even more pixels could be applied, but this can lead to artefacts such as ripples (*Gibbs' phenomenon*).

Each image has their background subtracted before combination, this process is the same as for `SExtractor` described in more detail in section **9.2.1**. The flux will be scaled using properly flatfielded data, and conserves the original flux ratios for each pixel. This is then scaled according to the specified photometric and astrometric parameters. The resulting scale will be either manually set using the `FSCALE_DEFAULT`, or is extracted from the FITS files.

The last, and most important step, is combining the individual images and their respective weight-maps together. Due to all the possible variances between them, several combining methods are available. Here are the simple, `MIN`, `MAX`, `SUM` and `MEDIAN`, followed by the more relevant ones, `MEAN`, `CHI2` and `WHEIGHTED`.

$$F_{\texttt{MEAN}} = \frac{1}{n_{w_i \neq 0}} \sum_i p_i f_i \qquad F_{\texttt{CHI2}} = \sqrt{\frac{1}{n_{w_i \neq 0}} \sum_i w_i f_i^2} \qquad F_{\texttt{WHEIGHTED}} = \frac{\sum_i w_i p_i f_i}{\sum_i w_i} \qquad (18)$$

Where for each pixel $i$, $p_i$ is the flux scaling factor, $w_i$ is the weight and $f_i$ is the flux. The `MEAN` provides the best S/N, and is the average over all non-zero weight pixels.

Weaver et al. (subm) uses a `CHI_MEAN` setting not described in this section, to create the $izYJHK_s$-detection image from the non PSF homogenized images and weight maps.

## 9.2   SExtractor

In order to extract individual sources from the images, `SExtractor` (Bertin and Arnouts, 1996) is used. The basics are a two step process, firstly creating a background map, and

comparing it to the data itself. The difference between the two sets will then be used for source extraction.

### 9.2.1   Creating a background

The first process is *segmentation*, which is separating the image into segments. This includes areas of for example varying colour, brightness or borders. For astronomical purposes, this process boils down to finding a group of connected pixels that exceed some threshold above the background. This is not trivial, as the background can vary widely, depending on parameters such as stars or artefacts. In addition to this, data is often noisy for very faint objects, so noise reduction or smoothing has to be applied. Blending of sources is also a major issue, a simple detection threshold will identify two sources as one if connected. In `SExtractor` de-blending is archived through multi-thresholding.

The actual creation of the background map is done by going over the entire frame and performing $\sigma\kappa$-clipping. Typically this means removing pixels that have larger value than the product $\sigma\kappa$, but `SExtractor` is somewhat more sophisticated here. The local background histogram is clipped iteratively until convergence at $\pm 3\sigma$ around its median, where the histogram mode is estimated using:

$$\text{Mode} = 3 \times \text{Median} - 2 \times \text{Mean} \tag{19}$$

$$\text{Mode} = 2.5 \times \text{Median} - 1.5 \times \text{Mean} \tag{20}$$

Equation (19) shows to be the most accurate with clipped distributions based on simulated images of random Gaussian meshes. If the mode and the median disagree by more than 30%, `SExtractor` reverts to using the more standard equation (20).

After the creation of this map, a median filter can be applied if needed, to smooth out any local overestimations due to bright features like stars. The final products are the background map made by bicubic-spline interpolation between the meshes of the grid. Additionally to this is a RMS background map, which is the noise standard deviation for the background images.

### 9.2.2   Parameters

There are some important parameters to chosen when creating a background map. For any given data set, a fitting `BACK_SIZE` has to be chosen. If the grid is too small, the map will be influenced by random detector noise, and if it is too coarse, it will not be able to represent the variations in the detector properly.

The user can also chose `BACK_FILTERSIZE`, to change over how many pixels a median filter will be applied. Typically this background filter will be $3 \times 3$, which is enough for most cases, but it can be increased to correct for a small mesh size for example. The option `BACK_FILTERTHRESH` lets the user chose a threshold above which the median filter is applied.

Alternatively a manually defined constant can also be subtracted from the entire image, by setting the `BACK_TYPE` to *manual* and specifying the value to subtract with the `BACK_VALUE` parameter. This might be done for very non-Gaussian noise images.

### 9.2.3   Weights

Noise level in astronomical images is often low, but in some cases like strong vignetting or composite images, more accurate weights are required. Too much noise can cause both

missed detections, and failed ones. SExtractor can handle variable noise images, which have to be the same size as the science image. They are internally stores as absolute variance, and if the variance is too high ($> 10^{30}$), the pixel will be considered completely unreliable and is ignored. Variance is used because it linearises many operations, greatly improving simplicity and efficiency.

This way of processing the noise from the detector, also means that the covariance between pixels is ignored. Normally all non-zero covariances within geometric limits should be taken into account. For astronomical CCD's however, this is mostly not an issue, because the correlation between the noise will be much smaller than the detected patterns. The weight map (WEIGHT_IMAGE) itself can be given to SExtractor in the form of RMS(standard deviation), variance or the weight, since they are all trivially associated with each other (21).

$$\text{var}_i = \sigma_i^2 = \frac{1}{w_i^2} \tag{21}$$

Apart from the obvious transformations between the three equations, SExtractor still has some procedures. If the WEIGHT_IMAGE is given as standard deviation, it will simply use the map as is. If given as variance, SExtractor will perform a robust scaling to an appropriate absolute level, compared to the low resolution variance map produced from the science image discussed in section **9.2.1**. The method is the same if the supplied WEIGHT_IMAGE is a weight map, it is first converted to variance using equation (21). Additionally SExtractor also supports the input of multiple weight maps, if RMS and weight are supplied, then the RMS will be used to adjust the detection threshold, and the weights to estimate the scaling on the error estimates.

**Effect of weight maps**

- If more than 50% of a given mesh are bad, the final mesh will be interpolated from the surrounding meshes.

- The detection threshold $t$ above the local sky background is adjusted for each pixel $i$ with variance $\sigma_i^2$. This might result in some pixels not being detected, if the WEIGHT_THRESH is set too high. This can be avoided with interpolation or filtering the data.

- The cleaning process takes the threshold for each pixel into account for fain detections.

- For most cases the image intensity is due to noise gain, then it is assumed that the local pixel gain is inversely proportional to $\sigma_i^2$. When the variance is dominated by read-out noise, the weight WEIGHT_GAIN-parameter should be set to $N$ (Used in Weaver et al. (subm)), instead of the default $Y$.

- Pixels with weights beyond the WEIGHT_THRESH are discarded by the masking process.

### 9.2.4   Source detection

When a source extraction is performed, SExtractor will give flags if there are any errors like saturated pixels, biases, de blending issues, memory overflows and so on. This is simply given as a 2 bit output, with integers corresponding to warnings.

| Value | Meaning |
|---|---|
| 1 | Aperture photometry is likely to be biased by neighbouring sources or by more than 10% of bad pixels in any aperture. |
| 2 | The object has been deblended |
| 4 | At least one object pixel is saturated |
| 8 | The isophotal footprint of the detected object is truncated (too close to an image boundary) |
| 16 | At least one photometric aperture is incomplete or corrupted (hitting buffer or memory limits) |
| 32 | The isophotal footprint is incomplete or corrupted (hitting buffer or memory limits) |
| 64 | A memory overflow occurred during deblending |
| 128 | A memory overflow occurred during extraction |

Table 8: The 2 bit flags generated by `SExtractor`.

When the sources are detected, and the background is removed, `SExtractor` can perform the measurements on each of them. This is done in three ways:

- **Isophotal**: Sources are detected only if they are above the detection threshold defined by `DETECT_THRESH`, any connected pixels above will be considered a source. For extended sources this can cause a lack of detection for the dimmer parts.

- **Full**: `SExtractor` now uses all pixels of the science image to detect sources. This is done after some cleaning and masking of the image, this relatively fast and is less sensitive to the detection threshold.

- **Model-fitting**: The last stage fits models to fainter sources, and takes by far the longest.

### 9.2.5   PSFEx

`PSFEx` is largely a program on its own, but it relies heavily on `SExtractor` to work properly. `PSFEx` uses photometry from source extracted images, in order to fit PSF models. This includes everything from vignetting due to over-saturated points sources, to fitting galaxy light curves directly.

`PSFEx` takes a created catalogue of sources in a photometric image, as well as image biases. The first step is then to find the likely points sources, this is done by looking for over-saturated sources and half light radii, which are then excluded from the fitting process. This will not be the case for point sources with visual vignetting, as they produce significant extended light artefacts. Then the entire source selected by `SExtractor` is loaded to memory, and a model is fit. The fitting process is iterated 4 times, to achieve a proper fit. There are options for some additional precision here, with more fitting iterations, and more advanced modelling of hidden dependencies in the PSF's. With sufficiently small residuals, the final PSF will be saved to the user's disk.

An important note here, is that `PSFEx` removes the artefacts created by atmospheric distortions and lenses, and not to fit specific galaxy models. Profile fitting is handled separately by `SExtractor`, after the PSF homogenisation is performed.
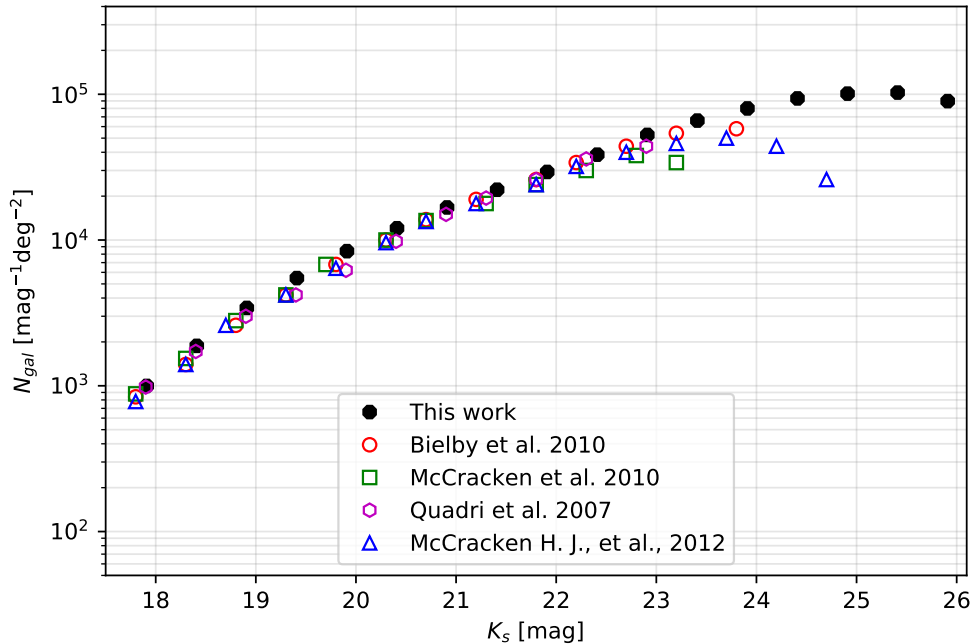
## 9.3   Recreating a histogram from COSMOS2020



Figure 9.1: Comparison of galaxies with previous studies.

Firstly to validate that the `Classic` data aligns with previous studies (Bielby et al., 2010; McCracken et al., 2010; Quadri et al., 2007; McCracken et al., 2012), a selection of bright $K_s$ selected galaxies was picked out. This was initially done to get a better understanding of the data, but also to better understand the SED structure of stars and how they move across colour-colour diagrams. Out of the 1.7M sources in `Classic`, 478,854 galaxies are brighter than 26 mag. About half of the sources in `Classic` overlap with the VISTA data, so this is no surprise. In figure 9.1 a 0.5 mag bin spacing histogram is presented of these galaxies. The bin height is normalized with a factor 1.8 square degrees[2]. Stars are removed beforehand through colour-colour selection. This step could also have been done using the flags from `Classic`, which uses half light radius and pixel saturation with HSC $i$ magnitude to identify them.

## 9.4   Colour-colour diagrams

The selection of stars in figure 9.1 was done through a $YJHK_s$ colour-colour diagram. In addition an $K_s$ magnitude cut of 24.5 is also applied, because stars are relatively bright. Sources brighter than this threshold are very clearly separated, which is due to the very specific spectral shape of stars. Most common sources will be somewhat flat, with some altercations relating to SFR and amount of dust and gas. This will place them around (0,0) in a colour-colour diagram, which will not be the case for stars. In figure 9.2 the IMF from Chabrier (2003) was used, assuming main sequence stars, the temperature was extrapolated and Planck curves were used. Normally it would be necessary to correctly evaluate the flux for such spectra, but scaling constants do not matter when comparing the ratio, and converting to magnitude. The randomly generated $10^7$ stars vary from $1500\,\mathrm{K}$ to $50\,000\,\mathrm{K}$, with the vast majority of them being in the lower range (see figure

---

[2]The ULTRA VISTA total detector area is calculated from $1.5° \times 1.2°$.
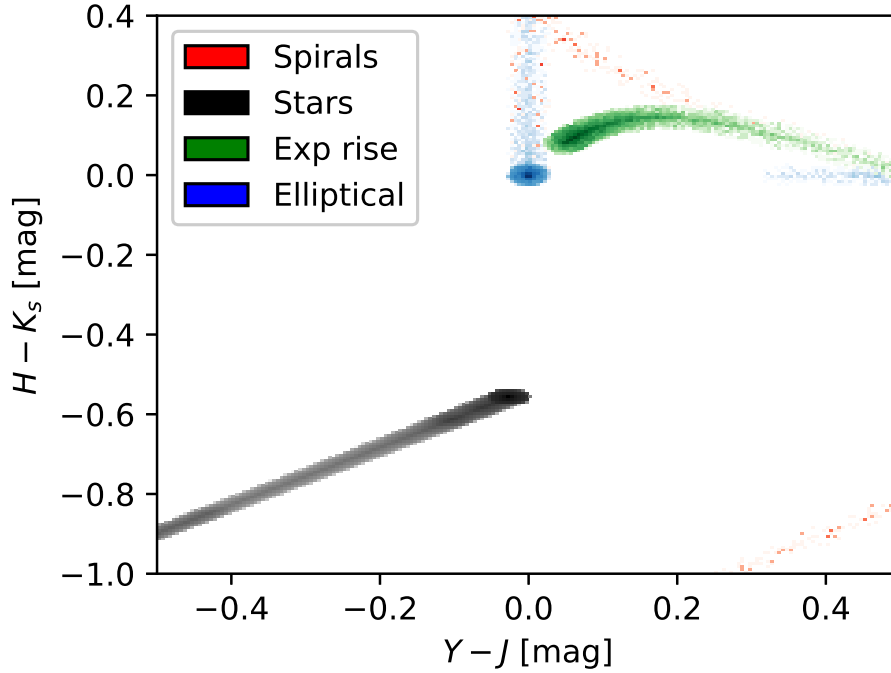
Figure 9.2: Colour-colour 2D histogram for some typical galaxy spectral shapes for $z \in [1, 9]$, together with Planck curves of varying temperatures based on IMF from eq.(3) to simulate stars.

2.2). Since most galaxies will be with approximately flat shape. The flat line shape was not included, since they would just subside around the origin. For spirals a steep slope at around 4000 Å was used, with a slowly declining flux from there on. The same was done for ellipticals, but after the peak it remains linear to simulate the sum of Planck curves that this type of galaxy is defined by. Lastly an exponentially approaching shape $\left(1 - e^{-kx}\right)$ was also included, again rising around that 4000 Å mark and then redshifted. For all shapes, random flux fluctuations are added to simulate the variation for each spectrum. If the noise is not added, all histograms would be thin lines. There are $10^4$ of each of the non-star spectre, due to the more intense calculations involved. We can take away from figure 9.2 that stars will always have a negative slope, because their peak is so far towards the blue compared to the $YJHK_s$ filters, which will place them in a unique position in this diagram. There might be some irregular galaxies that do get caught, and flagged as stars by this method however, because some irregular shapes can occur, and if they come close the Planck shape, and are bright enough in $K_s$.
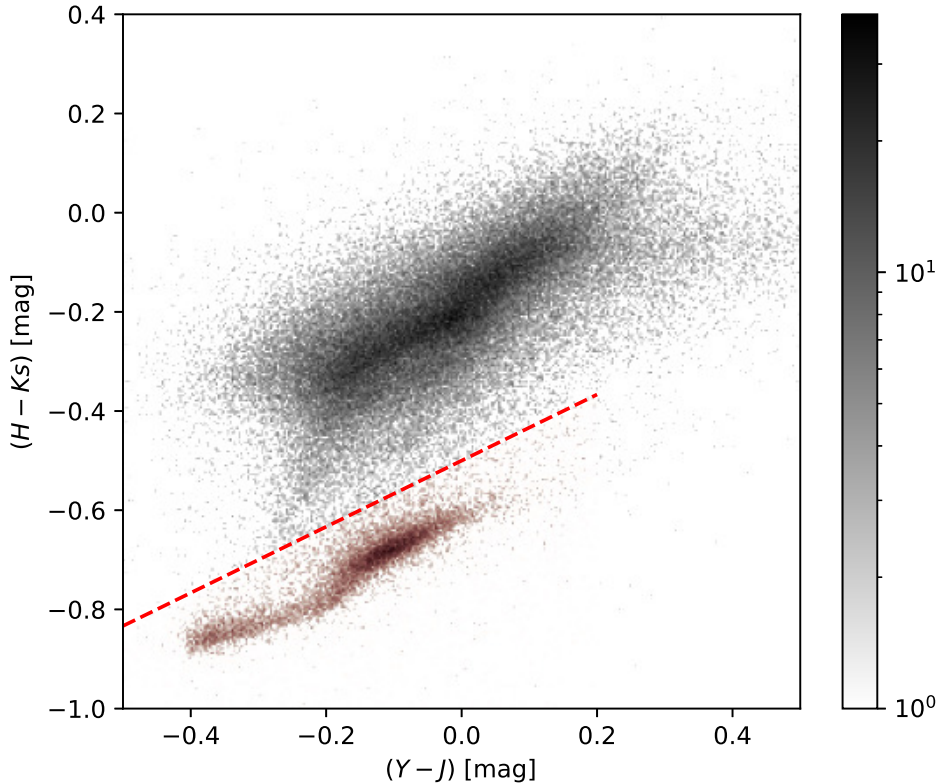
Figure 9.3: 2D $YJHK_s$ colour-colour histogram showing all sources brighter than $24.5\,\mathrm{mag}$ in $K_s$. Black represents all sources, with the colorbar indicating the number of sources in each bin. Sources selected through equation (22) are tinted slightly red.

From a total of 131947 sources, 17125 stars were filtered off through this method. The cut is done in AB magnitude $(Y-J)$ in the range $[-0.5, 0.2]$, with the straight line criteria:

$$(H-K_s) \leq \frac{2}{3}(Y-J) - \frac{1}{2} \tag{22}$$

and the constraint $K_s \geq 24.5$, also in AB magnitude. This cut was selected specifically because it is the faintest magnitude where stars are clearly separated from other sources. In figure 9.3 we see how most of the galaxies are near the origin in $(Y-J)$, and hint slightly more towards negative in $(H-K_s)$. This tells us that the majority of galaxies have a declining flux for longer wavelengths, but not as steep as stars. This is a compound effect of both extinction from interstellar dust, as well as redshift of the varying galaxies. Any source is also made up of likely millions of stars, and the summed light from varying temperature stars gives a somewhat flattened spectrum when compared to just a single one. If the $K_s$ magnitude cut is made to be more faint, i.e. increased, a lot of galaxies will also subside where the stars are. Luckily for us, galaxies with such a sharp IR drop off are typically faint, while the stars are not with their flat tail at long wavelength, allowing for removal of almost all stars using this method.

(a) $Y$, $NB$118, and $K_s$.          (b) $Y$, $J$, and $K_s$.
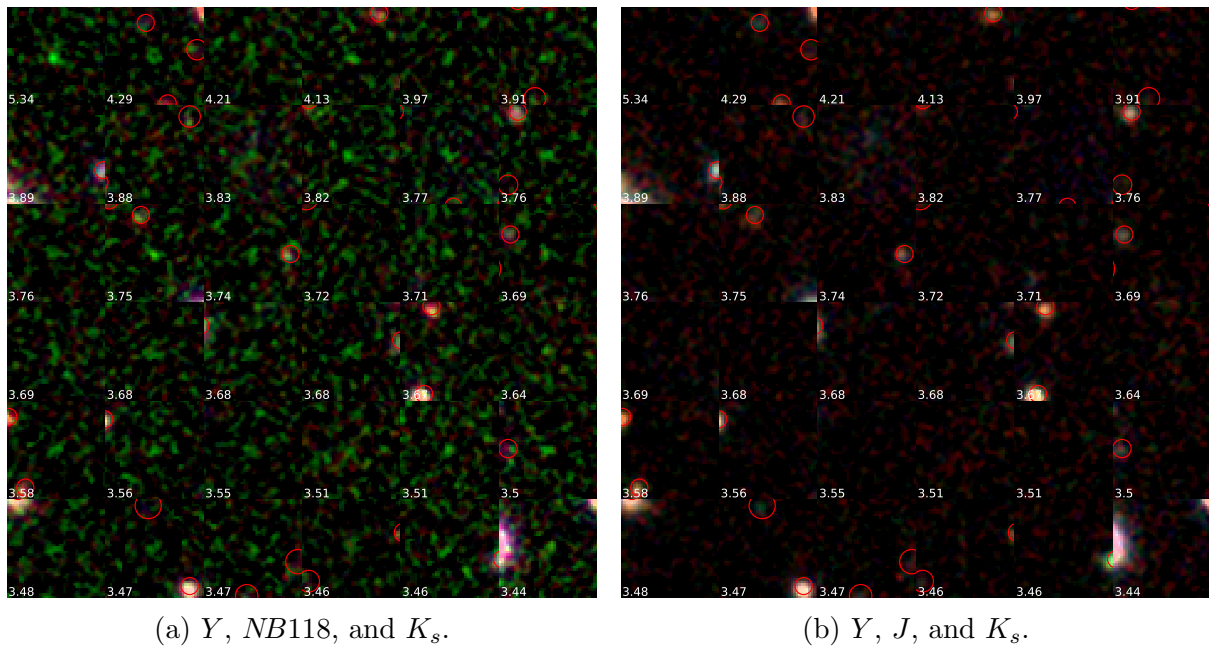
Figure 9.4: Composite RGB images ($6''$x$6''$) for the 16 most statistically significant narrowband image extractions. The red circles represent the sources from `Classic`. The number in the bottom left corner of each source represents how many sigma the source is detected at. For each thumbnail, the extracted object resides at the exact centre, this is purposely done in order to to avoid obscuring the target in question with a marker. Both figures represent the exact same area of the sky, with the only difference being the green channel of the Lupton RGB image. All thumbnails are produced using the same Q and stretch value.

(a) LS$_1$

(b) LS$_2$

(c) LS$_3$

(d) LS$_4$
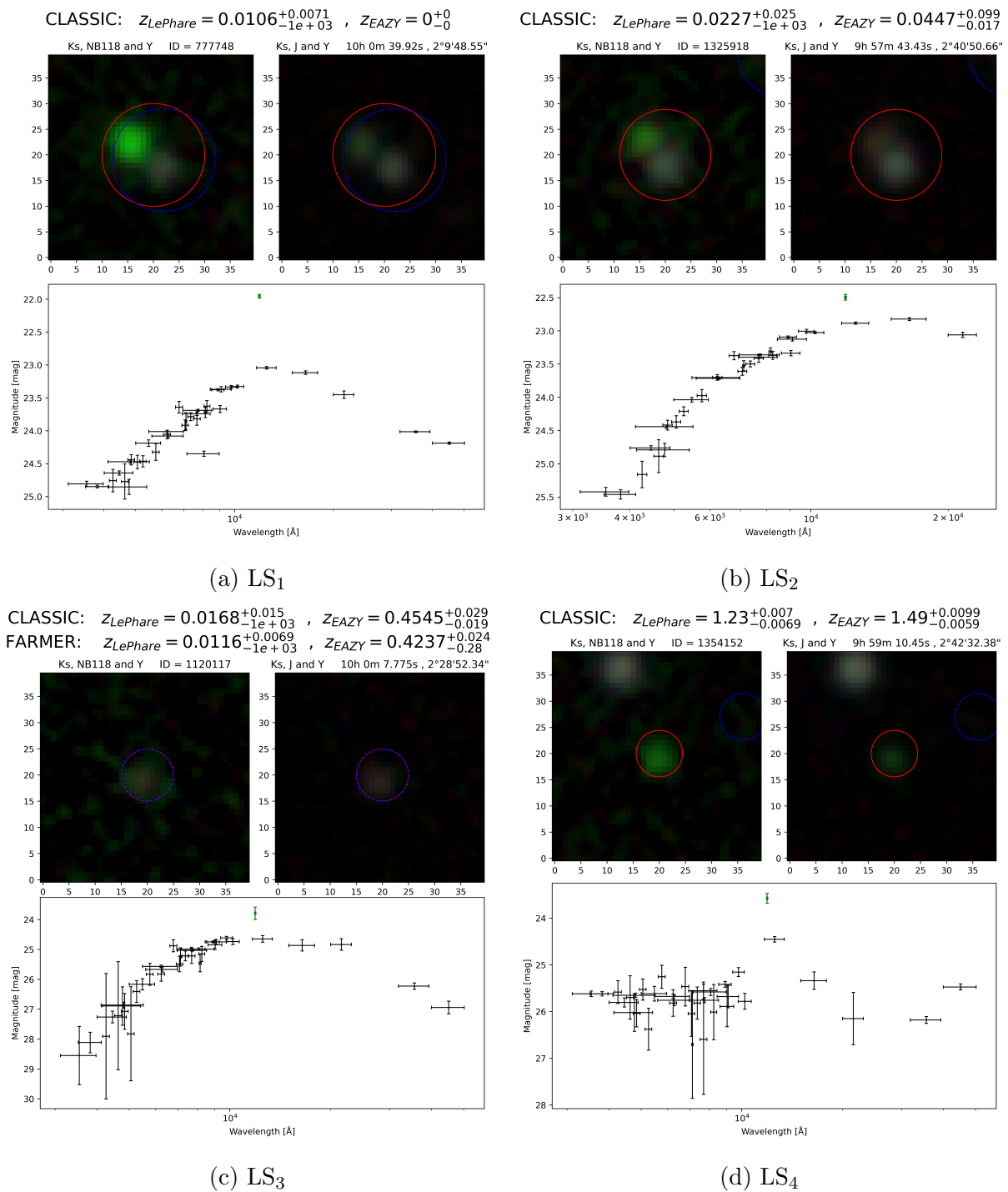
Figure 9.5: SED's and photometric redshift for figure 6.20. IF the source is detected in both catalogues, redshift for both will be shown, but only for `Classic` if that is not the case.
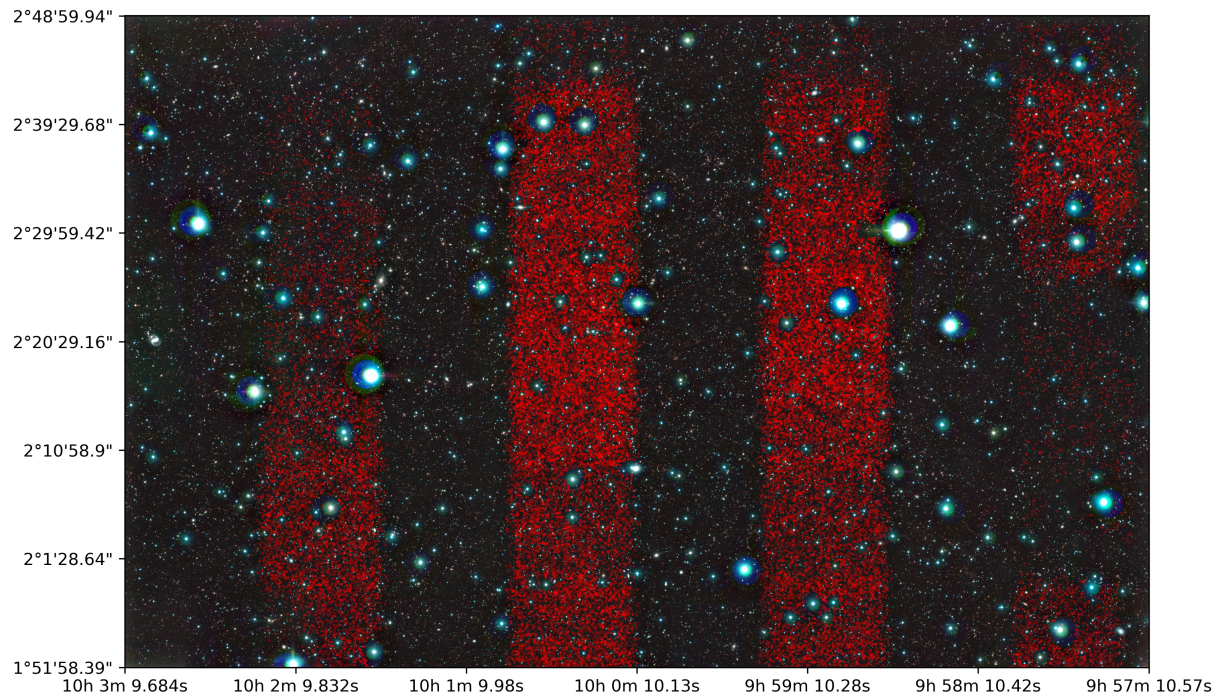
Figure 9.6: Sources extracted using the $NB118$ narrowband strips, fainter than 20 mag detected within $3\sigma$. The red stripes are all individual markers, representing each detected source within the criteria. This background depicts the $YJK_s$ composite RGB image Franx (2019). This figure serves purely as a visual representation of where in the COSMOS field, the narrowband strips are located, and how the varying sky brightnesses of the overlapping pawprints affect the significance of detection. This is the figure corresponding to figure 6.15.
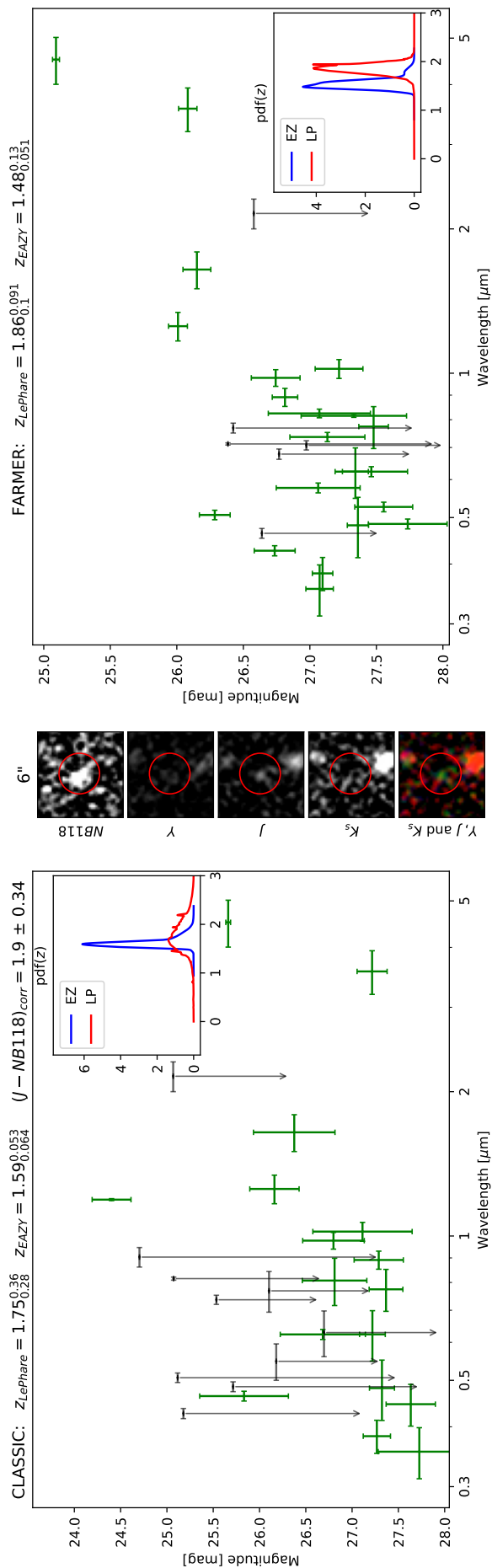
Figure 9.7: $S_1$ Classic ID: 538662, Ra,Dec (J2000)=(10h 0m 20.87s, 1° 56' 23.17'').
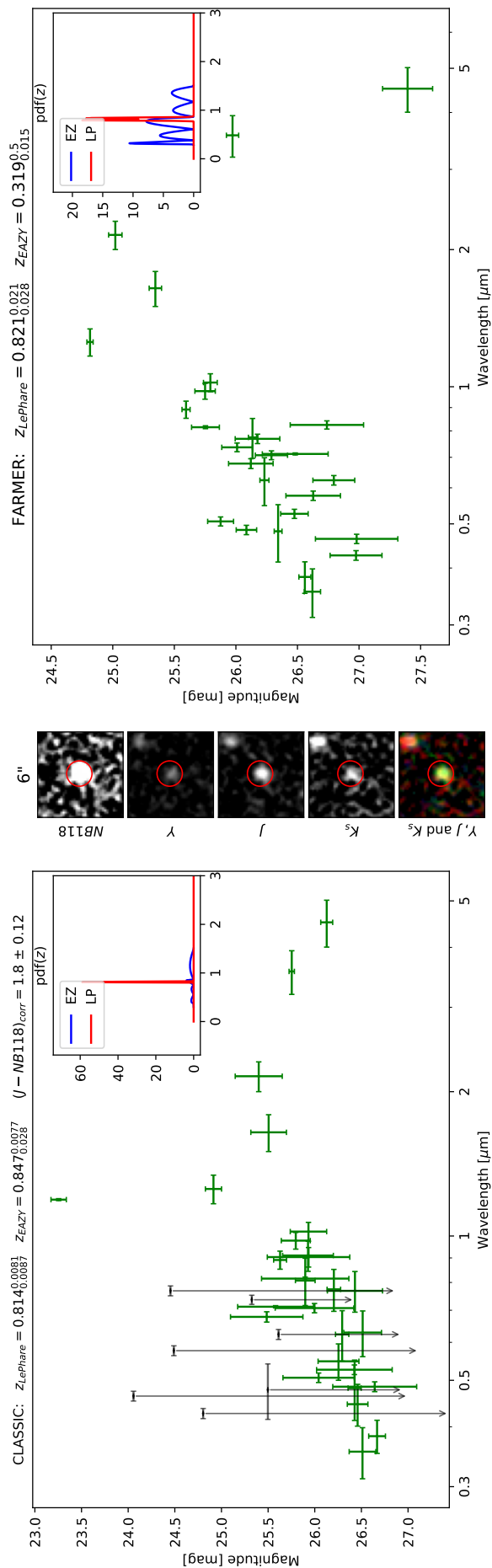
Figure 9.8: $S_2$ Classic ID: 589653, Ra,Dec (J2000)=(10h 2m 16.13s, 1° 59′ 15.67″).
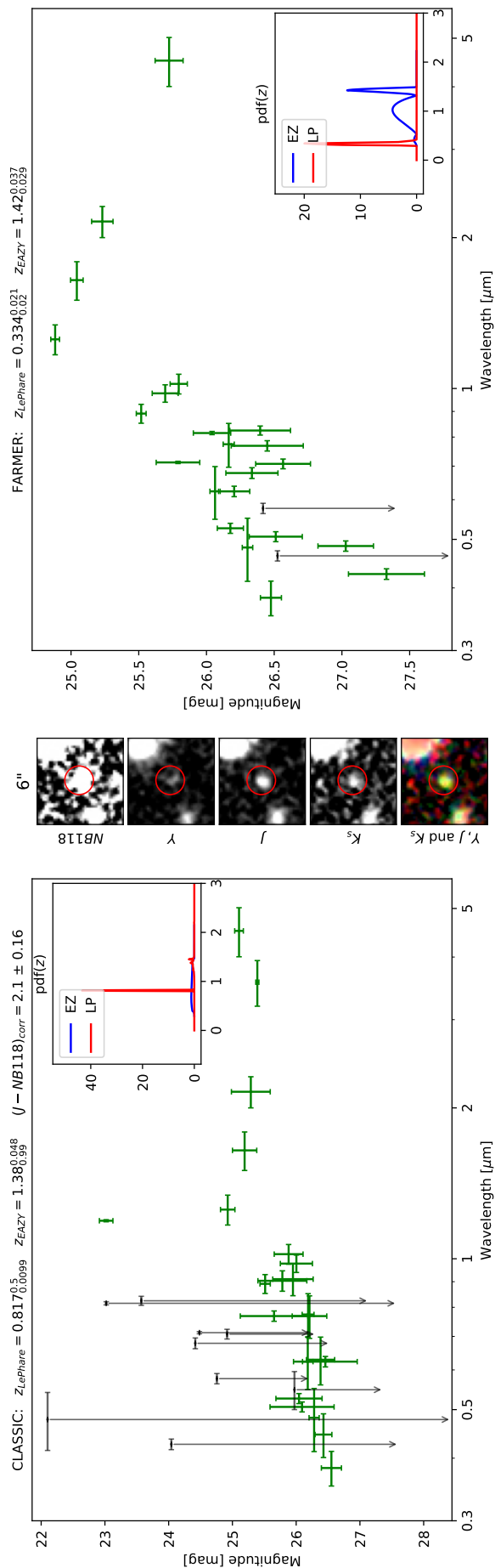
Figure 9.9: $S_3$ Classic ID: 767387, Ra,Dec (J2000)=(9h 57m 44.05s, 2° 9′ 17.21″).
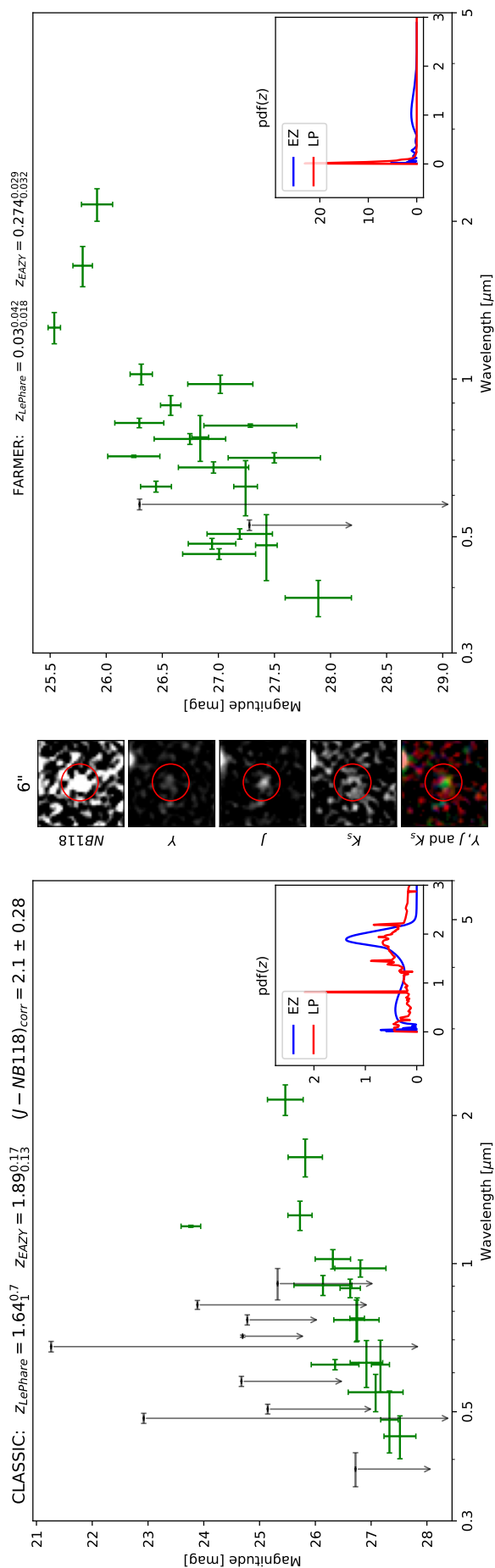
Figure 9.10: $S_4$ Classic ID: 1044607, Ra,Dec (J2000)=(9h 57m 46.34s, 2° 24′ 31.73″).
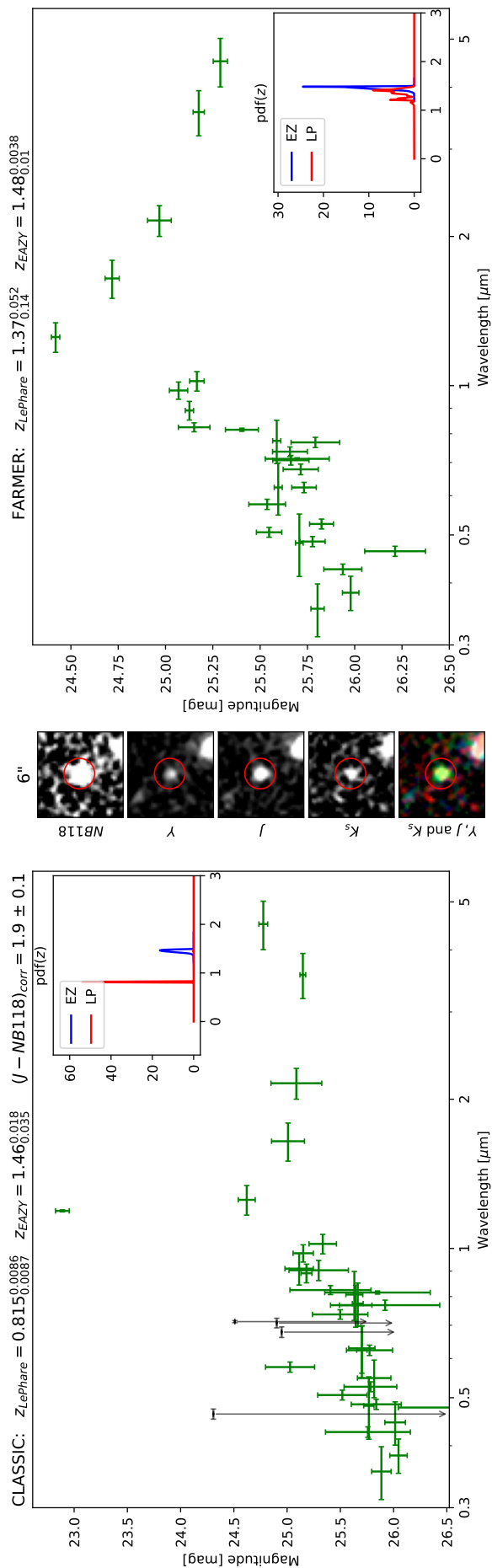
Figure 9.11: $S_5$ Classic ID: 1366525, Ra,Dec (J2000)=(9h 59m 20.94s, 2° 43′ 15.37″).
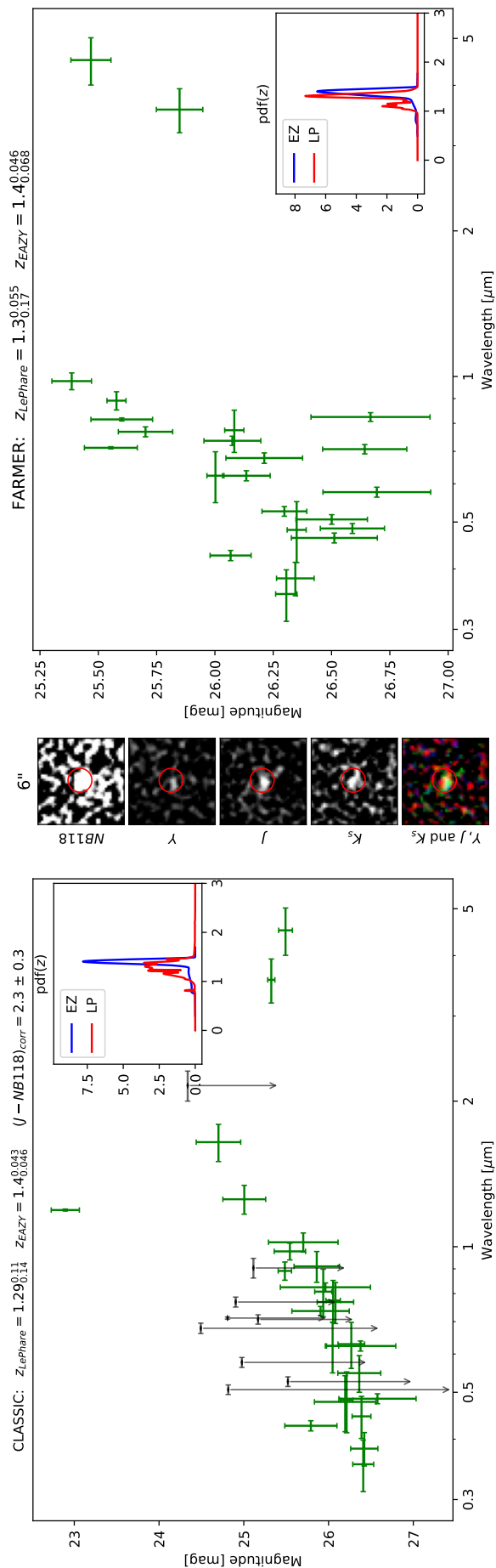
Figure 9.12: S$_6$ Classic ID: 1441184, Ra,Dec (J2000)=(9h 58m 38.8s, 2° 47' 40'').
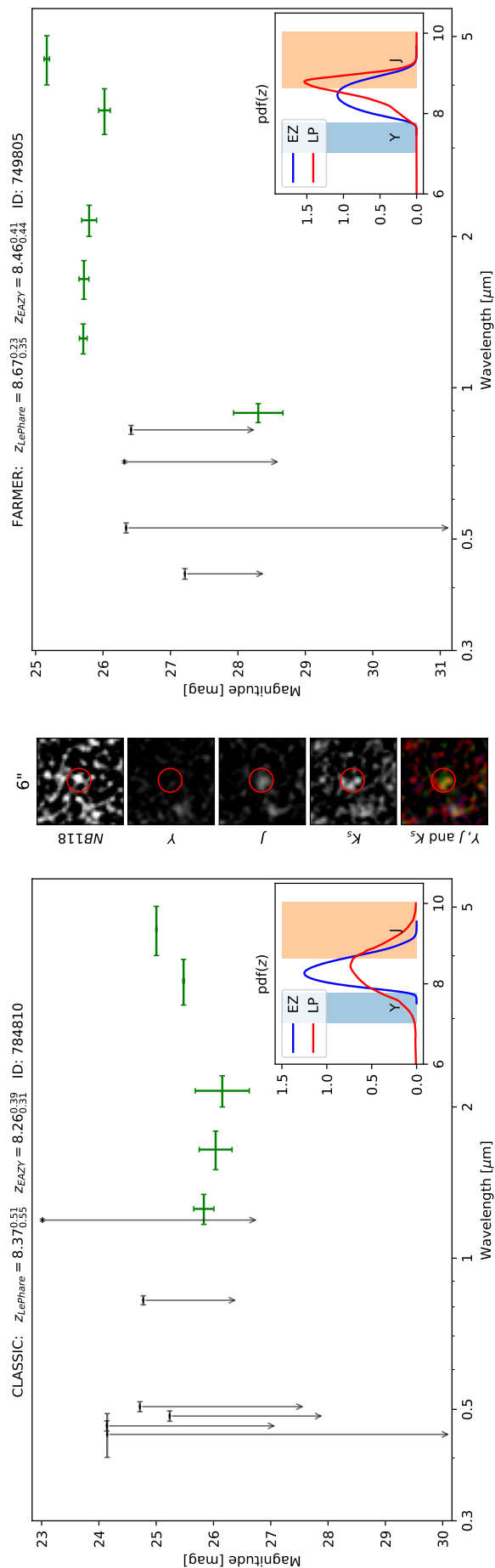
Figure 9.13: J$_1$ Classic ID: 784810, Ra,Dec (J2000)=(10h 1m 47.49s, 2° 10' 15.44'') and is part of the brightest $z \geq 8$ galaxies in Ultra-VISTA (Stefanon et al., 2019) and also known as UVISTA-Y10 and is UVISTA-598 from Bowler et al. (2020).
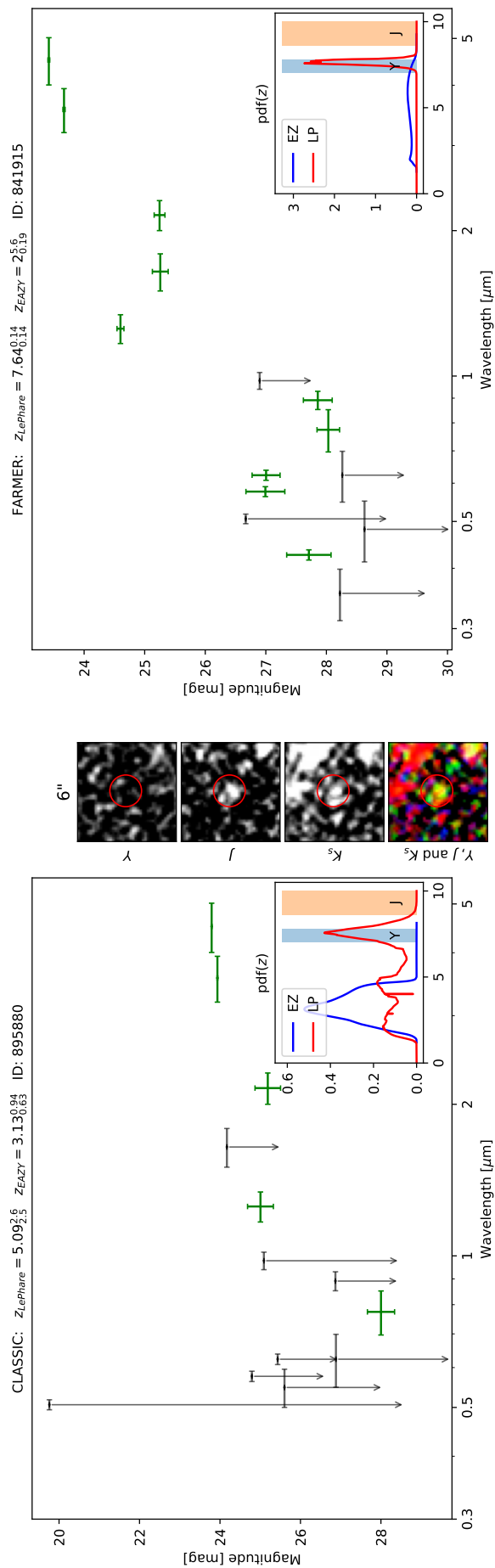
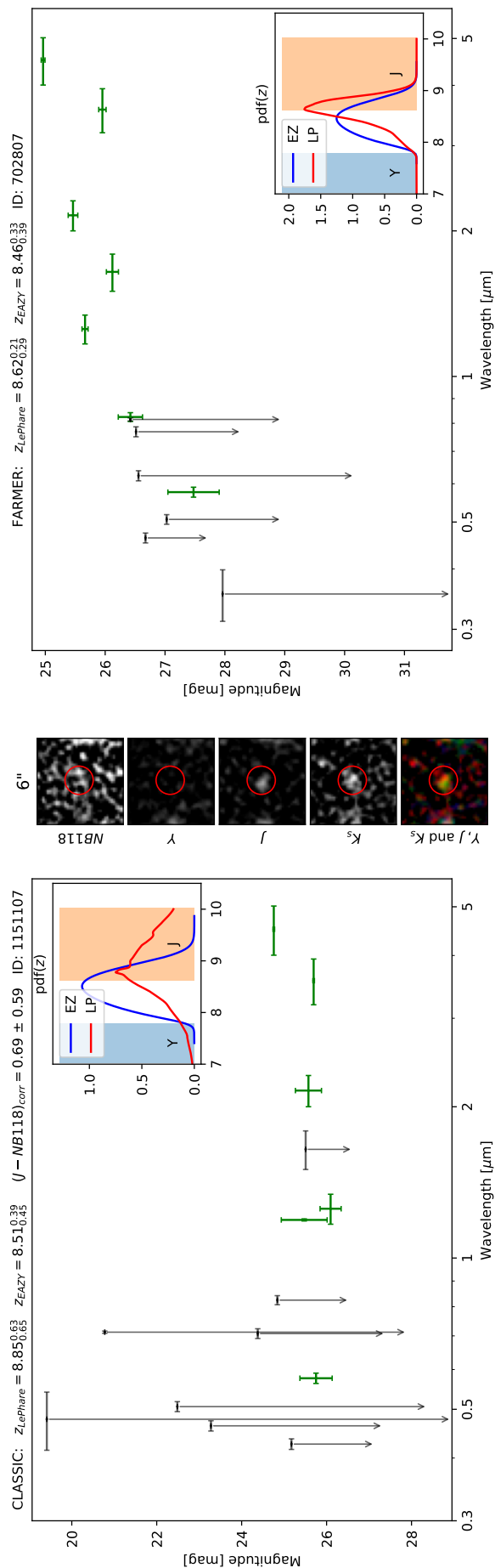Figure 9.14: $J_2$ Classic ID: 895880, Ra,Dec (J2000)=(10h 0m 1.874s, 2° 16′ 14.77″).

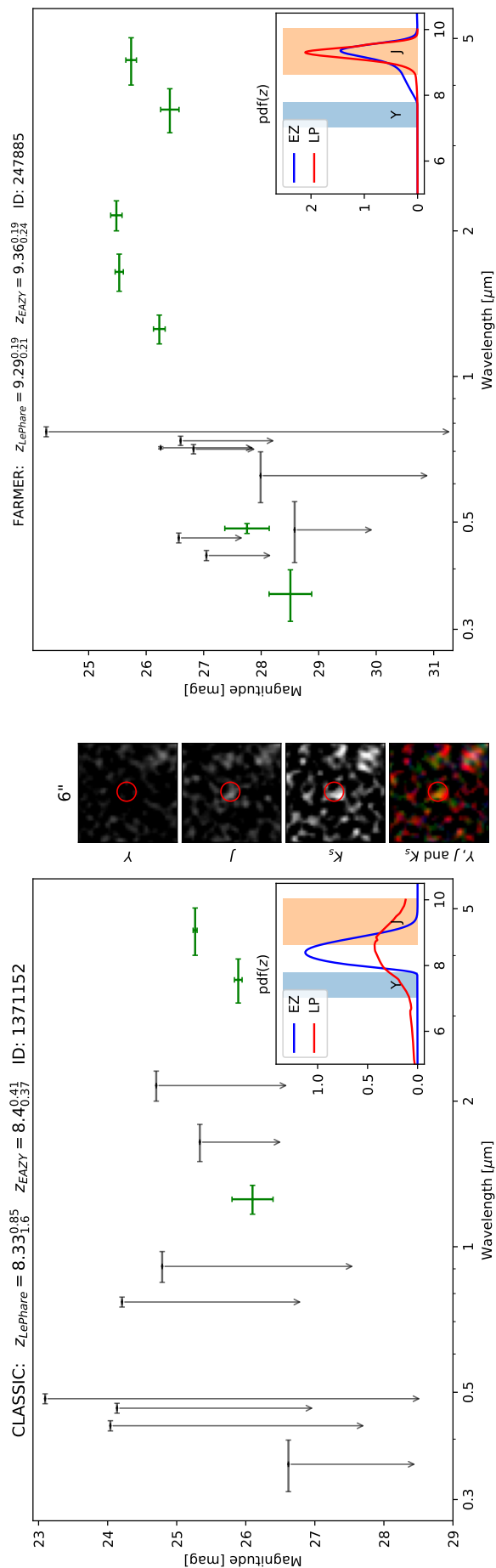Figure 9.15: J$_3$ Classic ID: 1151107, Ra,Dec (J2000)=(9h 58m 59.78s, 2° 30′ 44.46″).

Figure 9.16: $J_4$ Classic ID: 1371152, Ra,Dec (J2000)=(10h 0m 15.97s, 2° 43′ 32.92″) Also known as UVISTA-Y12 and is part of the brightest $z \geq 8$ galaxies in Ultra-VISTA from Stefanon et al. (2019).
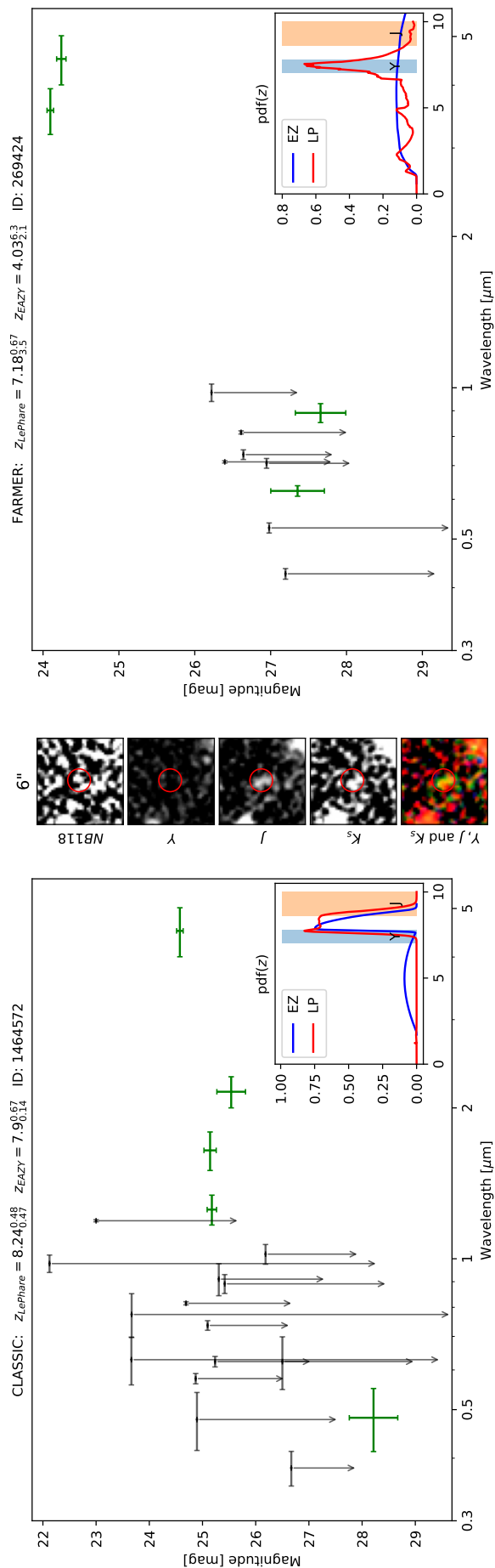
Figure 9.17: $J_5$ Classic ID: 1464572, Ra,Dec (J2000)=(9h 57m 43.02ss, 2° 49' 2.439'').