

## PAPER I.

MARAT MUSIN <sup>1,2</sup>, JIASHENG HUANG <sup>1</sup>, HAOJING YAN <sup>2</sup>,  
*Draft version October 23, 2019*

### ABSTRACT

Context: What can be the use of this catalog?

Aims: We present a catalog of XX galactic and 14 million extragalactic sources with photometric redshifts in Stripe 82 with deep data from SDSS and WISE.

Methods: In order to get consistent photometry, optical sources are convolved to the same PSF for the WISE data we perform "template fitting" - technique that uses high resolution optical sources as a prior to obtain total flux of the low resolution near-IR sources.

Results: T-PHOT code that does "template fitting" uses high resolution and low resolution PSFs to extract robust flux even when source have little flux in near-IR and would not appear in WISE catalog or when sources are blended. Morphological star-galaxy separation is performed and then for each source four photometric redshifts are calculated using SED fitting code (with and without near-IR bands) EAZY and machine learning code ANNz (using two different training set). Available data allows to retrieve photometric redshift up to  $z \sim 0.8$  with the average scatter XX.

Conclusions: Our final catalog with photometry and appropriate errors, CLASS\_STAR indices, photometric redshift and cross-matched spectroscopic data within Stripe 82 is publicly available at XXX.XXX.com

**Keywords:** infrared: galaxies — submillimeter: galaxies — galaxies: starburst — methods: data analysis

## 1. INTRODUCTION

### 1.1. *Lilly-Madau formalism*

### 1.2. *SED fitting as a standard technique of mass and redshift estimation*

SED fitting is now a standard technique of deriving stellar mass and photometric redshifts for a large set of galaxies. In this method multi-band photometry for a given galaxy is fitted to a series of a templates predicted by a certain stellar population synthesis (SPS) model. The best-fit template gives the parameters of the galaxy, including its redshift and mass. Historically, SPS models were using restframe optical photometry. One caveat is the degeneracy between the dust extinction and age of the stellar population, as both make the color of galaxy red, i.e. galaxy can be red because it is intrinsically red with no young massive star and ongoing star-formation, or it can be very dusty, or it can be metal-rich and metals effectively absorb light in the bluer bands. Solution to this is to implement restframe near-IR where light suffers much less extinction (comparing to restframe UV and optical) and thus the degeneracy can be broken. We aim to build the largest sample of galaxies with optical and near-IR photometry over a large sky area. The natural choice for us then is to use optical Sloan Digital Sky Survey (SDSS) and IR all-sky data from Wide-Field Infrared Survey Explorer (WISE).

### 1.3. *problems associated with construction of the catalog*

<sup>1</sup> Chinese Academy of Sciences South America Center for Astronomy (CASSACA), National Astronomical Observatories, Chinese Academy of Sciences, Beijing 100012, China

<sup>2</sup> Department of Physics & Astronomy, University of Missouri, Columbia, MO 65211, USA

Blending, poor spatial resolution in IR our method – template fitting

### 1.4. *Goal of this paper*

In this paper we present our technique for construction a catalog of galaxies with reliable SED data in optical and near-IR in Stripe 82 field. We discuss data selection, sources identification in different bands and problems associated with it.

## 2. DATA DESCRIPTION

### 2.1. *SDSS and Stripe 82*

The imaging component of the SDSS, which was done in five broad bands ( $u'g'r'i'z'$ ), has covered 14,555 deg<sup>2</sup>. In most area, the SDSS only scanned for one pass at an exposure time of 53.9 seconds per band, and thus is rather shallow (for example, the  $r'$ -band 5  $\sigma$  limiting magnitude is 22.2 mag). For this reason, in most cases the SDSS can only probe the normal galaxy population up to  $z \approx 0.4$ . However, the Stripe 82 region, which is a long stripe along the equator that spans  $20^h < RA < 4^h$  and  $-1.26^\circ < Dec < 1.26^\circ$ , totalling in  $\approx 300$  deg<sup>2</sup> is the exception. It was repeatedly scanned ( $\sim 70$ -90 times, depending on RA) for calibration purpose during the survey (Adelman-McCarthy et al. 2007), and thus the combined scans can reach much better sensitivities.

A number of teams have created deep Stripe 82 stacks and made them available to public. The first such stacks were produced by Annis et al. (2014) based on the data obtained up to December 2005 (20-35 runs), which achieved 1-2 magnitude deeper limits than the single-pass SDSS images. Several other teams (e.g., Jiang et al. 2009; Huff et al. 2014) produced different stacks using different procedures to optimize the image qualities.

Jiang et al. (2014; hereafter J14) released a new version of stacks using only the images that were taken under

the best weather conditions. These stacks are  $\sim 0.2$  mag deeper than those produced by Annis et al. (2014), reaching  $5\sigma$  limits of 23.9, 25.1, 24.6, 24.1, 22.8 mag in  $u'g'r'i'z'$ , respectively, and also have better PSF characteristics. We adopt these stacks in our work.

## 2.2. Structure of SDSS Stripe 82 files

We use description from J14 to present the structure of optical data. An SDSS run (strip) consists of six parallel scanlines, identified by camera columns (Figure ??). The scanlines are 13.5 arcmin wide, with gaps of roughly the same width, so two interleaving strips make a stripe that consists of total 12 scanlines (columns).

The size of each co-added SDSS image is 2854 x 2048 pixels, or roughly 18.8' x 13.5' (RA x Dec), with a pixel size of 0.396" and an average full width at half maximum (FWHM) of  $\sim 1.5''$  in u-band,  $\sim 1.3''$  in g-band, and  $\sim 1''$  in r-, i-, and z-bands. In total there are 401 SDSS images in each column and overall  $12 \cdot 401 \cdot 5 = 24,060$  SDSS images in all 5 bands. Each SDSS image has a corresponding *weight.fits* image, that records relative weights at individual pixels.

## 2.3. WISE and unWISE

WISE (Wright et al. 2010) is a near-to-mid IR space telescope launched in 2009 and has performed an all-sky imaging survey in four bands at 3.4, 4.6, 12, and 22  $\mu$ m (denoted as W1, W2, W3, and W4, respectively). During its original mission phase from 2010 January 7 to 2010 August 6 (the “4-band Cryogenic” phase), WISE surveyed the entire sky 1.2 times in all four bands simultaneously until the solid hydrogen coolant in the outer cryogen tank was depleted. It then entered the “3-band Cryogenic” phase for the next 54 days, during which time it mapped an additional 30% of the sky in W1, W2 and W3. When the coolant in the inner tank was also depleted by 2010 September, only W1 and W2 are operational. The NEOWISE project took over the mission on 2010 October 1 and brought it into the four-month “Post-Cryo” phase to survey the sky in these two bands for near-earth objects until 2011 February 1 (see Mainzer et al. 2014). The telescope was then put into hibernation for the next 35 months as the funding stopped. The extended NEOWISE project reactivated it in 2013 December to continue the two-band observations (“NEOWISE Reactivation”) through today.

The WISE team made three data releases separately for the 4-band Cryogenic, the 3-band Cryogenic and the NEOWISE Post-Cryo phases in 2012 March, 2012 June and 2013 May, respectively. To take the advantage of these repeated observations, the WISE team also made the “AllWISE Data Release” in 2013 November by combining all the WISE data available till then (see Cutri et al. 2013; for details). The included image products, known as the “Atlas Images” reach the nominal  $5\sigma$  limits of 0.054, 0.071, 0.73, and 5.0 mJy in the four bands, respectively.

To optimize the detection of isolated sources, the WISE team has been using a special treatment when combining images, namely, the single-exposure images are convolved with the individual point spread function (PSF) during the stacking process. However, this operation has the drawback that it reduces the spatial resolution of the fi-

nal stacks, which is not desirable in many applications. To deal with this problem, Lang (2014; hereafter L14) reprocessed all the WISE images independently without the PSF convolutions, and produced the stacks that preserve the original WISE spatial resolutions. These image products of L14, dubbed as the “unWISE” images, have the PSF full-width at half maximum (FWHM) values of 6" in W1, W2 and W3 and 12" in W4. Sensitivity of bands W3 and W4 is  $\sim 10$  times worse than that of bands W1 and W2 and is not sufficient so that at least 10% of the optical sources have secure W3 and W4 fluxes. We use unWISE W1 and W2 images for this work.

## 2.4. Structure of unWISE files - once again maybe I need to omit it in the paper

The unWISE coadds are on the same tile centers as the WISE tiles with 18,240 images per band, 1.56 x 1.56 degrees each. The tiles are named by their RA, Dec center: tile ”0591p530” is at RA = 59.1, Dec = +53.0 degrees; i.e., the first four digits of the tile name is  $int(RA \cdot 10)$ , then ”p” for +Dec and ”m” for -Dec, then three digits of  $int(abs(Dec) \cdot 10)$ . For each tile and band W1-w4, several images are produced, we shall list only the ones that we make use of:

- unwise-0000p000-W1-img-m.fits - ”Masked” image, 2048 x 2048 pixels, TAN projected at 2.75"/pixel. Background-subtracted, in units of ”Vega nanomaggies” per pixel:  $mag = -2.5 \cdot (\log_{10}(flux) - 9)$ . This is the science image, the word ”masked” means that some pixels have no unmasked pixels and no measurement at all: pixel value 0 and infinite uncertainty.
- unwise-0000p000-W1-std-m.fits - Sample standard deviation (scatter) of the individual-exposure pixels contributing to this coadd pixel.

Three unWISE images centered at the same RA cover the whole width of Stripe 82 in Dec ( $-1.26^\circ$  to  $+1.26^\circ$ ). We shall call three such unWISE images a frame. There may be up to 72 SDSS images within one frame.

## 2.5. Spectroscopic sample

Spectroscopic redshifts are used to perform star/galaxy separation and calibrate photometric redshift estimation: determine a set of templates and photometric offsets for the template-fitting code and train machine learning algorithms. We shall call such full sample of sources a star/galaxy catalog and a subset with only sources classified as galaxies or QSOs with  $z_{spec} > 0$  as spectroscopic catalog. Originally, we constructed our training set using spectroscopic data from the SDSS DR14 (Bolton et al. (2012)). Spectroscopic redshifts and classification into galaxies, stars and QSOs for the SDSS DR14 catalog are calculated using principal component analysis (PCA). The software `idlspc2d` is used to perform, at each potential redshift, a least-squares fit to each spectrum, using a fairly general set of models, for galaxies, for stars, for cataclysmic variables, and for quasars. The best fit model and redshift is chosen and assigned for the object.

Their spectroscopic sample consists of a wide variety of galaxies, stars and QSO with no cuts on color, although it is rather limited in terms of redshift (Strauss et al. (2002)). Stripe 82 data are 1-2 mag deeper than single-pass images and therefore potentially has a lot of sources

at higher redshifts. We decided to extend our training set by cross-matching galaxies in our photometric catalog with spectroscopic measurements of other, publicly available surveys, such as 6dF, WiggleZ, DEEP2, VVDS and VIPERS.

In Tab.1 we list auxiliary catalogs with selection criteria, median redshifts for galaxies and QSO and total number of sources that were added to our training set. For each source, we use published redshift quality flag to select trustworthy sources, but also add stars when possible (which sometimes have special flag or assigned with negative redshift).

We cross-matched the galaxies from the star/galaxy set with our photometric catalog using 2" as maximum separation in RA and Dec and this resulted in 268,587 sources - galaxies (59%), stars (29%), QSOs (11%) and not classified objects (1%).

### 3. OVERVIEW OF METHODS FOR ANALYSIS

The most critical factor in SED fitting is consistent photometry in the involved bands, i.e., the photometry should include the same fraction of light across all bands so that the colors are defined in a consistent manner. This is challenging in our case because the spatial resolutions of WISE are at least  $6\times$  worse than that of the SDSS. For this reason, the objects detected in WISE often suffer blending. Even for relatively isolated WISE sources, the photometric apertures appropriate for the (low resolution) WISE images cannot guarantee the same fraction of light being included as what is done in the (high resolution) SDSS images. Such a systematic offset, which is different for every galaxy, severely skews the SED fitting.

To best address this problem, we opt to use the T-PHOT software (Merlin et al. 2015), which recently emerged as a robust and flexible tool to perform “template fitting”. The basic idea is to use a high-resolution image (here an image from the SDSS) as the prior to build the morphological template of the source under question, convolve this template with the PSF of the low-resolution image (here the corresponding image from the unWISE), and fit this degraded template to the low-resolution image to obtain the total flux that is within the aperture as defined by the high-resolution image. In this way, we get reliable color information (i.e., flux ratio) in the most consistent manner.

While T – PHOT is much more user-friendly as compared to its predecessors (i.e. TFIT Laidler et al. (2007)), running this software is still non-trivial. It not only requires careful tuning of parameters but also several tedious preparatory steps with both the high- and the low-resolution images. Here we detail our procedures.

#### 3.1. Initial preparation of unWISE and SDSS images

T-PHOT requires that the low- and the high-resolution images have the same orientation and the same World Coordinate System (WCS) reference position, the latter of which is defined by the FITS keywords (CRVAL1, CRVAL2). It also requires that their pixel scale ratio must be an integer. To meet these prerequisites, we carried out the following procedures utilizing the SWarp software (Bertin et al. 2002), which can subsample or bin an image to any pixel scale and then re-project to an arbitrary orientation at any tangential point.

We first rescaled the unWISE images from 2.750"/pix to 2.772"/pix. As the scale of an SDSS image is 0.396"/pix, this makes the ratio of their pixel scales an integer ( $2.772/0.396 = 7$ ). We kept the same orientation, which is always North-up and East-left, and the same reference position for each unWISE image. This process was done for both the W1 and the W2 unWISE images, which are always aligned.

For a given SDSS image, we oriented it to North-up and East-left, and re-projected it at the tangential point as defined by the reference position of the unWISE images that it lies within. In other words, the FITS keywords (CRVAL1, CRVAL2) of the re-projected SDSS image are the same as those of the unWISE images. As an unWISE image covers much larger area and thus encompasses multiple SDSS footprints, the tangential projection point of a reprojected SDSS image is often outside of its coverage. In the extreme cases, it can be as far as 0.7° outside of the image itself.

About 30% of the SDSS images’ footprints lie across two adjacent unWISE fields, and therefore need to be treated separately. If an SDSS image has more than 60 arcmin<sup>2</sup> belonging to adjacent unWISE fields, such image is duplicated, and each copy is reprocessed with respect to the appropriate unWISE field as described above. This increases the number of SDSS images from 4,812 to 5,556 per band. All SDSS images in col02 and col11 have  $\sim 2.9'$  overlap with unWISE images centered at Dec=-1.7 and Dec=1.5 respectively. Processing of such small region requires construction of  $\sim 930$  PSFs per band and 2,000 hours of CPU time and is unviable. This excludes 11.34 deg<sup>2</sup> from the total area for which catalog is constructed.

We note that the above procedures were done for both the science images and the standard deviation (for the unWISE) or the weight (for the SDSS) images. After the subsampling and re-projection, the standard deviation or the weight value per pixel no longer preserves the absolute scale. In other words, the value of a given pixel on an unWISE (SDSS) reprojected standard deviation (weight) image no longer reflects the true standard deviation (weight) on that pixel. Fortunately, this does not affect the performance of T – PHOT, as it only uses these values in a relative sense (i.e., the absolute scale does not matter).

\*\*\* describe unWISE cutouts for each SDSS image?  
\*\*\*

We also note that one special treatment needs to be done for the saturated pixels in the unWISE standard deviation images. They are all assigned with zero values in the unWISE release, which is invalid for T – PHOT. We therefore use the IRAF/imcalc task to set such values to “9999” before reprocessing.

#### 3.2. Input SDSS source catalog

J14 produced object catalogs from their stacked images using SExtractor (Bertin & Arnouts 1996). While it is tempting to use them directly as the input source catalogs for T – PHOT on the unWISE images, several caveats prevented us from taking this approach. For example, these catalogs are not cleaned of duplicated sources from the overlapped areas between adjacent images; they are not matched among the bands and thus have different number of sources for each band; the object detection

**Table 1**  
Catalogs with spectroscopic redshifts.

Survey Name	Selection criteria	References	Number of sources in the star/galaxy <sup>1</sup> (spectroscopic <sup>2</sup> ) set	Median redshift of the spectroscopic set	Comments
SDSS DR14	$z_{\text{Err}} < 0.1$ and $z_{\text{Warning}} = 0$	Bolton et al. (2012)	225,076 (115,359)	0.316	
6dF	$Q = 3, 4, 6$	Jones et al. (2004), Jones et al. (2009)	375 (328)	0.054	no $z_{\text{Err}}$
WiggleZ	$Q_{\text{op}} = 3, 4, 5$	Drinkwater et al. (2010), Parkinson et al. (2012)	10,674 (9,555)	0.550	
DEEP2	$Z\text{QUALITY} = 3, 4$ and $z_{\text{Err}} < 0.01$ or $\text{CLASS} = \text{STAR}$	Davis et al. (2003), Newman et al. (2013)	15,730 (11,683)	0.957	
VVDS	$1 < Z\text{FLAGS} < 20$	Le Fèvre et al. (2013), Gavazzi et al. (2008)	10,137 (8,383)	0.602	no $z_{\text{Err}}$
VIPERS	$2 < z_{\text{flag}} < 5$ and $\text{class-Flag}=1$ and $z_{\text{spec}} > 0$	Scoville et al. (2018), Guzzo et al. (2014)	6,595 (5,460)	0.690	$\sigma_z = 0.00054 \cdot (1 + z)^3$
Full sample			268,587 (150,767)	0.409	

<sup>1</sup> Includes stars, galaxies, QSOs and unidentified sources with secure redshift. The sample is used for the star-galaxy separation.

<sup>2</sup> Includes galaxies and QSO and is used for EAZY and ANNz testing.

<sup>3</sup> No  $z_{\text{spec}}$  errors are provided, these values are given in the paper.

threshold was set too high and many faint objects were excluded; bright and saturated objects have clusters of false detection; etc. For these reasons, we constructed our own input source catalogs.

### 3.2.1. Rationale

In addition to providing input source information to T-PHOT, our SDSS catalog also provides optical SEDs for all the detected sources. For the latter, it is critical that the colors of a given source are measured consistently across all five bands, or in other words, the photometric aperture of a source must include the same fraction of total light in any given band. To achieve this goal, we took the standard approach by performing “PSF-matching” of the SDSS images. For a given SDSS field, we first matched the size of the PSF in g, r, i, and z-bands ( $\sim 1.0\text{--}1.3''$ ) to that of u-band, which always has the largest PSF ( $\sim 1.5''$ ; see J14). We then ran matched-aperture photometry using r-band as the reference band, which detects more sources than any other band.

The PSF-matching step was the most tedious part of the process, which required a lot of human intervention. To derive the convolving kernels between two images, we first must obtain the PSFs of both images. As the SDSS PSF varies from image to image, we had to construct it for each image individually. In total, we built 27,780 SDSS PSFs (5,556 per band).

### 3.2.2. PSF construction and PSF-matching

An empirical PSF is best derived by combining a large number of bright, isolated point sources (also known as the “PSF stars”) distributed over the entire image, given that PSF is constant across the image. The most robust method to derive a PSF is to run the IRAF task “psf” interactively on a list of candidate PSF stars and to retain only the best ones in the construction. Given the huge number of images involved, however, this was not practical. After extensive tests, we settled on an approach that would result in reliable PSF stars, which would then allow us to run the “psf” task non-interactively in most cases.

The key in the PSF star selection is to determine whether an object is a point source of high image quality. We first removed the objects that are saturated or too faint. The magnitude cuts are typically  $\sim xx$  and  $\sim xx$  mag at the bright and the faint ends, respectively. We then used the stellarity identifier produced by SExtractor, “CLASS\_STAR”, to select the potential point sources. The output value of this parameter ranges from 0 (most likely non-stellar) to 1 (most likely stellar), and we rejected the objects that have  $\text{CLASS\_STAR} < 0.85$ . The survived candidates were further refined by comparing their “core” magnitudes and the total magnitudes. The smaller the difference between the two means the more compact the object is. The size of the aperture depends on the particular band, and is  $\sim 3.56''$  in general. We adopted the criterion of  $< 0.xx$  mag for this difference in this refinement. Finally, we only retained the stellar objects that are not close to the image edge ( $> xx$  pixels to the boundary) and do not have a neighbor within  $XX''$ .

The retained point sources were then supplied to the IRAF/psf task for PSF construction. Typically, each SDSS image had more than 20 point sources in the end to build its PSF. Lower value appeared in only a few images, most of which are in the less sensitive u-band. On the other hand, there are plenty of images that have hundreds of qualified point sources, and using all of them would actually make the PSF construction exceedingly time-consuming. Therefore, we applied an upper limit of 160 point sources per image when running IRAF/psf.

\*\*\* Parameters, Moffat function \*\*\*

The output from this task was then converted by using IRAF/seepsf to the format that can be further used. At this point, we visually examined all the resulted PSFs. If a PSF showed some undesirable features indicative of contaminations (for example, elongation, gradient in the wing, neighbors not identified previously), we would manually reject the problematic point sources and rebuild the PSF. Finally, all the PSF images were normalized to the total count of unity using wcstools/sumpix (Mink 1998) and IRAF/imapath tasks.

Before the PSF matching could be carried out, we would need the transferring kernal between the two PSFs under question. This was done by suppling the PSFs to `IRAF/lucy` that uses the algorithm developed by Richardson (1972) and Lucy (1974). Finally, the PSF-matching to the u-band was performed by `IRAF/psfmatch` that used the g'r'i'z' images and the relevant kernels as the inputs. For the sake of simplicity, we shall refer to these u-band PSF-matched images as the "g2u", "r2u", "i2u", and "z2u" images, respectively.

A subtle problem with PSF matching is that the convolution introduces extra correlation among nearby pixels, which artificially suppresses the noise. In other words, the noise of the PSF-matched images would be underestimated if it were measured by the pixel-to-pixel background fluctuation. We will return to this point in the next section.

### 3.2.3. Optical catalog construction

The catalogs were produced by running SExtractor in the dual-image mode, where the r2u images were set as the detection images and the photometry was run in turn on the u, g2u, r2u, i2u, and z2u images.

\*\*\* You're missing a lot of details here!!! You should provide these:

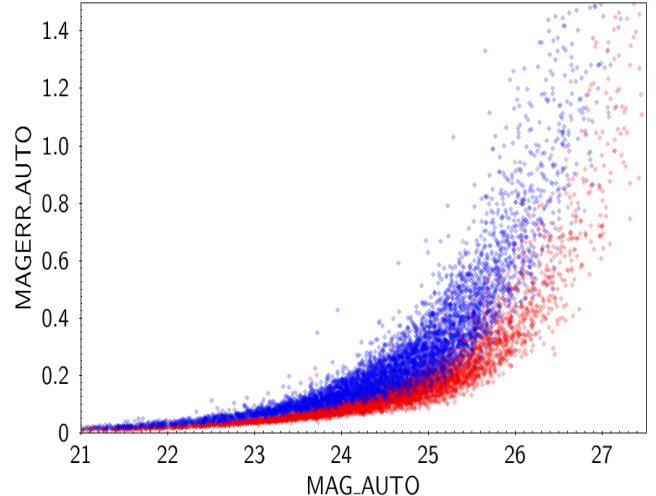
- What is the filter?
  - What is the detection threshold?
  - What is the number of connecting pixels (DETECT\_MINAREA)?
  - What flavors of magnitude are derived?
- \*\*\*

The image weighting scheme was set to "MAP\_WEIGHT", and we used the original weight maps from J14. Strictly speaking, these weight maps no longer reflect the pixel-to-pixel weights because the science images have been convolved. Nonetheless, this does not significantly impact the source detection because the general trend of the weight is still preserved (e.g., it is unlikely that the convolution would transform a low-weight pixel into a high-weight one).

However, as the correlation of pixels due to the PSF matching is not taken into account by these weight maps, the photometric errors reported in our `SExtractor` runs are underestimated (except in u-band because the u-band images were not convolved). This is demonstrated in the left panel of Figure 1, where we show the MAGERR\_AUTO vs. MAG\_AUTO diagram for a randomly chosen g-band image. The blue and the red points represent the photometry using the original g-band image and the g2u image, respectively. The same weight image was used in both cases. Clearly, the red points are below the blue ones, or in other words, the photometric errors on the PSF-matched image are significantly underestimated.

### 3.3. Kernels for T-PHOT

Running T-PHOT requires the convolution kernel between the Hi-res and the Lo-res PSFs, which in our case are the PSFs of the SDSS "r matched to u" images and the unWISE W1 and W2 images, respectively. While in theory the PSF of a "r matched to u" image should be the same as that of the u-band image, in reality the matching is never exact. Therefore, we could not use



**Figure 1.** Magnitude error vs. absolute g-band magnitude for one SDSS image. Magnitude errors in the g2u image(red) are underestimated as compared to the original SDSS g-band magnitude errors (blue).

the u-band PSFs but had to construct the PSFs of the "r matched to u" images to optimize the performance of T-PHOT. We already had a list of point sources when constructing the original r-band PSFs (see §3.2.2), which could be used here again. We visually examined them on the "r matched to u" images, removed all those that are contaminated by neighbors due to the larger PSF in "r matched to u", and used the refined list of point sources to derive the "r matched to u" PSFs in the same way as described in §3.2.2.

For the unWISE images, we opted to derive their PSFs independently in all tiles instead of using the WISE PSFs as published by the WISE team. While the WISE PSFs were rather stable throughout the mission, they should not be used directly for our purpose because the image stacks are the combination of multiple single-exposure images of varying orientations. As the result, the PSF of a WISE image stack is no longer that of a single exposure, and also changes significantly from tile to tile.

There are 240 unWISE image tiles within the Stripe 82 footprint, and thus we derived 480 unWISE PSFs (240 each in W1 and W2, respectively) using point sources that were derived in the same way as described in §3.2.2. All PSFs were constructed by running `IRAF/psf` interactively.

All the unWISE PSFs were normalized to unity total counts using `IRAF/imarith` and sub-sampled in size by the factor of 7 using `IRAF/imlintran` to match the pixel scale ratio between the unWISE and the SDSS images (see §3.1). Finally, we used `IRAF/lucy` to generate the individual convolution kernels between the "r matched to u" images and the unWISE images. In total there are 11,112 kernels: 5,556 as a result of convolving "r matched to u" to W1 band and 5,556 - convolving "r matched to u" to W2 band.

### 3.4. Running T-PHOT

With the r2u source catalogs (§3.2.3) and the transformation kernels between the r2u and the unWISE images, we performed T-PHOT on each r2u/unWISE pair. This was by far the most computing-intensive step, which was

carried out at the Lewis cluster of the High-Performance Computer (HPC) at the University of Missouri.

Following the recommendation of Merlin et al. (2015), we ran T-PHOT for two passes. The first pass performs template fitting “globally” for the Hi-res sources on the Lo-res image at the exact source locations as provided by the input Hi-res source catalog. In the meantime, it also allows the fitting position of each source to shift by a small amount around the prior to find the most optimal position that would result in even better fit than at the prior position. The second pass then uses these refined positions to perform the template fitting again. In all our testing cases, the results from the second pass were significantly improved as compared to those from the first pass, which justified spending extra computing time to finish two passes.

As mentioned in §3.1, there are 5,556 SDSS images per band in the Stripe 82. Therefore, there were 5,556 r2u/W1 pairs and the same number of r2u/W2 pairs. Each T-PHOT pass took about 3 hours of computing time at the cluster, and the whole process totaled  $\sim 66,700$  hours.

The output T-PHOT catalogs in W1 and W2 were merged with the SDSS optical catalogs for the future use. The sensitivity of WISE is not sufficient to detect all the SDSS sources. In total, 11,367,420 (39%) and 12,868,350 (44%) r2u sources are not detected in W1 and W2, respectively, and T-PHOT resulted in zero or negative fluxes for them. For the sake of completeness, these sources are still kept in the final catalogs and are assigned magnitude of -99.0 in W1 and/or W2.

\*\*\*\*\* These are numbers before masking, I am not sure what you mean by "cleaning" \*\*\*\*\*

### 3.5. Catalog post-processing

The merged photometric catalog from the above contains 29,046,660 entries, which would need to go through a few steps before it can be used further. In particular, our major goal is to produce a catalog of galaxies that is suitable for future applications. These steps are detailed below

#### 3.5.1. Removing duplicate sources

Adjacent SDSS images have 25" and 28" wide overlapping regions in RA and Dec, respectively. Similarly, adjacent unWISE images have 186" and 180" wide overlapping regions in RA and Dec, respectively. Sources in these regions appear in the catalog more than once and thus the duplications should be removed.

We performed an internal matching within the catalog, using a matching radius of 1.3". We found 1,423,979 duplications, which were subsequently removed from the catalog based on the SNR in the r-band (i.e. the source with the lower SNR is removed, while the source with the higher SNR is kept in the catalog).

#### 3.5.2. Masking around bright stars

Extremely bright objects severely contaminate their vicinities with various artifacts, such as bright halos, stripes of saturated pixels, diffraction spikes, ghost images, etc. As the sources found in such regions are often spurious and their photometry highly unreliable, we opted to remove them entirely from our catalog by mask-

ing out such regions. We used the SAO Bright Star Catalog (URL) and the Bright IR Stars Compilation (URL) as the basis of masking. There are around 400 (exact number) stars in both catalogs within Stripe 82, and our visual inspection of the T-PHOT residual images identified 300 (exact number?) additional objects (stars and very bright galaxies). We applied circular masks centered on these 705 sources, where the masking sizes depend on their V-band magnitudes if exist or our visual estimate of the halo sizes otherwise. The masking radius ranges from 50 to 600 arcsec. We removed the objects detected within the masked regions from our catalog. The masked area totals 3.866 deg<sup>2</sup>, which reduces the total sky area of the survey to 288.212 deg<sup>2</sup>.

#### 3.5.3. Star-Galaxy separation

Our catalog includes both galaxies and Galactic stars, the latter of which are not the subjects of our study and therefore should be flagged. There are two major ways to separate stars and galaxies, namely, morphological classification and photometric separation. The former distinguishes point-like sources (mostly stars, but also including QSOs) from extended ones, while the latter uses some certain color criteria to separate stars (and QSOs) from galaxies in the color space. We chose to use the morphological method as the primary criterion and the follow up cleaning will be done by the gentle photometric separation using r-i vs. z-w1 color diagram. This was again to utilize the SExtractor stellarity parameter CLASS\_STAR, which we used previously in selecting point sources for the SDSS PSF derivation (see §3.2). We ran SExtractor in the dual-image mode again, this time using the r2u images as the detection images and measuring the CLASS\_STAR parameter on the original g', r' and i' images. An accurate derivation of CLASS\_STAR needs an accurate input of the seeing FWHM value for each image, which we obtained based on the PSFs derived earlier. With the CLASS\_STAR values, we then used the spectroscopic sample of 254,190 sources in Stripe 82 classified by the SDSS team (Bolton et al. (2012)) to determine the optimal range of CLASS\_STAR. We found that it must simultaneously be less than 0.47, 0.43 and 0.75 in g', r' and i', respectively, to have the lowest fraction of missed galaxies (1.76%) and contamination from stars and QSO (2.39%).

After performing star-galaxy separation our catalog contains 14,418,304 galaxies and 13,423,475 point sources (stars and QSOs).

## 4. PHOTOMETRIC REDSHIFT ESTIMATION

A key goal of this work is to derive photometric redshifts ( $z_{ph}$ ) for the galaxies in our sample. To this end, we tested a number of softwares widely used in the literature. There are two broad categories of  $z_{ph}$  tools, one being based on SED fitting and the other being based on machine learning. In the SED fitting category, we tested HyperZ (citation), EAZY (citation) and LePhare (citation). In the machine learning category, we only used ANNz (citation), which takes a neural network approach. We compiled a large list of spectroscopic redshifts ( $z_{spec}$ ) in Stripe 82, and tested the performance of the different  $z_{ph}$  tools by comparing the derived  $z_{ph}$  and the available  $z_{spec}$ , quantified as  $\frac{|z_{ph} - z_{spec}|}{1 + z_{spec}}$ , and

also behavior of the sample around  $z \sim 0.4$  and around  $z \sim$ . The purpose of our test was to find the most appropriate tool(s) that would produce the smallest dispersion in terms of  $\Delta z/(1+z)$  and the least number of outliers. EAZY outperforms all other SED fitting codes and therefore we use it.

#### 4.1. Running EAZY

EAZY comes with several sets of SED templates, both theoretical and empirical, and they do not necessarily perform equally well with different data. After extensive testing, we chose the CWW + KIN set (??) because it performs the best according to the criteria outlined above. We show how CWW+KIN templates cover our data on a “color vs. spectroscopic redshift” plane using Fig. ???. Black points show g-r (left) and W1-W2 (right) colors for all sources from the spectroscopic sample with  $z_{spec} < 1.0$  and red lines show evolution of the of CWW+KIN templates’ colors with redshift.

We list the detailed parameter files that we used in our EAZY run.

##### 4.1.1. Redshift probability distribution

One is the redshift probability distribution prior. Due to the color-redshift degeneracy, it is not unusual that an SED fit by any  $z_{ph}$  tool could have multiple, widely separated peaks in its redshift probability distribution function (PDF), which would then create difficulty in assigning the most probable  $z_{ph}$ . Following Benítez (2000), EAZY deals with this problem by using a Bayesian prior of redshift probability distribution in the magnitude domain,  $P(z|m)$ , which essentially means that an object of magnitude  $m$  is unlikely to reside at a redshift that is too lower or too higher than  $z$  (see §2.2 of Brammer et al. 2008). We adopted the default prior used by EAZY, which is a function of  $R$ -band magnitude, and treated  $r$  as  $R$ .

##### 4.1.2. Magnitude zeropoint offset

The second is the magnitude zeropoint offset. The synthetic magnitudes ( $m_{syn}$ ) from the best-fit template can never exactly match the input magnitudes ( $m$ ), and this should not be a concern if the statistics shows that the difference ( $m_{syn} - m$ ) in a particular band is random. However, sometimes such a difference shows a systematic offset, which can be caused by a number of reasons, such as a systematic deviation of the templates with respect to the true SEDs, a slight error in the adopted system response curve, etc. To obtain the best results, one could manually correct this systematic difference by applying a zeropoint offset to the input magnitudes. Based on our test using the spectroscopic subset, we found that the offsets are -0.0749, 0.0193, 0.0019, 0.0074, -0.0351, 0.1574, 0.1460 AB in u, g, r, i, z, W1, and W2, respectively, for the EAZY run with all 7 bands and -0.0604, 0.0141, -0.0047, 0.0150, -0.0141 AB in u, g, r, i, z, respectively, for the EAZY run with only 5 SDSS bands. EAZY has the mechanism to apply such offsets, which we used in the fitting process. Our source catalogs, however, are not modified.

##### 4.1.3. EAZY with and without unWISE bands

The third point is about incorporating the W1/W2 photometry. For the EAZY runs, we did not find any significant offset of  $z_{ph}$  between the one including W1/W2 photometry and the one without (results from both EAZY runs are shown on Fig. 4. Left (right) panel shows  $z_{ph}$  vs  $z_{spec}$  for the EAZY run with (without W1/W2 bands). However, based on the spectroscopic subset, inclusion of W1/W2 photometry did improve the  $z_{ph}$  quality by reducing the standard deviation of  $\Delta z/(1+z_{spec})$  from 0.0296 to 0.0287 and reducing the number of outliers by 837 - from 8,099 to 7,262. Therefore, incorporating W1/W2 photometry is justified. In the catalog we opt to present both flavors of  $z_{ph}$  derived by EAZY - using all 7 bands and using only optical SDSS bands (denoted by the index *nowise*), but in the following discussion we only use 7 band catalog when refer to EAZY  $z_{ph}$ .

The inclusion of W1/W2 photometry does come with a price. Inspecting the results based on the spectroscopic subset, we found a small population of sources with  $z - W1 < -1.0$  mag and  $z < 19.9$  mag (Fig.5 left panel). We also found that the vast majority of such sources in the spectroscopic subset always had wrong  $z_{ph} \approx 0$  (Fig.5 right panel, green dots). After visually inspect the T-PHOT residual images, we concluded that these extreme colors are predominantly caused by the erroneous T-PHOT results (oversubtraction of the morphological templates). Therefore, we isolate such objects (143,810 or 1% of the catalog) and derive their  $z_{ph}$  without W1/W2 photometry, which improved the quality of the fit (Fig.5 right panel, blue dots).

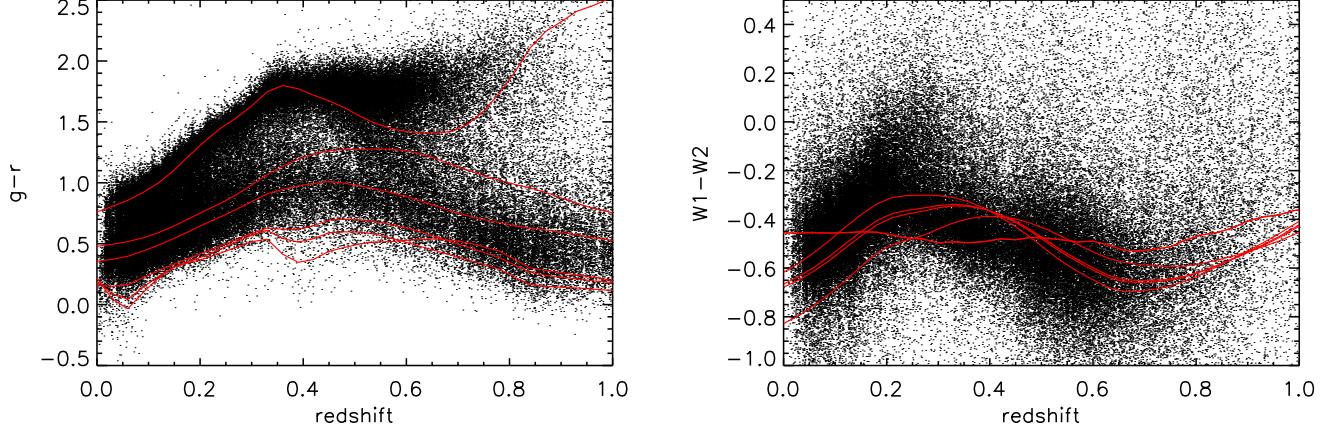
As can be seen on Fig.4 both EAZY (red) and EAZY<sub>nowise</sub> (blue) redshifts are in a good agreement with spectroscopic one up to  $z_{spec} \sim 0.85$  where the scatter becomes too large. We fix EAZY to fit photometric redshifts up to  $z_{max} = 1.0$  and discuss this upper limit in Sec. ?? in more details.

#### 4.2. Running ANNz

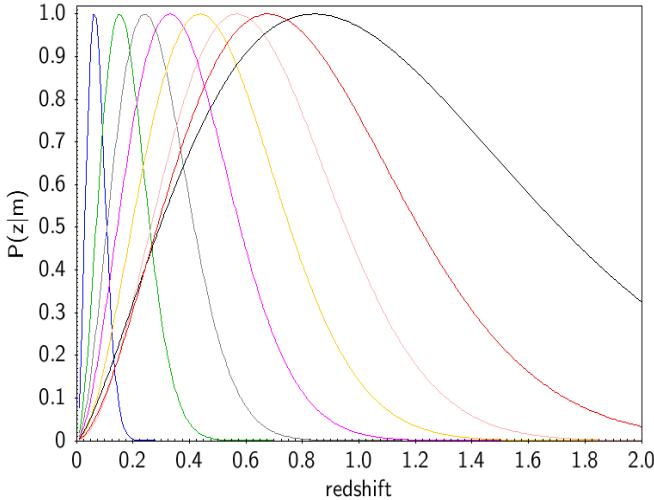
ANNz requires a sample of galaxies with known spectroscopic redshifts as the training set to build an empirical relation between measurable parameters and redshifts. While magnitude, surface brightness, color, size, morphology, etc. can all be used as an input node for ANNz, the most commonly used ones are magnitudes and colors, and these were what we adopted.

Quality of the set of sources used for training, validation and testing defines the performance of ANNz. Out of 150,767 sources from the spectroscopic sample we chose 148,567 with the redshift within  $0 < z_{spec} < 1.5$  (number density of the sources drops sharply after  $z_{spec} < 1.5$  and our set of bands cannot prevent the degeneracy in the redshift determination) and call such subset an ANNz spectroscopic sample. While auxiliary deep surveys (6df, DEEP2, WiggleZ, VIPERS and VVDS) compose only 19.5% of the total, they are constitute an key portion of the ANNz spectroscopic sample by providing important information about the faint (and thus presumably high-redshift) sources and also filling in the magnitude space in different regions than SDSS spectroscopic sample thus helping to more efficiently train ANNz.

Following the instructions of ANNz, we randomly divided this ANNz spectroscopic sample into three sets of roughly equal number of objects, namely, the training set



**Figure 2.** Color vs. spectroscopic redshift for the whole set of the photometric catalog sources (colored black) with available spectra. Colors of the six CWW+KIN templates as a function of redshift templates are plotted as red lines. These templates alone cover sufficient area on the plane and their linear combination returns reliable photometric redshift.



**Figure 3.** R-band redshift probability distribution prior used in this work for a range of magnitudes from AB 18 (blue) to AB 25 (black) with a step of 1 magnitude.

to train ANNz on our data, the validation set to validate the  $z_{ph}$  derivation, and the testing set to find the best choice of the input nodes. We tried following options: all seven magnitudes, only optical magnitudes, seven non-redundant colors, combination of seven magnitudes and seven colors, seven magnitudes and three optical colors (except for the  $u - r$ ) and seven colors with  $gri$  magnitudes. After extensive testing, we found that seven magnitudes gives the best result and use it as the input node.

Architecture of the ANNz network can be described as  $N_{input} : N_1 : N_2 : \dots : N_m : N_{output}$ , where  $N_{input} = 7$  is the number of input magnitudes for each object in our sample,  $N_1, N_2, \dots$ , and  $N_m$  are the number of nodes in each ANNz hidden layer for a total of  $m$  layers, and  $N_{output}$  is the number of sets for output photometric redshift, usually set to be 1. We again randomly divided the spectroscopic catalog into three subsamples and tested ANNz for the best combination of the nodes, the hidden layers and the number of iterations. We found that two layers with seven nodes in each layer (7:10:10:1 architec-

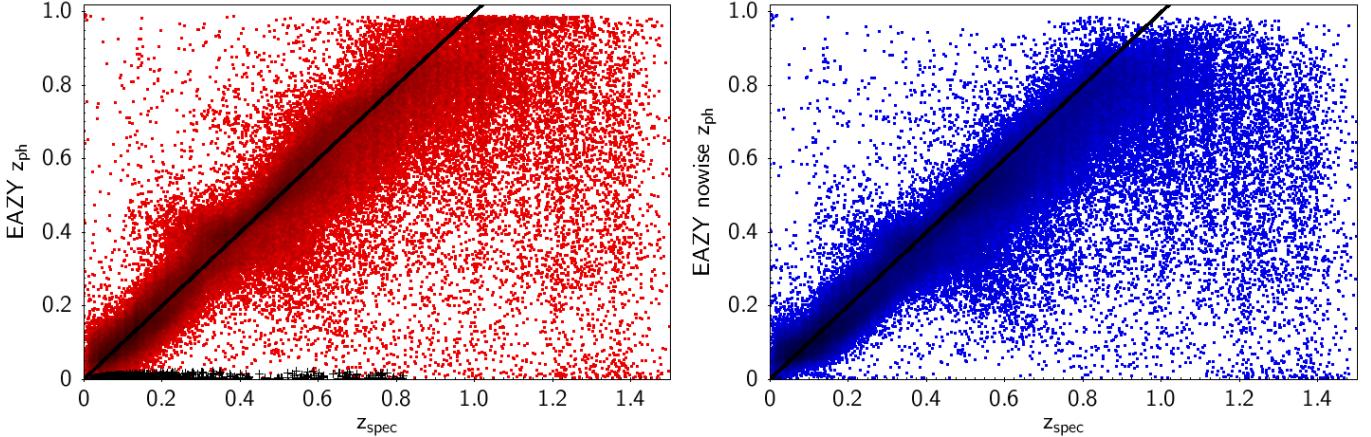
ture) with 700 iterations would perform the best with our data.

Depending on the particular seed for the random number generator, the number of iterations and the fraction of the spectroscopic sample used for the weights, the training process usually converges to different redshift local minima. However, it is not necessary to select one best network - in fact, ANNz can use the sub-optimal networks to improve overall accuracy: the mean of the individual outputs from a group of networks usually gives a better estimate for the photometric redshift than the outputs of any network alone. In order to get the full advantage of this feature and also to get a comprehensive validation of the ANNz performance (ANNz performance should only be tested on a subsamples that is not used for either training or validation) we used the following strategy to train ANNz:

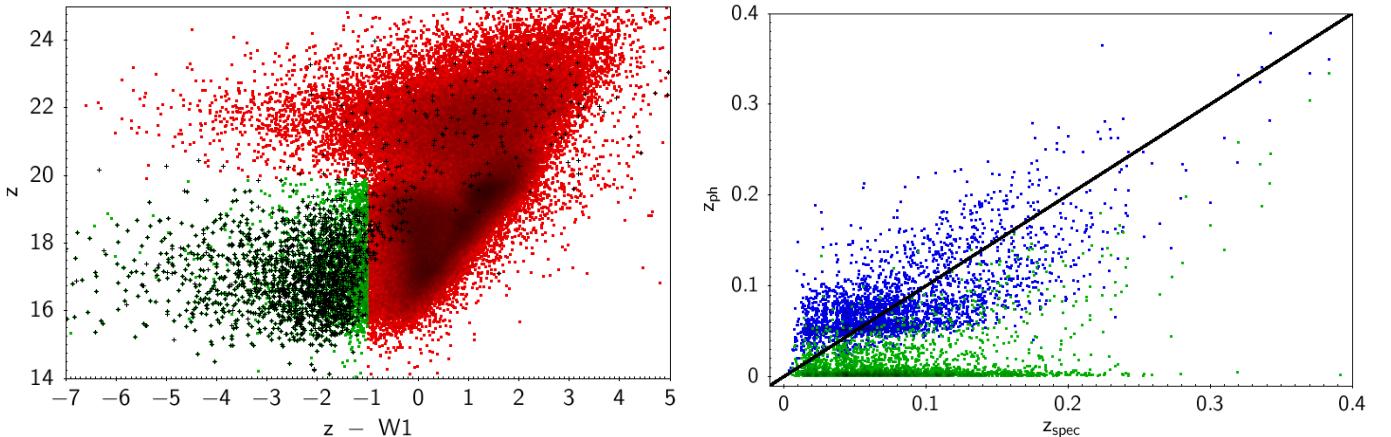
1. ANNz spectroscopic sample was shuffled in a random order
2. The catalog was divided into 3 sets: training set (70% of sources), validation set (10%) and testing set (20%).
3. With this set of data ANNz was trained 2 times using different seed for the random number generator and 2 weights files .wts are created.
4. Steps 2. and 3. were repeated 4 more times in such a way that each of the five testing sets uses different part of the ANNz spectroscopic sample and thus the whole sample can be independently tested.
5. ANNz was ran with 10 weight files on an ANNz spectroscopic sample and then 2,091 outliers identified as sources with  $\frac{|\Delta z|}{1 + z_{spec}} < 0.15$  were rejected.

6. Steps 2-4 were repeated for the ANNz spectroscopic sample without outliers and ANNz was ran with 10 new weight files on our Stripe 82 photometric catalog.

The performance of ANNz must be tested on a subset of sources that has not been used for the training, therefore it is not straightforward to test the quality of  $z_{ph}$  for the whole subsample of spectroscopic sources. In our approach the sample of sources with known  $z_{spec}$  is divided 5 times into training, validation and test subsets in such a way that 5 test subsets comprise a full ANNz spectro-



**Figure 4.**  $z_{ph}$  vs  $z_{spec}$  for the subsample of sources with spectroscopic redshifts from the training sample. Left panel shows results for all 7 bands used in EAZY run and right panel - with only SDSS bands. Black crosses with  $z_{ph} \sim 0$  are sources with erroneous fluxes in W1/W2. They are refitted by EAZY using only optical fluxes (see Fig.5). Black line shows 1:1 relation.



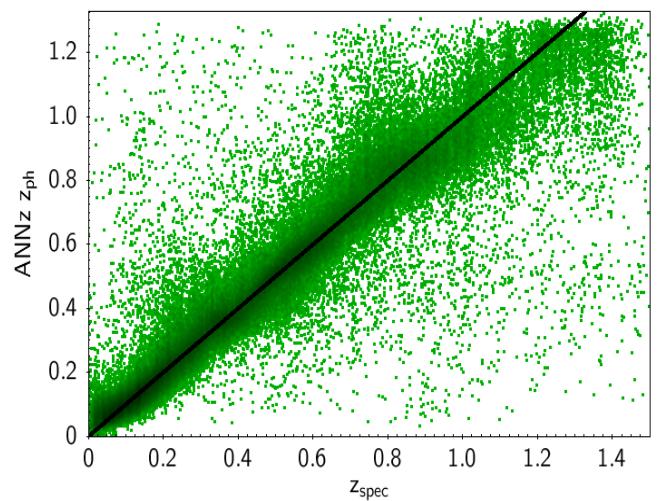
**Figure 5.** Left: sources with known  $z_{spec}$  are plotted on a color-magnitude plane. SED of the sources labeled with black crosses are fitted poorly by EAZY when photometry in W1 and W2 bands is included. In green color we show the region defined as  $z - W1 < -1.0$  mag and  $z < 19.9$  mag inside which sources will be fitted to EAZY using only optical fluxes. Right panel shows such poor fit on a  $z_{ph}$  vs  $z_{spec}$  plane (green) and this fit is improved when sources are fitted only with SDSS bands (blue). Black line shows 1:1 relation.

scopic sample. Then we run ANNz on each of these 5 test sets using two weight files per set that were derived using sources from the corresponding training and validation sets. In the end we have a full set of ANNz spectroscopic sample tested by ANNz in a correct way that is plotted on Fig.6. Standard deviation of the sources within  $\frac{|\Delta z|}{1 + z_{spec}} < 0.15$  (i.e. not outliers) is 0.0241 - better than for any EAZY run.

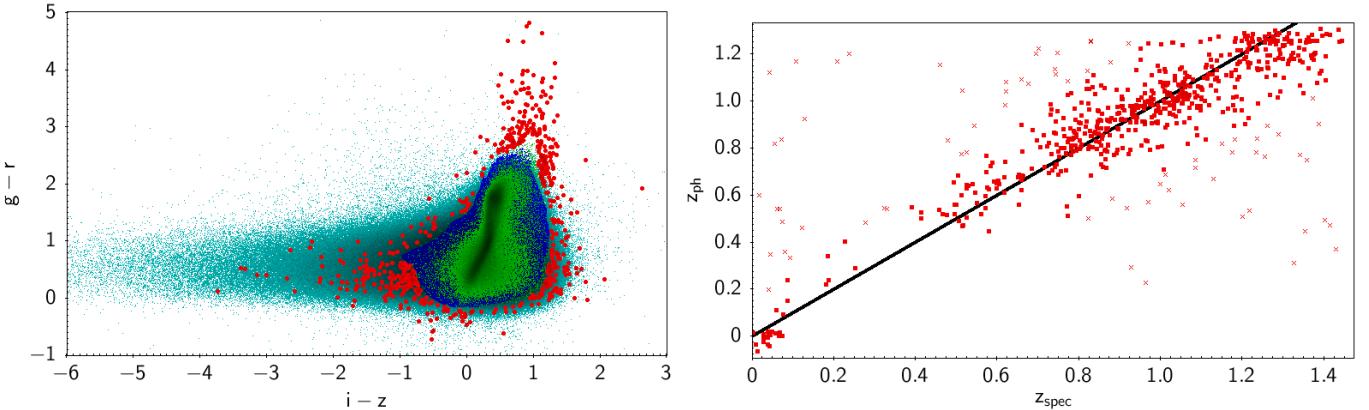
The only difference with the science run of ANNz is that our ANNz spectroscopic sample was tested using 2 corresponding weight files per set, while for the photometric catalog we used all 10 available .wts files, so reported standard deviation should be treated as the upper limit and expected the performance of ANNz on a full set of data is expected to be better than that.

#### 4.2.1. Beyond the training set

A limitation of machine learning is that the results are reliable only when the training set covers the full parameter space of the input nodes. Our spectroscopic sample only covers the brighter part of the catalog, and therefore only covers part of the parameter space of the



**Figure 6.** ANNz  $z_{ph}$  vs  $z_{spec}$  for the subsample of sources with spectroscopic redshifts from the training sample. Standard deviation of the sources that are not outliers is 0.0241. Black line shows 1:1 relation.



**Figure 7.** Left: Color-color plane of all extragalactic sources with  $S/N > 5$  in bands  $gri$  (blue), training set for ANNz (green) and the subsample of the sources from the training set that poorly cover the color-color plane (red). Right:  $z_{ph}$  vs  $z_{spec}$  for the subsample of 639 sources from the ANNz spectroscopic sample that poorly cover the color-color plane. Red dots - sources within  $\frac{\Delta z}{1 + z_{spec}} < 0.15$  with standard deviation 0.0353, and red crosses - outliers.

full catalog, regardless of what input values we use. It is impossible to estimate the performance of ANNz in the regions that are completely uncovered by the training set, but it is important to understand what it would be in the regions that are still populated by the sources from the training set, though sparsely.

For this purpose, we carry out a test by projecting our color space onto the  $g - r$  versus  $i - z$  color plane, which is shown in the left panel of Fig. 7. We plot all the sources whose  $z_{ph}$  are to be derived (i.e., those with  $S/N \geq 5$  in the  $gri$  bands) as the cyan dots, and roughly divide the region that they occupy into four parts according to the density of the objects from the spectroscopic sample: densely populated, intermediately populated, sparsely populated, and not-populated. To outline these regions, the objects from the spectroscopic sample in the first three regions are shown in green, blue, and red, respectively. We test the ANNz performance in the sparsely populated region (red), which includes 481,293 cyan points out of the total of 10,403,778. There are 639  $z_{spec}$  in this region (the red dots in the left panel of Fig. 7), and we compare them to the corresponding ANNz  $z_{ph}$  values in the right panel of Fig. 7. Only 75 sources are outliers based on our criterion of  $\Delta z/(1 + z_{spec}) > 0.15$ , which are shown as the red crosses. The rms value for the rest is 0.0353. For comparison, the rms value for the full ANNz test is 0.0241 after discarding the outliers (see §XXX above). Therefore, the ANNz performance in the sparsely populated region is still reasonable.

While we cannot evaluate the performance in the “not populated” region, we conclude that this is not a serious concern because the objects in this region are of a small fraction of the total. In this particular case, they constitute only 4.6% of the full catalog. The vast majority of them are faint sources, most of which have high  $z_{ph}$  and do not enter the final  $z_{ph}$  catalog because of the redshift limit that we impose to this catalog (see §5.3 below).

## 5. DISCUSSION

In this section, we further examine our  $z_{ph}$ , and compare them to other photometric redshift surveys in this field.

### 5.1. Cleaning of $z_{ph}$ Catalog

We derived  $z_{ph}$  for all the objects in our extragalactic catalog, however some objects should not be included in the final  $z_{ph}$  catalog for obvious reasons. For the objects that have photometry in less than three bands, their EAZY  $z_{ph}$  are deemed to be unreliable and are assigned  $z_{ph} = -99$ . There are only 2,946 (11,101) such objects when W1/W2 are (not) included.

Some objects from the ANNz run should also be discarded, but for a different reason. If a source is outside of the region in the input parameter space where the training set populates, the ANNz calculates  $z_{ph}$  by extrapolation. As the ANNz does not care about the physical meaning of any input or output parameters, sometimes the extrapolation could result in negative  $z_{ph}$ . There are 163 such objects with  $z_{ph}$ , and as it turns out, 27 of them have  $z_{spec}$ , which range from 0 to 0.045 while their  $z_{ph}$  goes from -0.014 to 0. The maximum photometric redshift error for these 27 sources is 0.085, mean photometric redshift error is 0.0374 and standard deviation is only 0.0215 - well below the value for the full photometric catalog.

As our photometric catalog is based on the r-band detections with  $S/N \geq 5$ , it contains many sources that are weaker or even invisible in other bands and their  $z_{ph}$  could be unreliable. Therefore, we only focus on the objects that have  $S/N \geq 5$  in at least the  $gri$  bands. There are 10,403,778 sources that meet this requirement, which constitute 72% of the total extragalactic catalog. Our discussion below will be based on this subsample.

### 5.2. Comparison between ANNz and EAZY photometric redshifts

In this section we compare  $z_{ph}$  derived with ANNz and EAZY, discuss how do they differ and what could be the reasons for it.

Both algorithms, SED fitting and machine learning, rely their photometric redshift estimation on a detection of characteristic features in a galaxy spectrum. Observed wavelength of this characteristic feature unambiguously defines the redshift of a source under question. Such features usually manifest themselves as a sharp drop in magnitudes - Ly limit,  $Ly\alpha$  break, PAH features, etc.

For our range of bands and expected range in the redshifts such a feature is a 4000Å/Balmer break that is caused by the absorption by hydrogen atoms of photons with energy above the Balmer limit (3,646Å) and also by the combination of the absorption features by ionized metals in the stellar atmospheres.

Because of the absence of other prominent features, relative location of a 4000Å/Balmer break explains almost every non-linear behavior of the  $z_{ph}$  vs  $z_{spec}$ .

At  $z \sim 0$  this break falls into the  $u$  band, one of the two most shallow SDSS bands, and causes the kink in the range  $0 < z < 0.1$ . This kink is the most prominent in the EAZY\_nowise sample, in which no information from the near-IR bands can be used to break this degeneracy. As local galaxies are not usually bright in the near-IR, ANNz machine learning algorithms also do not get much information that can be used to refine the fit. Surprisingly, adding W1/W2 bands to the EAZY helps to improve the fit - though CWW+KIN templates that we adopt here do not contain any near-IR features, steady decrease in the relative flux from  $\sim 1.2\mu$  in all templates helps to constrain the  $z_{ph}$  even in the case of no detection (what is exactly expected for the sources at  $z \sim 0$ ).

A well-known scatter around  $z \sim 0.4$  is caused by the 4000Å/Balmer break that falls in the dip between  $g$  and  $r$  bands - the construction of the instrument is such that this dip is the largest among all four dips in the optical.

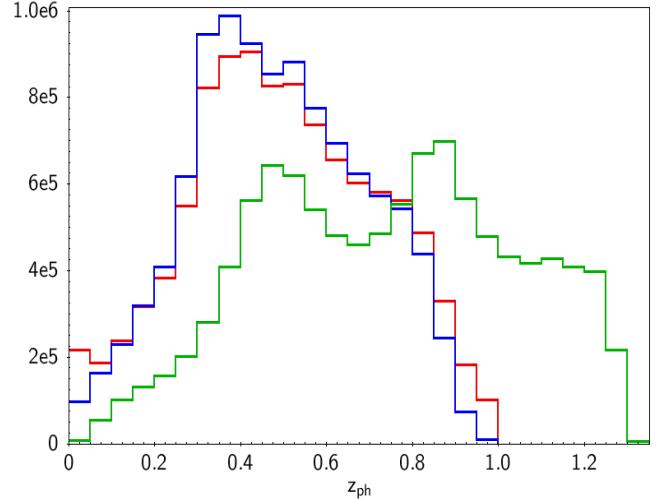
This break also defines the upper limit of the EAZY  $z_{ph}$  - at  $z \sim 0.9$  it falls between  $i$  band and  $z$  band which is the most shallow among SDSS, and the scatter of the  $z_{ph} - z_{spec}$  immediately increases. But also the SDSS depth starts to become important as the  $S/N$  ratio of high-z sources gets much worse.

This scatter is however not due to  $z_{max}$  that we assigned to EAZY, which sets the upper limit to the allowed photometric redshift - we tested larger values to confirm that it does not put high-z galaxies closer to the true  $z_{ph}$ , but on contrary - it assigns low-z galaxies with unrealistic  $z_{ph} \sim 1$  thus increasing the overall scatter. Thus our choice of  $z_{max} = 1$  as a reasonable limit for the EAZY is justified.

While ANNz behaves considerably better in this redshift range and (due to the inclusion of the deep surveys from the ANNz spectroscopic training sample) have a much tighter  $z_{ph}$  vs  $z_{spec}$  relation as compared to the EAZY, at  $z_{spec} \sim 1.1$  machine learning algorithm starts to fail and underpredicts  $z_{ph}$  (Fig.6) - result of the low number of the galaxies in the training set and an overall decrease of the  $S/N$  in all 7 input nodes.

### 5.3. Contamination with high-z sources

Considering similar behavior on the  $z_{ph}$  vs  $z_{spec}$  plots one could expect also the similar shape of the redshift distribution for all three flavors of derived  $z_{ph}$  for the full sample. But as one may see on Fig. 8, these distributions are very different - both EAZY  $z_{ph}$  have a sharp single peak around  $z \sim 0.4$  and a steady decrease in the number density up to the assigned  $z_{max} = 1$ , while ANNz produces a double peak distribution with the first one at a similar redshift as the EAZY one and the second, even larger peak being at  $z \sim 0.9$ . While this second peak is not fully nonphysical - we estimated stellar masses in



**Figure 8.** Photometric redshift distribution. Red - EAZY, blue - EAZY\_nowise, green - ANNz.

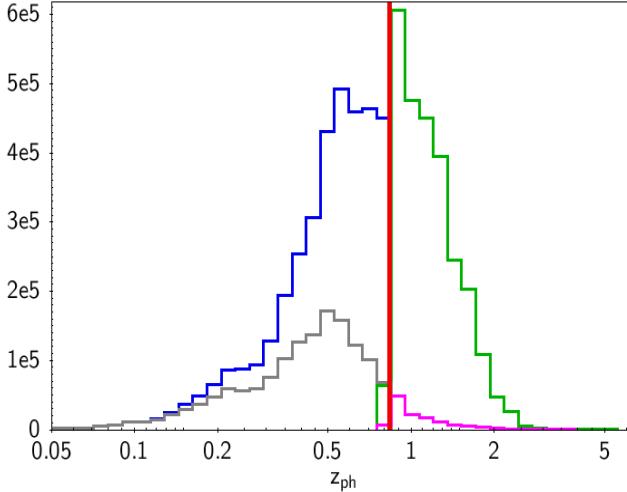
the equal comoving volume bins of  $20 \text{ Mpc}^3$  and verified that there is no stellar mass decrease - this drastic difference requires explanation. We anticipate that both algorithms and all 3 flavors of  $z_{ph}$  produce similar results *in the redshift range of our interest*, i.e.  $0 < z < 0.85$  and behaves differently at higher redshifts.

Stacked Stripe 82 images should contain a lot of high-z sources in contrast to the single-epoch SDSS data. We tested this hypothesis by comparing the number of low- and high-z sources that would be detected in a stacked Stripe 82 image and in a single-epoch SDSS image, using CFHTLS data. Subsamples were constructed in the following way: the galaxy (StarGal=0) that has  $S/N > 5$  in  $gri$  bands, and  $0 < phot\_z < 9$  is visible in Stripe 82 (SDSS) if it is  $r < 24.8$  (22.7) AB. It is a low-z sources if  $photo\_z \leq 0.834$  and a high-z source if  $photo\_z > 0.834$ . We chose  $z = 0.834$  as this is the redshift of an edge of a sphere with comoving volume of  $100 \text{ Mpc}^3$  that will shall bin into  $20 \text{ Mpc}^3$  slices to study evolution of the galaxy population.

Fig.9 shows a redshift distributions for the low-z Stripe 82 (blue), SDSS (grey) and for the high-z Stripe82 (green) and SDSS (magenta) sources. Vertical red lines defines our choice of separating high-z from the low-z sources. The drop in the number of sources is very sharp for SDSS - there are 11.5 times more low-z sources as compared to the high-z ones (1,434,791 vs. 124,905). The situation with the Stripe82 sources is quite different - there are only 1.4 times more low-z sources as compared to the high-z ones (3,684,728 vs. 2,629,236).

One more hint on a large fraction of the high-z sources in our data comes from the fact that 32.4% of the high-z sources from the spectroscopic sample are cross-matched with our photometric catalog with the matching radius of  $2''$ . For the set of filter used in this project both EAZY and ANNz cannot treat high-z sources properly and assign them correct redshift, therefore such sources should be removed from the catalog. Due to the high fraction of faint sources that populate large area on a color-color plane, such high-z sources cannot be removed based on their colors.

We propose to remove high-z fraction from our sample using the difference in the treatment of such sources by



**Figure 9.** Histogram shows the redshift distribution of the CFHTLS data for the low-z (blue) and high-z (green) sources detected in Stripe82 images as compared to the low-z (grey) and high-z (magenta) sources in SDSS single-epoch images. Red vertical line at  $z=0.834$  separates low-z from high-z galaxies. Stacked images are  $\sim 2$  magnitudes deeper and contain a lot of high-redshift sources that contaminate our sample and should be removed from the catalog

EAZY and ANNz - as we discussed above, while EAZY scatters high-z sources all over the allowed redshift range (this is fortified by the use of the redshift probability distribution introduced in Sec.4.1.1), while ANNz tends to assign such high-z sources with the maximum redshift for which it has sufficient number of galaxies from the training set ( $z \sim 1.1$ ). On a Fig.10 we plot EAZY  $z_{ph}$  - ANNz  $z_{ph}$  and draw a horizontal line at -0.4 - sources that fall below this line are considered as high-z with bad fit and are rejected from the catalog. This further reduces the number of sources in the catalog to 8,214,318 (blue dots on Fig. 10) and rejects 2,189,460 (green dots). Majority of the rejected sources are high-z sources and fall beyond our redshift of interest (denoted with the red vertical line).

I rewrote till this phrase

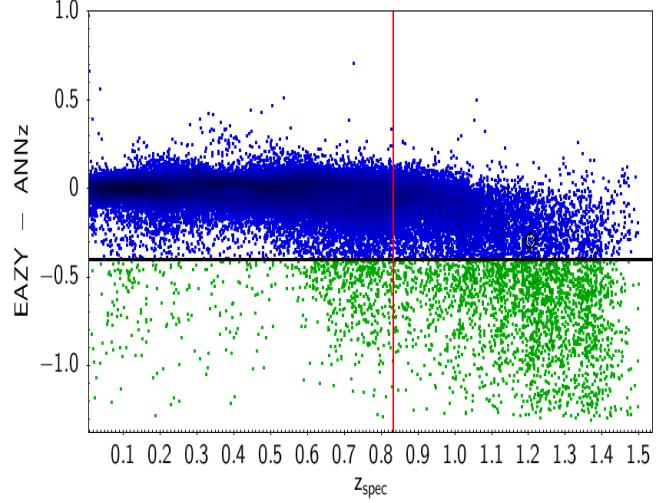
-Different photoz flavours treat these high-z source differently: EAZY full and EAZY\_nowise distribute them more or less homogeneously with a peak at  $\sim 0.75$  while ANNz puts all of the high-z sources at  $z \sim 0.89$  and this is a good way of removing them!

This behavior is significantly different to the EAZY one, that just scatters high-z sources all over the  $z_{ph}$  vs  $z_{spec}$  plane - we shall use this in the §5.3 when attempt to remove high-redshift sources from our sample.

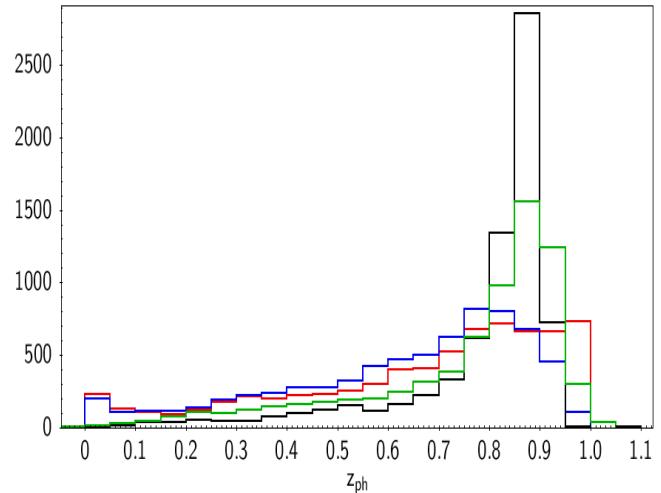
Reference to the subsection: §5.3

#### 5.4. Comparison with existing photometric redshifts

Testing our results against available surveys with photometric redshifts is not a direct way to test the quality of our results as errors are on both sides, but such surveys has much more sources as compared to the available spectroscopic samples and provide important statistical insights. We compared all three versions of photometric redshifts derived in this project to photometric redshifts in SDSS DR14 (?), photometric redshifts are updated in



**Figure 10.** Difference of the  $z_{ph}$  produced by EAZY and ANNz is used as a criteria for rejection of the high-z sources. Blue points lay above  $EAZY z_{ph} - ANNz z_{ph} = -0.4$  show similar behavior with both codes and are left in the catalog, while green dots - the one that lay below the black line are rejected - most of them are high-z sources that are poorly fitted by either EAZY, ANNz or both algorithms. Red line at  $z=0.834$  defines the redshift limit of our interest - our catalog is neither secure in terms of photometric redshift estimation, nor complete even at  $M_*$ .



**Figure 11.** OLD DATA! Photometric redshift distribution of the sources with  $z_{spec} > 1$ . Red - EAZY, blue - EAZY\_nowise, grey - ANNz\_full, green - ANNz\_sdss  
Both photometric redshifts produced by ANNz have sharper peaks and this can be used to eliminate such high redshift sources from our sample.

DR12 and are unchanged in DR13 and DR14, we refer to it as SDSS DR14); redshifts, derived in ? (we referred to it as Reis) using artificial neural network technique, redshifts based on imaging data of the SCUSS, SDSS and WISE surveys and presented in ? (we refer to it as SCUSS+SDSS+WISE) and Canada-France-Hawaii Telescope Legacy Survey “Wide” survey from data release T0007 (referred to as CFHTLS). Several relations between photometric redshifts are plotted on Fig.12. Table 2 contains information about number of matched sources and median redshift of the surveys (all sources in supplementary catalogs were matched to sources with  $S/N > 5$  in  $gri$  bands using a  $2''$  matching radius) and Table

**Table 2**  
Matched photometric redshift surveys

Survey	Number of matched sources	Median redshift
<b>SDSS DR15</b>	12 hours	
<b>Reis</b>	8,312,976	
<b>SCUSS+SDSS+WISE</b>	12 hours	
<b>CFHTLS</b>	infinite	

3 shows standard deviation for all pairs of photometric redshifts.

- we discuss, how this can influence the GSMD estimation.

- redshift bins: 20 Gpc<sup>3</sup> using standard cosmology; number of galaxies per bin; last 5th bin is not used in estimation of the GSMD, but to reject all high-z galaxies

- completeness of the catalog - as seen from the distribution, catalogs are incomplete after z=0.45. Thus Stripe 82 does not produce complete catalog at significantly deeper redshifts as was anticipated, but still has much more data at higher redshifts.

SCUSS, SDSS DR15, Reis et al, CFHTLS

### 5.5. Outliers

Check my notes p8. and section 4.8 of Dahlen et al 2013

Majority of outliers  $(z_{ph} - z_{spec})/(1 + z_{spec}) > 3\sigma$  are different for different codes, 1250 of them are outliers in all 4 codes and also outliers in SCUSS, Reis, CFHTLS and SDSS DR15 (which are photo-z catalogs). 460 of them have zWarning=4 from SDSS DR14 (chi-squared of best fit is too close to that of second best), so it is not necessary that our photo-z are wrong. So it is possible that the real fraction of outliers is even smaller than reported here.

## 6. GALACTIC CATALOG

That many stars based on CLASS\_STAR in gri and color-color diagram. all have SNR > 5 in r-band.

## 7. SUMMARY

### REFERENCES

- Adelman-McCarthy, J., M.A., A., S.S., A., et al. 2007, VizieR Online Data Catalog, 2276  
 Annis, J., Soares-Santos, M., Strauss, M. A., et al. 2014, ApJ, 794, 120  
 Benítez, N. 2000, ApJ, 536, 571  
 Bertin, E., & Arnouts, S. 1996, A&AS, 117, 393  
 Bertin, E., Mellier, Y., Radovich, M., et al. 2002, in Astronomical Society of the Pacific Conference Series, Vol. 281, Astronomical Data Analysis Software and Systems XI, ed. D. A. Bohlender, D. Durand, & T. H. Handley, 228  
 Bolton, A. S., Schlegel, D. J., Aubourg, É., et al. 2012, AJ, 144, 144  
 Brammer, G. B., van Dokkum, P. G., & Coppi, P. 2008, ApJ, 686, 1503  
 Cutri, R. M., Wright, E. L., Conrow, T., et al. 2013, Explanatory Supplement to the AllWISE Data Release Products, Tech. rep.  
 Davis, M., Faber, S. M., Newman, J., et al. 2003, in Proc. SPIE, Vol. 4834, Discoveries and Research Prospects from 6- to 10-Meter-Class Telescopes II, ed. P. Guhathakurta, 161–172  
 Drinkwater, M. J., Jurek, R. J., Blake, C., et al. 2010, MNRAS, 401, 1429  
 Garilli, B., Le Fèvre, O., Guzzo, L., et al. 2008, A&A, 486, 683  
 Guzzo, L., Scodéglio, M., Garilli, B., et al. 2014, A&A, 566, A108  
 Huff, E. M., Hirata, C. M., Mandelbaum, R., et al. 2014, MNRAS, 440, 1296  
 Jiang, L., Fan, X., Bian, F., et al. 2009, AJ, 138, 305  
 —. 2014, ApJS, 213, 12  
 Jones, D. H., Saunders, W., Colless, M., et al. 2004, MNRAS, 355, 747  
 Jones, D. H., Read, M. A., Saunders, W., et al. 2009, MNRAS, 399, 683  
 Laidler, V. G., Papovich, C., Grogin, N. A., et al. 2007, PASP, 119, 1325  
 Lang, D. 2014, AJ, 147, 108  
 Le Fèvre, O., Cassata, P., Cucciati, O., et al. 2013, A&A, 559, A14  
 Lucy, L. B. 1974, AJ, 79, 745  
 Mainzer, A., Bauer, J., Cutri, R., et al. 2014, in Lunar and Planetary Inst. Technical Report, Vol. 45, Lunar and Planetary Science Conference, 2724  
 Merlin, E., Fontana, A., Ferguson, H. C., et al. 2015, Astronomy and Astrophysics, 582, A15  
 Mink, D. J. 1998, in Bulletin of the American Astronomical Society, Vol. 30, AAS/Division of Dynamical Astronomy Meeting, 1144  
 Newman, J. A., Cooper, M. C., Davis, M., et al. 2013, ApJS, 208, 5  
 Parkinson, D., Riener-Sørensen, S., Blake, C., et al. 2012, Phys. Rev. D, 86, 103518  
 Richardson, W. H. 1972, Journal of the Optical Society of America (1917-1983), 62, 55  
 Scodéglio, M., Guzzo, L., Garilli, B., et al. 2018, A&A, 609, A84  
 Strauss, M. A., Weinberg, D. H., Lupton, R. H., et al. 2002, AJ, 124, 1810  
 Wright, E. L., Eisenhardt, P. R. M., Mainzer, A. K., et al. 2010, AJ, 140, 1868

**Table 3**  
 Standard deviation of  $\frac{\Delta z}{1+z}$  for pairs of photometric redshifts.

	EAZY_nowise	ANNz	SDSS DR14	Reis	SCUSS + SDSS + WISE	CFHTLS
<b>EAZY</b>	0.0667	0.1199	0.958			
<b>EAZY_nowise</b>		0.0953	0.0686			
<b>ANNz</b>			0.0697			

<sup>1</sup> Spectroscopic sample is limited to  $z_{spec} < 1$  for EAZY and to  $z_{spec} < 1.3$  for ANNz.

## APPENDIX

### *EAZY parameters*

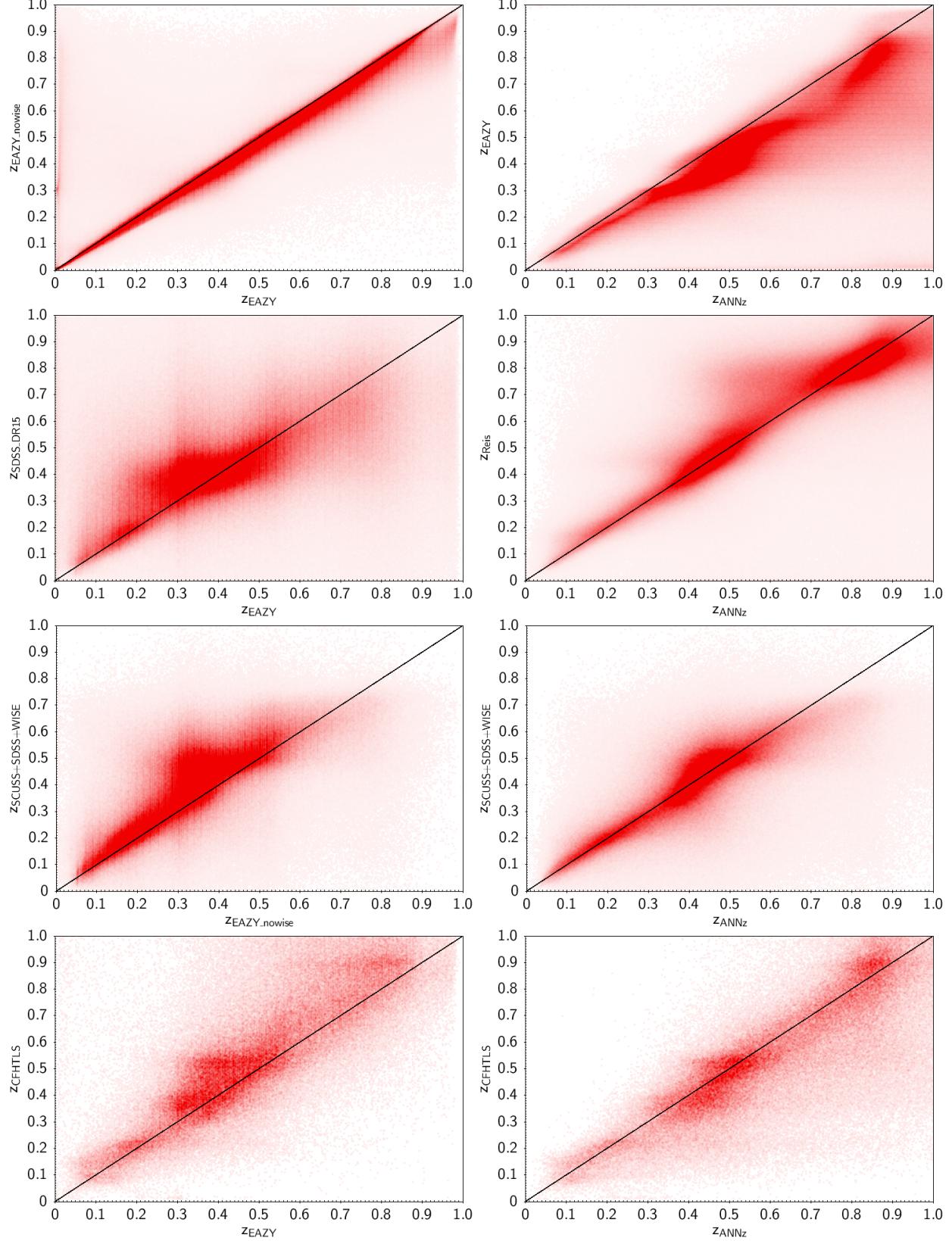
```

TEMPLATES_FILE CWW_KIN.spectra.param # Template definition file
TEMPLATE_COMBO a # Template combination option
APPLY_IGM y # Apply Madau 1995 IGM absorption
N_MIN_COLORS 3 # Require N_MIN_COLORS to fit
APPLY_PRIOR y # Apply apparent magnitude prior
PRIOR_FILE templates/prior_K_extend.dat # File containing prior grid
Z_MIN 0.01 # Minimum redshift
Z_MAX 1.0 # Maximum redshift
Z_STEP 0.005 # Redshift step size
Z_STEP_TYPE 1 # 0 = Z_STEP, 1 = Z_STEP*(1+z)
GET_ZP_OFFSETS y # Look for zphot.zeropoint file and compute zeropoint offsets

```

## ACKNOWLEDGMENTS

This publication makes use of data products from the Wide-field Infrared Survey Explorer, which is a joint project of the University of California, Los Angeles, and the Jet Propulsion Laboratory/California Institute of Technology, funded by the National Aeronautics and Space Administration. Part of our data processing and analysis were done using the HPC resources at the University of Missouri Bioinformatics Consortium (UMBC).



**Figure 12.** Comparison of 3 different versions of  $z_{ph}$  to each other and available surveys with known photometric redshifts: SDSS DR15, Reis, SCUSS and CFHTLS