

Photometric redshifts for X-ray-selected active galactic nuclei in the eROSITA era

M. Brescia¹,^{1★} M. Salvato²,² S. Cavuoti³,^{1,3,4} T. T. Ananna,^{5,6} G. Riccio¹,¹
S. M. LaMassa,⁷ C. M. Urry⁶ and G. Longo^{3,4}

¹INAF – Astronomical Observatory of Capodimonte, via Moiariello 16, I-80131, Napoli, Italy

²Max-Planck-Institut für Extraterrestrische Physik, Postfach 1312, Giessenbachstraße, D-85741 Garching, Germany

³Department of Physics ‘E. Pancini’, University Federico II, via Cinthia 6, I-80126 Napoli, Italy

⁴INFN section of Naples, via Cinthia 6, I-80126 Napoli, Italy

⁵Department of Physics, Yale University, PO Box 201820, New Haven, CT 06520-8120, USA

⁶Yale Center for Astronomy and Astrophysics, PO Box 208121, New Haven, CT 06520, USA

⁷Space Telescope Science Institute, 3700 San Martin Drive, Baltimore, MD 21218, USA

Accepted 2019 July 27. Received 2019 July 26; in original form 2019 May 6

ABSTRACT

With the launch of eROSITA (extended Roentgen Survey with an Imaging Telescope Array), successfully occurred on 2019 July 13, we are facing the challenge of computing reliable photometric redshifts for 3 million of active galactic nuclei (AGNs) over the entire sky, having available only patchy and inhomogeneous ancillary data. While we have a good understanding of the photo-*z* quality obtainable for AGN using spectral energy distribution (SED)-fitting technique, we tested the capability of machine learning (ML), usually reliable in computing photo-*z* for QSO in wide and shallow areas with rich spectroscopic samples. Using MLPQNA as example of ML, we computed photo-*z* for the X-ray-selected sources in Stripe 82X, using the publicly available photometric and spectroscopic catalogues. Stripe 82X is at least as deep as eROSITA will be and wide enough to include also rare and bright AGNs. In addition, the availability of ancillary data mimics what can be available in the whole sky. We found that when optical, and near- and mid-infrared data are available, ML and SED fitting perform comparably well in terms of overall accuracy, realistic redshift probability density functions, and fraction of outliers, although they are not the same for the two methods. The results could further improve if the photometry available is accurate and including morphological information. Assuming that we can gather sufficient spectroscopy to build a representative training sample, with the current photometry coverage we can obtain reliable photo-*z* for a large fraction of sources in the Southern hemisphere well before the spectroscopic follow-up, thus timely enabling the eROSITA science return. The photo-*z* catalogue is released here.

Key words: methods: data analysis – methods: statistical – galaxies: active – galaxies: distances and redshifts – X-rays: galaxies.

1 INTRODUCTION

Photometric redshifts (photo-*z*) are now routinely used in many applications, from galaxy evolution to cosmological studies. In particular, the present and planned photometric surveys over many thousand square degrees (e.g. DES, *Euclid*, LSST, eROSITA, and SphErEx) rely mostly on photo-*z* for their scientific exploitation. As known, there are basically two classes of methods commonly used to derive photo-*z*: the template spectral energy distribution (SED) fitting methods (e.g. Bolzonella, Miralles & Pelló 2000; Ilbert et al.

2006; Tanaka 2015) and the empirical or interpolative methods (e.g. Cavuoti et al. 2015b; Carrasco Kind & Brunner 2013). Both methods are characterized by advantages and shortcomings (for a complete review, see Salvato, Ilbert & Hoyle 2018), but essentially both rely on colour/magnitude–redshift maps, with the difference that SED based methods assume a priori knowledge of the map, while empirical methods, based on machine learning (ML), learn the map anew from the data every time. Recently more algorithms that merge the pros of the two techniques are developed and show very promising results also in computing photo-*z* for mixed populations of galaxies and active galactic nuclei (AGNs, e.g. Duncan et al. 2018), in some case being also able to successfully characterize the sources (Fotopoulou & Paltani 2018).

★ E-mail: massimo.brescia@inaf.it

SED-fitting techniques are able to provide all at once photo- z point estimates, photo- z probability density function (PDZ) and the spectral type of each source, at any redshift.

In the supervised ML techniques, the learning process is regulated by the spectroscopic information (i.e. redshift) available for a sub-sample of the objects. ML methods – ideal for the million of sources provided by multiwavelength surveys – are extremely precise as long as the spectroscopic sample is representative of the population for which the photo- z has to be computed. This means that they cannot provide accurate solutions outside z -spec range of the training set.

Even though the accuracy reached by empirical methods and SED fitting for inactive galaxies are comparable, is not the case for galaxies hosting AGNs. For these sources in fact, the amount of contribution from the AGN to the total emission in the various bands is a priori unknown. In photo- z computed via SED fitting this translates in the difficulty of defining for every survey the correct set of templates forming the library (e.g. Salvato et al. 2009; Cardamone et al. 2010; Luo et al. 2010; Salvato et al. 2011; Hsu et al. 2014; Ananna et al. 2017, hereafter A17). Similarly, in photo- z computed with empirical methods, it is necessary to have a very large and complete spectroscopic sample to use as training set. That is why in the past, there have been only few attempts to compute photo- z for AGN with ML-based methods (e.g. Budavári et al. 2001; Bovy et al. 2012; Brescia et al. 2013). All these pioneering works focused on the SDSS footprints (where the plethora of spectroscopic redshifts, hereafter z_{spec} , are well suited to ML techniques) and on optically selected QSOs, mostly at $z \geq 1.5$, which are dominated by the AGN, with little contribution from the host. For low-redshift and low-luminosity AGN (i.e. Seyfert galaxies), where the host galaxy contribution to the total emission is significant, both the accuracy and fraction of outliers of the photo- z computed with SED-fitting technique are comparable to the results for normal galaxies ($\sigma_{\text{NMAD}} \sim 1$ per cent; $\eta \sim 5$ per cent). However, this is true only in the fields where narrow and/or intermediate filter band photometry is available, like e.g. in COSMOS, CDFS, and Alhambra (e.g. Salvato et al. 2009, 2011; Marchesi et al. 2016; Cardamone et al. 2010; Matute et al. 2012). In fact, the narrow-/intermediate-band photometry easily pinpoints the emission lines, typical SED features for AGN. For a mixed set of AGNs and with only broad-band photometry, even with coverage from ultraviolet (UV) to mid-infrared (MIR), photo- z for AGN via SED-fitting reach an accuracy of about 6–8 per cent with about 18–25 per cent in fraction of outliers (e.g. Fotopoulou et al. 2012; Nandra et al. 2015, A17), with the accuracy of the extended sources better than those classified as point-like in optical images. The situation gets even more difficult when the survey is wide and the photometry is assembled from heterogeneous photometric catalogues rather than computed in a consistent way from homogenized images (e.g. COSMOS, Laigle et al. 2016; Ilbert et al. 2009).

ML methods are less affected by this problem because they account for the differences in the photometry defined in the various catalogues, but the photo- z computation must then be preceded by a search for the best features e.g. certain type of magnitudes, or specific photometric bands/colours (e.g. see Polsterer et al. 2014; D’Isanto et al. 2018; Fotopoulou & Paltani 2018; Ruiz et al. 2018). In addition, the quality degrades fast, when not sufficient photometry is available.

Photo- z can replace spectroscopic redshifts in AGN evolution studies or AGN clustering only if accompanied by the redshift PDZ (e.g. Miyaji et al. 2015; Georgakakis et al. 2014). While PDZs are routinely produced when computing photo- z via SED fitting,

their production with empirical methods requires an additional computational effort that only recently became more feasible (e.g. Sadeh, Abdalla & Lahav 2016; Cavauoti et al. 2017; Amaro et al. 2018; Brescia et al. 2018; Mountrichas et al. 2017; Duncan et al. 2018).

In terms of computational speed, ML outperforms the template-fitting techniques (Vanzella et al. 2004). This makes ML a natural choice for the computation of photo- z for the very large forthcoming deep and wide surveys such as *Euclid* (Laureijs 2010) and LSST (Ivezić et al. 2019), in which the computation of photo- z will be a real challenge. For AGN, the next challenge is presented by the ~ 3 million sources that eROSITA (extended Roentgen Survey with an Imaging Telescope Array; Merloni et al. 2012), the primary instrument on the Russian Spektrum–Roentgen–Gamma mission, will detect. eROSITA will provide an all-sky X-ray survey every 6 months for 4 yr, with a final expected depth of $1 \times 10^{-14} \text{ erg cm}^{-2} \text{ s}^{-1}$ ($3 \times 10^{-15} \text{ erg cm}^{-2} \text{ s}^{-1}$ at the poles) which is about 30 times deeper than *ROSAT* (Voges et al. 1999; Boller et al. 2016) in the soft band (0.5–2 keV). eROSITA will also provide for the first time ever an all-sky image in the hard band (2–10 keV), reaching an expected depth of $2 \times 10^{-13} \text{ erg cm}^{-2} \text{ s}^{-1}$ ($4 \times 10^{-14} \text{ erg cm}^{-2} \text{ s}^{-1}$ at the poles). With this depth, eROSITA will detect the low-luminosity AGN that are present in deep pencil-beam surveys, but also the bright and more rare objects that are observed in wide areas.

The recent release by A17 of the complete photometry of the counterparts to the X-ray sources detected in Stripe 82X (LaMassa et al. 2016, 2013a,b) offers the possibility to study the performances of photo- z for AGN via ML in a complete way. Stripe 82X covers an area of about 31 deg^2 , with a X-ray depth of $8.7 \times 10^{-16} \text{ erg cm}^{-2} \text{ s}^{-1}$ in the soft band and $4.7 \times 10^{-15} \text{ erg cm}^{-2} \text{ s}^{-1}$ in the hard band and it includes about 6000 sources, ~ 3000 of which are provided with reliable z_{spec} from SDSS. The photometric catalogue includes *GALEX*, SDSS, UKIRT, VHS, *Spitzer*/IRAC, and *WISE* with a depth sufficient to detect the X-ray sources at least at the depth of eROSITA. Thus, Stripe 82X can mimic eROSITA in terms of X-ray depth and ancillary data coverage that will be available in the whole sky, thanks to the coverage provided by e.g. PanStarrs, skyMapper, DES, VHS, UKIDSS, *WISE* (*Spitzer*/IRAC will be available only for patches of 100 deg^2) and with LSST and SpherEx in the future. A17 provides not only photometry, but also photo- z and related PDZs computed via SED fitting using LEPHARE (Arnouts et al. 1999; Ilbert et al. 2006), after splitting the sample in subgroups, each fitted with a dedicated library of templates.

The work presented in this paper consists of two parts. In the first part we evaluate and optimize the space of the parameters that improve the accuracy of ML. This is done by selecting the best features (Lal et al. 2006). In the literature (e.g. Guyon & Elisseeff 2003), there are plenty of algorithms aimed at the selection of the best features, such as principal component analysis (PCA; Jolliffe 2002), *filter* techniques, based on the evaluation of single features through a variety of significance tests, generally fast but less accurate (Gheyas & Smith 2010), *wrapper* methods, which make use of an arbitrary learning algorithm (such as neural networks or nearest-neighbour) to evaluate the relevance of feature sets (Kohavi & John 1997) and *embedded* methods, performing a feature selection during the prediction/classification model training procedure (a typical example is the Random Forest model, Breiman 2001). Here, we use a novel feature selection method, named Φ LAB (PhiLAB, Parameter handling investigation LABoratory), a hybrid approach that includes properties of both wrappers and embedded feature selection categories (Delli Veneri et al. 2019).

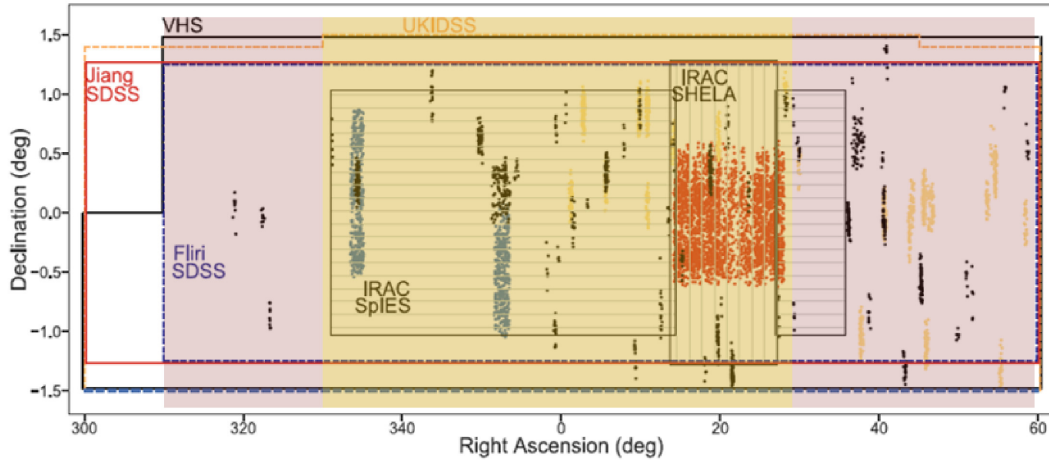


Figure 1. Map of the original multiwavelength coverage of Stripe 82X area discussed in A17. The total area extends for $\sim 2.5^\circ$ in Dec. and 120° in RA. The dots represent X-ray sources, respectively, from *XMM-Newton* AO13 (red), AO10 (blue), archival *XMM-Newton* sources (yellow), and *Chandra* sources (black). While standard photo-z are generated for the entire area (in red), the selection of the best features discussed in the first part of the paper is obtained considering only the sources in the yellow area.

For this purpose, the determination of the features, rather than the computation of the photo-z, is relevant. Because the best features are not available for the entire sample, our purpose here is to list them, for their use in other surveys. The computation of the photo-z via ML is discussed in the second part of the paper and for the reason described above it will be computed for the entire sample also in a more traditional way using MLPQNA (Brescia et al. 2013; Cavioti et al. 2012, 2015a). Missing some of the best features, will translate in the degradation of the accuracy.

For each of the sources, we provide also the PDZ through METAPHOR (Cavioti et al. 2017) and we will compare the photo-z and the PDZ with the results presented in A17. In particular, we will test the accuracy of the new photo-z at the depth of *eROSITA* all-sky survey.

Outline: in Section 2, we describe the data used in this work, while in Section 3, we briefly describe the methods involved in our analysis. In Section 4, we analyse the parameter space and its optimization, while in Section 5, we discuss the impact of X-ray flux, photometry, and morphology in the quality of photo-z. Sections 6 and 7 are devoted to the analysis of photo-z and PDZ estimation, respectively. Finally, the conclusions are drawn in Section 8. In Appendix A1, the catalogue of photo-z computed with MLPQNA and used for this work are made publicly available, while PDZs are available under request.

In this paper, unless differently stated, we use magnitude expressed in AB and adopt a cosmology of $H_0 = 70 \text{ km s}^{-1} \text{ Mpc}^{-1}$, $\Omega_M = 0.27$, and $\Lambda = 0.73$.

2 DATA

This section is dedicated to describe all photometric and spectroscopic data used in the experiments.

2.1 PHOTOMETRY

The photometry used in this work is extracted from the catalogue presented in A17, which lists the multiwavelength properties of the counterparts to the X-ray sources detected in Stripe 82X. Compared to the previous version of the catalogue presented in LaMassa et al. (2016), this new catalogue uses deeper multiwavelength data for

the identification of the counterparts and for the computation of the photometric redshifts via SED fitting. Although the catalogue of A17 includes 6187 X-ray sources, we focus here only on the 5990 for which a reliable counterpart was identified. All the details on the properties of the photometric data set are exhaustively presented in A17. Here, we provide only the list of data that we use for this paper, focusing on the central area of Stripe 82X, observed by *Spitzer*, as shown in Fig. 1. In particular, we considered:

- (i) FUV and NUV magnitudes and corresponding errors from *GALEX* all-sky survey (Martin et al. 2005). They were not used in this work due to the shallowness of the data;
- (ii) u, g, r, i, z SDSS AUTO magnitudes and corresponding errors from Fliri & Trujillo (2016);
- (iii) J, H , and K from VISTA (Irwin et al. 2004). As shown in A17 additional data in J_{UK}, H_{UK}, K_{UK} data from UKIDSS (Lawrence et al. 2007) are available for the same area but were not used in this paper;
- (iv) 3.6 and $4.5 \mu\text{m}$ magnitudes and corresponding errors from IRAC. Here, two complementary surveys are used: SPIES (Timlin et al. 2016) and SHELA (Papovich et al. 2016). Given the similarity of the two surveys, we do not differentiate sources belonging to one or another;
- (v) $W1, W2, W3$, and $W4$ magnitudes and corresponding errors from AllWISE (Wright et al. 2010).

In the first column of Table 1, we report the nominal depth of each photometric band considered.

The original catalogue is complemented by soft, hard, and full-band X-ray fluxes from *XMM-Newton* and *Chandra* (see LaMassa et al. 2016, for details). It also includes morphological information on the extension and variability of the sources in the optical band. We retain such information, as it has been already demonstrated in literature that they affect the accuracy of photo-z for AGN and can be used as priors for improving performance (e.g. Salvato et al. 2009). While these data are not used directly for the computation of the photo-z, they are employed to perform various experiments by creating sub-samples in X-ray flux and morphology.

Table 1. Summary table for depth, amount of sources, and redshift coverage. The first column refers to the nominal depth of the entire sample of reliable counterparts in Stripe 82X, as presented in A17. The following columns refer to the magnitudes reached in the various experiments, i.e. the faintest magnitude reported in the Stripe 82X catalogue for the various sub-samples for which the photo-z have been computed. The values in the column *BEST* represent the faintest magnitudes of the sub-sample of sources in the yellow area of Fig. 1, used for the features analysis performed with Φ LAB, (Section 3.1). The bands marked with a – symbol have been discarded from that specific experiment. [†] SPIES and SHELA have been used together (Section 2.1). The last rows of the table list, for each of the sub-samples, the total number of sources, their amount with spectroscopic redshift available, those with a X-ray flux brighter than 10^{-14} erg cm⁻² s⁻¹ and the final depth expected by eROSITA all-sky survey. Note that here we use only the sources for which the determination of the counterpart is secure, i.e. (SDSS,VHS,IRAC).REL.CLASS= SECURE in the catalogue of A17.

| Filter | Band depth | | | | | | | | |
|--|------------|--------------------|-------|--------------|--------------------|---------------|--------------------|-------------------|------------------------|
| | Nominal | BEST | SDSS | SDSS and VHS | SDSS and IRAC | SDSS and WISE | SDSS VHS and IRAC | SDSS VHS and WISE | SDSS VHS IRAC and WISE |
| <i>u</i> | 31.22 | 28.54 | 28.54 | 28.54 | 28.54 | 28.54 | 28.54 | 28.54 | 28.54 |
| <i>g</i> | 28.77 | 24.20 | 24.39 | 24.20 | 24.39 | 24.39 | 24.20 | 24.20 | 24.20 |
| <i>r</i> | 27.13 | 23.25 | 23.43 | 23.25 | 23.43 | 23.43 | 23.25 | 23.25 | 23.25 |
| <i>i</i> | 27.21 | 22.35 | 23.49 | 22.64 | 23.49 | 22.45 | 22.64 | 22.35 | 22.35 |
| <i>z</i> | 30.46 | 22.42 | 23.35 | 22.46 | 22.99 | 22.42 | 22.46 | 22.42 | 22.08 |
| <i>J</i> | 24.74 | 21.64 | – | 24.64 | – | – | 21.64 | 21.64 | 21.51 |
| <i>H</i> | 24.15 | 22.87 | – | 22.87 | – | – | 21.61 | 22.87 | 21.61 |
| <i>K</i> | 22.60 | 21.63 | – | 21.63 | – | – | 21.63 | 21.63 | 21.63 |
| CH1_SPIES | 24.27 | 20.82 [†] | – | – | 21.64 [†] | – | 21.06 [†] | – | 20.49 [†] |
| CH1_SHELA | 22.80 | – | – | – | – | – | – | – | – |
| CH2_SPIES | 22.88 | 20.49 [†] | – | – | 21.41 [†] | – | 21.07 [†] | – | 20.22 [†] |
| CH2_SHELA | 23.88 | – | – | – | – | – | – | – | – |
| W1 | 21.16 | 20.71 | – | – | – | 20.71 | – | 20.71 | 20.61 |
| W2 | 20.74 | 20.59 | – | – | – | 20.63 | – | 20.63 | 20.59 |
| W3 | 18.20 | 18.04 | – | – | – | 18.11 | – | 18.11 | 18.04 |
| W4 | 16.15 | 16.06 | – | – | – | 16.13 | – | 16.13 | 15.94 |
| Number of sources | 5990 | 2290 | 4855 | 3218 | 2293 | 3291 | 1620 | 2696 | 1380 |
| Number of sources | 2933 | 1686 | 2793 | 2218 | 1596 | 2160 | 1279 | 1935 | 1121 |
| w/z_{spec} | | | | | | | | | |
| Number of sources | 2351 | 1249 | 2025 | 1649 | 1051 | 1619 | 888 | 1445 | 793 |
| $w/F_X > 10^{-14}$ | | | | | | | | | |
| Number of sources | 1550 | 1025 | 1483 | 1309 | 857 | 1256 | 758 | 1174 | 683 |
| $w/F_X > 10^{-14}$ and z_{spec} | | | | | | | | | |

2.2 SPECTROSCOPY

The spectroscopic coverage of the field (see Fig. 2) is ideal for assessing the performances of photo-z for X-ray-selected sources via ML. The spectroscopic surveys BOSS (Dawson et al. 2013) and eBOSS (Delubac et al. 2017) in the lifetime of SDSS, provide reliable redshifts for about 50 percent of the sources (2, 962/5, and 990). In addition Stripe 82X was also suitable for a dedicated spectroscopic program during SDSS-IV, targeting specifically the counterparts to X-ray sources (LaMassa et al. 2019). There, the exposure time was of at least 2 h long, allowing the determination of the redshifts also for faint sources. The training and testing samples are formed only by sources with redshift available at the time of the publication of A17. However, as additional blind test, we checked the accuracy of the photo-z also using this new available spectroscopic sample of 257 sources (see Section 6).

3 THE ALGORITHMS

We first performed a feature analysis on the sub-sample of sources located in the yellow area of Fig. 1. This sample maximizes the number of sources, the number of photometric bands, their depth, and the faintest magnitudes available that are reported in the column *BEST* of Table 1. The feature analysis was performed

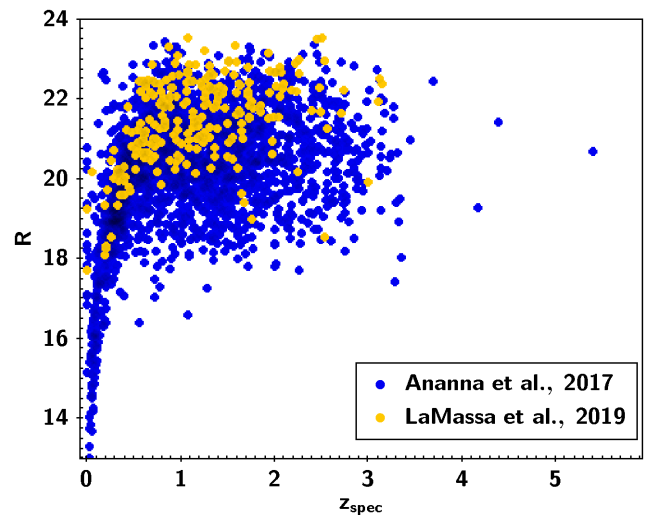


Figure 2. Redshift and magnitude distribution for the sources with spectroscopic redshift. The blue sources were presented in A17 and have been used in this work as training and blind test samples. The 258 yellow sources are on average fainter and were recently presented in LaMassa et al. (2019). They are used as additional blind test sample.

with the algorithm Φ LAB, developed by our group and described in Section 3.1. Obviously, the computation of the photo- z using the best features would provide the best accuracy, but it would dramatically limit the size of the sample for which the photo- z could be computed. In fact, because in this work magnitudes were used instead of fluxes, the sample includes many source missing data in various bands.¹ Rather than using specific but not yet fully tested methods for recovering the missing data (see Brescia et al. 2018, for a discussion), we computed the photo- z for subsets of sources that share the same photometric system (i.e. sources for which a certain combination of bands does not include missing data). The sub-samples naming convention is presented in Table 1, together with the breakdown of the sources and the depth reached in every filter. The photo- z were estimated with MLPQNA (described in Section 3.2). For each source, we also derived the redshift PDZ using METAPHOR, described in Section 3.3.

The three algorithms are described in the following subsections.

3.1 Φ LAB

Recently, in Delli Veneri et al. (2019), we presented a novel method suitable for a deep analysis and optimization of any parameter space, which provides in output the selection of the most relevant features. The algorithm developed by our group is called Φ LAB. It is a hybrid approach, including properties of both wrappers and embedded feature selection categories (Tangaro et al. 2015), based on two joined concepts, respectively: *shadow features* (Kursa & Rudnicki 2010) and *Naïve LASSO* (Least Absolute Shrinkage and Selection) statistics (Tibshirani 2013). Shadow features are randomly noised versions of the real ones and their importance percentage is used as a threshold to identify the most relevant features among the real ones. Afterwards, the two algorithms, based on LASSO and integrated into Φ LAB, perform a regularization, based on the standard L_1 norm, of a ridge regression on the residual set of weak relevant features (i.e. a shrinking of large regression coefficients to avoid overfitting). This has the net effect of *sparsifying* the weights of the features, effectively turning off the least informative ones.

LASSO acts by conditioning the likelihood with a penalty on the entries of the covariance matrix and such penalty plays two important roles. First, it reduces the effective number of parameters and, second, produces an estimate which is sparse. Having a regularization technique as part of a regression minimization law, represents the most evident difference with respect to more traditional parameter space exploration methods, like PCA (Pearson 2010). The latter is a technique based on feature covariance matrix decomposition, where the principal components are retained instead of the original features. The two concepts, *shadow features* and *Naïve LASSO*, are then combined within the proposed method by extracting the list of candidate most relevant features through the noise threshold imposed by the shadow features and by filtering the set of residual weak relevant features through the LASSO statistics.

Φ LAB is detailed in Delli Veneri et al. (2019), where the method has been used to investigate the parameter space in the case of the photometric determination of star formation rates in the SDSS.

¹Working with fluxes would have allowed to use also the faint sources at background level, with negative fluxes but associated to a large, positive photometric errors. By definition these sources are not present in a catalogue that lists magnitudes.

3.2 MLPQNA

MLPQNA (Multi Layer Perceptron trained with Quasi Newton Algorithm) is a Multi Layer Perceptron (MLP; Rosenblatt 1962) neural network trained by a learning rule based on the Quasi Newton Algorithm (QNA) which is among the most used feed-forward neural networks in a large variety of scientific and social contexts, such as electricity price (Aggarwal, Saini & Kumar 2009), detection of premature ventricular contractions (Ebrahimzadeh & Khazaei 2010), forecasting stock exchange movements (Mostafa 2010), landslide susceptibility mapping (Zare et al. 2013), etc.

Furthermore, it has been successfully applied several times in the context of photometric redshifts (see for instance Biviano et al. 2013; Brescia et al. 2013, 2014b; Cavuoti et al. 2015a; de Jong et al. 2017; Nicastro et al. 2018). The analytical description of the method has been discussed in the contexts of both classification (Brescia, Cavuoti & Longo 2015) and regression (Cavuoti et al. 2012; Brescia et al. 2013).

3.3 METAPHOR

METAPHOR (Machine-learning Estimation Tool for Accurate PHOtometric Redshifts; Cavuoti et al. 2017) is a modular workflow, designed to produce the redshift PDZs through ML. Its internal engine is the MLPQNA already described in Section 3.2. The core of METAPHOR lies in a series of different perturbations of the photometry in order to explore the parameter space of data and to grab the uncertainty due to the photometric error. In practice, the procedure to determine the PDZ of individual sources can be summarized in this way: we proceed by training the MLPQNA model and by perturbing the photometry of the given blind test set to obtain an arbitrary number N of test sets with a variable photometric noise contamination. Then, we submit the $N + 1$ test sets (i.e. N perturbed sets plus the original one) to the trained model, thus obtaining $N + 1$ estimates of photo- z . With such $N + 1$ values we perform a binning in photo- z (0.01 for the described experiments), thus calculating for each one the probability that a given photo- z value belongs to each bin. In this work, we used $N = 999$ to obtain a total of 1000 photo- z estimates.

3.4 Statistical estimators

For brevity, we define Δz as:

$$\Delta z = (z_{\text{phot}} - z_{\text{spec}})/(1 + z_{\text{spec}}) \quad (1)$$

Then, in order to be able to compare the accuracy with that available for other surveys/methods present in literature, we use the classical basic statistical estimators, applied on Δz , described as following:

- (i) mean (or bias);
- (ii) standard deviation σ ;
- (iii) $\sigma_{\text{NMAD}} = 1.4826 \times \text{median}(|\Delta z|)$;
- (iv) σ_{68} that is the width in which falls the 68 per cent of the Δz distribution;
- (v) η , defined as the fraction (percentage) of outliers or source for which $|\Delta z| > 0.15$.

Due to the limited number of data samples, a canonical splitting of the data set (or knowledge base) into training and blind test set cannot be applied. Therefore, in order to circumvent this problem the training + test process involves a k -fold cross-validation (Hastie, Tibshirani & Friedman 2009; Kohavi 1995): the knowledge base

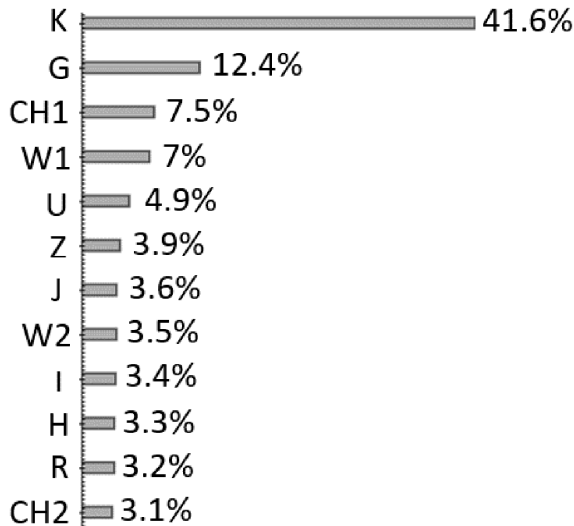


Figure 3. Results of the feature analysis performed with Φ LAB. The importance of each feature is estimated for the case in which only magnitudes are considered for the sample *BESTmagopt*.

has been manually split into four not-overlapped subsets. In this way by taking each time three of these subsets as training set and leaving the fourth as blind test set, an overall blind test on the entire knowledge base sample can be performed, i.e. each object of the available data sample has been evaluated in a blind way (i.e. not used for the training phase).

4 RESULTS ON PARAMETER SPACE OPTIMIZATION

We started our experiments by considering the *BEST* sample that maximizes the number of sources with spectroscopic redshift and with the maximum number of photometric points available (as discussed in Section 2.1). The *BEST* sample includes 15 bands:

- (i) u, g, r, i, z ;
- (ii) J, H, K ;
- (iii) $CH1, CH2$;
- (iv) $W1, W2, W3, W4$;
- (v) X-FLUX.

Using Φ LAB on the *BEST* sample, we valued the most relevant features, considering the parameter space of photometry alone, named *BESTmagopt* (Fig. 3). We then repeated the feature analysis on the *BESTmagopt* space, in which we added the direct derived colours (i.e. those obtained by couples of adjacent magnitudes only), obtaining the optimized parameter space named as *BESTmagcolopt*. Fig. 4 shows the impact of each selected feature in this case. Finally, (as done in Ruiz et al. 2018), we considered also the case of the full feature space, by including all *BESTmagopt* magnitudes and all possible derived colours, for a total of 78 features:

- (i) all five SDSS magnitudes and related 10 colours;
- (ii) all three VHS magnitudes and related three colours;
- (iii) all two IRAC magnitudes and related one colour;
- (iv) the two previously selected *WISE* bands $W1, W2$, and related one colour;
- (v) all 15 combinations of colours among SDSS and VHS;
- (vi) all 10 combinations of colours among SDSS and IRAC;

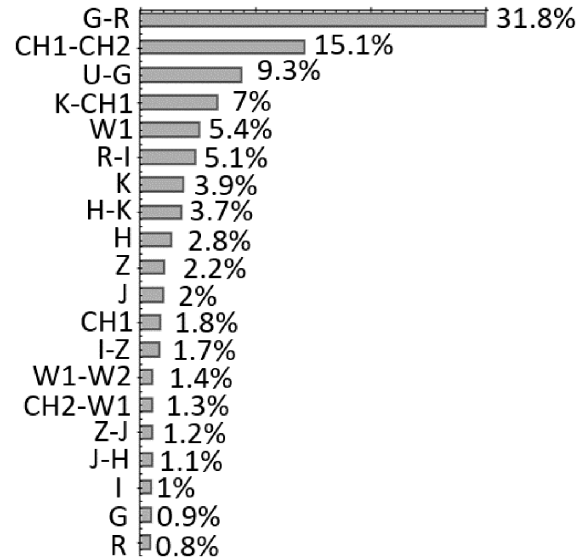


Figure 4. Results of the feature analysis performed with Φ LAB. The importance of each feature is estimated for the case in which magnitudes and colours are considered for the sample *BESTmagcolopt*.

- (vii) all 10 combinations of colours among SDSS and the two selected *WISE* bands;
- (viii) all six combinations of colours among VHS and IRAC;
- (ix) all six combinations of colours among VHS and the two selected *WISE* bands;
- (x) all four combinations of colours among IRAC and the two selected *WISE* bands.

In this case, our method extracted a set of 67 features considered suitable for the photo- z estimation. Table 2 reports the feature importance ranking for all the selected features.

Table 3 reports the statistical results about the photo- z predictions for the various selected feature spaces. The photo- z estimation experiment performed with the largest selected parameter space (i.e. using the 67 features selected by Φ LAB among the 78 available), provided statistical results comparable with the *BESTmagcolopt* case, thus inducing to consider the latter as the best candidate parameter space, due to its smaller number of dimensions. When only magnitudes are considered, the K band is by far the most important feature. The reason is easily understood keeping in mind the SED of a galaxy. The rest-frame K band indicates the knee of the SED and this clear feature is indeed suitable to determine the redshift. However, the optimal feature combination turned out to be the mixed one, i.e. including colours and some reference magnitudes belonging to different surveys (for instance, *GRIZ* for SDSS, *JHK* for VHS, *CHI* for IRAC, and *W1* for *WISE*). From the data mining viewpoint, this could appear rather surprising, since the kind of information should not be expected to change by introducing linear combinations between parameters, as colours are obtained by subtraction between magnitudes. However, from an astrophysical point of view, colours define the SED of the sources and their evolution with redshift, and for this reason, they are crucial in the process of determining photo- z . The additional reference magnitudes, on the other hand, contribute to minimize the degeneracy in the luminosity class for a specific object type.

By keeping in mind such arguments, it appears less surprising that the information importance carried by magnitudes within the parameter space devoid of colours, drastically decreases in the

Table 2. Results of the feature analysis (percentages of estimated feature importance) performed with Φ LAB in the case of the parameter space composed by considering all magnitudes and colours available.

| Feature | Importance (per cent) | Feature | Importance (per cent) |
|---------|--------------------------|---------|--------------------------|
| R–Z | 14.51 | J–K | 0.40 |
| G–I | 12.44 | U–CH1 | 0.35 |
| CH1–CH2 | 7.50 | H–CH1 | 0.34 |
| U–G | 6.00 | R–CH2 | 0.33 |
| Z–W1 | 5.84 | U–I | 0.33 |
| Z–CH1 | 4.24 | R–W2 | 0.33 |
| G–R | 4.03 | K–CH1 | 0.33 |
| K | 3.14 | R–W1 | 0.31 |
| G–Z | 3.03 | U | 0.30 |
| I–W1 | 2.00 | U–J | 0.30 |
| I–CH2 | 1.94 | G–W2 | 0.27 |
| H | 1.81 | G–CH2 | 0.24 |
| R–I | 1.67 | I–J | 0.23 |
| I–CH1 | 1.51 | CH1–W2 | 0.23 |
| J | 1.45 | G | 0.22 |
| H–K | 1.34 | J–CH2 | 0.21 |
| R | 1.21 | G–CH1 | 0.21 |
| I | 1.21 | G–K | 0.20 |
| W1 | 1.18 | J–W1 | 0.20 |
| I–Z | 1.08 | H–W2 | 0.20 |
| Z | 0.99 | K–CH2 | 0.17 |
| H–W1 | 0.97 | K–W2 | 0.16 |
| K–W1 | 0.83 | U–W1 | 0.16 |
| Z–W2 | 0.83 | Z–J | 0.16 |
| CH2–W1 | 0.77 | U–K | 0.15 |
| Z–CH2 | 0.68 | R–CH1 | 0.14 |
| U–R | 0.68 | H–CH2 | 0.13 |
| U–Z | 0.62 | CH1 | 0.13 |
| G–W1 | 0.56 | Z–H | 0.13 |
| J–CH1 | 0.54 | U–H | 0.12 |
| Z–K | 0.52 | J–W2 | 0.09 |
| I–W2 | 0.50 | I–K | 0.08 |
| J–H | 0.46 | R–K | 0.07 |
| W1–W2 | 0.45 | – | – |

Table 3. Accuracy of photo-z computed with MLPQNA on *BEST*, *BESTmagopt*, and *BESTmagcolopt* samples, after the optimization of the parameter spaces with the features analysis and selection performed with Φ LAB. All quantities are calculated on blind test sets only.

| ID | <i>BEST</i> | <i>BESTmagopt</i> | <i>BESTmagcolopt</i> |
|------------------------|-------------|-------------------|----------------------|
| Number of sources | 1686 | 1686 | 1686 |
| Bands | 14 | 12 | 20 |
| bias | 0.0159 | 0.0105 | 0.0102 |
| σ | 0.141 | 0.135 | 0.121 |
| σ_{NMAD} | 0.079 | 0.074 | 0.056 |
| σ_{68} | 0.091 | 0.083 | 0.069 |
| η | 16.09 | 13.88 | 12.74 |

mixed parameter space, where some representative colours result as the most significant features. In fact, the first four colour features of the *BESTmagcolopt* parameter space (Fig. 4), contain about 65 per cent of the importance carried by the whole set of magnitudes present in the *BESTmagopt* parameter space (Fig. 3). Furthermore, the strong relevance carried by the colours in the ranking list is reflecting the importance decreasing of related magnitudes, now represented by their colour combinations, causing in some cases

Table 4. Same of Table 3, but considering only objects with $F_X > 10^{-14}$.

| ID | <i>BEST</i> | <i>BESTmagopt</i> | <i>BESTmagcolopt</i> |
|------------------------|-------------|-------------------|----------------------|
| Number of sources | 1029 | 1029 | 1029 |
| Bands | 14 | 12 | 20 |
| bias | 0.0157 | 0.0183 | 0.0130 |
| σ | 0.144 | 0.138 | 0.122 |
| σ_{NMAD} | 0.078 | 0.072 | 0.057 |
| σ_{68} | 0.092 | 0.077 | 0.074 |
| η | 15.90 | 12.31 | 13.12 |

the rejection of some magnitudes (e.g. U, CH2, and W2) from the optimized parameter space.

4.1 Impact of feature analysis on photo-z

The identification and consequent rejection of the non relevant features, allow us to obtain more accurate photo-z. This is demonstrated in Table 3 where we report the accuracy and fraction of outliers for the photo-z computed with MLPQNA for the *BEST* sample, with and without the removal of unimportant features for both *BESTmagopt* and *BESTmagcolopt*. Table 4 is the same as Table 3, but this time the metrics are computed by limiting the samples to the sources that *eROSITA* will be able to detect. The comparison between the two tables points out something that is already well documented in literature in the case of photo-z computed via SED fitting: namely, the accuracy of photo-z for AGN increases when the sample includes faint AGN, dominated by the host galaxies, easier to be modeled. As soon as the sample is limited to bright AGN, the quality of photo-z decreases.

5 THE IMPACT OF X-RAY FLUX, PHOTOMETRY, AND MORPHOLOGY IN THE QUALITY OF PHOTO-Z

After having identified the most relevant features, we were interested in exploring under which range of parameters the photo-z could be further improved. We have investigated in this respect the impact that photometric errors, X-ray depth, and lack of information on optical morphology have on the results, again using the sample *BESTmagcolopt* as reference.

It is worth noting that ideally, we should have also checked whether the feature analysis itself was influenced by these parameters. However, the samples are not sufficiently large for such kind of test. Therefore, we assume that the feature analysis is not affected and simply quantify the accuracy in photo-z for different sub-samples extracted from *BESTmagcolopt*, using different selection criteria reported in the following sections.

5.1 Impact of photometric errors

The photometric catalogues used in Stripe 82X are relatively shallow, implying that the fainter sources in general have a large photometric error. Unlike SED fitting, until recently ML techniques could not handle the errors associated to the measurements and the same weight was incorrectly assumed for each photometric value (but see Reis, Baron & Shahaf 2019, for a counter example application). We tried to assess how this is impacting on the result, by reducing *BESTmagcolopt* to sub-samples of decreasing photometric errors in all the bands. More specifically, we considered only sources with an error in magnitude smaller than 0.3, 0.25, and

Table 5. Statistical results of the *BESTmagcolopt* photo- z estimation experiments after having removed objects with photometric errors larger than 0.3, 0.25, and 0.2 respectively.

| | <i>BESTmagcolopt</i> | with Mag_err limit | | |
|------------------------|----------------------|--------------------|--------|--------|
| | | 0.3 | 0.25 | 0.2 |
| Number of sources | 1686 | 1442 | 1372 | 1275 |
| bias | 0.0102 | 0.0152 | 0.0126 | 0.0122 |
| σ | 0.121 | 0.134 | 0.157 | 0.151 |
| σ_{NMAD} | 0.056 | 0.056 | 0.059 | 0.054 |
| σ_{68} | 0.069 | 0.065 | 0.069 | 0.065 |
| η | 12.74 | 11.93 | 12.90 | 12.24 |

Table 6. Comparison between statistics for the complete best sample and for the sub-sample limited to eROSITA flux also in the training sample.

| | <i>BESTmagcolopt</i> | At eROSITA depth |
|------------------------|----------------------|------------------|
| Number of sources | 1686 | 1029 |
| bias | 0.010 | 0.013 |
| σ | 0.121 | 0.142 |
| σ_{NMAD} | 0.056 | 0.064 |
| σ_{68} | 0.069 | 0.075 |
| η | 12.74 | 12.73 |

0.2, thus reducing the original sample by ~ 15 per cent, ~ 19 per cent, and ~ 24 per cent, respectively (see Table 5).

Considering only sources with small photometric errors provides the best accuracy ($\sigma_{\text{NMAD}} = 0.054$), but the bias increases with respect to the original sample. The best trade-off is obtained by keeping sources with a photometric error smaller than 0.3 mag.

5.2 X-ray depth

The general experiment on Stripe 82X has demonstrated that reliable photo- z of the same quality, or even better than those computed via SED fitting, can be obtained for X-ray-detected AGN also with ML, as long as a large number of photometric points is available and the spectroscopic sample is representative. But we are also interested to evaluate the accuracy obtained for a sub-sample of the sources, such as the brightest or the faintest detected in X-ray. And, in particular, the expected accuracy that can be reached for eROSITA. The final depth after 4 yr of observations will be of $\sim 10^{-14} \text{ erg s}^{-1} \text{ cm}^{-2}$ for the all-sky survey. The survey will detect AGN also at high redshift, but it will be dominated by bright, nearby AGN, for which the computation of the photo- z is typically more challenging. Table 4 already reported a partial answer to the questions: namely, the accuracy in photo- z for X-ray bright AGN is worse than for the entire sample. This means that the good results obtained in the second column are driven by the fact that the faint AGN, easier to fit because galaxy dominated, are more numerous than the bright AGN. However, in that experiment, the training in the photo- z computation was done using all the sources in the *BESTmagcolopt*, with the cut in X-ray flux done a posteriori on the output. In the following experiment instead, also the training sample is limited to the bright sources that eROSITA will detect. The result of test is presented in Table 6. By comparing the last column of that table with the last column of Table 4, we see that, while the bias and σ_{NMAD} remain unchanged, the fraction of outliers decreases. It means that we could improve our result, if, in addition to good photometry for the entire sample, we could increase the training sample of bright objects by the time when eROSITA survey will be available. Photo-

z computed via ML for X-ray-selected sources in 3XMM-DR6 and 3XMM-DR7 where recently presented also in Ruiz et al. (2018) and Meshcheryakov et al. (2018), respectively. While we are in overall agreement with the first, our results are less optimistic than those obtained by the second group. However, the results is not surprising when noting that their results are specifically obtained for QSO or type 1 only, having as targets sources in *ROSAT* and 3XMM-DR7 that are presented in the spectroscopic catalog SDSS-DR14Q (P  ris et al. 2018). In our work, there is no any pre-selection and the sample includes QSO, type 1 and type 2 AGN and galaxies.

5.3 Point-like versus extended

Given the resolution of the ground-based optical imaging, extended sources can only be found at low redshift ($z_{\text{spec}} \leq 1$), with not significant contribution to the emission by the host galaxy. In contrast, point-like sources are mostly dominating the high-redshift regime, with the emission due to the nuclear component. This is taken into account when computing photo- z via SED fitting by adopting a prior in absolute magnitude (e.g. Salvato et al. 2009, 2011; Fotopoulou et al. 2012; Hsu et al. 2014, A17). More recently, the separation of the sources in these two subgroups is becoming the standard also when computing photo- z via ML (e.g. Mountrichas et al. 2017; Ruiz et al. 2018). One limitation of this method is that it relies on images that are affected by the quality of the seeing, which can alter the morphological classification of the sources. This has been demonstrated in Hsu et al. (2014) where the authors shown how the sources can change classification (and thus their photo- z value), depending on whether the images used are from *Hubble Space Telescope* or ground based. In Stripe 82X, out of 1469 sources at $z > 1$, 77 (~ 5 per cent) are classified as extended. In the *BESTmagcolopt* sample, the fraction is approximately the same (27/704, ~ 4 per cent). Given the resolution of SDSS, this is clearly nonphysical.

In this section, we first measure separately the accuracy for the point-like and extended sources in the *BESTmagcolopt* sample. Here, a mixed training sample was used. In our second approach, we created two training samples: one that includes only sources classified as ‘extended’ and having redshift smaller than one; a second including only sources classified as ‘point-like’ and/or at redshift larger than one. The sources in the test samples were separated accordingly.

The resulting statistics are shown in Table 7, with the first column reporting for convenience the first one of the two previous tables.

The second and third columns show the accuracy for the same sources of the *BESTmagcolopt* sample, but this time divided according to their extension.

As expected, the photo- z for extended sources are more reliable, with 50 per cent less outliers than the point-like sources. Columns 4 and 5 show the extreme case, in which the training sample is split in two *ab initio*. In this case, the photo- z for extended sources have the same accuracy of the best photo- z for normal galaxies, virtually without outliers or bias.

In contrast, the photo- z for point-like sources show about 20 per cent of outliers. This is clearly understood, again thinking about the SED of these objects, by comparing Column 3 with Column 5 of the table. This suggests that the training sample for point-like sources must include also sources with a contribution from the host. The last column of the table shows which precision for the entire sample is achieved by two training samples (one specific for extended and nearby sources and the second for a generic one).

Table 7. Photo- z estimation accuracy for the sources in *BESTmagcolopt* using a unique training sample (Column 1), afterwards divided between extended (Column 2) and point-like (Column 3). The photo- z are also computed by splitting the sources between the two groups and training them separately. For this case, the accuracy for extended and point-like are presented in Columns 4 and 5. In Column 6, we recombine the sample. The improvement can be seen by comparing the column *Combined* with *BESTmagcolopt*. All quantities are calculated on the blind test set extracted from the *BESTmagcolopt* sample.

| | <i>BESTmagcolopt</i> | Limited to: | | With specific training: | | |
|------------------------|----------------------|-------------|------------|-------------------------|------------|----------|
| | | Extended | Point-like | Extended | Point-like | Combined |
| Number of sources | 1686 | 598 | 1088 | 598 | 1088 | 1686 |
| bias | 0.0102 | 0.0097 | 0.0107 | 0.0005 | 0.0168 | 0.0099 |
| σ | 0.121 | 0.089 | 0.136 | 0.051 | 0.142 | 0.109 |
| σ_{NMAD} | 0.056 | 0.042 | 0.068 | 0.029 | 0.082 | 0.053 |
| σ_{68} | 0.069 | 0.050 | 0.082 | 0.032 | 0.096 | 0.071 |
| η | 12.74 | 7.69 | 15.57 | 1.67 | 19.40 | 12.59 |

6 PHOTO-Z ESTIMATION AND RELIABILITY

In the previous section, we analysed the impact of different factors on the accuracy of photo- z for X-ray-selected sources. However, that analysis was done only on a small sample of the sources in Stripe 82X, for which all the photometric points were available. It will be possible to use all what we have learned on tens of thousands of square degrees of sky only when surveys such as LSST (Ivezić et al. 2019) and SpherEx (Dore 2018) will come online. For the time being, in case of photo- z estimated via ML using catalogues of photometric points expressed in magnitudes, we face the problem that for many sources some of the values are missing. For this reason, in order to provide a photo- z for most of the sources, we prioritized the sample divided in subgroups that share the same multiwavelength coverage, sorted by accuracy in terms of σ_{NMAD} :

- (i) SDSS, VHS, *WISE*, and IRAC (*sdssVWI*)
- (ii) SDSS, VHS, and *WISE* (*sdssVW*);
- (iii) SDSS, VHS, and IRAC (*sdssVI*);
- (iv) SDSS and *WISE* (*sdssW*);
- (v) SDSS and IRAC (*sdssI*);
- (vi) SDSS and VHS (*sdssV*);
- (vii) SDSS.

In addition, with the goal of comparing the results with those obtained via SED fitting in A17, we have created a sample, *MLPQNA_{merged}*, where for each source we consider the photo- z computed via MLPQNA using the data set with the highest accuracy available.

This sorting, with the obvious limitation of producing photo- z with different accuracy across the field, offers nevertheless the possibility to characterize their quality as a function of the amount of photometric bands and the wavelength coverage. The results of the metrics used for measuring the quality of the photo- z are presented in Table 8 for the entire sample and in Table 9 for the sources that are in common to all samples. A visualization of the results is also presented in Fig. 5, which provides the comparison between spectroscopic and photometric redshifts computed with MLPQNA for the sources in each sub-sample and in A17. Here, we plot only the sources that eROSITA will detect. The photometric coverage, limited to all optical bands, reduces the sample by 5 per cent, from 1535 down to 1471 and produces an excess of high-redshift values for sources that are actually at low redshift. The effect can be mitigated by adding redder bands, with the MIR bands from *WISE* being more efficient than the NIR bands from VHS. The combined addition of NIR and MIR photometric points removes most of the outliers, but it also reduces the original sample by 35 per cent (from 1535 to 1019 sources).

However, at the X-ray flux of eROSITA, and when SDSS, VHS, *WISE*, and IRAC data are simultaneously available, MLPQNA performs better than the SED-fitting technique, with a lower fraction of outliers and the reassuring absence of systematics (Fig. 6). The result is even more impressive if we consider the limited size of the training sample (contrary to Ruiz et al. 2018, only spectroscopy available within the Stripe 82X field is used).

At the depth of eROSITA, the two methods are equally performing, with essentially the same accuracy and fraction of outliers (Fig. 7).

ML facilitates the process of computing reliable photo- z when the number of photometric points is sufficient, because it avoids any assumption on the type of templates needed in the library when SED-fitting technique is used. As underlined in A17, this is a lengthy and risky procedure, since a slightly different set of template SEDs can produce vastly different results. However, when only a limited number of photometric points is available, SED fitting remains a better approach, as it provides more reliable results and smaller fraction of outliers for the entire sample.

Fig. 5 also shows that the performance of *sdssVWI* is higher compared to *sdssVW* and *sdssVI*. This is due to the fact that, as can be seen in Fig. 4 of the official *WISE* web site,² although W1 and W2 are centred almost at the same wavelength of IRAC/CH1 and IRAC/CH2, they are broader, with W1(W2) extending to shorter(longer) wavelength, with respect to CH1(CH2). Using simultaneously the four bands increases the characterization of the SEDs.

IRAC photometry is deeper and more precise (i.e. smaller photometric errors) than *WISE* photometry and this explains why, at the depth of eROSITA, *sdssVI* performs better than *sdssVW*. As we discussed in Section 5.1, a precise photometry helps in disentangling the correct SED of the sources and this is particularly true for the bright X-ray sources. In contrast, Table 8 shows that when the sample includes also faint sources, *sdssVW* performs better than *sdssVI*. This can be explained by considering that the sources start to be dominated by the host, with less differences in their SED and for this reason better determined using a larger wavelength coverage than a precise photometric point.

It is worth noting that, although IRAC data are available only for certain areas of the sky, the all-sky observations with *WISE* is ongoing. Already the most recent data release (Schlafly, Meisner & Green 2019) is 0.7 mag deeper than the data used here. So it is plausible to predict that the precision over the entire sky for eROSITA is well represented by the *sdssVI* case and will not be worse than the *sdssVW* case.

²http://wise2.ipac.caltech.edu/docs/release/prelim/expsup/sec4_3g.html

Table 8. Summary of all statistical results for the sub-samples with different photometric coverage, listed in decreasing order of reliability (based on σ_{NMAD}). In MLPQNA_{merged}, we list the best photo-z available for each source. All quantities are calculated on blind test sets. For comparison, on the last two columns we report σ_{NMAD} and η for the same sub-sample in A17.

| | Number of sources | $ bias $ | σ | σ_{68} | σ_{NMAD} | η | $\sigma_{\text{NMAD}_{A17}}$ | η_{A17} |
|------------------------------|----------------------|----------|----------|---------------|------------------------|--------|------------------------------|--------------|
| (1) <i>sdssVWI</i> | 1686 | 0.0102 | 0.121 | 0.069 | 0.056 | 12.74 | 0.059 | 13.3 |
| (2) <i>sdssVW</i> | 1889 | 0.0163 | 0.169 | 0.077 | 0.065 | 16.20 | 0.060 | 13.2 |
| (3) <i>sdssVI</i> | 1279 | 0.0210 | 0.134 | 0.079 | 0.067 | 16.50 | 0.057 | 13.4 |
| (4) <i>sdssW</i> | 2121 | 0.0193 | 0.190 | 0.078 | 0.069 | 16.55 | 0.061 | 13.8 |
| (5) <i>sdssI</i> | 1595 | 0.0152 | 0.170 | 0.089 | 0.077 | 17.49 | 0.059 | 15.1 |
| (6) <i>sdssV</i> | 2142 | 0.0247 | 0.264 | 0.108 | 0.089 | 23.24 | 0.060 | 13.9 |
| (7) <i>sdss</i> | 2747 | 0.0410 | 0.270 | 0.104 | 0.087 | 24.17 | 0.063 | 15.7 |
| (8) MLPQNA _{merged} | 2780 | 0.0178 | 0.173 | 0.082 | 0.068 | 16.51 | 0.063 | 15.9 |

Table 9. Same of Table 8, but using only sources in common across all samples.

| | $ bias $ | σ | σ_{68} | σ_{NMAD} | η |
|--------------------|----------|----------|---------------|------------------------|--------|
| (1) <i>sdssVWI</i> | 0.0075 | 0.119 | 0.070 | 0.059 | 11.99 |
| (2) <i>sdssVW</i> | 0.0147 | 0.166 | 0.077 | 0.065 | 16.57 |
| (3) <i>sdssVI</i> | 0.0164 | 0.126 | 0.078 | 0.065 | 15.07 |
| (4) <i>sdssW</i> | 0.0197 | 0.188 | 0.078 | 0.066 | 14.97 |
| (5) <i>sdssI</i> | 0.0096 | 0.143 | 0.088 | 0.075 | 15.07 |
| (6) <i>sdssV</i> | 0.0257 | 0.233 | 0.110 | 0.090 | 21.37 |
| (7) <i>sdss</i> | 0.0287 | 0.224 | 0.107 | 0.085 | 22.79 |
| (8) <i>A17</i> | - | - | - | 0.059 | 11.96 |

We usually assess the accuracy of photo-z using a single value and we tacitly assume that two independent methods with similar statistics will provide the same photo-z value for a given source. This is clearly not the case, as shown in Fig. 7, where, for the sources in *sdssVWI* with spectroscopic redshift, we plot the difference between photo-z and z_{spec} for the redshifts computed with MLPQNA (in cyan) and LEPHARE (in red). It is interesting to note that only a small number of sources are simultaneously outliers in both methods. The majority are method-dependent, thus ruling out the possibility that these sources are outliers because they are peculiar objects (e.g. varying objects).

Fig. 5 has already shown how, by adding more bands, the overall accuracy improves and the fraction of outliers decreases. But, are the outliers just being reduced or are there sources that became outliers with the increasing of the photometric bands? To this question we answer in Fig. 8. There, for each combination of bands, we see that there are sources having the correct photo-z computed with a limited number of bands and that become outliers when more bands are considered. These are a small fraction of the sources, and they can be explained by either unstable photometry (i.e. photometry with large errors), or by a non-representative spectroscopic sample.

Finally, the left-hand panel of Fig. 9 shows, for each source with spectroscopic redshift, the details of the relative comparison between the photo-z computed via MLPQNA for the various photometric sets and LEPHARE from A17. Because all the photo-z incarnations used the same spectroscopic sample, the bulk of the sources in each plot lies on the one-to-one relation (red dotted lines). This is generally not the case when the same comparison is done using the sources for which no spectroscopic redshift is available (right-hand panel of Fig. 9). However, it is reassuring that, despite the small size of the sample, the photo-z computed with MLPQNA, using SDSSVWI sample, agree with those computed in A17. In contrast, the second column of the right-hand panel of Fig. 9 shows

how computing the photo-z using only SDSS creates an excess of sources at $z \sim 0.7$ and ~ 1.1 . The remaining comparison cases are reported, for completeness, in Appendix A (Figs A1 and A2).

To further test our accuracy, we decided to compare the results against the 257 new spectroscopic redshifts newly presented in LaMassa et al. (2019). Table 10 shows the photo-z accuracy for MLPQNA and LEPHARE for this sub-sample of objects. Once again, the noticeable difference in the accuracy and fraction of outliers obtained (see Tables 8 and 9 for the comparison) for every data set, points to the importance of the training sample that must be representative of the entire population, in type, but also in fraction to the total. This new sample covers the same redshift range as the original one, but it is dominated by fainter sources (see Fig. 2). The same issue affects, although marginally, also the photo-z computed with SED fitting. This is not surprising if one takes into account the fact that the templates, to be considered in the library, are determined by looking at the properties of the spectroscopic sample. However, in the case of SED-fitting the effect is minor, because only the SED type is considered and not its frequency in the sample.

7 PDZ ESTIMATION AND RELIABILITY

It is well known that SED-fitting algorithms tend to underestimate the error associated to the photo-z (Dahlen et al. 2013). But what about the PDZ?

The PDZ is a standard product of SED-fitting algorithms and can be used in assessing how reliable a photo-z and consequently fitted SEDs are. Larger the PDZ, more secure the redshifts. In addition, the PDZ is now routinely used in luminosity functions (e.g. Buchner et al. 2015; Miyaji et al. 2015; Fotopoulou et al. 2016, to quote just a few). The computation of the PDZ is a novelty in ML and here we want to test the reliability of the PDZ computed using METAPHOR with respect to the PDZ computed in A17 with LEPHARE.

For the analysis described below, it is important to understand how the PDZ is represented in the two methods. For LEPHARE, the range covered is pre-defined by the user. In A17, the photo-z was searched between redshift 0 and 7, in bins of 0.01 up to redshift 6. Between redshifts 6 and 7, the bin sizes are set to 0.02. Thus, for each source we have a file of 651 bins. For each source, the higher PDZ is normalized to 1. It can happen that the PDZ has non-zero values only in a limited range of bins, when the photo-z solution is well defined. In METAPHOR, although the range covered could be pre-defined by the user, it is recommended to set it within the extremes of the z_{spec} distribution of the training sample. This because METAPHOR, having an empirical method as internal engine, does not produce results outside the limits of z_{spec} . Therefore, in this case,

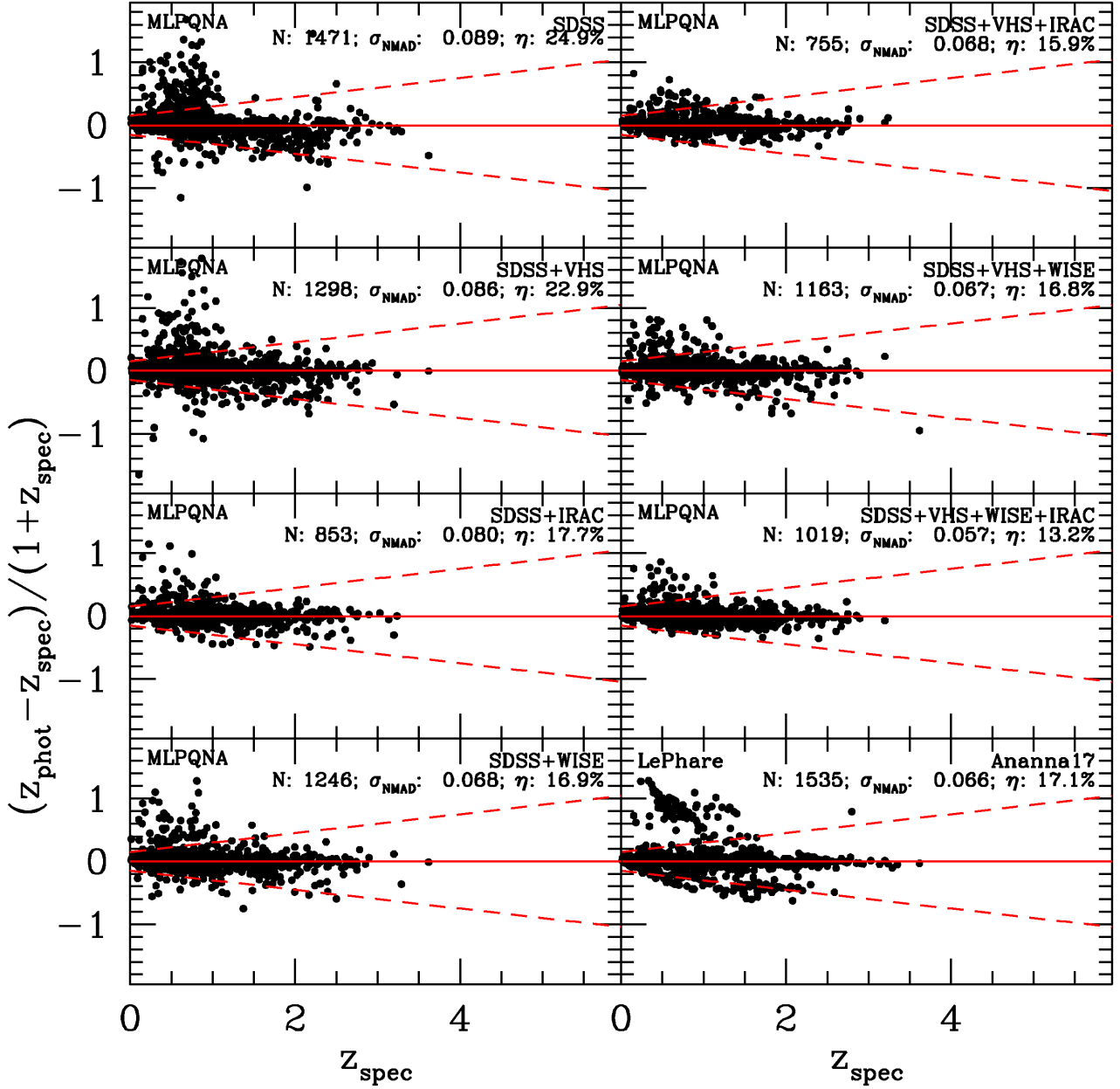


Figure 5. Comparison between spectroscopic redshift and photo-z for the sources cut at the eROSITA flux and divided on the basis of available photometric points. For comparison, the result from A17 is reported in the lower right panel of the figure. By comparing the accuracy and the fraction of outliers in every panel with the corresponding row in Table 8, we see that computing photo-z using only SDSS for bright X-ray sources is not recommended.

it has been set to $[0, 5.5]$ and since only one size of bins is allowed, we choose 0.01. It means that for each source we have a file of 551 bins. As for LEPHARE, also with METAPHOR, if the solution is well defined, the PDZ has non-zero values only in a limited range of bins.

As discussed in Amaro et al. (2018), a unique and universal method to evaluate the PDZ reliability is extremely difficult to find. One value often used is: $\text{PDZ_BEST} = \int_{z_{\text{best}}-0.1(1+z_{\text{best}})}^{z_{\text{best}}+0.1(1+z_{\text{best}})} \text{PDZ}(z) dz$ (cf. LEPHARE, Ilbert et al. 2006, documentation³). Here, in order to compare the results, we calculated PDZ_BEST also for METAPHOR. For illustrative purpose, Fig. 10 shows an example

of PDZ computed with METAPHOR and compares it with the PDZ computed with LEPHARE for the same random sources in the catalogue.

First, we tested how often the spectroscopic redshift is close to the peak of the PDZ. This is somewhat similar to what is done in Dahlen et al. (2013), where the fraction of sources with spectroscopic redshift is within 1σ or 3σ error from the photo-z. The analysis is shown in Table 11, where we split the cases in classes, with the following meaning:

- (1) z_{spec} is in the bin including the PDZ peak;
- (2) z_{spec} is in the bin beside to one including the PDZ peak;
- (3) z_{spec} is in a bin for which the value of PDZ is zero, but it is in a bin beside to the one containing the PDZ peak;

³<http://www.cfht.hawaii.edu/~arnouts/LEPHARE/DOWNLOAD/lephare.doc.pdf>

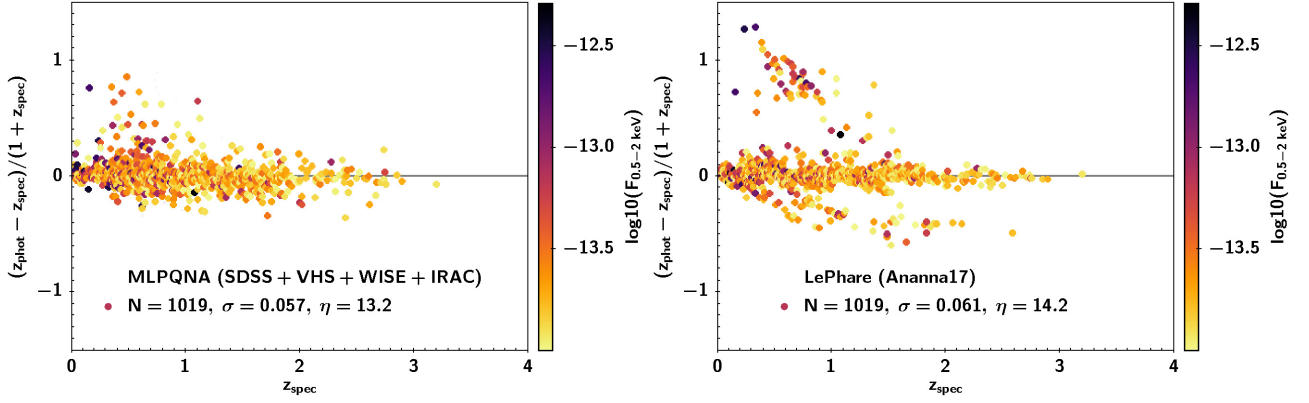


Figure 6. The same as the last two bottom right panels of Fig. 5, but this time limiting the sources from A17 to the same for which MLPQNA can compute photo-z using all the photometry available. The sources are colour coded as a function of their X-ray fluxes. With SED fitting, the outliers with overestimated redshift are basically X-ray bright.

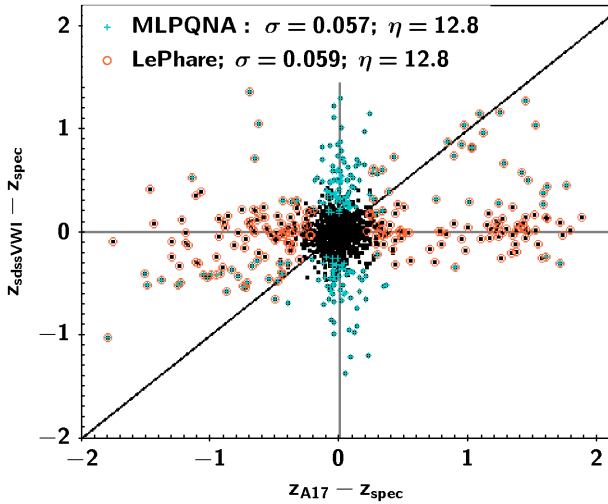


Figure 7. Difference between spectroscopic redshift and photo-z computed via MLPQNA and LEPHARE for the sub-sample of 1689 sources with SDSS, VHS, WISE, and IRAC photometry, regardless their X-ray flux. Sources that are outliers for MLPQNA (LEPHARE) are plot in cyan (orange). For this sub-sample, with complete photometry, the accuracy and fraction of outliers are very similar for the two methods. Nevertheless, the majority of the outliers are such only for one of the two algorithms. For the common outliers along the black one-to-one line, the two methods agree in terms of predicted photo-z.

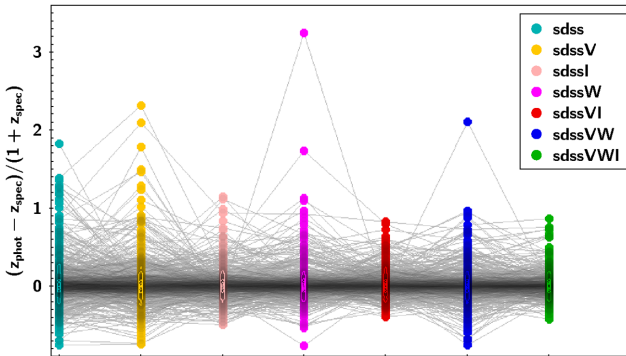


Figure 8. One-to-one comparison of accuracy for photo-z computed via MLPQNA with different combinations of photometry. For this plot only sources present in all the sub-samples have been used.

- (4) z_{spec} is in a bin for which the value of PDZ is more than zero (i.e. it contains classes 1 and 2);
- (5) z_{spec} is in a bin for which the value of PDZ is zero, but beside to a bin for which the value of PDZ is not zero;
- (6) z_{spec} is in a bin for which the value of PDZ is zero (contains classes 3 and 5).

For this test, we have considered only the sources with z_{spec} and for which the photometry in SDSS, VHS, WISE, and IRAC is always available. In this way, the two methods got access to the same information. From Table 11, it seems that METAPHOR is superior to the PDZ produced by LEPHARE in all the classes considered for the comparison. However, it is worth to keep in mind that the numbers presented in table are depending on the bin sizes chosen for the two experiments. We did not experiment with different bin sizes.

Second, in Fig. 11 we looked at the PDZ_BEST cumulative distribution for the sub-samples and compared with the results from A17, separating the sources with reliable photo-z (left-hand panel) from the outliers (right-hand panel).

Overall, METAPHOR is more conservative; while with LEPHARE, less than 10 per cent of the sources with spectroscopic redshift have $\text{PDZ_BEST} < 50$ per cent, the distribution for the sub-samples with PDZ computed by METAPHOR is completely different (32 per cent, 64 per cent, 98 per cent, 86 per cent, and 33 per cent for *sdss*, *sdssV*, *sdssW*, *sdssVW*, and *sdssVWI*, respectively).

The overconfidence of the PDZ computed by LEPHARE is even more evident in the right panel of Fig. 11, where we focus on the outliers. The various PDZ computed with METAPHOR are low for the large majority of the samples, with no PDZ computed at all for the outliers in *sdssVWI*. In contrast, LEPHARE assigns a $\text{PDZ_BEST} > 80$ per cent to the 42 per cent of its outliers.

This indicates that the PDZ computed via SED fitting, using only a limited number of broad-band photometric points, cannot be compared with the accuracy obtained in deep fields with many bands, where the precision of the PDZ is well tested (e.g. XMM-COSMOS Lusso et al. 2010). Here, with less than a dozen of photometric points, the PDZ_BEST is very high for the majority of the outliers. One can argue that the outliers are not more than 20 per cent overall. However, neither the outliers nor their photo-z are randomly distributed. Therefore, a too *optimistic* PDZ is more affecting certain regions of the mag/redshift/luminosity parameter space. In contrast, ML does not oversell the photo-z for the outliers and the PDZ_BEST remain very low in general.

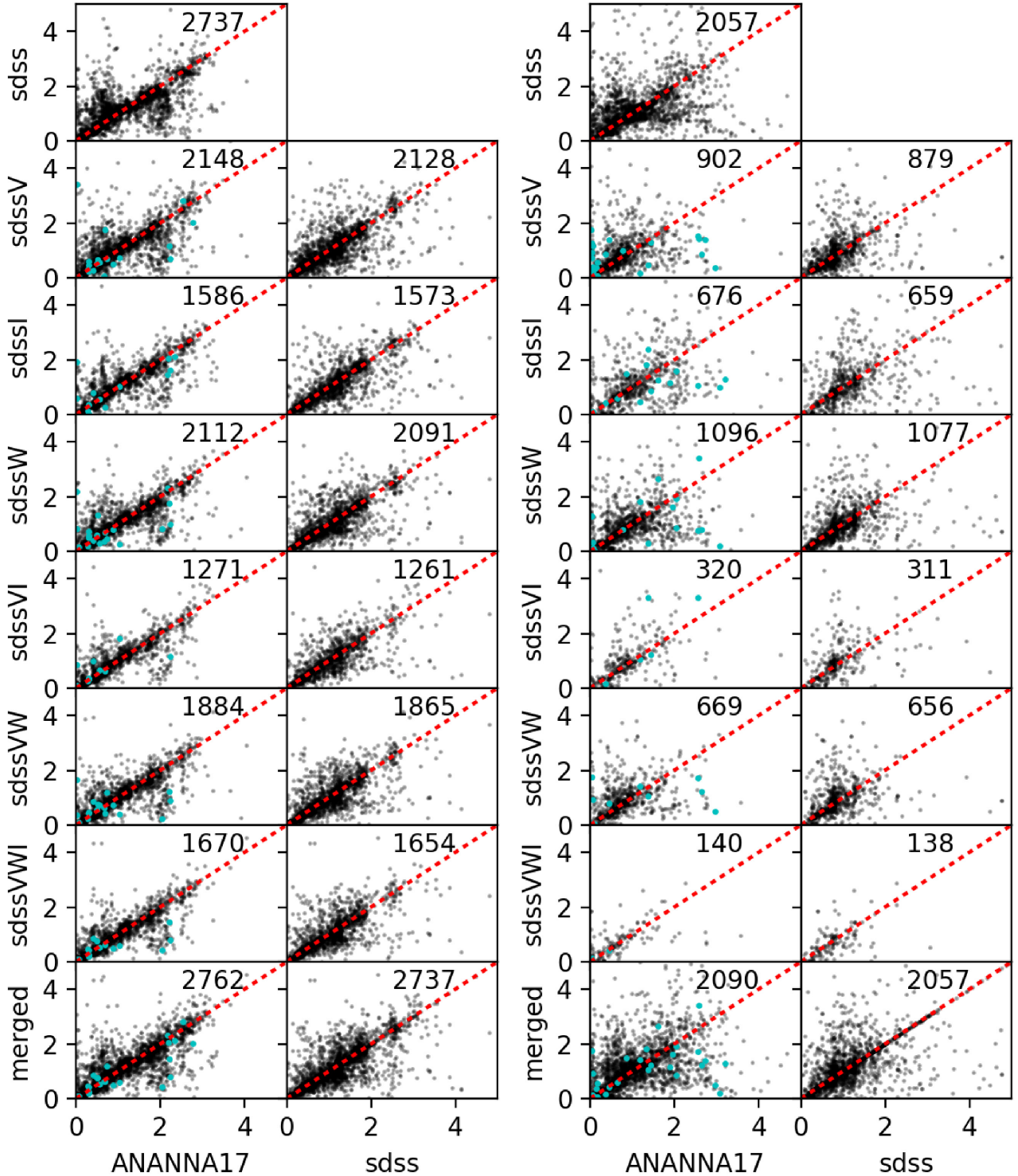


Figure 9. Details of the comparison between photo- z computed via SED fitting (A17) and MLPQNA for the sample for which spectroscopic information is, respectively, available (left-hand panel) and not available (right-hand panel). The cyan points indicate the sources for which the redshift could be computed only after considering supplementary photometry in addition to SDSS.

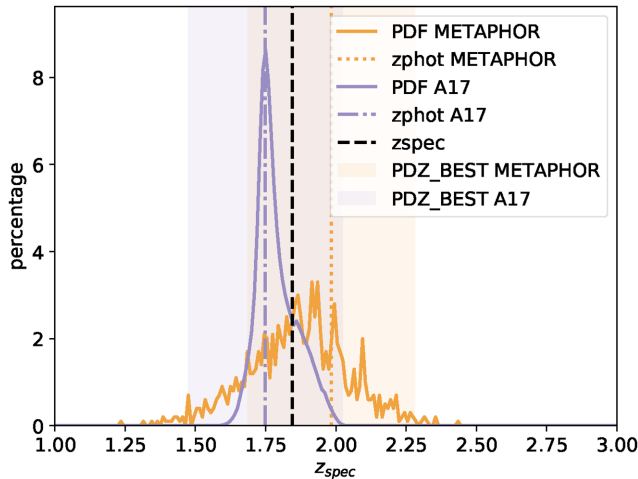
8 SUMMARY AND CONCLUSIONS

With the launch of *eROSITA*, we are facing the challenge of computing the photo- z in a reliable and fast manner for about 3 million sources distributed in the entire sky, with multiwavelength data that are non-homogeneous in depth and wavelength coverage.

Given that photo- z computed with ML are becoming the trend in cosmological surveys involving normal galaxies, we wanted to test whether this is a viable solution also for AGN. With this purpose in mind we have used the multiwavelength catalogue of the counterparts to the X-ray-selected sources detected in Stripe 82X (LaMassa et al. 2016), presented in A17. As ML method we have

Table 10. Summary of all statistical results for the new sample of 258 spectroscopic redshifts presented in LaMassa et al. 2019.

| | Number of sources | $ bias $ | σ | σ_{68} | σ_{NMAD} | η |
|--------------------------|-------------------|----------|----------|---------------|-----------------|--------|
| A17 | 258 | 0.0066 | 0.292 | 0.129 | 0.089 | 27.07 |
| sdss | 227 | 0.0037 | 0.367 | 0.158 | 0.129 | 33.48 |
| sdssV | 135 | 0.0357 | 0.322 | 0.211 | 0.149 | 41.48 |
| sdssW | 144 | 0.0073 | 0.288 | 0.173 | 0.137 | 34.03 |
| sdssI | 110 | 0.0119 | 0.202 | 0.184 | 0.163 | 40.91 |
| sdssVW | 111 | 0.0459 | 0.272 | 0.167 | 0.143 | 33.33 |
| sdssVI | 58 | 0.0343 | 0.255 | 0.161 | 0.116 | 32.76 |
| sdssVWI | 25 | 0.0298 | 0.151 | 0.152 | 0.104 | 32.00 |
| MLPQNA _{merged} | 229 | 0.0182 | 0.270 | 0.192 | 0.154 | 38.43 |

**Figure 10.** Example of PDZ obtained by METAPHOR and LEPHARE for the object *ID1431* (SDSS J221448.69 + 002508.7). The true redshift is represented by the black dashed line. The coloured areas represent the PDZ.BEST. In this specific case, the PDZs are both limited to the redshift range 1–3 for illustration purposes.**Table 11.** Distribution of sources with spectroscopy from A17 and in the sdssVWI sample, among classes defined as a function of the position of the redshifts within the PDZ.

| Class | METAPHOR (per cent) | A17 (per cent) |
|-------|---------------------|----------------|
| 1 | 5.4 | 2.8 |
| 2 | 14.0 | 7.8 |
| 3 | 0.0 | 0.1 |
| 4 | 99.9 | 96.8 |
| 5 | 0.0 | 0.3 |
| 6 | 0.1 | 3.2 |

tested MLPQNA (Brescia et al. 2013). The catalogue containing the photo- z computed for this paper, is released here. An excerpt of it is shown in Appendix A1.

We have compared our photo- z with those computed via SED fitting with LEPHARE (Arnouts et al. 1999; Ilbert et al. 2006), presented in A17. The main conclusions drawn from the comparison are:

(i) Before computing the photo- z with MLPQNA and ML in general, a feature analysis should be performed. Besides the obvious advantage of reducing the parameter space under analysis (i.e. to minimize the regression problem complexity), the feature selection mechanism aims also at finding an exhaustive subset of features,

able to maintain a high photo- z prediction accuracy, but avoiding any information redundancy occurrence and thus degeneracy in the results. The best features are not fixed but change, depending on the data set available. When only magnitudes are considered, the K band is by far the most important feature. The reason is easily understood keeping in mind the SED of a galaxy. The rest-frame K band indicates the knee of the SED and this clear feature can be used to determine the redshift. When also colours are available, the importance of single-band photometry is drastically reduced and the first four colours in Fig. 4 represents more than 63 per cent of the key feature relevance.

(ii) In this particular experiment, with a rich training sample able to represent the parent population and with optical, NIR, and MIR data available, the accuracy of the photo- z computed with MLPQNA is comparable to the accuracy obtained via SED fitting with LEPHARE. Comparable are also the fractions of outliers. When limiting the sample to the bright X-ray sources that eROSITA will detect (by comparing Table 8 and Fig. 6), MLPQNA performs slightly better (smaller fraction of outliers and absence of systematics). This is reassuring, given that these types of data are or will be available for the entire sky, thanks to the increasing depth of *WISE* (Schlafly et al. 2019) and the planned launch of *SpherEx* (Dore 2018).

(iii) Once that the training sample is large enough, the remaining limitation for ML algorithms is in the treatment of missing data. Currently, in our experiment, about 35 per cent are lacking a photo- z for this reason. However, the use of fluxes instead of magnitudes and adding the photometric errors to the list of parameters (using e.g. Reis et al. 2019) should solve the issue.

(iv) As for photo- z computed via SED fitting, the accuracy of photo- z can improve for AGN with small photometric error and for which information on the morphology (e.g. point-like or extended) is available.

(v) As mentioned already in literature, the completeness of the training sample is extremely crucial for ML algorithms. We can confirm that this is even more the case for AGN. It can be seen by comparing the accuracy obtained with the spectroscopic sample, available when this work started, with the accuracy obtained on the 257 new, fainter sources from LaMassa et al. (2019). The accuracy decreases and the fraction of outliers increases noticeably. The problem is not limited to ML, as also the results from A17 worsened for this new sample. As discussed in A17, the selection of the templates is based on the available spectroscopic sample. It means that irrespective of the method that will be used for computing the photo- z for eROSITA, we will need to make sure to define a training sample gathering from all the various surveys, (CDFs, COSMOS etc.), all the sources with X-ray properties typical of those that eROSITA will detect. This is similar to what Ruiz et al. (2018) has done for 3XMM.

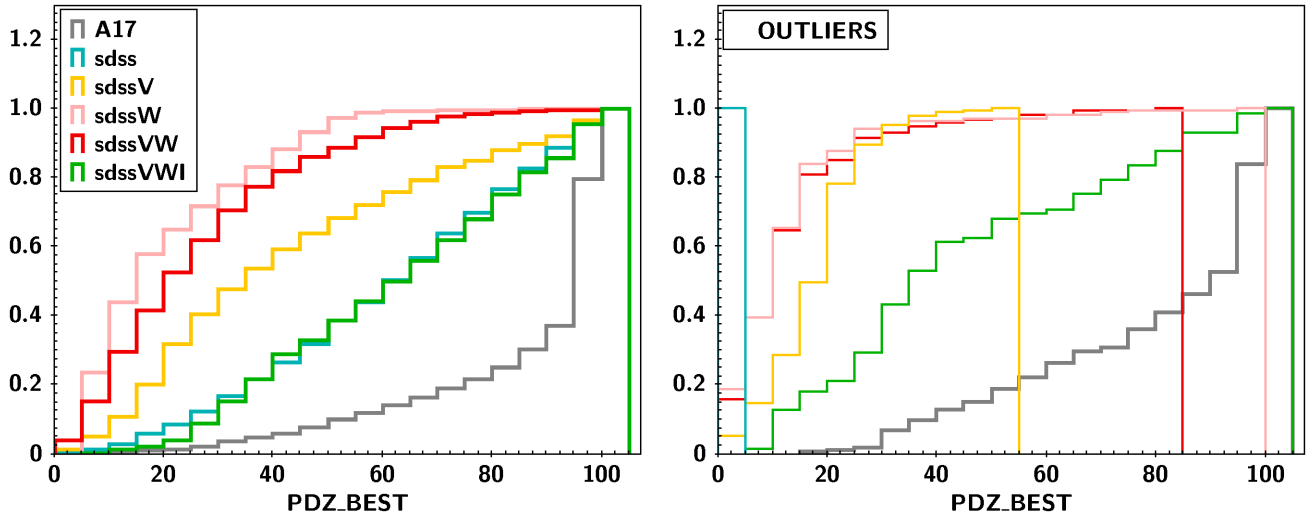


Figure 11. PDZ.BEST cumulative distribution for the entire spectroscopic sample (left) and for the outliers in the respective sub-samples used in this work (right), compared with the results from A17. The comparison is missing *sdssVI* and *sdssI*, only for brevity. While the majority of the sources have a high PDZ.BEST (e.g. larger than 85 percent) in A17, METAPHOR is much more conservative and only a handful of sources reach such value. Surprisingly, LEPHARE assigns a high reliability to the photo-*z* (high PDZ.BEST) also to the outliers, with more than 50 per cent of the outliers having PDZ.BEST > 80.

(vi) The PDZ computed by METAPHOR is reliable and tends to be in general more conservative than the one computed by LEPHARE. Even in the best case of complete multiwavelength coverage only for very few sources, the PDZ from METAPHOR is high. This is contrary to what happens with LEPHARE, where a high PDZ.BEST is obtained also for sources for which the photo-*z* is an outlier. While we recommended to always use the PDZ when working with photo-*z*, we also want to point out that in surveys covering wide areas with shallow data, the accuracy of the PDZ computed via SED fitting should not be taken for granted. It is important to underline that this is not the case for the PDZ provided, for example, in COSMOS, where not only the photo-*z* are reliable, but also the type 1/2/gal classification is mostly correct (e.g. Lusso et al. 2010). The limited reliability obtained here with PDZ computed via SED fitting is due to the photometry available only from broad-band filters, with a complete lack of photometry from intermediate- and narrow-band photometry (see Salvato et al. 2018, for a complete discussion). To be aware of the issue is important when, e.g. computing luminosity functions for sources with limited photometry. This was also noticed and pointed out in Buchner et al. (2015). There, the authors suggested a method to correct the PDZs, making them more realistic and showing how the correction needed was a factor of two larger for the survey with photo-*z* computed using only broad-band photometry (i.e. Aegis-X; Nandra et al. 2015), rather than for COSMOS (Salvato et al. 2011) and CDFS (Hsu et al. 2014), where also intermediate- and narrow-band photometry was available.

(vii) Assuming that a large and representative spectroscopic sample can be constructed for *eROSITA*, the only real limitation remains the multiwavelength coverage on the entire sky. Although many all-sky surveys exist (e.g. DES, SkyMapper, Pan-STARRS, VHS, AllWISE/unWISE, deep enough surveys, able to allow reliable results with ML, are still not available in all bands. Adopting Stripe 82X as reference, we expected that for *eROSITA*, at least for 2/3 of the final sample, reliable photo-*z* computed with ML can be obtained. SED-fitting will continue to provide more reliable results for the rest of the cases. The public catalogue with all our photo-*z* estimations is publicly released (see Appendix A for details.)

ACKNOWLEDGEMENTS

The authors thank the anonymous referee that with specific comments improved the manuscript. MB acknowledges the *INAF PRIN-SKA 2017 program 1.05.01.88.04* and the funding from MIUR Premiale 2016: MITIC. MB and GL acknowledge the H2020-MSCA-ITN-2016 SUNDIAL (*SURvey Network for Deep Imaging Analysis and Learning*), financed within the Call H2020-EU.1.3.1. SC acknowledges support from the project ‘Quasars at high redshift: physics and Cosmology’ financed by the ASI/INAF agreement 2017-14-H.O. TOPCAT (Taylor 2005) and STILTS (Taylor 2006) have been used for this work. This material is based upon work supported by the National Science Foundation under grant number AST-1715512. DAMEWARE has been used for ML experiments (Brescia et al. 2014a). c^3 has been used for catalogue cross-matching (Riccio et al. 2017). The SDSS Web site is <http://www.sdss.org/>. The scientific results reported in this article are based in part on data obtained from the *Chandra* and *XMM* Data Archive.

REFERENCES

- Aggarwal S. K., Saini L. M., Kumar A., 2009, *Int. J. Electr. Power Energy Syst.*, 31, 13
- Amaro V. et al., 2018, *MNRAS*, 482, 3116
- Ananna T. T. et al., 2017, *ApJ*, 850, 66
- Arnouts S., Cristiani S., Moscardini L., Matarrese S., Lucchin F., Fontana A., Giallongo E., 1999, *MNRAS*, 310, 540
- Biviano A. et al., 2013, *A&A*, 558, A1
- Boller T., Freyberg M. J., Trümper J., Haberl F., Voges W., Nandra K., 2016, *A&A*, 588, A103
- Bolzonella M., Miralles J.-M., Pelló R., 2000, *A&A*, 363, 476
- Bovy J. et al., 2012, *ApJ*, 749, 41
- Breiman L., 2001, *Mach. Learn.*, 45, 5
- Brescia M., Cavuoti S., D’Abrusco R., Longo G., Mercurio A., 2013, *ApJ*, 772, 140
- Brescia M. et al., 2014a, *PASP*, 126, 783
- Brescia M., Cavuoti S., Longo G., De Stefano V., 2014b, *A&A*, 568, A126
- Brescia M., Cavuoti S., Longo G., 2015, *MNRAS*, 450, 3893
- Brescia M., Cavuoti S., Amaro V., Riccio G., Angora G., Vellucci C., Longo G., 2018, *Commun. Comput. Inform. Sci.*, 822, 61

- Buchner J. et al., 2015, *ApJ*, 802, 89
- Budavári T. et al., 2001, *AJ*, 122, 1163
- Cardamone C. N. et al., 2010, *ApJS*, 189, 270
- Carrasco Kind M., Brunner R. J., 2013, in Friedel D. N., ed., ASP Conf. Ser. Vol. 475, Astronomical Data Analysis Software and Systems XXII
- Cavuoti S., Brescia M., Longo G., Mercurio A., 2012, *A&A*, 546, A13
- Cavuoti S., Brescia M., De Stefano V., Longo G., 2015a, *Exper. Astron.*, 39, 45
- Cavuoti S. et al., 2015b, *MNRAS*, 452, 3100
- Cavuoti S., Amaro V., Brescia M., Vellucci C., Tortora C., Longo G., 2017, *MNRAS*, 465, 1959
- D’Isanto A., Cavuoti S., Gieseke F., Polsterer K., 2018, *A&A*, 616, A97
- Dahlen T. et al., 2013, *ApJ*, 775, 93
- Dawson K. S. et al., 2013, *AJ*, 145, 10
- de Jong J. T. A. et al., 2017, *A&A*, 604, A134
- Delli Veneri M., Cavuoti S., Brescia M., Longo G., Riccio G., 2019, *MNRAS*, 486, 1377
- Delubac T. et al., 2017, *MNRAS*, 465, 1831
- Dore O., 2018, in 42nd COSPAR Scientific Assembly. p. E1.16
- Duncan K. J., Jarvis M. J., Brown M. J. I., Röttgering H. J. A., 2018, *MNRAS*, 477, 5177
- Ebrahimzadeh A., Khazaei A., 2010, *Measurement*, 43, 103
- Fliri J., Trujillo I., 2016, *MNRAS*, 456, 1359
- Fotopoulou S., Paltani S., 2018, *A&A*, 619, A14
- Fotopoulou S. et al., 2012, *ApJS*, 198, 1
- Fotopoulou S. et al., 2016, *A&A*, 587, A142
- Georgakakis A. et al., 2014, *MNRAS*, 443, 3327
- Gheys I. A., Smith L. S., 2010, *Pattern Recogn.*, 43, 5
- Guyon I., Elisseeff A., 2003, *J. Mach. Learn. Res.*, 3, 1157
- Hastie T., Tibshirani R., Friedman J., 2009, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, Second Edition. Springer Series in Statistics. Springer, New York
- Hsu L.-T. et al., 2014, *ApJ*, 796, 60
- Ilbert O. et al., 2006, *A&A*, 457, 841
- Ilbert O. et al., 2009, *ApJ*, 690, 1236
- Irwin M. J. et al., 2004, in Quinn P. J., Bridger A., eds, *Proc. SPIE Vol. 5493, Optimizing Scientific Return for Astronomy through Information Technologies*, SPIE, Bellingham, p. 411
- Ivezić Ž. et al., 2019, *ApJ*, 873, 111
- Jolliffe I. T., 2002, *Principal Component Analysis*. Springer, New York
- Kohavi R., 1995, in *Proceedings of the 14th International Joint Conference on Artificial Intelligence - Volume 2 (IJCAI’95)*. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, p. 1137
- Kohavi R., John G. H., 1997, *Artif. Intell.*, 97, 273
- Kursa M., Rudnicki W., 2010, *J. Stat. Softw.*, Articles, 36, 1
- Laigle C. et al., 2016, *ApJS*, 224, 24
- Lal T. N., Chapelle O., Weston J., Elisseeff A., 2006, *Embedded Methods*. Springer, Berlin Heidelberg, p. 137
- LaMassa S. M. et al., 2013a, *MNRAS*, 432, 1351
- LaMassa S. M. et al., 2013b, *MNRAS*, 436, 3581
- LaMassa S. M. et al., 2016, *ApJ*, 817, 172
- LaMassa S. M., Georgakakis A., Vivek M., Salvato M., Tasnim Ananna T., Urry C. M., MacLeod C., Ross N., 2019, *ApJ*, 876, 50
- Laureijs R., 2010, in *Joint European and National Astronomy Meeting (JENAM’10)*. p. 166
- Lawrence A. et al., 2007, *MNRAS*, 379, 1599
- Luo B. et al., 2010, *ApJS*, 187, 560
- Lusso E. et al., 2010, *A&A*, 512, A34
- Marchesi S. et al., 2016, *ApJ*, 817, 34
- Martin D. C. et al., 2005, *ApJ*, 619, L1
- Matute I. et al., 2012, *A&A*, 542, A20
- Merloni A. et al., 2012, preprint ([arXiv:1209.3114](https://arxiv.org/abs/1209.3114))
- Meshcheryakov A. V., Glazkova V. V., Gerasimov S. V., Mashechkin I. V., 2018, *Astron. Lett.*, 44, 735
- Miyaji T. et al., 2015, *ApJ*, 804, 104
- Mostafa M. M., 2010, *Expert Syst. Appl.*, 37, 6302
- Mountrichas G., Corral A., Masoura V., Georgantopoulos I., Ruiz A., Georgakakis A., Carrera F., Fotopoulou S., 2017, *A&A*, 608,
- Nandra K. et al., 2015, *ApJS*, 220, 10
- Nicastro F. et al., 2018, *Nature*, 558, 406
- Papovich C. et al., 2016, *ApJS*, 224, 28
- Pâris I. et al., 2018, *A&A*, 613, A51
- Pearson K., 2010, *London Edinburgh Dublin Philos. Mag. J. Sci.*, 2, 559
- Polsterer K. L., Gieseke F., Igel C., Goto T., 2014, in Manset N., Forshay P., eds, ASP Conf. Ser. Vol. 485, Astronomical Data Analysis Software and Systems XXIII. Astron. Soc. Pac., San Francisco, p. 425
- Reis I., Baron D., Shahaf S., 2019, *AJ*, 157, 16
- Riccio G., Brescia M., Cavuoti S., Mercurio A., di Giorgio A. M., Molinari S., 2017, *PASP*, 129, 024005
- Rosenblatt F., 1962, *Principles of Neurodynamics; Perceptrons and the Theory of Brain Mechanisms*. Spartan Books, Washington
- Ruiz A., Corral A., Mountrichas G., Georgantopoulos I., 2018, *A&A*, 618, A52
- Sadeh I., Abdalla F. B., Lahav O., 2016, *PASP*, 128, 104502
- Salvato M. et al., 2009, *ApJ*, 690, 1250
- Salvato M. et al., 2011, *ApJ*, 742, 61
- Salvato M., Ilbert O., Hoyle B., 2018, *Nat. Astron.*, 3, 212
- Schlaflly E. F., Meisner A. M., Green G. M., 2019, *ApJS*, 240, 30
- Tanaka M., 2015, *ApJ*, 801, 20
- Tangaro S. et al., 2015, *Computational and Mathematical Methods in Medicine*, Vol. 2015, Hindawi Publishing Corporation, p. 814104
- Taylor M. B., 2005, in Shopbell P., Britton M., Ebert R., eds, ASP Conf. Ser. Vol. 347, Astronomical Data Analysis Software and Systems XIV. Astron. Soc. Pac., San Francisco, p. 29
- Taylor M. B., 2006, in Gabriel C., Arviset C., Ponz D., Enrique S., eds, ASP Conf. Ser. Vol. 351, Astronomical Data Analysis Software and Systems XV. Astron. Soc. Pac., San Francisco, p. 666
- Tibshirani R. J., 2013, *Electron. J. Statist.*, 7, 1456
- Timlin J. D. et al., 2016, *ApJS*, 225, 1
- Vanzella E. et al., 2004, *A&A*, 423, 761
- Voges W. et al., 1999, *A&A*, 349, 389
- Wright E. L. et al., 2010, *AJ*, 140, 1868
- Zare M., Pourghasemi H. R., Vafakhah M., Pradhan B., 2013, *Arab. J. Geosci.*, 6, 2873

SUPPORTING INFORMATION

Supplementary data are available at *MNRAS* online.

PHOTOZ_ML_for_PAPER_2019JUL24.dat

Please note: Oxford University Press is not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.

APPENDIX A: PHOTO-Z COMPARISON AND PUBLIC CATALOGUE

A1 Catalogue release

The produced catalogue of photo-z, obtained by different cross-matches among available surveys as well as their final best combination, is made publicly available via the CDS VizieR facility. A sample of the internal structure is shown in Table A1. The catalogue is indexed on the first column, which can be used to retrieve all other information about spectroscopic redshifts and X-ray source counterparts, by cross-matching this catalogue with the one referred in A17. The other columns, from left to right, are respectively, RA and Dec. of optical counterparts, followed by the photo-z estimations obtained by all discussed combinations of surveys, i.e. SDSS, VHS, IRAC, and WISE. The last column is related to the best photo-z obtained from all the previous combinations, as explained in the main text.

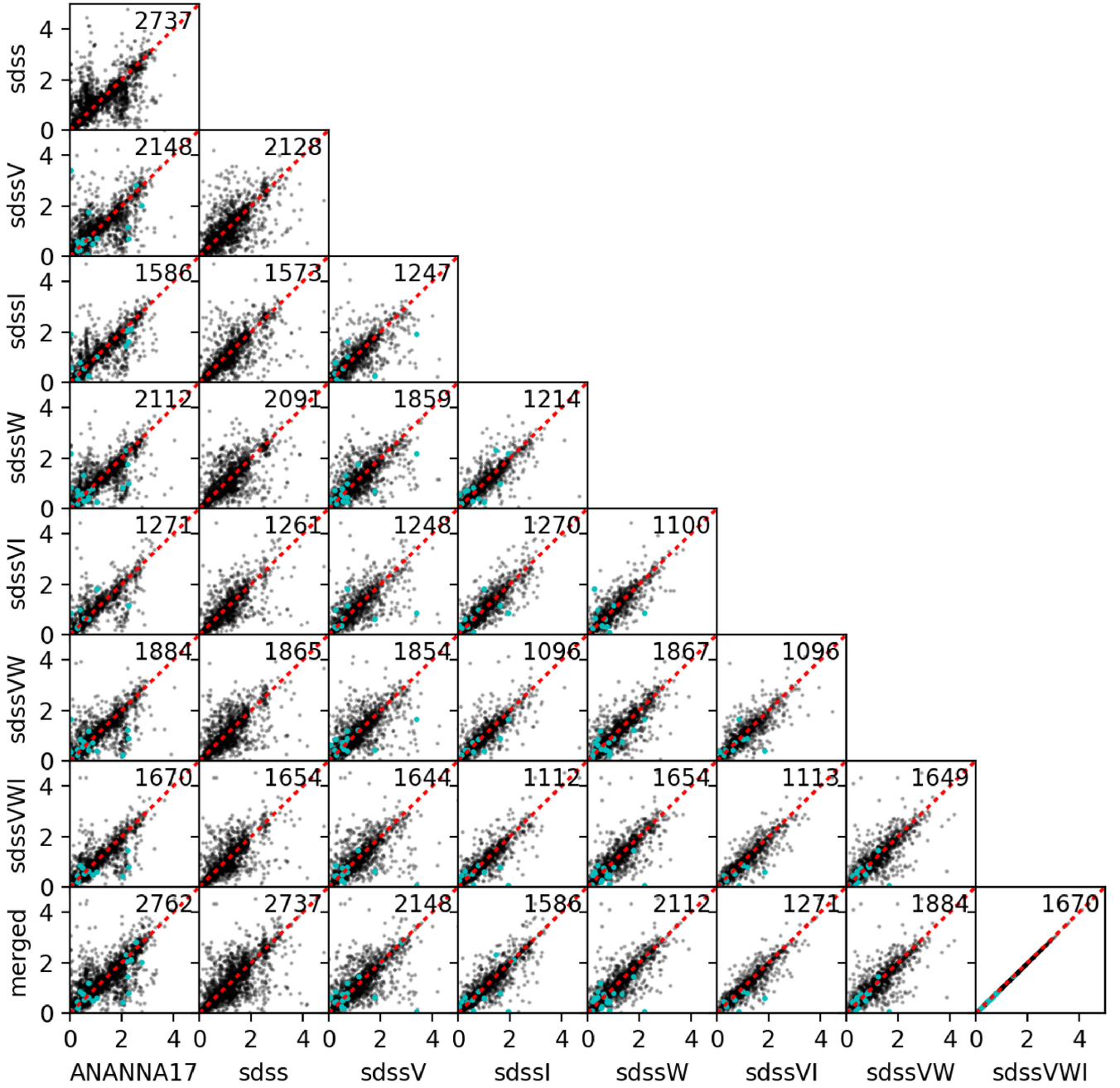


Figure A1. Comparison between photo- z computed via SED fitting (A17) and MLPQNA for the sample for which spectroscopic information is available. The cyan points indicate the sources for which the redshift could be computed only after considering supplementary photometry in addition to SDSS.

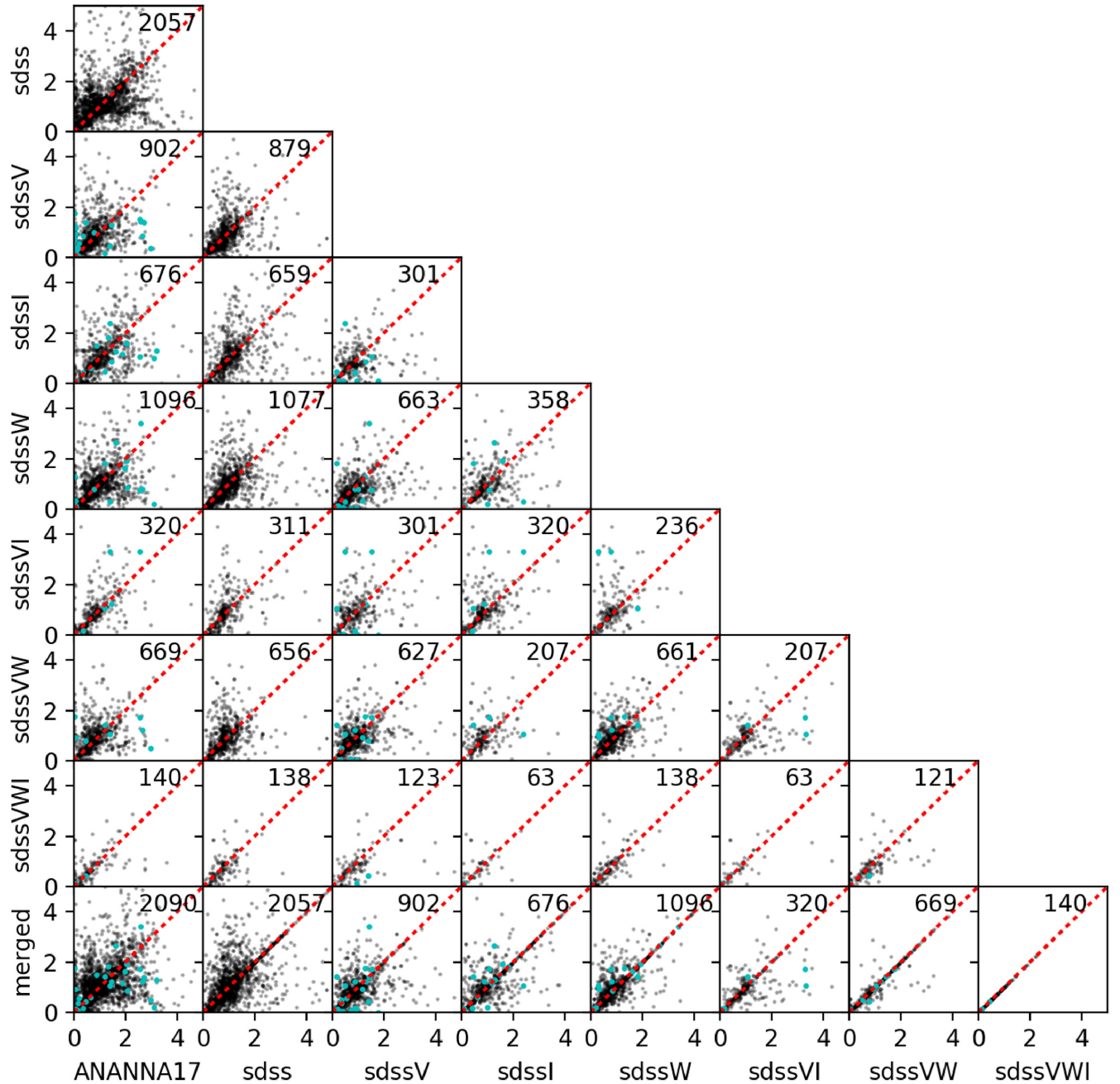


Figure A2. Comparison between photo- z computed via SED fitting (A17) and MLPQNA for the sample for which spectroscopic information is not available. The cyan points indicate the sources for which the redshift could be computed only after considering supplementary photometry in addition to SDSS.

Table A1. Example of contents of the photo- z catalogue made publicly available.

| REC_NO | CTP_RA | CTP_DEC | sdssVWI | sdssVW | sdssVI | sdssW | sdssI | sdssV | sdss | MLPQNA_merged |
|--------|-----------|-----------|---------|----------|--------|---------|-------|---------|---------|---------------|
| 1 | 0.980191 | 0.2046045 | 1.06487 | 1.11284 | −99.0 | 1.0867 | −99.0 | 1.16123 | 1.10542 | 1.06487 |
| 2 | 0.9812111 | 0.1268089 | 1.19282 | 0.955105 | −99.0 | 0.98219 | −99.0 | 1.10901 | 1.14375 | 1.19282 |
| 3 | 0.9939954 | 0.0559024 | −99.0 | −99.0 | −99.0 | −99.0 | −99.0 | −99.0 | 1.44947 | 1.44947 |
| 4 | 1.0114936 | 0.1948067 | −99.0 | 0.194302 | −99.0 | 0.10898 | −99.0 | 0.25149 | 0.16807 | 0.194302 |
| 5 | 1.0116661 | 0.1634697 | −99.0 | −99.0 | −99.0 | −99.0 | −99.0 | −99.0 | −99.0 | −99.0 |

This paper has been typeset from a \LaTeX file prepared by the author.