

# Improving the reliability of photometric redshift with machine learning

Oleksandra Razim <sup>1</sup>★, Stefano Cavuoti <sup>1,2</sup>, Massimo Brescia <sup>2</sup>, Giuseppe Riccio <sup>2</sup>, Mara Salvato <sup>3</sup>  
and Giuseppe Longo<sup>1</sup>

<sup>1</sup>Department of Physics, University Federico II, Strada Vicinale Cupa Cintia, 21, I-80126 Napoli, Italy

<sup>2</sup>INAF – Astronomical Observatory of Capodimonte, Salita Moiarillo 16, I-80131 Napoli, Italy

<sup>3</sup>MPI for Extraterrestrial Physics, Giessenbachstrasse 1, D-85748 Garching b. München, Germany

Accepted 2021 July 30. Received 2021 July 21; in original form 2020 November 11

## ABSTRACT

In order to answer the open questions of modern cosmology and galaxy evolution theory, robust algorithms for calculating photometric redshifts (photo- $z$ ) for very large samples of galaxies are needed. Correct estimation of the various photo- $z$  algorithms' performance requires attention to both the performance metrics and the data used for the estimation. In this work, we use the supervised machine learning algorithm MLPQNA (Multi-Layer Perceptron with Quasi-Newton Algorithm) to calculate photometric redshifts for the galaxies in the COSMOS2015 catalogue and the unsupervised Self-Organizing Maps (SOM) to determine the reliability of the resulting estimates. We find that for  $z_{\text{spec}} < 1.2$ , MLPQNA photo- $z$  predictions are on the same level of quality as spectral energy distribution fitting photo- $z$ . We show that the SOM successfully detects unreliable  $z_{\text{spec}}$  that cause biases in the estimation of the photo- $z$  algorithms' performance. Additionally, we use SOM to select the objects with reliable photo- $z$  predictions. Our cleaning procedures allow us to extract the subset of objects for which the quality of the final photo- $z$  catalogues is improved by a factor of 2, compared to the overall statistics.

**Key words:** methods: data analysis – techniques: spectroscopic – surveys – galaxies: distances and redshifts – catalogues –

## 1 INTRODUCTION

Knowledge of spatial distribution of galaxies is crucial for answering most of the open questions in observational cosmology and galaxy evolution. In particular, we need to know galaxy distances. The universal method of deriving them is to measure redshifts of galaxies, caused by the cosmological expansion. The most accurate way of doing that is to calculate spectroscopic redshifts (spec- $z$ , or  $z_{\text{spec}}$ ), but this method requires high-quality spectra for all investigated objects. Obtaining these spectra is time consuming and sometimes impossible for the faint sources. Currently, wide-field galaxy surveys cannot provide spec- $z$  catalogues of the same depth and richness as photometric catalogues. For example, for the 16th data release of the Sloan Digital Sky Survey (SDSS) the ratio between the number of detected galaxies and the number of available spectroscopic redshifts is of the order of 1/100,<sup>1</sup> going down to even smaller numbers for other surveys.

This difficulty of deriving galaxy redshifts from spectra led to the development of alternative techniques, collectively called *photometric redshift methods* (photo- $z$ , or  $z_{\text{phot}}$ ), first proposed in Baum (1957, 1962) and further investigated in a vast number of publications, e.g. Baldwin 1977; Butchins 1981; Koo 1985; Connolly et al. 1995; Gwyn & Hartwick 1996; Bolzonella, Miralles & Pelló 2000; Benítez 2000; Collister & Lahav 2004; Ball et al. 2008; Brammer, van Dokkum & Coppi 2008; Gerdes et al. 2010; Carrasco Kind & Brunner 2013, 2014; Hoyle 2016; Bonnett et al. 2016; Sadeh,

Abdalla & Lahav 2016; Bilicki et al. 2018; Pasquet et al. 2019, and many others.

These methods are based on the fact that multiband photometry can be treated as a low-resolution spectrum and hence as an approximation of the intrinsic spectral energy distribution (SED). Cosmological redshift causes a stretching of the SED and a change in the wavelengths of the prominent spectral features (Lyman or Balmer breaks, emission and absorption lines, etc.) that, moving in or out of the fixed bands of a given photometric system, cause changes in magnitudes and colours. From these changes, we can derive photo- $z$ .

There is a variety of photo- $z$  estimation methods. The majority of them use one of the two main approaches: a theoretical one, called SED template fitting, and an empirical one, largely based on machine learning (ML) paradigms (Euclid Collaboration 2020; Schmidt et al. 2020).

SED template fitting methods use spectral template libraries obtained from either observations or galaxy models. These templates are shifted across the expected redshift range (using an arbitrarily chosen step) and then convoluted with the transmission curves of the acquisition system in order to derive estimated magnitudes. The resulting magnitudes are then fitted to the observed data in order to find the combination of a template and redshift value that minimize the residuals. Depending on the particular approach, template fitting methods may include such steps as producing new templates via interpolation between the already existing ones, correcting systematic biases of the observed photometry, compensating for the absorption of the interstellar medium, etc. A downside of these methods is that they require template libraries that are conditioned by our knowledge of galaxy evolution and therefore by definition incomplete.

\* E-mail: [shr.razim@gmail.com](mailto:shr.razim@gmail.com)

<sup>1</sup>Based on the sizes of photometric and spectroscopic catalogues obtained via <http://skyserver.sdss.org/CasJobs/>.

Instead of using pre-defined templates, ML derives the correlation between photometry and redshift from the data themselves. To train the ML models, we need a knowledge base (KB), formed by both photometry and high-quality spec- $z$  for a considerable number of objects (how large depends on the homogeneity of the sample under scrutiny). With a KB large enough and fully covering the parameter space, the ML approach allows us to take into account all observational and physical effects automatically (Brescia et al. 2018). At the same time, providing a high-quality KB is not a trivial task by itself. For instance, due to the fact that ML methods have a limited extrapolating power, they cannot provide reliable photo- $z$  predictions outside the spectroscopic range covered by the KB. This is particularly relevant in the faint object domain, usually not properly covered by spectroscopic surveys.

To provide best performance, SED fitting and ML require different types of data, and for this reason the areas of application of the two methods are slightly different (Salvato, Ilbert & Hoyle 2019). In the low-redshift regime, ML photo- $z$  often perform better than SED fitting methods, provided that a well-representative spectroscopic KB exists. For high-redshift objects, SED template fitting usually shows better results, mostly due to the lack of a consistent KB for these galaxies. Furthermore, in the case of severe depth imbalance between spectroscopic and photometric information, SED fitting can be the only option.

Lately, hybrid photo- $z$  techniques that unite strengths of both approaches have appeared (see e.g. Newman 2008; Beck et al. 2016; Cavuoti et al. 2017b; Duncan et al. 2018). Still, at this point no method allows to estimate photo- $z$  with the same precision as high-quality spec- $z$ . Currently, the accuracy of spectroscopic redshift surveys can reach an error of  $\sigma \sim 10^{-3}$ , where  $\sigma$  is a standard deviation of the distribution of residuals between repeated measurements of spec- $z$  of the same objects (e.g. Le Fèvre et al. 2005; Hasinger et al. 2018). The accuracy of photo- $z$  catalogues obtained with broad-band photometry is characterized by  $\sigma \sim 0.02$  at best (LSST Science Collaboration 2009; Laureijs et al. 2011; Brescia et al. 2014; Salvato et al. 2019), where  $\sigma$  describes the distribution of residuals between spec- $z$  and photo- $z$ . Nevertheless, for the upcoming massive surveys, such as the Rubin Observatory Legacy Survey of Space and Time (LSST; Ivezić et al. 2019) and the *Euclid* (Laureijs et al. 2011), deriving photo- $z$  is the only realistic way to obtain distances for the majority of the observed galaxies, so further improvements of photo- $z$  methods are required.

Among the many factors that affect the quality of the photo- $z$  are the issues related to the completeness and quality of the spectroscopic catalogues used for SED fitting calibration and ML training. First, incompleteness of spec- $z$  sample limits the performance of photo- $z$  algorithms, especially the ML ones. Usually, spec- $z$  catalogues are incomplete in the faint part of the magnitude parameter space, but the selection function of the spec- $z$  observations may affect the performance in the whole range of magnitudes. Secondly, the miscalculated spec- $z$  affect the reliability of the performance metrics used for the estimation of quality of photo- $z$ . Some percentage of miscalculated or misidentified sources are present even in the spec- $z$  samples that are believed to be reliable. Typical high-confidence redshift quality flags provide us with 95–99 per cent of reliable sources, implying that 1–5 per cent of sources have unreliable spec- $z$ . Consequently, there is no way to say which of the photo- $z$  outliers are due to miscalculated photo- $z$  and which are related to incorrect spec- $z$  values.

For small-size surveys, the objects with noticeable difference between spec- $z$  and photo- $z$  can be manually inspected (see e.g. Lanzetta, Fernández-Soto & Yahil 1998 and Masters et al. 2017),

but for large-scale surveys this is not feasible. As a result, we have to find a way of disentangling the various contributions to the error budget, i.e. to distinguish the photo- $z$  prediction error induced by the defects of an estimation method from the uncertainty carried by the contamination of the spectroscopic data.

In this work, we present a data cleaning methodology, focused on handling these two data-related issues, namely

- (i) identifying unreliable spec- $z$ ;
- (ii) deriving from the photometric catalogue a set of objects that can be considered as well represented by the spec- $z$  catalogue, and therefore trusted in terms of photo- $z$  quality estimations.

As we will show, both procedures allow us to significantly improve the quality of the final photo- $z$  catalogues.

Our methodology is based on the Self-Organizing Map (SOM) algorithm, first proposed in Kohonen (1982), and on an approach proposed in Masters et al. (2015, 2017). For the photo- $z$  calculation, we use 30-band photometric data from the COSMOS2015 catalogue (Laigle et al. 2016), the master spec- $z$  catalogue provided by COSMOS collaboration (obtained via private communication with Mara Salvato), and the Deep Imaging Multi-Object Spectrograph (DEIMOS) spec- $z$  catalogue (Hasinger et al. 2018). To obtain photo- $z$ , we apply the Multi-Layer Perceptron with Quasi-Newton Algorithm (MLPQNA; Brescia et al. 2013, 2014). For additional testing of our data cleaning methodology, we also use the SED template fitting photo- $z$  catalogue described in Laigle et al. (2016).

This paper is organized as follows: Section 2 explains how we obtain and pre-process our data. In Section 3, we describe our algorithms, and in Section 4 we describe the experiments on calculating the photo- $z$  and improving their quality with the SOM. Section 5 contains the discussion of the results. The conclusions are stated in section 6. The Jupyter notebooks necessary to reproduce the work are available for download via GitHub repository [https://github.com/ShrRa/COSMOS\\_SOM](https://github.com/ShrRa/COSMOS_SOM). MLPQNA is available as a part of the PHOTRAPTOR (PHOTometric Research Application To Redshifts) software (Cavuoti et al. 2015a). The final photo- $z$  catalogue will be published via CDS Vizier facility.

## 2 DATA

In this section, we describe the following three catalogues used in this work:

- (i) COSMOS2015, which is our source of photometric measurements and SED fitting photo- $z$ . Its detailed description can be found in Laigle et al. (2016).
- (ii) A compilation of spectroscopic redshifts available in literature for the same COSMOS field. We will call this catalogue our main spec- $z$  catalogue.
- (iii) The DEIMOS spec- $z$  catalogue (Hasinger et al. 2018), used to perform an additional, independent test of our methods. Same as the main spec- $z$  catalogue, the DEIMOS sample covers most of the COSMOS2015 field.

### 2.1 COSMOS2015 photometric catalogue

The COSMOS2015 catalogue contains multiwavelength broad-range (from mid-IR to near-UV) photometry for about half million objects, for which X-ray and radio measurements, star formation rates, and other additional information are available. Moreover, this catalogue includes SED fitting photo- $z$ s, obtained with the LEPHARE software (Arnouts et al. 1999; Ilbert et al. 2006).

One of the most important aspects to consider in any ML-based experiment is the selection of the training parameter space, i.e. input features that maximize the amount of suitable information. In the case of ML methods for photo- $z$ , this problem has been widely discussed (cf. D’Isanto et al. 2018; Brescia et al. 2018, and references therein). From the ML point of view, an automatic feature selection is not a trivial task: Many algorithms for feature selection exist and have been tested on the problem (see Donalek et al. 2013; Baron 2019; Brescia et al. 2019), as well as some data-driven brute force approaches (Polsterer et al. 2014; D’Isanto et al. 2018), which may provide optimal results at the price of very high computational costs, making it impossible to use it on massive and high-dimensional data sets.

In this work, we do not perform an all-encompassing feature selection investigation. Instead, we try several feature sets, based on general physical considerations and our past experience. To try these different feature sets, we downloaded the COSMOS2015 photometric data consisting of 34 bands:

- (i) UV broad and medium bands: *NUVmag*, *FUVmag*, and *u*;
- (ii) optical and near-IR broad-bands: *B*, *V*, *ip*, *r*, *zp*, *zpp*, *Ks*, *Y*, *H*, *J*, *Hw*, *Ksw*, and *yH*;
- (iii) mid-IR broad-bands: *3.6mag*, *4.5mag*, *5.8mag*, and *8.0mag*;
- (iv) optical and near-IR medium bands: *IA484*, *IA527*, *IA624*, *IA679*, *IA738*, *IA767*, *IB427*, *IB464*, *IB505*, *IB574*, *IB709*, and *IB827*;
- (v) and optical and near-IR narrow bands: *NB711* and *NB816*.

For all of them, 2 and 3 arcsec apertures were available, except for *Spitzer* Large Area Survey with Hyper-Suprime-Cam (SPLASH; *3.6mag*, *4.5mag*, *5.8mag*, and *8.0mag*) and *GALEX* (*NUVmag* and *FUVmag*) bands.

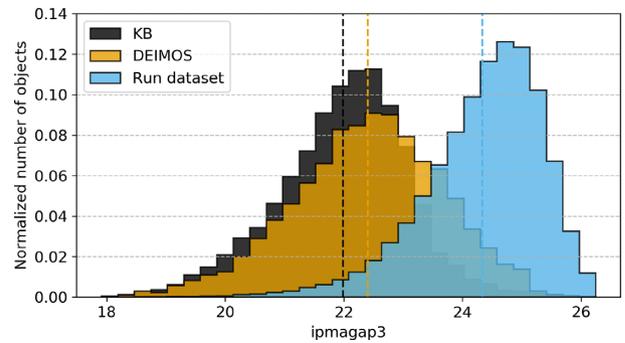
The pre-processing of this data set consisted of the following steps:

- (i) To ensure homogeneity of photometry and to enable a comparison with SED fitting photo- $z$ , we follow the procedure described in Laigle et al. (2016). We consider only those objects that lie within both UltraVISTA (this is done by using the condition `Area==0`) and COSMOS (`Cf1==1`) sky areas (see McCracken et al. 2012, Capak et al. 2007, and Scoville et al. 2007 for a detailed description of these regions). As a result, we extract 576 762 objects out of the initial 1182 108 samples available.
- (ii) We also exclude stars, X-ray, and unclassified sources (`OType==0`). This leaves us with 551 538 objects.
- (iii) Finally, we remove saturated sources, by rejecting objects masked in optical broad-bands (`Sat==0`). After this step, the final photometric catalogue consists of 518 404 objects.

## 2.2 Main spectroscopic catalogue

The main spec- $z$  catalogue is extracted from the spectroscopic COSMOS master catalogue maintained within the COSMOS collaboration and it includes only the publicly available redshifts prior Fall 2017. This catalogue contains 65 426 spectral redshifts, obtained with 27 different instruments in a spec- $z$  range of  $0 < z_{\text{spec}} < 6.5$ . The pre-processing consisted of the following steps:

- (i) To exclude stars, we remove sources with  $z_{\text{spec}} < 0.01$ ; we also remove objects with  $z_{\text{spec}} > 9$  to discard likely erroneous spec- $z$ ;
- (ii) Active galactic nuclei (AGNs) often pose a contamination problem for photo- $z$  algorithms (Norris et al. 2019). Therefore, we remove sources visible in X-ray, using a catalogue of AGN sources detected by *Chandra* in the COSMOS field (Civano et al. 2012).
- (iii) Then, we clean the resulting spec- $z$  catalogue from unreliable instances, using the available quality flags  $Q_f$ , described in Lilly et al. (2009). We select only robust spectroscopic redshifts (i.e. with



**Figure 1.** Normalized magnitude distributions for the KB, DEIMOS, and run data sets in the *ipmagap3* after standard cleaning but before any SOM filtering. Dashed lines show the mean value of the distribution.

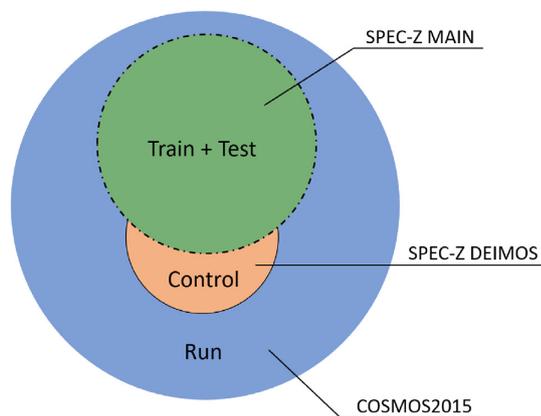
$\sim 99.6$  per cent of spectroscopic verification), using the conditions  $2 < Q_f < 5$  and  $22 < Q_f < 25$ .

It is important to note that the main spec- $z$  catalogue is a compilation of multiple catalogues. These data were obtained during different surveys with different targeting strategies and quality requirements during the last two decades, so the exact quality of the spectroscopic verification is impossible to estimate. As a result, the actual robustness of the final spec- $z$  set may be lower, and this is one of the issues that we will address in the following sections. Another nuance is that for some objects the main spec- $z$  catalogue contains multiple measurements made with different instruments, and in some cases these spec- $z$  values have large residuals between each other ( $>0.1$ ). At this stage, we do try to determine which measurements are correct, neither discard these objects. Instead, during the cross-match, we simply use the spec- $z$  measurements that are the best coordinate match to the COSMOS2015 objects. In Section 4.1, we analyse these objects to clarify the nature of the photo- $z$  outliers.

The resulting data set is cross-matched with the COSMOS2015 photometric catalogue, obtaining  $\sim 20\,000$  objects (the exact amount depends on the bands involved for limiting photometric errors, varying in the various experiments; see Section 2.4). The pre-processing does not noticeably affect the spec- $z$  distribution for  $z_{\text{spec}} \leq 4$ . However, the number of objects for  $z_{\text{spec}} > 1.2$  is approximately one order of magnitude lesser than the number of closer objects (see online Appendix A for the redshift distributions before and after the cleaning). In absolute numbers, we have only  $\sim 700$  galaxies in the redshift range of  $1.2 < z_{\text{spec}} < 4$ . Such number of objects is not enough to effectively train our photo- $z$  algorithm in this redshift range, and for this reason we limit our further analysis and the resulting catalogue to  $z_{\text{spec}} \leq 1.2$ .

## 2.3 DEIMOS spec- $z$ catalogue

Since we want to test our methodology for selecting the subset of photometric data that is well covered by the KB, we need an independent spec- $z$  catalogue different from the main spec- $z$  catalogue selection function. For this purpose, we used the catalogue of spectroscopic redshift presented in Hasinger et al. (2018), acquired with DEIMOS, within different programmes. This catalogue provides redshifts for sources that are not included in the main spec- $z$  catalogue and that are somewhat fainter (see Fig. 1). For these differences, it represents an excellent benchmark data for this study.



**Figure 2.** Venn diagram of the catalogues used in this work.

In the preparation of the DEIMOS spec-z catalogue, we follow the same procedure described in Section 2.2, aimed at discarding stars, AGNs, and unreliable sources.<sup>2</sup>

## 2.4 Final catalogues and SOM data cleaning prerequisites

After the basic pre-processing described in the previous sections, we produce the following data sets (see Fig. 2):

(i) KB, which is the intersection cross-match between the COSMOS2015 and main spec-z catalogues. It contains both photometry and spec-z. For the ML photo-z experiments, we randomly split this KB into train (70 per cent of KB) and blind test (30 per cent) data sets to provide reliable evaluation of the model performance (see online Appendix A for the comparison of the model performance on the train and test data sets). To avoid overfitting, we also use tenfold cross-validation during the training;

(ii) Run data set, which is the COSMOS2015 catalogue after excluding the objects from the KB. It contains only photometric data;

(iii) DEIMOS data set, which is an intersection of the run data set and DEIMOS spec-z catalogue (meaning that it does not contain the objects from the KB). We use the DEIMOS as a control data set to check how well our cleaning procedures work on independent data, occupying a different hyper-volume in the parameter space (see Section 2.3).

In order to ensure the quality of the trained model, a reliable photometry is required, so we excluded all objects with high magnitude errors. Some bands have too many objects with unreliable photometry; if applied to these bands, this part of the pre-processing would reduce the size of the data set by a factor of 2 or more, and narrow the area of parameter space where the photo-z algorithm would be applicable. To avoid this, we identified the bands affected by too many objects with large magnitude errors ( $e\_mag > 1$ ) and excluded them from the experiments. As a rule of thumb, this selection was operated by removing bands containing one order higher number of unreliable measurements than the others; these bands are  $5\_smag$ ,  $8\_omag$ ,  $NUVmag$ , and  $FUVmag$  (see online Appendix A). They contain thousands of objects with large magnitude uncertainties, while the other bands have up to hundreds of such galaxies.

<sup>2</sup>Note that the DEIMOS catalogue has two different quality flag columns. The one following the same scheme as in Lilly et al. (2009) is labelled ‘Qf’.

Afterwards, we clean the photometry in the remaining bands, by limiting the magnitude error within the range  $0 < e\_mag < 1$ , where  $e\_mag$  are magnitude errors for each band. The cleaning is performed for every feature set separately, since it allows us to preserve as many objects as possible (for example, in the experiments where we use only broad-bands we do not remove the objects with high magnitude errors in narrow bands). Then, in all ML experiments we independently normalize each band to the  $[0, 1]$  range.

In order to take care of the mentioned low extrapolative power of ML models, we have to make sure that the run data set is compliant with the KB in terms of parameter space coverage. Fig. 1 shows the magnitude distribution for the KB, DEIMOS, and run data sets in  $i\_pmagap3$ . As it can be seen, our run data set is noticeably deeper than the KB and DEIMOS data sets, as expected due to the spectroscopic bias induced by the targeting algorithms.

We then limit the magnitudes in the run data set bands by their corresponding maximum values present within the KB. However, this is just a preliminary solution, since it does not guarantee a full similarity between the KB and run data sets, with respect to the parameter space. In Section 3.5, we introduce a more accurate procedure to align the run data set to the parameter space of the KB.

The summary on the parameters of the data sets used in this work (except the preliminary experiments to determine the optimal feature set, described in Section 4.1) is given in Table 1.

## 3 METHODS

In this work, we use two neural network algorithms. To calculate the photo-z point estimations, we use the supervised ML algorithm MLPQNA (Brescia et al. 2013, 2014). Furthermore, in order to investigate and clean the data sets, we use an unsupervised ML approach, based on a modified version of the well-known dimensionality reduction algorithm SOM (Kohonen 2013).

### 3.1 MLPQNA algorithm

MLPQNA is a neural network based on the classical Multilayer Perceptron with two hidden layers (Rosenblatt 1963). Instead of the standard Backpropagation learning rule, this algorithm uses a more sophisticated quasi-Newton approximation of the Hessian error matrix (Nocedal 2006). MLPQNA was successfully used for photo-z estimation in a number of papers (such as Cavuoti et al. 2012a, 2015a, 2017a, b; Biviano et al. 2013; Brescia et al. 2013, 2014; Nicastro et al. 2018).

The model hyper-parameters were heuristically selected on the basis of our past experience and on an intensive test campaign. To determine the number of neurons in each layer, we follow the rule of thumb described in Brescia et al. (2013). This rule implies that the optimal number of neurons for the first hidden layer equals  $2N + 1$  and the optimal number of neurons in the second hidden layer is  $N - 1$ , where  $N$  is number of features (i.e. photometric bands). The MLPQNA hyper-parameters are reported in Table 2.

### 3.2 Metrics

In order to evaluate the quality of photo-z estimations, we calculate the residuals as

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

and apply the following set of statistical metrics:

- (i) Standard deviation  $\sigma(\Delta z)$ ;

**Table 1.** Main spectroscopic properties of the data sets. For the initial set of experiments, described in Section 4.1, different sets of bands were used. The filtering of the objects with large magnitude errors was performed for each feature set independently (see Section 2.4), meaning that the exact parameters of the data sets were varying from experiment to experiment. However, these variations were small (e.g. the variations of the sizes of the data sets were less than 1 per cent). The parameters given in this table describe the data set with an optimal feature set, which was determined in Section 4.1 and then used throughout the rest of the paper.

Data set	Size	Spec-z range	Median spec-z	imagap3 range	Median imagap3
KB	19 893	[0.020; 1.2]	0.57	[17.91; 26.19]	22.02
DEIMOS	2288	[0.025; 5.68]	0.89	[18.48; 25.90]	22.93
Run	228 361	NA	NA	[18.14; 26.25]	24.49

**Table 2.** MLPQNA model hyper-parameters settings (see Brescia et al. 2013, for the details).

Parameter name	Parameter value
Decay	0.001
Restarts	80
Threshold	0.01
Epochs	20 000
Activation function	tanh
nHiddenLayers	2
Number of neurons in the first layer	$(2 \cdot \text{num\_features}) + 1$
Number of neurons in the second layer	$\text{num\_features} - 1$
Cost function	Mean squared error

(ii) Normalized median absolute deviation  $\text{NMAD}(\Delta z) = 1.48 \times \text{median}(|\Delta z|)$ , which is less sensitive to catastrophic outliers than  $\sigma(\Delta z)$ ;

(iii) Bias, defined as  $\text{mean}(\Delta z)$ ;

(iv) Percentage of outliers  $\eta_{0.15}$ , defined as a number of sources with  $|\Delta z| \geq 0.15$ .

### 3.3 SOM

The SOM is a neural network algorithm first described in Kohonen (1982). The idea behind SOM is the following: A parameter space of an arbitrary dimensionality is projected on a map of lesser dimensionality, most commonly two-dimensional, in such a way that neighbour instances in the original parameter space remain neighbours on the resulting map. For this purpose, the SOM algorithm compares the feature vector of every object in the data set (in our case, the magnitude vector of every galaxy) with the weight vector of the same dimensionality associated with each cell on the 2D map. The object is then placed in the cell having the most similar weight vector (in terms of Euclidean distance). Such cell is called Best-Matching Unit (BMU). The weights of the BMU and its neighbour cells are updated in such a way that they become more similar to the feature vector of the object. This mechanism ensures that, at the end of training loop, the map learns the representation of the parameter space of the entire training data set in a self-organized way. That is why SOM is commonly used to perform an unsupervised data exploration. In photo- $z$  estimation applications, several authors demonstrated that SOM can be used for different tasks. For example, Geach (2012) and Way & Klose (2012) used it to obtain photo- $z$ , while Carrasco Kind & Brunner (2014) applied it to estimate photo- $z$  PDFs. Finally, Masters et al. (2015) adapted the SOM to check the coverage of the photometric parameter space by a given spectroscopic sample, thus indirectly creating a suitable method to optimize a spectroscopic follow-up strategy.

**Table 3.** SOM settings used for all experiments in this study. For more details about low- and high-resolution maps, see Section 3.3.

Parameter name	Parameter value
Width (low-resolution SOM)	25
Height (low-resolution SOM)	28
Width (high-resolution SOM)	67
Height (high-resolution SOM)	64
Num_features	10
Epochs	6000
Sigma	5
Learning rate	0.5
Neighborhood_function	bubble

In this work, we use the SOM for two different purposes: (i) to detect potentially unreliable spec- $z$  in the KB (see Sections 3.4 and 4.2); and (ii) to ensure that the run data set occupies the same part of the parameter space as the KB, i.e. to remove objects in the run data set that are photometrically different from those in the KB (see Sections 3.5 and 4.3). In our experiments, we use a modified version of MINISOM (Vettigli 2018).<sup>3</sup>

We use photometric bands as the input features for the SOM training. To analyse the data sets, we colour label the resulting maps using photo- $z$ , spec- $z$ , or the number of objects within each cell (also called the cell's occupation). The latter mapping is further referred as occupation maps, and it allows us to check how well the data set, used for the labelling, covers the parameter space of the SOM training data set.

Most of the SOM hyper-parameters (specifically, number of epochs, sigma, learning rate, and neighbourhood function) are chosen via grid search. Their final values are listed in Table 3. The choice of the shape (square, rectangular, or spherical) of the maps is somewhat arbitrary; in a number of previous papers, different shapes were tried and proven to be useful (see e.g. Carrasco Kind & Brunner 2014; Masters et al. 2015; Buchs et al. 2019). We opted for rectangular maps, since it is assumed to give the algorithm a preferred direction to align the data (Masters et al. 2015). It is worth noticing, though, that in the preliminary experiments we did not see any significant effect of the shape of the maps on the analysis, described in the next sections.

The size of the map is chosen in such a way that, on average, each cell contains more than 30 galaxies from the training set. The choice of this size is based on the best compromise between the two competing goals: the reliability of statistics within each cell, and the need to capture the data topology with the maximum finesse.

<sup>3</sup>The original version of MINISOM can be found in the repository <https://github.com/JustGlowing/minisom>. Our modified version is available at <https://github.com/ShrRa/minisom>.

A SOM with a small number of cells provides us a higher number of galaxies per cell, thus improving the reliability of the statistical indicators. On the other side, a larger SOM shows more details of the data distribution in the parameter space, but the statistics within some cells become unreliable. For this reason, we use SOM of different sizes. Specifically, we use small low-resolution maps to determine the anomalous sources within each cell, and large high-resolution maps to investigate the train data set coverage of the parameter space defined by the run catalogue.

### 3.4 Spec- $z$ cleaning with SOM

The trained SOM places objects with similar feature vectors in the same or neighbour cells. If some object has a photometry–spec- $z$  relationship that appears to be anomalous for its BMU, such object will be considered as an outlier in terms of spec- $z$  distribution of this BMU. In order to avoid the potential confusion between these in-cell outliers with the traditional outliers of a photo- $z$  distribution, hereafter we will refer the in-cell outliers as *anomalous sources* or *anomalies*.

We assume that these objects have a lower photo- $z$  reliability, so we drop them out. In order to exclude such objects, for every galaxy we calculate the coefficient:

$$K_{\text{spec}} = \frac{\langle z_{\text{spec}}^{\text{BMU}} \rangle - z_{\text{spec}}^{\text{obj}}}{\sigma(z_{\text{spec}}^{\text{BMU}})}, \quad (2)$$

where  $\langle z_{\text{spec}}^{\text{BMU}} \rangle$  is the mean spec- $z$  obtained by averaging over the objects falling in the same BMU,  $z_{\text{spec}}^{\text{obj}}$  is the spec- $z$  of the galaxy, and  $\sigma(z_{\text{spec}}^{\text{BMU}})$  is the standard deviation of the spec- $z$  distribution within the BMU cell. This coefficient has the same meaning as a standard score, or Z-score, often used in statistics. We choose the prefix  $K$  instead of  $Z$  to avoid confusion with redshifts. Typically, objects are considered to be outliers if  $|K_{\text{spec}}| > 3$ . Yet, in our experiments we also tried other criteria to see how it affects the photo- $z$  quality (see Section 4.2).

### 3.5 Parameter space verification with SOM occupation map

As we pointed out earlier, neural networks cannot perform extrapolation. It means that, in order to obtain reliable results, the run data set has to occupy the same area of the parameter space that is well sampled by the KB. Fig. 1 shows that, in terms of magnitude distribution, our train and run data sets are quite different. Therefore, in order to avoid biasing, we need to estimate the statistics for each magnitude bin independently. Furthermore, in order to obtain reliable results, we have to keep the parameter space of the run data set as similar as possible to the parameter space of the KB. Usually, this is done by limiting the magnitudes of the run data set to the maximum magnitudes of the KB, i.e. by cutting the faint-end ‘tail’ of the magnitude distributions (e.g. Cavuoti et al. 2015b; Eriksen et al. 2019; Wright et al. 2019). However, this approach poses a major problem: If extreme cuts are adopted, the coverage of the parameter space is ensured but at the risk of losing many objects with reliable photo- $z$ . With a soft limiting value, on the other hand, the run data set is affected by too many objects that lie outside of the KB parameter space.

Using SOM occupation maps, it is possible to implement a more accurate method of photometry filtering, based on the approach first investigated by Masters et al. (2015). The idea behind the method is simple: we train the SOM on the run data set and then project the KB on the trained map. Since the entire SOM represents a projection of

the parameter space of the run data set, the galaxies in the KB will be clustered only in a subset of cells, with a fairly large number of cells either poorly occupied or completely empty. For these cells, we do not have spectral information and hence photo- $z$  predictions cannot be trusted. In order to capture in detail the coverage of the run data set parameter space by the KB, we use a high-resolution SOM map of size  $67 \times 64$  (Table 3).

## 4 EXPERIMENTS

The experimental part of this work consists of the following four stages:

- (i) The first one, described in Section 4.1, is dedicated to identifying the optimal parameters for MLPQNA. For these experiments, we use our KB with only the standard pre-processing, as outlined in Section 2.
- (ii) The second set of experiments is described in Section 4.2. It aims to analyse how effective is our SOM in removing sources with anomalous spec- $z$ . In this set of experiments, we also use only the KB.
- (iii) At the third stage, discussed in Section 4.3, we check whether the SOM can be used to extract from a new data set (DEIMOS) only those sources that belong to the same part of the parameter space as the KB, i.e. photometrically similar to the sources from the KB. For this set of experiments, we use both the KB and the DEIMOS data sets.
- (iv) Finally, in Section 4.4 we apply the developed methods to estimate the photometric redshifts and their reliability for objects in the run data set.

### 4.1 Photo- $z$ before SOM cleaning

Here, we calculate the photometric redshifts for the KB after the standard pre-processing described in Section 2. We performed a number of experiments with different combinations of features, namely all photometric bands, only broad-bands, broad-bands plus one or more narrow bands, with five ‘SDSS-like’ bands, etc. The results of the most representative experiments are reported in Table 4.

Keeping in mind that COSMOS2015 has exceptionally rich and well-calibrated photometry, one could expect to obtain photo- $z$  catalogue of high precision. Our expectations were also based on the very precise results obtained with SED fitting in Laigle et al. (2016), and on the fact that in previous works [e.g. done for SDSS-DR9 (Cavuoti et al. 2017b) and KiDS (Brescia et al. 2014)], MLPQNA generally performed on a comparable or better level than other photo- $z$  algorithms.

Yet, in all the experiments, reported in Table 4, we did not reach the accuracy achieved in previous works. In particular, we have noticeably higher  $\sigma(\Delta z_{\text{ML}})$  and percentage of outliers than those reported in the aforementioned papers for KiDS and SDSS-DR9.

The COSMOS2015 is deeper than the catalogues used in those publications, so the cause of the deterioration of the overall statistics could be in the lower quality of photometry and spectroscopy for the high- $z$  sources. However, even for low- $z$  galaxies (i.e. with  $z_{\text{spec}} < 0.5$ ), both the percentage of outliers and  $\sigma(\Delta z_{\text{ML}})$  turned out to be of lower quality (see Table 5). In fact, there is an unexpected high percentage of outliers in all redshift bins; they are also uniformly distributed across the whole observed field, meaning that photo- $z$  errors are not caused by the contamination by the light of some nearby star. Also, various feature sets show similar percentage of outliers,

**Table 4.** Results of the experiments performed with the trained MLPQNA on the blind test set of the KB, after having just applied the standard KB cleaning procedure.

Exp ID	Description	Bands	Mean	$\sigma_{\Delta z/(1+z)}$	NMAD	$\eta_{0.15}$
exp001	SDSS-like, aper2	<i>u, B, r, ip, zpp</i>	0.000	0.053	0.025	1.999
exp002	SDSS-like, aper3	<i>u, B, r, ip, zpp</i>	−0.001	0.051	0.027	1.807
exp003	SDSS-like, aper_auto	<i>u, B, r, ip, zpp</i>	−0.003	0.055	0.027	2.364
exp004	SDSS-like, aper_ISO	<i>u, B, r, ip, zpp</i>	−0.004	0.059	0.028	2.346
exp005	Broad-bands, aper2 and aper3	<i>u, B, r, ip, zpp</i>	−0.001	0.051	0.025	1.756
exp006	Broad-bands, aper3	<i>B, H, J, Ks, V, Y, ip, r, u, zpp</i>	−0.002	0.048	0.018	2.138
exp007	Broad-bands + one narrow band, aper3	<i>B, H, J, Ks, V, Y, ip, r, u, zpp, IA574</i>	−0.002	0.048	0.019	1.642
exp008	All bands, aper3	<i>Ks, Y, H, J, B, V, ip, r, u, zp, zpp, IA484, IA527, IA624, IA679, IA738, IA767, IB427, IB464, IB505, IB574, IB709, IB827, NB711, NB816, Hw, Ksw, yH</i>	−0.002	0.049	0.017	1.877

**Table 5.** Statistics for the test set calculated in spec-*z* bins. ML photo-*z* used here were obtained during the exp007 from Table 4 using 10 broad and 1 medium bands.

Spec- <i>z</i> bin	Num. objects	$\sigma_{\Delta z/(1+z)}$	ML photo- <i>z</i>			SED photo- <i>z</i>			
			NMAD	Mean	$\eta_{0.15}$	$\sigma_{\Delta z}$	NMAD	Mean	$\eta_{0.15}$
[Overall]	5967	0.048	0.019	−0.0023	1.64	0.094	0.011	−0.0041	2.23
[0.0; 0.2]	422	0.100	0.027	−0.0416	8.29	0.278	0.012	−0.0552	7.82
[0.2; 0.4]	1445	0.043	0.016	−0.0081	1.31	0.056	0.008	−0.0034	1.94
[0.4; 0.6]	1158	0.038	0.016	0.0003	1.04	0.067	0.010	−0.0002	2.07
[0.6; 0.8]	1489	0.036	0.018	0.0001	1.01	0.052	0.012	0.0002	1.48
[0.8; 1.0]	1085	0.032	0.019	0.0041	0.55	0.044	0.014	0.0057	1.20
[1.0; 1.2]	368	0.051	0.028	0.0292	2.99	0.094	0.027	−0.0068	3.53

so it is unlikely that the source of the problem is the calibration of photometry in some particular band.

Another possible explanation is that the photo-*z* outliers can be either misinterpreted spectra (in a broad sense of this word, i.e. including blended galaxies, incorrect cross-match between photometric and spectroscopic catalogues, etc.) or exotic objects. Thanks to the fact that some objects in our main spec-*z* catalogue contain more than one spec-*z* measurement, we can estimate how many of the outliers belong to each of these categories. In order to do this, we select objects with more than one measurement of spec-*z* and calculate maximum difference between these measurements, calling it spec-*z* scatter. Then, we calculate the percentage of outliers separately for objects with a single spec-*z* measurement, multiple measurements and small (<0.1) scatter, and multiple measurements and large ( $\geq 0.1$ ) scatter. Table 6 reports these calculations for ML and SED fitting photo-*z*.

From this table, we can deduce several facts:

- (i) For the objects with multiple spec-*z* values, the percentage of outliers for ML and SED fitting photo-*z* is essentially the same.
- (ii) For the objects with small spec-*z* scatter, this percentage is significantly lower ( $\eta_{0.15} \sim 0.2$  per cent) than that for the objects with large spec-*z* scatter ( $\eta_{0.15} \sim 11$  per cent).
- (iii) For the objects with single spec-*z* measurement, the percentage of outliers for SED fitting is  $\sim 50$  per cent higher than that for ML.

The most probable explanation here is that for the photo-*z* outliers with large spec-*z* scatter, the specific spec-*z* measurement used to estimate the photo-*z* residual is incorrect. Consequently, the majority of such bona fide outliers are likely to have correct photo-*z* predictions. It also implies that a significant percentage of bona fide outliers with single spec-*z* measurements should be attributed to incorrect spec-*z* measurements. At the same time, we should not

assume that the percentage of incorrect spec-*z* measurements in this group is the same as for the objects with multiple measurements, since there is no guarantee of similarity of the selection functions for these two categories of objects.

Instead, we use the SOM cleaning procedure, described in Section 3.4, to select a set of objects with reliable spec-*z* even without multiple spec-*z* measurements.

#### 4.2 Photo-*z* statistics after removing in-cell anomalous spec-*z*

We train the SOM, using the KB with the set of broad-bands that gave us the best results for the MLPQNA experiments (*u, B, V, r, ip, zpp, Y, J, H, and Ks*; see Table 4). This feature set can be viewed as an approximation of the feature set that will exist for the combined LSST and *Euclid* data, which will also range from optical to NIR wavelengths (Rhodes et al. 2017). Fig. 4 shows the resulting SOM map, colour labelled with, respectively, the mean and standard deviation of spec-*z* and ML and SED fitting photo-*z*. In order to discard objects with anomalous spec-*z* values, we calculate the in-cell spec-*z* outlier coefficients  $K_{\text{spec}}$  for each source, as defined in equation (2); then, we bin the whole KB according to the value of this coefficient and calculate the statistics of photo-*z* residuals for each bin. This allows us to check how the quality of photo-*z* correlates with similarity between the spec-*z* of a given source and the mean spec-*z* of its BMU.

Fig. 3 shows that the majority of the objects have relatively small  $K_{\text{spec}}$  (second row of the figure), and the statistics are much better for them than for the objects with larger absolute values of  $K_{\text{spec}}$ .

In the majority of the bins, ML photo-*z*'s have lower standard deviations (third row from the top) and lower percentage of outliers (last row from the top) than SED photo-*z*'s, but higher NMAD (fourth row). Predictably, mean residuals (second row from the bottom) have inverse correlation with  $K_{\text{spec}}$ , implying that photo-*z* predictions, for

**Table 6.** Statistics for ML and SED outliers for the test set of the KB for objects with different number and scatter of spec- $z$  measurements. The upper part of the table reports the results before removing in-cell spec- $z$  anomalies, and the lower four rows report the results after  $|K_{\text{spec}}| \leq 1$  filtering.

Case	Num. objects	ML photo- $z$		SED photo- $z$	
		Num. outliers	Per cent outliers	Num. outliers	Per cent outliers
Total	5967	98	1.64	133	2.23
Single measurement	3745	63	1.68	98	2.62
Multiple measurements, spec- $z$ scatter $< 0.1$	1945	4	0.21	5	0.26
Multiple measurements, spec- $z$ scatter $\geq 0.1$	277	31	11.19	30	10.83
$ K_{\text{spec}}  \leq 1$ , total	4311	8	0.19	30	0.70
$ K_{\text{spec}}  \leq 1$ , single measurement	2683	5	0.19	24	0.89
$ K_{\text{spec}}  \leq 1$ , multiple measurements, spec- $z$ scatter $< 0.1$	1468	0	0.00	1	0.07
$ K_{\text{spec}}  \leq 1$ , multiple measurements, spec- $z$ scatter $\geq 0.1$	160	3	1.88	5	3.12

objects with spec- $z$  lower than the median spec- $z$  of their BMU, are biased towards higher values, and vice versa.

By limiting our data set to galaxies with absolute value of  $K_{\text{spec}}$  smaller than 1, we reduce the percentage of outliers from 1.64 to 0.19 for ML photo- $z$  and from 2.23 to 0.7 for SED fitting. The standard deviation of residuals also is reduced by a factor of  $\sim 2$  (see Table 7).

Remarkably, after this cleaning the statistics are effectively the same as for the objects with multiple spec- $z$  measurements and low spec- $z$  scatter. In other words, removing in-cell anomalous spec- $z$  leaves us with a reliable set of spec- $z$  with no need to use repeated spec- $z$  measurements, and this set is twice as large as the one with multiple measurements.

The lower part of Table 6 shows that while the percentage of outliers for the ML photo- $z$  drops not only for the objects with large spec- $z$  scatter, but for other categories as well, for the SED fitting the improvement for the objects with single spec- $z$  measurement is weaker.

Taking into account that the run data set does not contain spec- $z$ , it appeared useful to check whether the quality of the data set could be improved by using photo- $z$  in-cell anomaly detection instead of spec- $z$ . To do so, we calculated outlier coefficients for SED and ML photo- $z$  ( $K_{\text{SED}}$  and  $K_{\text{ML}}$  correspondingly), and applied the same filtering using these coefficients instead of  $K_{\text{spec}}$ . It turns out that such filtering improves the overall statistics, even though less than the  $K_{\text{spec}}$  cleaning (upper half of Table 7). As expected, the relative improvements are stronger for the SED fitting photo- $z$  than for the ML photo- $z$ ; in Section 5.3, we will discuss this aspect in better detail.

### 4.3 Photo- $z$ after parameter space verification with occupation map

In order to be able to discard the part of the run data set for which MLPQNA predictions are not likely to be reliable, we have to use occupation maps to identify and flag galaxies that are not photometrically similar to the galaxies of the train data set. To estimate the performance of such photometry cleaning, we made use of the DEIMOS data set. Considering that the DEIMOS data set is slightly deeper than the train data set, we expect that the cleaning process will mostly remove faint objects. Fig. 5 shows occupation maps for the train, test, DEIMOS, and run data sets, respectively. For each object of the test, DEIMOS, and run data sets, we derive the occupation of its BMU by the galaxies of the train data set. As expected, Fig. 6 demonstrates that the statistics tend to be better for the cells with higher occupation by the train data set. Based on the statistics for the DEIMOS data set (right-hand panel of Fig. 6), we

chose to leave only the objects that belong to the cells with occupation  $> 5$ , since more strict criteria do not bring further improvements.

Tables 7 and 8 report the overall statistics for the test and DEIMOS data sets, respectively. The statistics are given for the different stages of cleaning, including the cases of combination of cleaning with occupation map and removal of the objects with anomalous spectroscopic or photometric redshifts. From these tables, we see that the effect of occupation filtering is much stronger for the DEIMOS than for the test data set (since the test data set by construction covers the same regions of the parameter space as the train data set). Still, even on the test data set the effect of the filtering with occupation map is comparable or better than the effect of the filtering of the photo- $z$  anomalous sources. Also, occupation map filtering is more cost effective in a sense that it discards much less objects than spec- $z$  or photo- $z$  anomalous source filtering.

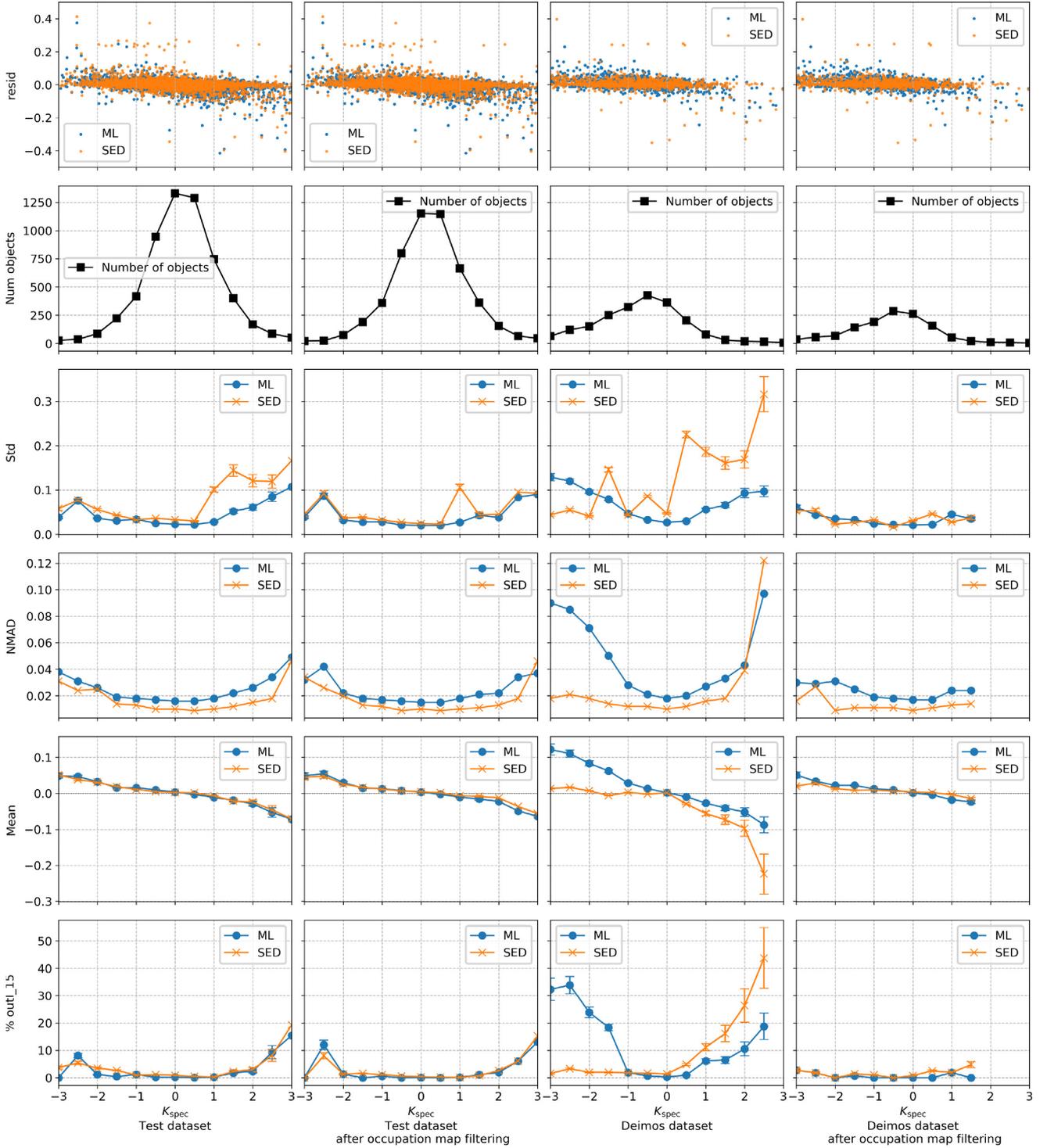
From Table 8, we see that without any filtering applied to the DEIMOS data set, our ML photo- $z$  has worse indicators than SED photo- $z$ . This can be easily understood by remembering that part of the DEIMOS data set lies outside of the parameter space of the train data set. However, for objects that belong to the cells with good occupation, the percentage of outliers and standard deviations for both ML and SED fitting photo- $z$  are very close.

After additional spec- $z$  anomalous source filtering, the statistics improves even more and reaches approximately the same values as for the KB; the scatter plots in the bottom part of Fig. 7 clearly illustrate the improvement. From there, it can be seen that the two cleaning procedures remove different outliers. The plots for the DEIMOS data set show that both ML and SED fitting photo- $z$  require occupation map filtering to select the objects with good predictions. Without it, SED fitting produces a lot of catastrophic outliers in the whole range of  $z_{\text{spec}}$ , and ML systematically fails for the objects with  $z_{\text{spec}} > 1.2$ .  $K_{\text{spec}}$  filtering eliminates most of the remaining outliers, distributed randomly across the  $z_{\text{spec}}$  range.

### 4.4 SOM cleaning of the run data set

The COSMOS2015 contains  $\sim 500\,000$  galaxies. After the standard pre-processing described in Section 2 and excluding objects that belong to the KB, the run catalogue consists of  $\sim 190\,000$  galaxies. For all these galaxies, we calculate ML photo- $z$ , and in order to determine the reliability of these redshifts, we calculate  $K_{\text{SED}}$ ,  $K_{\text{ML}}$ , and the occupation of their BMU by the objects in the train data set, as described in the previous subsections.

Since we do not have spec- $z$  for the galaxies in the run data set, we can only perform an accuracy test comparing SED and ML photo- $z$ . To do so, we construct ML/SED residuals  $\Delta z_{\text{ML/SED}} = (z_{\text{SED}} - z_{\text{ML}})/(1 + z_{\text{SED}})$ . Obviously, such a test suffers from biases introduced



**Figure 3.** Statistics for the test and DEIMOS data sets in  $K_{\text{spec}}$  bins before and after occupation map filtering. Bins with number of objects  $< 15$  are considered to be unreliable and excluded from these plots.

by both photo- $z$  methods, so it can be used only for qualitative estimation of the photo- $z$  robustness.

Table 9 compares the statistics for ML/SED photo- $z$  residuals for the DEIMOS and run catalogues after applying different filters. The cleaning with occupation map appears to be the most important step, since it allows us to reduce the percentage of outliers by almost an order of magnitude. The removal of SED fitting photo- $z$  anomalous

sources also improves the statistics. Maximum improvement for both data sets is achieved by a combination of occupation map filtering and removal of SED fitting photo- $z$  anomalous sources. It reduces the percentage of outliers from 11 per cent to 0.27 per cent for the DEIMOS data set, and from 40 per cent to 2 per cent for the run data set.  $\sigma(\Delta z)$  for the DEIMOS drops from 0.102 to 0.03, and from 0.238 to 0.056 for the run data set. NMAD changes from 0.033 to

**Table 7.** Statistics for ML and SED fitting photo- $z$  calculated for the test data set after different types of filtering. The upper part of the table presents the statistics calculated for the data set without spec- $z$  and photo- $z$  outliers. The lower part reports the effects of photometric filtering with occupation map.

Filtering	Num. objects	ML photo- $z$					SED photo- $z$			
		$\sigma_{\Delta z}$	NMAD	Mean	$\eta_{0.15}$	$\sigma_{\Delta z}$	NMAD	Mean	$\eta_{0.15}$	
No filtering	5967	0.048	0.019	-0.0023	1.64	0.094	0.011	-0.0041	2.23	
$ K_{\text{spec}}  \leq 1$	4311	0.025	0.017	0.0002	0.19	0.052	0.010	0.0006	0.70	
$ K_{\text{ML}}  \leq 1$	4071	0.045	0.018	-0.0020	1.28	0.077	0.010	-0.0026	1.65	
$ K_{\text{SED}}  \leq 1$	4133	0.043	0.018	-0.0020	1.06	0.061	0.010	-0.0021	1.43	
trainMapOccupation > 5	5167	0.041	0.018	-0.0021	1.18	0.058	0.010	0.0001	1.49	
$K_{\text{spec}} + \text{trainMapOccupation}$	3761	0.022	0.016	-0.0002	0.05	0.05	0.009	0.0018	0.35	
$K_{\text{ML}} + \text{trainMapOccupation}$	3587	0.039	0.017	-0.0015	1.06	0.062	0.010	-0.0003	1.17	
$K_{\text{SED}} + \text{trainMapOccupation}$	3624	0.038	0.017	-0.0016	0.94	0.038	0.010	0.0002	0.99	

**Table 8.** Statistics for ML and SED fitting photo- $z$  for the DEIMOS data set after different types of filterings. The upper part of the table describes statistics after spec- $z$  and photo- $z$  outlier removal, while the lower part reports the effects of photometry filtering with occupation map.

Filtering	Num. objects	ML photo- $z$					SED photo- $z$			
		$\sigma_{\Delta z}$	NMAD	Mean	$\eta_{0.15}$	$\sigma_{\Delta z}$	NMAD	Mean	$\eta_{0.15}$	
No filtering	2255	0.099	0.032	0.0347	10.86	0.142	0.014	-0.0082	5.06	
$ K_{\text{spec}}  \leq 1$	1075	0.035	0.020	0.0018	1.02	0.127	0.011	-0.0103	2.88	
$ K_{\text{ML}}  \leq 1$	1209	0.095	0.033	0.0392	12.57	0.090	0.013	-0.0031	4.55	
$ K_{\text{SED}}  \leq 1$	1183	0.085	0.029	0.0245	8.96	0.078	0.013	0.0005	3.30	
trainMapOccupation > 5	1382	0.058	0.023	0.0127	2.10	0.059	0.012	0.0085	2.68	
$K_{\text{spec}} + \text{trainMapOccupation}$	758	0.025	0.018	0.0017	0.13	0.031	0.010	0.0038	0.92	
$K_{\text{ML}} + \text{trainMapOccupation}$	724	0.064	0.022	0.0136	1.93	0.060	0.011	0.0080	2.62	
$K_{\text{SED}} + \text{trainMapOccupation}$	741	0.046	0.020	0.0063	1.48	0.044	0.011	0.0075	1.89	

**Table 9.** Statistics for ML/SED residuals for the DEIMOS and run data sets after different types of filtering. The upper part of the table describes statistics after spec- $z$  and photo- $z$  outlier removal, while the lower part reports the effects of photometry filtering with occupation map. Taking into account that for the run data set spectral information is absent, the rows corresponding to spec- $z$  cleanings for run data set are empty.

Filtering	DEIMOS data set					Run data set				
	Num. objects	$\sigma_{\Delta z}$	NMAD	Mean	$\eta_{0.15}$	Num. objects	$\sigma_{\Delta z}$	NMAD	Mean	$\eta_{0.15}$
No filtering	2255	0.102	0.033	0.0355	11.35	194509	0.238	0.129	0.1373	40.23
$ K_{\text{spec}}  \leq 1$	1075	0.070	0.023	0.0049	2.88	NA	NA	NA	NA	NA
$ K_{\text{ML}}  \leq 1$	1209	0.088	0.034	0.0392	11.75	137152	0.225	0.110	0.1243	37.76
$ K_{\text{SED}}  \leq 1$	1183	0.072	0.030	0.0224	7.44	146773	0.206	0.101	0.1217	34.93
trainMapOccupation > 5	1382	0.047	0.023	0.0031	1.52	43279	0.083	0.028	-0.0019	3.59
$K_{\text{spec}} + \text{trainMapOccupation}$	758	0.036	0.020	-0.0028	0.92	NA	NA	NA	NA	NA
$K_{\text{ML}} + \text{trainMapOccupation}$	724	0.043	0.022	0.0048	0.97	33296	0.080	0.028	-0.0029	3.48
$K_{\text{SED}} + \text{trainMapOccupation}$	741	0.030	0.022	-0.0015	0.27	35189	0.056	0.026	-0.0031	2.17

0.022 for the DEIMOS data set and from 0.129 to 0.026 for the run data set.

#### 4.5 Purely data-driven spec- $z$ sample

Apart from selecting reliable photo- $z$  predictions from the final catalogue, SOM filtering can be used to create a larger and potentially more representative KB. It can be done by including in the KB not only galaxies with ‘good’ spectral quality flags (see Section 2.2), but also those galaxies that have ‘non-robust’ quality flags, but good SOM quality indicator ( $|K_{\text{spec}}| \leq 1$ ), i.e. those that appear to have typical  $z_{\text{spec}}$  for their colour cell. In practice, it means that we omit the Q-f cleaning and use only  $|K_{\text{spec}}| \leq 1$  filtering. The KB constructed in this exclusively data-driven way consists of 24 058 objects instead of 19 893.

For this sample, we calculated photo- $z$  using the same MLPQNA set up as in the previous sections. For additional comparison, we calculated photo- $z$  for a random galaxy sample of the same size, taken from the non-cleaned spec- $z$  data set. Table 10 reports the results of these experiments.

From this table, it can be seen that by using only SOM in-cell anomaly filtering, we are unable to obtain the same quality of photo- $z$  as if we make preliminary selection using spectral quality flags. However, for the ML photo- $z$  the difference is rather small; with the combined standard and SOM cleaning, we have NMAD = 0.017 and  $\eta_{0.15} = 0.19$ , while with SOM-only cleaning NMAD = 0.018 and  $\eta_{0.15} = 0.56$ ; the mean residual is essentially the same for both cases. This deterioration of statistics comes with a benefit of increasing the size of the final spec- $z$  data set by  $\sim 67$  per cent. SED fitting photo- $z$  show slightly stronger degradation: While NMAD remains

**Table 10.** Photo- $z$  performance for the test sample with and without standard Q-f cleaning. The first two rows describe baseline exp007 experiment, which was done on the spec- $z$  sample cleaned with standard Q-f criteria (Section 2.2), before and after SOM in-cell spec- $z$  anomaly filtering. The third row describes an experiment done on the spec- $z$  sample that was not cleaned with Q-f, but only with SOM in-cell filtering. The last row reports an experiment done on a random sample from the spec- $z$  catalogue that was not cleaned, neither with Q-f nor SOM in-cell filtering.

Exp ID	Description	Num. obj.	$\sigma_{\Delta z}$	ML photo- $z$			SED photo- $z$			
				NMAD	Mean	$\eta_{0.15}$	$\sigma_{\Delta z}$	NMAD	Mean	$\eta_{0.15}$
exp007	Baseline	5967	0.048	0.019	-0.0023	1.64	0.094	0.011	-0.0041	2.23
exp007-clean	Baseline, $K_{\text{spec}} \leq 1$	4311	0.025	0.017	0.0002	0.19	0.052	0.010	0.0006	0.7
d_exp027	'Dirty' spec- $z$ , $ K_{\text{spec}}  \leq 1$	7218	0.033	0.018	-0.0009	0.57	0.135	0.012	-0.0134	1.93
d_exp028	'Dirty' spec- $z$ , random sample	7218	0.072	0.025	-0.0055	4.32	0.178	0.015	-0.0238	5.78

of the same order, the mean bias is worsened by two orders, from 0.0006 to -0.0134, and the percentage of outliers increases from 0.7 to 1.93. The experiment with a random sample taken from the uncleaned spec- $z$  catalogue demonstrates that without any form of data cleaning, photo- $z$  quality is significantly worse.

This purely data-driven selection of reliable spec- $z$  can be suitable not only for photo- $z$  prediction, where larger KB means better representation of the run data set, but also for a general verification and validation of the quality of spec- $z$  catalogues, useful in other contexts.

## 5 DISCUSSION

As we have shown in the previous section, there are several ways in which we can apply SOM to improve the quality of the photo- $z$  catalogues. In some cases, these cleaning procedures improve both the standard deviation of residuals and the percentage of outliers (e.g. Table 8). In this section, we compare the effects of the SOM cleaning on ML and SED fitting photo- $z$  (Section 5.1), consider the nature of spec- $z$  and photo- $z$  in-cell anomalies (Sections 5.2 and 5.3), and define general strategies for using the SOM cleaning methodology on other data sets (Section 5.4).

### 5.1 SED fitting versus ML

In all the experiments performed on the KB, the ML photo- $z$  distribution has a lower percentage of outliers and lower standard deviations than the SED fitting photo- $z$  distribution, but higher NMAD (see Table 7). For the DEIMOS data set, before SOM filtering, the situation is different: ML photo- $z$  have a significantly higher percentage of outliers ( $\sim 11$  per cent against 5 per cent for SED fitting), due to the fact that DEIMOS contains many objects lying outside of the boundaries of the parameter space sampled by the KB. Occupation map filtering discards the majority of these outliers and the statistics of both ML and SED fitting photo- $z$  become similar to those for the test data set (see Tables 7 and 8). In particular, for ML  $\sigma_{\Delta z}$  decreases from 0.099 to 0.058 (for the KB it equals 0.041), and for SED fitting  $\sigma_{\Delta z}$  from 0.142 to 0.059 (0.058 for the KB). The percentage of outliers also drops from 10.86 per cent to 2.1 per cent for ML photo- $z$  (1.18 per cent for the KB) and from 5.06 per cent to 2.68 per cent for SED fitting (1.49 per cent for the KB).

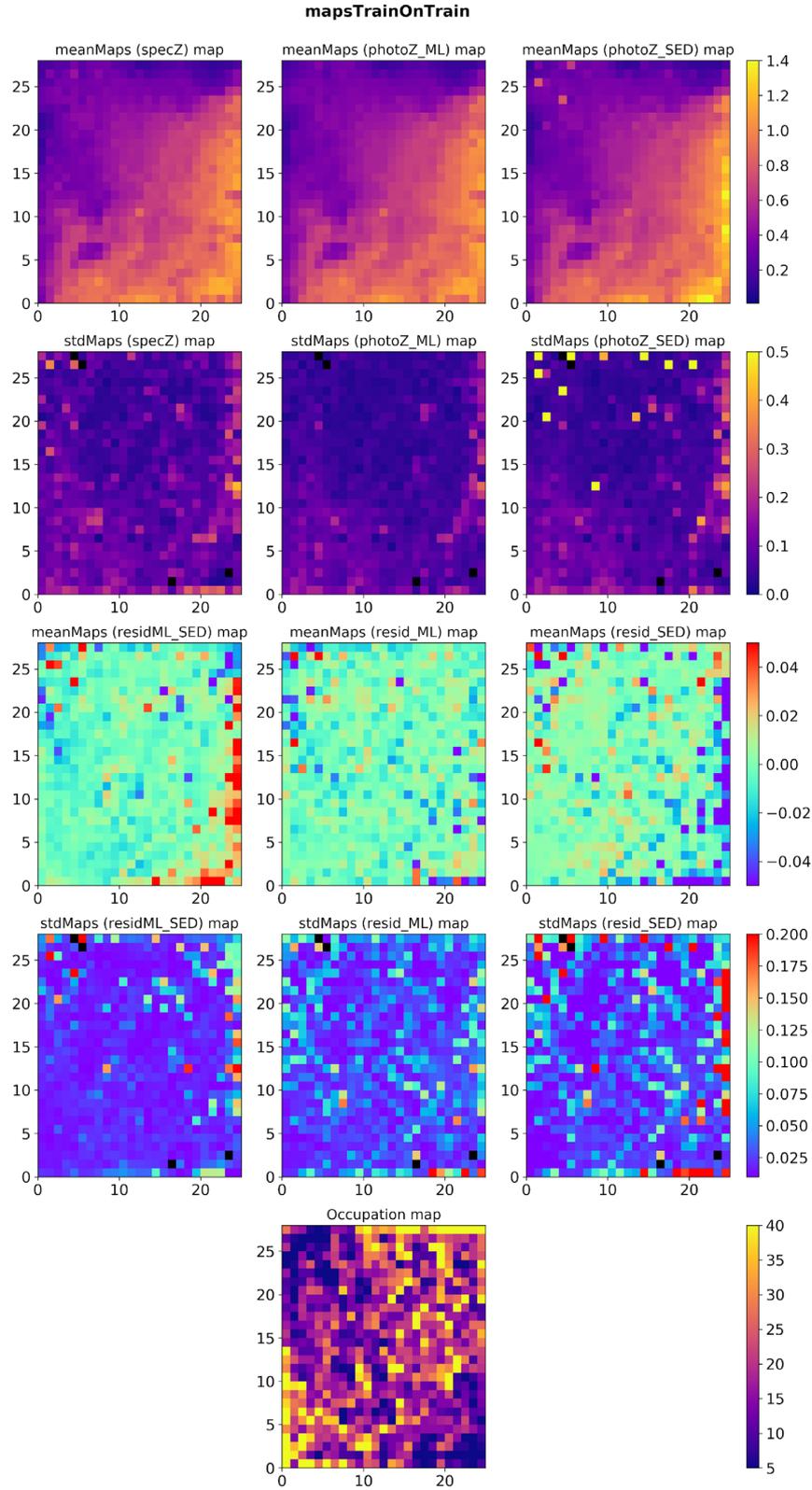
The difference between the two methods in terms of NMAD can be explained by the fact that SED fitting methods benefit from the inclusion of the narrow-band photometry. It allows the SED fitting methods to detect emission lines passing through a certain wavelength range (Ilbert et al. 2009). On the contrary, in the case of MLPQNA, the same inclusion of the additional narrow bands does not lead to significant improvements (Table 4). Earlier work by

several authors (see e.g. Heinis et al. 2016; Eriksen et al. 2020) demonstrated that the inclusion of the narrow-band photometry generally improves ML photo- $z$  only after additional preparations, such as transfer learning or a preliminary extensive phase of feature selection. A possible explanation to this is that in comparison to the broad-bands, narrow bands carry redshift signal for a smaller fraction of galaxies, namely only for those galaxies whose spectrum has emission or absorption lines passing through a particular narrow band. In order to learn to extract this signal, an ML model might need a much larger training set than when only broad-bands are used.

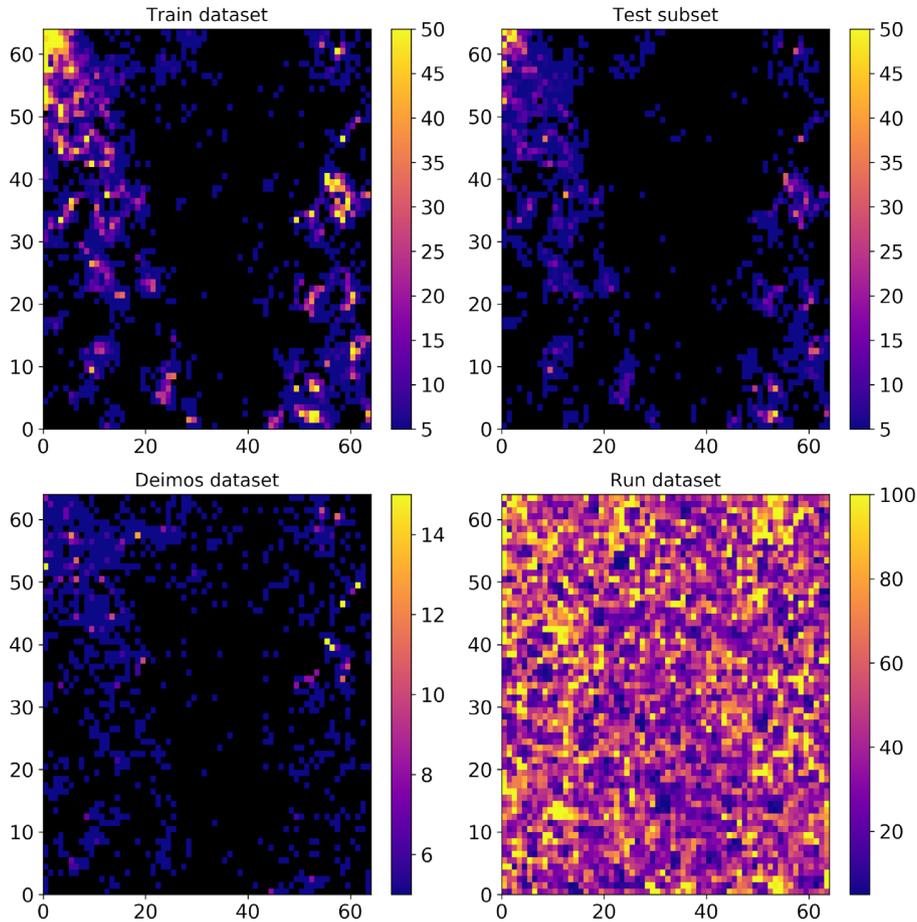
Even after occupation map filtering and removing in-cell anomalous spec- $z$ , SED fitting produces a higher number of catastrophic outliers than ML. Also, the residuals for these outliers are larger than those for ML outliers (see Fig. 7; for the test data set, the mean absolute value of residuals for SED fitting catastrophic outliers is 0.35, while for ML outliers it is 0.26; for the DEIMOS, the mean absolute residual for SED fitting is 0.29, and for ML it is 0.22). As can be seen from the second row of Fig. 4, the SOM map of the standard deviations of SED fitting photo- $z$  has a number of cells with significantly higher values than those that appear on the  $\sigma(z_{\text{spec}})$  and  $\sigma(z_{\text{ML}})$  maps. These are the cells where the majority of the catastrophic outliers are located, and it seems likely that these outliers appear due to the lack of a suitable SED template or SED fitting failure.

An unexpected result is that SED fitting seriously benefits from occupation map filtering. For the test data set, the percentage of outliers drops from 2.23 per cent to 1.49 per cent, while for the DEIMOS it goes from 5.06 per cent to 2.68 per cent, which is even better than what is obtained after spec- $z$  filtering. One possible explanation to this is that the objects in the areas of parameter space that are poorly covered by the spec- $z$  catalogue are likely to be faint and at a high spec- $z$ , thus having a less reliable photo- $z$  due either to the lack of suitable SED templates or to the photometry issues.

The aforementioned results for ML photo- $z$  are obtained using only 10 broad and 1 medium bands. In the case of SED fitting photo- $z$ , more than 30 (broad, medium, and narrow) bands were used; still, the statistics for SED fitting and ML are quite close. It implies some consequences for the observational strategies of the future surveys. Depending on the design of a survey, it can be more beneficial to invest resources into obtaining either a larger spectroscopic KB with ML photo- $z$  algorithms in mind or additional medium/narrow-band photometry for SED fitting. Both decisions lead to similar quality of the photo- $z$  catalogue with some differences in the NMAD and in the percentage of outliers. Obviously, obtaining both an extensive spectroscopic catalogue and medium/narrow-band photometry will allow us to use both techniques and to increase the reliability of the photo- $z$  predictions (e.g. Cavuoti et al. 2017b; also see Section 5.4).



**Figure 4.** SOMs built and labelled with KB. The first two rows from the top illustrate mean and standard deviation of the redshifts (spectroscopic on the left, ML in the centre, and SED fitting on the right) within each cell. The third and fourth rows illustrate mean and standard deviation of residuals: on the left, are the maps calculated for the residuals between ML and SED fitting photo-z, and in the centre and on the right are the residuals for ML and SED fitting photo-z. The last plot reflects how many objects are within each cell. Black cells on the occupation and mean maps imply that these cells are empty. On the maps of standard deviation, black cells mean that occupation of the cell equals 1 and standard deviation cannot be calculated.



**Figure 5.** Occupation maps for all data sets projected on SOM trained with run data set.

## 5.2 Spec- $z$ anomalous source filtering

The removal of the spec- $z$  anomalous sources drastically reduces the percentage of outliers for both photo- $z$  methods. As we showed in Section 4.1, the photo- $z$  outliers mostly appear to be misinterpreted spectra, and removing them from the analysis makes the performance estimation more correct. Yet, there are much more objects with high values of  $K_{\text{spec}}$  than ML or SED fitting outliers. A question arises, what are all these objects.

Looking at Fig. 3, we can see that in general the statistics (including NMAD, which is less sensitive to hard outliers) deteriorate smoothly with the increase of the absolute value of  $K_{\text{spec}}$ . This smooth dependence is present not only for ML photo- $z$ , but also for SED fitting photo- $z$ . It means that within each cell of SOM (which represents a small hyper-volume in the photometric parameter space), both ML and SED fitting perform better for ‘typical’ spec- $z$  than for atypical ones. This is true even for those regions of parameter space that are well covered by the spec- $z$  catalogue. In other words,  $K_{\text{spec}}$  works as an additional, finer indicator of whether a given galaxy is well represented in the spectroscopic KB, although it does not offer any insight into the reasons of why this galaxy is not well represented – whether it is due to the physical rarity of its type, to the catalogues’ selection function, or other reasons.

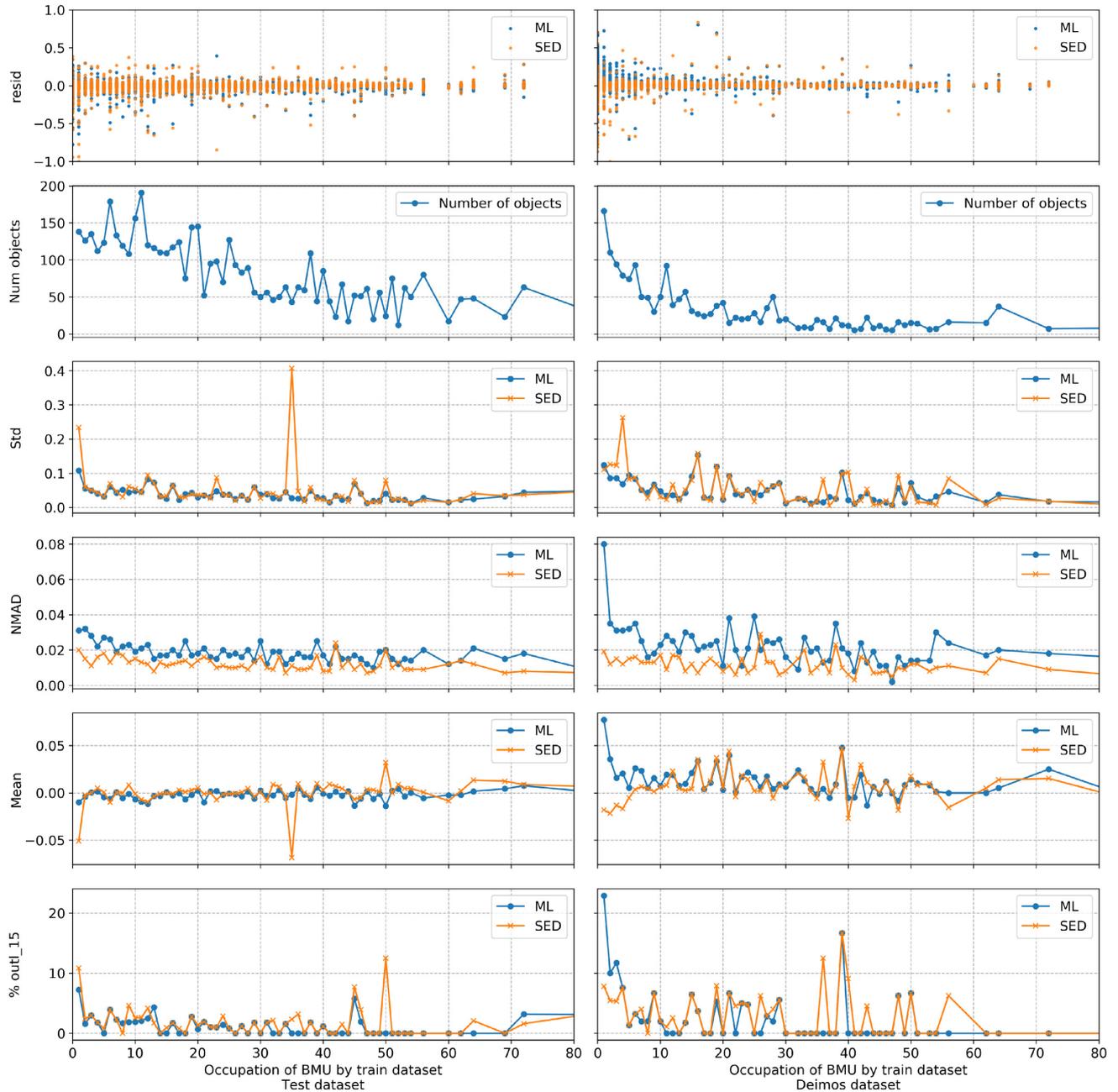
In general, removing objects with high  $K_{\text{spec}}$  allows us to create a reliable spec- $z$  sample, useful for comparing the performance of different photo- $z$  algorithms. In our case, for the test data set this sample is more than two times larger than the set of the objects with multiple spec- $z$  measurements that are in good agreement with each other.

A purely data-driven approach to the selection of reliable spec- $z$ , described in Section 4.5, allows us to enlarge this sample even more, with a comparatively small deterioration of the quality of the data. In every case, the SOM cleaning mostly does not change the overall shape of the redshift distribution of the samples, with an exception of slightly stronger filtering in the lowest and highest redshift bins, where the percentage of outliers is the highest (see online Appendix A for the plots of redshift distributions). Consequently, the usage of this method seems to be highly beneficial for the preparation of any future compound spec- $z$  catalogues, and possibly for the verification of the new spec- $z$  surveys against the old ones.

## 5.3 Photo- $z$ anomalous sources and occupation map filtering

To clean photo- $z$  values for the run data set, the only possibility is to use  $K_{\text{ML}}$  or  $K_{\text{SED}}$  to discard either ML or SED fitting photo- $z$  anomalies. However, the performance is different in the two cases. As it can be seen from Tables 7 and 8, discarding SED fitting anomalies improves the statistics slightly more than removing ML anomalous sources. Obviously, this is related to the large percentage of outliers obtained in SED fitting photo- $z$  and to their large residuals.

On the DEIMOS data set, photo- $z$  anomalous source filtering, together with occupation map cleaning, allows us to reach  $\sigma_{\Delta z} \approx 0.046$  and percentage of outliers  $\approx 1.48$  per cent for ML photo- $z$ , and  $\sigma_{\Delta z} \approx 0.044$  and percentage of outliers  $\approx 1.89$  per cent for SED fitting. At the same time, for ML/SED residuals on the DEIMOS the same cleaning procedure brings  $\sigma_{\Delta z} \approx 0.03$  and percentage



**Figure 6.** Dependence of the statistics for the test and DEIMOS data sets from the occupation of their BMU by objects from train data set. The x-axis is limited to occupation  $\leq 80$  since there are not enough objects in the cells with bigger occupations to calculate reliable statistics.

of outliers  $\approx 0.27$  per cent, which is quite close to the statistics for ML and SED photo- $z$  after occupation map filtering together with removing spec- $z$  anomalous sources.

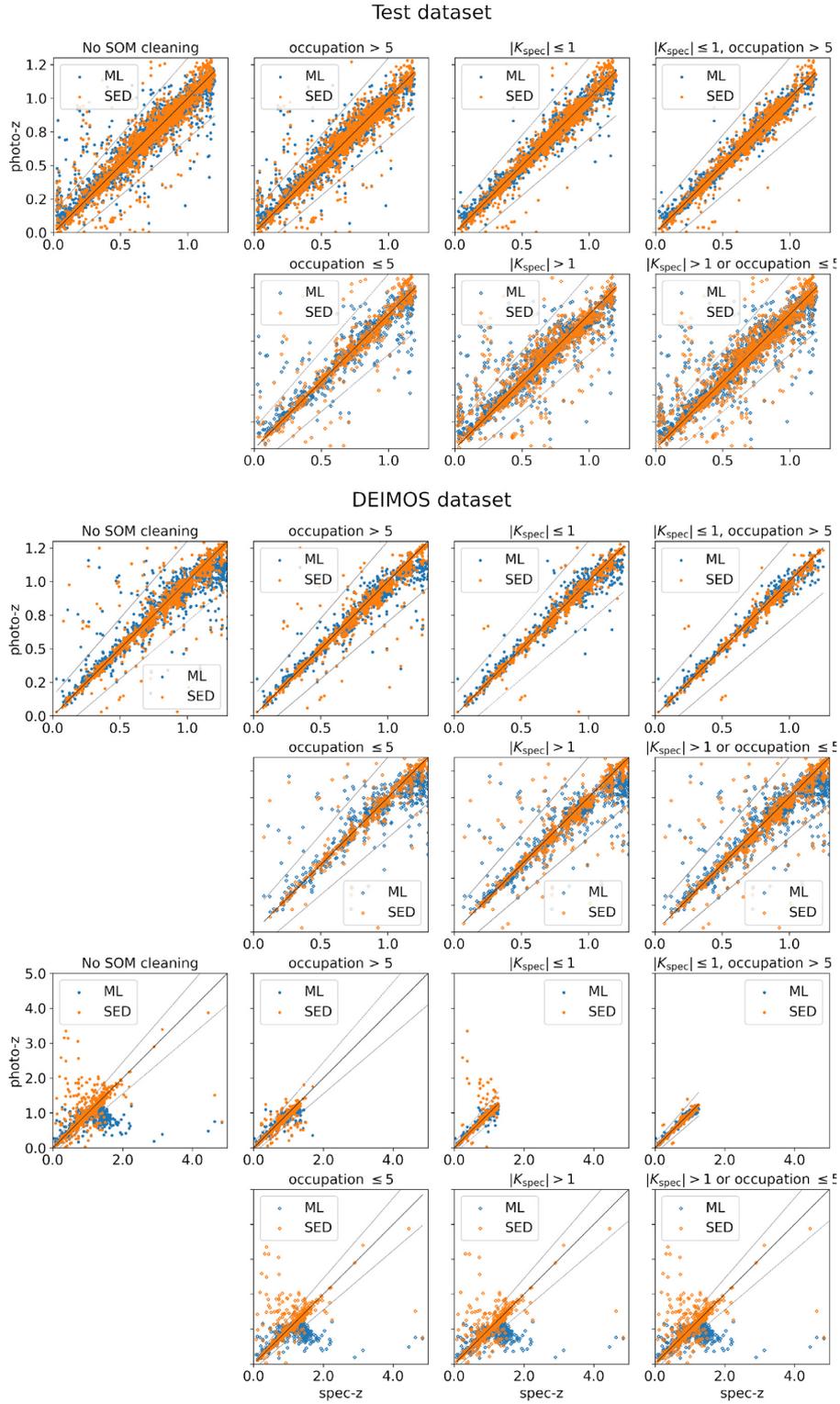
In Fig. 8 we plot the values of different statistics in  $ipmagap3$  magnitude bins at different stages of cleaning of the DEIMOS data set. As expected, with occupation map cleaning the improvements are mostly achieved due to the filtering in the fainter part of magnitude distribution. Instead of completely losing the faint objects, as it would have happened if we used the traditional cut-off procedure, we are preserving those that have fairly good quality of photo- $z$  predictions. Fig. 9 shows the same picture for the run data set. For what the standard deviation is concerned, the difference appears for

objects fainter than  $ipmagap3 \sim 21$ , but the most significant effect is observed for objects with  $ipmagap3 > 23$ .

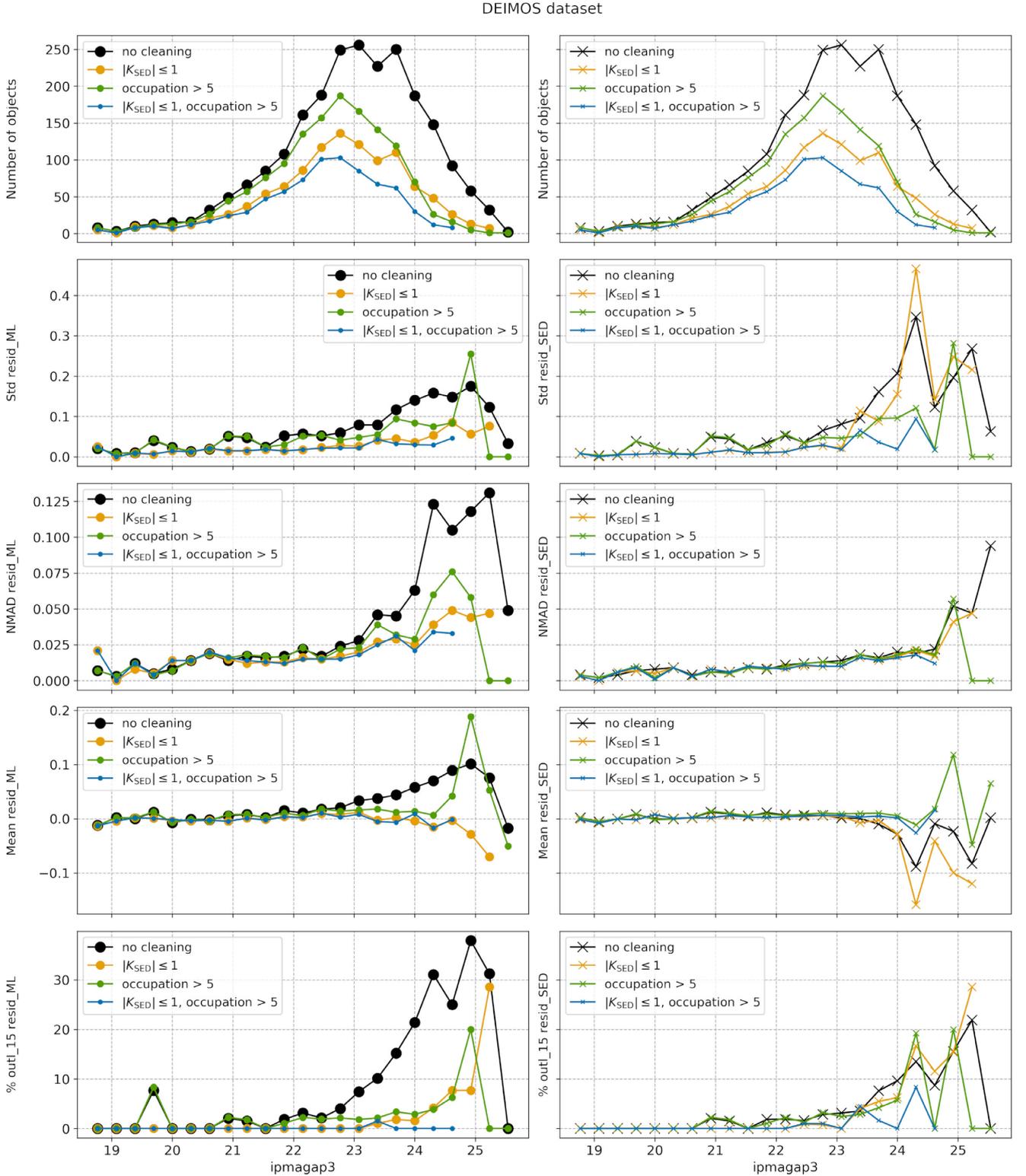
#### 5.4 Filtering strategies

The optimal strategy of the data cleaning with SOM depends on the nature of a data set and on the task at hand:

- (i) Identification of spec- $z$  anomalous sources is useful for finding unreliable spec- $z$  and underrepresented objects.
- (ii) Identification of photo- $z$  anomalies allows us to improve the quality of a photo- $z$  catalogue, especially for SED fitting photo- $z$ . It



**Figure 7.** Scatter plots of ML and SED fitting  $z_{\text{phot}}$  against  $z_{\text{spec}}$ . In the first column on the left are the data sets before SOM filtering procedures, in the second column are the results after only  $z_{\text{spec}}$  outlier filtering, in the third are the plots after only occupation map filtering, and in the last column are the data sets after the two cleaning together. The dotted lines show outlier boundaries defined as  $z_{\text{photo}} = z_{\text{spec}} \pm 0.15 \times (1 + z_{\text{spec}})$ . First two rows from the top correspond to the test data set, the next two rows report the results for the DEIMOS data set, limited by  $z_{\text{spec}} < 1.2$  and  $z_{\text{phot}} < 1.2$ , and the last two rows demonstrate the full DEIMOS data set ( $z_{\text{spec}} < 5$ ,  $z_{\text{phot}} < 5$ ). In every pair of rows, the first row shows the galaxies that remain in the data set after the filtering, and the second row shows the galaxies that have been removed.



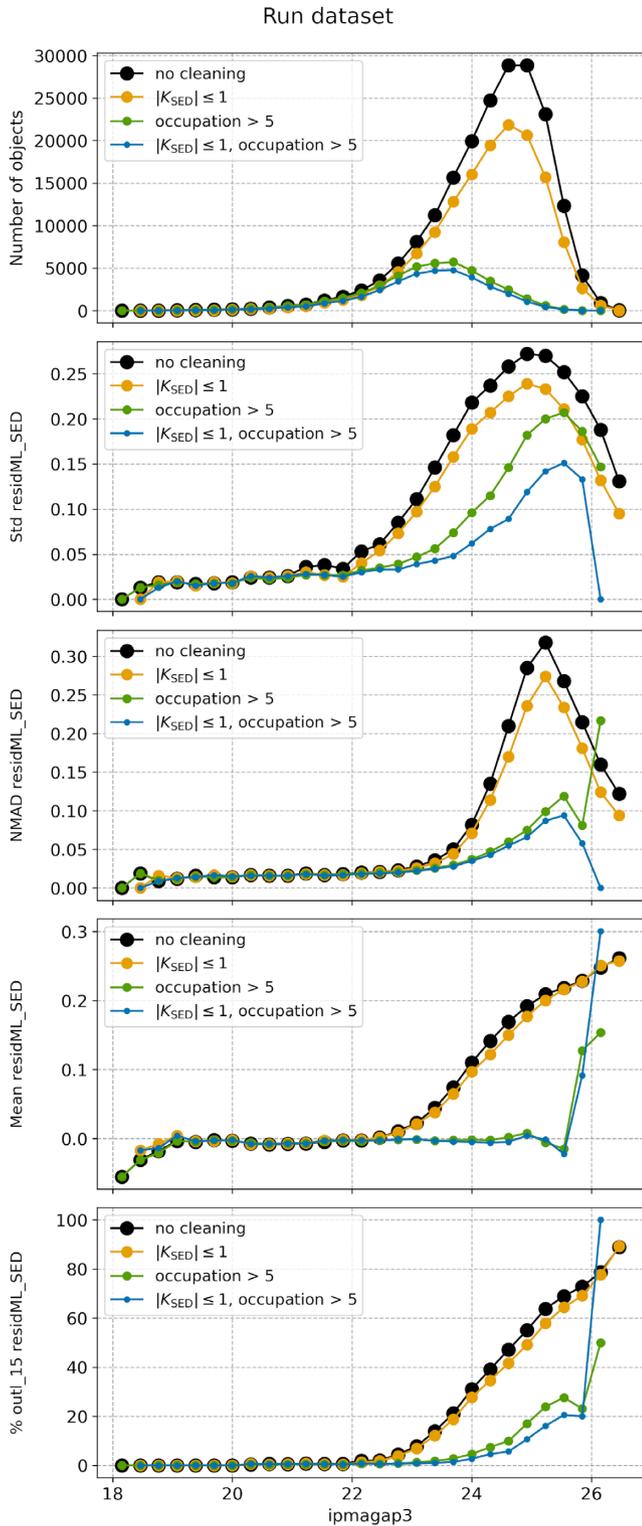
**Figure 8.** Statistics for the DEIMOS data sets in  $ipmagap3$  bins after applying different filters. Left-hand panel: ML photo- $z$  residuals; right-hand panel: SED fitting photo- $z$  residuals. Bins with number of objects < 15 are considered to be unreliable and excluded from these plots.

can be applied to the run catalogue (i.e. for objects without spectral information), but this procedure is effective only for data sets that are well sampled by the KB.

(iii) For the run data sets that are not well sampled by the KB, SOM filtering with an occupation map is the most important step. It

leads to significant improvements not only in the case of ML based methods but also in the case of SED fitting.

In this work, we use all three types of cleaning with the following thresholds:



**Figure 9.** Statistics for ML/SED residuals for run data set in `ipmagap3` bins after applying different filters. Bins with number of objects  $< 15$  are considered to be unreliable and excluded from these plots.

- (i) Spec- $z$  anomalies filtering:  $K_{\text{spec}} \leq 1$ .
- (ii) Occupation map filtering: `trainMapOccupation`  $> 5$ .
- (iii) Photo- $z$  anomalies filtering:  $K_{\text{SED}} \leq 1$ .

(iv) In cases when better NMAD is needed and percentage of outliers is less critical, SED fitting photo- $z$  are preferable. For the tasks that demand the lowest percentage of outliers, ML photo- $z$  show better results. Finally, it is possible to choose between the two values of photo- $z$ : When photo- $z$  predictions are similar (based on the grid search on the test and DEIMOS data sets we recommend to use the criteria of ML/SED residual  $< 0.5$ ), it is better to select SED fitting value, and when the residual is  $> 0.5$ , ML photo- $z$  are more likely to be correct. Table 11 reports the statistics for photo- $z$  selected in this way for the test and DEIMOS data sets. With such selection, we obtain both good NMAD and low percentage of outliers.

## 6 CONCLUSIONS

In this work, we calculated ML photo- $z$  for the COSMOS2015 catalogue using the MLPQNA algorithm. For the training and testing, we used multi-instrument spectroscopic KB with  $z_{\text{spec}} \leq 1.2$  and various sets of photometric bands, obtaining the best results with a feature set composed by 10 broad and 1 narrow bands. The comparison of the statistics for ML photo- $z$  and SED fitting photo- $z$ , calculated by Laigle et al. (2016) using the whole set of COSMOS2015 bands, showed that ML photo- $z$  has lower percentage of outliers and  $\sigma_{\Delta z}$ , but higher NMAD. Particularly, for the test data set without additional SOM cleaning MLPQNA produces photo- $z$  with  $\sigma_{\Delta z} = 0.048$ , NMAD = 0.019, and  $\eta_{0.15} = 1.64$  per cent, while SED fitting photo- $z$  has  $\sigma_{\Delta z} = 0.094$ , NMAD = 0.011, and  $\eta_{0.15} = 2.23$  per cent (Table 7).

We found that the ML photo- $z$  algorithm does not benefit from the inclusion of the most medium and narrow photometric bands (Table 4). Finding a way to exploit the information contained in those bands should be a subject of further work.

Our experiments demonstrated that a significant percentage of outliers have similar values of ML and SED fitting photo- $z$ . By analysing the objects with multiple spec- $z$  measurements, we discovered that the majority of such outliers have unreliable spec- $z$  values, which makes us to believe that the photo- $z$  prediction for these objects is correct and the actual percentage of outliers for both photo- $z$  methods is significantly lower. On the subset of galaxies with multiple similar spec- $z$  measurements,  $\eta_{0.15} \sim 0.2$  per cent for both photo- $z$  methods.

We tested the possibility of using SOMs for removing unreliable spec- $z$  and creating a high-quality spec- $z$  sample. To do this, we calculated a coefficient  $K_{\text{spec}}$  that quantifies how much a spectrum of a given galaxy differs from the mean spectra of all the galaxies belonging to the same SOM cell, e.g. of the galaxies that are most photometrically similar. We found that  $K_{\text{spec}}$  allows us to remove objects with incorrect spec- $z$ : The percentage of outliers for the test set of the KB drops from 1.64 to 0.19 for ML photo- $z$  and from 2.23 to 0.7 for SED fitting, and  $\sigma_{\Delta z}$  for both ML and SED fitting photo- $z$  improves almost by a factor of  $\sim 2$  (see Table 7). At the same time,  $K_{\text{spec}}$  is sensitive to the intrinsic inhomogeneity of the galaxy population, caused by physical reasons. In this way  $K_{\text{spec}}$  serves as a fine indicator of whether a given object is well represented within the KB.

To ensure that our run data set occupies the same area of the parameter space as the KB, we also used SOM following the methodology that we call *occupation map* cleaning. On the control DEIMOS spec- $z$  data set, which is slightly deeper than our KB spec- $z$  catalogue, we found that by using this cleaning, we reduce the percentage of outliers from 11 per cent to 2 per cent for ML photo- $z$

**Table 11.** Statistics for ML, SED fitting, and mixed (selected based on the ML/SED residual value) residuals for the test and DEIMOS data sets. Both data sets were cleaned with the following conditions:  $\text{trainMapOccupation} > 5$ ,  $|K_{\text{spec}}| \leq 1$ , and  $|K_{\text{SED}}| \leq 1$ .

Photo-z	Num. objects	Test data set				DEIMOS data set				
		$\sigma_{\Delta z}$	NMAD	Mean	$\eta_{0.15}$	Num. objects	$\sigma_{\Delta z}$	NMAD	Mean	$\eta_{0.15}$
ML	3508	0.021	0.016	-0.0001	0.03	725	0.024	0.018	0.0020	0
SED	3508	0.049	0.009	-0.0001	0.06	725	0.027	0.009	0.0022	0.55
Mixed	3508	0.018	0.009	0.0012	0.03	725	0.017	0.010	0.0051	0

and from 5 per cent to 3 per cent for SED fitting photo-z, with consequent improvements of other metrics. The details are reported in Table 8.

We find that we can also improve the statistics by excluding the objects with SED fitting photo-z values that are anomalous for their SOM cells. Removing objects with anomalous ML photo-z does not improve the results significantly. Using both occupation map and SED fitting photo-z in-cell anomalies filtering, we are able to bring the statistics for the DEIMOS data set to the order of those for the KB. To be more precise, the standard deviation drops from  $\sigma_{\Delta z} = 0.099$  to 0.046 for ML and from  $\sigma_{\Delta z} = 0.142$  to 0.044 for SED fitting, and the percentage of outliers lessens from  $\eta_{0.15} = 10.86$  to 1.48 for ML and from  $\eta_{0.15} = 5.06$  to 1.89 for SED fitting. This result allows us to select the parts of photometric catalogues for which our photo-z predictions, obtained with any algorithm, can be trusted.

All in all, the SOM in-cell anomaly detection, presented in this work, proved to be a viable method for selecting reliable spec-z samples from a contaminated catalogue and a good tool for identifying SED fitting photo-z outliers. The SOM occupation map filtering also seems to be recommendable for ensuring the reliability of the future photo-z catalogues. We plan to investigate the potential of these methods in application to the other kinds of the astronomical data sets in the future works.

## ACKNOWLEDGEMENTS

We are grateful to the anonymous referee for the suggestions and comments that helped us to improve this paper.

This work has received financial support from the European Union's Horizon 2020 research and innovation program under the Marie Skłodowska-Curie grant agreement number 721463 to the SUNDIAL ITN network. SC acknowledges financial support from FFABR 2017 (Fondo di Finanziamento per le Attività Base di Ricerca). OR thanks Valeria Amaro, Civita Vellucci, Maurizio D'Addona, and Kseniia Sysoliatina for useful discussions and technical help at different stages of this research. MB acknowledges financial contributions from the agreement *ASI/INAF 2018-23-HH.0*, *Euclid ESA mission – Phase D*, and the *INAF PRIN-SKA 2017 programme 1.05.01.88.04*.

## DATA AVAILABILITY

The source catalogues for this work can be obtained as described in Section 2. The final catalogue, containing MLPQNA photo-z and SOM-produced parameters that can be used for selecting objects with high-confidence predictions, will be published via CDS Vizier facility. The code for reproducing this work is available in the GitHub repository ([https://github.com/ShrRa/COSMOS\\_SOM](https://github.com/ShrRa/COSMOS_SOM)). The

MLPQNA software is available within the PHOTRAPTOR<sup>4</sup> (Cavuoti et al. 2015a) package.

## REFERENCES

- Arnouts S., Cristiani S., Moscardini L., Matarrese S., Lucchin F., Fontana A., Giallongo E., 1999, *MNRAS*, 310, 540
- Baldwin J. A., 1977, *ApJ*, 214, 679
- Ball N. M., Brunner R. J., Myers A. D., Strand N. E., Alberts S. L., Tchong D., 2008, *ApJ*, 683, 12
- Baron D., 2019, preprint ([arXiv:1904.07248](https://arxiv.org/abs/1904.07248))
- Baum W. A., 1957, *AJ*, 62, 6
- Baum W. A., 1962, in McVittie G. C., ed., Proc. IAU Symp. 15, Problems of Extra-Galactic Research. Kluwer, Dordrecht, p. 390
- Beck R., Dobos L., Budavári T., Szalay A. S., Csabai I., 2016, *MNRAS*, 460, 1371
- Benítez N., 2000, *ApJ*, 536, 571
- Bilicki M. et al., 2018, *A&A*, 616, A69
- Biviano A. et al., 2013, *A&A*, 558, A1
- Bolzonella M., Miralles J. M., Pelló R., 2000, *A&A*, 363, 476
- Bonnett C. et al., 2016, *Phys. Rev. D*, 94, 042005
- Brammer G. B., van Dokkum P. G., Coppi P., 2008, *ApJ*, 686, 1503
- Brescia M., Cavuoti S., D'Abrusco R., Longo G., Mercurio A., 2013, *ApJ*, 772, 140
- Brescia M., Cavuoti S., Longo G., De Stefano V., 2014, *A&A*, 568, A126
- Brescia M., Cavuoti S., Amaro V., Riccio G., Angora G., Vellucci C., Longo G., 2018, preprint ([arXiv:1802.07683](https://arxiv.org/abs/1802.07683))
- Brescia M., Salvato M., Cavuoti S., Ananna T. T., Riccio G., LaMassa S. M., Urry C. M., Longo G., 2019, *MNRAS*, 489, 663
- Buchs R. et al., 2019, *MNRAS*, 489, 820
- Butchins S. A., 1981, *A&A*, 97, 407
- Capak P. et al., 2007, *ApJS*, 172, 99
- Carrasco Kind M., Brunner R. J., 2013, *MNRAS*, 432, 1483
- Carrasco Kind M., Brunner R. J., 2014, *MNRAS*, 438, 3409
- Cavuoti S., Brescia M., Longo G., Mercurio A., 2012a, *A&A*, 546, A13
- Cavuoti S., Brescia M., De Stefano V., Longo G., 2015a, *Exp. Astron.*, 39, 45
- Cavuoti S. et al., 2015b, *MNRAS*, 452, 3100
- Cavuoti S., Amaro V., Brescia M., Vellucci C., Tortora C., Longo G., 2017a, *MNRAS*, 465, 1959
- Cavuoti S. et al., 2017b, *MNRAS*, 466, 2039
- Civano F. et al., 2012, *ApJS*, 201, 30
- Collister A. A., Lahav O., 2004, *PASP*, 116, 345
- Connolly A. J., Csabai I., Szalay A. S., Koo D. C., Kron R. G., Munn J. A., 1995, *AJ*, 110, 2655
- D'Isanto A., Cavuoti S., Gieseke F., Polsterer K. L., 2018, *A&A*, 616, A97
- Donalek C. et al., 2013, preprint ([arXiv:1310.1976](https://arxiv.org/abs/1310.1976))
- Duncan K. J., Jarvis M. J., Brown M. J. I., Röttgering H. J. A., 2018, *MNRAS*, 477, 5177
- Eriksen M. et al., 2019, *MNRAS*, 484, 4200
- Eriksen M. et al., 2020, *MNRAS*, 497, 4565
- Euclid Collaboration, 2020, *A&A*, 644, A31
- Geach J. E., 2012, *MNRAS*, 419, 2633
- Gerdes D. W., Sypniewski A. J., McKay T. A., Hao J., Weis M. R., Wechsler R. H., Busha M. T., 2010, *ApJ*, 715, 823

<sup>4</sup>[http://dame.oacn.inaf.it/dame\\_photoz.html#photoraptor](http://dame.oacn.inaf.it/dame_photoz.html#photoraptor)

- Gwyn S. D. J., Hartwick F. D. A., 1996, *ApJ*, 468, L77
- Hasinger G. et al., 2018, *ApJ*, 858, 77
- Heinis S. et al., 2016, *ApJ*, 821, 86
- Hoyle B., 2016, *Astron. Comput.*, 16, 34
- Ilbert O. et al., 2006, *A&A*, 457, 841
- Ilbert O. et al., 2009, *ApJ*, 690, 1236
- Ivezić Ž. et al., 2019, *ApJ*, 873, 111
- Kohonen T., 1982, *Biol. Cybern.*, 43, 59
- Kohonen T., 2013, *Neural Netw.*, 37, 52
- Koo D. C., 1985, *AJ*, 90, 418
- Laigle C. et al., 2016, *ApJS*, 224, 24
- Lanzetta K. M., Fernández-Soto A., Yahil A., 1998, in Livio M., Fall S. M., Madau P., eds, *The Hubble Deep Field*. Cambridge Univ. Press, New York, NY, p. 143
- Laureijs R. et al., 2011, preprint ([arXiv:1110.3193](https://arxiv.org/abs/1110.3193))
- Le Fèvre O. et al., 2005, *A&A*, 439, 845
- Lilly S. et al., 2009, *ApJ*, 184, 218
- LSST Science Collaboration, 2009, preprint ([arXiv:0912.0201](https://arxiv.org/abs/0912.0201))
- McCracken H. J. et al., 2012, *A&A*, 544, A156
- Masters D. et al., 2015, *ApJ*, 813, 53
- Masters D. C., Stern D. K., Cohen J. G., Capak P. L., Rhodes J. D., Castander F. J., Paltani S., 2017, *ApJ*, 841, 111
- Newman J. A., 2008, *ApJ*, 684, 88
- Nicastro F. et al., 2018, *Nature*, 558, 406
- Nocedal J., Wright S. J., 2006, *Numerical Optimization*. Springer, Berlin
- Norris R. P. et al., 2019, *PASP*, 131, 108004
- Pasquet J., Bertin E., Treyer M., Arnouts S., Fouchez D., 2019, *A&A*, 621, A26
- 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
- Rhodes J. et al., 2017, *ApJS*, 233, 21
- Rosenblatt F., 1963, *Am. J. Psychol.*, 76, 705
- Sadeh I., Abdalla F. B., Lahav O., 2016, *PASP*, 128, 104502
- Salvato M., Ilbert O., Hoyle B., 2019, *Nat. Astron.*, 3, 212
- Schmidt S. J. et al., 2020, *MNRAS*, 499, 1587
- Scoville N. et al., 2007, *ApJS*, 172, 1
- Vettigli G., 2018, *Minisom: Minimalistic and NumPy-Based Implementation of the Self-Organizing Map*. <https://github.com/JustGlowing/minisom/>, access date July 2019
- Way M. J., Klose C. D., 2012, *PASP*, 124, 274
- Wright A. H. et al., 2019, *A&A*, 632, A34

## SUPPORTING INFORMATION

Supplementary data are available at *MNRAS* online.

**Appendix A.** Additional information on data processing and its effects on the redshift distributions.

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.

This paper has been typeset from a  $\text{\TeX}/\text{\LaTeX}$  file prepared by the author.