The many flavours of photometric redshifts

Mara Salvato[1]*, Olivier Ilbert[2]* and Ben Hoyle[1,3]*

Since more than 70 years ago, the colours of galaxies derived from flux measurements at different wavelengths have been used to estimate their cosmological distances. Such distance measurements, called photometric redshifts, are necessary for many scientific projects, ranging from investigations of the formation and evolution of galaxies and active galactic nuclei to precision cosmology. The primary benefit of photometric redshifts is that distance estimates can be obtained relatively cheaply for all sources detected in photometric images. The drawback is that these cheap estimates have low precision compared with resource-expensive spectroscopic ones. The methodology for estimating redshifts has been through several revolutions in recent decades, triggered by increasingly stringent requirements on the photometric redshift accuracy. Here, we review the various techniques for obtaining photometric redshifts, from template-fitting to machine learning and hybrid schemes. We also describe state-of-the-art results on current extragalactic samples and explain how survey strategy choices affect redshift accuracy. We close with a description of the photometric redshift efforts planned for upcoming wide-field surveys, which will collect data on billions of galaxies, aiming to investigate, among other matters, the stellar mass assembly and the nature of dark energy.

The distance of an extragalactic source needs to be calculated before any more meaningful physical quantities can be inferred. The primary observable used to measure a distance of an object is its electromagnetic spectral energy distribution (hereafter SED), which is composed of a continuum together with emission and absorption lines. The expansion of the Universe stretches the SED toward longer wavelengths by a factor $1 + z$, with z being the galaxy's redshift. The observational difficulty in distance estimation consists of identifying a pair of characteristic features in the SED and measuring the amount they have been stretched. The measured redshift is then related to a proper distance, assuming a cosmological model. Emission and absorption lines are sharp features that can be easily identified in the SED. In addition, two well-known features shape the SED continuum (see Fig. 1): the Balmer break below 4,000 Å, which is explained by the absorption by hydrogen atoms of photons more energetic than the Balmer limit at 3,646 Å and the combination of numerous absorption lines by ionized metals in stellar atmospheres; and the Lyman break below 1,216 Å, explained by the absorption of light below the Lyman limit at 912 Å and absorption by the intergalactic medium along the line of sight.

When SEDs of sufficient wavelength resolution (that is, spectra) are available, the emission/absorption lines can be identified and the redshift precision can be measured to better than 10^{-3} for a resolving power $R = \lambda/\Delta\lambda > 200$ (for example ref. [1]). Despite the efficiency of the current generation of multi-object spectrographs (MOS; for example, VLT/VIMOS, KECK/DEIMOS, SUBARU/FMOS), we can obtain meaningful spectra for only a few per cent of the sources detected in deep imaging surveys, even on 8-m class telescopes[1]. A minimum of two well-identified spectral features is required to obtain a robust redshift measurement. Given the limited spectral coverage of MOS and the limited signal-to-noise ratio in spectra for faint objects, the success rate of measuring spectroscopic redshifts (hereafter spec-z) can be lower than 50–70% in deep spectroscopic surveys[2].

Alternatively, by measuring the flux of an object in broader filters, we can obtain a sparse sampling of the SED sufficient to constrain the continuum shape, define the extragalactic nature of the sources and estimate the redshift from broad features such as the Lyman and Balmer breaks, or strong emission and absorption lines. Such a low-resolution distance estimate is called a 'photometric redshift' (hereafter photo-z). This principle was first applied[3] to measure the photo-z of elliptical galaxies at $z \approx 0.4$. A modern version of the photo-z method has been published by Puschell et al.[4], who were the first to apply a template-fitting procedure to radio galaxies.

The main advantage of using photo-z is that we can derive a distance measurement for all sources identified in an imaging survey. The price paid is the lower redshift precision, which is typically a factor of 10–100 times worse than that obtained with a low-resolution spectrograph[5]. Any photo-z-based study needs an accurate assessment of the photo-z performance, which depends on the image properties (for example, depth and wavelength coverage) and whether the sources are galaxies, active galaxy nuclei (AGN) or stars. Assessing photo-z performance requires a deep and representative spectroscopic sample, which highlights the complementarity of photometric and spectroscopic redshift surveys.

The photo-z method has been adopted as a common tool to estimate galaxy distances, and a simple search of NASA's Astrophysical Data System shows that the fraction of refereed publications including the term "photometric redshift" has increased by a factor of 10 in the past two decades. This can be partly explained by the increasing number of multi-wavelength surveys. In addition, confidence in the method has grown, thanks to reassuringly good comparisons with spec-z results for an always-increasing number of galaxies as faint as the limit of the imaging surveys. Finally, several mature and well-tested photo-z codes are both publicly available and well documented, making the technique readily accessible.

The quality of the photo-z results defines the range of their possible scientific applications. Photo-z estimates are used to study the formation and evolution of galaxies, allowing a statistical analysis of larger samples than those allowed by spec-z, and also offering results in those regions of the colour–magnitude space difficult to populate with spectroscopy (for example, the range $1.4 < z < 2$ is becoming accessible to spectroscopy only now that near-infrared MOS such as KECK/MOSFIRE are available). Some recent applications
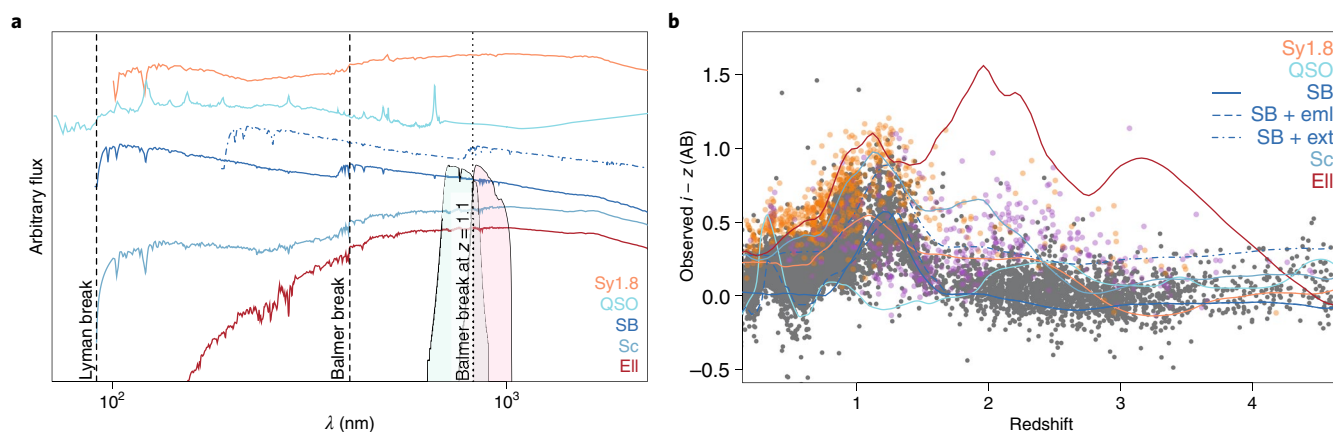
Fig. 1 | The colour–redshift mapping principle. a, Examples of SEDs for various types of galaxies (Ell, elliptical; SB, starburst; Sc, spiral with small bulge) and AGN (QSO, luminous AGN; Sy1.8 (Seyfert 1.8), low-luminosity, obscured AGN). The Lyman and Balmer breaks, among the key features in determining the redshifts, are indicated by vertical dashed lines at rest-frame. One template and the Balmer break are also plotted at redshift 1.1 (dashed-dot blue line and vertical dotted line, respectively). For clarity, the transmission curves of i and z filters, covering the wavelength range between 700 (green) and 1,100 (pink) nm, are also indicated. **b**, The $i − z$ colour as a function of redshift. Galaxies with reliable spectroscopic redshifts are represented by filled grey circles; AGN obscured and unobscured by dust are represented by orange and violet circles, respectively. The solid lines represent the expected redshift evolution of the $i − z$ colour for the templates presented in the left panel, without any extinction. The starburst galaxy is also shown when considering extinction (SB + ext) and considering the contribution from emission lines (SB + eml).

heavily relying on accurate photo-z are the cosmic time evolution of galaxy properties[6], the search for primordial galaxies[7], the study of the relation between galaxy properties and their dark-matter halos[8], and cluster identifications[9]. Galaxy environment or the evolution of galaxy pairs can also be investigated with photo-z (for example, refs [10–12]), although with limited precision[13,14]. Similarly, photo-z results are used for studying the evolution of galaxies hosting an AGN[15] or more peculiar objects such as blazars[16].

Photo-z estimates are also becoming an important tool for studying the properties of dark energy. The weak lensing tomography approach[17] has become one of the main cosmological probes in current and future cosmological experiments such as the Dark Energy Survey (DES)[18], the Hyper Suprime-Cam (HSC) survey[19], Euclid[20] and the Large Synoptic Survey Telescope (LSST)[21] (see section 'The future'). For these applications, the required photo-z precision is not as stringent as for studying galaxy evolution, but several other challenges need to be overcome. Among those are the amount of information to be processed; the need for homogeneous photo-z performances over thousands of square degrees; and the need to characterize the redshift distribution of a population with extreme precision.

In this Review, we provide an overview of the ground principles of the photo-z technique, with details on the most common methods, and with an eye to current and future surveys and scientific application challenging the current state of the art.

Methods for estimating photo-z
We use Fig. 1 to introduce the principles of the photo-z technique. In Fig. 1a, we show how key features such as the Lyman or Balmer breaks in galaxy SEDs can vary in position and shape. The photo-z technique relies on the capacity to isolate the wavelength position of these redshifted features. The breaks correspond to a rapid increase of the flux continuum from the blue to the red part of the SED. The photo-z technique aims to detect gradients between observed fluxes in adjacent filters, which would reveal the presence of a break. Figure 1b illustrates this key principle. The difference between the magnitude seen through the i-band and the z-band filters (the $i − z$ colour) varies with redshift and reaches a maximum at $z = 1.1$ when the Balmer break falls between the i and z bands. Thus, the first rule

when designing a photometric survey is that the filter set should be chosen to encompass key features at the redshift range of interest.

However, the same $i − z$ colour could correspond to multiple redshifts, implying degeneracies in the redshift solution. Such degeneracies can be broken by combining several colours. A second important rule is that the multi-wavelength coverage should be as broad as possible to limit the risk of photo-z degeneracies[22].

The flow-chart of Fig. 2 shows the fundamental steps and ingredients used in all photo-z techniques. At the core there is always a model of the mapping between various colours (or fluxes) and redshift. The redshift solution and its associated probability distribution function (PDF$_z$) are obtained by comparing this mapping and the observed fluxes of the studied source. In the case of template-fitting methods, the redshift–colour mapping is based on physical knowledge built by scientists over time. With machine learning methods, the mapping is obtained anew every time, using a representative 'training sample' of galaxies with both photometry and known redshifts. For both methods the results can be improved by using additional priors (see later in this section for more details).

Although the basic principle is simple, the mapping and its application to the data can be established in many different ways. In this section, we highlight the most commonly used methods.

Physically motivated methods. In this branch of photo-z techniques, the mapping between fluxes and redshift is predicted for extragalactic sources, taking into account physical processes regulating the observed light emission.

Thus, the most basic ingredient is the definition of a set of SED templates, which could be obtained either from theory or from observations. Theoretical templates are generated using stellar population synthesis models (see for example refs [23–26]) which have improved considerably in recent decades. Such theoretical templates rely on numerous assumptions, for instance on the star formation histories of the galaxy. Empirical templates are extracted from observed spectra[27,28], extended over the entire wavelength range with models[29,30]. The quality of photo-z depends not only on the type of templates but also on their optimal coverage of the colour–redshift space[31].
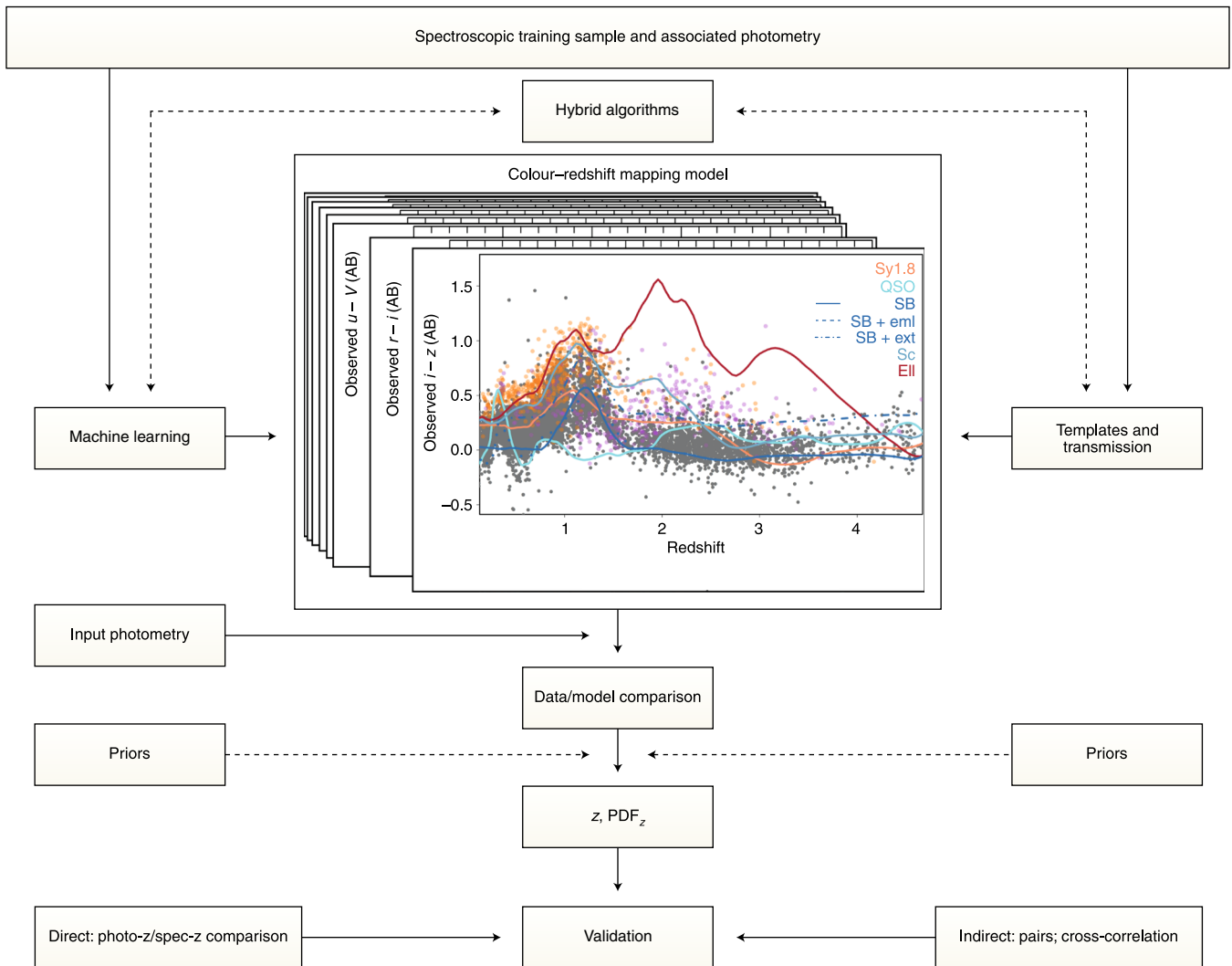
Photo-z in a nutshell

Fig. 2 | Photometric redshift technique in a nutshell. Flowchart representing the various steps involved in computing photometric redshift. The core of machine learning, template-fitting and hybrid methods is the multi-dimensional colour–redshift relation. The spectroscopic sample, essential for machine learning, is optional (although desirable for some tuning) for template-fitting techniques. With the knowledge accumulated over the past decade, we can define reliable priors that are greatly improving our results. After the computation of photo-z and PDF_z, huge effort goes into the validation of the results.

Nebular emission lines emitted by H II regions should also be considered in the templates[32–34]. Considering the contribution of the nebular emission can improve the accuracy of photo-z by a factor of 2.5 for sources at redshift $z \leq 1.5$ (ref. [5]). The importance of this component is easily understood by looking at Fig. 1b. There, the SED of a starburst galaxy is plotted with and without a contribution from emission lines, with the former better matching the colour–redshift map of spectroscopically confirmed galaxies.

To obtain accurate photo-z at $z > 1$, it is important to model the effect of dust attenuation, because optical data will sample the rest-frame ultraviolet part of the SED which is most affected. The dust grains present in the interstellar medium absorb and scatter the stellar light, reddening the SED continuum. Depending on the amount of dust and its grain properties, the impact can be severe (see Fig. 1b). When computing photo-z with template-fitting codes, the dust attenuation is typically modelled as a free parameter, using one of several dust attenuation laws[35,36]. The templates are then corrected for the absorption of the galaxy light when crossing the neutral gas present in the intergalactic medium[37,38].

The light is also attenuated by the dust from our Milky Way along the optical path between the source and the observer. Normally, catalogues are corrected for this effect using dust extinction maps (see, for example, ref. [39]), but a better proposed treatment[40] would also account for the galaxy SED, which should be considered in the correction.

Finally, the redshifted templates should be integrated through the filter transmission curves to produce the modelled fluxes. All other factors that could modify the light distribution should be considered, such as the impact of the Earth's atmosphere, the optics of the telescope, the shape and efficiency of the filter curves, and charged coupled device (CCD) transmission. These various components are usually integrated into one single transmission curve per filter and stored within the code libraries.

As a final note, although this section has discussed the modelling of extragalactic sources, template-fitting techniques also need to consider stellar templates, because we cannot rule out the Galactic nature of the studied sources. Typically, the fit to the stellar templates is performed independently and the extragalactic/Galactic nature of the source is decided a posteriori.
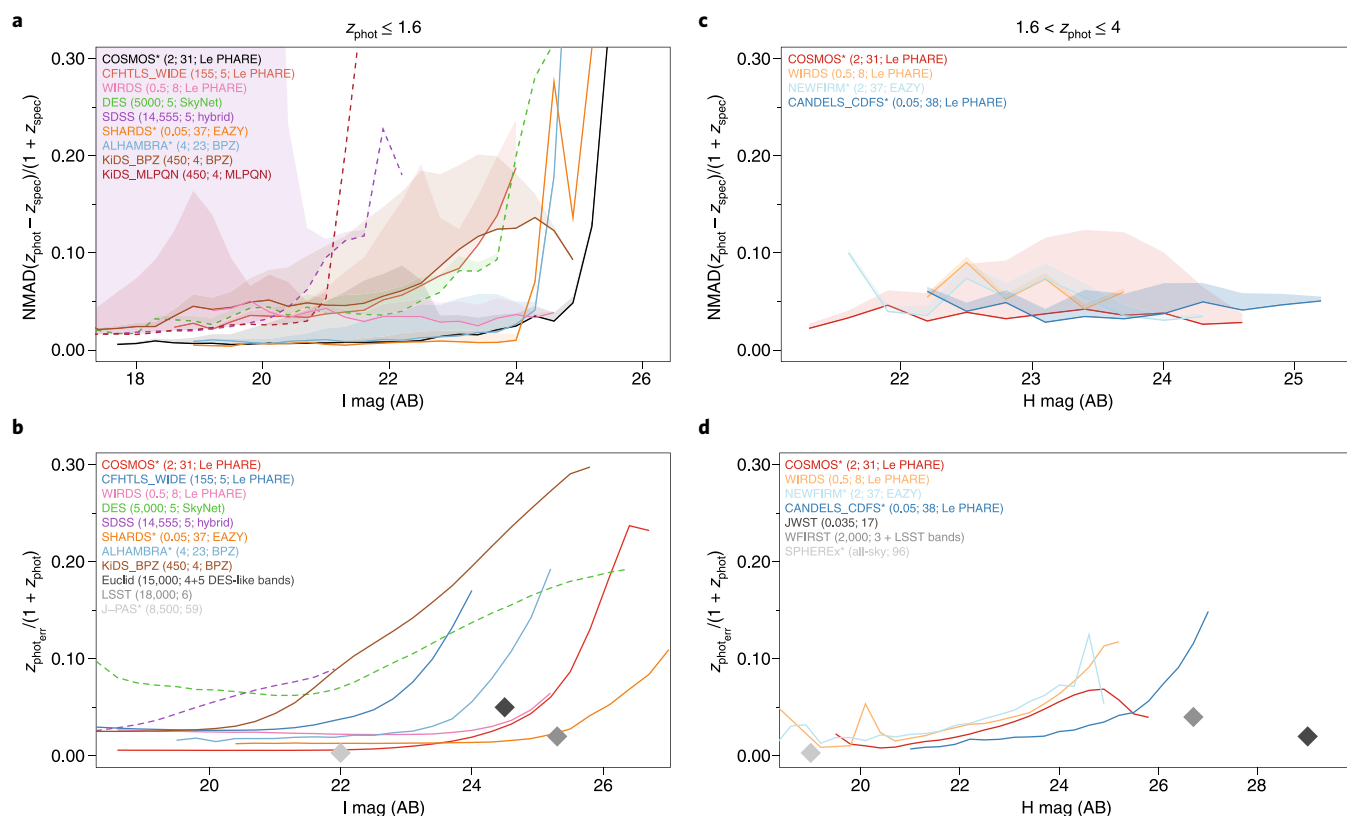
Fig. 3 | Accuracy of various photo-z surveys as a function of their depth and number of bands. a,b, The low-redshift range, where the accuracy is measured as a function of the I band. **c,d,** For the high-redshift range, the accuracy is measured as a function of the near-infrared depth in the H band. In all panels, only sources with a photometric error (denoted phot$_{err}$) smaller than 0.3 mag are considered. The accuracy is measured with normalized median absolute deviation (NMAD), with the solid lines indicating the surveys for which photo-z were computed by template-fitting, and the dashed lines indicating the surveys with photo-z computed by machine learning or hybrid methods. Surveys with intermediate or narrowband photometry (COSMOS[88]; SHARDS[80]; ALHAMBRA[81]; CANDELS/CDFS[78]; NEWFIRM[93]) are highlighted with an asterisk. The height of the coloured area in each magnitude bin is proportional to the number of reliable spectroscopic redshifts used for the training, at that magnitude, and the numbers in brackets beside each survey indicate their size in square degrees, the number of bands and the code used. In general, the less accurate photo-z are at the faint end of each survey. A measure of the accuracy beyond the availability of spectroscopic redshift is provided in **b** and **d**, where the solid lines show the median photo-z error divided by $(1 + z_{phot})$, as a function of the depth of the survey. It is clear that the accuracy in all panels is better in the region well populated by the spectroscopy, where the sources are brighter. Surveys with intermediate and narrowband photometry perform better than those having only broadband photometry (WIRDS[135]; CFHTLS_WIDE[89]; KiDS[136]; SDSS[133]; DES[85]). Finally, the grey symbols indicate the expected accuracy for future cosmological surveys.

Data-driven methods. Rather than relying on any physical models, one may estimate photo-z using data-driven approaches. The most common method is machine learning (hereafter ML). ML algorithms can be categorized into unsupervised and supervised learning, the difference being that unsupervised learning requires only photometry, whereas supervised methods require both photometry and reliable spec-z for a data sample during the training of the algorithm. In the photo-z framework, these training samples allow the algorithm to learn the mapping between colour and redshift (see, for example, Fig. 1).

In general, all supervised ML methods perform function approximation, and for photo-z this entails finding a function that maps between the multi-dimensional photometry space (for example, fluxes or colours) and the redshift values of the training sample. The function is assumed to be 'well behaved' between the multi-dimensional locations of the training data, and thus supervised ML methods can be viewed as performing high-dimensional interpolation. Once this function has been learned, the photometry of a source with an unknown redshift is localized in the multi-dimensional photometry space and paired with a corresponding redshift value or redshift probability distribution. Owing to the interpolative nature of supervised ML approaches, one must ensure that the

space of photometric properties of the sample for which predictions will be made is well sampled by the training examples[41–43]. Failing this, the algorithms may perform extrapolation and lose accuracy. Furthermore, the complexity of the (unknown) manifold to be approximated using ML provides an estimate of the number density of training data that are required in each region of colour space.

Two popular supervised ML systems are often used for photo-z computation: random forests[44] and neural networks[45], which we briefly describe below.

Random forests[46,47] are aptly named because they are constructed from collections of 'decision trees', and each tree is grown using a random subset of the training data. Decision trees are a set of logical if-statements that group the properties of the training sample into cells, which are defined to minimize the spec-z dispersion of the data in each cell. Successively, each cell is assigned the average value of the spec-z. Once all the trees have been grown, predictions are made on the redshift of the sources in a test sample by passing their photometric properties through each tree and averaging over all of the obtained redshifts. The number of trees and the maximum number of cells are some of the variables of the algorithm.

Neural networks are a class of popular ML algorithms that were inspired by how neurons in animal brains process information

and learn. Neural network algorithms[42,48–51] make increasingly complex nonlinear matrix transformations of the input properties. The structure consists of different layers, and each layer can be viewed as a new transformation of those that come before it. The output of the neural network is defined by the user and represents the target value of interest, such as the redshift value. The neural network algorithm tunes the matrix transformations using the training data, such that the output value minimizes the residuals between true redshift and predicted redshift.

Finally, deep machine learning (DML), the current state of the art in computer science, has also been applied to the problem of photo-z estimation[52,53]. DML is based on normal neural networks but with many thousands of neurons in each hidden layer. In particular, DML may also work directly from galaxy images. One benefit of using DML comes from working directly with the images, which brings the redshift analysis pipeline closer to the data. This is in contrast with template techniques, which must standardize catalogue-level photometry before combination, for example by remeasuring fluxes in a particular aperture.

Given that supervised ML methods require training data with both photometry and spectroscopic redshifts to approximate the mapping between colours and redshift, they are of course limited by the nature of the training data, often to low redshift and bright luminosity (see Fig. 3). Given the dearth of both high-redshift galaxies (for example, $z > 1$) and very faint galaxies, ML methods should be used with caution in these regimes.

Unlike supervised ML, unsupervised ML algorithms do not use spec-z in the training phase, but rather perform clustering in the input data space to identify groups of similar objects. We note that unsupervised methods can not produce redshift estimates without further post-processing steps using labelled data. The most popular example in the photo-z literature is the self-organizing map (SOM). A nice introduction to SOMs can be found in refs [54,55].

All ML algorithms used for photo-z are able to generate PDFs of varying sophistication. The simplest method is to randomly sample from values and errors of each input parameter many times, and associate the normalized distribution of the resulting redshifts as a PDF[47,55,56].

Other exploratory techniques from ML have also been ported to photo-z estimation, such as anomaly detection, data augmentation and feature importance[57–59]. In turn, these techniques allow ML methods to identify problematic training data, provide some extrapolation ability outside the original training sample and motivate the choice of input parameters that provide the most photo-z predictive power. As in the case of template-fitting methods, ML can also improve on the results by considering additional information as a prior, often in the form of re-weighting the training data[60].

Using additional data as prior information. Many photo-z codes produce a redshift z that is the maximum of the likelihood $L(z|T,D)$ given the data D and galaxy type T or choice of SED. This redshift inference can be improved by adding other prior information based on the empirical knowledge that we have acquired over time of the redshift evolution of various galaxy properties.

In Bayesian inference, the redshift posterior distribution is computed by multiplying the likelihood function with a prior and marginalizing over all other parameters. In this integration, the priors on the parameters define how to weigh different volumes of the parameter space a priori. Such a Bayesian approach has been established in the public code BPZ[61] by introducing a prior on the redshift distribution per galaxy spectral type and magnitude bin. Other choices of prior have also been recently made, such as the luminosity function for the GOODZ code[62] and the relation between galaxy mass and star formation rate in ref. [63]. One of the most recent public codes is BEAGLE[31], which is among the few that adopt a fully Bayesian approach, and thus a prior can be assigned

to each parameter. Priors should be used with care given the risk of introducing inaccurate information (for example, hardly anything is known about the redshift distribution beyond the spectroscopic limits). Still, in the case of sparse wavelength coverage which produces several redshift peaks in the PDFs, Benitez et al.[61] showed that using the most probable solution based on prior knowledge helped to reduce the fraction of catastrophic failures for the average population, at a price of associating the wrong photo-z to specific populations (such as high-z sources, or AGN). Independent of the type of prior, we believe that it is the responsibility of authors to describe in detail the priors that they adopt and to discuss the implications that the prior may have for the photo-z results.

Introducing spatial information. The correlation of source positions can also be used as a redshift estimation technique. This method uses the fact that galaxies are not distributed randomly in the Universe but reside in large-scale structures. By using the spec-z of a reference sample, the redshift distribution of an unknown sample can be found by maximizing the spatial cross-correlation signal between the unknown and the reference samples. As early as 1979[64], galaxy clustering was used to extract information on the distances of quasi-stellar objects or QSOs (bright AGN outshining their host). This method has been mainly revived in the past decade to estimate the mean redshift of photo-z selected samples for weak lensing applications[65]. Although such applications usually consider only the large-scale clustering, Menard et al.[66] generalize the method to use clustering on smaller scales as well. However, such an approach requires the unknown sample to be preselected in a narrow redshift range, in photo-z or in colours[67]. By considering an extremely narrow sample of preselected galaxies, it even becomes possible to measure individual redshifts[68].

Aragon-Calvo et al. have proposed[69] a new estimate of photo-z using the cosmic web. Such a method requires a dense spectroscopic coverage to characterize accurately void regions, filaments, walls and clusters. The redshift is obtained from the product between the photo-z derived from the colours, the cosmic web and the density field. One limitation of the method is the need to use accurate photo-z (better than 1%) from colours to avoid a catastrophic association between the photo-z and the wrong cosmic web structure. For 3% of the SDSS multi-colour catalogue, Aragon-Calvo et al.[69] were able to apply this method, and show that an accuracy as good as 10^{-5} to 10^{-4} was possible.

Which method for which survey

But how can one decide between different methods? Is template-fitting better than ML, or vice versa? Or is there a favoured one among the SED-fitting and ML codes? In general, direct and fair comparison between performances of different template-fitting and ML codes is possible only within tests specifically designed for this purpose[41,43,70–72]. Here we list the most used and/or public codes, together with their key features, pointing to papers that attempted a more quantitative comparison of the various algorithms. More on this is discussed later.

Template-fitting codes. Numerous template-fitting codes are publicly available, with Hyperz[73] being a precursor. Although the ultimate common goal of the codes is to compute photo-z, each code has developed its own specificity. For example, template libraries change from code to code. EAZY[74] combines basic templates and creates new ones on-the-fly, whereas ZEBRA[75] trains the templates using a spectroscopic sample (a risk when the spectroscopic sample is not representative of the entire population; see figure 11 and related text in ref. [76]). Some codes (for example BPZ[61]) include dust attenuation in their templates whereas others such as Hyperz include dust as a free parameter. As yet, no papers have clearly demonstrated that one template set is superior to another. Still,
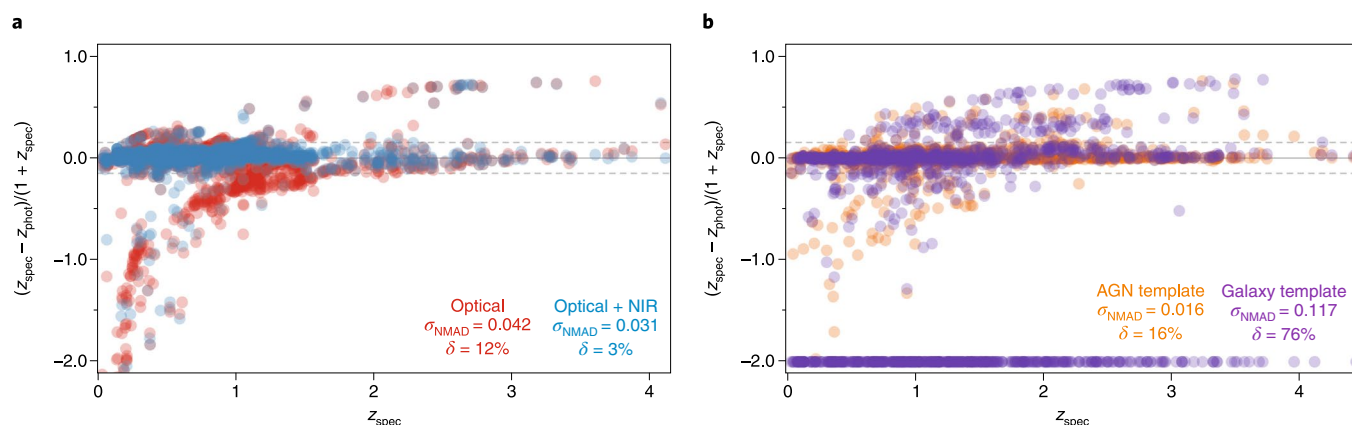
Fig. 4 | Impact of multi-wavelength coverage and choice of templates. a, Comparison of photo-z for galaxies computed using the optical bands (*ugriz*; red) or the optical and near-infrared (NIR) (*ugriz*JHK; blue). The addition of NIR photometry greatly improves the precision at $z > 1$. Sources external to the dashed lines at $\pm 0.15(1 + z_{spec})$ are considered outliers. **b**, Difference between photo-z obtained using either galaxy (violet) or AGN (orange) templates for sources that are X-ray detected in Chandra Legacy COSMOS and with a reliable spec-z. For the 33% of sources for which no normal galaxy template could be used to compute the photo-z, we artificially set the photo-z to −2.

EAZY includes an error function associated to the templates, which improves the quality of the PDF.

Depending on the dataset and the scientific objective, some features are crucial to improve the photo-z accuracy. It is now widely accepted that galaxy photo-z performance improves when emission lines are included in the templates[33], as in Le PHARE[32,77] or EAZY. This is particularly true at high redshift and when photometry from narrow and intermediate bands is used, like in COSMOS[5,78], MUSYC[79], SHARDS[80] and ALHAMBRA[81].

Another example of adapting the code configuration to the photometric data available is the use of priors. For a faint survey at $I_{AB} <$ 25 mag and with only *g*, *r*, *i* and *z* bands[63], it was possible to decrease the fraction of outliers from 45% to 25%. But if the survey provides extensive multi-wavelength coverage including NIR, adding a prior is usually not considered, meaning that an equal probability is associated to each value of the parameter. Another important aspect is the sensitivity of the photo-z bias to the absolute flux calibration of a survey: Ilbert et al.[32] showed that a few per cent uncertainty in the absolute calibration of the photometry could create a bias of the same order as the photo-z precision even for bright sources. This calibration issue can be corrected using a spec-z training sample, with a overall improvement in accuracy by a factor of 2 for sources with $I_{AB} < 24$ mag. The zero-point calibration has also been pointed out[71] as a way to reduce the fraction of outliers and to increase accuracy. We caution the reader that such calibration could also be created by an incomplete or biased colour–redshift model.

ML codes. Choosing which ML algorithm will provide the best performance for a particular dataset or problem is nontrivial. Results from computer science show that if the training sample is large enough, and the training time of each algorithm is long enough, then the performance from different algorithms converges (see, for example, refs [43,58] for a photo-z setting). If the training samples are small, or the training times are limited, then the performance of different codes depends on the complexity of the underlying surface to be learned. Given the plethora of ML algorithms on the market, each group typically considers only a few algorithms for any given problem, often chosen from those with which they have obtained good results in the past. Furthermore, the algorithms that are finally implemented in large-scale astronomical surveys, such as the TPZ[47], SkyNet[51] and aNNz2[82] methods for DES, are often those in which the code authors were part of the collaboration and had a vested interest in seeing their codes advanced. A good starting point for

deciding which ML algorithms should be applied to new datasets is to use those that are very fast to implement so that any subsequent performance can be easily benchmarked. For this purpose, we would suggest starting with decision-tree-based algorithms, such as random forests.

SED-fitting or ML?. Generally, the choice between template-fitting and ML methods strongly depends on the scientific application and on the spec-z sample available for the training. For instance, ML applied to SDSS data outperforms template-fitting for low-redshift sources. This is because the relatively few optical bands are still sufficient to correlate accurately with redshift, and the spectroscopic sample used for the training is rich and complete. But in deep pencil-beam surveys covering a large redshift range, the reverse is the case[41]. In other words, the use of ML is limited to those surveys with a sufficient training set, whereas if the scientific objective is to study galaxy population with limited spectroscopic coverage, template-fitting code should be favoured.

Template-fitting codes could also model uncertainties in the absolute calibration, but ML methods are insensitive to photometry biases depending on magnitude or colours. Therefore if the goal is to limit biases over a large field well covered with spectroscopy, ML would normally be the favoured approach. In terms of speed, ML clearly outperforms the template-fitting techniques[49]. One of the greatest benefits of ML algorithms comes from their optimization to work smoothly on massive datasets. They have been specifically developed to process massive volumes of data and can easily accommodate current and future astronomical-sized datasets (see, for example, ref. [49]). Finally, missing photometric points affect the estimation of photo-z and affect the PDFs. All techniques should correctly account for missing data, and although there is consensus on the correct approach within template methods, this is still an evolving topic for ML methods[83].

Validation and state of the art
An extensive characterization of the photo-z performance should be made for every catalogue and sub-sample of interest. It is typical to compare a photo-z point prediction (such as the mean or mode of the PDF, z_{phot}) with the spec-z (z_{spec}) as a way to assess performance. Note that any spec-z used for ML training, or prior construction, should generally be excluded from the validation sample.

The following three measures are often adopted to assess photo-z performances:

- Bias: defined as $<z_{phot} - z_{spec}>$, this characterizes the average separation between the predicted and the true redshift[84].
- Precision (σ_{zp}): is often defined as the standard deviation of ($z_{phot} - z_{spec}$)/(1 + z_{spec}) or as 1.48 × median($|z_{phot} - z_{spec}|$/(1 + z_{spec})) (ref. [32]), the latter being less sensitive to outliers. The precision describes expected scatter between predictions and truths.
- Outliers (or catastrophic failures) fraction: usually defined as the fraction of sources for which $|z_{phot} - z_{spec}| > N\sigma_{zp}$, where N is decided by the authors and σ_{zp} is the standard deviation (for example ref. [71]), or as $|z_{phot} - z_{spec}|$/(1 + z_{spec}) > 0.15 (for example ref. [41]). It highlights the fraction of sources with unexpectedly large errors.

These metrics are not unique (see, for example, discussion in section 4 of ref. [71]), and they could be science-dependent. Because they are the most used, however, they allow an easy comparison between surveys.

We note that photo-z performance measured by using a direct comparison between photo-z and spec-z is less meaningful if the spectroscopic coverage is not fully representative of the entire photo-z sample—for example if the spec-z sample is biased toward bright sources (see, for example, refs [42,85]).

When this happens, a statistical mapping between the photometric parameter space covered by the two samples should be defined[55]. For regions of photometric space without spectroscopic coverage, alternative methods have been developed to establish the photo-z performances. For example, the galaxy close-pairs technique[78,86] uses the fact that close pairs have a significant probability of being associated and that they therefore should have similar redshifts. The comparison of the spatial cross-correlation between two redshift bins with that expected from a model may also be used to measure redshift precision and contamination[87].

Characterizing the performance of the full photo-z PDF is slowly becoming popular, as it can highlight the lack of templates or training sample, or sub-optimal redshift prediction routines. One common problem affecting PDFs is caused by underestimated photometric uncertainties, creating a PDF peak that is too narrow[71]. Authors often validate the PDFs by ensuring that 68% of the spec-z falls within the 1σ uncertainties derived from the PDF. Bordoloi et al.[84] introduced the statistical method of the continuous rank probability score (CRPS) to the photo-z community as a means of testing PDF performance. The CRPS asserts that the value of the cumulative PDF evaluated at the location of the spectroscopic redshift must be uniform for an ensemble of galaxies.

State of the art. The wavelength coverage (see Fig. 4a) and the quality of the input data determine the accuracy of the photo-z, independently of the merits of each redshift code. One of the main advantages of photo-z is the ability to estimate distance information for faint sources, provided that the photometric errors associated with the measures are small enough to constrain the redshift by limiting the degeneracy in the solutions. Hence, because the photometric errors are larger when the sources in an image are faint, the only way to keep the photometric errors small (and thus keep the accuracy of photo-z high) is to obtain images as deep as possible (see Fig. 3). The photo-z estimates in CANDELS represent one of the deepest photo-z samples available today, with the precision σ_{zp} increasing from 0.040 (8%) to 0.055 (28%) between H-band magnitudes H_{AB} < 24 and 26 < H_{AB} < 28, respectively. Over large fields of several square degrees, $\sigma_{zp} \approx 0.05$ is routinely reached at I-band magnitude I_{AB} < 26 (see, for example, ref. [88]).

With broadband photometric data alone, the best σ_{zp} is limited to approximately 0.025 (see Fig. 3 for results on pencil-beam and wide-area surveys), and this does not seem to improve even for large signal-to-noise values (for example, >40)[89]. In the past decade, medium-band data with filter widths around 400 Å have enabled a breakthrough in precision by improving the SED resolution. COMBO-17[90] was the first survey to use medium-band imaging to produce a photo-z catalogue. There, σ_{zp} of bright sources (I_{AB} < 22) reached about 0.02, thanks also to the ability to locate the Balmer break precisely. The COSMOS, SHARDS and ALHAMBRA surveys improved upon this precision to about 0.01 out to a redshift of about 1.5 with deeper medium-band photometry and by introducing emission lines into the templates.

The increase in sensitivity of NIR detectors has allowed astronomers to extend their study beyond the redshift desert by following both the Balmer break in the range 1.6 < z < 4 (for example, ref. [91]) and the Lyman break above z = 8 (ref. [92]). In this way, high-precision photo-z ($\sigma_{zp} \approx 0.02$) has been reached for 1.6 < z < 4 for massive galaxies in the NEWFIRM survey[93].

Photo-z results enable the identification of galaxy samples across thousands of square degrees. For example, the DES survey is reaching depths of I ≈ 24 over 5,000 deg², with a photo-z accuracy expected to be below about 0.08 (ref. [94]). The main challenge for wide-field surveys is the calibration of photometric noise precision over large areas being observed over timescales of many years with varying sky conditions, together with potential degradation in quality of the instruments. Also, the spectroscopic training sample should not be limited to a small area of the sky but homogeneously distributed.

Finally, beside these other difficulties, the procedure used to extract the photometry from images is also crucial. Source extraction is commonly performed with SExtractor[95], but the region within which the galaxy flux is measured is important because too large an area would compromise the signal-to-noise of faint sources[96,97]. Moreover, the light should be produced by the same regions of the studied sources despite the point spread function (PFS) variation from one band to another. In cases of PFS variation between the various bands, PFS homogenization is necessary[5,88,96,98], and efficient tools are being developed to produce fluxes measured in a consistent way using the high-resolution images as reference (see, for example, PYGFIT[99] Synmag[100], TFIT[101], T-PHOT[102]).

Photo-z of exotic sources

The discovery that virtually every galaxy hosts a supermassive black hole[103] suddenly increased the interest of the scientific community in those galaxies hosting an AGN, now seen as a key ingredient in galaxy evolution models. Therefore, computing their redshifts became crucial. But, although the number density of AGN is high enough that spec-z follow-up with single-slit spectroscopy would be too time-consuming, they are so sparse that a multi-object spectroscopic campaign would be inefficient. Thus, the photo-z technique is favoured, although these sources are particularly challenging. In fact, their SED is characterized by the sum of two unknown relative contributions—host and AGN—and, in any photometric band, this depends on the type of host and the type and strength of the AGN (see, for example, figure 5 of ref. [104], and related text). When these constraints are ignored, and for example only templates of galaxies are used (that is, assuming a dormant black hole), the photo-z obtained with SED-fitting will be characterized by a high number of degenerated photo-z solutions, or, worse, catastrophic failures. This is demonstrated in Fig. 4b, where we show the difference between photo-z computed using either AGN[105] or galaxy[88] templates, for a sample of 1,672 X-ray selected AGN with a secure spec-z from the Legacy Chandra COSMOS field[106]. In particular, the discrepancy between the photo-z solutions often increases with the strength of the AGN[105].

The choice of templates to include in a library is dictated by the type of AGN that are treated (for example, AGN classified according to their X-ray hardness ratio[107]), by the selection technique (for example, excess flux, variability) used for the identification and by the selection band (for example, X-ray, radio, optical, infrared) and
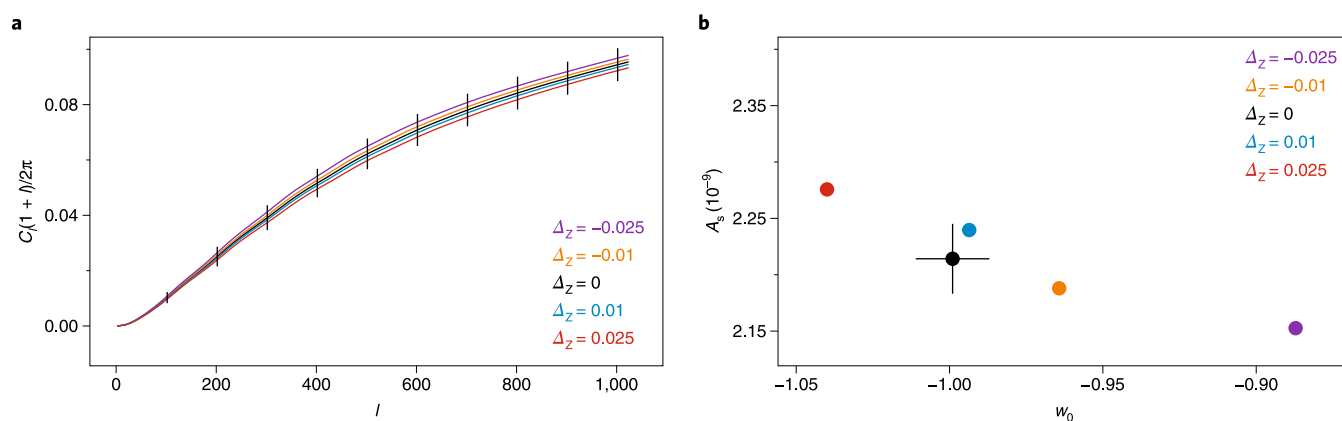
Fig. 5 | Cosmological parameter biases incurred by photo-z inaccuracy. a, The predicted autocorrelation function (C_l) as a function of inverse scale l, is shown for a simulated Euclid-like redshift distribution (black line) and for a redshift distribution modified by a translation of ±0.01 and ±0.025. **b**, The subsequent biases induced in the cosmological parameters, w_0 (the equation of state of dark energy) and A_s (the amplitude of the fluctuations on the surface of last scattering) by the redshift translations as shown by the different values of Δ_z. The plotted bars are representative of the typical uncertainties for all the points.

its depth. This is clearly demonstrated in figure 14 of ref. [78], in which the X-ray flux distribution of the AGN detected in various X-ray surveys is shown. The population of sources that are represented in the wide but shallow XMM-COSMOS[108] is very underrepresented in the deep but narrow Chandra observations of the Chandra Deep Field South field[76]. Reciprocally, most of the sources in the latter field are not represented at all in the former one. Similar results have been shown recently[72] for AGN identified at radio frequencies. This is particularly relevant considering that wide or all-sky surveys will observe the sky at X-ray wavelengths (for example, eROSITA[109]) or in the radio (for example, EMU[110]). These surveys are expected to detect millions of AGN that will be used as cosmological probes and for evolutionary studies.

The problem is alleviated in rich photometric datasets that are inclusive of both intermediate and narrowband filters because they help to identify emission lines[78,79,111], which are the only feature visible in the power-law continuum of a QSO for which colours are independent of redshift (see, for example, the cyan track in Fig. 1b).

Priors are a crucial ingredient when computing photo-z for AGN by template-fitting. Specifically, given the absolute magnitude of these sources (typically brighter than -22 in blue optical bands), a prior based on the apparent size allows one to limit the range in redshift solution and thus the degeneracy[105,111,112].

ML techniques have been also used to estimate photo-z for AGN, but the most reliable results are for QSOs at high redshift, where the host contribution can be neglected. Several examples of successful applications[113–115] are focused on the SDSS footprints, in which the plethora of spectroscopic follow-up data are well suited to ML techniques. Very recently, ML has been applied[116] to a sample of bright X-ray selected AGN, providing reliable results even at low redshift.

An additional problem when computing photo-z for AGN is that variability across the spectrum is an intrinsic property of the sources, with different cadence and intensity. This makes it difficult to estimate photo-z reliably when multi-wavelength data are collected over timescales of years[117].

Variability affects not only QSOs but also objects such as BL Lacertae objects (BL Lacs), gamma-ray bursts (GRBs) and supernovae. BL Lacs belong to the family of AGN, in particular blazars — sources that are observed through the jet launched from the centre — and for this reason are characterized by a featureless power-law continuum, unusable for assessing redshift through spectroscopy. Here, simultaneous observations in multiple bands are required in order to measure the absorption bluer than Lyman α by the intervening material[118]. GRBs and supernovae are sudden, rapidly

occurring explosive events that allow us to measure the distance of galaxies that would be otherwise too faint to be detected. For GRBs, we rely on simultaneous observations in the optical and the NIR and rely on SED-fitting templates[119], while the most reliable photo-z for supernovae rely on the use of priors[120].

The future

Where next for the photo-z technique?

Reaching the deep Universe. The photo-z technique gained momentum when the first images from the Hubble Deep Field became available in 1995. A first photo-z catalogue[121] was published to study 1,683 galaxies out to $z = 6$, which was an incredible leap forward in studying the high-redshift Universe. Such analysis was followed by numerous attempts to make the photo-z estimation technique more robust[73,77] and to exploit the following generations of HST deep surveys[71]. The template-fitting technique has been central, as only sparse spectroscopic coverage is possible in these deep fields. The study of primordial galaxies will trigger a new burst of activity in the next years with the launch of the James Webb Space Telescope (JWST)[122]. With the efficient near-infrared NIRCAM camera, surveys could be conducted at depths of AB magnitude around 30–31 by adopting a strategy similar to the CANDELS survey. NIRCAM will allow reliable measurement of photo-z above 5 and out to ~20. (ref. [123]).

Based on the HST experience, template-fitting will probably remain the main photo-z technique in these extremely faint fields. However, new difficulties will arise because of the increasing contribution of the emission lines with increasing redshift[124] and an evanescent Balmer break feature. Although emission lines create gradients in the colour–redshift relation that could be used to measure a redshift[125], this also makes the estimate more challenging because it is more prone to degenerate solutions[123]. The knowledge of the galaxies and space at redshift beyond $z = 10$ will soon be revealed, thanks to the deep (28 AB magnitude in NIR for a signal-to-noise ratio of 10, in 10^4 seconds) for spectroscopy that JWST will perform in optical and NIR. Template-fitting methods will then have to incorporate the new knowledge in the form of new templates and new physics.

Building blocks to unveil new cosmology. The next decade will see many exciting imaging surveys dedicated to cosmology. The goal of these surveys is to constrain the nature of the dark energy through the combination of various probes. Several probes rely on

photo-z, such as galaxy cluster counts, weak lensing tomography and baryon acoustic oscillations. These probes require large statistical samples of galaxies and QSOs over large areas to extract their cosmological signals, which has triggered this period of gigantism for the next generation of cosmological surveys. The ongoing DES survey will cover 5,000 square degrees in five bands and gathering information for 300 million sources, but the next generation of surveys will increase the number of galaxies by an order of magnitude. For instance, LSST will begin operations in 2021 and cover 18,000 square degrees of the sky, gathering information on 4 billion sources down to $r_{AB} = 27.5$, after 10 years of operation. The Euclid mission will be launched in 2021 and will remain in operation for 6 years; it will use its visible-light camera (VIS) to measure the galaxy shapes of 1.5 billion of galaxies over 15,000 square degrees.

Computation of photo-z and storage of the PDF will be as challenging as maintaining the required photometry quality across such a large area. For instance, one requirement of the LSST survey is to obtain band-to-band calibration errors no larger than 0.005 mag and no more than 0.01 mag variation across the sky. Success in maintaining such relative calibration is necessary to ensure homogeneous performance.

As described above, the multi-wavelength coverage is also crucial to define the redshift range of interest. In the new generation of imaging surveys, unless SPHEREx[126] with its 96 photometric points between 0.7 and 5 μm is approved, Euclid will be the only survey to map the Universe in NIR using three filters between 0.9 and 2 μm. The goal of these filters is to ensure precise photometric redshifts at $z > 1.3$, which is impossible without NIR (see Fig. 4a). However, Euclid will need to be complemented with ground-based data at optical wavelengths.

Although most of the cosmological imaging surveys are performed with broad bands (HSC, DES, LSST, Euclid), two surveys use medium bands: the PAU and the J-PAS surveys. For instance, J-PAS[127] will cover 8,500 square degrees with 54 narrowband filters. J-PAS is expected to observe 300 million galaxies with a photo-z precision of 0.3%. With this precision, one of the main objectives of this survey will be to measure baryon acoustic oscillations.

Finally, the variable Universe will become more easily accessible with LSST, owing to the 1,000 repeated observations of each location of the sky over 10 years, starting in 2022, in up to six bands. In particular, it will include the U band, making it capable of breaking the degeneracy between low- and high-redshift solutions. More generally, LSST will be extremely powerful at discriminating between stars, AGN and exotic sources.

Evolution in photo-z technique and synergy with spectroscopic surveys. The evolution of the photo-z methods will depend on the spectroscopic data that will become available. For the next generation of cosmological surveys, the imaging and spectroscopic surveys are often conceived in parallel. For instance, the HSC imaging survey will be complemented with spectroscopy using the multi-object Prime Focus Spectrograph[128] instrument covering a wavelength range 0.38–1.3 μm and gathering redshift for millions of galaxies at $0.8 < z < 6$. Similarly, the Euclid survey is complemented by the spectroscopy performed by the NISP instrument, able to perform NIR (1.1–2 μm) slit-less spectroscopy to detect the Hα emission for more than 50 million galaxies at $0.7 < z < 2$. With such a bright future for spectroscopic surveys, ML techniques will become competitive for looking at the high-redshift Universe. It must also be stressed that because of the massive photometric surveys designed as cosmological probes, the algorithms and techniques for computing reliable photo-z are in continuous development. The most recent methods are hybrids that combine the best of the ML and SED-fitting techniques (see, for example, refs [49,129–133]).

Specific needs will also trigger new techniques. Scientific applications based on weak lensing tomography require the true redshift distribution to be characterized with extreme precision in order to map the result of a weak lensing shear analysis onto cosmological parameters. In a very simplified setting, we show in Fig. 5 how a biased photo-z distribution expected for one tomographic redshift bin for a Euclid-like survey will affect the predicted autocorrelation function, and therefore will lead to incorrect best-fitting cosmological parameters. The actual effect of such an offset depends strongly on the redshift of the galaxy sample, and the full shape of the redshift distribution function. In a weak lensing setting, we direct the reader to ref. [134] and figure 1.3 in ref. [2], showing how a bias on the mean redshift translates into an error on the time evolution component of the dark energy equation of state in the LSST case. With current methods, such low bias can not be achieved using photo-z alone. The development of new techniques, able to combine photo-z with the source positions (see 'Introducing spatial information' above), has the potential to revolutionize the field.

References

1. Le Fevre, O. et al. The VIMOS VLT deep survey. First epoch VVDS-deep survey: 11 564 spectra with 17.5 IAB 24, and the redshift distribution over $0 \leq z \leq 5$. *Astron. Astrophys.* **439**, 845–862 (2005).
2. Newman, J. A. et al. Spectroscopic needs for imaging dark energy experiments. *Astropart. Phys.* **63**, 81–100 (2015).
3. Baum, W. A. Photoelectric determinations of redshifts beyond 0.2 c. *Astron. J* **62**, 6–7 (1957).
4. Puschell, J. J., Owen, F. N. & Laing, R. A. Near-infrared photometry of distant radio galaxies: spectral flux distributions and redshift estimates. *Astrophys. J. Lett.* **257**, 57–61 (1982).
5. Ilbert, O. et al. Cosmos photometric redshifts with 30-bands for 2-deg. *Astrophys. J. Suppl. Series* **189**, 1236–1249 (2009).
6. Fontana, A. et al. photometric redshifts and selection of high-redshift galaxies in the NTT and Hubble deep fields. *Astron. J* **120**, 2206–2219 (2000).
7. Dunlop, J. S. et al. A critical analysis of the ultraviolet continuum slopes (beta) of high-redshift galaxies: no evidence (yet) for extreme stellar populations at $z > 6$. *Mon. Not. R. Astron. Soc.* **420**, 901–912 (2012).
8. Coupon, J. et al. The galaxy-halo connection from a joint lensing, clustering and abundance analysis in the CFHTLenS/VIPERS field. *Mon. Not. R. Astron. Soc.* **449**, 1352–1379 (2015).
9. Finoguenov, A. et al. The XMM-Newton Wide-Field Survey in the COSMOS field: statistical properties of clusters of galaxies. *Astrophys. J. Suppl. Series* **172**, 182–195 (2007).
10. López-Sanjuan, C. et al. The dominant role of mergers in the size evolution of massive early-type galaxies since $z \sim 1$. *Astron. Astrophys.* **548**, A7 (2012).
11. Man, A. W. S., Zirm, A. W. & Toft, S. Resolving the discrepancy of galaxy merger fraction measurements at $z \sim 0$–3. *Astrophys. J.* **830**, 11–25 (2016).
12. Etherington, J. et al. Environmental dependence of the galaxy stellar mass function in the Dark Energy Survey science verification data. *Mon. Not. R. Astron. Soc.* **466**, 228–247 (2017).
13. Etherington, J. & Thomas, D. Measuring galaxy environments in large-scale photometric surveys. *Mon. Not. R. Astron. Soc.* **451**, 660–679 (2015).
14. Malavasi, N., Pozzetti, L., Cucciati, O., Bardelli, S. & Cimatti, A. Reconstructing the galaxy density field with photometric redshifts. I. Methodology and validation on stellar mass functions. *Astron. Astrophys.* **585**, 116 (2016).
15. Miyaji, T. et al. Detailed shape and evolutionary behavior of the X-ray luminosity function of active galactic nuclei. *Astrophys. J.* **804**, 104–129 (2015).
16. Padovani, P., Giommi, P. & Rau, A. The discovery of high-power high synchrotron peak blazar. *Mon. Not. R. Astron. Soc.* **422**, 48–52 (2012).
17. Hu, W. Power spectrum tomography with weak lensing. *Astrophys. J. Lett* **522**, 21–24 (1999).
18. Abbott, T. et al. The Dark Energy Survey: more than dark energy—an overview. *Mon. Not. R. Astron. Soc.* **460**, 1270–1299 (2016).
19. Aihara, H. et al. The Hyper Suprime-Cam SSP Survey: overview and survey design. *Publ. Astron. Soc. Japan* **70**, (2018).
20. Laureijs, R. et al. Euclid Definition Study Report. Preprint at https://arxiv.org/abs/1110.3193 (2011).
21. Ivezic, Z. et al. LSST: from science drivers to reference design and anticipated data products. Preprint at https://arxiv.org/abs/0805.2366 (2008).

22. Benitez, N. et al. Optimal filter systems for photometric redshift estimation. *Astrophys. J. Lett.* **692**, 5–8 (2009).

23. Fioc, M. & Rocca-Volmernage, B. PEGASE: a UV to NIR spectral evolution model of galaxies. Application to the calibration of bright galaxy counts. *Astron. Astrophys.* **326**, 950–962 (1997).

24. Bruzual, G. & Charlot, S. Stellar population synthesis at the resolution of 2003. *Mon. Not. R. Astron. Soc.* **344**, 1000–1028 (2003).

25. Maraston, C. Evolutionary population synthesis: models, analysis of the ingredients and application to high-z galaxies. *Mon. Not. R. Astron. Soc.* **362**, 799–825 (2005).

26. Conroy, C. On the birth masses of the ancient globular clusters. *Astrophys. J.* **758**, 21–34 (2012).

27. Coleman, G. D., Wu, C.-C. & Weedman, D. W. Colors and magnitudes predicted for high redshift galaxies. *Astrophys. J. Suppl. Series* **43**, 393–416 (1980).

28. Kinney, A. L. et al. Template ultraviolet to near-infrared spectra of star-forming galaxies and their application to k-corrections. *Astrophys. J.* **467**, 38 (1996).

29. Polletta, M. et al. Spectral energy distributions of hard X-ray selected active galactic nuclei in the XMM-Newton Medium Deep Survey. *Astrophys. J.* **663**, 81–102 (2007).

30. Noll, S. et al. The FORS Deep Field spectroscopic survey. *Astron. Astrophys.* **418**, 885–906 (2004).

31. Chevallard, J. & Charlot, S. Erratum: Modelling and interpreting spectral energy distributions of galaxies with BEAGLE. *Mon. Not. R. Astron. Soc.* **464**, 2349 (2017).

32. Ilbert, O. et al. Accurate photometric redshifts for the CFHT legacy survey calibrated using the VIMOS VLT deep survey. *Astron. Astrophys.* **457**, 841–856 (2006).

33. Schaerer, D. & de Barros, S. in *The Spectral Energy Distribution of Galaxies* (eds Tuffs, R. J. & Popescu, C. C.) IAU Symp. 284, 20 (IAU, 2012).

34. Pacifici, C., Charlot, S., Blaizot, J. & Brinchmann, J. Relative merits of different types of rest-frame optical observations to constrain galaxy physical parameters. *Mon. Not. R. Astron. Soc.* **421**, 2002–2024 (2012).

35. Calzetti, D. et al. The dust content and opacity of actively star-forming galaxies. *Astrophys. J.* **533**, 682–695 (2000).

36. Prevot, M. L., Lequeux, J., Prevot, L., Maurice, E. & Rocca-Volmerange, E. The typical interstellar extinction in the Small Magellanic Cloud. *Astron. Astrophys.* **132**, 389–392 (1984).

37. Madau, P. Radiative transfer in a clumpy universe: the colors of high-redshift galaxies. *Astrophys. J.* **441**, 18–27 (1995).

38. Draine, B. T. *Physics of the Interstellar and Intergalactic Medium*. (Princeton Univ. Press, Princeton, 2011).

39. Schlegel, D. J., Finkbeiner, D. P. & Davis, M. Maps of dust infrared emission for use in estimation of reddening and cosmic microwave background radiation foregrounds. *Astrophys. J.* **500**, 525–553 (1998).

40. Galametz, A., Saglia, R., Paltani, S., Apostolakos, N. & Dubath, P. SED-dependent galactic extinction prescription for Euclid and future cosmological surveys. *Astron. Astrophys.* **598**, 20 (2017).

41. Hildebrandt, H. et al. PHAT: PHoto-*z* accuracy testing. *Astron. Astrophys.* **523**, 31 (2010).

42. Cavuoti, S., Brescia, M., Longo, G. & Mercurio, A. Photometric redshifts with the quasi Newton algorithm (MLPQNA) results in the PHAT1 contest. *Astron. Astrophys.* **546**, 13 (2012).

43. Beck, R. et al. On the realistic validation of photometric redshift. *Mon. Not. R. Astron. Soc.* **468**, 4323–4339 (2017).

44. Breiman, L. Random forests. *Mach. Learn.* **45**, 5–32 (2001).

45. Van Der Malsburg C. In *Brain Theory* (eds Palm G. & Aertsen A.) 245–248 (Springer, Berlin, 1986); https://doi.org/10.1007/978-3-642-70911-1_20

46. Carliles, S., Budavári, T., Heinis, S., Priebe, C. & Szalay, A. S. Random forests for photometric redshifts. *Astrophys. J.* **712**, 511–515 (2010).

47. Carrasco Kind, M. & Brunner, R. J. TPZ: photometric redshift PDFs and ancillary information by using prediction trees and random forests. *Mon. Not. R. Astron. Soc.* **432**, 1483–1501 (2013).

48. Collister, A. A. & Lahav, O. ANNz: estimating photometric redshifts using artificial neural networks. *Publ. Astron. Soc. Pacif.* **116**, 345–351 (2004).

49. Vanzella, E. et al. Photometric redshifts with the Multilayer Perceptron Neural Network: application to the HDF-S and SDSS. *Astron. Astrophys.* **423**, 761–776 (2004).

50. Brescia, M., Cavuoti, S. & Longo, G. Automated physical classification in the SDSS DR10. A catalogue of candidate quasars. *Mon. Not. R. Astron. Soc.* **450**, 3893–3903 (2015).

51. Bonnett, C. Using neural networks to estimate redshift distributions. An application to CFHTLenS. *Mon. Not. R. Astron. Soc.* **449**, 1043–1056 (2015).

52. Hoyle, B. Measuring photometric redshifts using galaxy images and Deep Neural Networks. *Astron. Comput.* **16**, 34–40 (2016).

53. D'Isanto, A. & Polsterer, K. L. Photometric redshift estimation via deep learning. *Astron. Astrophys.* **609**, 111 (2018).

54. Carrasco Kind, M. & Brunner, R. J. SOMz: photometric redshift PDFs with self-organizing maps and random atlas. *Mon. Not. R. Astron. Soc.* **438**, 3409–3421 (2014).

55. Masters, D. et al. Mapping the galaxy color–redshift relation: optimal photometric redshift calibration strategies for cosmology surveys. *Astrophys. J.* **813**, 53–68 (2015).

56. Cavuoti, S. et al. METAPHOR: A machine-learning-based method for the probability density estimation of photometric redshifts. *Mon. Not. R. Astron. Soc.* **465**, 1959–1973 (2017).

57. Hoyle, B. et al. Anomaly detection for machine learning redshifts applied to SDSS galaxies. *Mon. Not. R. Astron. Soc.* **452**, 4183–4194 (2015).

58. Hoyle, B., Rau, M. M., Seitz, S. & Weller, J. Data augmentation for machine learning redshifts applied to Sloan Digital Sky Survey galaxies. *Mon. Not. R. Astron. Soc.* **450**, 305–316 (2015).

59. Hoyle, B., Rau, M. M., Zitlau, R., Seitz, S. & Weller, J. Feature importance for machine learning redshifts applied to SDSS galaxies. *Mon. Not. R. Astron. Soc.* **449**, 1275–1283 (2015).

60. Lima, M. et al. Estimating the redshift distribution of photometric galaxy sample. *Mon. Not. R. Astron. Soc.* **390**, 118–130 (2008).

61. Bentez, N. Bayesian photometric redshift estimation. *Astrophys. J.* **536**, 571–583 (2000).

62. Dahlen, T. et al. A detailed study of photometric redshifts for GOODS-South galaxies. *Astrophys. J.* **724**, 425–447 (2010).

63. Tanaka, M. Photometric redshift with Bayesian priors on physical properties of galaxies. *Astrophys. J.* **801**, 20–39 (2015).

64. Seldner, M. & Peebles, P. J. E. Statistical analysis of catalogs of extragalactic objects. XI - Evidence of correlation of QSOs and Lick galaxy counts. *Astrophys. J.* **227**, 30–36 (1979).

65. Newman, J. A. Calibrating redshift distributions beyond spectroscopic limits with cross-correlations. *Astrophys. J.* **684**, 88–101 (2008).

66. Ménard, B. et al. Clustering-based redshift estimation: method and application to data. Preprint at https://arxiv.org/abs/1303.4722 (2013).

67. Scottez, V. et al. Clustering-based redshift estimation: application to VIPERS/CFHTLS. *Mon. Not. R. Astron. Soc.* **462**, 1683–1696 (2016).

68. Rahman, M., Ménard, B., Scranton, R., Schmidt, C. B. & Morrison, C. B. Clustering-based redshift estimation: comparison to spectroscopic redshifts. *Mon. Not. R. Astron. Soc.* **447**, 3500–3511 (2015).

69. Aragon-Calvo, M. A., van de Weygaert, R., Jones, B. J. T. & Mobasher, B. Submegaparsec individual photometric redshift estimation from cosmic web constraints. *Mon. Not. R. Astron. Soc.* **454**, 463–477 (2015).

70. Sánchez, C. et al. Photometric redshift analysis in the Dark Energy Survey Science Verification data. *Mon. Not. R. Astron. Soc.* **445**, 1482–1506 (2014).

71. Dahlen, T. et al. A critical assessment of photometric redshift methods: a CANDELS investigation. *Astrophys. J.* **775**, 93 (2013).

72. Duncan, K. J. et al. Photometric redshifts for the next generation of deep radio continuum surveys—I: Template fitting. *Mon. Not. R. Astron. Soc.* **473**, 2655–2672 (2018).

73. Bolzonella, M., Miralles, J.-M. & Pelló, R. Photometric redshifts based on standard SED fitting procedures. *Astron. Astrophys.* **363**, 476–492 (2000).

74. Brammer, G. B., van Dokkum, P. G. & Coppi, P. EAZY: a fast, public photometric redshift code. *Astrophys. J.* **686**, 1503–1513 (2008).

75. Feldmann, R. et al. The Zurich Extragalactic Bayesian Redshift Analyzer and its first application: COSMOS. *Mon. Not. R. Astron. Soc.* **372**, 565–577 (2006).

76. Luo, B. et al. Identifications and photometric redshifts of the 2 ms Chandra Deep Field-South sources. *Astrophys. J. Suppl. Series* **187**, 560–580 (2010).

77. Arnouts, S. et al. Measuring and modelling the redshift evolution of clustering: the Hubble Deep Field North. *Mon. Not. R. Astron. Soc.* **310**, 540–556 (1999).

78. Hsu, L.-T. et al. CANDELS/GOODS-S, CDFS, and ECDFS: photometric redshifts for normal and X-ray-detected galaxies. *Astrophys. J.* **796**, 60 (2014).

79. Cardamone, C. N. et al. The Multiwavelength Survey by Yale-Chile (MUSYC): deep medium-band optical imaging and high-quality 32-band photometric redshifts in the ECDF.-S. *Astrophys. J. Suppl. Series* **189**, 270–285 (2010).

80. Pérez-González, P. G. et al. SHARDS: an optical spectro-photometric survey of distant galaxies. *Astrophys. J.* **762**, 46 (2013).

81. Molino, A. et al. The ALHAMBRA survey: Bayesian photometric redshifts with 23 bands for 3 deg. *Mon. Not. R. Astron. Soc.* **441**, 2891–2922 (2014).

82. Sadeh, I., Abdalla, F. B. & Lahav, O. ANNz2: photometric redshift and probability distribution function estimation using machine learning. *Publ. Astron. Soc. Pacif.* **128**, 104502 (2016).

83. Brescia, M. et al. The astronomical data deluge and the template case of photometric redshifts. Preprint at https://arxiv.org/abs/1802.07683 (2018).

84. Bordoloi, R., Lilly, S. J. & Amara, A. Photo-z performance for precision cosmology. *Mon. Not. R. Astron. Soc.* **406**, 881–895 (2010).

85. Bonnett, C. et al. Redshift distributions of galaxies in the Dark Energy Survey Science Verification shear catalogue and implications for weak lensing. *Phys. Rev. D* **94**, 1168 (2016).

86. Quadri, R. F. & Williams, R. J. Quantifying photometric redshift errors in the absence of spectroscopic redshifts. *Astrophys. J.* **725**, 794–802 (2010).

87. Benjamin, J., van Waerbeke, L., Ménard, B. & Kilbinger, M. Photometric redshifts: estimating their contamination and distribution using clustering information. *Mon. Not. R. Astron. Soc.* **408**, 1168–1180 (2010).

88. Laigle, C. et al. The COSMOS2015 Catalog: Exploring the 1 < z < 6 Universe with half a million galaxies. *Astrophys. J. Suppl. Series* **224**, 24 (2016).

89. Coupon, J. et al. Photometric redshifts for the CFHTLS T0004 deep and wide fields. *Astron. Astrophys.* **500**, 981–998 (2009).

90. Wolf, C. et al. A catalogue of the Chandra Deep Field South with multi-colour classification and photometric redshifts from COMBO-17. *Astron. Astrophys.* **421**, 913–936 (2004).

91. Cimatti, A. et al. The K20 survey. I. Disentangling old and dusty star-forming galaxies in the ERO population. *Astron. Astrophys.* **381**, 68–72 (2002).

92. Bouwens, R. J. et al. UV luminosity functions at redshifts z ~ 4 to z ~ 10: 10,000 galaxies from HST legacy fields. *Astrophys. J.* **803**, 34 (2015).

93. Whitaker, K. E. et al. The NEWFIRM medium-band survey: photometric catalogs, redshifts, and the bimodal color distribution of galaxies out to z ~3. *Astrophys. J.* **735**, 86 (2011).

94. Drlica-Wagner, A. et al. Dark Energy Survey Year 1 results: photometric data set for cosmology. Preprint at https://arxiv.org/abs/1708.01531 (2017).

95. Bertin, E. & Arnouts, S. SExtractor: software for source extraction. *Astron. Astrophys. Suppl. Series* **117**, 393–404 (1996).

96. Hildebrandt, H. et al. CFHTLenS: improving the quality of photometric redshifts with precision photometry. *Mon. Not. R. Astron. Soc.* **421**, 2355–2367 (2012).

97. Moutard, T. et al. The VIPERS Multi-Lambda Survey. I. UV and near-IR observations, multi-colour catalogues, and photometric redshifts. *Astron. Astrophys.* **590**, 102 (2016).

98. Grazian, A. et al. A comparison of LBGs, DRGs, and BzK galaxies: their contribution to the stellar mass density in the GOODS-MUSIC sample. *Astron. Astrophys.* **465**, 393–404 (2007).

99. Mancone, C. L., Gonzales, A. H., Moustakas, L. A. & Price, A. PyGFit: a tool for extracting PSF matched photometry. *Publ. Astron. Soc. Pacif.* **125**, 1514–1524 (2013).

100. Bundy, K. et al. SYNMAG photometry: a fast tool for catalog-level matched colors of extended sources. *Astron. J.* **144**, 188 (2012).

101. Laidler, V. G. et al. TFIT: a photometry package using prior information for mixed-resolution data sets. *Publ. Astron. Soc. Pacif.* **119**, 1325–1344 (2007).

102. Merlin, E. et al. T-PHOT: a new code for PSF-matched, prior-based, multiwavelength extragalactic deconfusion photometry. *Astron. Astrophys.* **582**, 15 (2015).

103. Magorrian, J. et al. The demography of massive dark objects in galaxy centers. *Astron. J* **115**, 2285–2305 (1998).

104. Bongiorno, A. et al. Accreting supermassive black holes in the COSMOS field and the connection to their host galaxies. *Mon. Not. R. Astron. Soc.* **427**, 3103–3133 (2012).

105. Salvato, M. et al. Dissecting photometric redshift for active galactic nucleus using XMM- and Chandra-COSMOS samples. *Astrophys. J.* **742**, 61 (2011).

106. Marchesi, S. et al. The Chandra COSMOS Legacy survey: optical/IR identifications. *Astrophys. J.* **817**, 34 (2016).

107. Zheng, W. et al. Photometric redshift of X-ray sources in the Chandra Deep Field South. *Astrophys. J. Suppl. Series* **155**, 73–87 (2004).

108. Brusa, M. et al. The XMM-Newton Wide-field Survey in the Cosmos Field (XMM-COSMOS): demography and multiwavelength properties of obscured and unobscured luminous active galactic nuclei. *Astrophys. J.* **716**, 348–369 (2010).

109. Merloni, A. et al. eROSITA Science Book: Mapping the Structure of the Energetic Universe. Preprint at https://arxiv.org/abs/1209.3114 (2012).

110. Norris, R. et al. EMU: Evolutionary Map of the Universe. *Publ. Astron. Soc. Pacif.* **28**, 215–248 (2011).

111. Salvato, M. et al. Photometric redshift and classification for the XMM-COSMOS sources. *Astrophys. J.* **690**, 1250–1263 (2009).

112. Kitsionas, S., Hatziminaoglou, E., Georgakis, A. & Georgantopoulos, I. On the use of photometric redshifts for X-ray selected AGNs. *Astron. Astrophys.* **434**, 475–482 (2005).

113. Bovy, J. et al. Photometric redshifts and quasar probabilities from a single, data-driven generative model. *Astrophys. J.* **749**, 41 (2012).

114. Brescia, M., Cavuoti, S., D'Abrusco, R., Long, G. & Mercurio, A. Photometric redshifts for quasars in multi-band surveys. *Astrophys. J.* **772**, 140 (2013).

115. Budavari, T. et al. Photometric redshifts from reconstructed quasar templates. *Astron. J* **122**, 1163–1161 (2001).

116. Mountrichas, E. A. et al. Estimating photometric redshifts for X-ray sources in the X-ATLAS field, using machine-learning techniques. *Astron. Astrophys.* **608**, 39 (2017).

117. Simm, T. et al. Pan-STARRS1 variability of XMM-COSMOS AGN. I. Impact on photometric redshifts. *Astron. Astrophys.* **584**, 106 (2015).

118. Rau, A. et al. BL Lacertae objects beyond redshift 1.3—UV-to-NIR photometry and photometric redshift for Fermi/LAT blazars. *Astron. Astrophys.* **538**, A26 (2012).

119. Krühler, T. et al. Photometric redshifts for gamma-ray burst afterglows from GROND and Swift/UVOT. *Astron. Astrophys.* **526**, 153 (2011).

120. Palanque-Delabrouille, N. et al. Photometric redshifts for type Ia supernovae in the supernova legacy survey. *Astron. Astrophys.* **514**, 63 (2010).

121. Lanzetta, K. M., Yahil, A. & Fernández-Soto, A. Star-forming galaxies at very high redshifts. *Nature* **381**, 759–763 (1996).

122. Finkelstein, S. et al. The case for a James Webb Space Telescope extragalactic key project. Preprint at https://arxiv.org/abs/1512.04530 (2015).

123. Bisigello, L. et al. The impact of JWST broadband filter choice on photometric redshift estimation. *Astrophys. J. Suppl. Series* **227**, 19 (2016).

124. Schaerer, D. & de Barros, S. On the physical properties of z ~ 6–8 galaxies. *Astron. Astrophys.* **515**, 73 (2010).

125. Labbe, I. et al. The spectral energy distributions of z ~ 8 galaxies from the IRAC Ultra Deep Fields: emission lines, stellar masses, and specific star formation rates at 650 Myr. *Astrophys. J. Lett.* **777**, 19–25 (2013).

126. Doré, O. et al. Science impacts of the SPHEREx all-sky optical to near-infrared spectral survey: report of a community workshop examining extragalactic, galactic, stellar and planetary science. Preprint at https://arxiv.org/abs/1606.07039 (2016).

127. Benitez, N. J-PAS: The Javalambre-Physics of the Accelerated Universe Astrophysical Survey. Preprint at https://arxiv.org/abs/1403.5237 (2014).

128. Tamura, N. et al. Prime Focus Spectrograph (PFS): a very wide-field, massively multi-object, optical and near-infrared fiber-fed spectrograph on the Subaru Telescope. *Publ. Astron. Soc. Pacif.* **507**, 387 (2016).

129. Carrasco-Kind, M. & Brunner, R. J. Exhausting the information: novel Bayesian combination of photometric redshift PDFs. *Mon. Not. R. Astron. Soc.* **442**, 3380–3399 (2014).

130. Leistedt, B. & Hogg, D. W. Data-driven, interpretable photometric redshifts trained on heterogeneous and unrepresentative data. *Astrophys. J.* **838**, 5 (2017).

131. Speagle, J. & Eisenstein, D. J. Deriving photometric redshifts using fuzzy archetypes and self-organizing maps—I. Methodology. *Mon. Not. R. Astron. Soc.* **469**, 1186–1204 (2017).

132. Duncan, K. J., Jarvis, M. J., Brown, M. J. I. & Rottgering, H. J. A. Photometric redshifts for the next generation of deep radio continuum surveys II. Gaussian processes and hybrid estimates. Preprint at https://arxiv.org/abs/1712.04476 (2018).

133. Beck, R., Dobos, L., Budavári, T., Szalay, A. S. & Csabai, I. Photometric redshifts for the SDSS Data Release 12. *Mon. Not. R. Astron. Soc.* **460**, 1371–1381 (2016).

134. Amendola, L. et al. Cosmology and fundamental physics with the Euclid satellite. *Living Rev. Relativ.* **16**, 6 (2013).

135. Bielby, R. et al. The WIRCam Deep Survey. I. Counts, colours, and mass-functions derived from near-infrared imaging in the CFHTLS deep fields. *Astron. Astrophys.* **545**, 23 (2012).

136. De Jong, J. T. A. et al. The third data release of the Kilo-Degree Survey and associated data products. *Astron. Astrophys.* **604**, 134 (2017).

Acknowledgements

Author contributions

The authors contributed equally to the writing of this Review Article.

Competing interests

The authors declare no competing interests.

Additional information