Photometric redshifts for the next generation of deep radio continuum surveys - I: Template fitting

Kenneth J Duncan[1]*, Michael J. I. Brown[2,3], Wendy L. Williams[4], Philip N. Best[5], Veronique Buat[6], Denis Burgarella[6], Matt J. Jarvis[7,8], Katarzyna Małek[6,9], S. J. Oliver[10], Huub J. A. Röttgering[1], Daniel J. B. Smith[4]

[1]*Leiden Observatory, Leiden University, NL-2300 RA Leiden, Netherlands*
[2]*School of Physics, Monash University, Clayton, Victoria 3800, Australia*
[3]*Monash Centre for Astrophysics, Monash University, Clayton, Victoria, 3800, Australia*
[4]*Centre for Astrophysics Research, School of Physics, Astronomy and Mathematics, University of Hertfordshire,*
College Lane, Hatfield AL10 9AB, UK
[5]*SUPA, Institute for Astronomy, Royal Observatory, Blackford Hill, Edinburgh, EH9 3HJ, UK*
[6]*Aix-Marseille Université, CNRS - LAM (Laboratoire d'Astrophysique de Marseille) UMR 7326, 13388 Marseille, France*
[7]*Astrophysics, University of Oxford, Denys Wilkinson Building, Keble Road, Oxford, OX1 3RH*
[8]*Physics and Astronomy Department, University of the Western Cape, Bellville 7535, South Africa*
[9] *National Centre for Nuclear Science, ul. Hoza 69, 00-681 Warsaw, Poland*
[10]*Astronomy Centre, Department of Physics and Astronomy, University of Sussex, Falmer, Brighton BN1 9QH, UK*

28 September 2017

ABSTRACT

We present a study of photometric redshift performance for galaxies and active galactic nuclei detected in deep radio continuum surveys. Using two multi-wavelength datasets, over the NOAO Deep Wide Field Survey Boötes and COSMOS fields, we assess photometric redshift (photo-z) performance for a sample of $\sim 4,500$ radio continuum sources with spectroscopic redshifts relative to those of $\sim 63,000$ non radio-detected sources in the same fields. We investigate the performance of three photometric redshift template sets as a function of redshift, radio luminosity and infrared/X-ray properties. We find that no single template library is able to provide the best performance across all subsets of the radio detected population, with variation in the optimum template set both between subsets and between fields. Through a hierarchical Bayesian combination of the photo-z estimates from all three template sets, we are able to produce a consensus photo-z estimate which equals or improves upon the performance of any individual template set.

Key words:

1 INTRODUCTION

Photometric redshifts are a vital tool for estimating the distances to large samples of galaxies observed in extragalactic surveys. At almost all survey scales, from large area surveys such as the Sloan Digital Sky Survey (SDSS; York et al. 2000) or the Dark Energy Survey (DES; Dark Energy Survey Collaboration 2005) to deep pencil-beam Hubble Space Telescope (HST) surveys such as CANDELS (Grogin et al. 2011), it is impractical to obtain spectroscopic redshifts for more than a small fraction of photometrically detected sources. For the vast majority of sources that are currently detected or will be detected in future photometric surveys, we are therefore reliant on photometric redshift techniques to estimate their distance or extract information about the intrinsic physical properties (Laureijs et al. 2011).

While this statement is applicable to photometric surveys across all of the electromagnetic spectrum, the latest generation of deep radio continuum surveys by Square Kilometre Array (SKA) precursors and pathfinders such as the Low Frequency Array (LOFAR; van Haarlem et al. 2013), the Australian SKA Pathfinder (ASKAP; Johnston 2007) and MeerKAT (Booth et al. 2009) pose a new challenge. Probing to unprecedented depths, these surveys will increase the detected population of radio sources by more than an order of magnitude and probe deep into the earliest epochs of galaxy formation and evolution (Rottgering 2010; Jarvis 2012; Norris et al. 2013).

* E-mail: duncan@strw.leidenuniv.nl

The population of radio detected sources is itself extremely diverse - with radio emission tracing both black hole accretion in active galactic nuclei (AGN) and star formation activity. With the majority of these sources lacking useful radio morphology information (being unresolved in radio continuum observations), classifying and separating the various sub-populations of radio sources will rely on photometric methods (e.g. Chung et al. 2014; Calistro Rivera et al. 2017). Accurate and unbiased photometric redshift estimates for the radio source population will therefore be essential for studying the faint radio population and achieving the scientific goals of these deep radio continuum surveys.

Since the publication of the first widely used photometric redshift (photo-z) estimation tools (e.g. Arnouts et al. 1999; Benítez 2000; Bolzonella et al. 2000), both the accuracy of photo-z estimates and our understanding of their biases and limitations has significantly improved. The development and testing of photometric redshift techniques has been driven not just by studies of galaxy evolution at high redshifts (Dahlen et al. 2013), but also by the next generation of tomographic weak lensing cosmology surveys (Carrasco Kind & Brunner 2014b; Sanchez et al. 2014); specifically, the need for computationally fast, accurate and unbiased photometric redshifts for unprecedented samples of galaxies.

Detailed studies have shown that while it is possible to produce accurate photo-zs for X-ray selected AGN (Salvato et al. 2008, 2011; Hsu et al. 2014), care must be taken to correct for the effects of optical variability on photometric data which have been observed over long time periods. Similarly, various studies have been increasingly successful in estimating accurate photometric redshifts for large photometric quasar samples such as the SDSS (York et al. 2000), e.g. Richards et al. (2001), Weinstein et al. (2004), Ball et al. (2008), Bovy et al. (2012), Zhang et al. (2013) and Brescia et al. (2013). However, fundamental to all of these efforts is the large representative spectroscopic sample upon which the empirical redshift estimation algorithms are trained.

Several studies have illustrated that the AGN populations selected at different wavelengths (X-ray, optical, IR, radio) are often distinct, with only some overlap between different selection methods (Hickox et al. 2009; Kochanek et al. 2012; Chung et al. 2014). The optimal photometric redshift techniques and systematics identified for one particular AGN population are therefore not necessarily applicable to an AGN sample selected by other means.

In this paper we aim to quantify some of these systematic effects and find the optimum strategy for estimating accurate photometric redshifts for radio selected populations. Specifically, we want to understand how the photometric redshift accuracy of radio sources varies as a function of radio luminosity and redshift. Do the current methods and optimization strategies developed for 'normal' galaxies or other AGN populations in optical surveys extend to radio selected galaxies? Finally, based on the results of these tests, we wish to construct an optimised method which can then be applied successfully to other survey fields in preparation for the next generations of radio continuum surveys (e.g. LOFAR/MIGHTEE: Rottgering 2010; Jarvis 2012) and the millions of radio sources they will detect (Shimwell et al. 2017).

The paper is structured as follows: Section 2 outlines the multi-wavelength datasets used in this analysis, including details of optical data used for the photometric redshift estimates and the corresponding radio continuum and spectroscopic redshift datasets. Section 3 then describes how the individual photometric redshift estimates used in this comparison were determined and the choice of software, templates and settings used. Section 4 presents the detailed comparison and analysis of these photo-z methods in the context of deep radio continuum surveys. In Section 5 we outline the improved photometric redshift method devised for the LOFAR survey. Section 6 presents a discussion of the results presented in Section 4 and their implications for future galaxy evolution and cosmology studies with the forthcoming generation of radio continuum surveys. Finally, Section 7 presents our summary and conclusions. Throughout this paper, all magnitudes are quoted in the AB system (Oke & Gunn 1983) unless otherwise stated. We also assume a Λ-CDM cosmology with $H_0 = 70$ kms^{-1}Mpc^{-1}, $\Omega_m = 0.3$ and $\Omega_\Lambda = 0.7$.

2 DATA

To maximise the parameter space explored in this analysis, we make use of two complementary datasets. Firstly, we make use of the extensive multi-wavelength data over the large ~ 9 deg^2 NOAO Deep Wide Field Survey in Boötes (NDWFS: Jannuzi & Dey 1999). Secondly, we also include data from the COSMOS field which extends to significantly fainter depths across all wavelengths but over a smaller ~ 2 deg^2 area.

2.1 Wide' field - Boötes Field

2.1.1 Optical photometry

The Boötes photometry used in this study is taken from the PSF matched photometry catalogs of available imaging data in the NDWFS (Brown et al. 2007, 2008). The full catalog covers a wide range of wavelengths, spanning from 0.15 to 24 μm.

The photometry included in the subsequent analysis is based primarily on the deep optical imaging in B_W, R and I-bands from Jannuzi & Dey (1999). At optical wavelengths there is also additional z band coverage from the *zBoötes* survey (Cool 2007). Near-infrared observations of the field are provided by NEWFIRM observations at J, H and K_s (Gonzalez et al. 2010).

Filling in two critical wavelength ranges not previously covered by the existing NOAO Boötes data is additional imaging in the $U_{\rm spec}$ ($\lambda_0 = 3590$) and y ($\lambda_0 = 9840$) bands from the Large Binocular Telescope (Bian et al. 2013), covering the full NDWFS observational footprint. Finally, IRAC observations (Fazio et al. 2004) at 3.6, 4.5, 5.8 and 8 μm are provided by the *Spitzer* Deep Wide Field Survey (SDWFS, Ashby et al. 2009).

Although the available GALEX NUV data cover a significant fraction of the NDWFS field and reach depths comparable to the NOAO B_W data, the large point-spread function (PSF) with full-width half maximum (FWHM) equal to $\sim 4.9''$ could result in significantly increased source confusion relative to the other bands used in the catalog. As such,

the NUV data were not included for the purposes of photometric redshift estimation.[1]

Finally, we also include the *u, g, r, i* and *z* imaging from SDSS (Alam et al. 2015). Although the limiting magnitudes reached by the SDSS photometry are not as faint as the NDWFS optical dataset at comparable wavelengths, the different central wavelengths of the SDSS filters provide valuable additional colour information for bright sources and are therefore worth including.

The matched aperture photometry in the catalogs produced by Brown et al. are based on detections in the NOAO *I* band image as measured by SExtractor (Bertin & Arnouts 1996). Forced aperture photometry was then performed on each of the available UV to infrared images for a range of aperture sizes. The optical/near-infrared images were all first gridded to a common pixel scale and smoothed to a matched PSF. The common PSF chosen was that of a Moffat profile with $\beta = 2.5$ and a FWHM equal to $1.35''$ for the B_W, R, I, Y, H and Ks filters and a larger $1.6''$ for u, z and J.

For the matched catalog for photometric redshift estimation, we use fluxes in $3''$ apertures for all optical/near-IR bands and $4''$ for the IRAC bands. These aperture fluxes were then corrected to total fluxes using the aperture corrections based on the 1.35 or $1.6''$ Moffat profiles or the corresponding IRAC PSF curves of growth. Tests performed using 2, 3 and $4''$ apertures for the optical bands indicate that for Boötes the 3 and $4''$ aperture-based photometry perform almost identically for photometric redshift estimation while the $2''$-based photometry performed significantly worse. The choice of $3''$ over $4''$ apertures is based solely on consistency with the 'Deep' data presented in the following sub-section.

2.1.2 Spectroscopic redshifts

Spectroscopic redshifts for sources in Boötes are taken from a compilation of observations within the field (Brown, *priv. communication*). The majority of redshifts within the sample come from the AGN and Galaxy Evolution Survey (AGES; Kochanek et al. 2012) spectroscopic survey, with additional samples provided by numerous follow-up surveys in the field including Lee et al. (2012, 2013, 2014), Stanford et al. (2012), Zeimann et al. (2012, 2013), Dey et al. (2016) and Hickox, R. C. et al (priv. communication).

The spectroscopic redshift catalog was matched to the combined multi-wavelength catalog based on their quoted physical coordinates in the two respective catalogs and using a maximum separation of $1''$. In total, the combined sample consists of 22830 redshifts over the range $0 < z < 6.12$, with 88% of these at $z < 1$.

It is important to raise a caveat to the analysis in the following sections, namely that while the spectroscopic sample used here represents one of best available in the literature and includes a diverse range of galaxy types, it may still not be fully representative of the radio source population. As with any non spectroscopically complete sample, the subset of sources with available spectroscopic redshifts represents

a somewhat biased sample with respect to both the overall photometric sample and the radio selected galaxy population. In particular, low excitation radio galaxies (LERGS) may be under-represented within the spectroscopic sample due to the lack of strong emission lines available for redshift estimation.

2.1.3 Radio fluxes

Radio observations for the Boötes field are taken from new LOw Frequency ARray (LOFAR; van Haarlem et al. 2013) observations presented in Williams et al. (2016). Full details of the radio data and reduction are presented in Williams et al. (2016), including details of the methods used during calibration and imaging to correct for direction-dependent effects (DDEs) caused by the ionosphere and the LOFAR phased array beam.

In summary, the observations consist of 8 hr of data taken with the LOFAR High Band Antennae (HBA) and covering the frequency range 130-169 MHz, with a central frequency of ≈ 150 MHz. The resulting image covers 19 deg^2, with a rms noise of $\approx 120 - 150 \mu$Jy beam^{-1} and resolution of 5.6×7.4 arcsec (see Williams et al. 2016, for details on the source extraction and catalog properties).

Within the LOFAR field of view, the final source catalog contains a total of 6267 separate 5σ radio sources. Of these sources, 3902 fall within the boundaries of the I-band optical imaging and can therefore be matched to the optical catalogs. Matches between the LOFAR radio observations and the optical catalog were estimated using a multi-step likelihood ratio technique. Full details of the visual classifications, radio positions and likelihood ratio technique are presented in Williams et al. (in prep). However, the key steps are as follows. Firstly, radio sources were visually classified into distinct morphological classes. Next, optical counterparts for each radio source are determined through a likelihood ratio technique based on the positions, positional uncertainty and brightness of the optical and radio sources (with the radio centroid position and uncertainty dependent on the radio morphology classification). For the small subset of large extended sources where the automated likelihood ratio matching technique cannot be applied, matches were determined individually based on source morphology and visual comparison with the optical imaging.

The cross-matching process yields a total of 2971 matches to sources within the full list of sources within the Brown et al. (2007) optical catalog. However, of these 2971 matches, 578 are matches to optical sources which are flagged as potentially being affected by bright stars/extended sources or are on chip edges. Of the ~ 1000 sources which lie within the I-band optical footprint for which no reliable counterpart could be found, a large fraction represent faint un-resolved radio sources for the optical counterpart is too faint. These sources may be optically faint either due to being at high redshift or as a result of having intrinsically red SEDs. The Boötes sample used in this analysis is potentially biased against radio sources with optically faint counterparts. However, thanks to the deep near-IR imaging which forms the basis of the 'Deep' COSMOS field we are still able to explore the photo-z properties for these sources.

---

[1] Initial tests with EAZY also found including the NUV data made no appreciable improvement.

2.2 'Deep' field - COSMOS

2.2.1 Optical photometry

The optical/near-IR data used in the COSMOS field are taken from the C OSMOS2015 catalog presented in Laigle et al. (2016). Laigle et al. (2016) outline fully the details of the optical dataset, including data homogenisation, source detection and extraction. We therefore refer the interested reader to said paper for further detail.

For the analysis in this paper, we use the seven optical broad bands (B, V, g, r, i, z^+/z^{++}), 12 medium bands ($IA427$, $IA464$, $IA484$, $IA505$, $IA527$, $IA574$, $IA624$, $IA679$, $IA709$, $IA738$, $IA767$, and $IA827$) and two narrow bands ($NB711$, $NB816$) taken with the Subaru Suprime-Cam (Capak et al. 2007; Taniguchi et al. 2015). Also included at optical wavelengths are the u^*-band data from the Canada-France-Hawaii Telescope (CFHT/MegaCam) as well as Y-band data taken with the Subaru Hyper-Suprime-Cam. As with the Boötes field, we do not include GALEX UV data in the fitting. At longer wavelengths we include the UltraVISTA $YJHK_s$-band data (McCracken et al. 2012) and 3.6, 4.5, 5.8 and 8μm *Spitzer*-IRAC bands. We make use of the aperture corrected 3″ flux estimates for all optical to near-IR bands in combination with the deconfused IRAC photometry as outlined in Laigle et al. (2016).

2.2.2 Spectroscopic redshifts

Spectroscopic redshifts for the COSMOS field were taken from the sample of redshifts compiled by and for the Herschel Extragalactic Legacy Project (HELP; Vaccari 2016, PI: S. Oliver).[2] The compilation includes the large number of publicly available redshifts in the field (see Laigle et al. 2016, and references therein) and a small number of currently unpublished samples. In total, the sample comprises 44,875 sources extending to $z > 6$ and with $\sim 12,000$ sources at $z > 1$.

Thanks to the optical depths probed by both the photometric and spectroscopic data available in the COSMOS field, the 'Deep' spectroscopic samples a range of galaxy types magnitudes, redshifts which may be missing from the 'Wide' sample. Despite this, the subset of radio detected galaxies with available spectroscopic redshifts may still be biased against towards brighter sources and populations with higher spectroscopic success rates.

2.2.3 Radio fluxes

Radio observations for the COSMOS field were taken from the recently released deep VLA observations presented in Smolčić et al. (2017a); the VLA-COSMOS 3 GHz Large Project. Reaching a median rms $\approx 2.3\mu$Jy/beam over the COSMOS field with at a resolution of 0.75″, these observations represent the deepest currently available deep extragalactic radio survey covering a representative volume.

---

[2] The goal of HELP is to produce a comprehensive panchromatic dataset for studying the galaxy population at high redshift - assembling multi-wavelength data and derived galaxy properties over the ~ 1200 deg^2 surveyed by the Herschel Space Observatory.

Radio sources from the Smolčić et al. (2017a) catalog were matched to their optical counterparts based on the optical matches to Laigle et al. (2016) provided in the companion paper Smolčić et al. (2017b). Within the spectroscopic redshift subsample there are a total of 3400 radio detected sources. While a comparison of the difference in the redshift distribution and source types between a 150 MHz and 3GHz selected survey may be of scientific interest, it is a topic which we do not intend to address here. To facilitate direct comparison with the LOFAR 150MHz fluxes, we convert the observed 3 GHz fluxes to estimated 150MHz fluxes assuming a median 3000 to 150MHz spectral slope of $\alpha = -0.7$ (Smolčić et al. 2017a; Calistro Rivera et al. 2017).

2.3 Flagging of known X-ray sources and known IR/Optical AGN

In deep radio continuum surveys, the radio detected population includes a very diverse range of sources, ranging from rapidly star-forming galaxies to radio quiet quasars and massive elliptical galaxies hosting luminous radio AGN. To fully characterise the diverse radio population and to facilitate comparison between the radio population and other AGN selection methods, we classify all sources in the spectroscopic comparison samples using the following additional criteria:

• *Infrared AGN* are identified using the updated IR colour criteria presented in Donley et al. (2012). In addition to the colour criteria outlined by Donley et al. (2012), we split the IR AGN sample into two subsets based on their signal to noise in the IRAC 5.6 and 8μm bands. To be selected as a candidate IR AGN, we require that all sources have $S/N > 5$ at 3.5 and 4.6μm and $S/N > 2$ at 5.6 and 8μm. The subset of robust AGN sources is then based on a stricter criteria of $S/N > 5$ at 5.6 and 8μm.

• *X-ray* selected sources in the Boötes field were identified by cross-matching the positions of sources in our catalog with the X-Boötes *Chandra* survey of NDWFS (Kenter et al. 2005). We matched the x-ray sources through a simple nearest-neighbour match between the optical photometry catalog used in this analysis and the position of most-likely *optical* counterpart for each x-ray source presented in (Brand et al. 2006). In COSMOS, we make use of the compilation of matched x-ray data presented in Laigle et al. (2016) and the corresponding papers detailing the X-ray sources and optical cross-matching (Civano et al. 2016; Marchesi et al. 2016a).

For both fields we calculate the x-ray-to-optical flux ratio, $X/O = \log_{10}(f_X/f_{\rm opt})$, based on the i^+ or I band magnitude following Salvato et al. (2011) and Brand et al. (2006) respectively. To be selected as an X-ray AGN, we require that an x-ray source have $X/O > -1$ or an x-ray hardness ratio > 0.8 (Bauer et al. 2004).

• Bright, known *Optical AGN* were also identified through two additional selection criteria. Where available, any sources which have been spectroscopically classified as AGN are flagged. Secondly, we also cross-match the optical catalogs with the Million Quasar Catalog compilation of optical AGN, primarily based on SDSS (Alam et al. 2015) and other literature catalogs (Flesch 2015). Objects in the million quasar catalog were cross-matched to the photometric catalogs using a simple nearest neighbour match in RA
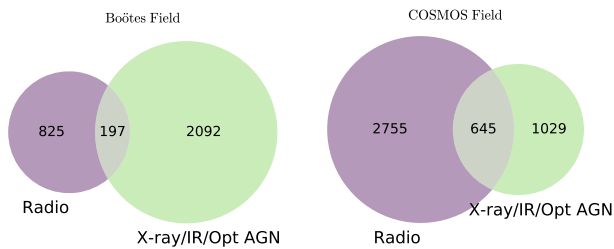
Figure 1. Multi-wavelength classifications of the sources in the full spectroscopic redshift samples. The 'Radio' and 'X-ray/IR/Opt AGN' subsets correspond respectively to radio detected sources and identified X-ray sources and optical/spectroscopic/infra-red selected AGN (see Section 2.3). As illustrated in previous studies, the X-ray, IR AGN and radio source population are largely distinct populations with only partial overlap.
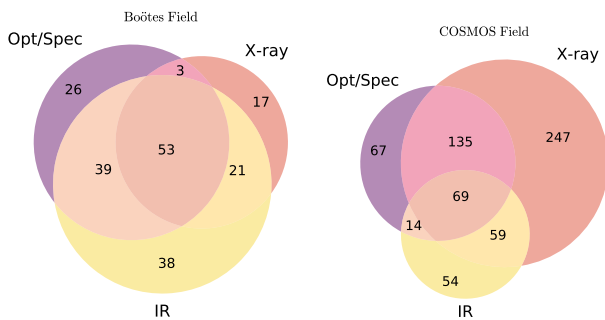

Figure 2. Multi-wavelength classifications of the radio detected sources within the spectroscopic redshift sample. For the 214 and 711 radio detected sources in the Boötes and COSMOS fields respectively, subsets which satisfy the X-ray, Optical ('Opt/Spec') and IR AGN criteria as defined in Section 2.3.

and declination and allowing a maximum separation of $1''$. Simulations using randomised positions indicate that at $1''$ separation, the chance of a spurious match with an object in the optical catalog is less than 5%. While this value is relatively high due to the depth of the optical catalogs, the actual median separation between matches is $\approx 0.2''$ and are highly unlikely to be spurious. Visual inspection of the quasar catalog sources with no optical counterpart in our catalog indicates that the majority fall within masked regions of the optical catalog (e.g. around bright stars and artefacts) and thus are not expected to have a match.

In Fig. 1 we show the relative numbers of radio detected sources and sources which satisfy any of the X-ray/Optical IR AGN critera within the full spectroscopic subsets. In both the Boötes and COSMOS spectroscopic samples, there are large numbers of radio or X-ray detected sources, as well as large numbers of sources classified as IR AGN. For the Boötes field, the large number of IR AGN is due to the specific selection criteria targeting these sources within the AGES spectroscopic survey (Kochanek et al. 2012).

Within the subset of radio detected sources itself there is a clear diversity in the nature of sources. In Fig. 2 we show
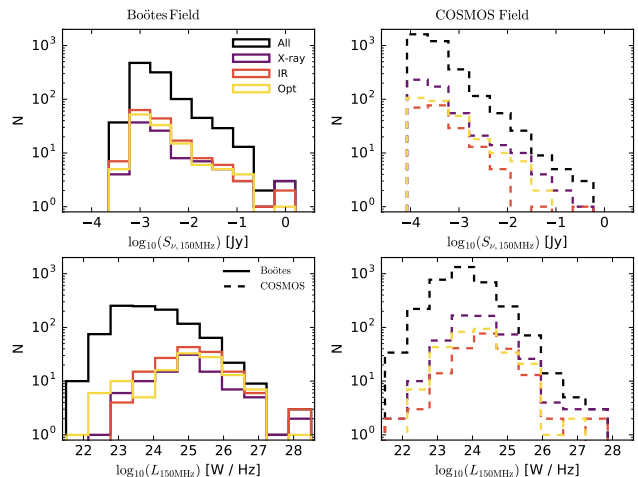

Figure 3. Flux and luminosity distributions (top and bottom rows respectively) of the radio detected sources within the Boötes and COSMOS (left and right columns) spectroscopic sample used in this analysis. Plotted in both pairs of histograms are the fluxes (or luminosities) of all radio detected sources (black lines) as well as those which are also X-ray sources (Brand et al. 2006, purple lines), infrared AGN following the criteria of Donley et al. (2012, orange lines), or optical/spectroscopically identified AGN (yellow lines).
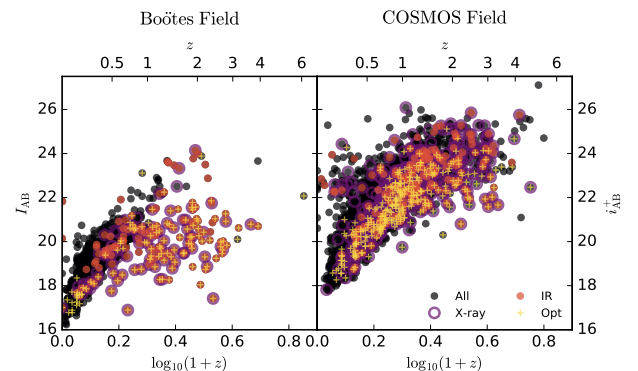

Figure 4. Optical magnitude vs redshift distributions of the radio detected sources within the Boötes and COSMOS (left and right panels respectively) spectroscopic sample used in this analysis. Plotted in both plots are the I (i^+) magnitude and spectroscopic redshift of all radio detected sources (black circles). For each source, additional markers are added to illustrate whether it is X-ray detected (purple ring), an infrared AGN following the criteria of Donley et al. (2012, orange circle), or an optical/spectroscopically identified AGN (light yellow crosses).

the multi-wavelength classifications of the respective radio samples.

Inspecting the radio flux and luminosity distributions of the two samples (Fig. 3) reveals that the X-ray detected sources and IR AGN typically have a higher radio luminosity than the sample median - in line with the expected dominance of AGN at $L_{150MHz} \gtrsim 10^{24}$ W/Hz (Jarvis & Rawlings 2004; Padovani 2016). However as seen in Fig. 2, of the most radio luminous sources, e.g. $L_{150MHz} > 10^{25}$ W/Hz, only $\sim 40-50\%$ also satisfy another AGN selection criterion. Of all X-ray and IR AGN sources in our samples, we note that

$\approx 10 - 20\%$ are radio detected, broadly consistent with the measured radio-loud fraction of optical quasars (Jiang et al. 2013).

Finally, to illustrate the magnitude and redshift parameter space probed by our spectroscopic redshift comparison samples, in Fig. 4, we plot the apparent $I(i^+)$ band magnitudes and redshifts of the radio detected populations. By construction, the 'Deep' COSMOS sample probes to significantly fainter magnitudes than the wide area Boötes sample. Between the two samples we are able to sample a wide range of magnitudes in the redshift range $0 < z < 3$ ($0 < \log_{10}(1 + z) \lesssim 0.6$). We also caution that due to the nature of the AGES spectroscopic survey selection criteria (Kochanek et al. 2012), the majority of spectroscopic redshifts at $z > 1$ in the Boötes field are known AGN. Conclusions on the photo-zs for sources at $z > 1$ will therefore largely be driven by the less biased COSMOS sample.

3 PHOTOMETRIC REDSHIFT METHODOLOGY

Photometric redshift estimation techniques fall broadly into two distinct categories. Firstly, one can use redshifted empirical or model template sets fitted to the model photometry through χ^2-minimisation or maximum likelihood techniques (e.g. Arnouts et al. 1999; Bolzonella et al. 2000; Benítez 2000; Brammer et al. 2008). Alternatively, one can take a representative training set of objects that has known spectroscopic redshifts and use any of a wide variety of supervised or un-supervised machine learning algorithms to estimate the redshifts for the sample of galaxies for which the redshift is unknown (e.g. Collister & Lahav 2004; Brodwin et al. 2006; Carrasco Kind & Brunner 2013, 2014a; Almosallam et al. 2016a,b).

In recent years, empirical methods based on training sets have been shown to produce redshift estimates that can have lower scatter and outlier fractions than template-based methods (Sanchez et al. 2014; Carrasco Kind & Brunner 2014b). Furthermore, because the computationally expensive training step only occurs once these methods can also be significantly faster than template fitting when applied to very large datasets.

However, the drawback of training sample methods is that they are very dependent on the parameter space covered by the training sample and its overall representativeness of the sample being fitted (Beck et al. 2017). While template-fitting methods do benefit from additional optimisation through spectroscopic training samples (Section 3.2, see also Hildebrandt et al. 2010; Dahlen et al. 2013), they can be applied effectively with no prior redshift knowledge and tested without spectroscopic samples for comparison (Quadri & Williams 2010).

Fully representative training samples for the rare sources of interest are not yet readily available for many different fields. Contributing to this problem is the inhomogeneous nature of the photometric data both within and across the various deep survey fields. While deep spectroscopic samples are available for fields such as COSMOS, the variation in filter coverage between survey fields makes it impractical to fully apply this training sample to other fields. Given these constraints, we believe that template based pho-

tometric redshifts still represent the best starting point when estimating photo-zs for the datasets and science goals of interest. Future work will explore the application of such empirical photo-z estimates to the widest tiers of the LOFAR survey. For this study we base the photometric redshift estimates on the EAZY photometric redshift software presented in Brammer et al. (2008).

As mentioned above, several different template fitting photometric redshift codes have been published and have been widely used in the literature, e.g. BPZ (Benítez 2000), LePhare (Arnouts et al. 1999; Ilbert et al. 2005) or HyperZ (Bolzonella et al. 2000). The key differences in approach (and potential outcomes) between these codes are primarily the choice in default template sets as well as their treatment of redshift priors based on magnitude or spectral type. Both of these assumptions can be changed either within EAZY itself or in subsequent analysis of its outputs. We are therefore confident that our choice of specific photometric redshift code does not strongly bias the results of our analysis and we note that alternative template fitting codes could be used without systematically affecting the results.

3.1 Template sets

The three template sets used in this analysis are as follows:

(i) Default EAZY reduced galaxy set ('EAZY'):
The first set used are the updated optimised galaxy template set provided with EAZY and we refer the reader to Brammer et al. (2008) for full details of how these templates were generated. In the latest version of the software, this template set has been updated to incorporate nebular emission lines and includes both an additional dusty galaxy template and an extremely blue SED with strong line emission.

Because the EAZY template set includes only stellar emission it gives poor fits at wavelengths where the overall emission is typically dominated by non-stellar radiation (e.g. rest-frame mid-infrared; dust emission/PAH features). To minimise the effect of this potential bias, observed filters with wavelengths greater than that of IRAC channel 2 (4.5μm) are not included when fitting.

(ii) Salvato et al. (2008) 'XMM-COSMOS' templates:
Our second set of templates is that presented by Salvato et al. (2008, 2011) in their analysis of photometric redshifts for X-ray AGN. Based on the templates presented in Polletta et al. (2007, see also references within), this template set includes 30 SEDs and covers a wide range of galaxy spectral types in addition to both AGN and QSO templates. In contrast to the EAZY templates, the XMM-COSMOS templates include both dust continuum and PAH features as well as power-law continuum emission for the appropriate AGN templates. We therefore do not exclude the IRAC 5.8 and 8.0 μm photometry when fitting with these templates.

(iii) Brown et al. (2014) Atlas of Galaxy SEDs ('Atlas'):
Finally, we make use of the large atlas of 129 galaxy SED templates presented in Brown et al. (2014, referred to as 'Atlas' hereafter). These templates are based on nearby galaxies and cover a broad range of galaxy spectral types including ellipticals, spirals and luminous infrared galaxies (both starburst and AGN). Constructed from panchromatic synthetic SED models (da Cunha et al. 2008) and optical to

infrared photometry and spectroscopy, the library has been constructed to minimise systematic errors and span the full gamut of nearby galaxy colours. As with the template set 2, because the templates include rest-frame mid-infrared spectral and continuum features, IRAC 5.8 and 8.0 μm photometry were also used when fitting with this library.

These three template libraries were selected either because of their common use within the literature (EAZY/XMM-COSMOS) or because of their explicit intention to fully represent the range of colours observed in local galaxies (Atlas). They are however not directly comparable in the intrinsic galaxy types they include and there are some key differences which could affect their potential performance for the radio galaxy population. As mentioned above, the EAZY template set models only stellar emission and does not include any templates with contributions from AGN. We may therefore expect the EAZY template set to perform very poorly for galaxies with SEDs which are dominated by AGN components.

In contrast, while the Atlas library does include templates with significant AGN contributions (primarily at longer wavelengths), it does not include any bright optical quasars due to its local galaxy selection. The XMM-COSMOS library is therefore the only set included in this analysis which includes the full range of optical AGN classes.

3.2 Photometric zeropoint offsets

The addition of small magnitude offsets to the observed photometry of some datasets has been shown to improve photometric redshift estimates (Dahlen et al. 2013). While typically small ($\lesssim 10\%$), these additional offsets can often substantially reduce the overall scatter or outlier fractions for photo-z estimates.

To calculate the appropriate photometric offsets we use the commonly followed strategy of fitting the observed SEDs of a subset of galaxies while fixing their redshift to the known spectroscopic redshift. For a training sample of 80% of the available spectroscopic sample, the offset for each band is then calculated from the median offset between the observed and fitted flux values for sources with $S/N > 3$ in that band.

To ensure that spurious offsets are not being applied based on a small number of catastrophic failures in the photometry we perform a bootstrap analysis to calculate the scatter in estimated zeropoint offsets. The zeropoint offset is calculated for 100 iterations of a random subset of 10% of the spectroscopic training sample, with the standard deviation of this distribution then taken as the uncertainty in the zeropoint offset. An offset is then only applied to a given band if the offset is significant at the 2σ level.

We apply this procedure to each template set individually, with the zeropoint offsets applied in all subsequent analysis steps. Using the remaining 20% of spectroscopic redshifts as a test sample we are able to verify that for each template set the inclusion of the zeropoint offsets in the fitting produces an overall improvement in the various photometric redshift quality metrics.

Finally, before including the estimated photometric offsets in the fitting process for the full photometric samples we assess any potential adverse effects they could have. For the two example fields used in this study we find that there is

no strong bias in the photometric offsets introduced by the redshift distribution of the spectroscopic sample. That is to say, applying photometric offsets based on a spectroscopic sample with $\overline{z_s} \approx 0.3$ to a sample of photometric galaxies at higher redshift will not strongly bias the resulting redshifts. Such biases could arise either from aperture effects (due to the larger angular size of nearby galaxies) or from differences in the age-dependent features (e.g. 4000Åbreak) in the SEDs; a problem which which may be most acute for the local galaxy based Atlas template library. However, we find that for the extreme example of applying photometric offsets calculated for a spectroscopic sample at $z \sim 0.2$ to a test sample at $z > 1$, the photometric redshift quality of the test sample with the 'biased' offsets applied is not significantly worse than when no offsets were applied.

3.3 Fitting methods

The EAZY template set is fitted following their intended use, using fits of N-linear combinations of templates and allowing all templates to be included in the fit.

In contrast, the XMM-COSMOS templates are used in a way which best matches their implementation in LePhare (Arnouts et al. 1999) and their intended use (Salvato et al. 2011). A range of dust attenuation levels ($0 \leq A_V \leq 2$) is applied to each of the 32 unique templates, using both the Calzetti et al. (2000) starburst attenuation law and the Pei (1992) Small Magellanic Cloud (SMC) extinction curve. The extended set of dust attenuated templates are then fitted using single template mode in EAZY.

Due to the large number of unique templates already included (making fits of N-linear combinations impractical), the Atlas template set is fitted in a similar manner to the XMM-COSMOS set. To allow for finer sampling of the rest-frame UV/optical emission in the empirical Atlas SEDs, we also apply additional dust attenuation to the empirical templates as was done for the XMM-COSMOS set. Due to the wider range of dust extinction already intrinsic to the empirical templates, we apply a smaller range of additional dust attenuation ($0 \leq A_V \leq 1$) and assume only the Calzetti et al. (2000) starburst attenuation law. We note that the maximum dust extinction of $A_V = 1$ may be unrealistic for some of the galaxy archetypes included in the Atlas library (e.g. blue compact dwarfs), but dust ranges tuned to individual template type is beyond the scope of this work. As for the XMM-COSMOS template set, the extended Atlas of Galaxy SEDs template set is then fit in single template mode.

For all three template sets, additional rest-frame wavelength dependent flux errors are also included through the EAZY template error function (see Brammer et al. 2008). These errors are added in quadrature to the input photometric errors and vary from $< 5\%$ at rest-frame optical wavelengths to $> 15\%$ at rest-frame UV and near-IR where template libraries are more poorly constrained.

Finally, although EAZY allows for the inclusion of a magnitude dependent prior in the redshift estimation, we choose not to include it at this stage. A summary of these three different photo-z fitting estimates is presented in Table 1 for reference.

Table 1. Summary of the three template sets used in the photometric redshift analysis, including details of how the templates were fitted, whether any dust attenuation was applied to the original template sets and whether the template set includes contributions from AGN emission.

Template Set	N Templates	Fitting Mode	Dust Attenuation Applied	AGN Included
EAZY Brammer et al. (2008)	9	N-linear combinations	N/A	No
XMM-COSMOS Salvato et al. (2008, 2011)	32	Single Template	$0 \leq A_V \leq 2, \delta A_V = 0.2$ Calzetti et al. (2000), Pei (1992, SMC)	Yes
Atlas of SEDs Brown et al. (2014)	129	Single Template	$0 \leq A_V \leq 1, \delta A_V = 0.2$ Calzetti et al. (2000)	Yes

4 RESULTS

To explore the performance of the three template sets on the two spectroscopic samples, we first look at the statistics of the best-fit photometric redshift estimates relative to the measured spectroscopic redshifts. Within the literature there are a wide range of statistical metrics used to quantify the quality of photometric redshifts (see Dahlen et al. 2013; Sanchez et al. 2014; Carrasco Kind & Brunner 2014b). In this analysis we choose to adopt a subset of the metrics outlined in Dahlen et al. (2013), including three measures of the redshift scatter, one measure of the bias and one of the outlier fraction (see Table 2 for details of these definitions and their notation).

We also introduce an additional metric, the continuous ranked probability score (CRPS) and the corresponding mean values for a given sample (Brown 1974; Matheson & Winkler 1976). Widely used in meteorology, the CRPS is designed for evaluating probabilistic forecasts. We refer to Hersbach & Hersbach (2000, see also: Polsterer et al. (2016)) for full details of the metric and its behaviour, but its definition is presented in Table 2 and represents the integral of the absolute difference between the cumulative redshift distributions of the predicted value $(\mathrm{CDF}(z))$ and true values $(\mathrm{CDF}_{z_s}(z)$: i.e. a Heaviside step function at z_s). A key advantage over the more widely used metrics is that CRPS takes into account the full PDF rather than just a simple point value when evaluating a model prediction (i.e. the photometric redshift).

4.1 Overall photometric redshift accuracy

Before analysing the photometric redshift properties of the radio source population specifically, it is useful to first verify the overall redshift accuracy of the estimates in the respective fields. For galaxy evolution studies (where the overall bias is less critical), the two most important metrics are typically the robust scatter, σ_{NMAD}, and outlier fraction, O_f. Figures 5 and 6 illustrate how these metrics vary with redshift and magnitude for the full Boötes and COSMOS spectroscopic samples. In Table 3 we also present the corresponding photo-z quality metrics for the full spectroscopic sample and all subsets of radio detected sources.

As expected given the availability of medium band observations, the COSMOS photo-zs (Fig. 5) typically have lower scatter than the Boötes dataset at any given redshift. However, at $z \lesssim 1$ the photo-zs for all template sets perform
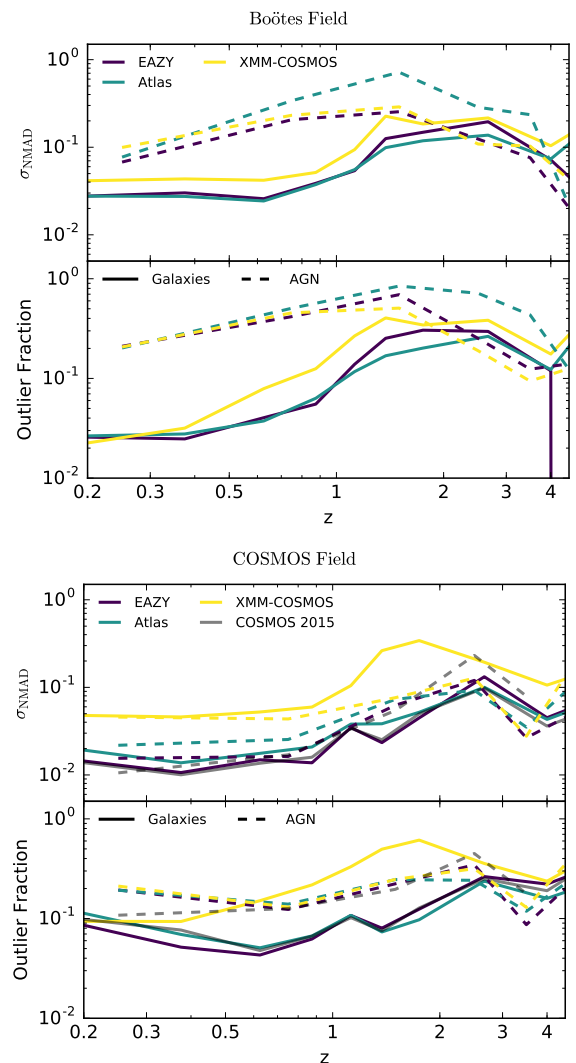

Figure 5. Photometric redshift scatter (σ_{NMAD}) and outlier fraction as a function of spectroscopic redshift for the Boötes field (top) and COSMOS fields (bottom) respectively. In both plots, dashed lines show the results for sources which pass any of the X-ray/Optical/IR AGN criteria outlined in Section 2.3 and solid lines show the results for sources which do not satisfy any of these criteria. The 'COSMOS 2015' line corresponds to the combined literature COSMOS photometric redshift values from Laigle et al. (2016) and Marchesi et al. (2016b).

Table 2. Definitions of statistical metrics used to evaluate photometric redshift accuracy and quality along with notation used throughout the text.

Metric		Definition
σ_f	Scatter - all galaxies	$\mathrm{rms}(\Delta z/(1 + z_{\mathrm{spec}}))$
σ_{NMAD}	Normalised median absolute deviation	$1.48 \times \mathrm{median}(|\Delta z|/(1 + z_{\mathrm{spec}}))$
Bias		$\mathrm{median}(\Delta z)$
O_f	Outlier fraction	Outliers defined as $|\Delta z|/(1 + z_{\mathrm{spec}}) > 0.2$
σ_{O_f}	Scatter excluding O_f outliers	$\mathrm{rms}[\Delta z/(1 + z_{\mathrm{spec}})]$
$\overline{\mathrm{CRPS}}$	Mean continuous ranked probability score	$\overline{\mathrm{CRPS}} = \frac{1}{N}\sum_{i=1}^{N}\int_{-\infty}^{+\infty}[\mathrm{CDF}_i(z) - \mathrm{CDF}_{z_s,i}(z)]^2 dz$ - Hersbach & Hersbach (2000)

Table 3. Photometric redshift quality statistics for the Boötes (left) and COSMOS (right) spectroscopic samples. The statistical metrics (see Table 2) are shown for the full spectroscopic sample, the radio detected sources and for various subsets of the radio population. For each subset, values from the best performing are highlighted in bold font.

			Boötes						COSMOS			
Templates	σ_f	σ_{NMAD}	Bias	O_f	σ_{O_f}	$\overline{\mathrm{CRPS}}$	σ_f	σ_{NMAD}	Bias	O_f	σ_{O_f}	$\overline{\mathrm{CRPS}}$
			All Sources						All Sources			
EAZY	0.772	0.040	0.007	0.111	0.049	0.293	**0.365**	**0.019**	**-0.002**	**0.096**	**0.048**	**0.157**
Atlas	0.879	**0.037**	**-0.005**	0.128	**0.046**	0.314	0.460	0.025	-0.004	0.103	0.051	0.174
XMM-C	**0.658**	0.055	-0.010	**0.095**	0.060	**0.218**	0.384	0.070	-0.029	0.214	0.071	0.249
			All Radio Sources						All Radio Sources			
EAZY	0.573	0.038	0.010	**0.120**	0.044	0.274	**0.241**	**0.016**	**-0.001**	**0.081**	**0.042**	**0.115**
Atlas	**0.471**	**0.037**	-0.005	0.144	0.046	0.257	0.241	0.020	-0.001	0.088	0.045	0.121
XMM-C	0.566	0.056	**-0.003**	0.124	0.059	**0.210**	0.236	0.038	-0.013	0.104	0.054	0.133
			Radio Sources - Non X-ray/IR/Opt AGN						Radio Sources - Non X-ray/IR/Opt AGN			
EAZY	0.490	0.029	0.005	0.030	**0.041**	0.203	**0.365**	**0.019**	**-0.002**	**0.096**	**0.048**	**0.157**
Atlas	**0.407**	**0.027**	-0.003	**0.027**	0.042	**0.132**	0.460	0.025	-0.004	0.103	0.051	0.174
XMM-C	0.551	0.047	**-0.002**	0.049	0.054	0.151	0.384	0.070	-0.029	0.214	0.071	0.249
			Radio Sources - X-ray Detected						Radio Sources - X-ray Detected			
EAZY	0.769	0.648	0.307	0.606	**0.055**	0.734	0.296	**0.028**	0.006	**0.169**	0.053	**0.225**
Atlas	0.848	0.594	-0.064	0.681	0.067	0.858	0.283	0.052	**0.001**	0.175	0.062	0.240
XMM-C	**0.569**	**0.261**	**0.010**	**0.436**	0.089	**0.487**	**0.273**	0.056	-0.009	0.186	0.063	0.231
			Radio Sources - IR AGN						Radio Sources - IR AGN			
EAZY	0.742	0.503	0.236	0.609	**0.066**	0.702	0.516	**0.079**	0.016	**0.303**	**0.064**	0.445
Atlas	0.751	0.546	-0.298	0.775	0.083	0.965	**0.454**	0.132	**-0.005**	0.316	0.078	0.433
XMM-C	**0.695**	**0.288**	**-0.012**	**0.510**	0.097	**0.505**	0.513	0.142	-0.036	0.393	0.080	**0.406**
			Radio Sources - Opt AGN						Radio Sources - Opt AGN			
EAZY	0.661	0.406	0.131	0.545	**0.069**	0.646	**0.162**	0.041	**-0.002**	0.186	**0.056**	**0.248**
Atlas	0.770	0.603	-0.317	0.769	0.070	1.046	0.175	0.064	-0.014	0.211	0.060	0.292
XMM-C	**0.500**	**0.187**	**0.000**	**0.430**	0.077	**0.503**	0.214	0.055	-0.020	0.233	0.064	0.364
			Radio Sources - $\log_{10}(L_{150\mathrm{MHz}}[\mathrm{W/Hz}]) > 25$						Radio Sources - $\log_{10}(L_{150\mathrm{MHz}}[\mathrm{W/Hz}]) > 25$			
EAZY	**0.477**	0.081	**0.014**	0.341	**0.046**	0.478	**0.218**	**0.015**	**0.000**	**0.094**	**0.041**	**0.123**
Atlas	0.539	0.290	-0.049	0.512	0.052	0.765	0.233	0.021	-0.001	0.099	0.044	0.135
XMM-C	0.498	**0.074**	-0.033	**0.279**	0.075	**0.403**	0.225	0.036	-0.011	0.104	0.053	0.137

well in both fields, with $0.03 \lesssim \sigma_{\mathrm{NMAD}} \lesssim 0.05$ (Boötes) and $0.01 \lesssim \sigma_{\mathrm{NMAD}} \lesssim 0.03$ (COSMOS).

For both samples we find that the redshift estimates for X-ray detected and Opt/IR AGN population (dashed lines) typically perform worse on average than the remaining galaxy population at $z < 2$. However, at $z \gtrsim 2.5$ the populations begin to converge to equivalent levels of scatter and outlier fraction. While the 'normal' galaxies at $z > 2$ deteriorate in quality (likely a result of decreasing S/N - Figure 6), the photo-z estimates for sources in the X-ray/Opt/IR AGN sample begin to improve. This convergence at higher redshift is potentially driven by the increasing importance of

the common Lyman break feature in determining the fitted redshift.

While the primary goal of this paper is to draw conclusions on the relative photo-z accuracies for different source populations, it is also useful to compare the absolute accuracy of the photo-zs produced relative to those of high-quality sets available in the literature. Therefore, in addition to the comparison between template sets, in Fig. 5 we also present the quality metrics of the published COSMOS2015 photometric redshift set (Laigle et al. 2016) for the same spectroscopic sample (the catalog 'photoz' column; grey lines). Because the COSMOS2015 estimates are
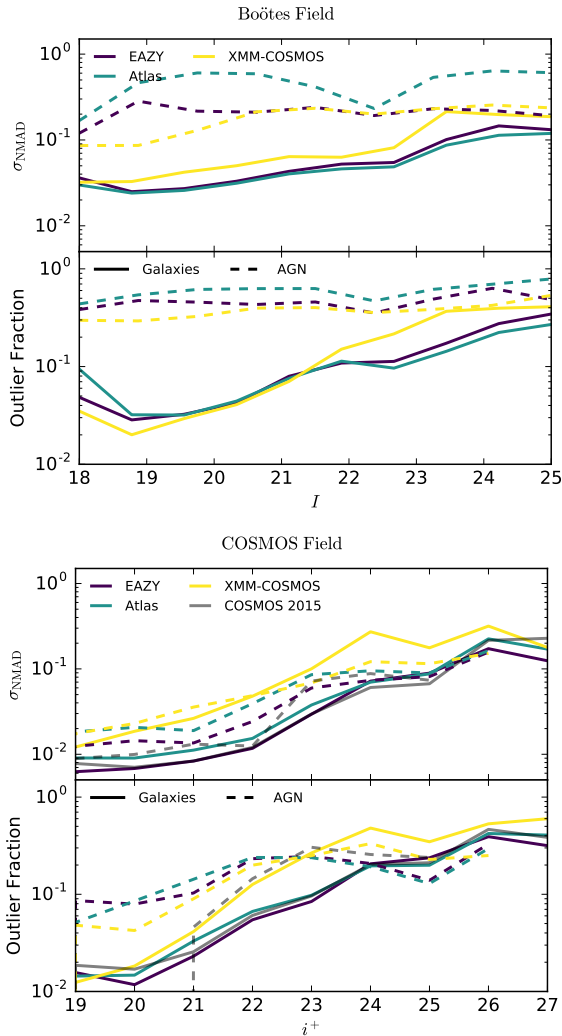
Figure 6. Photometric redshift scatter (σ_{NMAD}) and outlier fraction as a function of I, or i^+, magnitude for the Boötes field (top) and COSMOS fields (bottom) respectively. In both plots, dashed lines show the results for sources which pass any of the X-ray/Optical/IR AGN criteria outlined in Section 2.3 and solid lines show the results for sources which do not satisfy any of these criteria. The 'COSMOS 2015' line corresponds to the combined literature COSMOS photometric redshift values from Laigle et al. (2016) and Marchesi et al. (2016b).

not optimised for AGN (and exclude estimates for some X-ray sources), photo-zs for X-ray detected galaxies are taken from the results of Marchesi et al. (2016b). For the 'normal' galaxy population, the scatter and outlier fractions of the EAZY and Atlas template sets perform comparably to the official COSMOS2015 estimates. In contrast, for the Laigle et al. (2016) estimates alone the X-ray/IR/opt AGN sample perform significantly worse than the best estimates from this analysis. Incorporating the photo-zs for X-ray sources from Marchesi et al. (2016b, as shown in Fig. 5) the combined literature photo-zs performance improves, with scatter and outlier fractions at $z < 2$ comparable to the best estimates from this analysis, but a poorer performance above this range.

For both the Boötes and COSMOS samples we find that

the EAZY and Atlas template sets perform comparably and typically produce the lowest scatter (σ_{NMAD}) in both the full spectroscopic sample and the full radio selected population. However, in the sub-samples of X-ray detected sources or IR AGN, we find no consistency between the two different datasets. In the wide area dataset, the XMM-COSMOS template set performs significantly better in almost all metrics than the other two sets, for AGN populations (see Table 3 . Conversely, in the deep field the XMM-COSMOS set performs worst for the key σ_{NMAD} and O_{f} metrics in the subset of X-ray/Opt/IR AGN sources.

Given the consistent methodology used for both datasets, the underlying reason for this discrepancy is likely due to the differences in the source populations included in the relevant spectroscopic samples (see Section 2.3). As seen in Fig. 4, the Boötes X-ray/Opt/IR AGN source population is typically significantly optically brighter than that probed in COSMOS and may therefore have intrinsically different SEDs.

One clear conclusion that can be drawn from Figures 5/6 and Table 3 is that there is no single template set which performs consistently best across all subsets and datasets. Differences in the redshifts estimated by the three different template sets are found to systematically depend strongly both on optical magnitude (a proxy for overall S/N) and redshift. Specifically, as sources become optically fainter the range between the highest and lowest predicted redshifts systematically increases. As a function of redshift, this range of predicted photo-zs also increases significantly between $1 \lesssim z \lesssim 3$; above $z \sim 3$ the estimates begin to converge again. We see these trends in both the wide and deep fields, leading us to conclude that the redshift effect is not due to the systematics of the available optical data itself (e.g. the relatively shallow near-IR data in Boötes).

4.2 Relative photo-z accuracy for radio and non-radio sources

It is clear that the absolute values for photometric redshift quality metrics are strong functions of the redshifts being probed, along with relative depth (S/N), resolution and wavelength coverage of the photometry available. The fundamental question for photometric redshift estimates in deep radio continuum surveys is how does the redshift accuracy differ between the radio detected and radio undetected source populations?

To understand how the different intrinsic source populations affect the resulting photo-z accuracy, we therefore measured the relative photo-z scatter and outlier fraction between the radio detected and non detected populations as a function of redshift. To minimise the effects of known biases or photo-z quality dependencies, we first carefully match the two samples in redshift, magnitude and colour space.

Within the 3 dimensional parameter space of the spectroscopic redshift, I-band (i^+) magnitude and $I - 3.6\mu m$ ($i^+ - 3.6\mu m$) colour, we calculate the 10 nearest neighbours for each radio detected source. Due to the limited number of spectroscopic redshifts available, sources in the non-radio sample are allowed to be in the matched sample for more than one radio source. Next, for each redshift bin, we calculate σ_{NMAD} and O_f for the two matched samples and use
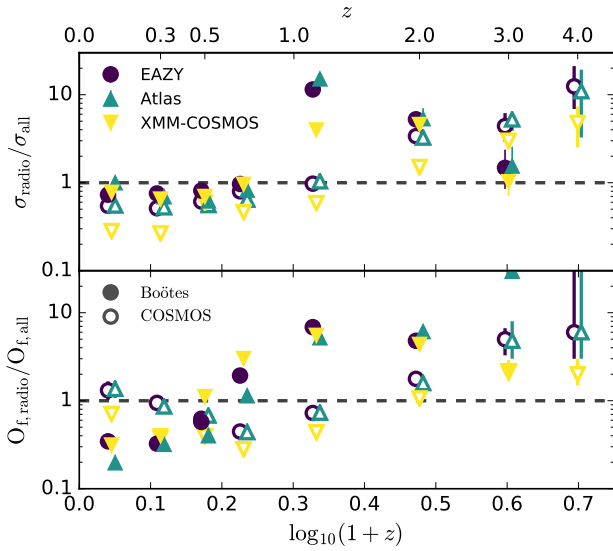
Figure 7. Photometric redshift quality for radio detected sources σ_{radio} ($O_{\mathrm{f,\ radio}}$) relative to matched samples of sources with no radio detection σ_{all} ($O_{\mathrm{f,\ all}}$) as a function of redshift, where σ and O_{f} correspond to the normalised median absolute deviation and the outlier fraction as defined in Table 2. Details of sample matching procedure are outlined in Section 4.2. In each plot we show the values for the EAZY (circles), Atlas (upward triangles) and XMM-COSMOS (downward triangles) template estimate for both the Boötes field (filled symbols) and COSMOS fields (empty symbols).

a simple bootstrap analysis to estimate the corresponding uncertainties in these metrics.

In Fig. 7 we show the relative scatter and outlier fractions of these two matched samples. We find that up to $z \sim 1$ (where both spectroscopic samples are fairly representative), photometric redshifts estimated for radio sources have typically lower scatter and outlier fraction than galaxies with no radio detection that have similar magnitudes. This trend is true for both datasets and for all three template sets.

Above $z \sim 1$, photo-zs for radio sources are significantly worse than their matched non-radio detected counterparts. This trend of increasing scatter/outlier fraction with redshift is not unexpected given as redshift increases the radio detected sources are increasingly luminous AGN for which photo-z estimates are expected to struggle.

In the Boötes field specifically (filled symbols), we see that at $z \sim 1$ there is a significant jump in the measured scatter for radio sources. Inspecting the magnitude-redshift distribution of the radio sample reveals that $z \sim 1$ ($\log_{10}(1 + z) \approx 0.3$) marks the transition where the AGES spectroscopic sample become dominated by the AGN selection criteria and almost all sources are classified as either X-ray sources or IR AGN. We note however that the sample bias towards X-ray and IR AGN sources is true for both the radio and matched non radio samples, indicating that at higher redshift the radio-loud subset of X-ray/IR AGN sources is systematically more difficult to fit than the radio-faint population of similar magnitude.
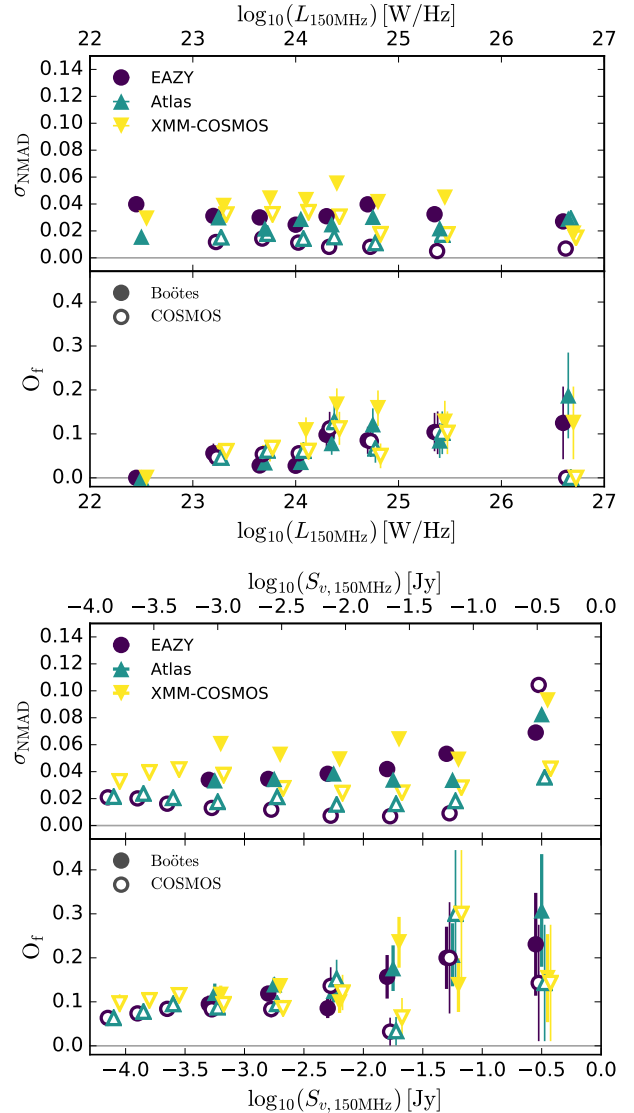

Figure 8. Photometric redshift scatter (σ_{NMAD}; upper panels) and outlier fraction (O_f; lower panels) as a function of 150MHz radio luminosity (top) and flux (bottom) for galaxies within the redshift range $0.2 < z < 0.9$. In each plot we show the values for the EAZY (circles), Atlas (upward triangles) and XMM-COSMOS (downward triangles) template estimate for both the Boötes field (filled symbols) and COSMOS fields (empty symbols). Symbols have been offset horizontally only for clarity, all luminosity/flux bins are identical. Error-bars plotted for the outlier fractions illustrate the binomial uncertainties on each fraction.

4.3 Photometric redshift accuracy as a function of radio power

In Section 4.2 we saw that the photo-zs for the radio detected population becomes systematically worse at high redshift. If this trend is driven by the evolution in sample radio luminosities from the flux limited samples, we expect to observe the same trend when looking at a fixed redshift but evolving radio luminosity. In Fig. 8 we present the evolution in σ_{NMAD} and O_f as a function of $\log_{10}(L_{150\mathrm{MHz}})$ for sources with spectroscopic redshift in the range $0.2 < z < 0.9$. We

choose this redshift range because based on Fig. 5 we know that the scatter and outlier fraction of the full sample do not evolve strongly across this range. In the COSMOS field, the scatter remains relatively constant with redshift for both the AGN and normal galaxy samples while the outlier fraction actually decreases slightly over this range.

In contrast to the naive expectation of increasing scatter with radio luminosity, it is evident that there is no clear evolution in σ_{NMAD} across the ~ 4 orders of magnitude in radio luminosity probed by our samples. The measured scatter follows a similar trend when examined as a function of radio flux. Between < 1 mJy and 100 mJy, σ_{NMAD} remains effectively constant for all three template libraries in both and both datasets. It is only at the very brightest radio fluxes $(0.1 < S_{\nu,150\mathrm{MHz}} < 1 \mathrm{\ Jy})$ where scatter increases for this redshift regime.

Some evidence of increasing outlier fraction as a function of $\log_{10}(L_{150\mathrm{MHz}})$ exists for both the deep and wide fields, with O_f rising by ~ 2 between $\log_{10}(L_{150\mathrm{MHz}}) \approx 23$ and $\log_{10}(L_{150\mathrm{MHz}}) \gtrsim 26$. As a function of radio flux, the trend of increasing outlier fraction is even more pronounced. Although there is significant scatter in the outlier fraction values and the small samples available at high radio power result in significant uncertainties, the trend is consistent across all three template sets.

In both fields, the overall AGN selected population has a higher outlier fraction across the redshift range $(0.2 < z < 0.9)$ than the galaxy population. The rise in outlier fraction with radio flux illustrated in Fig. 8 may therefore be a result of the increasing radio AGN fraction with radio flux (/luminosity), see e.g. Bonzini et al. (2013).

As radio surveys push to lower radio luminosities at higher redshift (e.g. $\log_{10}(L_{150\mathrm{MHz}}) < 25$ and $z \sim 2$), these results suggest that photometric redshifts for the large samples of intermediate power AGN will be very comparable to those of 'normal' galaxies. However, what we cannot measure here is the effect of how the intrinsic SEDs of these sources might change over these redshifts and the resulting effects on photo-z estimates.

4.3.1 *Best-fitting templates vs radio power*

As radio luminosity increases and the source population becomes dominated by increasingly powerful AGN, a plausible expectation is that photo-z template sets incorporating AGN/QSO templates will perform better than stellar-only template sets. Equally, at low radio luminosities where the population becomes dominated by star-forming galaxies, one would expect the template sets optimised for stellar emission to provide the most accurate photo-zs. Based on the results of Fig. 8 however, there is no clear dependence of the preferred template on radio luminosity.

This result is in line with our expectation on the radio source population; namely its extremely diverse nature. Across a broad range of radio luminosities, the observed spectral energy distributions are consistent with sources ranging from radio galaxies with old stellar populations to star-forming galaxies and luminous QSOs.
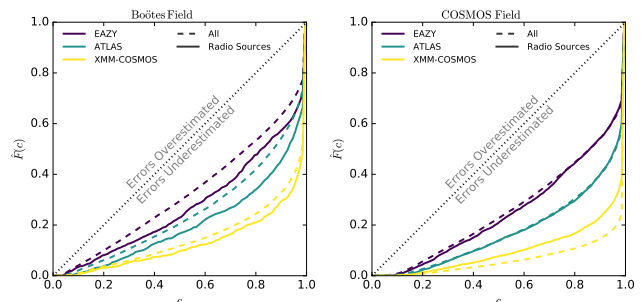

Figure 9. Q-Q ($\hat{F}(c)$) plots for the redshift PDFs for each template set, as directly produced by EAZY. Plotted are the cumulative distributions for the full spectroscopic redshift sample (solid line) and for the sub-sample of LOFAR detected sources (dashed line).

4.4 Accuracy of the redshift PDFs

While the scatter between the estimated photometric redshift (whether that is the peak or median of the $P(z)$) and the spectroscopic redshifts is a useful metric for judging their accuracy, this does not take into account the uncertainties on individual measurements nor the potential complexities of the $P(z)$ itself (e.g. multiple peaks or asymmetry). In addition to ensuring the minimum scatter and outlier fraction possible, it is therefore essential that the estimated $P(z)$ accurately represent the true uncertainty of the photometric redshifts. Even with the inclusion of additional photometry errors, it is common for template fitting photo-z codes to be over-confident in the predicted redshift accuracy (Dahlen et al. 2013; Wittman et al. 2016).

To quantify the over- or under-confidence of our photometric redshift estimates, we follow the example of Wittman et al. (2016) and x where the spectroscopic redshift is just included. For a set of redshift PDFs which perfectly represent the redshift uncertainty (e.g. 10% of galaxies have the true redshift within the 10% credible interval, 20% within their 20% credible interval, etc.), the expected distribution of c values should be constant between 0 and 1. The cumulative distribution, $\hat{F}(c)$, should therefore follow a straight 1:1 relation. Curves which fall below this expected 1:1 relation therefore indicate that there is overconfidence in the photometric redshift errors; the $P(z)$s are too sharp.

In Figure 9, we show the $\hat{F}(c)$ distributions (Q-Q plots) for the uncorrected $P(z)$ output of each template set. For both the full spectroscopic samples (dashed lines) and radio detected samples (solid lines), all three template sets show significant overconfidence in the photometric redshift errors. The $P(z)$ estimates based on the EAZY template set are the most accurate while the XMM-COSMOS template set performs the worst. We also find that despite having significantly lower scatter relative to the spectroscopic sample, the COSMOS field redshift estimates are noticeably more overconfident than those in the Boötes field.

Using a training subset of each population (AGN vs non-AGN), we smooth the redshift PDFs to minimise the euclidean distance between the observed $\hat{F}(c)$ and the desired 1:1 relation. To do this we define the rescaled redshift PDF for a galaxy, i, as

$$P(z)_{\mathrm{new},i} \propto P(z)_{\mathrm{old},i}^{1/\alpha(m_i)}, \tag{1}$$

where $\alpha(m)$ is a constant, c, below some characteristic apparent magnitude, m_c, and follows a simple linear relation above this magnitude, e.g.

$$\alpha(m) = \begin{cases} \alpha_c & m \le m_c \\ \alpha_c + \kappa \times (m - m_c) & m > m_c. \end{cases} \quad (2)$$

For both datasets, we use the equivalent I/i^+ band optical magnitude for calculating the magnitude dependence. We also assume a characteristic magnitude of $i^+ = 20$ for the COSMOS sample (Laigle et al. 2016) and $I = 18$ for the shallower Boötes sample. The parameters c and k are then fit using the EMCEE Markov Chain Monte Carlo fitting tool (MCMC Foreman-Mackey et al. 2013).

In Fig. 10 and 11 we show the resulting $\hat{F}(c)$ distributions for the all sources as well as the radio detected sources (and subsets thereof) after the redshift PDFs have been calibrated using the *full spectroscopic redshift sample*. For the Boötes field, the consensus PDFs from all three template sets are significantly improved for the full spectroscopic sample. The 0 to 50% credible interval ranges are all very accurately measured, with only a small remaining overconfidence within the 80% credible interval. For the COSMOS field, all three template sets perform well within the 50% credible interval but the tails of the distributions are not as accurate as those for the Boötes field.

Although the calibrated redshift PDFs for the radio detected subsets are somewhat improved by the calibration procedure, they do not match the same accuracy as the wider spectroscopic redshift sample. Of the three template sets, the calibrated PDFs from the XMM-COSMOS template set are the most accurate overall for the Boötes field. However, for the COSMOS field, the calibrated PDFs for the XMM-COSMOS set are under-confident for the overall radio sample.

For both the AGN and non-AGN calibration subsets the smoothing applied to the XMM-COSMOS set is significantly higher than required for the EAZY and Atlas template sets (with the exception of the Boötes AGN sample for the Atlas library; Table 4). The resulting PDFs (while accurately representing the uncertainty for the overall sample) are much broader than the other template sets. The typical 1-2σ uncertainties on an individual galaxy redshift solution are $\approx 2 - 3$ smaller for the EAZY/Atlas template sets compared to XMM-COSMOS.

5 OPTIMIZED REDSHIFTS THROUGH HIERARCHICAL BAYESIAN COMBINATION

As illustrated in Sections 4.1-4.3, no single template set can perform well for all types of radio-detected galaxy. To obtain the best photometric redshift estimates for sources in future deep radio continuum surveys one would therefore ideally like to pre-classify every galaxy and fit it with the optimum method for that source type (as successfully implemented by Fotopoulou et al. 2016). However, some of the key properties necessary for such *a priori* classification are potentially not going to be known at the time photometric redshifts are fitted.

A potential solution to this problem lies in the combination of multiple photo-z within a Bayesian framework such

Table 4. Magnitude-dependent redshift PDF smoothing parameters derived during the error calibration procedure. The smoothing is applied following Eq. 4.4 and 4.4 with the respective values of α_c and κ. Note, $m_c = 18$ and 20 for the Boötes and COSMOS fields respectively.

Templates	Boötes		COSMOS	
	α_c	κ	α_c	κ
'Galaxies'				
EAZY	1.50	0.41	0.57	0.71
Atlas	1.04	1.29	2.47	0.76
XMM-COSMOS	5.54	1.17	6.36	4.73
'AGN'				
EAZY	9.42	2.57	3.21	0.50
Atlas	25.5	1.38	3.00	1.59
XMM-COSMOS	10.7	2.22	12.2	0.45

as hierarchical Bayesian (HB) combination (Dahlen et al. 2013; Carrasco Kind & Brunner 2014b) or Bayesian model combination/averaging (Carrasco Kind & Brunner 2014b). Both of these ensemble methods for photometric redshifts have been illustrated to improve estimates for normal galaxy populations, with the combined redshift PDF more accurate and less biased than any individual photo-z determination incorporated in the analysis.

To further improve the photometric redshifts for radio continuum sources we therefore combine the estimates from each template set through a hierarchical Bayesian combination.

5.1 Hierachical Bayesian combination of redshift PDFs

Following the method outlined in Dahlen et al. (2013), a consensus $P(z)$ is determined while accounting for the possibility that individual *measured* redshift probability distributions $P_m(z)_i$ are incorrect. The possibility that an individual $P(z)$ is incorrect is introduced as a nuisance parameter, f_{bad}, that is subsequently marginalised over.

Following Dahlen et al. (2013), we define for each redshift estimate, i,

$$P(z, f_{bad})_i = P(z|bad\ measurement)_i f_{bad} \\ + P(z|good\ measurement)_i (1 - f_{bad}), \quad (3)$$

where $P(z|bad\ measurement)$ ($U(z)$ hereafter for brevity) is the redshift probability distribution assumed in the case where the estimated $P_m(z)_i$ is incorrect and $P(z|good\ measurement) \equiv P_m(z)_i$ is the case where it is correct. The choice of $U(z)$ is explored in detail in the following section. For now, given a sensible choice of $U(z)$, the combined $P(z, f_{bad})$ for all n measurements is then given by

$$P(z, f_{bad}) = \prod_{i=1}^{n} P(z, f_{bad})_i^{1/\beta}, \quad (4)$$

where the additional hyper-parameter, β, is a constant that defines the degree of covariance between the different measurements. For completely independent estimates $\beta = 1$, while for estimates that are fully covariant $\beta = n$ (= 3 in this work). In this work we expect some reasonable degree
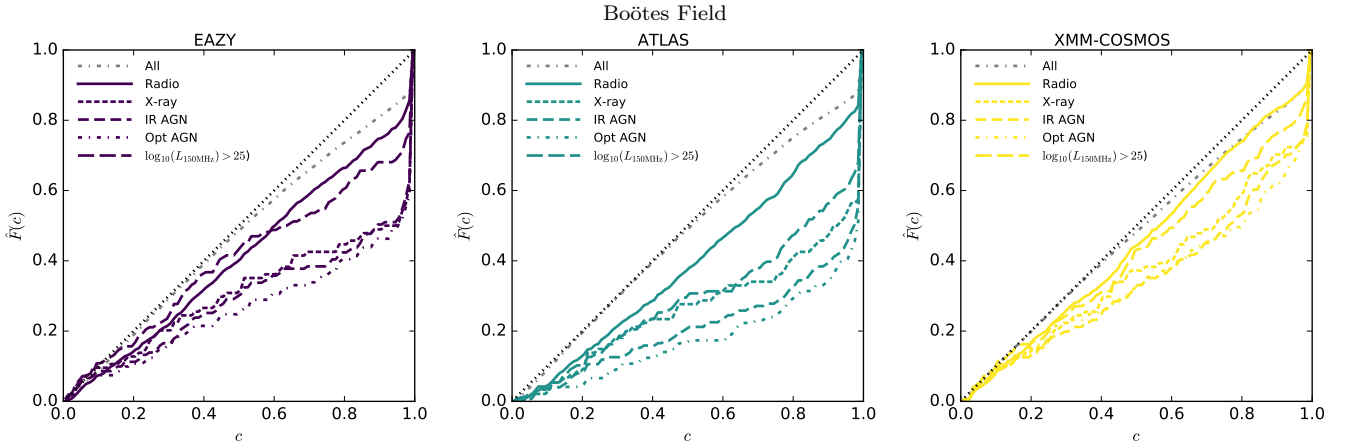
Boötes Field

Figure 10. Q-Q ($\hat{F}(c)$) plots for the redshift PDFs for each template set in the Boötes field after smoothing the raw EAZY PDFs to minimise the distance between the observed distribution and the ideal. Plotted are the cumulative distributions for all radio detected sources (solid lines) as well the subsets of the radio sample which are X-ray detected (short dashed lines), infrared AGN (medium dashed lines) and high power radio sources (long dashed lines).
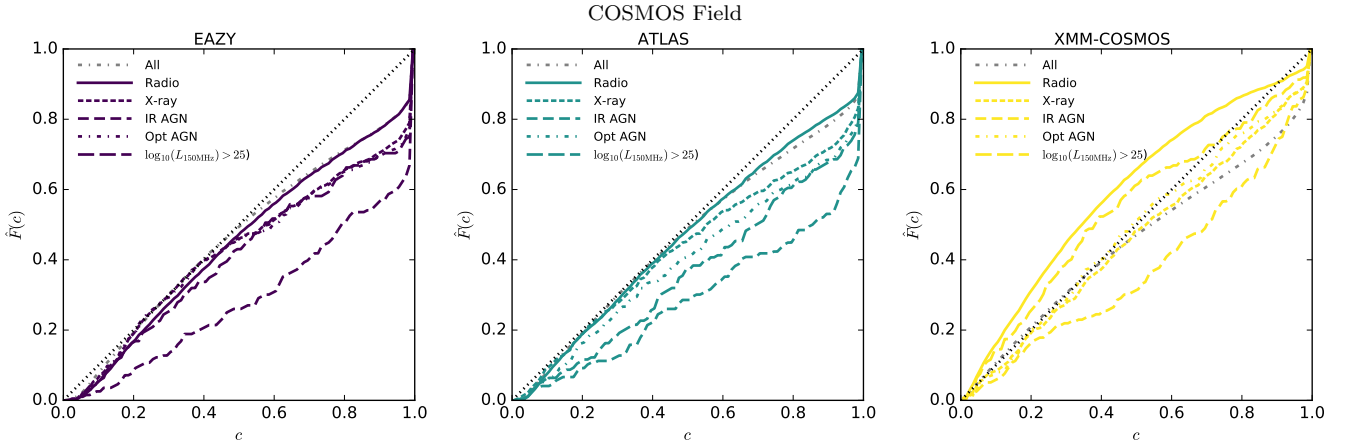
COSMOS Field

Figure 11. Q-Q ($\hat{F}(c)$) plots for the redshift PDFs for each template set in the COSMOS field after smoothing the raw EAZY PDFs to minimise the distance between the observed distribution and the ideal. Lines as described in Fig. 10.

of covariance between the three estimates as a result of the common photometric data and fitting algorithms used. Although the peak of the final redshift distribution is independent of β, changes in β do have an effect on the distribution widths. As part of the hierarchical Bayesian combination, β can also therefore be tuned such that posterior redshift distributions more accurately represent the redshift uncertainties.

Finally, we marginalise over $f_{\rm bad}$ to produce the consensus redshift probability distribution for each object

$$P(z) = \int_{f_{\rm bad}^{\rm min}}^{f_{\rm bad}^{\rm max}} P(z, f_{\rm bad}) df_{\rm bad}, \qquad (5)$$

where $f_{\rm bad}^{\rm min}$ and $f_{\rm bad}^{\rm max}$ are the lower and upper limits on the fraction of bad measurements. While fixed by definition to lie in the range $0 \leq f_{\rm bad} \leq 1$, the exact limits used when marginalising over $f_{\rm bad}$ can also be tuned using the training sample (Carrasco Kind & Brunner 2014b).

5.1.1 Assumptions for the $U(z)$ prior

During the calculation of the consensus redshift PDF, it is necessary to make an assumption on what the redshift prior is in the case where a given measurement is bad. The simplest assumption for $U(z)$ is that in cases where the measurement is bad, we have zero information on the redshift of a given object. Therefore, $U(z)$ is a flat prior, whereby $U(z) = 1/N$ for redshifts in the range of fitting $0 < z < N$.

As is discussed by Dahlen et al., we can also assume a more informative prior such as one which is proportional to the redshift dependent differential comoving volume, $dV(z)/dz$. Given the nature of the deep multi-wavelength surveys being used and the broad redshift range of interest, a volume prior increases the likelihood of sources being at higher redshifts and disfavours low-redshift solutions where $dV(z)/dz$ is very small.

Alternatively, as we adopt in our analysis, magnitude information for each source can also be incorporated through the use of an empirical or model-based magnitude prior (Benítez 2000). The benefit of incorporating magnitude de-

pendent redshift priors in template fitting has been well illustrated in the literature (Benítez 2000; Hildebrandt et al. 2012; Molino et al. 2014) and so the assumption of a magnitude dependent $U(z)$ is therefore well motivated.

Our empirical redshift prior is based on a modified version of the functional form outlined in Benítez (2000). Using subset of the spectroscopic training set, we fit the observed redshift - magnitude relation, $p(z|m_I)$, with the function:

$$p(z|m_I) \propto \left(c + \frac{dV(z)}{dz} \right) \times \exp \left\{ - \left[\frac{z}{z_m(m_I)} \right]^\alpha \right\}. \quad (6)$$

As in Benítez (2000), the prior distribution at high redshifts is determined by an exponential cut-off above a magnitude-dependent redshift z_m. However, rather than a linear dependence on m_I, we assume $z_m = z_0 + k_1 z + k_2 z^2$. Additionally, in place of the power-law term z^α, we use the differential comoving volume element $dV(z)/dz$. Following Hildebrandt et al. (2012), we also include the additional parameter, c, to allow for non-vanishing likelihood as $z \to 0$. We make the assumption that for AGN selected sources, $c = 0$, while for all other sources $c = 0.001$.

The parameters α, z_0, k_1 and k_2 are estimated by fitting the functional form to the desired subset of test galaxy samples using MCMC (Foreman-Mackey et al. 2013). Fig. 12 illustrates the resulting redshift priors as a function of I magnitude for 'normal' galaxies (top) and X-ray/IR/opt selected AGN (bottom). At bright apparent optical magnitudes there is a clear difference in the redshift distribution between the two source populations, with AGN sources having both a higher median redshift and a more extended tail at higher redshifts ($z > 3$).

5.1.2 *Tuning of hyper-parameters using spectroscopic sample*

In addition to the assumption of $U(z)$, it is also necessary to assume or fit the additional hyper-parameters, β, f_{bad}^{min} and f_{bad}^{max}. Our assumptions for f_{bad}^{min} and f_{bad}^{max} are based on the measured and expected outlier fractions for the relevant source populations. For non-AGN, we therefore marginalise f_{bad} over the range $0 < f_{bad} < 0.05$ while for the X-ray/IR sample, we marginalise over the range $0 < f_{bad} < 0.5$.

As discussed in the previous section, β can be tuned to maximise the accuracy of the resulting consensus $P(z)$ estimates. We therefore fit β using the spectroscopic training sample. Specifically, we find the β which minimises the distance between measured $\hat{F}(c)$ and the desired 1:1 relation within 80% HPD CI. The restriction of fitting only within the 80% HPD CI region is motivated by the observation by Wittman et al. (2016) that even for well calibrated photometric data, non-gaussian errors and uncalibrated template systematics can result in the tails of $P(z)$ distribution being poorly described. By restricting the optimisation to only $< 80\%$ HPD CI, we prevent over smoothing of the $P(z)$ due to these low probability tails.

5.2 Optimised photometric redshift properties

In Fig. 13 and 14 we illustrate the σ_{NMAD}, O_f and $\overline{CRPS}$ performance of the new consensus redshift estimates in each of the source population subsets (see also Table 5). For the

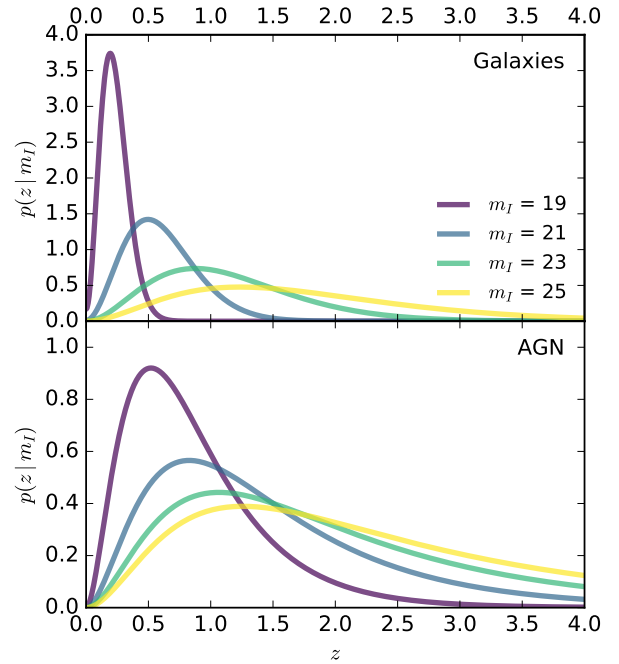**Figure 12.** Empirical redshift priors for 4 different *I*-band magnitudes based on the functional form in Equation 6. Top: redshift priors for sources which are not classified *a priori* as AGN based on X-ray or IR AGN criteria. Bottom: redshift priors for sources which are classified as either X-ray sources or IR AGN (see Section 2.3).

Table 5. Photometric redshift quality statistics for the derived combined consensus PDFs. The statistical metrics (see Table 2) are shown for the full spectroscopic sample, the radio detected sources and for various subsets of the radio population.

| Field	σ_f	σ_{NMAD}	Bias	O_f	σ_{O_f}	$\overline{CRPS}$
		All Sources				
Boötes	0.629	0.034	-0.003	0.090	0.048	0.178
COSMOS	0.329	0.021	-0.005	0.094	0.048	0.137
		All Radio Sources				
Boötes	0.442	0.034	-0.003	0.111	0.044	0.174
COSMOS	0.215	0.015	-0.002	0.077	0.042	0.098
	Radio Sources - Non X-ray/IR/Opt AGN					
Boötes	0.438	0.027	-0.001	0.020	0.040	0.104
COSMOS	0.200	0.014	-0.002	0.063	0.039	0.081
		Radio Sources - X-ray AGN				
Boötes	0.552	0.283	-0.023	0.500	0.072	0.444
COSMOS	0.219	0.030	0.002	0.145	0.055	0.178
		Radio Sources - IR AGN				
Boötes	0.509	0.323	-0.114	0.603	0.081	0.556
COSMOS	0.389	0.095	-0.001	0.241	0.071	0.293
		Radio Sources - Opt AGN				
Boötes	0.505	0.401	-0.108	0.595	0.071	0.607
COSMOS	0.173	0.038	-0.008	0.209	0.053	0.269
	Radio Sources - $\log_{10}(L_{150MHz}[W / Hz]) > 25$					
Boötes	0.398	0.122	-0.039	0.412	0.056	0.525
COSMOS	0.182	0.014	-0.001	0.088	0.041	0.105
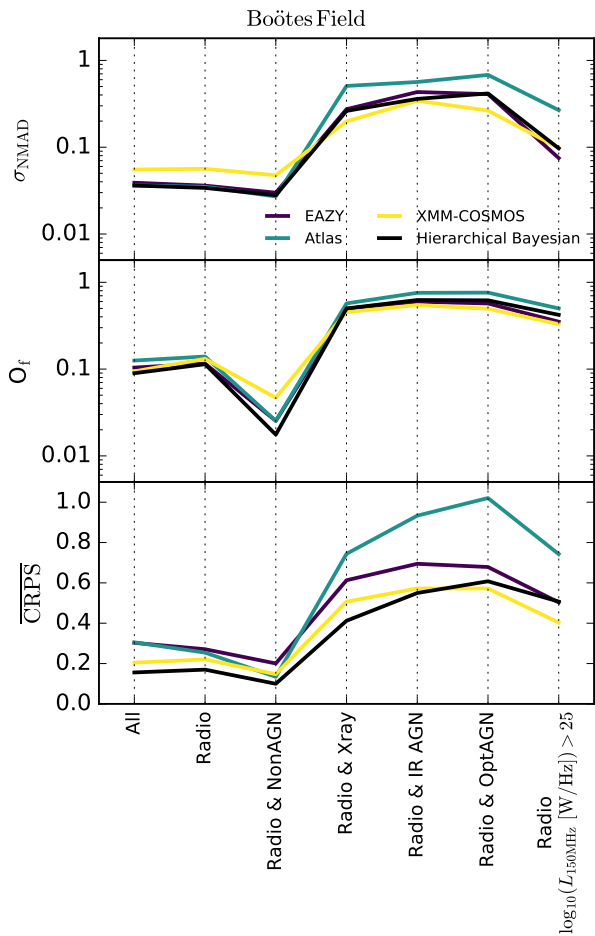
Figure 13. Visualised photometric redshift performance in three metrics (σ_{NMAD}, O_{f}, $\overline{\mathrm{CRPS}}$; see Table 2) for the different Boötes field radio source subsamples.

Figure 14. Visualised photometric redshift performance in three metrics (σ_{NMAD}, O_{f}, $\overline{\mathrm{CRPS}}$; see Table 2) for the different COS-MOS field radio source subsamples..

full spectroscopic samples and almost all subsets of the radio detected populations, the HB photo-z estimates either match the scatter and outlier fraction performance of the best individual template set (c.f. Table 3) to within 10% or outperform all of the estimates. The high performance of the HB photo-z is consistent across both data sets and substantially improves upon the individual template sets in several areas.

The improved performance of the consensus photo-zs also extends to the $P(z)$ distributions. The performance of the HB redshifts in the mean continuous ranked probability score ($\overline{\mathrm{CRPS}}$) is significantly better than any individual redshift estimate. In Figures 15 and 16 we show the $\hat{F}(c)$ distributions for the Boötes and COSMOS samples respectively. For both fields, not only is the overall $P(z)$ accuracy for the full spectroscopic redshift sample improved, but the $P(z)$ accuracy for the radio detected population (and all subsets) are also improved. Variation in $P(z)$ accuracy between the different radio subsets is significantly reduced.

Finally, we note that the average uncertainty on individual source for the HB photo-z estimates remain competitive with those of the best individual template set. In Fig. 17, we present the median 80% highest probability density con-
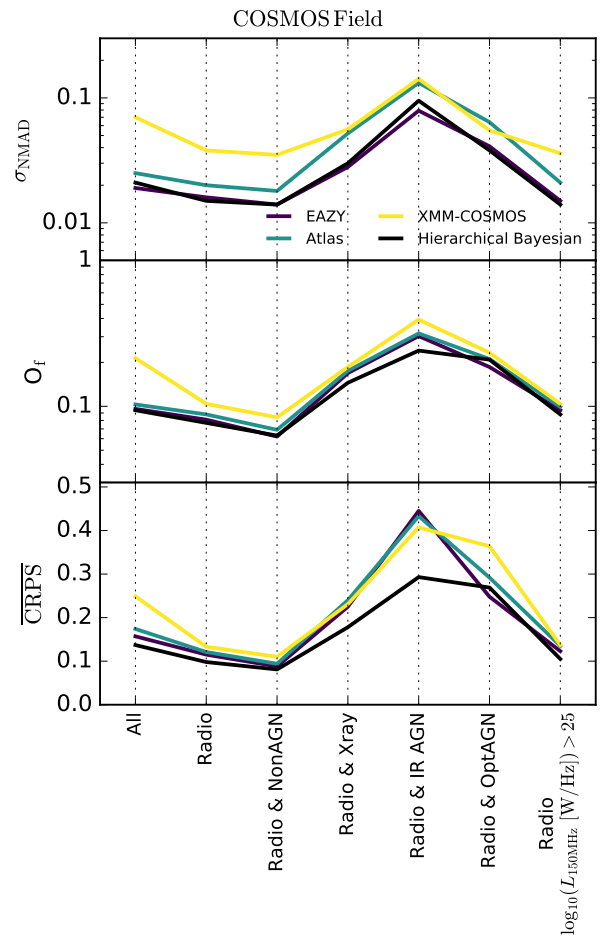
fidence intervals, Δ_{z_1}, around the primary redshift solution, z_1, as a function of $I(i^+)$ magnitude.

It is important to note here that the observed improvement in redshifts for the HB consensus photo-zs results primarily from the combining of multiple estimates and is not driven by the magnitude prior. When folding in the magnitude priors to each individual estimate, there is only a very minor improvement in the photo-zs (namely a small reduction in outlier fraction).

6 DISCUSSION

6.1 Radio surveys for studying galaxy and AGN evolution

In the preceding two sections we have presented a large amount of detailed analysis on photometric redshift estimates for two deep radio continuum surveys, including a wealth of statistics and comparisons that can be somewhat abstract. To interpret the results presented in these sections, it is worth revisiting the questions we specifically posed in the introduction. Firstly, how does the photometric redshift accuracy of radio sources vary as a function of redshift and radio luminosity? And secondly, do the current methods and
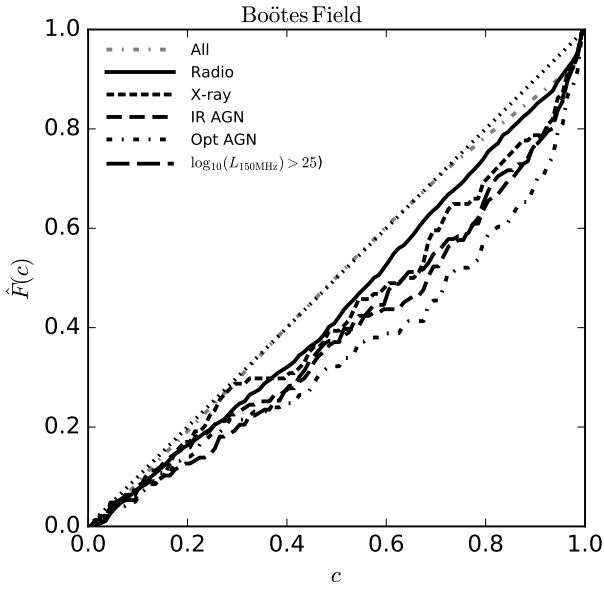
Figure 15. The $\hat{F}(c)$ plots for the consensus redshift PDF produced by the hierarchical Bayesian combination of the three different template sets. Plotted are the cumulative distributions for all LOFAR detected sources (solid line) as well the subsets of the radio sample which are X-ray detected (short dashed line), infrared AGN (medium dashed line) and high power radio sources (long dashed line).
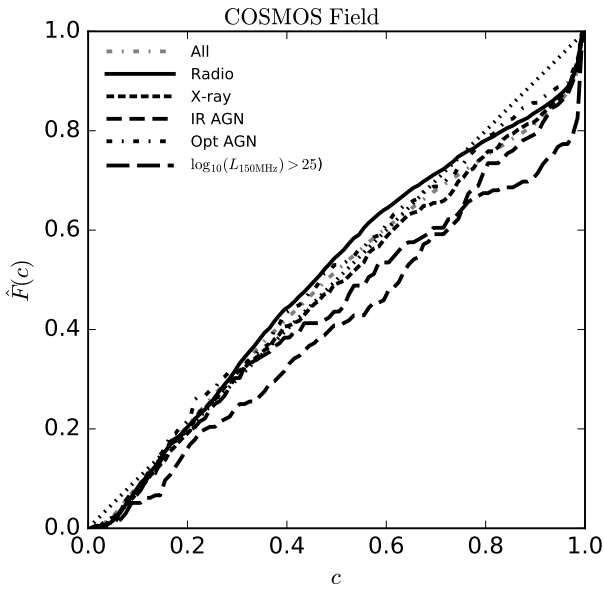

Figure 16. The $\hat{F}(c)$ plots for the consensus redshift PDF produced by the hierarchical Bayesian combination of the three different template sets. Plotted are the cumulative distributions for all spectroscopic sources (orange dash-dot line), all radio detected sources (solid line) as well the subsets of the radio sample which are X-ray detected (short dashed line), infrared AGN (medium dashed line) and high power radio sources (long dashed line).
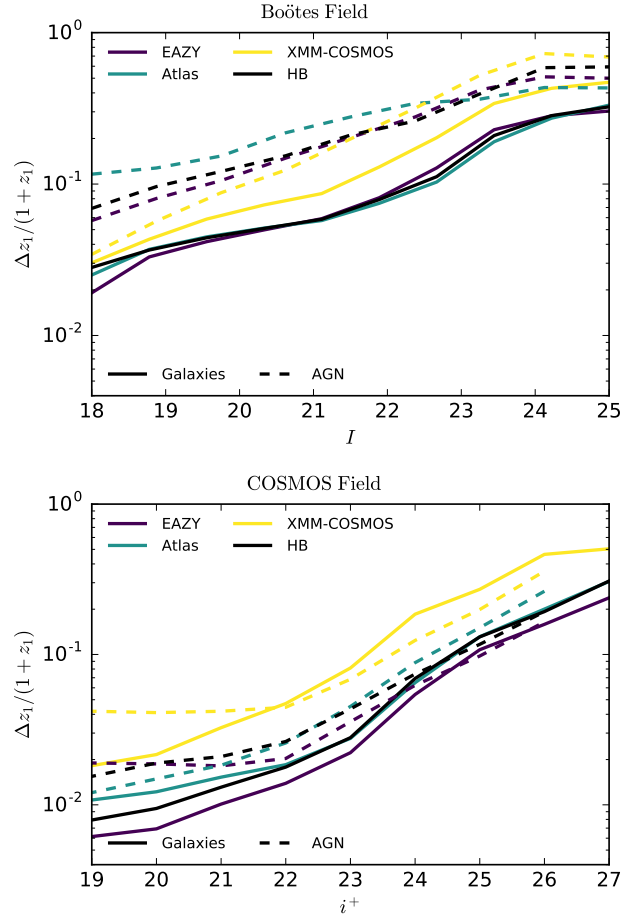

Figure 17. Median positive 80% highest probability density confidence intervals, Δ_{z_1}, above the primary redshift solution, z_1, as a function of $I(i^+)$ magnitude for both the Boötes (top) and COSMOS (bottom) fields. We illustrate only the upper error bounds to improve clarity by allowing a logarithmic scale. Within the primary peak, positive and negative errors are found to be very symmetrical; negative errors for each estimate follow the same magnitude trends.

optimization strategies developed for 'normal' galaxies or other AGN populations in optical surveys extend to radio selected galaxies?

The answer to the first question is partially provided in Section 4. Across a wide range in radio luminosity, the measured photo-z scatter remains approximately constant, regardless of which template set is used. In contrast, there is a much stronger evolution in the measured outlier fraction, which increases significantly between $23 < \log_{10}(L_{150\mathrm{MHz}}) < 27$. At the very faintest fluxes probed by the LOFAR and VLA data used in this analysis (i.e. sub-mJy), the photo-zs for the radio source population perform very well; comparable to the overall properties of the radio undetected source population.

Based on the results of our hierarchical Bayesian combination photo-z estimates presented in Section 5, our answer to the second question posed has to be a yes. Consensus redshifts from hierarchical Bayesian combination *do* improve the redshift accuracy for these populations whilst

sacrificing no accuracy for the non-radio population. The redshifts produced perform better than any individual template set/method for both the full spectroscopic sample and the radio detected population. The success of the ensemble redshifts is an excellent illustration that the techniques developed to provide marginal gains for the 'normal' galaxy population (Dahlen et al. 2013; Carrasco Kind & Brunner 2014b) can also provide very significant improvements for more diverse datasets.

While the consensus estimates do improve on the photo-z predictions for radio sources which also satisfy optical and IR AGN criteria (and to a lesser degree have strong X-ray flux), the overall quality of these estimates still remains very poor compared to those of the general radio source population. Salvato et al. (2008, 2011) illustrated that although the use AGN-dominated SEDs in the photometric redshift fitting can improve results for AGN photo-zs, additional steps are required to maximise the accuracy: namely strict magnitude priors based on optical morphology and corrections for variability in optical magnitudes between observations at different epochs. Such steps can only be taken in a very select number of fields (in the case of variability) or in very small areas with high-resolution optical imaging (for morphology selection), making it impractical to incorporate these steps in our photo-z strategy[3]. However, it is also these populations have been shown to benefit from empirical (or machine learning) photo-z estimates (e.g. Brodwin et al. 2006). In future, the expansion of the hierarchical Bayesian analysis to include more redshift estimates tailored for the difficult quasar populations should therefore offer further significant improvements.

Further improvements to photo-z estimates for the radio population will also be greatly aided by the forthcoming WEAVE-LOFAR spectroscopic survey (Smith et al. 2016). WEAVE-LOFAR will obtain $> 10^6$ spectra for radio sources from the LOFAR 150 MHz survey; providing spectroscopic redshifts and source classifications for an unprecedented number of radio sources. In particular, the deepest tier of the survey will target sources as faint as $S_{\nu,150} \approx 100\mu$Jy beam^{-1} over several deep fields. Such a sample will provide an extensive and unbiased training sample that can be used to improve photo-z estimates for the sub-mJy radio sources in the widest tier of the LOFAR survey and many others besides.

In addition to providing samples for photo-z evaluation/training which are more representative, significantly larger samples of spectroscopic redshifts will also offer potential improvements in the HB combination procedure outlined in this paper. By allowing for the calibration of photo-zs and tuning of hyper-parameters in smaller subsets (e.g. split into bins in several parameters; optical magnitude/redshift/radio power etc.), further gains to the consensus redshift accuracy and precision will be gained (Carrasco Kind & Brunner 2014b).

---

[3] As optical surveys with long-term variability measurements (e.g. PanSTARRs Medium Deep Survey or the Large Synoptic Survey Telescope (LSST Collaboration 2012)) reach the depths required for deep extragalactic surveys, such corrections for variability will become significantly easier to implement and offer significant photo-z improvements for some source types.

6.2 Radio surveys for cosmology

One of the key scientific capabilities provided by the next generation of radio interferometers like the Square Kilometre Array is as a tool for studying cosmology (Camera et al. 2012; Ferramacho et al. 2014; Jarvis et al. 2015). One such avenue for studying cosmology with radio surveys is through weak lensing (WL) experiments (Brown et al. 2015). Thanks to the different systematics offered by radio continuum observations (both in the intrinsic populations explored and the instrumentation), weak lensing studies with the SKA will be highly complementary to the increasingly powerful optical weak lensing studies planned for the next decades (e.g. EUCLID; Laureijs et al. 2011).

As with their optical counterparts, weak lensing studies with the SKA will be heavily reliant on photometric redshift estimates based on the all sky photometric data available. While extensive effort is being invested in reaching the redshift accuracy requirements for optical weak lensing experiments (Sanchez et al. 2014; Carrasco Kind & Brunner 2014b, e.g.), a key question for the SKA experiments is what effect does the radio selection have on the expected photo-z accuracies and biases?

The radio continuum depths explored in the study do not reach the faint fluxes expected for planned weak lensing studies outlined in Brown et al. (2015), so we cannot conclusively say what the full effects might be. Nevertheless, based on the available results for the brighter radio population it is relatively clear that the prognosis for SKA WL studies remains tied to that of the comparable optical studies.

The radio source population for which our photometric redshift estimates are at least as accurate are those radio sources which are also identified as luminous X-ray sources or host dust obscured AGN. In this regard the SKA WL experiments will be limited by the same source types as will effect the optical WL experiments. The radio source population which are *not* classified as likely AGN were found to have more accurate photometric redshift estimates than the population not detected by radio imaging. Provided the techniques and selection criteria developed for removing the bias of AGN in optical WL experiments can be applied equally to the SKA WL samples, the radio continuum selection should not present any critical problems. More studies will be required to test this once the SKA pathfinder and precursor surveys reach their full depths and large samples of un-biased spectroscopic redshifts are available.

7   SUMMARY

We have presented a study of template based photometric redshift accuracy for two samples of galaxies drawn from a wide area (NDWFS Boötes; Jannuzi & Dey 1999) and a deep (COSMOS; Laigle et al. 2016) survey field. We calculate photometric redshifts using three different galaxy template sets from the literature. The three template sets represent libraries which are either commonly used in photometric redshift estimates within the literature (Brammer et al. 2008; Salvato et al. 2008) or are designed to cover the broad range of SEDs observed in local galaxies (Brown et al. 2014).

Exploring the photometric redshift quality as a function of galaxy radio properties, we find:

• At low-redshift ($z < 1$), radio detected galaxies typically have better photo-z scatter and outlier fractions than galaxies with comparable magnitudes, redshifts and colours but are undetected in radio. However, as redshift increases, radio-detected galaxies perform worse than their radio undetected counterparts.

• Within a redshift range where photo-z quality remains relatively constant, the outlier fraction of all photo-z estimates increases towards the highest radio powers (and radio flux) while scatter remains roughly the same. This trend is independent of survey field and template set.

• Photo-zs for radio sources not identified as AGN through X-ray, optical or IR selection criteria perform comparably to radio un-detected sources at the same redshifts.

• Without additional calibration, the redshift PDFs for all three template sets are overconfident; producing error estimates which are significantly underestimated.

By combining all three photo-z estimates through hierarchical Bayesian combination (Dahlen et al. 2013; Carrasco Kind & Brunner 2014b) we are able to produce a new consensus estimate which outperforms any of the individual estimates which went into it. The consensus redshift estimates match or better the measured scatter or outlier fraction of the best individual estimate for most radio population subsets while also providing improved predictions on the redshift uncertainties. Nevertheless, while offering some improvement, the overall quality of photo-z estimates for radio sources which are X-ray sources or optical/IR AGN is still relatively poor; with high outlier fractions ($> 20\%$) and very large scatter ($\sigma_{\mathrm{NMAD}} > 0.2 \times (1 + z)$). Future work tailored to improving our photo-z estimates for IR/optically selected AGN will be required to achieve the some of the key science goals for deep radio continuum surveys. In the second paper in this series we will explore the improvements offered by photometric redshift estimates from gaussian processes (Almosallam et al. 2016b), both in isolation and when combined with the template based estimates through our hierarchical bayesian combination procedure.

REFERENCES

Alam S., et al., 2015, ApJS, 219, 12

Almosallam I. A., Lindsay S. N., Jarvis M. J., Roberts S. J., 2016a, MNRAS, 455, 2387

Almosallam I. A., Jarvis M. J., Roberts S. J., 2016b, MNRAS, 462, 726

Arnouts S., Cristiani S., Moscardini L., Matarrese S., Lucchin F., Fontana A., Giallongo E., 1999, MNRAS, 310, 540

Ashby M. L. N., et al., 2009, ApJ, 701, 428

Ball N. M., Brunner R. J., Myers A. D., Strand N. E., Alberts S. L., Tcheng D., 2008, ApJ, 683, 12

Bauer F. E., Alexander D. M., Brandt W. N., Schneider D. P., Treister E., Hornschemeier A. E., Garmire G. P., 2004, The Astronomical Journal, 128, 2048

Beck R., et al., 2017

Benítez N., 2000, ApJ, 536, 571

Bertin E., Arnouts S., 1996, Astron. Astrophys. Suppl. Ser., 117, 393

Bian F., et al., 2013, ApJ, 774, 28

Bolzonella M., Miralles J. M., Pello R., 2000, A&A, 363, 476

Bonzini M., Padovani P., Mainieri V., Kellermann K. I., Miller N., Rosati P., Tozzi P., Vattakunnel S., 2013, MNRAS, 436, 3759

Booth R. S., de Blok W. J. G., Jonas J. L., Fanaroff B., 2009, eprint arXiv:0910.2935, pp –

Bovy J., et al., 2012, ApJ, 749, 41

Brammer G. B., van Dokkum P. G., Coppi P., 2008, ApJ, 686, 1503

Brand K., et al., 2006, ApJ, 641, 140

Brescia M., Cavuoti S., D'Abrusco R., Longo G., Mercurio A., 2013, ApJ, 772, 140

Brodwin M., et al., 2006, ApJ, 651, 791

Brown T. A., 1974

Brown M. J. I., Dey A., Jannuzi B. T., Brand K., Benson A. J., Brodwin M., Croton D. J., Eisenhardt P. R., 2007, ApJ, 654, 858

Brown M. J. I., et al., 2008, ApJ, 682, 937

Brown M. J. I., et al., 2014, ApJS, 212, 18

Brown M., et al., 2015, Advancing Astrophysics with the Square Kilometre Array (AASKA14), pp –

Calistro Rivera G., et al., 2017, arXiv, p. arXiv:1704.06268

Calzetti D., Armus L., Bohlin R. C., Kinney A. L., Koornneef J., Storchi-Bergmann T., 2000, ApJ, 533, 682

Camera S., Santos M. G., Bacon D. J., Jarvis M. J., Mcalpine K., Norris R. P., Raccanelli A., Röttgering H., 2012, MNRAS, 427, 2079

Capak P., et al., 2007, ApJS, 172, 99

Carrasco Kind M., Brunner R. J., 2013, MNRAS, 432, 1483

Carrasco Kind M., Brunner R. J., 2014a, MNRAS, 438, 3409

Carrasco Kind M., Brunner R. J., 2014b, MNRAS, 442, 3380

Chung S. M., et al., 2014, ApJ, 790, 54

Civano F., et al., 2016, ApJ, 819, 62

Collister A. A., Lahav O., 2004, 116, 345

Cool R. J., 2007, ApJS, 169, 21

Dahlen T., et al., 2013, ApJ, 775, 93

Dark Energy Survey Collaboration T., 2005, arXiv, p. arXiv:0510346

Dey A., Lee K.-S., Reddy N., Cooper M., Inami H., Hong S., Gonzalez A. H., Jannuzi B. T., 2016, ApJ, 823, 11

Donley J. L., et al., 2012, ApJ, 748, 142

Fazio G. G., et al., 2004, ApJS, 154, 10

Ferramacho L. D., Santos M. G., Jarvis M. J., Camera S., 2014, MNRAS, 442, 2511

Flesch E. W., 2015, Publ. Astron. Soc. Aust., 32, e010

Foreman-Mackey D., Hogg D. W., Lang D., Goodman J., 2013, 125, 306

Fotopoulou S., et al., 2016, A&A, 592, A5

Gonzalez A. H., et al., 2010, American Astronomical Society Meeting Abstracts #216, 216

Grogin N. A., et al., 2011, ApJS, 197, 35

Hersbach H., Hersbach H., 2000, Weather and Forecasting

Hickox R. C., et al., 2009, ApJ, 696, 891

Hildebrandt H., et al., 2010, A&A, 523, A31

Hildebrandt H., et al., 2012, MNRAS, 421, 2355

Hsu L.-T., et al., 2014, ApJ, 796, 60

Ilbert O., et al., 2005, A&A, 439, 863

Jannuzi B. T., Dey A., 1999, The Young Universe: Galaxy Formation and Evolution at Intermediate and High Redshift, 191, 111

Jarvis M. J., 2012, African Skies, 16, 44

Jarvis M. J., Rawlings S., 2004, New Astronomy Reviews, 48, 1173

Jarvis M., Bacon D., Blake C., Brown M., Lindsay S., Raccanelli A., Santos M., Schwarz D. J., 2015, in Proceedings of Advancing Astrophysics with the Square Kilometre Array (AASKA14). 9 -13 June. p. 18

Jiang C., Jing Y., Han J., 2013, eprint arXiv:1307.3322

Johnston S., 2007, arXiv, pp 174–188

Kenter A., et al., 2005, ApJS, 161, 9

Kochanek C. S., et al., 2012, ApJS, 200, 8

LSST Collaboration 2012, arXiv, p. arXiv:1211.0310

Laigle C., et al., 2016, ApJS, 224, 24

Laureijs R., et al., 2011, arXiv, p. arXiv:1110.3193

Lee K.-S., Alberts S., Atlee D., Dey A., Pope A., Jannuzi B. T., Reddy N., Brown M. J. I., 2012, The Astrophysical Journal Letters, 758, L31

Lee K.-S., Dey A., Cooper M. C., Reddy N., Jannuzi B. T., 2013, ApJ, 771, 25

Lee K.-S., Dey A., Hong S., Reddy N., Wilson C., Jannuzi B. T., Inami H., Gonzalez A. H., 2014, ApJ, 796, 126

Marchesi S., et al., 2016a, ApJ, 817, 34

Marchesi S., et al., 2016b, ApJ, 830, 100

Matheson J. E., Winkler R. L., 1976, Management Science, 22, 1087

McCracken H. J., et al., 2012, A&A, 544, A156

Molino A., et al., 2014, MNRAS, 441, 2891

Norris R. P., et al., 2013, Publ. Astron. Soc. Aust., 30, e020

Oke J. B., Gunn J. E., 1983, ApJ, 266, 713

Padovani P., 2016, Astron Astrophys Rev, 24, 13

Pei Y. C., 1992, ApJ, 395, 130

Polletta M., et al., 2007, ApJ, 663, 81

Polsterer K. L., D'Isanto A., Gieseke F., 2016, arXiv, p. arXiv:1608.08016

Quadri R. F., Williams R. J., 2010, ApJ, 725, 794

Richards G. T., et al., 2001, The Astronomical Journal, 122, 1151

Rottgering H. J. A., 2010, in "Proceedings of the ISKAF2010 Science Meeting. June 10 -14 2010. Assen. pp 50–

Salvato M., et al., 2008, ApJ, 690, 1250

Salvato M., et al., 2011, ApJ, 742, 61

Sanchez C., et al., 2014, MNRAS, 445, 1482

Shimwell T. W., et al., 2017, A&A, 598, A104

Smith D. J. B., et al., 2016, eprint arXiv:161102706, pp –

Smolčić V., et al., 2017a, arXiv, p. arXiv:1703.09713

Smolčić V., et al., 2017b, arXiv, p. arXiv:1703.09719

Stanford S. A., et al., 2012, ApJ, 753, 164

Taniguchi Y., et al., 2015, Publ. Astron. Soc. Jpn, 67, 104

Vaccari M., 2016, The Universe of Digital Sky Surveys, 42, 71

Weinstein M. A., et al., 2004, ApJS, 155, 243

Williams W. L., et al., 2016, MNRAS, 460, 2385

Wittman D., Bhaskar R., Tobin R., 2016, MNRAS, 457, 4005

York D. G., et al., 2000, The Astronomical Journal, 120, 1579

Zeimann G. R., et al., 2012, ApJ, 756, 115

Zeimann G. R., et al., 2013, ApJ, 779, 137

Zhang Y., Ma H., Peng N., Zhao Y., Wu X.-b., 2013, The Astronomical Journal, 146, 22

da Cunha E., Charlot S., Elbaz D., 2008, MNRAS, 388, 1595

van Haarlem M. P., et al., 2013, A&A, 556, A2

This paper has been typeset from a TEX/LATEX file prepared by the author.