

## Photometric redshifts for the next generation of deep radio continuum surveys - II. Gaussian processes and hybrid estimates

Duncan, K.J.; Jarvis, M.J.; Brown, M.J.I.; Röttgering, H.J.A.

## Citation

Duncan, K. J., Jarvis, M. J., Brown, M. J. I., & Röttgering, H. J. A. (2018). Photometric redshifts for the next generation of deep radio continuum surveys - II. Gaussian processes and hybrid estimates. *Monthly Notices Of The Royal Astronomical Society*, 477(4), 5177-5190. doi:10.1093/mnras/sty940

Version:Not Applicable (or Unknown)License:Leiden University Non-exclusive licenseDownloaded from:https://hdl.handle.net/1887/68562

Note: To cite this publication please use the final published version (if applicable).

## Photometric redshifts for the next generation of deep radio continuum surveys - II. Gaussian processes and hybrid estimates

Kenneth J Duncan<sup>1\*</sup>, Matt J. Jarvis<sup>2,3</sup>, Michael J. I. Brown<sup>4,5</sup>, Huub J. A. Röttgering<sup>1</sup>

<sup>1</sup>Leiden Observatory, Leiden University, NL-2300 RA Leiden, Netherlands

<sup>2</sup>Astrophysics. University of Oxford, Denys Wilkinson Building, Keble Road, Oxford, OX1 3RH

<sup>3</sup>Physics and Astronomy Department, University of the Western Cape, Bellville 7535, South Africa

<sup>4</sup>School of Physics and Astronomy, Monash University, Clayton, Victoria 3800, Australia

<sup>5</sup>Monash Centre for Astrophysics, Monash University, Clayton, Victoria, 3800, Australia

10 November 2018

#### ABSTRACT

Building on the first paper in this series (Duncan et al. 2018), we present a study investigating the performance of Gaussian process photometric redshift (photo-z) estimates for galaxies and active galactic nuclei detected in deep radio continuum surveys. A Gaussian process redshift code is used to produce photo-z estimates targeting specific subsets of both the AGN population - infrared, X-ray and optically selected AGN and the general galaxy population. The new estimates for the AGN population are found to perform significantly better at z > 1 than the template-based photo-z estimates presented in our previous study. Our new photo-z estimates are then combined with template estimates through hierarchical Bayesian combination to produce a hybrid consensus estimate that outperforms either of the individual methods across all source types. Photo-z estimates for radio sources that are X-ray sources or optical/IR AGN are significantly improved in comparison to previous template-only estimates, with outlier fractions and robust scatter reduced by up to a factor of  $\sim 4$ . The ability of our method to combine the strengths of the two input photo-z techniques and the large improvements we observe illustrate its potential for enabling future exploitation of deep radio continuum surveys for both the study of galaxy and black hole co-evolution and for cosmological studies.

Key words: galaxies: distances and redshifts – galaxies: active – radio continuum: galaxies

#### INTRODUCTION

Photometric redshifts (photo-zs hereafter) have become a fundamental tool for both the study of galaxy evolution and for modern cosmology experiments. The main driving force behind recent developments in photometric redshift estimation methodology has been the stringent requirements set by the coming generation of weak-lensing cosmology experiments (e.g. EUCLID; Laureijs et al. 2011). However, the need for accurate and unbiased redshift estimates for large samples of galaxies ( $\approx 10^6 - 10^9$ ) represents a near universal requirement for all future extra-galactic surveys.

Through either template based (e.g. Arnouts et al. 1999;

Bolzonella et al. 2000; Benítez 2000; Brammer et al. 2008) or empirical/'machine learning' (e.g. Collister & Lahav 2004; Carrasco Kind & Brunner 2013, 2014a) estimation techniques, it is now possible to produce the precise and reliable photometric redshifts required for optically selected galaxy samples (Bordoloi et al. 2010; Sanchez et al. 2014; Carrasco Kind & Brunner 2014b; Drlica-Wagner et al. 2017). However, typically such methods are applied to, or optimised for, the galaxy emission due to stellar populations, with galaxies dominated by emission from active galactic nuclei (AGN) either removed from the analysis (where possible) or not explicitly accounted for. This therefore presents a problem in surveys where a larger fraction of the population is composed of AGN, for example in radio-continuum selected surveys (and for the  $\sim$  3 million X-ray selected AGN and

<sup>\*</sup> E-mail: duncan@strw.leidenuniv.nl

QSOs observed by the *eRosita* mission Merloni et al. 2012). The population of radio detected sources is extremely diverse - with radio emission tracing both black hole accretion in AGN and star formation activity.

Probing to unprecedented depths, deep radio continuum surveys from MeerKAT (Booth et al. 2009), the Australian SKA Pathfinder (ASKAP; Johnston et al. 2007) and the Low Frequency Array (LOFAR; van Haarlem et al. 2013) 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; Norris et al. 2013; Jarvis et al. 2017). Accurate and unbiased photometric redshift estimates for the full radio source population will be essential for studying the faint radio population and achieving the scientific goals of these deep radio continuum surveys - both for galaxy evolution and cosmological studies.

In Duncan et al. (2018, hereafter Paper I), we investigated the performance of template-based photometric redshift estimates for the radio-continuum detected population over a wide range of optical and radio properties. Specifically, three photometric redshift template sets, representative of those available in the literature, were applied to two optical/IR datasets and their performance investigated as a function of redshift, radio flux/luminosity and infrared/Xray properties.

Furthermore, by combining all three photo-*z* estimates through hierarchical Bayesian combination (Dahlen et al. 2013; Carrasco Kind & Brunner 2014b) we were able to produce a new consensus estimate that outperforms any of the individual estimates that went into it. Although the consensus redshift estimates were found to offer some improvement, the overall quality of template photo-*z* estimates for radio sources that are X-ray sources or optical/IR AGN was still relatively poor; with outlier fractions and scatter relative to the spectroscopic training sample that are unacceptable for many science goals. An alternative methodology is therefore needed to either replace the template-based photo-*z* estimates for these difficult populations or help to improve the consensus estimate.

Thankfully, empirical (or machine learning) photo-z estimates have already been shown to offer a potential solution for improving photo-zs for the AGN population (e.g. Richards et al. 2001; Brodwin et al. 2006; Bovy et al. 2012). In this paper we investigate how such machine learning photo-z techniques perform when applied to the same samples and data where template-based methods were found to struggle the most in Paper I. Specifically, we explore the use of Gaussian processes (GP) using the framework presented by Almosallam et al. (2016a,b, GPz). GPz offers several key advantages that make it an ideal choice for tackling the problems posed by large samples of radio selected galaxies. Firstly, it has been shown to outperform other empirical photo-z tools in the literature when applied to sparse datasets. Secondly, it incorporates cost-sensitive learning, i.e. the ability to give more or less weight to certain sources during the optimisation procedure. These additional weights potentially allow for biases in the available training sample to be accounted for. Finally, by modelling the non-uniform noise intrinsic in photometric datasets it offers estimations of the variance on the predicted photo-zs - meaning that its

outputs can also be easily incorporated into the hierarchical Bayesian combination framework presented in Paper I.

This paper is organized as follows: Section 2 presents the data used in this study along with details of how the Gaussian process photometric redshift framework of Almosallam et al. (2016a). Section 3 then outlines the application of the GPz framework to photometric data from deep survey fields such as those explored in Paper I and the improvements that can be made in photometric redshift quality for the most difficult radio source populations. In Section 4, we present the results of incorporating the new GP photo-zs within the Bayesian combination framework presented in Paper I. Finally, Section 5 presents a brief summary of the results in this paper and the key conclusions that can be drawn.

Throughout this paper, all magnitudes are quoted in the AB system (Oke & Gunn 1983) unless otherwise stated. We also assume a  $\Lambda$ -CDM cosmology with  $H_0 = 70$  kms<sup>-1</sup>Mpc<sup>-1</sup>,  $\Omega_m = 0.3$  and  $\Omega_{\Lambda} = 0.7$ .

#### 2 PHOTOMETRIC REDSHIFT METHODOLOGY

#### 2.1 Data

In Paper I we made use of two samples of galaxies drawn from both a wide area survey (NDWFS Boötes; Jannuzi & Dey 1999) and a smaller but deeper survey field (COSMOS; Laigle et al. 2016). In this paper we will make use of just the 'Wide' field sample in our subsequent analysis. The reasons for this are two-fold: Firstly, the optical filter coverage and depth of the available photometry in the field is more representative of the large survey fields that are being observed with deep radio continuum surveys such as LOFAR. The exceptional wavelength coverage and depth of the COS-MOS field photometry could give misleading expectations of the photo-z accuracy when the method is applied to other fields. Secondly, the targeted selection criteria of the AGN and Galaxy Evolution Survey (AGES; Kochanek et al. 2012) spectroscopic survey in the field results in a larger sample of AGN sources (see Fig.1 of Paper I) for training and testing the GP redshift estimates.

We refer the reader to Paper I and references therein for full details on the photometric catalog itself, along with details on the spectroscopic redshift information available in the field. As in Paper I, the radio continuum observations from this field are taken from the LOFAR observations presented in Williams et al. (2016). Details of the crossmatching procedure between the radio data and the optical catalog used in this work can be found in Williams et al. (2017).

Given its importance in the subsequent analysis it is worth summarising the multi-wavelength AGN classifications applied to the data. 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).

• X-ray AGN in the Boötes field were identified by cross-matching the positions of sources in our catalog with the X-Böotes Chandra survey of NDWFS (Kenter et al. 2005). We calculate the x-ray-to-optical flux ratio, X/O =



**Figure 1.** Multi-wavelength classifications of the sources in the full spectroscopic redshift sample for the Boötes dataset used in this study. 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 2.1). As illustrated in previous studies, the X-ray, IR AGN and radio source population are largely distinct populations with only partial overlap.

 $\log_{10}(f_X/f_{opt})$ , based on the *I* band magnitude following Brand et al. (2006) and for a source 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).

• Optical AGN were also identified through crossmatching the optical catalog with the Million Quasar Catalog compilation of optical AGN, primarily based on SDSS (Alam et al. 2015) and other literature catalogs (Flesch 2015).

Note however, these classifications are not expected to be distinct physical classifications but rather selection methods through which a wide variety of the most luminous AGN can be identified. Depending on data available in a given field, further sub-classifications or alternative criteria might be warranted. As shown in Fig. 1, there is significant overlap between different selection criteria with the majority of radio sources selected as AGN belonging to at least two of the subsets. Despite these overlaps, there is also potentially a very wide variety intrinsic spectral energy distributions within the full AGN sample, both between these subsets of AGN and within the subsets themselves.

As in Paper I, spectroscopic redshifts for sources in Boötes are taken from a compilation of observations within the field comprising primarily of the results of AGN and Galaxy Evolution Survey (AGES; Kochanek et al. 2012) spectroscopic survey, with additional redshifts provided by a large number of smaller surveys in the field including Lee et al. (2012, 2013, 2014), Stanford et al. (2012), Zeimann et al. (2012, 2013) and Dey et al. (2016).

In total, the combined sample consists of 22830 redshifts over the range 0 < z < 6.12, with 88% of these at z < 1. Due to the nature of the AGES target selection criteria, identified AGN sources have a higher degree of spectroscopic completeness than the general galaxy population ( $\approx 11\%$ of AGN have spectroscopic redshifts available compared to  $\approx 1\%$  of the rest of the galaxy population). Nevertheless, as is the case in most spectroscopic training samples the available sources do not necessarily sample the full photometric colour space. In the following section we present the weighting strategy employed to minimise the potential effects caused by the biased training sample. The limitations of the training sample and ways in which this can be mitigated in the future will also be revisited in Section 4.3.

### 3 GAUSSIAN PROCESS PHOTOMETRIC REDSHIFTS FOR AGN IN DEEP FIELDS

#### 3.1 GPz Method

Detailed descriptions of the theoretical background and methodology of GPz are presented in Almosallam et al. (2016a) and Almosallam et al. (2016b). In this section, we therefore outline only the details of how GPz was applied to our dataset.

Although the three different AGN selection criteria outlined in Section 2.1 contain significant overlap in their populations, we choose to train and calibrate the GP estimates of each subset seperately. Due to both inhomogeneity in the coverage of different filters and the relatively shallow depth of some of these observations, only a small fraction of sources are detected in all of the filters available in the field. For example, only  $\approx 9\%$  of the full photometric catalog has magnitude values available in the 13-bands extending from u-band to IRAC  $8\mu$ m. The number and combination of magnitudes input to GPz for each subset were therefore chosen to cover as broad a wavelength as possible whilst trying to ensure as many sources as possible were detected in the corresponding bands. Starting with the detection band of the multi-wavelength catalog (I), additional filter choices were added and the fraction of sources with magnitudes available in those filters calculated until the fraction fell to  $\sim 80\%$ . For cases where several different filter combinations offer a similar number of available sources, the combination that produces the best estimates in limited trials is chosen. We note however that systematic searches for the best filter combinations have not been performed. We also note that an extension to GPz is being developed to account for missing data in a fully consistent way (Almosallam et al. in prep) such that these issues will be further minimised in future.

The resulting filter selections and the sizes of the corresponding training samples are as follows:

• Infrared AGN – For the subset of IR AGN, the input dataset includes the optical R and I magnitudes in addition to the four IRAC magnitudes used in the colour selection of the subset. In the spectroscopic training set and full photometric IR AGN subsets, 98.9% and 82.6% of sources respectively have magnitudes in these bands. Of the 1751 spectroscopic sources classified as IR AGN, the final training, validation and test samples therefore consist of 1385, 173 and 173 sources respectively.

• X-ray AGN – The final filter choice for the X-ray AGN sources is  $B_w$ , R, I,  $K_s$  and Spitzer/IRAC 3.6 and  $4.5\mu$ m. Detection fractions in the spectroscopic and full photometric samples are almost identical to the IR AGN subset, with fractions of 98.8% and 82.7% respectively. There are 1133 spectroscopic sources classified as X-ray AGN, resulting in training, validation and test samples of 895, 112 and 112 respectively.

• Optical AGN – Although optically bright by definition, the chosen filter selection for the optical AGN subset consists of I in combination with the near and mid-infrared bands of J, Ks, Spitzer/IRAC 3.6/4.5 $\mu$ m and Spitzer/MIPS 24 $\mu$ m. In these filters, the available training and full sample fractions are 96.6% and 84.2% respectively. For the 1382 optical AGN sources in the spectroscopic training sample, this results in 1067, 134 and 134 sources in the training, validation and test samples.

In addition to the three GPz estimators targeted at subsets of the AGN population, we also produce an additional estimator trained on optical sources that do not satisfy any of the AGN selection criteria - corresponding to the significant majority of both the training sample and photometric catalog. As illustrated in the bottom panel of Fig. 2 (dashed blue line), the magnitude distribution for the full 'galaxy' sample extends to significantly fainter magnitudes than those in the AGN subsets. To find the optimum combination of optical bands we systematically calculated the fraction of sources with measured magnitudes in every possible combination of five bands out of those available in the field. The two sets of filters that would allow estimates for the largest fraction of catalog sources are  $\{u, B_w, R, I, z\}$ and  $\{B_w, R, I, z, 3.6\mu m\}$ , with 38.3% and 34.2% of the full photometric catalog respectively (87.3% and 92.8% of the training samples).

In all four cases, GPz was trained using 25 basis functions and allowing variable covariances for each basis function (i.e. the 'GPVC' of Almosallam et al. 2016a). We choose these parameters based on the tests of Almosallam et al. (2016a) who found minimal performance gain above 25 basis functions and significant improvements when using fully variable covariances compared to other assumptions. Finally, we also follow the practices in outlined Section 6.2 of Almosallam et al. (2016a) by pre-processing the input data to normalise the data and de-correlate the features (also known as 'sphering' or 'whitening').

#### 3.2 Weighting scheme

One of the key advantages offered by GPz with respect to some other empirical methods in the literature is its option of using cost-sensitive learning; allowing for potential biases in the training sample to be taken into account or certain regions of parameter space to be prioritised if desired. In this work we make use of two different weighting schemes. As a reference we first employ a flat weighting scheme (i.e. the 'Normal' weighting of Almosallam et al. 2016a). Secondly, we employ a weighting scheme that takes into account the colour and magnitude distribution of the training sample with respect to the full corresponding photometric sample.

Our colour based weighting scheme is based on the method presented in Lima et al. (2008) and successfully employed elsewhere in the photoz literature (e.g. Sanchez et al. 2014). Firstly, for all galaxies in the spectroscopic training set and the photometric sample we construct separate arrays consisting of the normalised distribution of I-band magnitudes and two photometric colours. The colour and magnitude distributions are both normalised based on the 99th percentile range observed in the full photometric sample. This renormalisation ensures that each observable is given equal importance in the subsequent weighting scheme and that the distribution is not severely affected by outliers.

Next, for each galaxy, i, in the spectroscopic training set, we compute the distance to the 9th nearest neighbour,  $r_{i,9}$ , in the colour-magnitude space of the training set<sup>1</sup> We



Figure 2. Illustration of the colour-magnitude based weighting scheme applied to each the training subsets employed in this work. In each plot, the dashed blue line shows the magnitude distributions for the full photometric sample while the thin black and thick gold lines show the training sample before and after weighting. The optical/infrared filter corresponding to each magnitude distribution is labelled in the upper right corner of each plot - 'ch1' and 'ch2' correspond to the Spitzer/IRAC  $3.6\mu$ m and  $4.5\mu$ m filters respectively.

then find the corresponding number of objects,  $N_{\rm P}(\mathbf{m}_i)$ , in the full photometric sample that fall within a volume with radius equal to  $r_{i,9}$ . The weight for a given training galaxy,

more localisation in the colour-magnitude space than the 16th nearest neighbour chosen in Lima et al. (2008) while still minimising the effects of small-number statistics. However, as illustrated by the minimal effect on results for 4 < n < 64 (Lima et al. 2008), we do not expect this choice to have any significant effect on the results presented.

<sup>&</sup>lt;sup>1</sup> The 9th nearest neighbour was chosen to provide marginally

 $W_i$ , is then defined following Equation 24 of Lima et al. (2008) such that

$$W_i = \frac{1}{N_{\rm P,tot}} \frac{N_{\rm P}(\mathbf{m}_i)}{N_{\rm T}(\mathbf{m}_i)},\tag{1}$$

where  $N_{\rm T}(\mathbf{m}_i)$  is the number of objects in the training sample within the same volume (by definition 8 in this work) and  $N_{\rm P,tot}$  the total number of objects in the photometric training sample. Finally, any training-set object with zero weight is removed from the sample and the weights renormalised such the  $\sum_i W_i = 1$  to meet the convention required by GPz.

In Fig. 2 we illustrate the results of this weighting scheme for each of the training sample subsets used in our analysis. For the three magnitudes used in the weighting scheme, Fig. 2 shows the magnitude distribution of the full photometric sample compared to that of the training sample before and after the weighting scheme has been applied.

The bias within the training sample is clearly strongest for both the IR AGN and normal galaxy populations, with the majority of training galaxies significantly brighter than those in the full photometric samples. In both cases, the weighting scheme does a good job of reproducing the distribution of the full photometric sample. However, as there are very few spectroscopic redshifts available at the very faintest optical magnitudes, the weighted training sample becomes somewhat noisy due to the small number of faint training objects being assigned high weights. Possible methods of minimising the effects of very small samples of faint training objects will be discussed further in Section 4.3.

#### 3.3 GPz photo-z Results

In Fig. 3 we present the results of our two GPz photo-z estimates in comparison to the consensus estimates produced through template-fitting in Paper I. In each set of figures we show the distribution of photo-z vs spectroscopic redshift for the consensus template estimates from Paper I, left, the GPz estimate with no weighting included in the cost-sensitive learning (centre) and the GPz estimate incorporating the colour and magnitude dependent weights as presented in Section 3.2 (right). The sample plotted in each row contains only the subset of test sources not included in the training of the GPz classifiers. In Table 1 we also present a subset the corresponding photo-z quality metrics (defined in Table 2) for each of the AGN/galaxy subsamples.

Visually, the poor performance of the template estimates for AGN populations between  $1 \leq z \leq 3$  is clear in the left-hand column of Fig. 3. Within this spectroscopic redshift range, many AGN sources are erroneously pushed towards  $z \sim 2$ , albeit with large uncertainties that keep the photo-z estimate within error of the true estimate. Alternatively, sources at  $1 \leq z \leq 3$  can have template estimates that are catastrophic failures, leading to estimated redshifts at  $z \ll 1$ .

Statistically, the overall improvement offered by the GPz estimates is illustrated in the reduction in scatter for the IR and optically selected AGN samples by a factor of two. The improvement in scatter for the X-ray selected AGN subset is less drastic but still very significant - again most noticeably at z > 1. As noted by Salvato et al. (2008, 2011) many X-ray selected AGN are more accurately described

**Table 1.** 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.

Estimate	$\sigma_{ m NMAD}$	Bias	$\mathcal{O}_f$	
IR AGN				
Template consensus	0.2429	0.0159	0.4425	
GPz - Unweighted	0.1431	-0.0187	0.2184	
GPz - Weighted	0.1183	-0.0072	0.1494	
X-ray AGN				
Template consensus	0.1067	0.0185	0.3214	
GPz - Unweighted	0.1241	0.0090	0.1339	
GPz - Weighted	0.0882	0.0090	0.0893	
Optical AGN				
Template consensus	0.2351	0.0169	0.4552	
GPz - Unweighted	0.1280	0.0195	0.1970	
GPz - Weighted	0.1147	0.0084	0.2313	
Galaxies				
Template consensus	0.0287	-0.0037	0.0416	
GPz - Unweighted	0.0323	0.0038	0.0220	
GPz - Weighted	0.0343	0.0033	0.0265	

by purely stellar SEDs - the template based photo-zs may therefore be expected to perform better for this subset than for the IR or optical AGN population. Improvement in the measured outlier fractions is consistent across all three subsets, with the outlier fraction,  $O_f$  (Table 2), measured for the GPz estimates typically a factor of two lower.

When applied to the remaining majority of galaxies that do not satisfy any of our AGN selection criteria, GPz is not able to significantly improve upon the estimates produced through template fitting – at least not when restricted to using a set of filters that maximises the number of sources that can be fitted. The performance of GPz with respect to the consensus template estimates is mixed, with  $\approx 20\%$ worse scatter but  $\approx 20 - 40\%$  better outlier fractions for the machine learning estimates.

#### 3.3.1 Accuracy of the error estimates

Following Paper I, we quantify the accuracy of the redshift PDFs by examining the cumulative distribution of threshold credible intervals, c, in a q-q plot. In the case of GPz, which provides only uni-modal Gaussian uncertainties with centre  $z_{i,\text{phot}}$  and width  $\sigma_i$ , c can be calculated for an individual galaxy analytically following

$$c_i = \operatorname{erf}\left(\frac{|z_{i,\operatorname{spec}} - z_{i,\operatorname{phot}}|}{\sigma_i}\right).$$
(2)

For each GPz estimate we also implement the additional magnitude-dependent error calibration in a similar fashion to Paper I, varying the width of the Gaussian errors in order to minimise the Euclidean distance between the calculated distribution and the optimum 1:1 relation (see also Gomes et al. 2017, for a similar analysis on uncertainty calibration for GPz estimates). During the error calibration procedure, we make use of the training, validation and test defined when training GPz. Although GPz includes the accuracy of the uncertainties within the metric it aims to minimise, the redshift PDFs output still typically underestimate

## 6 K. J. Duncan et al.



Figure 3. Comparison of photometric redshift estimates versus the spectroscopic redshifts for each of the three AGN population subsets. The left column shows the consensus template-based photo-z as calculated in Paper I. The centre and right-hand columns shows the results from the Gaussian process estimates when trained using the flat and colour-based weighting schemes respectively.

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

Metric		Definition
$\sigma_{ m NMAD}$	Normalised median absolute deviation Bias	$\frac{1.48 \times \text{median}( \Delta z  / (1 + z_{\text{spec}}))}{\text{median}(\Delta z)}$
$\frac{O_f}{CPPS}$	Outlier fraction	Outliers defined as $ \Delta z  / (1 + z_{\text{spec}}) > 0.2$ $\overline{\text{CPDS}} = \frac{1}{2} \sum_{n=1}^{N} \int_{-\infty}^{+\infty} [\text{CPE}_{n}(z_{n}) - \text{CPE}_{n}(z_{n})]^{2} dz = 2$

the photometric redshift uncertainty. This overconfidence is consistent across all three AGN estimators but is noticeably worse when using the colour-magnitude weights in the costsensitive learning.

In Fig. 4, we present the q-q plots of the raw and calibrated error distributions for each of the three AGN estimators - plotting only the validation and test subsets not included in the fitting of the magnitude dependent error calibration. After the error calibration procedure has been applied, we see significant improvement in the accuracy of the redshift PDFs in almost all cases and errors that are close to the ideal solution.



Figure 4. Q-Q ( $\hat{F}(c)$ ) plots for the redshift PDFs for the two Gaussian process photo-z estimates using unweighted (blue) and colourmagnitude weighted (red) training samples. The dot-dash and continuous lines show the results for the raw (as estimated by GPz) and calibrated distributions respectively.

#### 3.4 'Features' in the optical photometry

The strong performance of the Gaussian process redshift estimates in the regime where those from template fitting struggle raises the question of what features in the optical photometry is GPz using to derive the redshift information? And secondly, are those features missing from the template sets employed in the previous photo-z estimates? Or is the failure due to other factors such as variability in the photometry?

Investigating the cause of each template-based photoz failure individually is beyond the scope of this paper. However, we can very easily verify the existence of redshiftdependent colour or magnitude relations upon which the empirical photo-zs might be deriving their results. To illustrate this, in Fig. 5 we show how two example colours and corresponding apparent magnitudes evolve with redshift for the IR selected AGN population. In the redshift regime of 1 < z < 2 where GPz performs exceedingly well, it is clear that there is a strong evolution in the  $3.6\mu m - 4.5\mu m$  colour (with a strong feature at  $z \sim 1.7$ ) while the typical  $I - 3.6 \mu \text{m}$ also become increasingly blue over this range. Coupled with the colour-redshift relations are complementary magnituderedshift relations for the optical and mid-IR bands - the evolution of *I*-band magnitude for a fixed  $I - 3.6 \mu m$  colour with redshift at  $z \gtrsim 1$  remains relatively constant while the apparent  $3.6\mu m$  magnitude shows a much clearer trend of fainter magnitudes at higher redshift. Altogether it is therefore clear that at least for the IR AGN population, there are redshift dependent magnitude or colour features to which we can anchor empirical photo-z estimates.

The follow-up question raised at the beginning of this section was whether the features GPz is basing its redshift predictions from are absent within the templates. Sticking with the example of IR AGN, the bump in  $3.6\mu m - 4.5\mu m$  at  $z \sim 1.7$  is not well represented in the Brown et al. (2014) library - which does not include powerful AGN. But as illustrated by the colour tracks in Fig. 5, the Salvato et al. (2011, see also Hsu et al. (2014)) template set does a good of filling the broad colour region of interest.

There are areas within the colour inhabited by the IR-selected AGN population that the template do not cover,

specifically they do not extend to blue enough  $I - 3.6\mu$ m colours at z > 1 and at  $3.6\mu$ m  $- 4.5\mu$ m the templates are no longer representative for this population in this colour-space. Nevertheless, these deficiencies alone are unlikely to account for the very poor template performance at z < 2 and there may be an additional root causes for these failures. Examination of the average residuals measured for the best-fit templates (both for the free redshift determination and when the redshift is fixed to the known spectroscopic redshift) find no clear indication that any one individual band or colour is responsible for the causing incorrect fits.

Future extensions to the existing template libraries that better sample the full AGN colour space (Brown et al. in preparation) will still likely offer significant improvements in this regime. Due to the focus of this study on the GPz estimates, we defer any further investigation of the AGN template properties to future studies and instead concentrate the rest of our analysis on the machine learning estimates and those derived from them.

### 4 'HYBRID' PHOTO-ZS - COMBINING GP REDSHIFT ESTIMATES WITH TEMPLATE ESTIMATES

One of the key conclusions of Paper I and earlier studies in the literature (e.g. Dahlen et al. 2013; Carrasco Kind & Brunner 2014b) was that no single photometric estimate can perform the best for all source types or in all metrics. Furthermore, the combination of multiple estimates within a statistically motivated framework can yield consensus estimates that perform better than any of the individual inputs. Given the very different limitations and systematics observed in the template and GPz photoz estimates, a consensus photo-z that compounds the advantages of both methods is clearly desirable.

To incorporate the GPz predictions within the hierarchical Bayesian (HB) combination framework presented in Paper I, normal distributions based on position and corrected variance estimate for each source are evaluated onto the same redshift grid as used during the template fitting procedure. Any source in the full training sample that does



Figure 5. Selected observed colours as a function of redshift for the IR-selected AGN population. The upper panel shows the optical to mid-IR colour between the *I* and IRAC 3.6 $\mu$ m bands while the lower panel shows the mid-IR colour between the IRAC 3.6 $\mu$ m and 4.5 $\mu$ m bands. In each panel, the colour of the datapoints corresponds to the apparent magnitude in one of the observed bands. Dashed red lines indicate the colour-tracks as a function of redshift for the XMM-COSMOS (Salvato et al. 2011) templates which satisfy the IR AGN selection criteria of Donley et al. (2012) at any redshift up to z = 3.

not have a photo-z estimate for a given GPz estimator (either through not satisfying the selection criteria for a given subset or lack of observations in a required band) is assumed to have a flat redshift PDF. These sources therefore contribute no information in the HB combination procedure, so in the cases where only one estimate exists the consensus estimate is entirely based on that single prediction.

For comparison with the template-based consensus estimates from Paper I, we calculate two different HB estimates from our GPz estimates. Firstly, we calculate the HB consensus photo-z based only on the four separate GPz estimates. Secondly, we then calculate the HB consensus estimate incorporating all three of the template based estimates calculated in Paper I and the four machine learning estimates from this paper to produce a hybrid estimate. In both cases we follow the practice of Paper I and adopt a magnitude based prior when an observation is assumed to be 'bad'.

In Fig. 6 we present the photo-z vs spectroscopic redshift distribution of the three separate HB consensus estimates. To better illustrate the overall uncertainty and scatter given the large number of sources, we show the stacked redshift probability distributions within a spectroscopic redshift bin rather than individual point estimates. The left panel of Fig. 6 illustrates the previously known limitations of template-based photoz estimates for most AGN sources. At z < 1 the template estimates perform well, but between 1 < z < 3 the photo-z probability distributions are extremely broad; possibly due to the lack of strong photometric features in the optical SEDs in this regime. Additionally, the degradation of the template photo-z quality towards higher redshift may be a result of differences in the source population selected at higher redshift; the galaxy template work well for the low-luminosity AGN but fail for higher luminosity AGN where the host galaxy no longer dominates the optical emission. At  $z \gtrsim 3$ , the template-based estimates begin to perform well again due to the redshifted Lymancontinuum break moving into the observed optical bands.

It is worth noting that the extent of the template photoz issues at 1 < z < 3 are partly field specific, in that the relative depths of the near-infrared data available in the Boötes field are shallow with respect to the optical and mid-infrared data at wavelengths either side. As such, sources which may have high signal-to-noise (S/N) detections in the optical regime may still have very low S/N in the near-IR bands that probe the rest-frame optical features (both in spectral breaks and emission lines) at  $z \gtrsim 1$ ). Figure 5 of Paper I shows that in fields with deeper photometry and finer wavelength coverage (e.g. the COSMOS field Laigle et al. 2016) the trends are not as extreme, particularly at 1 < z < 2. Nevertheless, the improvement seen here is particularly encouraging for photo-z estimates in surveys without the same levels of exceptional filter coverage as available in COSMOS.

In contrast to the trends observed in our template estimates, and consistent with the trends seen in individual AGN estimates shown in Fig. 6, the GPz-only consensus estimates perform best in the region of  $1 \leq z \leq 2$ . At lower  $(z \leq 0.5)$  and higher  $(z \geq 2.5)$  redshifts, the GPz consensus estimate becomes increasingly biased. It is these wavelength regimes in which the training samples for the AGN population are most sparse, as can be seen visually in the right hand column of Fig.3.

Most encouraging however is the HB consensus estimate incorporating both the template and machine learning based predictions (right panel of Fig. 6). Visually, it is immediately clear that the total combined consensus estimate combines the advantages of both of the input methods.

This improvement can also be seen more quantitatively by looking at the measured photo-z scatter and outlier fraction for the AGN population as a function of redshift (Fig. 7). At z < 1, the hybrid estimates match or improve upon the scatter from the template estimates. Then, at 1 < z < 3, the hybrid estimates match the improved scatter and outlier fractions of the GPz estimates while the template-based estimates perform very poorly. Finally, at  $z \gtrsim 3$  when strong continuum features result in improved template estimates, the hybrid estimates are still able to perform comparably.

Fig. 8 shows the measured scatter and outlier fraction as a function of apparent *I*-band magnitude. At all magnitudes brighter than  $I \approx 23.5$ , the hybrid estimates perform better than either the template or GPz only estimates. The observed improvement in scatter for the GPz only estimates at the very faintest magnitudes (as compared to the template or hybrid method) likely results from the cost-sensitive



Figure 6. Stacked probability distributions for the combined AGN population (IR, X-ray or optically selected) as a function of spectroscopic redshift for each consensus HB photo-z estimate. To improve the visual clarity at higher redshifts where there are few sources within a given spectroscopic redshift bin, the distributions have been smoothed along the x-axis. The same smoothing has been applied to all three estimates consistently. The superior performance of the hybrid template + GPz estimates is well illustrated by the side-by-side comparison.



Figure 7. Photometric redshift scatter ( $\sigma_{\text{NMAD}}$ ) and outlier fraction as a function of spectroscopic redshift for AGN sources in the Boötes field. Lines show the results for sources that pass any of the X-ray/Optical/IR AGN criteria outlined in Section 2.1.

learning increasing the importance of these faint AGN during the optimisation procedure. However, it is evident that the hybrid estimates are most similar in performance to the template-only estimates in this regime, with the rise in scatter and outlier fraction at I > 23 closely mirroring the observed rise. The apparent inability of the hybrid consensus estimates to mirror the performance of the best performing estimate could be seen as a failure of the hierarchical Bayesian combination method at faint magnitudes.

Without a detailed inspection of all individual estimates before and after Bayesian combination, it is not immediately obvious what is causing this behaviour. Nevertheless, there are two possible explanations. Firstly, the statistics shown in



**Figure 8.** Photometric redshift scatter ( $\sigma_{\text{NMAD}}$ ) and outlier fraction as a function of *I* magnitude for AGN sources in the Boötes field. Lines show the results for sources that pass any of the X-ray/Optical/IR AGN criteria outlined in Section 2.1. At almost all redshift ranges, the hybrid photo-*z* performance is comparable or better to the best input methodology.

Fig. 8 do not take into account the additional signal-to-noise criteria that are implicit to the GPz estimates, i.e. the lack of estimates for sources that do not have magnitude estimates in all bands required for GPz. The metric may therefore not be a fair comparison at faint magnitudes. However, given that all three AGN-targeted GPz estimates have detections in the required bands for > 95% of the training sample, this effect should not be particularly large.

Alternatively, if the individual errors on the template estimates are systematically smaller than those of the GPz estimates for the same sources (i.e. the redshift PDFs are narrower), then the consensus estimate itself will converge



Figure 9. Median positive 80% highest probability density confidence intervals,  $\Delta_{z_1}$ , above the primary redshift solution,  $z_1$ , as a function of I magnitude for the AGN sources in the Boötes field. 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.

towards those estimates. Evidence for this can be seen in Fig. 9, where we show the median positive error as a function of magnitude for the three consensus estimates. Despite having a larger scatter and outlier fraction across almost all magnitudes, the average quoted error on an individual source is smaller for the template based estimates.

#### 4.1 Comparison to Brodwin et al. (2006)

As mentioned in the introduction, this study is not the first to attempt to combine the different strengths of templatebased and empirical photo-z estimates. In addition to the comparison of different methods for Bayesian combination of template and machine learning estimates presented in Carrasco Kind & Brunner (2014b), Brodwin et al. (2006) have also previously explored a hybrid photo-z method aimed at improving estimates for AGN within the Boötes field.

Based on predominantly the same underlying photometry as used in this analysis, Brodwin et al. (2006) estimated photo-zs using two approaches - firstly using template fitting and secondly employing an empirical method using neural networks (Collister & Lahav 2004). The most direct comparison we are able to make between the results of Brodwin et al. (2006) and those presented in this work is via their quoted estimates of the 95%-clipped photo-z scatter.

For AGN between 0 < z < 3 in the AGES (Kochanek et al. 2012) spectroscopic sample, Brodwin et al. (2006) find a scatter of  $\sigma_{95\%}/(1+z) = 0.12$  and for galaxies between 0 < z < 1.5 a lower scatter of  $\sigma_{95\%}/(1+z) = 0.047$ . Restricting our spectroscopic sample to contain only those from AGES and requiring a  $4.5\mu m$  detection to best match the Brodwin et al. selection criteria, our hybrid photo-z estimate have comparable 95%-clipped scatters of  $\sigma_{95\%}/(1+z) = 0.11$  and  $\sigma_{95\%}/(1+z) = 0.045$  for sources classified by AGES as AGN and galaxies respectively.

When comparing the two results it is important to recognise that the template-fitting and the GPz estimates trained for the galaxy population make use of additional photometry not available at the time of Brodwin et al. (2006, e.g. u, z and y). Some small improvement is therefore to be expected.

A key improvement offered by the Bayesian combination framework employed in this work is that it is able to make maximal use of the redshift information available for a given source. In Brodwin et al. (2006), the choice of template or neural-network based estimates for a given source is a binary based on where a source with respect to the Stern et al. (2005) IRAC colour criteria (similar to the criteria we have used for selecting IR AGN). As seen in Fig. 3 the performance of machine learning estimates for these sources is significantly better over the redshift range of interest, so this choice is well motivated. However, at higher redshifts the machine learning estimates become increasingly biased due to the sparsity of the training samples in this regime. This bias is clearly visible both in Fig. 5 of Brodwin et al. (2006) and in the centre panel of Fig. 6 of this work. Although still imperfect, the hierarchical Bayesian combination procedure is able to fall back on the more accurate and reliable template-based estimates at  $z \gtrsim 2.5$ .

# 4.2 Hybrid photo-*z* performance for the radio source population

Given our motivation in producing the best possible photo-z estimates for the diverse population selected objects in forthcoming radio continuum surveys, it is interesting to see how the improvement seen in the optical/IR/X-ray selected AGN population propagates through into the hybrid photo-z performance for radio selected objects. In Fig. 10 we illustrate the  $\sigma_{\rm NMAD}$ ,  $O_{\rm f}$  and CRPS performance of the template, GPz and hybrid consensus redshift estimates in each of the source population subsets. Across all subsets of the radio detected populations, the hybrid photo-z estimates either match or significantly improve upon the scatter and outlier fraction performance of the best single method.

Furthermore, across all subsets of the radio population the scatter is now  $\sigma_{\rm NMAD} \lesssim 0.1$ , an improvement of up to a factor of four compared to the template estimates. Despite them not performing significantly better than the template estimates for sources not optically classified as AGN, the inclusion of GPz estimates in the hierarchical Bayesian photozs results in a factor of ~ 2 improvement in outlier fraction for the radio-detected subset of these sources.

Exploring the key quality statistics as a function of radio luminosity (Fig. 11) and flux (Fig. 12) we can see more clearly that the greatest gain in improvement is for the most luminous radio sources. For a given apparent radio flux, the GPz and hybrid estimates offer no clear improvement in terms of scatter but do improve the outlier fraction. This behaviour is something we would expect to see, bearing in mind that lower luminosity sources at low redshift dominate the spectroscopic sample we are comparing ( $\approx 90\%$  of the spectroscopic sample is at z < 1). The rarer high luminosity radio sources for which GPz produces more accurate photo-z estimates have a broad range of apparent fluxes and therefore the robust scatter is not strongly affected but the outlier fraction is.

The performance of the GPz-only estimates compared to the template-only estimates as a function of radio power could shed further light on the discussion in Section 3.4 on



Figure 10. Visualised photometric redshift performance in three metrics ( $\sigma_{\text{NMAD}}$ ,  $O_{\text{f}}$ ,  $\overline{\text{CRPS}}$ ; see Table 2) for the different Boötes field radio source subsamples. For all subsets of the radio-detected population, the hybrid method performs better than either template or GPz alone.

the causes of failures in the template fitting. That GPz performs best for the most luminous radio AGN could support the idea that our selected template fits struggle most in the regime where the AGN dominates the optical emission. Although in the local Universe the most powerful radio sources are typically host-dominated in their optical emission, at higher redshifts the population is dominated by QSO/Seyfert-like sources (e.g. Heckman & Best 2014, and references therein). Within a deep survey field such as that used in this work, the larger volume probed at high redshift means that z > 1 sources are the dominate the high luminosity end of our sample. Further exploration of the different methods as a more detailed function of radio luminosity and redshift would clearly be valuable in better understanding our methods and their strengths and limitations, however the currently limited training sample makes this impractical.

# 4.3 Prospects and strategies for further improvements

Despite the substantial improvement in photo-z accuracy and reliability for the GPz and hybrid estimates, the inho-



Figure 11. Photometric redshift scatter ( $\sigma_{\text{NMAD}}$ ; upper panel) and outlier fraction ( $O_f$ ; lower panel) as a function of 150MHz radio luminosity for all radio detected sources within the spectroscopic redshift range 0 < z < 3. In each plot we show the values for the template-only (circles), GPz-only (upward triangles) and combined (downward triangles) consensus estimates. Symbols have been offset horizontally only for clarity, luminosity bins for all estimates are identical. Error-bars plotted for the outlier fractions illustrate the binomial uncertainties on each fraction. The hybrid estimate performs significantly better than either the template or GPz-only estimates across the full range of radio luminosities probed in this field, with particularly large improvement at the greatest radio powers.



Figure 12. As in Fig. 11 but for 150MHz radio flux - for all radio detected sources within the spectroscopic redshift range 0 < z < 3. Due to the majority of the spectroscopic training sample probing low redshift sources where the template estimates perform well, the improvement in the scatter for the hybrid estimates is not significant. However, the number of catastrophic outliers in the hybrid estimates is lower than the template-only estimates at all fluxes.

mogenous photo-z quality across the sub-populations within the radio detected subset indicate that there is still potential for further improvements to be gained. With regards to the GPz and resulting hybrid estimates, such improvements could potentially come from several different aspects of the methodology.

Firstly, as is the case in all empirical photo-z estimates, the accuracy of GPz is limited by the training sample being used. Key to the production of accurate photo-zs based on training samples is not necessarily the sheer size of the training sample, but rather its ability to fully represent the parameter space probed by the catalogs to which the method will be applied. The effect of limited training samples can be seen in the performance of GPz at both the very lowest and highest redshifts, the regimes in which the training sample is particularly sparse. Although our implementation of colour and magnitude based weights within the cost-sensitive learning is able to mitigate some effects of the biased training sample, it will never be able to account for regions of parameter space which are entirely absent from the training data.

In coming years, the problems caused by limited training samples will be solved by forthcoming large-scale spectroscopic surveys. In Paper I we discussed how for the radiocontinuum selected population the  $> 10^6$  radio source spectra provided WEAVE-LOFAR (Smith et al. 2016) will provide an ideal reference and training sample for photo-*z* estimates in all-sky radio surveys. While helpful for improving the template-based estimates, such a training sample will be transformational for machine-learning photo-*z* estimates of radio sources in future continuum surveys.

In the short term however, it should be possible to better leverage the spectroscopic redshift samples already available in the literature. The Herschel Extragalactic Legacy Project (HELP: Vaccari 2016) is bringing together all publicly available multi-wavelength datasets within the regions of the sky observed in extragalactic Herschel surveys. The collation and homogenisation of these many datasets offers the possibility to leverage the extensive spectroscopic datasets in some survey fields to significantly improve estimates in other fields where training samples are particularly sparse.

Secondly, in deep fields such as Boötes, the heterogenous nature of the optical data means that GPz in its current form is not able to make full use of the available information. This problem is illustrated in Section 3.1, with the only 38.3% of sources having magnitude information available in five filters and significantly fewer when additional available bands are included. In the cases where magnitude information is missing as a result of non-detections in the data, training and fitting the photo-zs on fluxes rather than magnitudes would largely solve this problem provided the algorithms being used still perform well in the linear regime. In many other cases however, the missing data can be a result of instrumental effects (e.g. masked regions due to bright stars or diffraction spikes) or differences in the survey coverage.

The flexibility of the hierachical Bayesian combination procedure outlined in this paper allows for the possibility of training GPz on any/all combinations of the photometric data and combining those estimates to produce a consensus estimate given all the available information. However, such a procedure would rapidly become impractical in some fields. Recent developments of the GPz algorithm whereby missing data can be jointly predicted with the redshift (Almosallam et al. in prep) will be of great benefit in the future and could result in significant improvements to the empirical photo-z estimates in these heterogenous deep fields.

Finally, there is also potential for further improvements which can be made to the Bayesian combination. With additional improvements to the input redshifts themselves, suboptimal combinations of the various estimates such as those seen at  $z \sim 3$  in Fig. 7 will have less of an effect on the final consensus redshifts. Nevertheless, more informative priors could be incorporate into the combination procedure which gives more weight to individual estimates in regions of parameter space in which they are known to perform better. Such an improvement is illustrated in Carrasco Kind & Brunner (2014b), with the performance of Bayesian Model Averaging (BMA) and Bayesian Model Combination (BMC) exceeding that of hierarchical Bayesian combination in their implementation. However, in the context of photo-zs for AGN, we believe these gains will be very small compared to the other strategies outlined in this section.

#### 5 SUMMARY

Building on the first paper in this series which explored the performance of template-based estimates (Duncan et al. 2018, Paper I), we have presented a study exploring how new estimates from machine learning can be used to significantly improve photo-z estimates for both the radio continuum selected population and the wider AGN population as a whole. Using the Gaussian process redshift code, GPz, we have produced photo-z estimates targeted at different subsets of the galaxy population - infrared, X-ray and optically selected AGN - as well as the general galaxy population. The GPz photo-z estimates for the AGN population perform significantly better at z > 1 than photo-z estimates produced through template fitting presented in Paper I. Compared to the template-based photo-zs, GPz estimates for the IR/Xray/Optical AGN population have lower scatter and outlier fractions by up to a factor of four.

By combining these specialised GPz photo-z estimates with the existing template estimates through hierarchical Bayesian combination (Dahlen et al. 2013; Carrasco Kind & Brunner 2014b) we are able to produce a new hybrid consensus estimate that outperforms either of the individual methods across all source types. The overall quality of photo-z estimates for radio sources that are X-ray sources or optical/IR AGN are vastly improved with respect to Paper I, with outlier fractions and scatter with respect to spectroscopic redshifts reduced by up to a factor of  $\sim 4$ .

For both the radio detected population with no strong optical signs of AGN (i.e. radio AGN hosted in quiescent galaxies or star-forming sources) our new methodology also provides significant improvement. Despite the template and GPz estimates performing very comparably when treated separately, the combination of the two sets of estimates yields outlier fractions which are a factor of  $\approx 2$  lower. Investigating the new photo-z estimates as a function of radio property, we find that the improvement observed for the radio selected population can likely be attributed to the highest luminosity radio sources for which the GPz estimates

The success of the method despite the small training samples and heterogeneous datasets available is encouraging for future exploitation of deep radio continuum surveys for both the study of galaxy and black hole co-evolution and for cosmological studies.

#### ACKNOWLEDGEMENTS

The research leading to these results has received funding from the European Union Seventh Framework Programme FP7/2007-2013/ under grant agreement number 607254. This publication reflects only the author's view and the European Union is not responsible for any use that may be made of the information contained therein. KJD and HR acknowledge support from the ERC Advanced Investigator programme NewClusters 321271.

#### 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
- 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
- Benítez N., 2000, ApJ, 536, 571
- Bolzonella M., Miralles J. M., Pello R., 2000, A&A, 363, 476
- Booth R. S., de Blok W. J. G., Jonas J. L., Fanaroff B., 2009, eprint arXiv:0910.2935, pp-
- Bordoloi R., Lilly S. J., Amara A., 2010, MNRAS, 406, no
- 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
- Brodwin M., et al., 2006, ApJ, 651, 791
- Brown M. J. I., et al., 2014, ApJS, 212, 18
- 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
- Collister A. A., Lahav O., 2004, 116, 345
- Dahlen T., et al., 2013, ApJ, 775, 93
- 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
- Drlica-Wagner A., et al., 2017, arXiv, p. arXiv:1708.01531
- Duncan K. J., et al., 2018, MNRAS, 473, 2655
- Flesch E. W., 2015, Publ. Astron. Soc. Aust., 32, e010
- Gomes Z., Jarvis M. J., Almosallam I. A., Roberts S. J., 2017, arXiv
- Heckman T. M., Best P. N., 2014, ANNU REV ASTRON ASTR, 52, 589
- Hsu L.-T., et al., 2014, ApJ, 796, 60  $\,$
- Jannuzi B. T., Dey A., 1999, The Young Universe: Galaxy Formation and Evolution at Intermediate and High Redshift, 191, 111
- Jarvis M. J., et al., 2017, arXiv, p. arXiv:1709.01901
- Johnston S., et al., 2007, Publ. Astron. Soc. Aust., 24, 174
- Kenter A., et al., 2005, ApJS, 161, 9

- Kochanek C. S., et al., 2012, ApJS, 200, 8
- 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
- Lima M., Cunha C. E., Oyaizu H., Frieman J., Lin H., Sheldon E. S., 2008, MNRAS, 390, 118
- Merloni A., et al., 2012, arXiv, p. arXiv:1209.3114
- Norris R. P., et al., 2013, Publ. Astron. Soc. Aust., 30, e020
- Oke J. B., Gunn J. E., 1983, ApJ, 266, 713
- 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
- Smith D. J. B., et al., 2016, eprint arXiv:161102706, pp -
- Stanford S. A., et al., 2012, ApJ, 753, 164  $\,$
- Stern D., et al., 2005, ApJ, 631, 163
- Vaccari M., 2016, The Universe of Digital Sky Surveys, 42, 71
- Williams W. L., et al., 2016, MNRAS, 460, 2385
- Williams W. L., et al., 2017, arXiv, p. arXiv:1711.10504
- Zeimann G. R., et al., 2012, ApJ, 756, 115
- Zeimann G. R., et al., 2013, ApJ, 779, 137
- van Haarlem M. P., et al., 2013, A&A, 556, A2

This paper has been typeset from a  $T_{\rm E}X/I\!\!\!{}^{\rm A}T_{\rm E}X$  file prepared by the author.