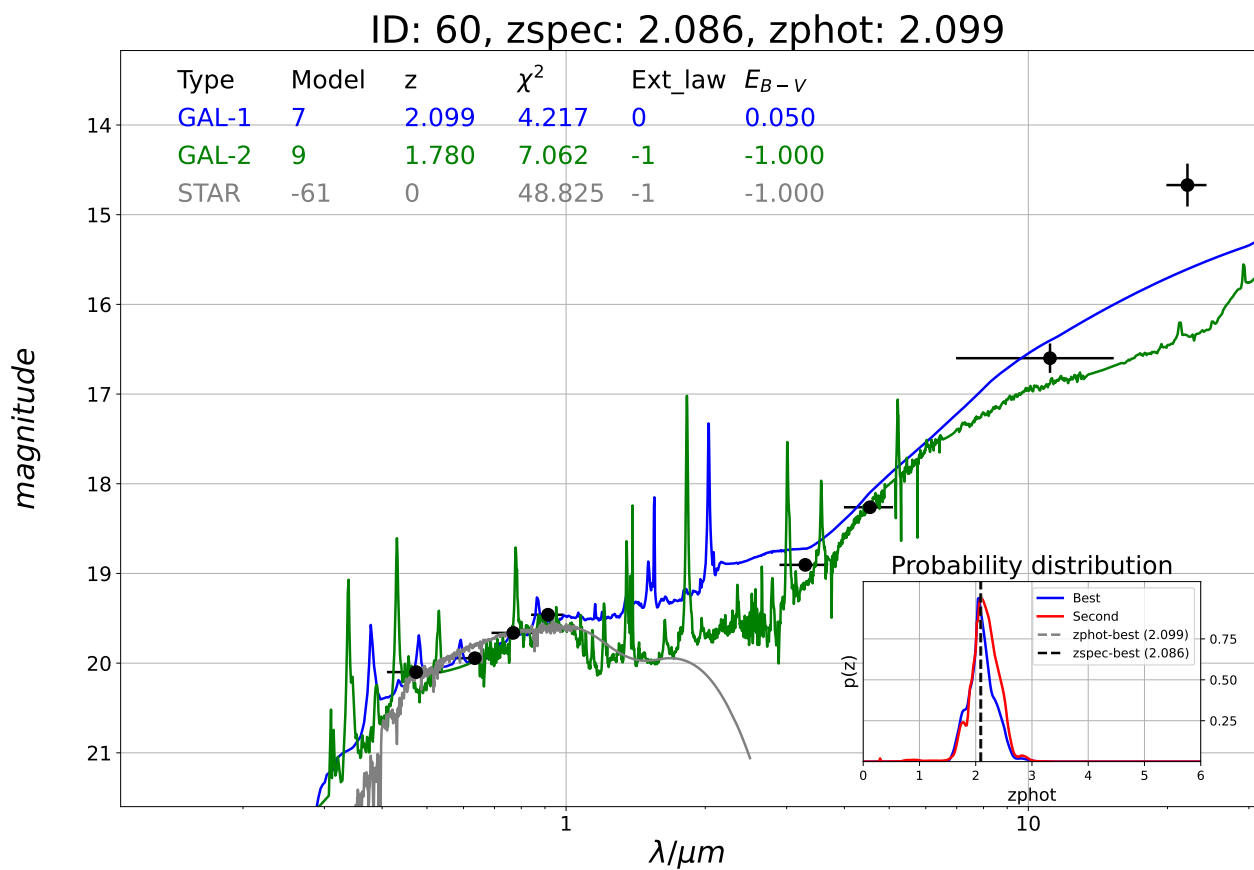


Quasar selection for 4Hi-Q using photometric redshifts

Fabian Balzer



Template fit to the photometry of a pointlike source (see sec. 3.4.3 for further explanation).

Abstract

Context.

With the advent of the new, powerful spectroscopic survey facility 4MOST at the Visual-Infrared Survey Telescope for Astronomy (VISTA), a variety of surveys to exhaust its capabilities have been proposed.

One of these proposals is the 4Hi-Q project. By analysing absorption lines (Mg II, C IV and H I) in the spectra of high-redshift quasars observed using the high-resolution spectroscopic fibres of 4MOST, valuable information about the circumgalactic medium (CGM) of galaxies along the line of sight is going to be gathered. This will enable a census of the cosmic metals and improve our understanding of the cosmic baryon cycle.

Goals.

In this thesis, a data pipeline is developed to provide the input catalogue with the more than 900 000 quasar sources needed for this project.

To be able to observe the absorption lines of the CGM in the foreground of the quasar spectra, it needs to be ensured that all observed sources have a true redshift $z > 0.55$, as the absorbers of interest are located at redshifts $0.4 < z < 0.55$ for the low-redshift part of the 4Hi-Q, and at even higher z for the high-redshift part.

This poses a problem: As most quasars in the available catalogues have only been observed via flux measurements, no spectroscopic redshift information is available.

As an alternative, one can use photometric redshifts, which are obtained by analysing the colours of galaxies derived from flux measurements in a sample of different photometric bands. These can be seen as a coarse spectral energy distribution, allowing the identification of key spectral features and thus a redshift estimation. The precision of this estimation improves with more photometric bands.

Methods.

The data pipeline set up in this work therefore gathers photometric data from a variety of catalogues, processes it and estimates the redshift for all sources. The computation of the photometric redshift is carried out with the LePhare algorithm (Arnouts et al. 1999; Ilbert et al. 2006).

This thesis emphasises on optimising the various possible input parameters to improve on these redshifts.

More specifically, preliminary analysis is carried out in the eROSITA Final Equatorial Depth Survey (eFEDS) field. The 85 % reliability infrared selected quasar catalogue by Shu et al. (2019) is used as a parent catalogue. Photometric data from the DESI Legacy Imaging surveys (Schlegel et al. 2021) (g , r , z and W_1 - W_4 bands), the Vista Hemisphere Survey (McMahon et al. 2013) (Y , H , J , K_s bands) and the Galaxy Evolution Explorer (Bianchi et al. 2017) (FUV and NUV bands) are then used to provide the input photometry for LePhare. Spectroscopic verification data is obtained from a catalogue of X-ray selected quasars in the eFEDS field (Salvato et al. 2021).

With the upcoming 10th data release of the DESI Legacy Imaging Surveys, the improvements of the photometric redshifts upon inclusion of additional i band data is assessed by analysing HyperSuprimeCam (Miyazaki et al. 2018) and Kilo-Degree Survey (Kuijken et al. 2019) catalogue data.

Results.

With the improved templates, substitute i band data from the HyperSuprimeCam survey and preliminary i data from the DR10 of the Legacy Surveys, an outlier fraction of $\eta_{\text{out}} = 0.244$ (0.299), accuracies of $\sigma_{\text{NMAD}} = 0.107$ (0.125) and false positive fractions of $\psi_{\text{Pos}} = 0.013$ (0.005) could be achieved for the pointlike (extended) sub-sample in the eFEDS field.

The pipeline to extract the photometric data from the different surveys and to perform the photometric redshifts is in place and ready to generate the input catalogue for 4Hi-Q on the whole sky.

Eidesstattliche Erklärung

Ich versichere, dass ich die beigelegte schriftliche Masterarbeit selbstständig angefertigt und keine anderen als die angegebenen Hilfsmittel benutzt habe.

Alle Stellen, die dem Wortlaut oder dem Sinn nach anderen Werken entnommen sind, habe ich in jedem einzelnen Fall unter genauer Angabe der Quelle deutlich als Entlehnung kenntlich gemacht. Dies gilt auch für alle Informationen, die dem Internet oder anderer elektronischer Datensammlungen entnommen wurden.

Ich erkläre ferner, dass die von mir angefertigte Masterarbeit in gleicher oder ähnlicher Fassung noch nicht Bestandteil einer Studien- oder Prüfungsleistung im Rahmen meines Studiums war.

Die von mir eingereichte schriftliche Fassung entspricht jener auf dem elektronischen Speichermedium.

Ich bin damit einverstanden, dass die Masterarbeit veröffentlicht wird.

Hamburg, 09.07.2022

Ort, Datum

Falcan Bülzer

Unterschrift

Contents

1. Introduction	6
1.1. 4MOST and the 4Hi-Q survey	6
1.1.1. The 4-metre Multi-Object Spectroscopic Telescope	6
1.1.2. 4MOST High Resolution Quasars	6
1.1.3. The goal of this thesis: Building an input catalogue for 4Hi-Q	8
1.2. Photometric redshifts	10
1.2.1. Inferring redshift from photometry	11
1.2.2. Templates and template fitting	12
1.2.3. A note on photometric redshifts for quasars	13
1.2.4. Validation of photometric redshifts	14
2. Procedure	17
2.1. Working with the photometric redshift code LePhare	18
2.1.1. Reading and processing filter libraries	18
2.1.2. Providing extinction information	21
2.1.3. Building an SED template library	21
2.1.4. Template processing and the creation of magnitude libraries	21
2.1.5. Fitting photometric redshifts	22
2.2. Preparing the input catalogue for LePhare	25
2.2.1. The AGN parent catalogue	26
2.2.2. Optical and infrared: The DESI Legacy Imaging Surveys	27
2.2.3. Near-Infrared: The VISTA Hemisphere Survey	29
2.2.4. Ultraviolet: The Galaxy Evolution Explorer	31
2.2.5. Validation data: Matched spectroscopy from eRASS	32
2.2.6. Supplementary <i>i</i> band data: HSC	33
2.2.7. Supplementary <i>i</i> band data: KiDS	35
2.2.8. Overview	36
2.3. Calculating photometric redshifts with LePhare	39
2.3.1. Collecting filter information	39
2.3.2. Selection of input templates	39
2.3.3. Running LePhare with priors	41
2.4. Detailed description of the Code	41
2.4.1. Requirements	41
2.4.2. Available parameters in the configuration file	42
2.4.3. Running the code	42
3. Improvements	44
3.1. First results in the eFEDS field	44
3.2. Improving the template selection	45
3.3. Introducing <i>i</i> band photometry	48
3.4. Further discussion	51
3.4.1. Analysing the used bands	51
3.4.2. Template coverage	52
3.4.3. Taking a closer look at certain SEDs	53
3.4.4. Overview on the possible sources of insufficiencies	53
4. Summary and outlook	57

A. Code snippets	58
A.1. Downloading the SWEEP files in a given region	58
A.2. Querying for VHS data	59
A.2.1. Querying for HSC data	59
A.2.2. Querying for KiDS data	60
A.2.3. The configuration file for the code	60
B. Additional tables and figures	62
B.1. Code configuration parameters	62
B.2. Templates	65
B.3. Band usage overview	67
B.4. More individual SED plots	67
C. Acknowledgements	72
C.1. Personal acknowledgements	72
C.2. Legacy Surveys	72
C.3. GALEX	73
C.4. KiDS data	73
C.5. HyperSuprimeCam	73

1. Introduction

1.1. 4MOST and the 4Hi-Q survey

1.1.1. The 4-metre Multi-Object Spectroscopic Telescope

One of the most promising upcoming scientific projects for astronomers covering a broad range of use-cases is the 4-metre Multi-Object Spectroscopic Telescope (4MOST), which is a high-multiplex, wide-field spectroscopic survey facility (de Jong et al. 2019) currently being developed and planned to be installed at the four-metre-class Visible and Infrared Survey Telescope for Astronomy (VISTA) at Paranal in Chile. Currently, the expected start of scientific observations is the first quarter of 2024.

The facility is going to have a field of view of 4.2 square degrees with 1624 independent fibres connected to two low-resolution (low-res) spectrographs and 812 fibres feeding a high-resolution (high-res) spectrograph. This combination allows for simultaneous spectroscopy of ~ 2400 objects in the optical wavelength range.

The spectral resolution

$$R := \frac{\Delta\lambda}{\lambda} \quad (1.1)$$

is a commonly used quantity to characterise a spectrograph. The 4MOST spectrographs are planned to have $R \sim 6500$ (low-res) and $R \sim 20000$ (high-res) with small variations among the respective spectral ranges.

In a five-year observation schedule, 4MOST is expected to observe a large portion of the Southern sky two to three times, resulting in more than 30 million spectra. Around 70 % of the observations are going to be Guaranteed Time Observations (GTOs) awarded to so-called Consortium Surveys. The consortium is the set of institutes which are involved in the construction of the facility and the set up of the data pipelines.

The remaining observation time is awarded to Public Surveys by the ESO community and the Chilean host country, chosen via a peer-reviewed selection process. The corresponding proposals were submitted by the end of 2020, the results of the process were published by the end of 2021.

1.1.2. 4MOST High Resolution Quasars

One of the accepted Public Surveys is the 4MOST High Resolution Quasars (4Hi-Q) survey. This project aims to observe a large sample of quasars with the high-resolution spectrograph of 4MOST.

Its objective is to shed new light onto the mechanisms involved in the baryon cycle and the connection of gas, stars and metals by investigating the Circumgalactic Medium (CGM), as currently more than 90 % of the baryons are found in low-density gas (compare Péroux and Howk 2020a), which is hard to detect via its emission features.

Observational methods

The 4Hi-Q aims to utilise absorption line spectroscopy to gather the information about the metal content of the CGM. In this technique, high-resolution spectra of bright background sources are analysed for characteristic absorption lines that arise from the light passing through intervening gas clouds^I along the way. Once a spectral line of an element has been identified, the redshift of the absorber can be inferred.

Since the background sources have to be further away than the absorbers, they need to be

^Iboth inside of galaxies, close to them (the CGM), and intergalactic

sufficiently bright. While in principle there are other possible objects and phenomena^I, the 4Hi-Q survey plans on observing bright, high-redshift quasars as the background sources. Via absorption line spectroscopy, the column density of atoms or molecules in the line of sight can be measured. As an example, the observed HI column density $N(\text{HI})$ depends on the neutral hydrogen particle density n_{HI} :

$$N(\text{HI}) = \int_0^{d_{\text{bs}}} n_{\text{HI}}(s) ds, \quad (1.2)$$

where d_{bs} is the distance to the background source and s the path along the line of sight through the gas.

An advantage of this approach is that the minimum observable gas density is only set by the quasar’s brightness and the detection efficiency, being independent of the redshift of the absorber. On the one hand, this allows for much lower gas densities to be detected than via emission line measurements, but on the other hand, such an observational strategy has the downside that these measurements are limited to pencil-beams along the line of sight and require the bright background sources to be present.

Thus, to be able to infer meaningful conclusions on the CGM, a large number of these background quasars needs to be observed to assemble a meaningful statistical sample of absorbers by combining the multiple lines of sight.

Scientific goals

The following goals have been outlined in the initial 4Hi-Q survey proposal:

- **The low-redshift sample** ($z < 1.04$):

Quasar absorbers are cross-correlated with galaxies detected in emission in the same fields. More specifically, the radial profile, covering fraction and optical depth of neutral hydrogen and metals in the CGM are going to be measured.

The *Circum-Galactic Medium* (CGM) is defined as the gas inside the virial radius of galaxies not associated with their discs or interstellar media (see e. g. Tumlinson et al. 2017). Observations and simulations indicate that its influence on galaxy evolution is significant: It both stores the gas that fuels star formation and is the location where galactic feedback and recycling takes place.

Peebles et al. (2014) suggest that the CGM might be the prevalent reservoir of heavy elements, with a mass comparable to or even higher than the mass of the galaxies themselves, indicating that further investigation might be key to better understand star formation, feedback processes and its contribution to the baryon budget.

As the gas in the CGM is diffuse, it has not been studied as thoroughly as other components of galaxies. Thus, there are still many important questions not yet answered properly which are addressed by the 4Hi-Q cross-correlation of absorbers and galaxies:

- What is the distribution of metals in the CGM?

The 4Hi-Q team aims to measure the radial profile of neutral hydrogen and metals as a function of the galactocentric radius. This will enable the computation of the number density of metal absorbers $\frac{dn}{dz}$ with respect to z . These measurements will result in constraints on the extent of the CGM and provide limits for the scales of metal mixing in the CGM.

- What is the baryon budget of the universe?

As the CGM consists of hot, metal-rich, and cold, metal-poor gas, an assessment of the mass of these phases and the baryon budget in general is instructive.

The 4Hi-Q team aims to compute the optical depth of the metal doublets for

^Ie. g. gamma-ray bursts, fast radio bursts and even other bright galaxies.

different phases of the gas in the $T = 10^4$ K to $T = 10^5$ K range (from Mg II to C IV). As the number of available absorber-galaxy pairs will see a 1000-fold increase with respect to previous observations, this will result in improved 2D-maps of the optical depth around galaxies similar to those obtained from cosmological EAGLE simulations (Turner et al. 2017).

- How do the properties of the CGM evolve with cosmic time, and how dependent are they on galaxy type?

With the large dataset of quasar-absorber pairs acquired, these questions and the impact of AGN feedback on the CGM metal enrichment can be pursued.

The direct cross-correlation of the CGM to the corresponding galaxies is convenient as, for some areas, they are going to be observed by GTO surveys^I at the same time using 4MOST’s low resolution fibres.

- **The high-redshift sample ($z > 2.58$):**

The second goal is to build new catalogues of H I absorbers which trace both the neutral phase of the gas and its metal content, providing a census of the cosmic metals.

The evolution of the metal contents in cosmic environments is directly linked to the processes driving galaxy formation and evolution.

While Pettini (1999) identified that the co-moving density of the observed metals at $z \approx 2$ was an order of magnitude lower than what would be expected to be produced by stars (the so-called *missing metals problem*), Péroux and Howk (2020b) note that, using recent observations, the expected metal content of the Universe is likely accounted for. Still, there are redshift ranges ($1 \lesssim z \lesssim 2.5$) where the census of metals is yet to be done (for $z < 1$, 50 % of the metals are contributed by stars, while for $z > 2.5$ 90 % of the metals are in the neutral gas).

The 4Hi-Q team aims to provide an estimate for the number density per unit redshift $\frac{dn}{dz}$ for LLS and DLAs^{II}.

Via further investigations, the missing metals problem can be answered for the intermediate redshift ranges not yet accounted for, enabling a census of the metals $\Omega_{\text{metals}} = \Omega_{\text{gas}} \langle Z \rangle$.

Survey Strategy

In the 4Hi-Q project, the quasars are going to be observed using the high-resolution fibres of 4MOST that provide a spectral resolution of $R \sim 20000$. To be as efficient as possible, the 4Hi-Q survey is planned such that the redshift distribution and area of GTO low-resolution surveys, namely the S7 Wide-Area VISTA Extragalactic Survey (WAVES) deep, the S8 Cosmic Redshift Survey (CRS), Luminous Red Galaxies (LRG), Emission Lines Galaxies (ELG), S5 and S6 cluster surveys, and S9 1001MC Magellanic Cloud survey, is matched (see tab. 1.1).

1.1.3. The goal of this thesis: Building an input catalogue for 4Hi-Q

As outlined above, the 4Hi-Q survey is in need of a sample of bright, non-transient background objects, i. e. quasars, to apply the absorption line spectroscopy and gain information about the CGM gas along the line of sight.

^Ie. g. the WAVES (Driver et al. 2019) Deep, CRS (Richard et al. 2019) Luminous Red Galaxies and Emission Lines Galaxies surveys.

^{II}To distinguish the neutral gas, one defines Lyman Limit Systems (LLS) for which the optical depth is greater than unity below the Lyman limit at 912 Å, which can be identified by a distinct flux break in the quasar’s spectrum, and Damped Lyman α (DLA) absorbers, which are specified by an H I column density $N(\text{H I}) > 2 \times 10^{20}$ atoms/cm².

Table 1.1: Summary of the 4Hi-Q sub-surveys

Sub-survey	Minimum redshift	Related GTO surveys	Area [deg ²]
Mg II Blue Deep	0.55	S7 WAVES deep	66
Mg II Blue Wide	0.55	Wide area ^a	10 ⁴
Mg II Green	1.04	S8 CRS ELG	10 ³
DLA	2.59	Wide area	10 ⁴
LLS	3.77	Wide area	10 ⁴

^a The wide area consists of the following GTO surveys:
S8 CRS LRG, S6 AGN, S5 Clusters, S9 1001 MC.

For this task, both optically and X-ray selected quasars need to be considered.

Due to exposure time constraints, quasars with $r_{AB} < 21^m$, and $r_{AB} < 22^m$ (Mg II Blue Deep) are expected to be observable.

The quasars are required to have minimum redshifts of $z > 0.55$ to be in the background of the objects of interest (see tab. 1.1), which poses a problem as a priori knowledge of their redshifts is necessary.

Since spectroscopic redshifts have not yet been obtained for most of them^I, other means of deriving them in advance have to be considered^{II}.

Generally, the redshift z of a source can be inferred by analysing its spectral energy distribution (SED) as the restframe wavelength λ_0 of any spectral feature is stretched by a factor of $1 + z$. If the spectrum is not known with high accuracy, the redshift can be estimated via photometric data of the sources, which provide a coarse impression of the SED (further discussed in sec. 1.2). This process is more precise for galaxies than for quasars (further discussed in sec. 1.2.3) due to more recognisable features in the SEDs, but for the purposes of ensuring $z < 0.55$, the photometric redshift estimation of quasars should be sufficient for all sources where spectroscopic redshifts are not available.

Therefore, the goal of this thesis is to provide the pipeline for an input catalogue for the expected area (overlapping the area of the other surveys, compare tab. 1.1), ideally with reliable information on the SEDs of the objects. To incorporate photometric redshift estimation for the sources without spectroscopic redshifts available, a data pipeline is necessary which

- considers different parent catalogues for quasar sources,
- cross-matches these sources to catalogues with photometric data available, and
- performs the photometric redshift analysis with an optimal selection of templates and photometric bands.

^IIn fact, low-resolution spectroscopy of X-ray selected AGN is a goal of the S6 consortium survey, and the 4Hi-Q survey will improve the availability of spectroscopic redshifts as well.

^{II}A designated preparatory spectroscopic survey would be beyond the scope of the 4Hi-Q project.

1.2. Photometric redshifts

Note: *This section closely follows an overview on photometric redshifts provided by Salvato et al. (2019).*

As mentioned above, the redshift of an extragalactic source such as a galaxy can be inferred by analysing its spectral energy distribution (SED). The composition of the SED varies depending on the type of object observed, usually consisting of a continuum combined with specific emission and absorption lines.

Once a pair of known absorption or emission lines is identified, the precision of the redshift z only depends on the spectral resolution R (compare eq. (1.1)) and the signal-to-noise ratio (S/N) of the instrument used to obtain the SED.

For a resolving power of $R > 200$, redshifts can be measured with a precision better than 10^{-3} (see e. g. Le Fèvre et al. 2005).

Although quite a number of multi-object-spectrographs (MOSs) have been in operation (e. g. VLT/VIMOS, SDSS 2dF, KECK/DEIMOS, SUBARU/FMOS), their source coverage is never as exhaustive as it is for imaging surveys, which observe a multifold of the objects.

Another problem is that MOSs usually only observe in a very limited spectral range and have a low S/N ratio for faint objects. This can make it difficult to identify the two minimum characteristic features needed to measure the redshift. This leads to a success rate as low as 60 - 75 % for the spectroscopic redshift (spec- z) of the faintest of objects (for a detailed discussion of the demand for spec- z , even for photo- z projects, see e. g. Newman et al. 2015).

SEDs of galaxies and quasars

The spectra of galaxies are usually an overlay of stellar spectra within temperature ranges of 3000...40000 K and absorption and re-emission caused by gas and dust. Galaxies with high star formation rates usually appear brighter in the UV and FIR due to increased heating of the interstellar medium.

In addition to absorption and emission lines, there are two distinct breaks in the spectra of galaxies that can be identified in their SEDs, the first one being the **Balmer break**, the second one being the **Lyman break**.

The Balmer break occurs at around 3646 Å in the rest frame spectrum, as photons more energetic than that carry sufficient energy to ionise hydrogen atoms with electrons in the second energy state (bound-free absorption).

The same mechanism happens at the Lyman break below 1216 Å, where the photons become energetic enough to excite ground state hydrogen in the Lyman series. With wavelengths below the Lyman limit at 912 Å, they are able to completely ionise ground state hydrogen.

While these breaks are great features to identify in galactic spectra, they are usually not present in spectra of AGN (although they might be present in composite objects where galaxies host an active nucleus).

Most non-dwarf galaxies host a supermassive black hole in their central region, and in some galaxies, these cause a steady flow of infalling matter which is heated up due to friction. Due to their observational appearance, different sub-groups of AGN have been defined. Here, quasars are the most luminous AGN, outshining their host galaxies. Their luminosity can show high variability even on short timescales. Seyfert AGN are less luminous, and one differentiates between Seyfert 1 and Seyfert 2 AGN depending on the broadness of the emission lines in their optical spectra. Another type of AGN are those with high radio fluxes, usually hosted by ellipticals. Urry and Padovani (1995) proposed a widely accepted composite model, explaining the differing appearance of AGN with the inclination angle towards the observer. The fact that AGN consist of different components (such as the SMBH, accretion disk, corona, and obscuring torus) with varying sizes depending on the mass of the SMBH, and that their

appearance relies on the inclination angle leads to varying shapes of SEDs. Quasars in particular usually display a power-law-like continuum with strong emission lines caused by the infalling heated gas.

A sample of SEDs of both galaxies and quasars can be found in fig. 1.1.

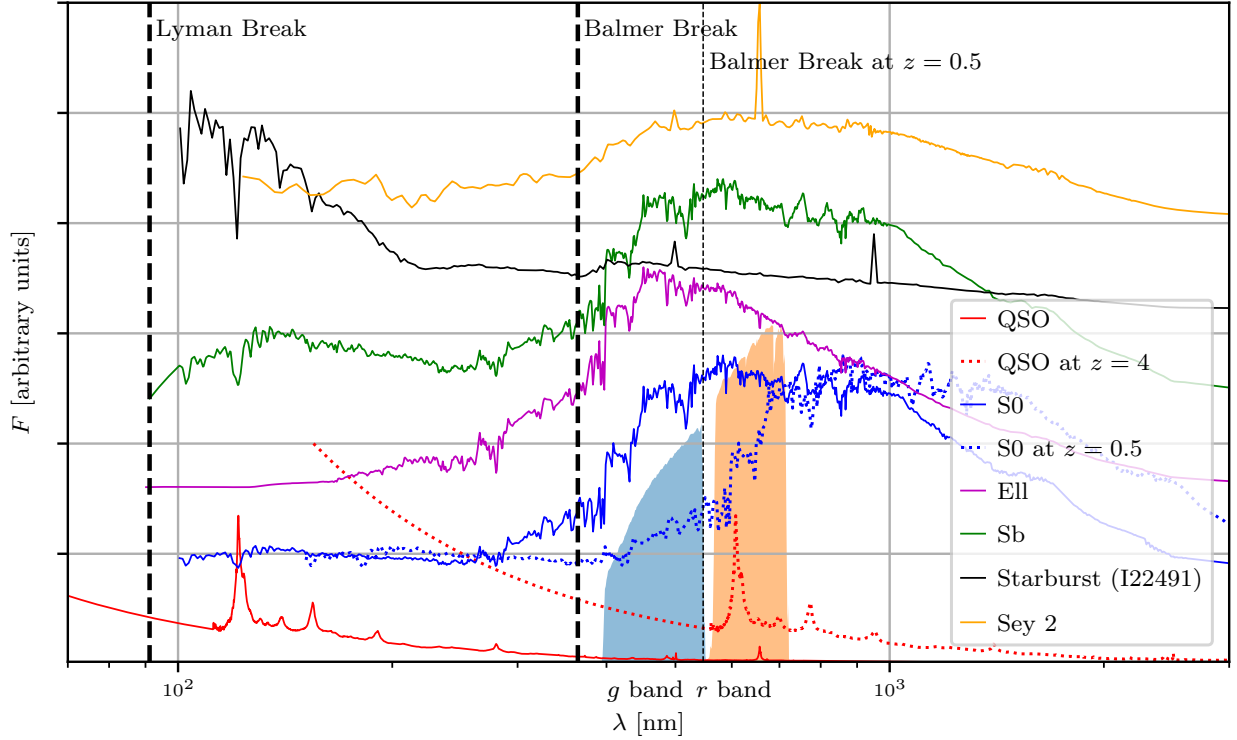


Figure 1.1: Comparison of SEDs of different objects (QSO: luminous AGN, S0: lenticular galaxy, Ell: ellipical, Sb: Spiral bar, Starburst, and Sey 2: Seyfert, from Ilbert et al. (2009), Salvato et al. (2009) and Brown et al. (2019)) with the inclusion of the g and r band transmission curves, inspired by fig. 1 in the review by Salvato et al. (2019).

For the galactic SEDs (e. g. S0), an increase of flux can be seen starting at the Balmer Break. For an S0 at $z \approx 0.5$, this break would lie right in between the g and r , leading to an increased $g - r$ colour^I. The same increase in colour could be observed for the AGN at a redshift of $z \approx 4$, as here the Lyman alpha emission line would be shifted into the r band. Note that for low redshifts, almost no $g - r$ colour features can be found in the power-law-like SED of the AGN.

1.2.1. Inferring redshift from photometry

Because the determination of spectroscopic redshifts is usually infeasible for the bulk of sources that are observed in all-sky imaging surveys, but needed as a selection parameter of input catalogues (as in the case of this thesis for 4Hi-Q), other means to derive it have been developed.

Given that extensive photometric information (that is, flux measurements in different broad-band filters) is available, a method called **photometric redshift** (photo- z) estimation can be applied. An early overview can be found e. g. in Koo (1999), a more recent, concise review has been provided by Salvato et al. (2019).

Here, the redshift (or rather a probability distribution $p(z)$ for the redshift) of a source is

^INote that the colour refers to magnitudes and not flux

estimated using its imaging information in the different filters as a crude low-resolution spectrum, relying on the fact that characteristic spectral features can be recognised by analysing the colour differences between filters.

This method was proposed by Baum (1957) to determine the redshift of elliptical galaxies at $z \approx 0.4$ ¹.

In their review, Salvato et al. (2019) suggest two general principles for surveys suitable for photometric redshift calculations:

1. The set of photometric bands should cover key features of likely sources at the redshift range of interest.

If key features such as the Balmer break are not included in the colours that can be calculated using the bands, it might be hard to have any indicators for the program, resulting in inaccurate estimations.

2. The amount of bands in different wavelength ranges should be as high as possible to minimise degeneracies.

If there would e. g. be a rapid increase in flux between the i and the z band, corresponding to a high $i - z$ colour, this could have a number of reasons:

The increase could be caused by the Balmer break typical for a starburst galaxy at redshift $z = 1.1$, an elliptical at the same or a higher redshift, or even just the appearance of an unfortunate strong emission line in a quasar spectrum with an arbitrary redshift. To break this degeneracy, as many different colours as possible are needed to narrow down the type of object and thus its redshift.

With this information, a mapping between the observed fluxes and the probability distribution of the individual redshifts can be obtained, although this process proves to be more tricky for exotic objects such as quasars (see sec. 1.2.3 for a short discussion).

There are several variations of the photo- z estimation technique which can broadly be divided into the following:

- **Template fitting:**

Empirical or synthetic template spectra are fit to the photometry by minimising the χ^2 or maximising the likelihood.

- **Machine learning:**

A neural network or random forest classifier is trained on a training sample to predict the redshift of a given photometry.

- **Hybrid methods:**

Both of the above approaches are combined, e. g. with a neural network supporting the selection of templates.

The performance of the photometric redshifts obtained via any method need to be validated before they can be used for scientific application (sec. 1.2.4).

In this work, the template-fitting method is utilised, which was first applied by Puschell et al. (1982) to estimate redshifts of radio galaxies.

1.2.2. Templates and template fitting

As the name suggests, for this branch of photo- z estimation, physically motivated templates of galaxy and stellar spectra are produced to map an observed flux distribution to redshift probabilities.

The basic idea is to model the physical processes for each likely type of object (extragalactic

¹Where, at the given time, no robust spectroscopy could be obtained, making this method a promising way to extrapolate to higher redshifts.

sources such as galaxies, quasars, or galaxies with quasar contributions) to generate a set of representative SEDs (such as the ones in fig. 1.1).

In addition to the extragalactic sources, stellar templates also need to be considered, as a Galactic origin of an observed source can not be completely ruled out. Usually, fits to both stellar and galactic templates are performed independently and the type of each source is determined afterwards.

The galaxy and quasar model SEDs can be obtained empirically, where they are taken from the observed spectra of such objects and extrapolated to missing parts of the spectrum, or in a theoretical way, where the SEDs are taken from stellar population synthesis models, requiring good comprehension of the underlying physical processes. They rely on several assumptions like e. g. the star formation history, which is linked to the adopted stellar evolution models and require a precise transformation to the observables (Maraston 2005).

In addition to having these observed or modelled templates, nebular emission lines by H II regions need to be considered, especially for low-redshift ($z \leq 1.5$) sources.

Furthermore, the effect of dust attenuation needs to be taken into account for $z > 1$, as the ultraviolet part of the SEDs (where the extinction and reddening effects are highest) is shifted to the optical or further.

Some template fitting codes tackle this problem by modelling the dust attenuation as an additional free parameter, adopting a proper attenuation law. With this approach, the templates can be corrected for the absorption in the intergalactic medium.

Apart from the intergalactic dust attenuation, one also has to account for the effects of the dust in the Milky Way. While most catalogues are already corrected for this, utilising common dust extinction maps, Galametz et al. (2017) propose an alternate method that takes the SED of each object into account when calculating the extinction.

After all of these corrections have been considered, each of the filter transmission curves needs to be applied to the templates via integration, producing a model flux representing how the template would appear in the filter in its rest frame.

A typical filter transmission curve needs to not only account for the physical filter, but also for all other possible modifications to the incoming light, including the Earth's atmosphere, telescope, instrument, and detector efficiency.

Then, the model fluxes are usually propagated with small redshift steps, and χ^2 or maximum likelihood fitting is performed to generate a redshift probability distribution for a given observed flux measurement.

1.2.3. A note on photo- z for quasars

With the rising interest in the role of active galactic nuclei¹ (AGN) in galaxy evolution, fuelled by reviews such as the one by Kormendy and Richstone (1995), redshifts of galaxies hosting AGN are becoming more and more important.

As the 4Hi-Q project shows, quasars can also serve as bright, non-transient background objects in absorption line surveys, and redshift information is necessary to ensure that they are in the background.

Although their redshifts can be obtained spectroscopically, it usually is an expensive task as

¹While quasars (which is short for *quasi-stellar radio sources*, sometimes also QSOs for *quasi-stellar-objects*) had at first been discovered separately, after realising that the broad emission lines in the spectrum of one of these objects were just redshifted Balmer and Mg II lines (Schmidt 1963), they were associated with the active nuclei of galaxies.

While there are other types of AGN than quasars (a brief summary and history is e. g. provided in Shields (1999)), for the purpose of this thesis, these terms are used interchangeably, keeping in mind that not all AGN are quasars.

the number densities are too high for single-slit spectroscopy but too low for multi-object spectroscopy with the sole goal of obtaining them¹.

Thus, setting out for photo- z determination is a logical step.

Unfortunately, this analysis is challenging because of multiple reasons:

- **Featureless SEDs:**

The typical SED of a quasar follows a power-law continuum with some emission lines as the only remarkable features (compare fig. 1.1). These emission lines are best identified with the use of intermediate and narrow-band filters.

The absence of strong features such as the Balmer break causes a large amount of degeneracy.

- **Contribution to the observed SED:**

The total observed SED has contributions from both the AGN component and the host galaxy in unknown proportions, leading to a drastic increase in required templates and possible degeneracies for the analysis.

- **Dependency on selection methods:**

The templates need to reflect the types of AGN, which depend on the depth, the selection technique (e. g. via variability or excess flux) and the band used for selection (X-ray selected AGN populations differ drastically from radio-selected ones).

- **Variability:**

Many AGN are intrinsically variable in intensity (see e. g. Wold et al. 2007) even in short time scales, which is a problem if the photometric data is collected over the course of several years with different bands being observed at different times.

Because of these problems, AGN-centred photo- z analysis is generally more inaccurate, albeit still possible as e. g. shown by Salvato et al. (2009) (who analysed the impact of variability) or Brescia et al. (2019) (who compared different techniques for X-Ray selected AGN).

One approach to improve the template-fitting process is to add priors, which take into account additional information such as the morphology.

1.2.4. Validation of photo- z

The best way to assess the performance of photometric redshift estimations is to compare them to redshifts obtained spectroscopically for a representative sub-sample of the sources.

There are four commonly used quantities to evaluate the performance of the photo- z method for N sources with spec- z and photo- z available:

- The **precision** σ_x as a measure for the scatter between predicted and true redshift, which is the standard deviation of $x := \frac{z_{\text{phot}} - z_{\text{spec}}}{1 + z_{\text{spec}}}$.

$$\sigma_x = \left\langle \frac{z_{\text{phot}} - z_{\text{spec}}}{1 + z_{\text{spec}}} \right\rangle_2 = \sqrt{\frac{1}{N} \sum_{i=1}^N (\langle x \rangle - x_i)^2}. \quad (1.3)$$

- The **accuracy** σ_{NMAD} , which is the Normalised Median Absolute Deviation that utilises the median, is used as an alternative to the precision throughout the literature:

$$\sigma_{\text{NMAD}} = 1.48 x_{\text{med}} = 1.48 \text{med}_i (|x_i|) = 1.48 \text{med}_i \left(\frac{|z_{\text{phot},i} - z_{\text{spec},i}|}{1 + z_{\text{spec},i}} \right). \quad (1.4)$$

- The **outlier fraction** η_{out} describes the fraction of sources which have been critically misjudged by the photo- z estimation.

¹although some of the 4MOST surveys are very promising on that front.

The criterion for categorising a source as an outlier varies throughout the literature. Most commonly (e. g. Salvato et al. 2009; Dahlen et al. 2013), sources with

$$|x| = \frac{|z_{\text{phot}} - z_{\text{spec}}|}{1 + z_{\text{spec}}} > 0.15 \quad (1.5)$$

are classified as outliers.

Thus,

$$\eta_{\text{out}} = \frac{N_{\text{out}}}{N} \quad (1.6)$$

is the outlier fraction with N_{out} being the number of outliers according to eq. (1.5).

An alternative classification using

$$|z_{\text{phot}} - z_{\text{spec}}| > n\sigma_x \quad (1.7)$$

with $n \in \{1, 2, 3\}$ being left as a choice to the authors is also possible.

Note that Shu et al. (2019) classify each source with $|x| > 3\sigma_{\text{NMAD}}$ as an outlier.

- The **Bias** B is the average separation of the predicted and true redshift:

$$B = \langle z_{\text{phot}} - z_{\text{spec}} \rangle = \frac{1}{N} \sum_{i=1}^N (z_{\text{phot},i} - z_{\text{spec},i}). \quad (1.8)$$

Throughout this thesis, of these quality measures, only the accuracy σ_{NMAD} and outlier fraction η_{out} are used as the bias does not add meaningful information.

In addition to these traditional metrics, it makes sense to assess the performance in terms of the goals for the input catalogue of 4Hi-Q:

- The **false positive fraction** ψ_{Pos} describes the fraction of sources with

$z_{\text{phot}} > 0.5$, but $z_{\text{spec}} < 0.5$.

These are sources that in reality are closer than estimated and would thus falsely be considered as input quasars.

It is hereafter defined as

$$\psi_{\text{Pos}} := \frac{N((z_{\text{spec}} < 0.5) \wedge (z_{\text{phot}} > 0.5))}{N}. \quad (1.9)$$

- The **false negative fraction** ψ_{Neg} describes the fraction of sources with

$z_{\text{spec}} > 0.5$, but $z_{\text{phot}} < 0.5$.

These are sources that in reality are further than estimated and would thus falsely not be considered as input quasars.

It is hereafter defined as

$$\psi_{\text{Neg}} := \frac{N((z_{\text{spec}} > 0.5) \wedge (z_{\text{phot}} < 0.5))}{N}. \quad (1.10)$$

It is important to note that all of these quantities can only properly assess the photo- z performance if the sub-sample with spectroscopy is representative of the entire sample - a bias (e. g. toward bright sources for the spectroscopy) is especially critical if the choice of templates is optimised on these quantities.

If spectroscopic redshifts are systematically missing for a certain kind of objects, this needs to be accounted for. Masters et al. (2015) suggest a statistical mapping between the multi-colour distribution and the available spectroscopy to identify the parameter space of the sources for which spectroscopy is missing.

If no spectroscopic information is available at all, there are still validation techniques relying e. g. on cross-correlation and neighbouring redshifts, but these are not part of this thesis. In addition to the pure comparison of the best-fit photo- z , it is advisable to take a closer look at the probability distribution functions as this can show signs of insufficient templates. Also, the second-best fit might in some cases point toward the true redshift.

2. Procedure: Photometric redshift via SED fitting

As outlined in the previous sections, the 4Hi-Q survey requires a catalogue of quasars with $z > 0.55$ that can be used for the absorption line spectroscopy. In addition to that, information about the expected SED for each source is required by 4MOST to estimate observation times and plan the survey appropriately.

Therefore, this thesis provides the pipeline for an input catalogue in the expected area in the southern hemisphere (compare tab. 1.1), along with information on the SEDs of the objects. To assemble the quasar sources, a parent catalogue of optically/mid-IR selected quasars is utilised. Using the photo- z estimation code provided by LePhare (sec. 2.1), each of the sources is fit to a template to estimate its redshift.

Since the amount of photometric bands severely limits the reliability of this task, a variety of catalogues carrying diverse photometric information for the area of interest are used.

As discussed in sec. 1.2.3, the determination of photo- z for quasars is inherently challenging due to the unknown contributions of AGN and host galaxy. Therefore, it is beneficial to additionally consider morphological information of the sources to restrict their parameter space of possible redshifts.

If, on the one hand, the appearance of a source in the sky is pointlike, which can be assumed if its two-dimensional imaging data is best fit by a point-spread function (PSF), it either is a luminous high-redshift source such as a quasar, or a Galactic source that can be fit by one of the stellar templates.

If, on the other hand, the appearance of a source is extended, which can be assumed whenever it can be resolved and a typical surface brightness profile (reflecting the typical components of galaxies, e. g. an exponential disc or De Vaucouleurs profile) is the best fit, the source is likely to be a resolved galaxy, implying a redshift of $z < 2$.

Limitation for testing

Since the catalogue data for the area of $> 10^4 \text{deg}^2$ is too large to handle efficiently, only a small portion of it is focused on first in which the necessary photometric data is already available alongside good spectroscopic coverage for validation of the photo- z .

A seemingly suitable area for this task is the eROSITA Final Equatorial-Depth Survey (eFEDS; Brunner et al. 2021) field, which is a region of ~ 140 square degrees that has been used to probe the performance of the eROSITA^I mission. It covers the region of $126^\circ \lesssim \text{RA} \lesssim 146^\circ$, $-3^\circ \lesssim \text{Dec} \lesssim 6^\circ$.

In addition to spectroscopy for verification being available, a catalogue of X-Ray selected quasars is available in this region, providing a complementary parent catalogue for future analysis.

Data pipeline

The data pipeline constructed in this work is built in such a way that it can easily be extended from this small testing area to the whole sky.

The tool of choice for handling the catalogue data is the STILTS^{II} package (Taylor 2006), in particular the **jython** implementation which makes use of common **python** syntax. STILTS is optimised for operations on large-scale tables.

To tie the input catalogue assembly to the processing with LePhare, a **python** (van Rossum 1995) script is provided, additionally making use of both the **pandas** (Wes McKinney 2010) and **matplotlib** (Hunter 2007) packages for data analysis and visualisation.

The following section is laid out as follows:

^Iextended ROentgen Survey with an Imaging Telescope Array, (Predehl et al. 2021)

^{II}Starlink Tables Infrastructure Library Tool Set, available [here](#).

- First, the main functionalities of the photo- z -estimation tool LePhare are discussed (sec. 2.1).
- Following that, the steps taken to cross-match and assemble the data are described, including descriptions of the considered catalogues and details on the matching techniques used (sec. 2.2).
- Then, the adopted parameters for the photometric redshift estimation are discussed, including information on the filters, templates, and fitting parameters (sec. 2.3).
- Although a thorough description of the data pipeline code is then provided in sec. 2.4, many of the underlying calculations are already described in sec. 2.2 and sec. 2.3.

2.1. Working with the photo- z code LePhare

There are numerous codes capable of producing reliable photometric redshifts, with each specialising on some aspects more than others.

One notable difference between them are the template libraries, which are partially created for each run with the EAZY code (which allows for linear combinations of different templates Brammer et al. 2010), are trained using a spectroscopic sample with the ZEBRA code (which utilises iterative techniques to find fitting templates Feldmann et al. 2006), or have to be provided externally for others.

While dust attenuation is already accounted for in the templates for some of these codes, others treat it as a free parameter.

A code notable for the inclusion of galaxy emission lines and good performance with quasar templates is LePhare (Arnouts et al. 1999; Ilbert et al. 2006), which is therefore used throughout this thesis. Apart from that, LePhare allows for the use of luminosity priors to rule out unlikely redshifts, and offers a way to automatically correct for intergalactic extinction.

More precisely, LePhare is a set of Fortran (or, in the newer version LePhare++, C++, used throughout this thesis) programs specifically designed for photometric redshift calculation and analysis with the capabilities to

- read and process filter libraries (sec. 2.1.1) and provide extinction information (sec. 2.1.2),
- read a collection of SEDs to build a SED template library (sec. 2.1.3),
- compute the apparent magnitudes of each template in each filter for each redshift step, along with dust attenuation and nebular emission line parametrisation (sec. 2.1.4), and finally
- perform a fit of given observed fluxes to the templates using a χ^2 fitting process (sec. 2.1.5).

An overview of LePhare’s structure and the usual data flow is laid out in fig. 2.1.

In addition to these general features, the code comes with several tools to analyse and improve the quality of the photometric redshift output.

The coming sections describe the main steps in detail, following the structure of the LePhare++ manual^I.

Note that all of the parameters specified in each of the commands may be supplied via the configuration file, which is named **baseline_in.para** in the examples.

2.1.1. Reading and processing filter libraries

As the flux measurements of the observed sources have been performed with filters unique to each telescope, the respective transmission functions need to be accounted for when perform-

^IAvailable [here](#).

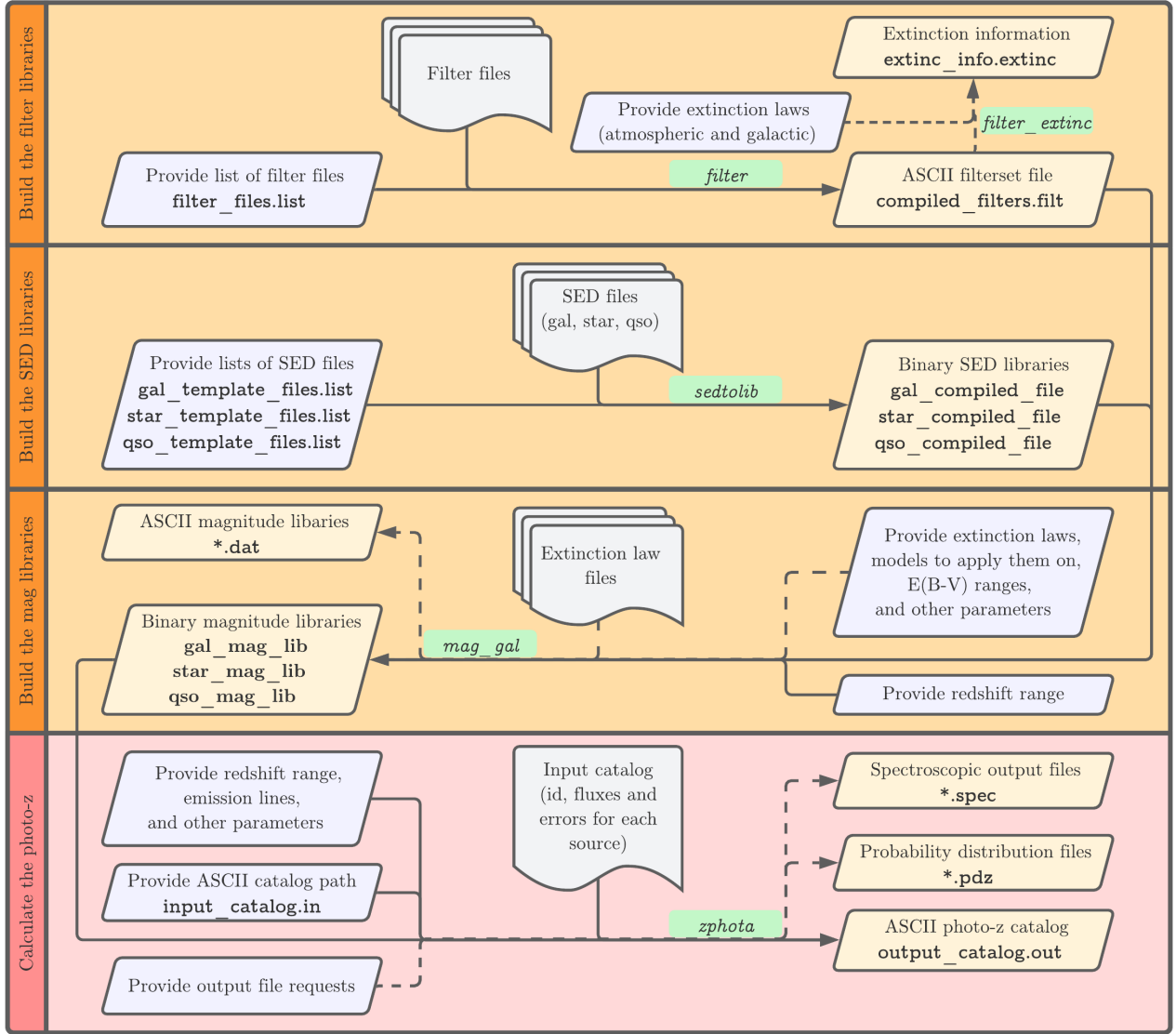


Figure 2.1: Overview of the LePhare package structure (adopted from fig. 1 of the LePhare manual).

Executable names (located in `$LEPHAREDIR/source`) have a light green background.

Dashed arrows indicate optional routines.

ing the template fits.

Thus, each SED template has to be processed using these transmission functions.

LePhare already provides several sets of filter response curves from a variety of telescopes, but one can also add custom filters corresponding to the ones used for the photo- z analysis.

Usage of the filter-routine in LePhare

Via the **filter** command of LePhare, a given list of such filters needs to be compiled into a single library.

While a filter transmission curve $T(\lambda)$ is generally dimensionless, where each value maps a transmission factor to the corresponding wavelength, these factors can either refer to a transmission in energy or in photons.

In a filter, the observed magnitude m_X (relative to Vega with flux F_{Vega}) of a source with flux distribution $F_X(\lambda)$ can be calculated using

$$m_X = -2.5 \log_{10} \left(\frac{\int F_X(\lambda) R(\lambda) d\lambda}{\int F_{\text{Vega}}(\lambda) R(\lambda) d\lambda} \right). \quad (2.1)$$

Here, $R(\lambda)$ is the response function of the instrument. If the transmission curve refers to energy, $R(\lambda) = T(\lambda)$.

This is usually not the case, as in real photometric measurements, photon counting detectors are used.

If the given transmission curve corresponds to photon counts N_γ , a transformation has to be applied to the integrals. Since the total number of photons can be obtained via

$$N_\gamma = \int \frac{F(\lambda)}{h\nu} d\lambda = \int \frac{F(\lambda)\lambda}{hc} d\lambda, \quad (2.2)$$

one can calculate the observed magnitude as

$$m_X = -2.5 \log_{10} \left(\frac{\int F_X(\lambda) \lambda T(\lambda) d\lambda}{\int F_{\text{Vega}}(\lambda) \lambda T(\lambda) d\lambda} \right). \quad (2.3)$$

Thus, for photon count transmission curves, $R(\lambda) = \lambda T(\lambda)$.

In LePhare, the transmission type tt of the provided transmission curves can be specified via the keyword **TRANS.TYPE**, leading to the calculation of

$$R(\lambda) = \left(\frac{\lambda}{\langle \lambda \rangle} \right)^{tt} T(\lambda). \quad (2.4)$$

Thus, $tt = 0$ corresponds to the energy type while $tt = 1$ corresponds to the photon count type.

According to the manual, this modification is more important for filters with longer wavelengths, but usually not the prevalent source of errors.

Instead, errors regarding the quantum efficiency of the charged-coupled devices (CCDs), telescope transmission and others usually dominate the uncertainty.

Other than accounting for the transmission type, LePhare offers functionality to apply calibrations via a correction factor f_c using the **FILTER.CALIB** keyword, where LePhare assumes no correction ($\frac{B(\nu)}{B(\nu_0)} = 1$, see below). For longer wavelengths such as in the WISE filters, a filter calibration corresponding to a flat spectrum is recommended.

The correction factor is derived from the reference spectrum $B(\nu)$ used to calibrate the filter, it is defined as

$$f_c = \frac{\int R(\nu) d\nu}{\int \frac{B(\nu)}{B(\nu_0)} R(\nu) d\nu} = \frac{\int \frac{1}{\lambda^2} R(\lambda) d\lambda}{\frac{1}{\lambda_0^2} \int \frac{B(\lambda)}{B(\lambda_0)} R(\lambda) d\lambda}, \quad (2.5)$$

where λ_0 is the effective wavelength of the filter calculated via

$$\lambda_0 = \langle \lambda \rangle = \frac{\int R(\lambda) B(\lambda) \lambda d\lambda}{\int R(\lambda) B(\lambda) d\lambda}. \quad (2.6)$$

Further details concerning this process are outlined in the manual. For this work, no correction factor is used for any bands except for the WISE ones, where it is set to 1.

An example code for the compilation of the necessary filter information might look as follows:

```
$LEPHAREDIR/source/filter -c baseline.in para \ # Name of the configuration
file
-FILTER_LIST baseline.filters.list \ # list of the desired filters
-TRANS_TYPE 1 \ # for photon transmission
-FILTER_CALIB 0,0,0,0,1,1,1,1 \ # Filter calibration with 1 for longer
wavelengths
-FILTER_FILE baseline.filters.compiled.filt
```

After creating such a filter file, its contents can be viewed using the command

```
$LEPHAREDIR/source/filter_info -f baseline.filters.compiled.filt
```

2.1.2. Providing extinction information¹

As outlined in sec. 1.2.2, it is important to account for atmospheric and galactic extinctions when applying the fits.

LePhare already provides a set of extinction curves which mostly differ around the 2170 Å bump.

Usage of the `filter_extinc`-routine in LePhare

A utility program to provide an overview of the extinction information can be used via the `filter_extinc` command, where the atmospheric extinction and galactic extinction curves can be specified using the keywords `EXT_CURVE` and `GAL_CURVE`.

An example use case might look as follows:

```
$LEPHAREDIR/source/filter_extinc -f baseline_filters_compiled.filt \
-e extinc_eso.dat \ # file to use for atmospheric extinction
-g SB_Calzetti.dat \ # file to use for galactic extinction
-o extinction_info.extinc # File to save the information in
```

2.1.3. Building an SED template library

As described in sec. 1.2.2, a set of SED templates representative of the population of objects is needed to perform the photometric redshift estimation via template fitting.

LePhare already provides sets of templates for stars, galaxies and QSOs gathered from a variety of releases (e. g. Salvato et al. 2009).

LePhare generally expects each template file to come in an ASCII format of $(\lambda [\text{\AA}], F(\lambda) [\text{erg/s/\AA/cm}^2])$, although it allows for the use of output files from stellar population synthesis models (Pegase and BC03, see Le Borgne et al. 2011).

Usage of the `sedtolib`-routine in LePhare

The `sedtolib` command compiles a given list of template SEDs of a certain type (stars, galaxies, or QSOs) into one single binary file, in addition to providing a documentation file and (for galaxies only) a file with physical information.

For example, the code for compiling a list of galactic templates might look as follows:

```
$LEPHAREDIR/source/sedtolib -t G \ # Here, the SED type (galaxy) is
    specified
-c baseline_in.para \ # Name of the configuration file
-GAL_SED baseline_templates.list \ # Path of the file that stores the
    references to the templates
-GAL_LIB baseline_templates_compiled_file # Name of the output file
```

2.1.4. Template processing and the creation of magnitude libraries

After the SED library for a given type (galaxy, star or QSO) has been created, the magnitude libraries needed can be prepared for the fitting process.

This step includes the propagation of each SED with redshift and (if requested) the inclusion of extinction laws.

¹Note: At the time of writing this thesis, this part of the program had not been implemented into the C++ version of LePhare. This section thus corresponds to the Fortran version.

Usage of the `mag_gal`-routine in LePhare

For these tasks, LePhare offers the `mag_gal` routine, which requires both a filter file (sec. 2.1.1) and the combined SED library (sec. 2.1.3) as a minimal input.

Using this command creates a library of magnitudes in each filter for each model and redshift step.

One can also apply extinction laws with varying E_{B-V} values using the `EXTINC_LAW` parameter to specify the laws, `MOD_EXTINC` to specify which models to use them on, and `EB_V` to specify the requested reddening colour excess E_{B-V} values.

In addition to that, LePhare can generate nebular emission lines in case the models themselves did not already account for them. By specifying `EM_LINES`, one of three different modes can be chosen, treating the calculation in a slightly different way (laid out in the LePhare manual). An important thing to note is that the magnitude or flux library can quickly exceed the disk space as its number of rows N_{rows} is determined by the number of models N_{Models} , the age range N_{ages} (usually not considered), the number of redshift steps $N_{z \text{ step}}$, the number of extinction laws $N_{\text{Ext laws}}$ and the number of E_{B-V} values.

An upper limit^I of N_{rows} can roughly be estimated via the formula

$$N_{\text{rows}} = N_{\text{Models}} N_{\text{ages}} N_{z \text{ step}} N_{\text{Ext-laws}} N_{E(B-V)}. \quad (2.7)$$

The code to generate a magnitude library with magnitudes for each of the galactic models might look as follows:

```
$LEPHAREDIR/source/mag_gal -t G \ # Again, the type of object (G for galaxy
) is specified
-c baseline_in.para \ # Name of the configuration file
-Z_STEP 0.04,0.6 \ # Redshift steps from z=0 to z=6 with dz=0.04
-FILTER_FILE baseline_filters_compiled.filt \ # Name of the filter file
-GAL_LIB_IN baseline_templates_compiled.file \ # Name of the model library
-GAL_LIB_OUT baseline_templates_mag_lib \ # Name of the output file
-EXTINC_LAW SMC_prevot.dat,SB_calzetti.dat \ # Name of the requested
extinction law
-MOD_EXTINC 1,5,4,8 \ # Ranges of models these extinction laws need to be
used on
-EB_V 0.,0.1,0.2,0.3 \ # Requested E(B-V) values
-LIB_ASCII YES \ # A well readable output file is additionally generated
```

2.1.5. Fitting photometric redshifts

After the fluxes $F_{\text{temp}, i}$ for each model, redshift step, and band i have been created, the fitting process for the redshift estimation can be initiated.

Here, the observed fluxes $F_{\text{obs}, i}$ and uncertainties σ_i in each band are compared to the template fluxes using χ^2 as a measure:

$$\chi^2 := \sum_i \left(\frac{F_{\text{obs}, i} - s F_{\text{temp}, i}}{\sigma_i} \right)^2. \quad (2.8)$$

Since the template fluxes are normalised, a scaling factor s is introduced to scale them accordingly.

^Isince the extinction laws are usually only applied to a range of the models

As s needs to minimise χ^2 , it can be derived by setting $\frac{d\chi^2}{ds} = 0$ using the chain rule¹:

$$\begin{aligned} \frac{d\chi^2}{ds} &= \sum_k (-2F_{\text{temp}, k}) \left(\frac{F_{\text{obs}, k} - sF_{\text{temp}, k}}{\sigma_k^2} \right) \stackrel{!}{=} 0 \\ \Leftrightarrow \quad s \sum_j \frac{F_{\text{temp}, j}^2}{\sigma_j^2} &= \sum_k \frac{F_{\text{obs}, k} F_{\text{temp}, k}}{\sigma_k^2} \\ \Leftrightarrow \quad s &= \frac{\sum_k (F_{\text{obs}, k} F_{\text{temp}, k} \sigma_k^{-2})}{\sum_j (F_{\text{temp}, j}^2 \sigma_j^{-2})}. \end{aligned} \quad (2.9)$$

To start the photo- z analysis in LePhare, careful preparation of the input catalogues is necessary.

For each source, an identification number, the fluxes (set to $-99.$ if not available and provided either as a magnitude in AB or VEGA, or as a flux in $\text{erg/s/cm}^2/\text{Hz}$), and the errors for each of the bands have to be provided. Additionally, a quantity called **CONTEXT** might be specified along with the spec- z and additional information.

The context $C \in \mathbb{N}$ encodes the information which of the bands should be used for the calculation of the photometric redshifts, defined by

$$C := \sum_{i=1}^N 2^{i-1} \delta_{ij} \quad j \in I. \quad (2.10)$$

Here, N is the total number of bands used for the analysis and I the set of band numbers requested for use.

Thus, each band is simply assigned a unique exponentiation of 2. The values of the requested bands are then summed up for easier storage.

Example for the context

If the available passbands for the analysis were the g , r , z , W_1 , W_2 , W_3 , and W_4 bands, they would have the following band contexts C_i :

Band	g	r	z	W_1	W_2	W_3	W_4	Total context C
Band number i	1	2	3	4	5	6	7	-
Band context $C_i = 2^{i-1}$	1	2	4	8	16	32	64	127
Example band selection	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	27

If for example only the g , r , W_1 and W_2 bands are supposed to be considered for a certain analysis (i. e. to check whether the photo- z really worsens in comparison to full band usage), a context of $C = 27$ corresponding to these bands could be provided.

Such a context can be set both globally and on a source by source basis.

Although a context value referring to it might suggest that a band is always used, its photometric information is not considered if both flux and error are negative.

Usage of the `zphota`-routine in LePhare

As the calculation of photo- z is the core task of LePhare, its **zphota**-routine that performs the χ^2 analysis can be used with a variety of options.

One of the most important options is the consideration of prior information for the fitting process.

In addition to the band selection via the context keywords, a redshift probability distribution function (PDF) can be requested for each source using the **PDZ.OUT** parameter along with the

¹and seeing that $\frac{d\chi^2}{ds} = \sum_k \frac{F_{\text{temp}, k}^2}{\sigma_k^2} > 0$, implying that eq. (2.9) indeed minimises χ^2

name of an output file.

Apart from that, LePhare provides the option to create an extra output file for each individual source by setting the **SPEC_OUT** parameter to **YES**. These output files can be used for further analysis as they provide detailed information on the photometry, the PDF, and the best-fit templates.

When carrying out the photo- z analysis, an example code might look as follows:

```
$LEPHAREDIR/source/zphota -ZPHOTLIB compiled_template_file_gal,
    compiled_template_file_star \ # Names of the libraries used
-c baseline_in para \ # Name of the configuration file
-CAT_IN ascii_input_catalog.in \ # Name of the input catalog
-CAT_OUT ascii_output_catalog.out \ # Name of the output catalog
-PARA_OUT baseline_out para \ # Name of the parameter file used for the
    output catalog
-ZPHOTLIB baseline_templates_mag_lib,star_flux_redshift_grid \ # Name of the
    files with the propagated templates
-ERR_SCALE 0.05,0.05,0.1,0.1,0.2,0.2,0.3,0.4 \ # Errors for the given
    filters
-GLB_CONTEXT 27 # only consider certain bands
```


2.2. Preparing the input catalogue for LePhare

To maximise the quality of the photometric redshifts, data from different surveys and thus input catalogues has been utilised. The focus was set on data available for the whole portion of the southern hemisphere planned to be observed by the 4Hi-Q survey, with tests conducted in the eFEDS area.

Therefore, optically- and mid-infrared selected quasars (sec. 2.2.1) are used as a baseline.

Based on their right ascension and declination, they are cross-matched to photometric data in the optical and infrared (sec. 2.2.2), in the near-infrared (sec. 2.2.3), and in the ultraviolet (sec. 2.2.4) range of the spectrum, as well as to spectroscopic data for validation (sec. 2.2.5). In addition to that, *i* band data from different supplementary surveys has been considered to assess a possible performance improvement, since an upcoming data release of the legacy surveys is going to feature *i* photometry (sec. 2.2.6 and sec. 2.2.7).

After the matched catalogue is set up, the data is processed and split into two sub-catalogues based on the morphology of the sources (depending on the model used to fit them in LS data), as the distinction between pointlike and extended sources increases the quality of the photo-*z*-fitting.

An overview of the general data flow is provided in fig. 2.2.

An analysis of the distribution of available input photometry is given in sec. 2.2.8, while the filter properties and therefore spectral coverage are discussed in sec. 2.3.1, see fig. 2.12.

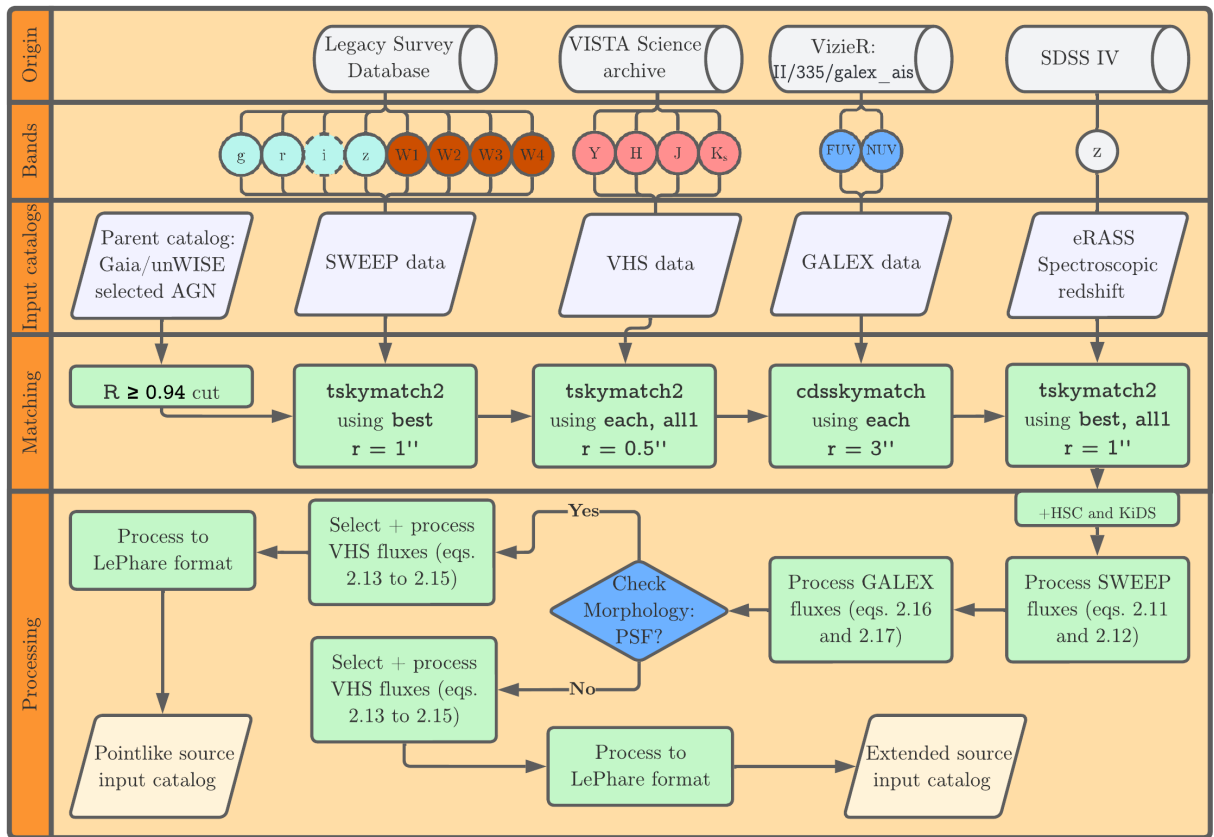


Figure 2.2: Overview of the data pipeline.

The *i* band has only partially been considered as a benchmark, indicated by the dashed line, HSC and KiDS data have only partially been added.

The referenced surveys and connected equations are explained in the respective sections below.

2.2.1. The AGN parent catalogue

Optically and mid-infrared selected quasars

As a parent catalogue to draw the sources from, a catalogue of AGN candidates (Shu et al. 2019) selected using a random forest classifier from the Data Release 2 of the Gaia mission and from the unWISE catalogue of the Wide-field Infrared Survey Explorer (WISE) is used. As a result of the classifier, each source is assigned a probability P_{RF} representing the likelihood of it belonging to the AGN class. In order to have a sample with 85 % reliability, one can use their catalogue with 85 % completeness^I and impose a cut at $P_{\text{RF}} \geq 0.94$.

Of the 2 734 464 sources in the catalogue, 2 182 193 (79.8 %) have $P_{\text{RF}} \geq 0.94$. Of these, roughly 13 418 are located in the eFEDS footprint ($126^\circ \lesssim \text{RA} \lesssim 146^\circ$, $-3^\circ \lesssim \text{Dec} \lesssim 6^\circ$).

In addition to the AGN classification, the Shu et al. (2019) already provided photometric redshift estimates using a random forest regressor on the Gaia and WISE magnitudes and colours. The accuracy (eq. (1.4)) of these redshifts is $\sigma_{\text{NMAD}} = 0.11$ (related to the galaxy-dominated galaxy sample).

While these might already be sufficient to distinguish between quasars at redshifts > 0.55 and quasars at lower redshifts, no SED information is supplied as the photo- z estimation was conducted using machine learning.

Their random forest photo- z (fig. 2.3) are a good benchmark to compare the results of this thesis to.

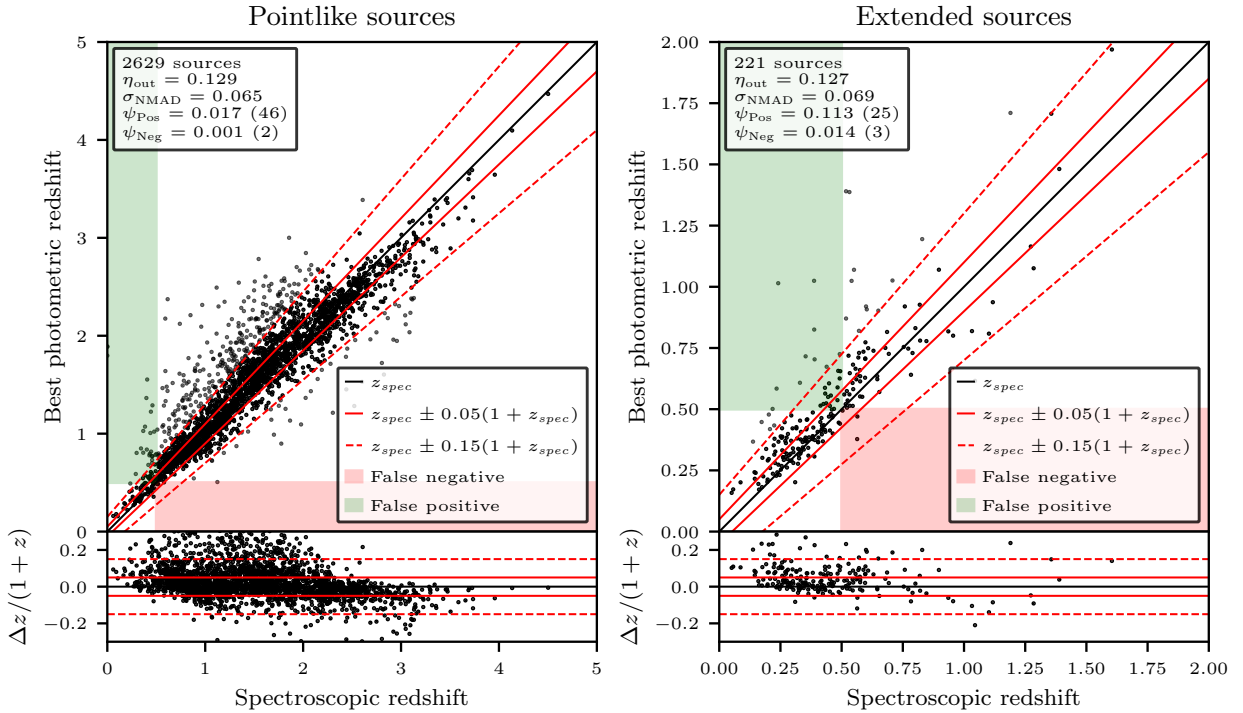


Figure 2.3: Results of Shu et al. (2019) for the sources in the eFEDS field adopting the random forest regressor.

The plot shows the relation between spec- z and photo- z for each of the sources, along with the quality measures introduced in sec. 1.2.4.

^Iavailable at https://www.ast.cam.ac.uk/~ypshu/AGN_Catalogues.html

2.2.2. Optical and infrared: The DESI Legacy Imaging Surveys

The DESI Legacy Imaging Surveys (Schlegel et al. 2021, Legacy Surveys/LS hereafter) unite three public imaging projects (the Beijing-Arizona Sky Survey (BASS), the Dark Energy Camera Legacy Survey (DECaLS), the Mayall z band Legacy Survey (MzLS)) to image the sky visible from the northern hemisphere (although sufficient data in the southern hemisphere is available as well due to the location of DECaM). Apart from calibrated images, a catalogue constructed using a probabilistic inference-based approach to estimate source shapes and brightness is provided, including both photometry from the grz bands and the mid-infrared bands of the Wide-field Infrared Survey Explorer (WISE) satellite.

At the time this thesis was started, the ninth Data Release (DR9) was the latest Data Release (DR) available.

In May 2022, a preliminary catalogue for the tenth DR that additionally includes some i band data for a portion of the sources observed with DECaM, was made available.

Since the first photo- z -results that only considered the LS DR9 data had high outlier fractions, analysis making use of i data of similar surveys was carried out, awaiting the DR10 data.

By analysing HSC and KiDS (see sec. 2.2.6 and sec. 2.2.7) data, it could be shown that the photo- z quality can be significantly improved if i band data is taken into account (discussed in sec. 3.3). Therefore, the data pipeline and data analysis are designed such that they are going to suit the data of the upcoming full tenth data release.

Target selection and photometry

In the Legacy Surveys, the imaging data is post-processed through the so-called *legacypipe*¹, which can process one $0^\circ 25' \times 0^\circ 25'$ region (called a *brick*) at a time.

For this, it takes the CCDs used to obtain the imaging into account, performs background subtraction and PSF analysis (using PSFEx by Bertin (2011)) for each CCD, and creates a weighted image stack for the source detection. Then, sources are detected using a thresholding algorithm, and modelled by *The Tractor* (Lang et al. 2016).

This modelling is performed as follows:

For each source, different parametric light profiles (a delta function for point sources, a de Vaucouleurs profile $\propto r^{-\frac{1}{4}}$, an exponential disk, and a composite profile with both de Vaucouleurs and exponential contributions) are fit to the pixel-level data of all images (in the grz bands) containing the source.

The best-fit model is found by minimising the residuals for all images. It is then assumed that the model is the same for all bands, implying that a point source's photometry will be based on a PSF model, while extended sources are modelled using the same light profile to extract their position, shape parameters, and photometry.

The photometry for the mid-infrared WISE bands is determined afterwards by forcing location and shape of the model while convolving with the WISE PSF for the fit to the stacked image.

Thus, the LS sources are selected using the high resolution of the optical bands but might therefore lack photometry of sources in the WISE bands that have no optical counterparts.

Notable shortcomings of *The Tractor* are the modelling of spatially large, extended sources with complex morphologies and crowded fields, as well as possible detections of spurious sources such as particle events or satellite trails. Also, a small number of sources with particularly low surface brightness or small distances to brighter stars or galaxies are not detected.

¹Available at <https://github.com/legacysurvey/legacypipe>.

Working with the catalogue data

Since the LS output catalogue is split up in bricks of 0.25×0.25 that are available at NERSC^I, additional efforts are made to join them for single-catalogue usage.

To minimise the data required to be downloaded, only the reduced *SWEEP* catalogues^{II}, which contain only the most important columns of *The Tractor* output catalogues in *.fits* files for each brick, are considered.

Each brick is characterised by the right ascension (RA) and declination (Dec) ranges in the following format:

sweep–<brickmin>–<brickmax>.fits, where <brickmin> = <AAA>c<BBB> encodes the minimum of RA (<AAA>) and Dec (c<BBB>, having c $\in \{p, m\}$ for *plus* or *minus* denote a positive or negative declination).

As an example, **sweep**–**010m035**–**020m030**.fits corresponds to a region with $10^\circ \leq \text{RA} \leq 20^\circ$ and $-35^\circ \leq \text{Dec} \leq -30^\circ$.

To collect all necessary catalogue data in a specific region, multiple of these files need to be downloaded, taking this pattern into account. The code for selecting and downloading all necessary files for a region can be found in sec. A.1.

After the catalogues have been downloaded, they can be merged into a single file using STILTS, resulting in 16 673 219 sources for the eFEDS region. The fluxes $F_{X, \text{SWEEP}}$ for a given band X correspond to the model fluxes determined by *The Tractor*. They are provided in Nanomaggies^{III} and are not corrected for extinction. For extinction correction, the catalogue provides information on the transmission via the **MW_TRANSMISSION_X** =: M_X columns.

Thus, the corrected flux $F_{X,c}$ converted to the desired unit erg/s/cm²/Hz is calculated by applying the following transformation with the corresponding error:

$$F_{X,c} = \frac{F_{X, \text{SWEEP}}}{M_X} 3.631 \cdot 10^{-29} \frac{\text{erg/s/cm}^2/\text{Hz}}{\text{Nanomaggies}} \quad (2.11)$$

$$\sigma_{F_{X,c}} = \frac{1}{\sqrt{\text{IVAR}_{X, \text{SWEEP}}}} \frac{1}{M_X} 3.631 \cdot 10^{-29} \frac{\text{erg/s/cm}^2/\text{Hz}}{\text{Nanomaggies}}. \quad (2.12)$$

Here, $\text{IVAR}_{X, \text{SWEEP}}$ is the inverse variance of the fluxes that is provided in the catalogues.

Morphological classification

Along with the fluxes in the different filters, the LS catalogues contain information on the profile that was adopted to fit the photometry using *The Tractor* for each source, reflecting the morphology. This information is provided in the **TYPE** column.

For the DR9, five morphological types are possible to have been used for fitting:

Point sources (*PSF*), round exponential galaxies with variable radius (*REX*), De Vaucouleurs (*DEV*) profiles, exponential (*EXP*) profiles, and Sersic (*SER*) profiles. Apart from that, *DUP* denotes Gaia sources that are coincident with extended sources which do not have any optical flux assigned to them.

For the analysis in this thesis, this morphological information is used to split the input catalogue for LePhare into two parts.

All sources having *PSF* as their **TYPE** are assumed to be pointlike sources (which could either be stars or very distant galaxies with a luminous quasar component), while all sources that are associated with other **TYPE**s are assumed to be resolved (extended) galaxies at lower redshifts.

^I<http://legacysurvey.org>

^{II}Available at <https://portal.nersc.gov/cfs/cosmo/data/legacysurvey/dr9/south/sweep/> for DR9 in the southern hemisphere

^{III}A Nanomaggie is defined as $3.631 \cdot 10^{-23} \cdot 10^{-9}$ eerg/s/cm²/Hz, see also [the SDSS website](#) for further explanation.

Matching to the parent catalogue

The data in the joint SWEEP catalogues is cross-matched to the parent catalogue using the **tskymatch2** routine provided by STILTS.

Since the LS coverage is the most extensive of the catalogues both in source density and photometry, and since the photometry of the other catalogues would yield unsatisfying photo- z -results if used on its own, all sources of the parent catalogue not having a counterpart in the LS data are discarded, but since Shu et al. (2019) is based on the LS data, all sources are found to have a counterpart, resulting in 13 418 sources with LS photometry.

For the matching task, a symmetric nearest-neighbour best match was applied such that each source in each table can at most be used once. In addition to that, an inner join was performed to discard all sources without counterparts.

The maximum separation for allowing two objects to be matched was set to just $r_{\max} = 0''.01$ (due to the same origin), yielding almost identical RA and Dec values^I.

2.2.3. Near-Infrared: The VISTA Hemisphere Survey

As can be seen for example in fig. 4a) in Salvato et al. (2019), the quality of photo- z improves considerably by incorporating near-infrared (NIR) bands for the redshift fitting process.

Thus, NIR data of the VISTA Hemisphere Survey (McMahon et al. 2013, , VHS hereafter), which is a survey carried out at the Visible and Infrared Survey Telescope (VISTA), is used. The aims of the VHS were to observe the southern hemisphere (~ 20000 square degrees) in at least two NIR wavebands such that median 5σ point source limits (Vega) of $J = 20.2^m$ and $K_s = 18.1^m$ could be achieved. Complementary data was taken from other public VISTA surveys to accomplish sufficient coverage. Along with the J and K_s bands, the Y and H bands have been observed.

Apart from studying properties of the Milky Way, one of the scientific goals was to observe the most luminous quasars at all redshifts in the southern hemisphere.

Target selection and photometry

In VHS, the source detection and extraction software (Irwin et al. 2004) uses background-subtracted aperture fluxes for each detected source with circular apertures.

The catalogues provide different kinds of photometry, which is especially important for extended sources as the surface brightness profiles of resolved galaxies can be quite complex and do not have sharp edges. Available types of photometry are Petrosian (Petrosian 1976) photometry^{II}, 2D Sersic profile photometry and simple aperture photometry with different radii of $2''$, $2''.8$ and $5''.7$. When deriving the photo- z for X-ray-selected AGN candidates, Ananna et al. (2017) found that the best results were achieved when using the $2''.8$ radius aperture magnitude (**AperMag4** in VHS) for their pointlike sources and $5''.7$ (**AperMag6** in VHS) for the extended sources. Since they are considering X-ray-selected AGN, this might not be optimal for the photo- z in this thesis, but it is reasonable to use them as a starting point.

Working with the catalogue data

The data products of the VHS are publicly available at the VISTA Science Archive^{III}, accessible via free-form SQL queries.

Therefore, the data necessary for this work has been retrieved using the query provided in sec. A.2, resulting in 1 406 548 sources in the eFEDS area.

In the data pipeline, the catalogue data is again matched using a symmetric best nearest-

^IThe maximum positional offset for a matched source is of order $10^{-9''}$, implying that these are just numerical artefacts.

^{II}a brief, comprehensive explanation can be found e. g. [here](#).

^{III}<http://horus.roe.ac.uk/vsa/>

neighbour match via the `tskymatch2`-routine of STILTS.

Without extinction and with sufficient depth and accuracy of the VHS, a counterpart for each AGN in the LS data should be available in its data, but this is not the case.

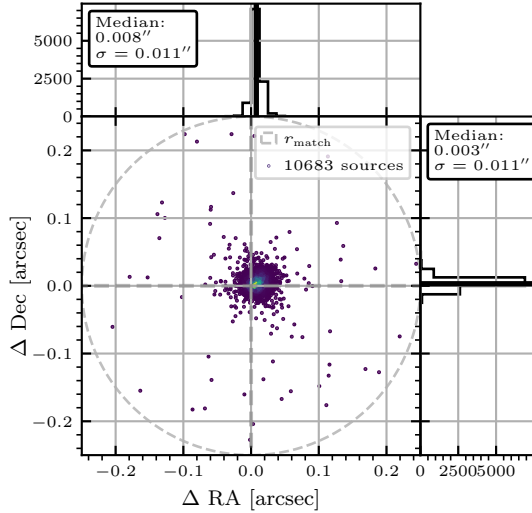
Keeping only the sources with photometry in all bands would severely limit the number of available sources, so instead of performing an inner join as done between the LS data and the parent catalogue, a left join is applied, keeping all sources independent of whether a VHS counterpart is found.

In the tests conducted in the eFEDS area, out of the 13 418 AGN available in the LS sample, 3 721 were at first found to have a counterpart in the VHS sample with a matching radius of $r < 0''.5$ (see fig. 2.4).

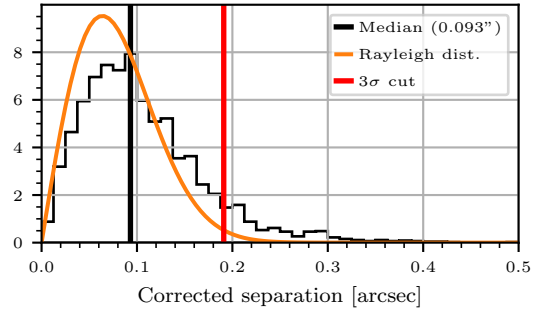
An analysis yields median positional offsets of $0''.044$ in the direction of the right ascension and $0''.034$ in the direction of declination.

After correcting the individual offsets by these medians, the standard deviation of the corrected offsets is $\sigma = 0''.064$.

To ensure that only true matches are considered for the photometric redshift estimation, VHS photometry of all sources at a corrected separation of more than 3σ is discarded, leading to only 3 393 matches with a corrected separation less than $3\sigma = 0''.19$ being taken into account for further analysis. Thus, VHS photometry is only available for around 25 % of all sources. A reason for this shortcoming is that the VHS data is shallower at a 5σ detection limit of $L_{AB,K} \sim 20.3^m$ in the K_s band.



(a) Positional offsets of VHS and LS sources



(b) Corrected offsets and the corresponding Rayleigh distribution

Figure 2.4: Positional offsets from the VHS DR6 data to the LS DR9 data for the 3 721 nearest-neighbour matched sources in the eFEDS field using a matching radius of $0''.5$.

The left panel shows the distributions of offsets in RA and Dec, while the right panel shows the distribution of the separation of sources after they have been corrected by the median offsets in RA and Dec.

Matches having separations $< 3\sigma = 0''.192$ (enclosed by the $r_{3\sigma}$ circle [left panel] and left to the red 3σ cut [right panel]) are considered as true matches.

As mentioned above, VHS provides several options for the type of aperture or model used to measure the photometry in each of the bands.

Thus, after the catalogue is split in pointlike and extended sources using the morphological classification described in the previous section, the fluxes in the VHS $YJHK_s$ bands are processed separately in preparation for the usage in LePhare.

In VHS, the magnitude values $m_{X,\text{VHS}}$ are provided in the Vega system both for the pointlike and for the extended sample.

Therefore, they need to be converted to AB magnitudes using proper correction values $m_{X,\text{corr}}$ before being transformed to the flux $F_{X,c}$ in $\text{erg/s/cm}^2/\text{Hz}$.

The correction values are taken from the LePhare filter analysis program and are applied as follows:

$$\begin{aligned} Y_{\text{AB}} &= Y_{\text{Vega}} + m_{Y,\text{corr}} = Y_{\text{Vega}} + 0.60^m \\ J_{\text{AB}} &= J_{\text{Vega}} + 0.92^m \\ H_{\text{AB}} &= H_{\text{Vega}} + 1.37^m \\ K_{s,\text{AB}} &= K_{s,\text{Vega}} + 1.83^m. \end{aligned} \quad (2.13)$$

The transformation from AB magnitude to flux (in $\text{erg/s/cm}^2/\text{Hz}$) for the VHS bands is then given by

$$F_{X,c} = 10^{-0.4(m_{X,\text{VHS}} + m_{X,\text{corr}} + 48.6^m)} \text{ erg/s/cm}^2/\text{Hz} \quad (2.14)$$

$$\sigma_{F_{X,c}} = \sigma_{m_{X,\text{VHS}}} 10^{-0.4(m_{X,\text{VHS}} + m_{X,\text{corr}} + 48.6^m)} 0.4 \ln(10) \text{ erg/s/cm}^2/\text{Hz}, \quad (2.15)$$

where the errors are propagated using Gaussian error propagation, and the correction values $m_{X,\text{corr}}$ are taken from eq. (2.13).

As outlined above, for the first iteration, **AperMag4** is used for the sources classified as pointlike, and **AperMag6** is used for the extended sources, although this is a topic that needs further investigation.

2.2.4. Ultraviolet: The Galaxy Evolution Explorer

In the ultraviolet (UV) range of the spectrum, ground-based photometry is not possible due to the absorption of the atmosphere. Therefore, the orbiting space-based Galaxy Evolution Explorer (GALEX) was built and launched to explore the origins of stars and heavy elements in the redshift range of $0 \leq z \leq 2$. During its up-time between 2003 and 2013, it conducted all-sky surveys with varying coverage and depth (see e. g. Morrissey et al. 2007; Bianchi 2014). The UV data does nicely complement the optical and infrared data collected thus far, which is why GALEX data is used whenever available.

Target selection and photometry

The GALEX target selection pipeline is described in e. g. Morrissey et al. (2007). Along with PSF, aperture and Kron-like (Kron 1980) photometry^I, the GR6+7 input catalogue provides a *calibrated* magnitude for each source, where the optimal photometric approach has been chosen by the pipeline. This calibrated magnitude is recommended by the authors and therefore used throughout this thesis.

Working with the catalogue data

As the most recent catalogue (GR6+7, described in Bianchi et al. (2017)) is available on the Vizier database^{II} of the Centre de Données astronomiques de Strasbourg (CDS), the **cdsskymatch** routine provided in STILTS can be utilised, allowing for a nearest-neighbour cross-matching to the sources of the remote table on the CDS servers.

The GALEX all sky imaging survey is relatively shallow, corresponding to a detection limit L_{AB} (5σ) of $L_{\text{AB},\text{FUV}} \sim 20^m$ and $L_{\text{AB},\text{NUV}} \sim 21^m$.

For the tests in the eFEDS field, of the 13 418 sources available in the LS sample, 5 465 (40.7

^Ia brief, comprehensive explanation can be found e. g. [here](#).

^{II}Available [here](#).

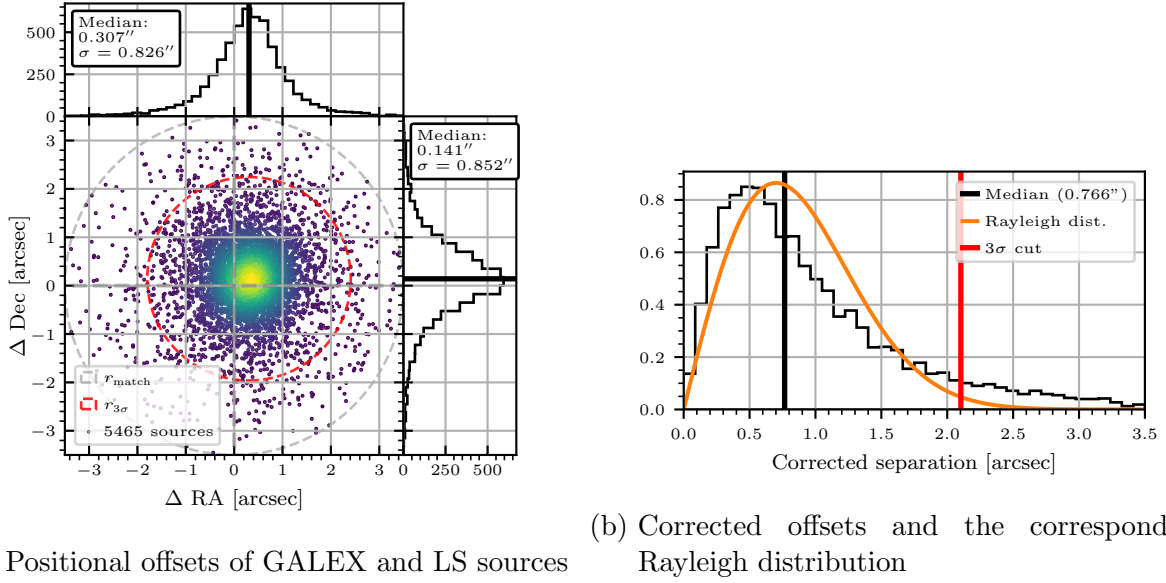


Figure 2.5: Positional offsets from the GALEX data to the LS DR9 data for the 5 465 nearest-neighbour matched sources in the eFEDS field using a matching radius of $3''.5$. The left panel shows the distributions of offsets in RA and Dec, while the right panel shows the distribution of the separation of sources after they have been corrected by the median offsets in RA and Dec. Matches having separations $< 3\sigma = 2''.103$ (enclosed by the $r_{3\sigma}$ circle [left panel] and left to the red 3σ cut [right panel]) are considered as true matches.

%) sources have a counterpart in the GALEX catalogue if a matching radius of $r_{\text{GALEX}} = 3''.5$ (see fig. 2.5) is employed.

The median positional offsets are $0''.307$ in the direction of the right ascension and $0''.857$ in the direction of declination. After correcting the individual offsets by these medians, the standard deviation of the corrected offsets is $\sigma = 0''.701$.

To ensure that only true matches are considered for the photometric redshift estimation, GALEX photometry of all sources at a corrected separation of more than 3σ is discarded, leading to only 5 013 matches with a corrected separation less than $3\sigma = 2''.103$ being taken into account for further analysis.

For reddening correction, Bianchi et al. (2017) suggest to use reddening correction factors of $A_{FUV} = 8.06E_{B-V}$ and $A_{NUV} = 7.95E_{B-V}$ for standard Milky Way extinction. Since the fluxes are already provided in mJy, the corrected and converted flux $F_{X,c}$ for band X (in $\text{erg/s/cm}^2/\text{Hz}$) can be calculated as follows¹:

$$F_{X,c} = F_{X,\text{Cat}} 10^{-0.4A_X - 29} \frac{\text{erg/s/cm}^2/\text{Hz}}{\text{mJy}} \quad (2.16)$$

$$\sigma_{F_{X,c}} = \sigma_{F_{X,\text{Cat}}} 10^{-0.4A_X - 29} \frac{\text{erg/s/cm}^2/\text{Hz}}{\text{mJy}}. \quad (2.17)$$

2.2.5. Validation data: Matched spectroscopy from eRASS

In preparation of the analysis of the coming data release for the eROSITA All Sky Survey (eRASS), which is going to provide a census of X-ray-emitting objects, Salvato et al. (2021) focused on the data release of the eROSITA Final Equatorial Depth Survey (eFEDS) to analyse the means of finding counterparts of the X-ray sources, characterise them, and probe the

¹Note that $1 \text{ mJy} = 10^{-29} \text{ erg/s/cm}^2/\text{Hz}$.

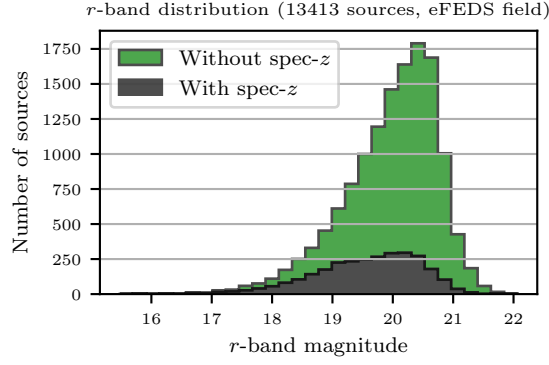


Figure 2.6: Stacked histogram of the r band magnitude for the sources in the eFEDS field, including the availability of spectroscopic redshifts.

The distribution of the sources without spec- z is shifted towards the fainter end, introducing a slight bias when relying on the spec- z for validation.

calculation of photometric redshifts.

For this purpose, they additionally gathered spectroscopic redshifts from publicly available spectroscopic surveys previously performed in the area (SDSS phases I-IV (Ahumada et al. 2020) and the SDSS-IV-SPIDERS campaign (Comparat et al. 2020)), linked them up to the sources of their base catalogue (the DR8 of the Legacy Surveys), and provided a quality flag corresponding to the reliability.

As these spectroscopic redshifts are available in the same area that is used for the testing purposes in this thesis for testing, the ones flagged as *secure* are used to validate the photometric redshifts obtained using LePhare.

Working with the catalogue data

After downloading the **eFEDS MAIN PS CTP** catalogue¹ containing the X-ray detected sources and their counterparts, a nearest-neighbour positional cross-match was performed. Since both catalogues are based on the DESI Legacy Imaging Surveys, a matching radius of $r < 10^{-3''}$ was sufficient, ensuring to match only true counterparts with the same coordinates.

Any sources not present in the optically selected AGN catalogue were excluded, leading to 2850 of the sources having reliable spec- z available.

Note: *The spectroscopy is generally more complete for the brighter sources, introducing a slight bias for the verification. This can especially be seen when looking at the distribution of r band magnitudes in fig. 2.6. Still, the general source population is represented fairly well by the spectroscopy, especially considering that a cut at $r < 21^m$ is going to be introduced for the 4Hi-Q catalogue input data.*

2.2.6. Supplementary i band data: HSC

As noted in sec. 2.2.2, during most of the time of this thesis being written, the DR10 of the LS had not fully been released, with a pre-release only being made available in May 2022.

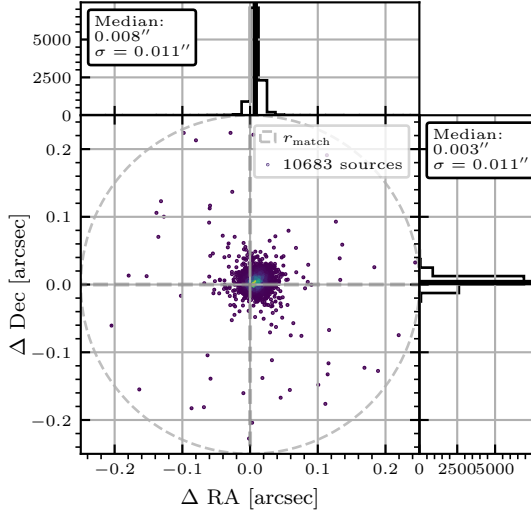
To still assess the impact of additional i band photometry even without this data release, two surveys containing similar data in the eFEDS field have been considered (see also sec. 2.2.7). The Hyper Suprime-Cam survey (Miyazaki et al. 2018, HSC hereafter) conducted different kinds of observations, e. g. wide broad-band observations in the g , r , i , z and y band close to the celestial equator with an emphasis on the i , along with deep and ultra-deep observations in smaller fields.

The footprint of its wide-field survey overlaps with the eFEDS area such that its data can

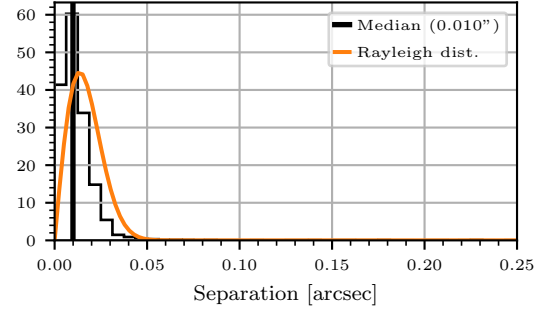
¹Available at <https://erosita.mpe.mpg.de/edr/eROSITAObservations/Catalogues/>

easily be used along the LS data gathered so far to assess the impact of additional i photometry.

Due to a strong radial dependence in the initial implementation of the i filter, the HSC survey team exchanged the i filter with an improved $i2$ filter in the middle of the operational period, leading to a slight change in the transmission curve, with some sources having been observed through both of them. Therefore, an estimate of the contribution of the initial i is provided in the `i_filterfraction_weighted` column.



(a) Positional offsets of HSC and LS sources



(b) Offsets and the corresponding Rayleigh distribution

Figure 2.7: Positional offsets from the HSC data to the LS DR9 data for the 10683 nearest-neighbour matched sources in the eFEDS field using a matching radius of $0''.25$. The left panel shows the distributions of offsets in RA and Dec, while the right panel shows the distribution of the separation of sources. No correction on the radius was applied since the matching seemed sufficiently accurate.

Working with the catalogue data

After querying (see sec. A.2.1) the HSC `pdr3—citus—columnar` catalogue¹ (leading to a total of 54 932 898 sources in the eFEDS area), a nearest-neighbour positional cross-match to the main catalogue was performed.

For the tests in the eFEDS field, of the 13 418 sources available in the LS sample, 10 683 (79.6 %) sources have a counterpart in the HSC catalogue if a matching radius of $r_{\text{HSC}} = 0''.25$ (see fig. 2.7) is employed.

Since the exact filter transmission curve is important for the template fitting step in LePhare, both of these filters have to be considered separately. The `i_filterfraction_weighted` data η_i is used to decide which filter to use for the fits: All sources with $\eta_i \leq 0.25$ are fit using the i filter, all sources with $\eta_i \geq 0.75$ are fit using the $i2$ filter. Photometry for all sources with $0.25 < \eta_i < 0.75$ is discarded (affecting approximately 16 % of the sources), reducing the number of available sources to 9002.

For the pointlike sources, PSF photometry (see sec. 4.9.5 in Bosch et al. 2018) was adopted, while CModel photometry (see sec. 4.9.9 in Bosch et al. 2018) was used for extended sources. The flux for each of the sources (provided in nJy) in the corresponding filter X (i and $i2$)

¹Available via an online registration at <https://hsc-release.mtk.nao.ac.jp/doc/index.php/tools/>

was corrected using the extinction values provided in the `A_i` column:

$$F_{X,c} = F_{X,Cat} 10^{-0.4A_i - 32} \frac{\text{erg/s/cm}^2/\text{Hz}}{\text{nJy}} \quad (2.18)$$

$$\sigma_{F_{X,c}} = \sigma_{F_{X,Cat}} 10^{-0.4A_i - 29} \frac{\text{erg/s/cm}^2/\text{Hz}}{\text{nJy}}. \quad (2.19)$$

2.2.7. Supplementary *i* band data: KiDS

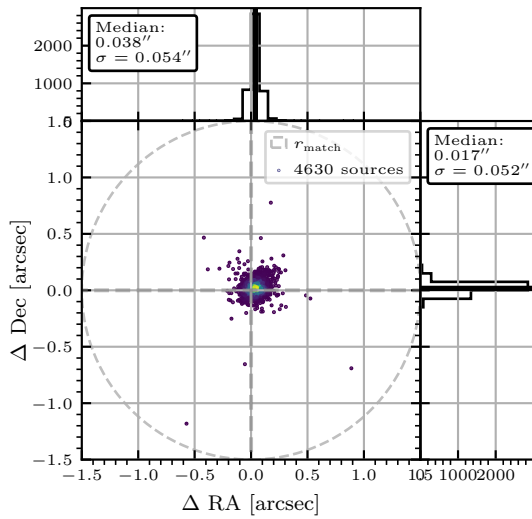
Due to a conversion error that had not been detected at first, the photo-*z* achieved with LePhare initially did not seem to improve. Therefore, it was decided that *i* band data from another survey could be used to find the origin of the error. Although the conversion error was later found and fixed, the results employing the KiDS data are still shown.

Due to its survey footprint overlapping with the eFEDS field, the Kilo-Degree Survey (Kuijken et al. 2019, KiDS hereafter) was chosen as the second catalogue for this assessment.

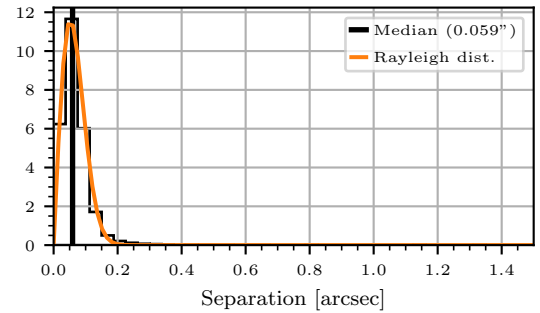
The KiDS is a wide-field survey conducted at the VLT survey telescope aiming to image 1500deg² in the *u*, *g*, *r* and *i* bands.

At the time of this thesis, it was advised to use the DR4.1.

After querying (see sec. A.2.2) the KiDS `KiDS_DR4.1_ugrizYJHKs_cat.fits` catalogue¹ (leading to a total of 5 481 230 sources in the eFEDS area), a nearest-neighbour positional cross-match to the main catalogue was performed.



(a) Positional offsets of KiDS and LS sources



(b) Offsets and the corresponding Rayleigh distribution

Figure 2.8: Positional offsets from the KiDS data to the LS DR9 data for the 4630 nearest-neighbour matched sources in the eFEDS field using a matching radius of 1''.5.

The left panel shows the distributions of offsets in RA and Dec, while the right panel shows the distribution of the separation of sources.

No correction on the radius was applied since the matching seemed sufficiently accurate.

For the tests in the eFEDS field, of the 13 418 sources available in the LS sample, 4630 (34.5 %) sources have a counterpart in the KiDS catalogue if a matching radius of $r_{\text{KiDS}} = 1''.5$ (see fig. 2.8) is employed.

The median positional offsets are 0''.038 in the direction of the right ascension and 0''.052 in the direction of declination. It can be assumed that this matching approach is accurate enough

¹Available at <https://kids.strw.leidenuniv.nl/DR4/index.php>

for this thesis, leading to 4630 matches being taken into account for further analysis.

In the catalogue, processed (and even extinction-corrected) i band magnitude data is available in the **MAG_GAAP_i** column. Therefore, it only has to be converted to erg/s/cm²/Hz such that it can be used as an input for LePhare:

$$F_{X,c} = 10^{-0.4(m_{X,\text{KiDS}}+48.6^m)} \text{ erg/s/cm}^2/\text{Hz} \quad (2.20)$$

$$\sigma_{F_{X,c}} = \sigma_{m_{X,\text{KiDS}}} 10^{-0.4(m_{X,\text{KiDS}}+48.6^m)} 0.4 \ln(10) \text{ erg/s/cm}^2/\text{Hz}. \quad (2.21)$$

2.2.8. Overview

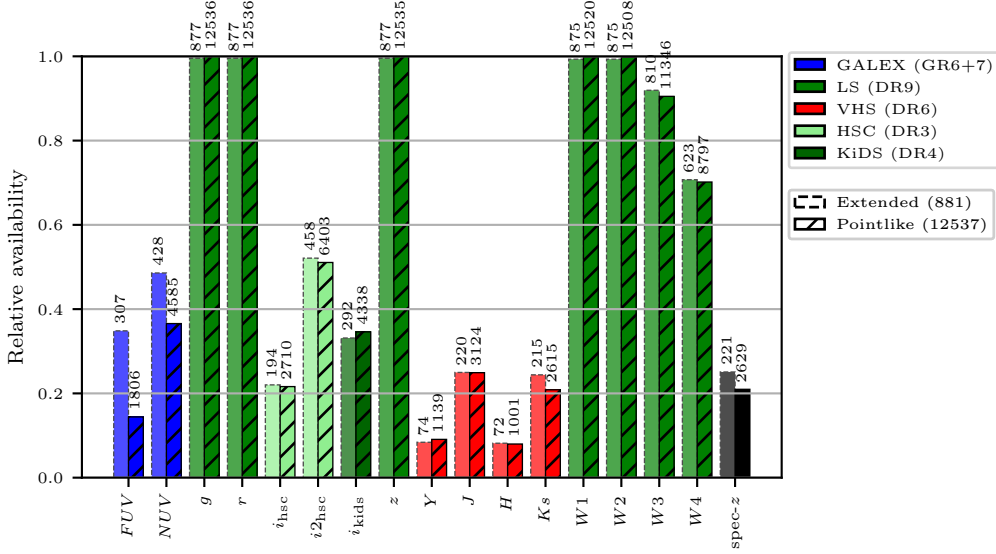


Figure 2.9: Relative availability of photometry in each band for the pointlike (12 537 sources in the eFEDS field) and extended (881) sub-samples.

When looking at the availability of photometry (fig. 2.9), g , r , z , W_1 and W_2 data is available for almost all of the 881 extended and 12 537 pointlike sources since Shu et al. (2019) used a prior data release of the LS to construct their optical/infrared AGN parent catalogue.

Generally, there is not much discrepancy between the extended and pointlike sub-samples except for the GALEX photometry (the FUV and NUV), which is more complete for the extended sources as the high-energy part of AGN spectra might be more obscured and red-shifted for the sources further away, while the galactic component has high UV luminosities, especially for star-forming galaxies.

Also, a slightly larger fraction of extended sources have reliable spectroscopic redshifts associated with them. This is simply due to more spectroscopy having been taken of galaxies.

In addition to looking at the availability of photometry in the respective bands, one can calculate the number of bands available for each source. This is shown in fig. 2.10. Due to the LS DR9 providing g , r , z , W_1 , W_2 , W_3 and W_4 band data for a majority of the sources, most have seven or more bands available.

Other interesting parameters are the distributions of the magnitudes of the sources in each of the bands. This is shown in fig. 2.11, along with source number counts for the pointlike and extended sub-sample in each of the bands.

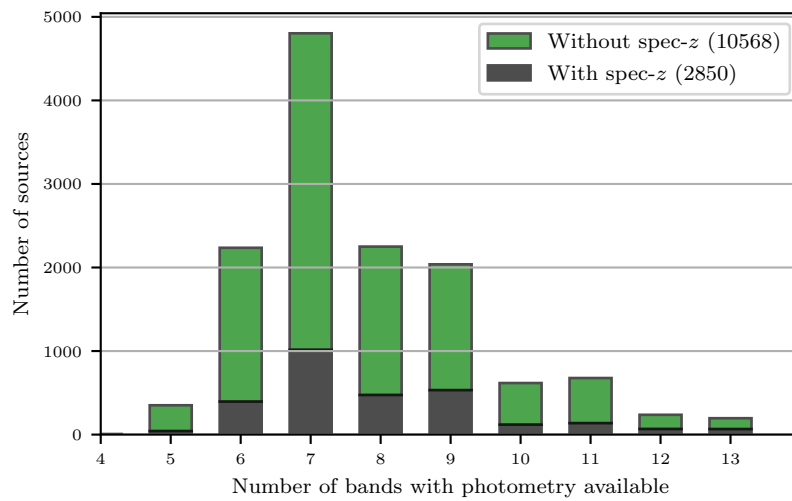


Figure 2.10: Histogram of the number of bands (excluding the i bands) available for each of the sources, colour-coded with the availability of spectroscopic redshifts.

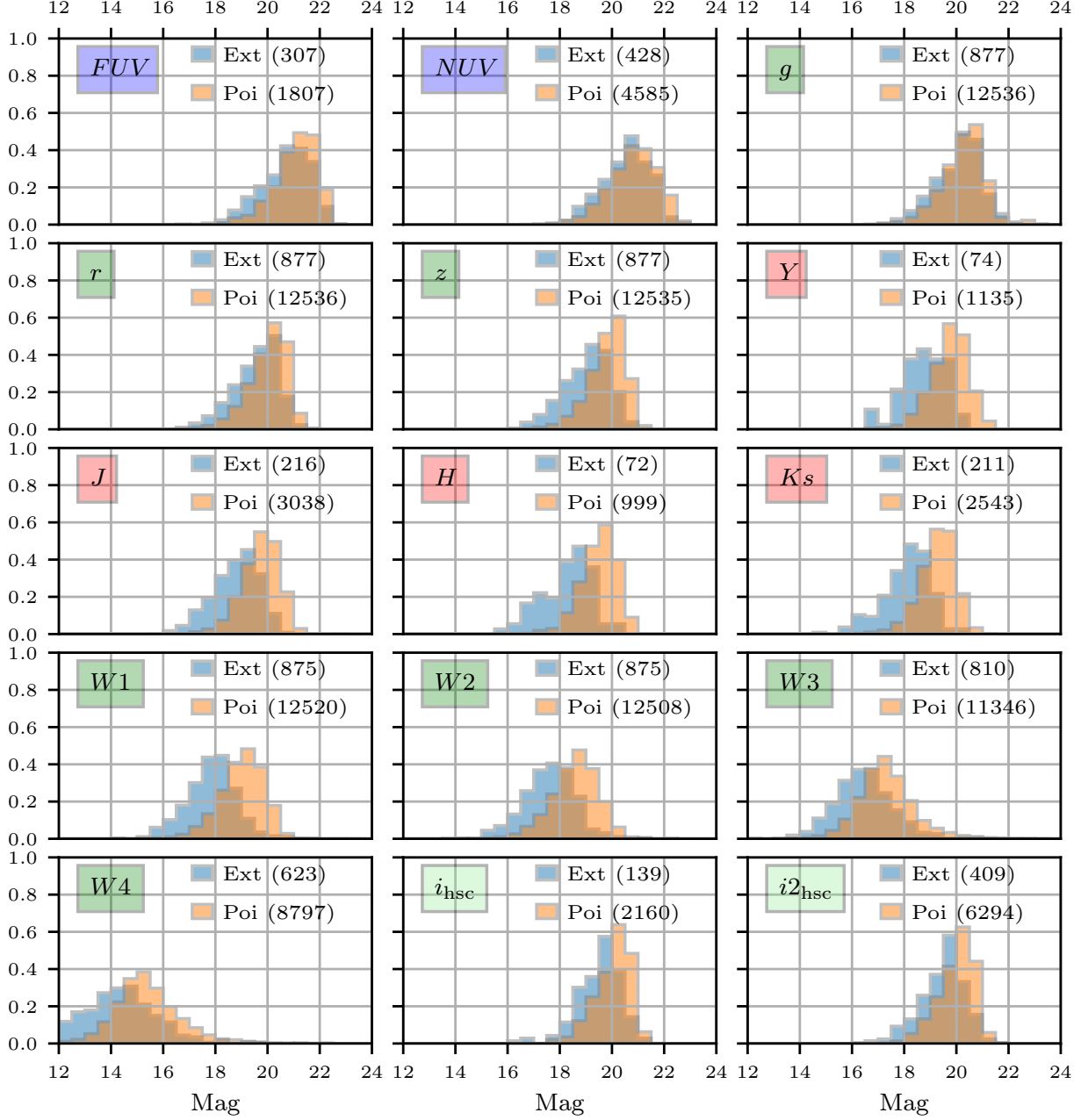


Figure 2.11: Magnitude distribution for the sources classified as pointlike (Poi.) and extended (Ext.) after extinction correction in each of the relevant photometric bands. The bands are colour-coded by the surveys they belong to, compare fig. 2.9. Generally, the pointlike sources are fainter.

2.3. Calculating photometric redshifts with LePhare

The following section describes the individual steps taken for the photometric redshift calculation, mirroring the general procedure laid out in sec. 2.1:

Before LePhare is able to calculate photometric redshifts, information on the filters and the templates needs to be provided and processed.

Then, the two morphology-separated input catalogues each have to be provided in the correct formats.

2.3.1. Collecting filter information

The filter information has to be provided to LePhare (compare sec. 2.1.1) such that the respective transmission curves can be applied to the templates, emulating the instrument's response.

The transmission files of the filters used in this work have in part kindly been provided by Mara Salvato and in part been taken from LePhare's native filter library.

An overview of the features of the used filter transmission files can be found in tab. 2.1, their respective transmission curves are depicted in fig. 2.12.

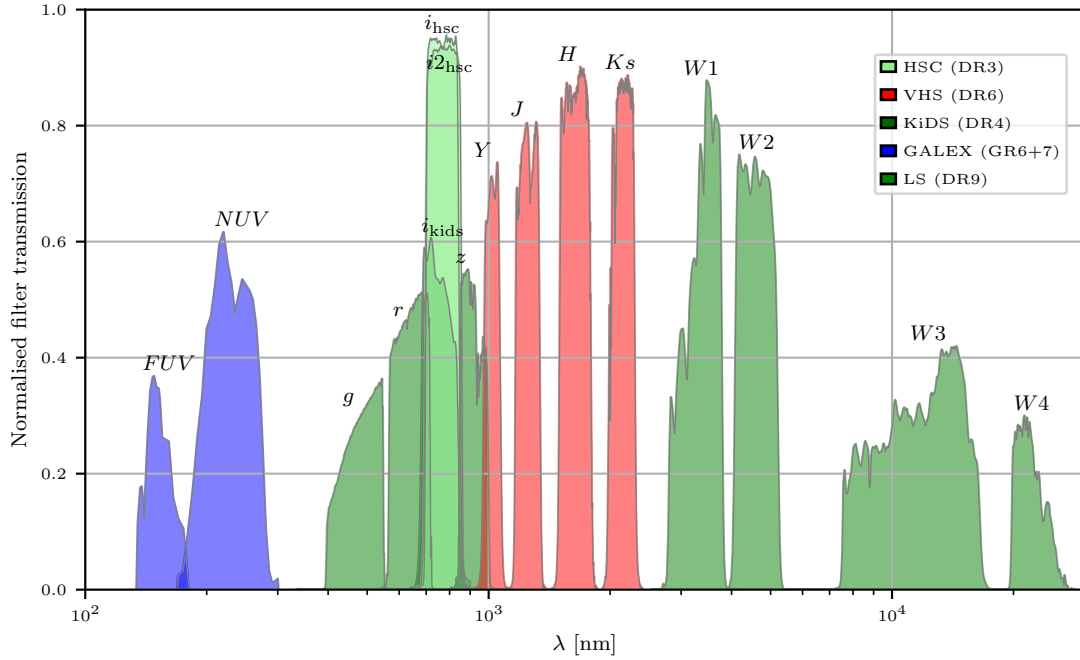


Figure 2.12: Spectral coverage of the filters used in this work.

Note that the HSC and KiDS bands (i_{HSC} , $i2_{\text{HSC}}$, and i_{KiDS}) were just used as a proof-of-concept for the LS i_{LS} .

2.3.2. Selection of input templates

In the next step, the template restframe SEDs of objects were selected, convolved through each of the specified filters, and propagated in redshift such that they could be used for the fitting (compare sec. 2.1.3).

One of the most important aspects about photo- z -fitting is the optimal selection of input templates.

While the entire population of objects should be reflected by the template library, too many models introduce degeneracy that can cause catastrophic outliers.

Table 2.1: Overview of the important parameters of the filters.

ID ^a	Band	Survey	$\langle\lambda\rangle^b$ [μm]	FWHM ^c [μm]	$m_{\text{AB}, c}^d$	m_{Vega}^e	$M_{\odot, \text{AB}}^f$
1	<i>FUV</i>	GALEX GR6+7	0.15	0.02	2.35	−20.68	18.98
2	<i>NUV</i>	GALEX GR6+7	0.23	0.08	1.76	−20.95	10.12
3	<i>g</i>	LS DR9	0.48	0.13	−0.09	−20.72	5.09
4	<i>r</i>	LS DR9	0.64	0.15	0.18	−21.62	4.62
5	<i>z</i>	LS DR9	0.92	0.15	0.52	−22.74	4.51
6	<i>Y</i>	VHS DR6	1.02	0.09	0.60	−23.05	4.51
7	<i>J</i>	VHS DR6	1.25	0.17	0.92	−23.81	4.56
8	<i>H</i>	VHS DR6	1.65	0.29	1.37	−24.85	4.70
9	<i>K_s</i>	VHS DR6	2.15	0.31	1.83	−25.90	5.13
10	<i>W₁</i>	LS DR9	3.38	0.79	2.66	−27.69	5.94
11	<i>W₂</i>	LS DR9	4.63	1.11	3.30	−29.02	6.61
12	<i>W₃</i>	LS DR9	12.33	8.33	4.72	−32.44	7.99
13	<i>W₄</i>	LS DR9	22.25	4.40	6.32	−35.44	9.58
14	<i>i_{HSC}</i>	HSC DR3	0.77	0.15	0.39	−22.23	4.53
15	<i>i_{2HSC}</i>	HSC DR3	0.78	0.16	0.40	−22.25	4.53
16	<i>i_{KIDS}</i>	KiDS DR4	0.76	0.16	0.37	−22.16	4.54
17	<i>i_{LS}</i>	LS DR10	0.78	0.15	0.41	−22.28	4.53

^a The context C_i of the filter with ID i can be calculated using $C_i = 2^{i-1}$.

^b $\langle\lambda\rangle = \frac{\int R(\lambda)\lambda d\lambda}{\int R(\lambda) d\lambda}$ is the mean wavelength of the filter.

^c Full Width at Half Maximum.

^d AB correction factor with $m_{\text{AB}} = m_{\text{Vega}} + m_{\text{AB}, c}$.

^e Vega Magnitude $m_{\text{Vega}} = -2.5 \log_{10} \left(\frac{\int R(\lambda) d\lambda}{\int R(\lambda) F_{\text{Vega}}(\lambda) d\lambda} \right)$.

^f Absolute magnitude of the sun in this filter.

The initial template libraries were provided by Mara Salvato, who recommended them from research on X-ray selected AGN. They contained 29 different templates for the pointlike and 86 for the extended library.

Since the sources analysed in this thesis are optically selected AGN, they still needed to be optimised for both the pointlike and extended sources. The refinement process for later runs is further described in sec. 3.2.

The template library for stars

For the stellar templates, LePhare’s recommended list of stellar templates `STAR_MOD_ALL.list` is adopted.

It contains a catalogue of templates from Pickles (1998) with spectral coverage from 1150 to 10620 Å for all components and to 25000 Å for some, mainly later types of solar abundance, Brown Dwarves, low mass stars, and White Dwarves from Chabrier et al. (2000).

Since these include more than 250 templates, they are not listed in this document, but an overview can be found in `baseline_templates.star.list` in the git code associated with the thesis.

The template library for galaxies and quasars: pointlike appearance

Since the objects classified as pointlike are expected to have a higher quasar contribution, strong galactic features such as the Balmer Break might be missing completely. Therefore, a range of synthetic quasar models and some models with contributions from both the host galaxy and the quasar are used.

In addition to considering completely different sets of templates, the effect of galactic ex-

tion may not be negligible, which also has to be reflected in the final magnitude library. Thus, for the pointlike templates, a range of E_{B-V} values from 0 to 0.4 in steps of 0.05 is considered applying an extinction law for the SMC by Prevot et al. (1984), as recommended by Bolzonella et al. (2000).

Then, each template is propagated in redshift as usual. While the inclusion of extinction covers a wider range of possible appearances, it also introduces an extra layer of degeneracy. In sec. 3.2, an in-depth discussion of how the final templates were selected is provided. The final selection of used pointlike templates can be found in tab. B.5.

The template library for galaxies and quasars: extended appearance

For the objects classified as being extended, it is reasonable to use templates with higher contributions of the host galaxies. For this task, a high portion of AGN templates of single objects by Brown et al. (2019) are considered, along with composite models of starburst galaxies with AGN contributions.

In sec. 3.2, an in-depth discussion of how the final templates were selected is provided. The final selection of used extended templates can be found in tab. B.6.

After the templates had been selected for stars, pointlike and extended sources, the expected flux in each band was calculated for redshifts in a range of $0.0 < z < 6.0$ in steps of $\Delta z = 0.02$.

2.3.3. Running LePhare with priors

As discussed in sec. 2.1.5, the photometric redshift fitting routine **zphota** of LePhare is utilised after the input catalogue and magnitude libraries are available.

In addition to the different magnitude libraries gathered for the pointlike and the extended sub-sample, magnitude priors are applied to avoid fits to unrealistic magnitudes for the galaxies.

In LePhare, a range of allowed absolute magnitudes referring to one of the photometric bands can be supplied via the **MAG_ABS** flag.

As a reference, the g band magnitude of the fitted template is used.

For the pointlike sample, the magnitude range is set to $[-30^m, -20^m]$, reflecting on the assumption that the pointlike sources are remote, luminous objects.

For the extended sample, the magnitude range is set to $[-24^m, -8^m]$, reflecting on the assumption that the extended sources are resolved galaxies.

2.4. Detailed description of the Code

The code **sel-4Hi-Q** accompanying this thesis is designed to be easy to access and use.

It is capable of processing the data provided by the different catalogues and to perform the photometric redshift fitting via LePhare on it.

In addition to that, various plotting routines are provided to perform analysis of the joint input catalogue and the output coming from LePhare.

2.4.1. Requirements

The latest version of **sel-4Hi-Q** can be requested from the author^I and will be available on gitlab shortly after this thesis is finished.

The following programs are necessary for **sel-4Hi-Q** to work:

- The current version of LePhare++^{II}, along with proper setup of the **LEPHAREDIR** and

^Iat fabian.balzer@studium.uni-hamburg.de

^{II}LePhare++ is available [here](#)

LEPHAREWORK environment variables^I.

- An installation of **jystilts**^{II}, which is the **jython** installation of STILTS Taylor (2006). The location of **jystilts** should be saved employing the **JYSTILTSDIR** environment variable.
- A **python** (version 3.8+) installation, including the **pandas**, **matplotlib** and **numpy** modules^{III}.
- A directory containing the input catalogues in a directory with the environment name **CATPATH**.

After the repository has been cloned, the program **startup.py** should be run using

```
>>> python startup.py
```

This routine sets up the proper directory structure, provides information about missing components, and it also resets the basic configuration files **general_config.ini** (which contains general information about the system and should not be altered) and **custom_config.ini**.

*Note: After additional components have been installed and new environment variables have been set, the **startup.py** program needs to be run again!*

2.4.2. Available parameters in the configuration file

Once the program has been installed, its behaviour can be manipulated via keywords in the configuration file **custom_config.ini**.

The available parameters are structured in different sections, depending on the roles they fulfil:

- The **General** settings include settings that influence all parts of the program (tab. B.1).
- The **Catalogue assembly** settings manage the behaviour of the routines to collate the input catalogue data. They are relying on the **assemble_cat** setting to be set to **True** (tab. B.2).
- The **LePhare** settings determine the invocation of the different LePhare subroutines of assembling the filter files and the template library, and whether the actual photo-*z* program should be run (tab. B.3).
- The **Plotting** settings determine which plots are produced along with the run (tab. B.4).

For better readability, the tables explaining the individual keywords have been moved to the appendix.

2.4.3. Running the code

The whole code can then be run from the main directory using

```
>>> python run_all.py
```

This will execute the code based on the parameters set in **custom_config.ini**. If multiple different configuration files are desired to be used, they need to be stored in the **config** directory and the active one needs to be specified in **general_config.ini**, changing the parameter for **current_config** to the name of the new configuration file.

^I**bash** users can just include **export LEPHAREDIR="path_to_lephare"** in their **.bashrc** file, Windows users can set environment variables manually.

^{II}**jystilts** is available [here](#)

^{III}which are part of most common **python** distributions like Anaconda

There are several sub-modules that might be interesting to use (such as `input_plot_container.py` to produce more flexible plots). For this, one needs to ensure to access them from the main directory or to add the main directory to their `PYTHONPATH`.

In addition to that, an adaptation of the `spec.py` file that comes with LePhare++ can be found and utilised via

```
>>> python analyse_spec_files.py -h
```

The parameters of this program need to be set via the command line, the flag `-h` will invoke a help message.

`analyse_spec_files.py` can be used for a quick analysis of the `.spec` output files that LePhare provides whenever the `SPEC_OUT` parameter is set to `YES` during a `zphota` run.

Examples of these plots are shown in sec. 3.4.3.

Note: *This code both adapts the existing `spec.py` file and a similar file provided to the author by Mara Salvato.*

3. Improving the photometric redshifts

3.1. First results in the eFEDS field

To assess the performance of the photo- z obtained in the eFEDS area, both the outlier fraction η_{out} (eq. (1.6)) and the accuracy σ_{NMAD} (eq. (1.4)) are used as a metric.

The assessment is only possible for the 2850 sources with spec- z available and thus might be biased toward the brighter sources as discussed above.

First results considering just the LS DR9, VHS and GALEX photometry (at maximum 13 bands) are shown in fig. 3.1.

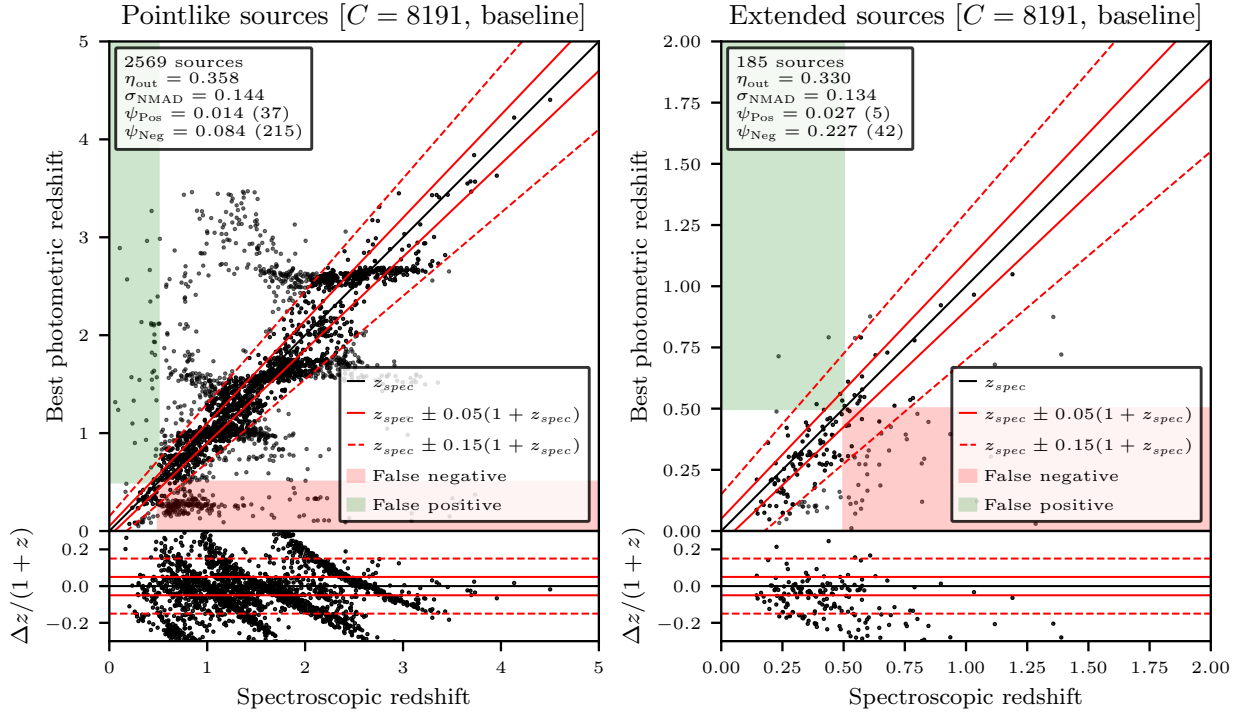


Figure 3.1: Comparison between spec- z and photo- z for the sources in the eFEDS field from the first LePhare run, adopting only the 13 basic bands with the initial templates found in `baseline_pointlike.list` and `baseline_extended.list`.

Here, each dot corresponds to a source with both spec- z and a best photo- z (determined by LePhare) available.

An outlier fraction of $\eta_{\text{out}} > 0.3$ is really high even for photo- z for quasars, implying that there are several improvements possible:

- The templates used for the photometric redshift fitting are not yet optimised for the optically selected quasars, which is implied by the horizontal features at $z_{\text{phot}} \approx 2.7$ and $z_{\text{phot}} \approx 1.6$ in the plot for the pointlike sources. Here, many of the quasars are falsely fit to the same template, hinting that the available templates do not yet reflect their population.

Attempts to improve the template selection are discussed in sec. 3.2.

- Another way to lift degeneracies is to introduce more photometry. The aspect of introducing i band photometry to improve the photo- z quality is explored in sec. 3.3.
- Other reasons for the poor photo- z quality are discussed in sec. 3.4

Due to the shallowness of the VHS and the templates not being optimised for the population, the overall performance is quite bad in comparison to literature, although it should be noted

that the 4Hi-Q project especially requires the photo- z to have a small rate of objects ψ_{Pos} with redshifts classified as being too low (eq. (1.9)).

3.2. Improving the template selection

The correct selection of templates is a crucial part of the photo- z fitting process: If the population of objects is not well represented or if there are too many templates that appear similar at different redshifts, a large fraction of sources is fit to the wrong template and therefore possibly a completely wrong redshift.

Additionally, wavelength-dependent extinction has to be factored in, since high-redshift galaxies might be obscured by dust, leading to high-energy flux being re-emitted at higher wavelengths. LePhare provides several reddening laws adopted from the *hyperz* code (Bolzonella et al. 2000). The authors of this code also note that the extinction laws of the LMC and the Milky Way did not seem appropriate for sources at $z > 2$, ruling them out for the pointlike sources. They note that considering either the extinction laws for the SMC (Prevot et al. 1984) or starburst galaxies (Calzetti et al. 2000) improves the correct assignment of high-redshift objects. A side effect of applying extinction laws is that some lower-redshift objects might be classified as being at higher redshifts due to the high extinction at short wavelengths.

Table 3.1: Overview of the performance with the different template files (adopting the 13 standard bands).

Template file ^a	N_j ^b	η_{out} ^c	σ_{NMAD} ^d	ψ_{Pos} ^e	ψ_{Neg} ^f
<code>baseline_pointlike.list</code>	29	0.358	0.144	0.014	0.084
<code>baseline_extended.list</code>	86	0.330	0.134	0.027	0.227
<code>ananna_pointlike.list</code>	25	0.442	0.166	0.019	0.061
<code>ananna_extended.list</code>	15	0.250	0.116	0.041	0.173
<code>ananna_score_reduced_pointlike.list</code>	10	0.430	0.163	0.022	0.030
<code>ananna_score_reduced_extended.list</code>	8	0.236	0.117	0.028	0.180
<code>baseline_score_reduced_pointlike.list</code>	10	0.332	0.139	0.023	0.054
<code>baseline_score_reduced_extended.list</code>	34	0.314	0.138	0.021	0.216
<code>combined_pointlike.list</code>	16	0.316	0.135	0.018	0.068
<code>combined_extended.list</code>	40	0.304	0.131	0.021	0.209

^a Explanations of the filenames can be found above.

^b number of templates, ^c eq. (1.6), ^d eq. (1.4), ^e eq. (1.9), ^f eq. (1.10)

To have a template selection as complete and with the least degeneracies as possible, several tests have been conducted employing the 13 basic photometric bands to improve upon the selection¹:

1. As a baseline, the templates in the `baseline_**type**.list` files kindly provided by Mara Salvato were used, yielding the results shown in fig. 3.1.
2. Another possible starting point are the templates used by Ananna et al. (2017), who carried out photo- z estimation for X-ray selected AGN in the Stripe 82X area. Due to the different selection methods, these might not be representative of the optically selected AGN. The templates for this test are located in the `ananna_**type**.list` files.

¹In the following, `**type**` is used to denote both the pointlike and extended sub-samples. The lists of templates representing these different objects differ - in general, more galactic contributions are expected for the extended objects, as discussed in sec. 1.2.3.

As evident in tab. 3.1, their template selection significantly improves the photo- z for the extended sources. Their pointlike template selection does not seem to properly represent the population of the optically selected sources.

3. To further improve on these templates, the number of outliers and correct redshifts was further analysed where each template j was awarded a score based on the following heuristic formula:

$$S_j = \frac{1}{N_j} \sum_{i=1}^{N_j} \exp \left(-\frac{|z_{\text{phot},i} - z_{\text{spec},i}|}{1 + z_{\text{spec},i}} - \frac{\chi_{\text{min},i}^2}{2} \right), \quad (3.1)$$

where N_j are the number of sources fit by the template.

This measure is the mean of the offset of each source towards its true redshift ($\exp(-|x_i|)$), weighted with a measure of the probability of the fit ($\exp(-\chi_i^2/2)$) such that good fits with high probabilities achieve the highest scores.

After these scores were awarded, only the 75 % of the templates with the highest scores were kept, in addition to templates that were not fit to any outliers. All templates that were only fit to outliers were cut as well. This procedure yielded the **baseline_score_improved**type**.list** and **ananna_score_improved**type**.list** files, see fig. 3.2 and fig. 3.3) for plots of this selection method.

4. After these tests, a joint list of templates was put together with all templates from these lists combined (see **combined**type**.list**).

As discussed in sec. 2.2.5 and evident by looking at fig. 2.6, an optimisation procedure making use of the spec- z introduces slight biases towards brighter sources.

A summary of the performances of the LePhare runs adopting the different template selections is provided in tab. 3.1.

Note: *To see which templates actually represent the population of sources, the **ZFIX** option of LePhare can additionally be applied. If this option is enabled during the **zphota** routine, instead of trying to find the best photo- z , the template best representing the source at the provided spec- z is assigned to it.*

Reducing the template lists to only these templates was tested, but the possibility of excluding the wrong important templates was considered too dangerous.

Which template list should be used?

As evident from the results in tab. 3.1, the baseline templates could be further improved with the inclusion of templates by Ananna et al. (2017).

Although their templates are not complete for the pointlike population (reflected by the high outlier fraction), adding them to the pool of possible templates improved both outlier fraction and accuracy.

While **ananna_score_improved_extended.list** yields the best results concerning both outlier fraction and accuracy, it contains only 8 templates and is thus unlikely to properly represent the multitude of possible extended objects and might be heavily biased on the validation sample. For this reason, the combined extended template list of 40 templates is adopted for future tests, although the outlier fraction is higher.

The results of the final template improvement run are shown in fig. 3.4. Comparing this to the results in fig. 3.1, some degeneracies at $z_{\text{spec}} \approx 1$ seem to have been lifted for the pointlike sources while one large degeneracy remains a $z_{\text{phot}} \approx 2.7$. It might be impossible to lift these due to the power-law nature of the SEDs of the sources, but an attempt for that is made in sec. 3.3 by including an additional photometric band.

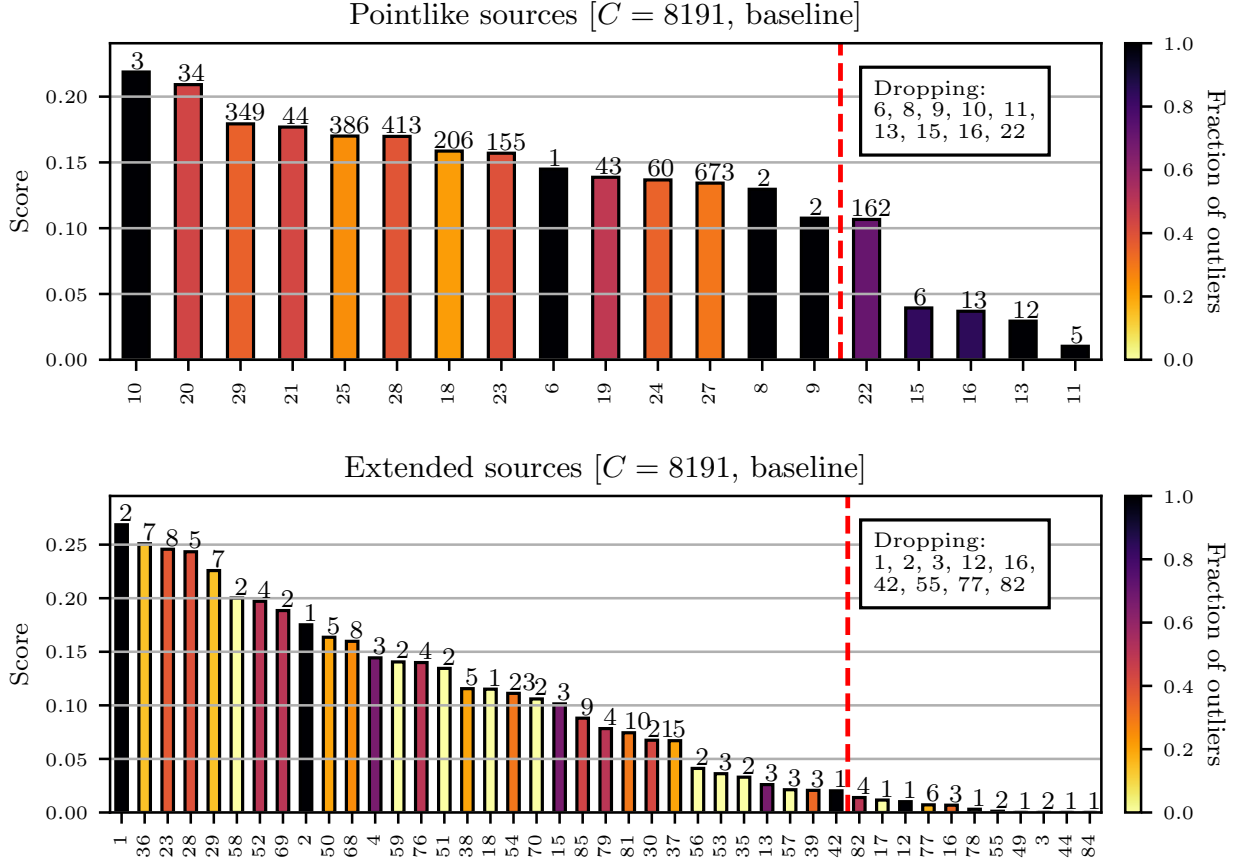


Figure 3.2: Score (eq. (3.1)) distribution for the templates in `baseline_**type**.list`, coloured by the fraction of outliers. Templates with outlier fractions of 1 and templates with the lowest scores (right of the red cutoff line) were cut for `baseline_score_reduced_**type**.list`.

The numbers above the bars indicate the amount of sources that were fit by that template, the box in the upper left denotes the templates that were dropped.

Further analysis of the templates

In addition to the reduction of the templates, it is instructive to look at the number of times different templates are used for certain fits (fig. 3.5).

For the pointlike sub-sample, template 10 (which is a hybrid consisting mainly of a luminous QSO and in part of a starburst galaxy) seems to be the main template that is fit for most of the sources, while the distribution among the extended sources is much more diverse, reflecting the different kinds of galactic contributions. Here, template 16 (Mrk817, which is a seyfert 1.2 galaxy) seems to be fitted most.

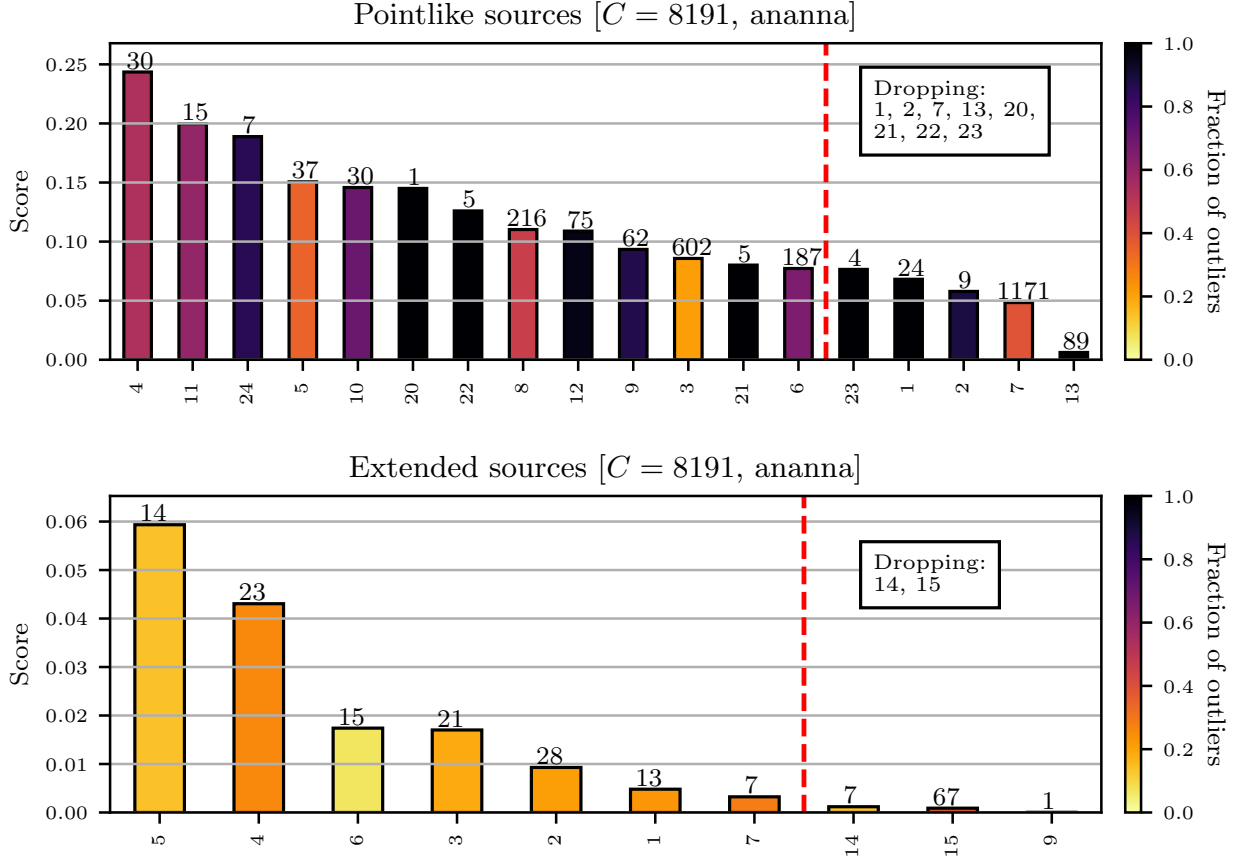


Figure 3.3: Score (eq. (3.1)) distribution for the templates in `ananna.**type**.list`, coloured by the fraction of outliers. Templates with outlier fractions of 1 and templates with the lowest scores (right of the red cutoff line) were cut for `ananna_score_reduced.**type**.list`.

The numbers above the bars indicate the amount of sources that were fit by that template, the box in the upper left denotes the templates that were dropped.

Interestingly, the score for the extended sources was generally a lot lower than for any of the other template selections due to high χ^2 values.

3.3. Introducing i band photometry

One way to improve the performance of the photometric redshifts is to include photometry of more bands, which mainly reduces possible degeneracies in the colour-redshift space. As discussed in sec. 2.2.6 and sec. 2.2.7, a preliminary catalogue containing i band data of the LS DR10 was made available to the author shortly before this thesis was finished.

In preparation of this, the impact of the inclusion of the i of the HSC survey and KiDS was first investigated, followed by a brief consideration of the LS DR 10 data.

The plots in fig. 3.6 show that the $i_{\text{HSC}} - z$ colour is excellent to lift degeneracies as the templates diverge especially in the range of $1 < z < 2$.

Results with i band data from HSC

If i band data from the HSC survey is taken into consideration along with the 13 basic bands, the results of the photometric redshifts are definitely more reliable, although the degeneracy at $z_{\text{phot}} \approx 2.6$ remains. The outlier fraction of the pointlike sources decreases to $\eta_{\text{out}} = 0.277$, while the outlier fraction of the extended sources decreases to $\eta_{\text{out}} = 0.299$. Especially notable for the extended sources is the drastic decrease in the false positive fraction ψ_{Pos} , which becomes as low as 0.005, and is probably the most important parameter for the 4Hi-Q

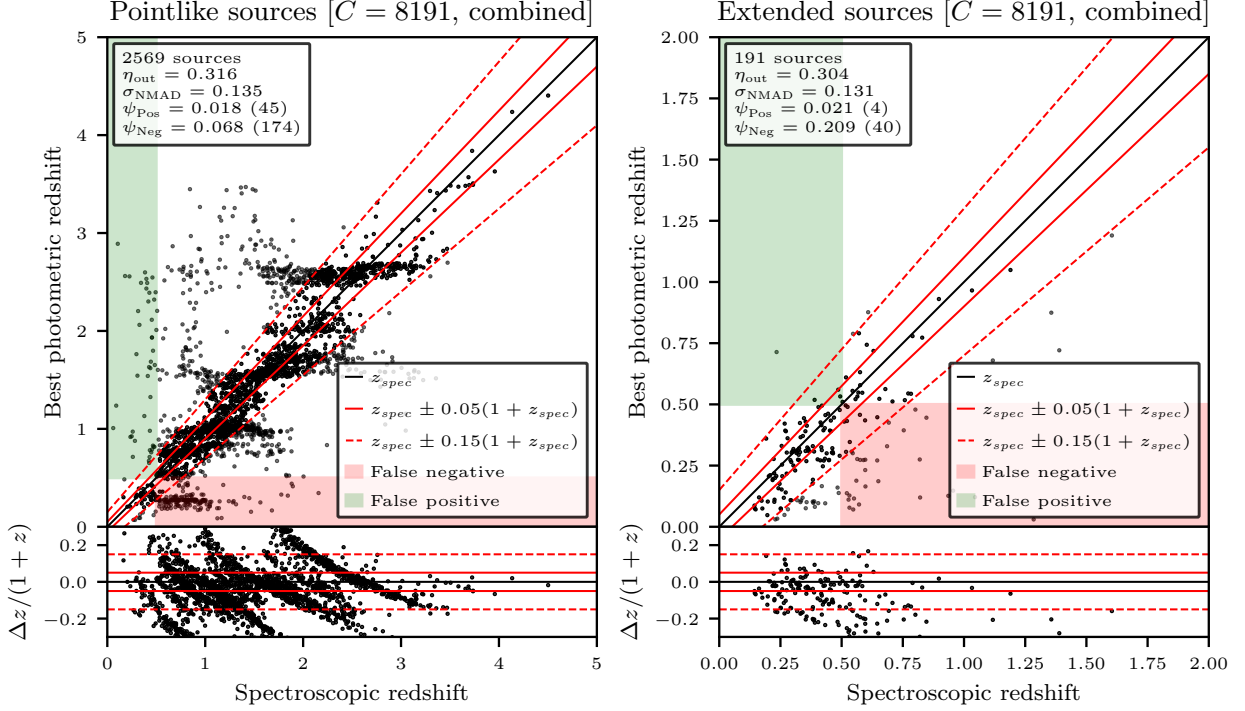


Figure 3.4: Comparison between spec- z and photo- z for the sources in the eFEDS field from the first LePhare run, adopting only the 13 basic bands with the improved templates of in `combined_pointlike.list` and `combined_extended.list`.

survey.

If one takes a look at the subset of only the sources with HSC data available, the outlier fraction decreases even further for the pointlike sub-sample, while it is slightly higher for the extended sub-sample¹.

Note: Since LePhare may automatically neglect photometric information if the errors are too high, one should be careful interpreting the look at this subset. This may also be the reason for η_{out} being slightly higher for the extended sub-sample.

Results with i band data from KiDS

If the i band from the KiDS survey is used along with the 13 basic bands, similar improvements are achieved, although (probably due to the lower relative amount of available KiDS photometry), the overall accuracy is not enhanced as much as with the HSC bands (see tab. 3.2 for the exact numbers. A plot of the photo- z vs. the spec- z is omitted here since it is very similar to fig. 3.6).

Results with i band data from LS DR10

Unfortunately, the amount of counterparts with i band photometry from the preliminary DR10 of the Legacy Surveys is really low (with only 692 of the 13418 (5.2 %) sources, that fraction carrying over to the validation sample).

Therefore, the results need to be considered with care.

For the 156 sources in the pointlike sub-sample with spec- z and i_{LS} band data available, the outlier fraction decreases to $\eta_{\text{out}} = 0.244$ (with at only $\psi_{\text{Pos}} = 0.013$). This result emphasises on the need for accurate i band data.

Still, for the 16 sources in the extended sub-sample with spec- z and i_{LS} band data, both η_{out} and ψ_{Pos} increase. This can be neglected due to the low sample size.

¹which is interesting, as the outlier fraction for the total sample decreases.

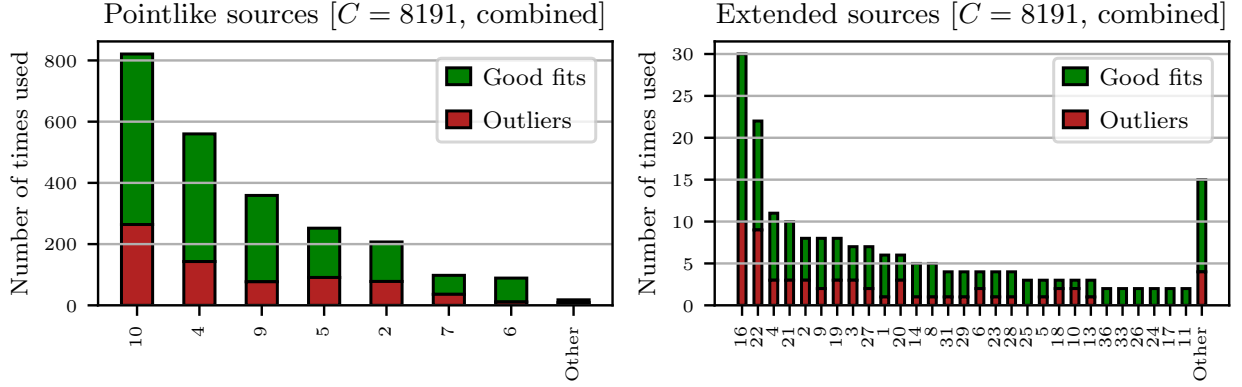


Figure 3.5: Comparison of the amounts each of the templates in `combined.**type**.list` (tab. B.5 and tab. B.6) were used for the fits, split by the quality of their fits. Outliers are defined by eq. (1.5). Templates with less than 50 (pointlike) and 2 (extended) total fits were pooled in the *other* bar.

Overview and comparison

An overview of the photo- z results employing each of the different i bands is provided in tab. 3.2.

The reason that the pointlike photo- z making use of the i_{LS} band data are even better than those making use of the i_{HSC} or i_{KIDS} band data might be that the Legacy Surveys apply the same kind of forced photometry in all bands, which is a huge advantage over data patched together from different surveys.

¹usually, that should not be the case, especially if the errors are high.

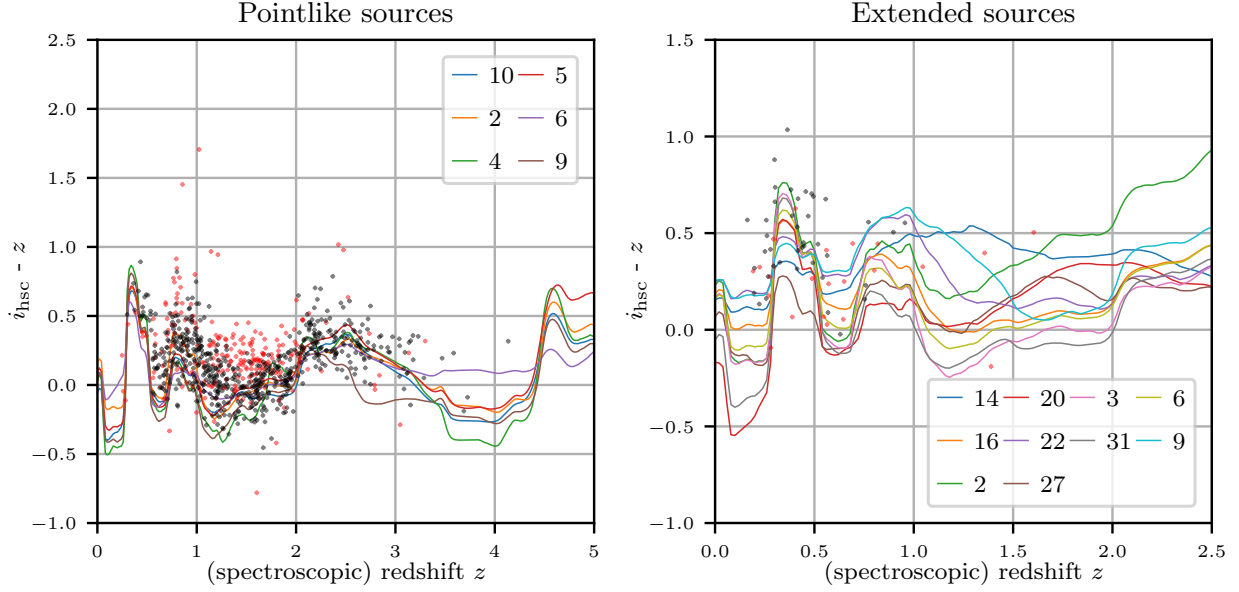


Figure 3.6: Plot of how the $i_{\text{HSC}} - z$ colour changes with regard to redshift.

The black and red dots represent sources that were fit by LePhare employing the 13 basic bands and i from the HSC survey, where red denotes outliers.

Note that additional reddening is applied to the pointlike templates to account for extinction, which is not shown here as this would clutter the plot too much.

3.4. Further discussion

In this section, the results of the photo- z employing the 13 standard bands and additionally the HSC i band are further discussed. For these tests, the final combined template lists are considered (see sec. 3.2).

3.4.1. Analysing the used bands

One metric to look at is the outlier fraction with respect to the amount of available photometric bands for the LePhare fitting (fig. 3.8).

Surprisingly, the outlier fraction does not decrease monotonically with increasing numbers of photometric bands, but rather increases again with 13 or 14 bands available.

Note: *As explained above, there is no possibility to learn which of the bands LePhare actually used for the fits; only the amount is provided as a feedback.*

For the extended sub-sample, the outlier fractions for the highest amounts of bands are not as meaningful since they each only contain < 13 sources.

For this reason, a closer look at the performance measures of subsets of the data makes sense. A table with an overview of these measures can be found in the appendix (tab. B.7).

One notable takeaway for the extended bands is that η_{out} is significantly lower for the subsets with GALEX band data (as low as $\eta_{\text{out}} = 0.222$), and significantly higher for the subsets with VHS (as high as $\eta_{\text{out}} = 0.500$). This is a hint that the **AperMag6** photometry prescription for sources with VHS might be false. In addition to that, the photometry might be contaminated by a higher amount of background and overlapping sources.

On the other hand, availability of VHS data seems to improve the quality of pointlike photo- z to outlier fractions as low as $\eta_{\text{out}} = 0.232$, hinting that using **AperMag4** photometry is the correct way.

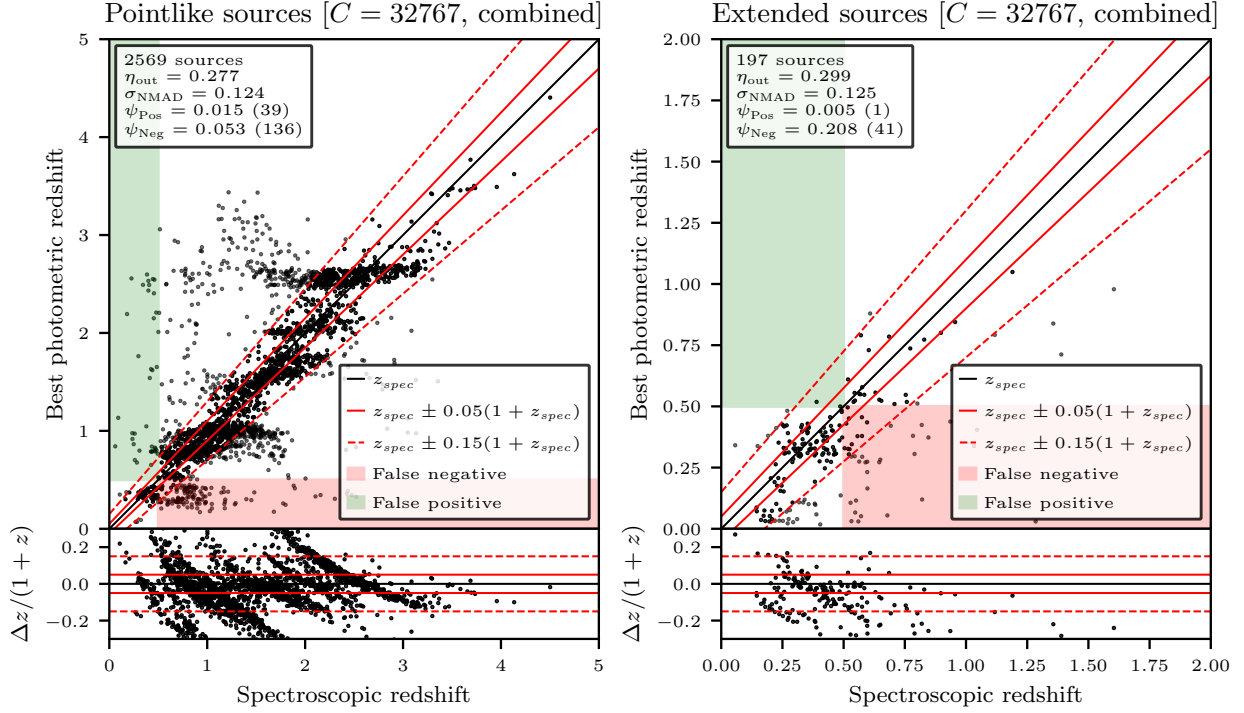


Figure 3.7: Comparison between spec- z and photo- z for the sources in the eFEDS field from the first LePhare, adopting the 13 basic bands + HSC data with the combined templates.

3.4.2. Template coverage

In addition to analysing the scores and outlier fractions achieved with fits to individual templates, it makes sense to analyse their coverage of the colour-redshift space of the sources.

As can be seen in fig. 3.10a, the $FUV - r$ colour-space does not seem to be covered properly in the redshift range between $1 < z_{\text{spec}} < 2$, even if extinction is applied. Notably, a majority of the sources not covered are outliers.

The $g - Y$ colour evolution (fig. 3.10b) is interesting for the extended sources; here, some sources at $z_{\text{spec}} \approx 0.5$ are not covered in the given colour-redshift space, but this might just be an artefact of a low amount of sources actually having Y band data available.

The last of the three plots shows the evolution of the $W_1 - W_2$ colour with respect to redshift. The colour distribution is pretty slim compared to the other plots, but the templates seem to cover this colour-redshift space well.

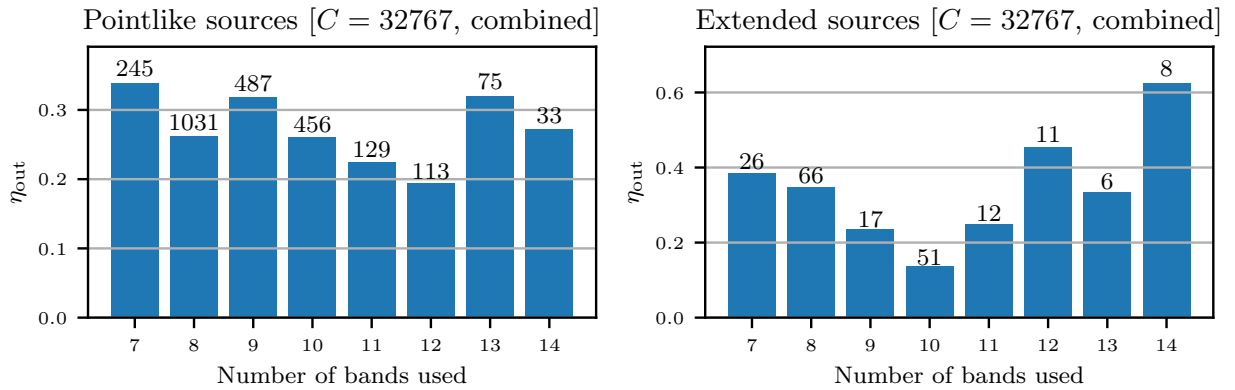


Figure 3.8: Outlier fractions w. r. t. the amount of photometric bands available for the fit.

Table 3.2: Comparison of the different results employing the i band for the sources with spec- z available.

The rows that were reduced to the respective bands may not be meaningful as no information is available on whether LePhare actually employed this band for all sources during the fitting process¹.

pointlike					
Bands	N_{source}	η_{out}	σ_{NMAD}	ψ_{Pos}	ψ_{Neg}
Total (just basic bands)	2569	0.3157	0.1351	0.0175	0.0677
Total (with HSC)	2569	0.2772	0.1242	0.0152	0.0529
Reduced to subset with i_{HSC} (joint)	1972	0.2693	0.1210	0.0122	0.0472
Total (with KiDS)	2569	0.3005	0.1260	0.0183	0.0662
Reduced to subset with i_{KiDS}	1561	0.2992	0.1258	0.0154	0.0679
Total (with LS DR 10)	2480	0.3129	0.1341	0.0165	0.0677
Reduced to subset with i_{LS}	156	0.2436	0.1071	0.0128	0.0641
extended					
Bands	N_{source}	η_{out}	σ_{NMAD}	ψ_{Pos}	ψ_{Neg}
Total (just basic bands)	191	0.3037	0.1307	0.0209	0.2094
Total (with HSC)	197	0.2995	0.1251	0.0051	0.2081
Reduced to subset with i_{HSC} (joint)	142	0.3099	0.1274	0.0000	0.2183
Total (with KiDS)	189	0.2910	0.1158	0.0212	0.2063
Reduced to subset with i_{KiDS}	124	0.2823	0.1163	0.0242	0.1935
Total (with LS DR 10)	184	0.2935	0.1173	0.0217	0.2011
Reduced to subset with i_{LS}	16	0.3750	0.1369	0.0000	0.1875

3.4.3. Taking a closer look at certain SEDs

Another way to analyse the nature of the photometric redshift estimations of LePhare is provided by the **SPEC.OUT**-keyword.

If this is set to **YES** during the **zphota** routine, LePhare produces a **.spec** file for each source, supplying information on the best-fit-templates, the probability distribution and the photometry.

While this thesis was written, the analysis of these files was a handy way to find various sources of errors, both in the photometry and the template selection.

One of these plots, with a correctly fit source, is shown in fig. 3.9. This source was fit to a template containing a composite of a starburst galaxy (20 %) and quasar (80 %), applying extinction with an E_{B-V} value of 0.05. The probability distribution shows a clear peak at the correct redshift. Note that the best stellar template is not able to reproduce the observed SED.

Some of the more interesting of these plots have been collected and are shown in sec. B.4.

3.4.4. Overview on the possible sources of insufficiencies

Compared to photo- z analyses in the literature (e. g. Salvato et al. (2009), Ananna et al. (2017)), the final outlier fractions and accuracies found for the sources in this thesis are worse than most results.

There are several reasons likely to be responsible for this discrepancy:

- **Photometric redshifts of quasars:**

As discussed in sec. 1.2.3, determining photo- z for quasars proves to be quite difficult due to their variability, their various types of appearances, their power-law-like SEDs with low amounts of features, and the unknown composition of quasar and host.

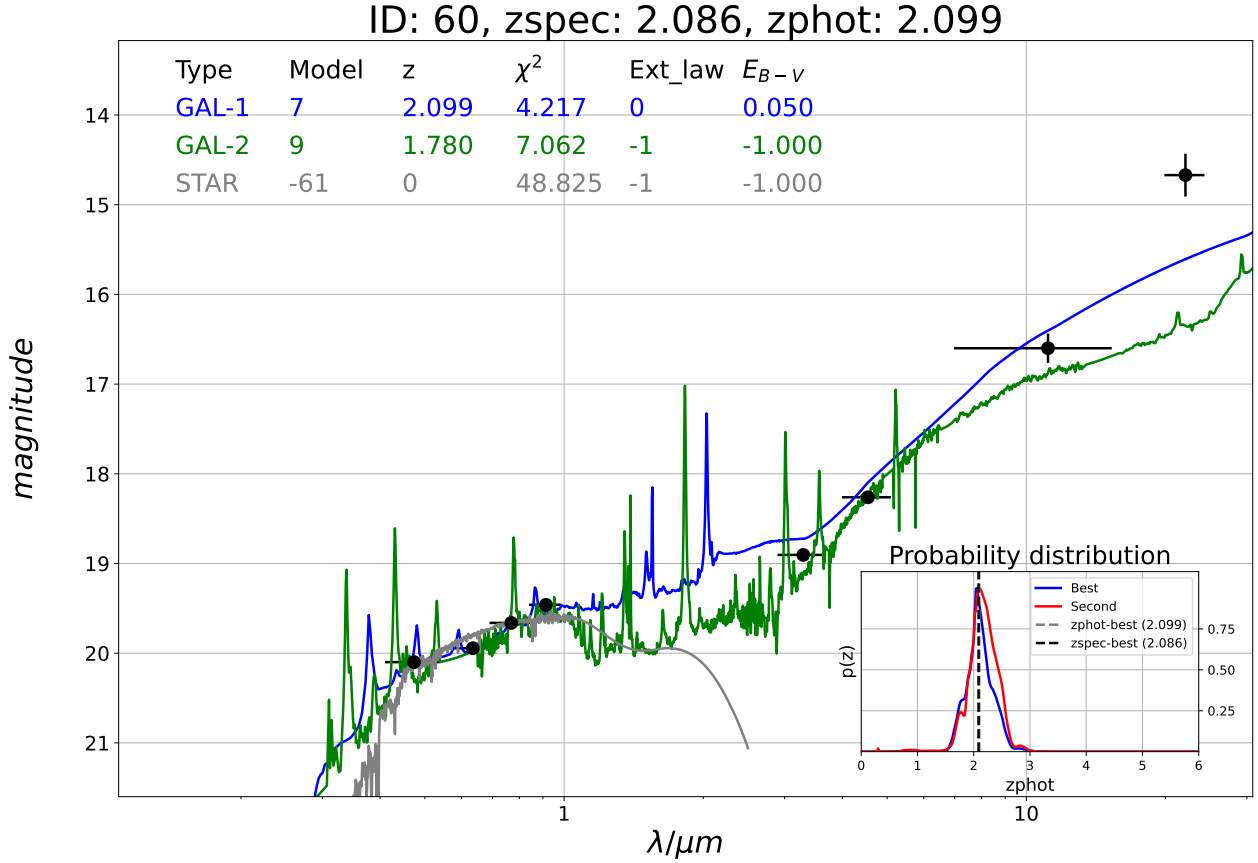


Figure 3.9: Spectral energy distribution of the best-fit templates for an exemplary source classified as pointlike.

The inset in the bottom right shows the probability distribution, while the black dots represent the photometric data points. The photometric errors are displayed by the vertical error bars while the horizontal error bars correspond to the effective widths of the respective filters (see tab. 2.1)

Nevertheless, e. g. Salvato et al. (2009) have shown that outlier fractions as low as $\eta_{\text{out}} = 0.063$ are possible for QSO samples when accounting for variability and using the right templates.

- **Insufficient templates:**

The templates that were used in this work were previously mostly applied to X-ray selected AGN, so utilising them for optically/mid-IR selected AGN might be the wrong approach^I. As shown e. g. in fig. 3.10a, some parts of the colour-redshift space are not yet covered properly. With a careful selection of additional templates and further optimisation, the performance of these photo- z might be improved.

- **Cross-matching different catalogues:**

Another source of errors might be erroneous matches between the different photometric catalogues, and generally the fact that photometry of miscellaneous catalogues is used^{II} for the photo- z . While the fitting parameter χ^2 does scale all templates with the same scaling factor s (compare eq. (2.8)), systematic offsets and shifts between the photometry of different surveys can lead to inaccurate fits.

- **Inclusion of background flux and wrong photometry:**

^Ias Menzel et al. (2016) note, the overlap between those different types of AGN might be smaller than anticipated.

^{II}especially since the source detection and flux extraction algorithms vary from catalogue to catalogue

As discussed in sec. 3.4.1, using the **AperMag6** VHS photometry for the extended sources might have led to a worse performance.

Especially for the GALEX photometry, the pointing errors are large. Therefore, unwanted flux of background sources might be included, distorting the observed SED and leading to inaccurate fits.

- **Split of the pointlike and extended sub-samples:**

For this thesis, the morphological classification was solely done by considering the **TYPE** of the source in the LS DR9 data.

Since both the templates and the magnitude priors rely on this distinction being correct, errors might be fatal and could explain a portion of the catastrophic outliers.

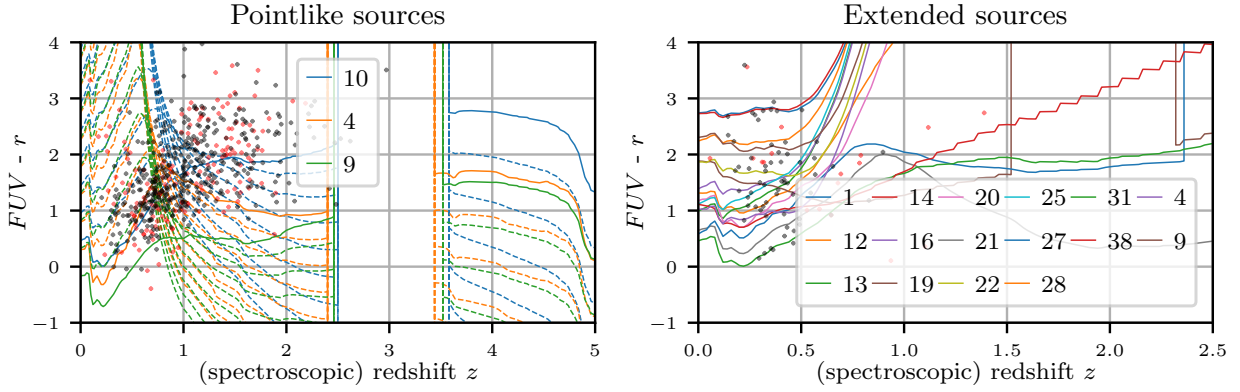
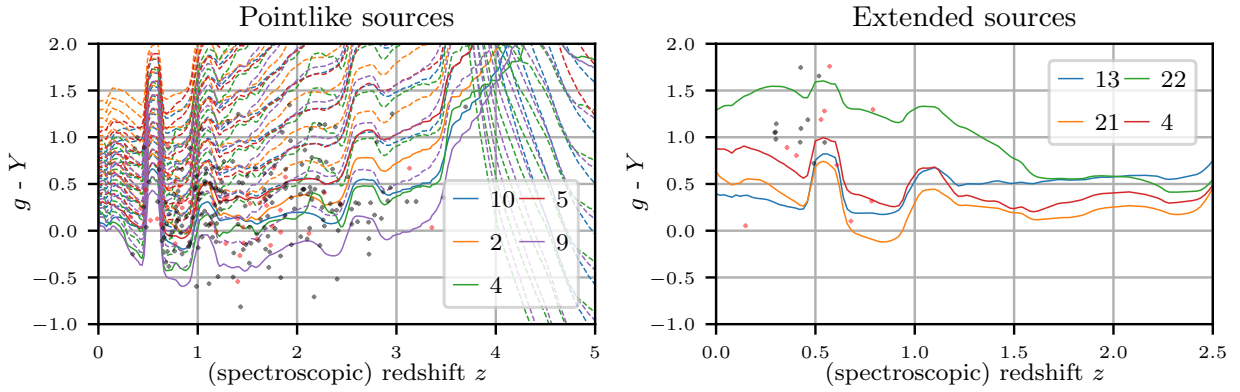
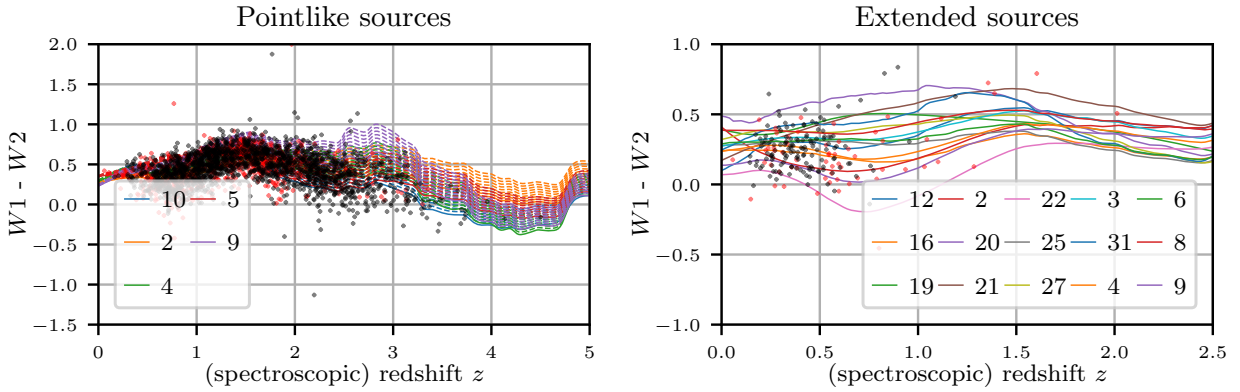
(a) Evolution of the $FUV - r$ colour with redshift.(b) Evolution of the $g - Y$ colour with redshift.(c) Evolution of the $W_1 - W_2$ colour with redshift.

Figure 3.10: Sample of different colour-redshift plots.

The black dots are sources that were fit within the outlier boundaries, the red dots symbolise sources considered as outliers due to their photo- z fit.

Only the templates that were fit are displayed and labelled with the numbers corresponding to tab. B.5 (pointlike) and tab. B.6 (extended). For the pointlike sources, the reddened templates for the different E_{B-V} are displayed as well.

4. Summary and outlook

In this thesis, the data pipeline for an input catalogue for the 4Hi-Q survey has been successfully assembled, with in-depth analysis carried out in the eFEDS (Brunner et al. 2021) field. The requirement of quasar sources with redshifts $z > 0.55$ implied the implementation of a photometric redshift routine since many of the identified quasars have been identified via imaging surveys and therefore have no spectroscopic redshifts available.

For the task of calculating the photometric redshifts, the method of template fitting was applied, utilising the LePhare++ code package (Ilbert et al. 2006, 2009).

To have the possibility of performing meaningful fits, photometry of different input catalogues available in the area of interest^I needed to be considered, cross-matched against a parent catalogue (Shu et al. 2019), and processed such that LePhare could accomplish the photometric redshift estimation.

Depending on the morphology, the sources were divided in a pointlike and an extended sub-sample.

To improve the results, which at first yielded an outlier fraction of $\eta_{\text{out}} = 0.358$ (0.330) for the pointlike (extended) sub-sample, the lists of templates were optimised, leading to an improved $\eta_{\text{out}} = 0.316$ (0.304).

Following up on these results, the possible impact of i band photometry on the photometric redshifts was analysed, yielding the best outlier fraction and accuracy of

$$\eta_{\text{out}} = 0.277 \text{ (0.299)} \quad \sigma_{\text{NMAD}} = 0.124 \text{ (0.125)}$$

employing the HSC (Miyazaki et al. 2018) survey i band data, and even an outlier fraction of $\eta_{\text{out}} = 0.244$ with preliminary i band data of the LS DR10.

The best false positive fraction ψ_{Pos} that could be achieved employing the improved templates and HSC i band data, which is the most important parameter concerning the 4Hi-Q input catalogue, was

$$\psi_{\text{Pos}} = 0.015 \text{ (0.005)}.$$

Possible improvements

With the pipeline available and described in sec. 2.4, the assembly of the input catalogue can be conducted for the whole sky with only a few adjustments^{II}.

The quality of the photometric redshifts might be improved even further, especially if one focuses on the errors described in sec. 3.4.4.

Final steps for the input catalogue include magnitude cuts to reach the source densities required for the 4Hi-Q survey. An alternative to magnitude cuts would be cuts that improve the reliability of the photometric redshifts, as discussed in sec. 3.4.1.

^Ifor the 4Hi-Q survey

^{II}due to the limited time, unfortunately this is not part of this thesis

A. Code snippets

A.1. Downloading the SWEEP files in a given region

```
# download_sweep_files.py
import math
import urllib

# Use the eFEDS field as an example
ra_min, ra_max = 126.5, 145.5
dec_min, dec_max = -3.2, 6.2

# Right ascension is taken in steps of 10, declination in steps of five
ra_min = 10*math.floor(ra_min/10)
ra_max = 10*math.ceil(ra_max/10)
dec_min = 5*math.floor(dec_min/5)
dec_max = 5*math.ceil(dec_max/5)

def sgnstr(dec):
    """Returns 'p' if dec is positive, 'm' if it's negative."""
    return "p" if dec >= 0 else "m"

def give_region_string(ra, dec):
    """Returns a string conforming to the naming convention of the SWEEP
    catalogues following the <AAA>c<BBB> pattern where AAA is the ra, c
    the p or m for the sign of dec and BBB is the dec."""
    return f"{ra:03}{sgnstr(dec)}{abs(dec):03}"

web_base = "https://portal.nersc.gov/cfs/cosmo/data/legacysurvey/dr9/south/
sweep/9.0/"
count = (ra_max - ra_min)/10*abs(dec_max-dec_min)/5
print(f"Fetching {int(count)} regions with ra between {ra_min} and {ra_max},
"
      f" dec between {dec_min} and {dec_max}")
directory = "sweep/" # Directory to store the data in
i = 1
for ra in range(ra_min, ra_max, 10):
    for dec in range(dec_min, dec_max, 5):
        reg_min, reg_max = give_region_string(ra, dec), \
            give_region_string(ra + 10, dec + 5)
        region = f"sweep-{reg_min}-{reg_max}.fits"
        website = web_base + region
        try:
            site = urllib.request.urlopen(website)
            print(f"Downloading {region}... Filesize:"
```

```

f" {round((site.length/1024**3),2)} GB")
urllib.request.urlretrieve(website, f"{directory}/SWEEP/{region}"
)
print(f"Downloaded and stored {region} in {directory} ({i} of {
int(count)})")
except urllib.error.HTTPError:
    print(f"Couldn't find {region}. Moving on to the next file.")
i += 1

```

A.2. Querying for VHS data

```

-- VHS query for the eFEDS field
-- -> YHJKs information
-- Query website: http://horus.roe.ac.uk:8080/vdfs/VSQL_form.jsp
-- (Note: Select VHS: VISTA Hemisphere Survey in the survey box!)
SELECT
    ra, dec, pStar, pGalaxy,
    yAperMag4, yAperMag4Err, yAperMag6, yAperMag6Err,
    jAperMag4, jAperMag4Err, jAperMag6, jAperMag6Err,
    hAperMag4, hAperMag4Err, hAperMag6, hAperMag6Err,
    ksAperMag4, ksAperMag4Err, ksAperMag6, ksAperMag6Err,
    eBV, aY, aJ, aH, aKs
FROM
    vhsSource
WHERE
    ra BETWEEN 126.5 AND 146.2
    AND dec BETWEEN -3.2 AND 6.2
    AND priOrSec=0

```

A.2.1. Querying for HSC data

```

-- HSC query for the eFEDS field
-- -> i band information (psf and cmodel flux)
-- Query website: https://hsc-release.mtk.nao.ac.jp/datasetsearch/?
-- (Note: Select PDR3 Citrus columnar in the release box!)
SELECT
    object_id, pdr3_wide.forced.ra, pdr3_wide.forced.dec, pdr3_wide.forced.
    a_i,
    pdr3_wide.forced2.i_psf_flux, pdr3_wide.forced2.i_psf_flux_err,
    pdr3_wide.forced.i_cmodel_flux, pdr3_wide.forced.i_cmodel_flux_err,
    pdr3_wide.meas.i_filterfraction_weighted
FROM
    pdr3_wide.forced
    LEFT JOIN pdr3_wide.forced2 USING (object_id)
    LEFT JOIN pdr3_wide.meas USING (object_id)
WHERE
    isprimary
    AND boxSearch(coord, 126.5, 146.2, -3.2, 6.2)

```

```

/* is equivalent to 'ra BETWEEN 126.5 AND 146.2 AND dec BETWEEN -3.2
AND 6.2',
but boxSearch() is much faster */
    AND (pdr3_wide.forced2.i_psf_flux IS NOT NULL OR pdr3_wide.forced
        .i_model_flux IS NOT NULL)

```

A.2.2. Querying for KiDS data

```

-- KiDS query for the eFEDS field
-- --> i band information (GAAP flux)
-- Query website: http://archive.eso.org/programmatic/#TAP
-- (Note: Select tap_cat in the TAP Service box!)
SELECT
    RAJ2000, DECJ2000, CLASS_STAR, MAG_GAAP_i, MAGERR_GAAP_i,
    EXTINCTION_i, Z_B, ODDS
FROM
    KiDS_DR4_1_ugriZYJKs_cat_fits
WHERE
    RAJ2000 between 126.5 and 146.2 AND DECJ2000 between -3.2 and 6.2
    AND FLAG_GAAP_i=0
    AND MAG_GAAP_i IS NOT NULL

```

A.2.3. The configuration file for the code

```

; Custom config.ini file with default parameters, explanations in the tables below
[GENERAL]
logging_level = 20
use_pointlike = True
use_extended = True
print_commands_only = False
ask_overwrite = True

[CAT_ASSEMBLY]
assemble_cat = False
cat_stem = baseline_input
use_matched = False
use_processed = False
reduce_to_specz = False
write_lephare_input = True
write_info_file = True

[LEPHARE]
para_stem = baseline
run_filters = False
filter_stem = baseline_filters
run_templates = False
extinc_range_pointlike = 0,0,0,30
template_stem = baseline_templates
run_zphota = False
forbidden_bands = ['i_hsc', 'i2_hsc', 'i_kids', 'i_ls10']

```

```
input_stem = baseline_input
output_stem = baseline_output
give_stats = True
spec_out = False
```

[PLOTTING]

```
input = False
sep = False
filters = False
output = True
template = False
```

B. Additional tables and figures

B.1. Code configuration parameters

Table B.1: General settings

Keyword	Default value	Behaviour
logging_level	20	This sets the minimum logging level. 10 corresponds to DEBUG, 20 to INFO, 30 to WARNING, 40 ERROR and 50 to CRITICAL ¹
use_pointlike	True	If set to False, all pointlike sources are not going to be considered for the run.
use_extended	True	If set to False, all extended sources are not going to be considered for the run.
print_commands_only	False	Only print shell commands instead of executing them. Useful for debugging purposes.
ask_overwrite	True	If set to False, files will be overwritten without extra permission.

¹More about logging levels can be found in the documentation of the **logging** module, which can be found [here](#).

Table B.2: Catalogue assembly settings

Keyword	Default value	Behaviour
<code>assemble_cat</code>	<code>False</code>	If set to <code>True</code> , an attempt to assemble the sources is performed, assuming the necessary catalogues are provided in the <code>CATPATH</code> directory.
<code>cat_stem</code>	<code>baseline</code>	File stem used for catalogue files.
<code>use_matched</code>	<code>False</code>	After the first run, an intermediary matched file is written, which may be used for further assemblies.
<code>use_processed</code>	<code>False</code>	After the first run, an intermediary processed file is written, which may be used for further assemblies.
<code>reduce_to_specz</code>	<code>False</code>	If set to <code>True</code> , all sources without spec- <i>z</i> are ignored.
<code>write_lephare_input</code>	<code>True</code>	Determine whether an actual input file for LePhare should be written.
<code>write_info_file</code>	<code>True</code>	If <code>True</code> , a small summary file about the input catalogue is written after its assembly.

Table B.3: LePhare settings

Keyword	Default value	Behaviour
<code>para_stem</code>	<code>baseline</code>	The name of the LePhare parameter files (one with a <code>_in.param</code> and one with a <code>_out.param</code> extension are expected).
<code>run_filters</code>	<code>False</code>	Determine whether the LePhare filter file should be assembled.
<code>filter_stem</code>	<code>baseline</code>	The name of the filter file (expected to be <code>baseline_filters.filter</code>).
<code>run_templates</code>	<code>False</code>	Determine whether the LePhare template library should be assembled for pointlike and extended sources with the <code>sedtolib</code> and <code>mag_gal</code> routines. Expects a filter file to be available.
<code>template_stem</code>	<code>baseline</code>	The stem of the template files. If <code>run_templates</code> is specified, <code>baseline_**type**.list</code> is expected for pointlike and extended.
<code>run_zphota</code>	<code>False</code>	Determine whether the photo- <i>z</i> program of LePhare should be run. Expects filter and template library files to be available.
<code>forbidden_bands</code>	<code>['i_kids', 'i_ls10', 'i_kids', 'i_ls10']</code>	Specify which of the 17 available bands should not be used. This is converted to a context for LePhare.
<code>input_stem</code>	<code>baseline</code>	The name of the LePhare input file (expects <code>baseline.in</code>).
<code>output_stem</code>	<code>baseline</code>	The name of the LePhare output files (with <code>.out</code> and <code>.fits</code> extensions).
<code>spec_out</code>	<code>False</code>	If set to <code>True</code> , the <code>SPEC_OUT</code> keyword will be set for LePhare, yielding <code>.spec</code> files for each source. Therefore, if this is specified, only the first 10 rows of the catalogue are going to be considered for the <code>zphota</code> run.

Table B.4: Settings concerning plotting

Keyword	Default value	Behaviour
<code>input</code>	<code>False</code>	Produce plots of the input distribution of sources.
<code>sep</code>	<code>False</code>	Produce separation plots with regard to the LS coordinates.
<code>filters</code>	<code>False</code>	Produce a joint filter transmission plot.
<code>output</code>	<code>True</code>	
<code>template</code>	<code>False</code>	

B.2. Templates

Table B.5: Final template library **combined_pointlike.list** for the pointlike sources.
The assembly of this final list is laid out in sec. 3.2.

ID	Name ^a	Type
1	pl_TQSO1	High IR lum. QSO ^b
2	NGC3783_Central_00	Composite AGN ^c
3	gh_s250_10_pl_TQSO1_90c	Hybrid ^d
4	Mrk110	Composite AGN ^c
5	NGC4151_Central	Composite AGN ^c
6	pl_QSOH	High lum. QSO ^e
7	gh_s800_20_pl_TQSO1_80c	Hybrid ^d
8	pl_QSO_DR2_029_t0	QSO ^e
9	3C120	Composite AGN ^c
10	pl_I22491_10_TQSO1_90	Starburst + Type 1 AGN ^f
11	S0_90_QSO2_10	Hybrid ^d
12	pl_I22491_20_TQSO1_80	Starburst + Type 1 AGN ^f
13	I22491_50_TQSO1_50	Starburst + Type 1 AGN ^f
14	pl_I22491_30_TQSO1_70	Starburst + Type 1 AGN ^f
15	gh_s180_30_pl_TQSO1_70c	Hybrid ^d

^a These names are the same as in their respective libraries.

^b Synthetic Type 1 AGN by Polletta et al. (2007).

^c Brown et al. (2019), available i. e. [here](#).

^d Hybrid models as presented in Salvato et al. (2009).

^e Template available at <http://classic.sdss.org/dr5/algorithms/spectemplates/>, presented in Salvato et al. (2009).

Composites of the I22491 starburst galaxy and a QSO template by Polletta et al. (2007) as presented in Salvato et al. (2009).

Table B.6: Final template library **combined_extended.list** for the extended sources.
The assembly of this final list is laid out in sec. 3.2.

ID	Name ^a	Type and reference
1	NGC4051_Central.00	Composite AGN ^b
2	NGC3227_Central	Composite AGN ^b
3	NGC3783_Central.00	Composite AGN ^b
4	NGC3783_Central.01	Composite AGN ^b
5	NGC4151_Central.08	Composite AGN ^b
6	NGC4151_Central.00	Composite AGN ^b
7	NGC5548_Central.02	Composite AGN ^b
8	NGC4051_Central	Composite AGN ^b
9	NGC5548_Central.01	Composite AGN ^b
10	3C390.3	Composite AGN ^b
11	NGC7469	Composite AGN ^b
12	NGC4051_Central.01	Composite AGN ^b
13	NGC3227_Central.04	Composite AGN ^b
14	NGC4151_Central.01	Composite AGN ^b
15	NGC5548_Central.04	Composite AGN ^b
16	Mrk817	Composite AGN ^b
17	pl_BQSO_Co19_sl-8	Type 1 AGN ^c
18	PG1211+143	AGN ^b
19	PG2349-014	AGN ^b
20	NGC3783_Central.04	Composite AGN ^b
21	NGC4151_Central.04	Composite AGN ^b
22	Mrk279	Composite AGN ^b
23	NGC4051_Central.02	Composite AGN ^b
24	Mrk926	Composite AGN ^b
25	PG0052+251	Quasar ^b
26	NGC4051_Central.64	Composite AGN ^b
27	NGC4051_Central.16	Composite AGN ^b
28	I22491_60_TQSO1_40	Starburst + Type 1 AGN ^d
29	I22491_50_TQSO1_50	Starburst + Type 1 AGN ^d
30	I22491_80_TQSO1_20	Starburst + Type 1 AGN ^d
31	pl_I22491_20_TQSO1_80	Starburst + Type 1 AGN ^d
32	pl_I22491_10_TQSO1_90	Starburst + Type 1 AGN ^d
33	M82	Starburst ^e

^a These names are the same as in their respective libraries.

^b A high portion of the templates are SEDs of single objects (including their names) and hybrids with different ratios of galactic and AGN portions by Brown et al. (2019), available i. e. [here](#).

^c Synthetic Type 1 AGN, similar to TQSO templates presented in Polletta et al. (2007), enhanced and used by Salvato et al. (2021) with less enhanced MIR flux.

^d Composites of the I22491 starburst galaxy and a QSO template by Polletta et al. (2007) as presented in Salvato et al. (2009).

^e Polletta et al. (2007).

B.3. Band usage overview

Table B.7: Comparison of the different results in the i band for the sources with spec- z available.

The rows that were reduced to the respective bands may not be meaningful as no information is available on whether LePhare actually employed this band for all sources during the fitting process.

pointlike					
Bands	N_{source}	η_{out}	σ_{NMAD}	ψ_{Pos}	ψ_{Neg}
All bands ^a	2569	0.2772	0.1242	0.0152	0.0529
Reduced to FUV	582	0.2835	0.1184	0.0258	0.0911
Reduced to NUV	1189	0.2817	0.1187	0.0168	0.0749
Reduced to Y	241	0.2324	0.1124	0.0124	0.0581
Reduced to J	486	0.2449	0.1115	0.0165	0.0597
Reduced to H	215	0.2605	0.1187	0.0140	0.0651
Reduced to K_s	450	0.2422	0.1115	0.0178	0.0622
Reduced to W_3	2434	0.2777	0.1235	0.0160	0.0534
Reduced to W_4	1918	0.2956	0.1288	0.0182	0.0563
Reduced to i_{HSC} (joint)	1972	0.2693	0.1210	0.0122	0.0472
extended					
Bands	N_{source}	η_{out}	σ_{NMAD}	ψ_{Pos}	ψ_{Neg}
All bands ^a	197	0.2995	0.1251	0.0051	0.2081
Reduced to FUV	81	0.2222	0.0963	0.0000	0.1235
Reduced to NUV	95	0.2632	0.1112	0.0000	0.1684
Reduced to Y	21	0.4762	0.1795	0.0000	0.2857
Reduced to J	44	0.3182	0.1431	0.0000	0.2273
Reduced to H	20	0.5000	0.2066	0.0000	0.3000
Reduced to K_s	45	0.3333	0.1464	0.0000	0.2444
Reduced to W_3	186	0.3011	0.1315	0.0054	0.2043
Reduced to W_4	144	0.2778	0.1191	0.0069	0.1875
Reduced to i_{HSC} (joint)	142	0.3099	0.1274	0.0000	0.2183

^a equivalent to g , z , r , W_1 and W_2 band reduced data.

B.4. More individual SED plots

The following plots show the spectral energy distribution of the best-fit templates for exemplary sources.

The inset in the bottom right shows the probability distribution, while the black dots represent the photometric data points. The photometric errors are displayed by the vertical error bars while the horizontal error bars correspond to the effective widths of the respective filters (see tab. 2.1).

Discussion of these types of plots is provided in sec. 3.4.3.

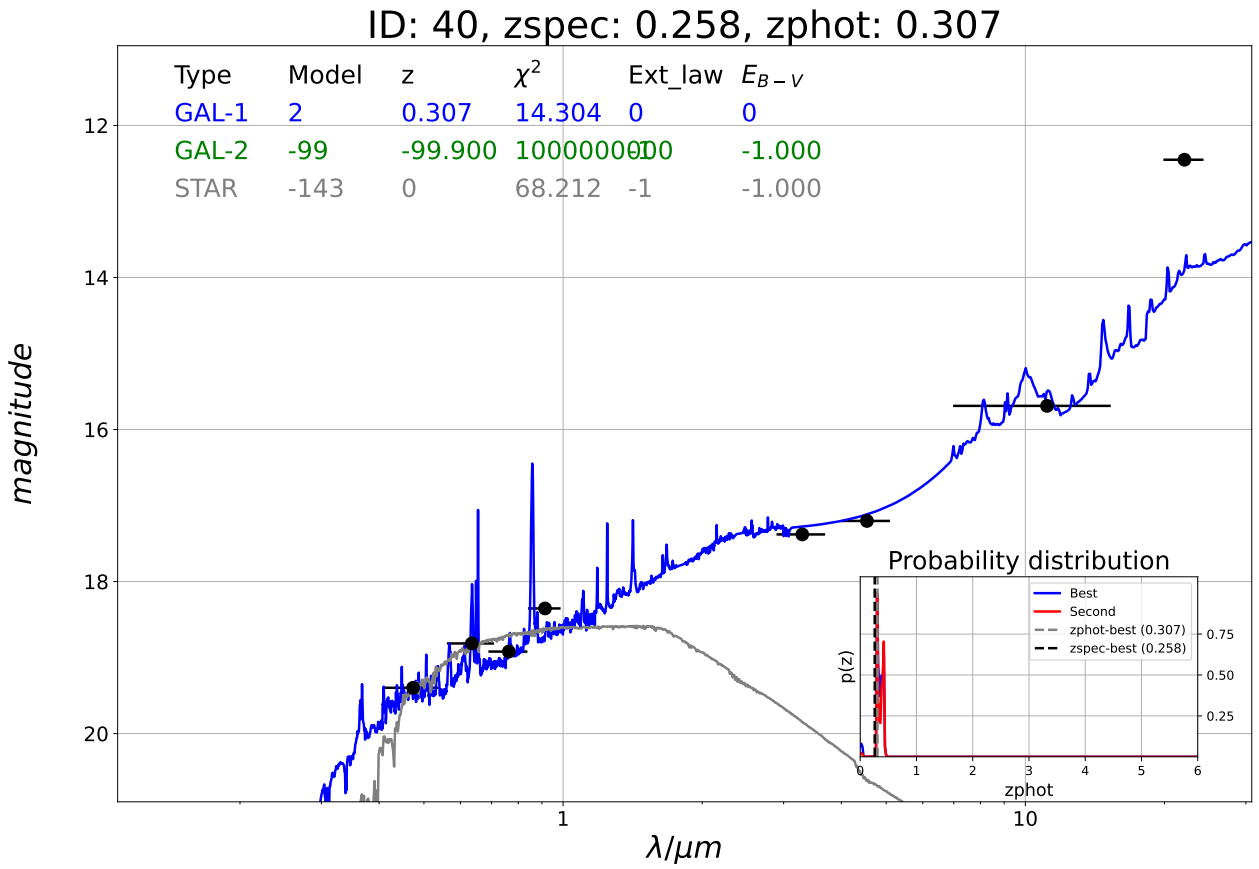


Figure B.1: An extended source. The best-fit model seems to represent the source very well, but the most likely redshift is still off.

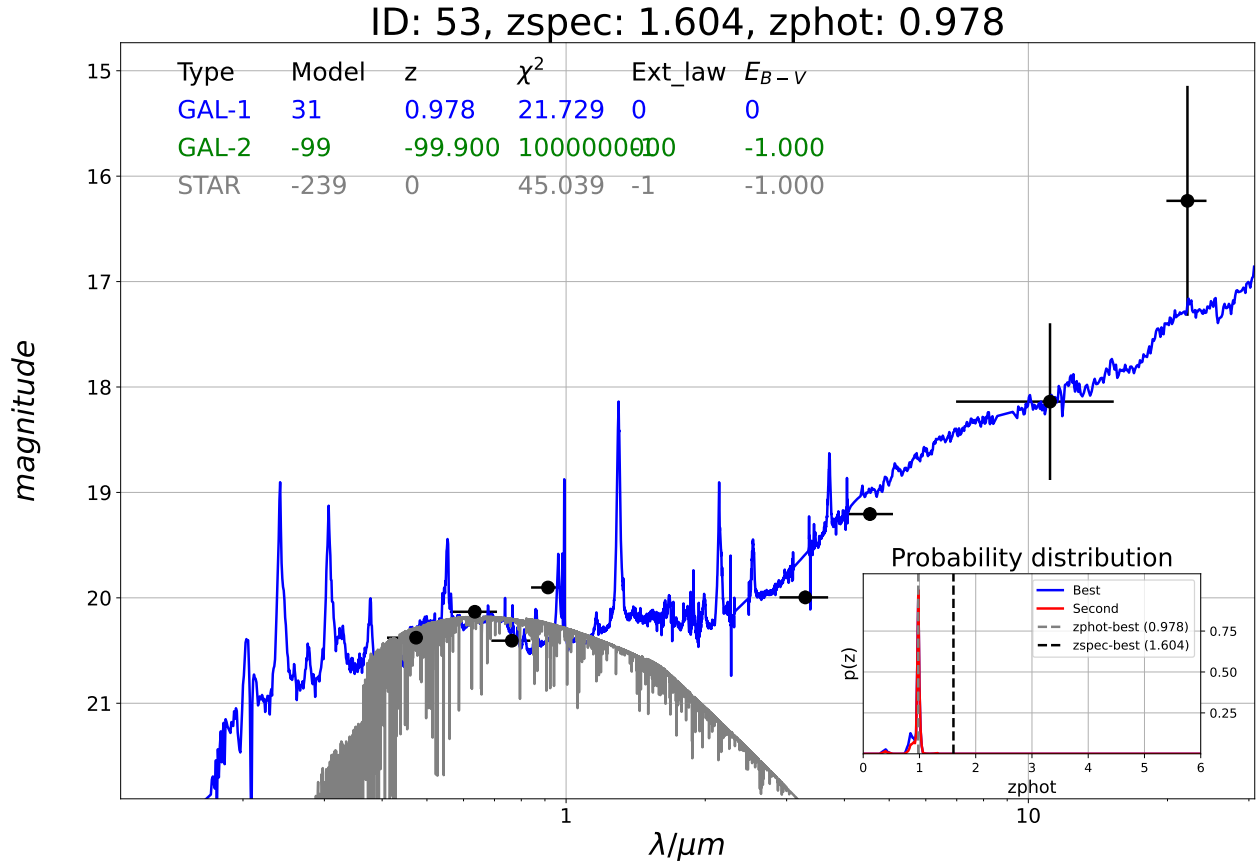


Figure B.2: An extended source. Here, the source seems to have been fit to the wrong template. A template actually reflecting the SED might not have been available.

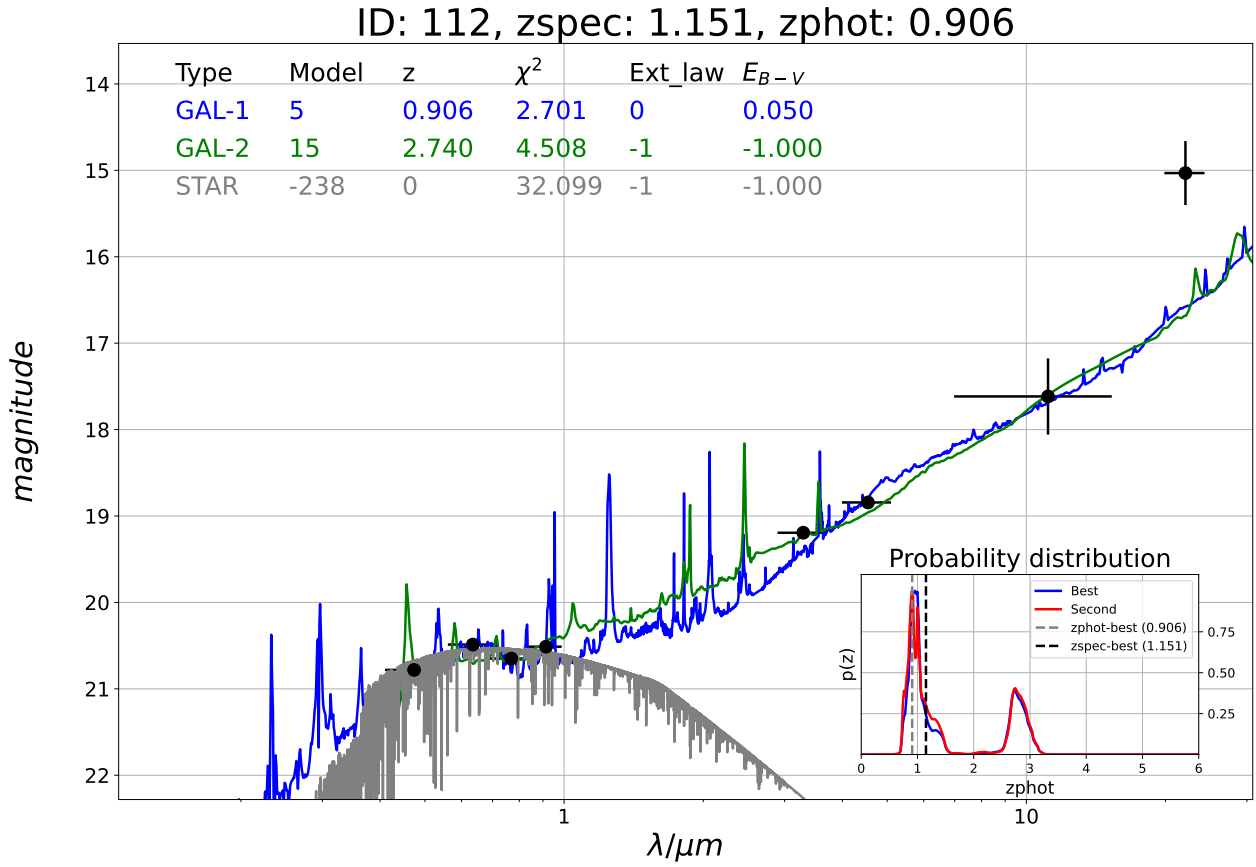


Figure B.3: A pointlike source. Here, a secondary model at a completely different (and wrong) redshift seems to fit the source' photometry almost as well as the best fit with low χ^2 .

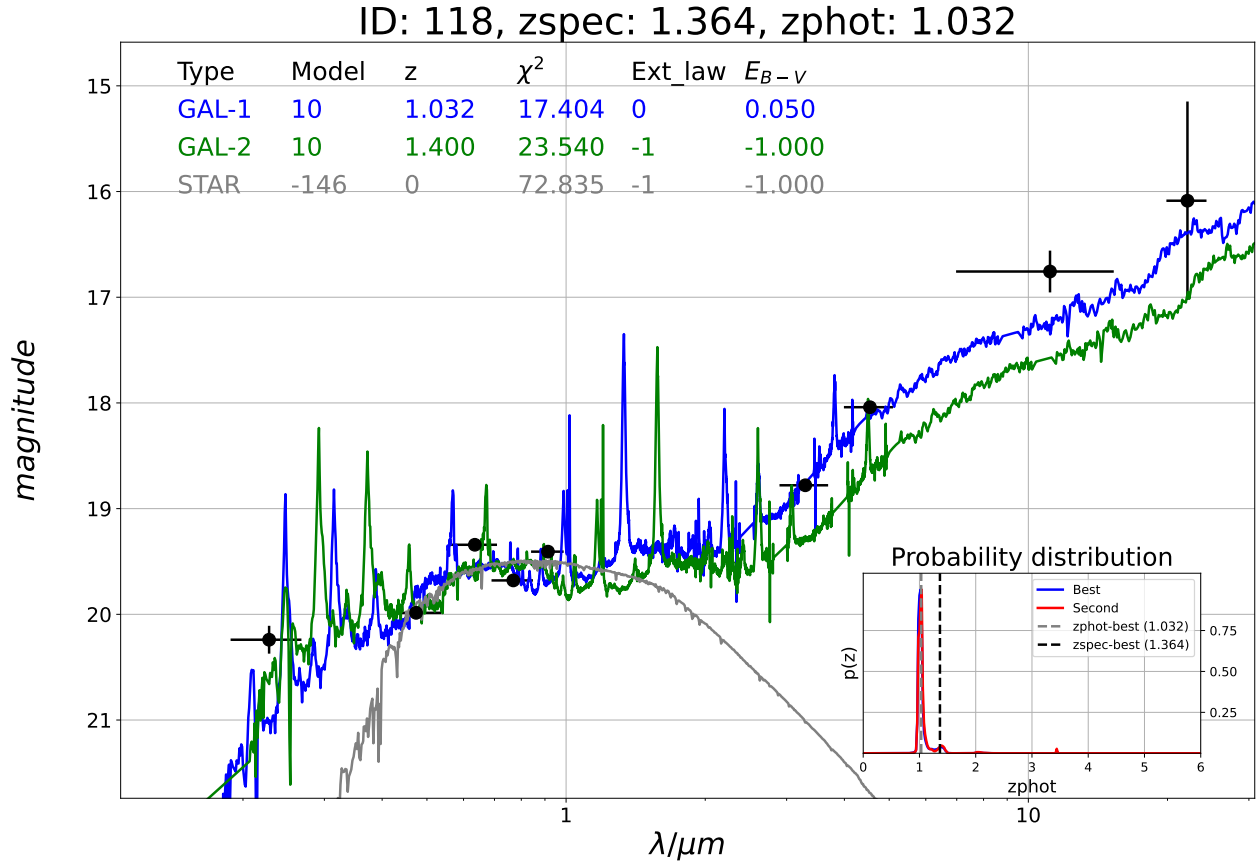


Figure B.4: A pointlike source. Here, the template with ID 10 was able to fit the source' both with a redshift of $z_{\text{phot}} = 1.0$ (with an E_{B-V} of 0.05) and $z_{\text{phot}} = 1.4$. According to z_{spec} , the latter would have been the better fit, but the degeneracy was too high.

C. Acknowledgements

C.1. Personal acknowledgements

Here, I'd like to give a huge thanks to all of the great people that supported me with this thesis, most notably Mara Salvato.

I'm also very thankful for the comments and additional guidance of Celiné Peroux and Jochen Liske.

Additionally, I'd like to thank my quick proof readers that hopefully caught all of the ever-so-subtle mistakes, which are Colin Ihlenfeldt, Jörn Bach, Annalena Friedrich, and Julian David Kuhlmann.

C.2. Legacy Surveys

The Legacy Surveys consist of three individual and complementary projects: the Dark Energy Camera Legacy Survey (DECaLS; Proposal ID #2014B-0404; PIs: David Schlegel and Arjun Dey), the Beijing-Arizona Sky Survey (BASS; NOAO Prop. ID #2015A-0801; PIs: Zhou Xu and Xiaohui Fan), and the Mayall z-band Legacy Survey (MzLS; Prop. ID #2016A-0453; PI: Arjun Dey). DECaLS, BASS and MzLS together include data obtained, respectively, at the Blanco telescope, Cerro Tololo Inter-American Observatory, NSF's NOIRLab; the Bok telescope, Steward Observatory, University of Arizona; and the Mayall telescope, Kitt Peak National Observatory, NOIRLab. The Legacy Surveys project is honored to be permitted to conduct astronomical research on Iolkam Du'ag (Kitt Peak), a mountain with particular significance to the Tohono O'odham Nation.

NOIRLab is operated by the Association of Universities for Research in Astronomy (AURA) under a cooperative agreement with the National Science Foundation.

This project used data obtained with the Dark Energy Camera (DECam), which was constructed by the Dark Energy Survey (DES) collaboration. Funding for the DES Projects has been provided by the U.S. Department of Energy, the U.S. National Science Foundation, the Ministry of Science and Education of Spain, the Science and Technology Facilities Council of the United Kingdom, the Higher Education Funding Council for England, the National Center for Supercomputing Applications at the University of Illinois at Urbana-Champaign, the Kavli Institute of Cosmological Physics at the University of Chicago, Center for Cosmology and Astro-Particle Physics at the Ohio State University, the Mitchell Institute for Fundamental Physics and Astronomy at Texas A&M University, Financiadora de Estudos e Projetos, Fundacao Carlos Chagas Filho de Amparo, Financiadora de Estudos e Projetos, Fundacao Carlos Chagas Filho de Amparo a Pesquisa do Estado do Rio de Janeiro, Conselho Nacional de Desenvolvimento Cientifico e Tecnologico and the Ministerio da Ciencia, Tecnologia e Inovacao, the Deutsche Forschungsgemeinschaft and the Collaborating Institutions in the Dark Energy Survey. The Collaborating Institutions are Argonne National Laboratory, the University of California at Santa Cruz, the University of Cambridge, Centro de Investigaciones Energeticas, Medioambientales y Tecnologicas-Madrid, the University of Chicago, University College London, the DES-Brazil Consortium, the University of Edinburgh, the Eidgenossische Technische Hochschule (ETH) Zurich, Fermi National Accelerator Laboratory, the University of Illinois at Urbana-Champaign, the Institut de Ciencies de l'Espai (IEEC/C-SIC), the Institut de Fisica d'Altes Energies, Lawrence Berkeley National Laboratory, the Ludwig Maximilians Universitat Munchen and the associated Excellence Cluster Universe, the University of Michigan, NSF's NOIRLab, the University of Nottingham, the Ohio State University, the University of Pennsylvania, the University of Portsmouth, SLAC National Accelerator Laboratory, Stanford University, the University of Sussex, and Texas A&M Uni-

versity.

BASS is a key project of the Telescope Access Program (TAP), which has been funded by the National Astronomical Observatories of China, the Chinese Academy of Sciences (the Strategic Priority Research Program “The Emergence of Cosmological Structures” Grant # XDB09000000), and the Special Fund for Astronomy from the Ministry of Finance. The BASS is also supported by the External Cooperation Program of Chinese Academy of Sciences (Grant # 114A11KYSB20160057), and Chinese National Natural Science Foundation (Grant # 11433005).

The Legacy Survey team makes use of data products from the Near-Earth Object Wide-field Infrared Survey Explorer (NEOWISE), which is a project of the Jet Propulsion Laboratory/California Institute of Technology. NEOWISE is funded by the National Aeronautics and Space Administration.

The Legacy Surveys imaging of the DESI footprint is supported by the Director, Office of Science, Office of High Energy Physics of the U.S. Department of Energy under Contract No. DE-AC02-05CH1123, by the National Energy Research Scientific Computing Center, a DOE Office of Science User Facility under the same contract; and by the U.S. National Science Foundation, Division of Astronomical Sciences under Contract No. AST-0950945 to NOAO.

C.3. GALEX

This research made use of the cross-match service provided by CDS, Strasbourg.

C.4. KiDS data

Based on observations made with ESO Telescopes at the La Silla Paranal Observatory under programme IDs 177.A-3016, 177.A-3017, 177.A-3018 and 179.A-2004, and on data products produced by the KiDS consortium. The KiDS production team acknowledges support from: Deutsche Forschungsgemeinschaft, ERC, NOVA and NWO-M grants; Target; the University of Padova, and the University Federico II (Naples).

C.5. HyperSuprimeCam

The Hyper Suprime-Cam (HSC) collaboration includes the astronomical communities of Japan and Taiwan, and Princeton University. The HSC instrumentation and software were developed by the National Astronomical Observatory of Japan (NAOJ), the Kavli Institute for the Physics and Mathematics of the Universe (Kavli IPMU), the University of Tokyo, the High Energy Accelerator Research Organization (KEK), the Academia Sinica Institute for Astronomy and Astrophysics in Taiwan (ASIAA), and Princeton University. Funding was contributed by the FIRST program from Japanese Cabinet Office, the Ministry of Education, Culture, Sports, Science and Technology (MEXT), the Japan Society for the Promotion of Science (JSPS), Japan Science and Technology Agency (JST), the Toray Science Foundation, NAOJ, Kavli IPMU, KEK, ASIAA, and Princeton University.

This paper makes use of software developed for the Large Synoptic Survey Telescope. We thank the LSST Project for making their code available as free software at <http://dm.lsst.org>

The Pan-STARRS1 Surveys (PS1) have been made possible through contributions of the Institute for Astronomy, the University of Hawaii, the Pan-STARRS Project Office, the Max-Planck Society and its participating institutes, the Max Planck Institute for Astronomy, Heidelberg and the Max Planck Institute for Extraterrestrial Physics, Garching, The Johns Hopkins University, Durham University, the University of Edinburgh, Queen’s University

Belfast, the Harvard-Smithsonian Center for Astrophysics, the Las Cumbres Observatory Global Telescope Network Incorporated, the National Central University of Taiwan, the Space Telescope Science Institute, the National Aeronautics and Space Administration under Grant No. NNX08AR22G issued through the Planetary Science Division of the NASA Science Mission Directorate, the National Science Foundation under Grant No. AST-1238877, the University of Maryland, and Eotvos Lorand University (ELTE) and the Los Alamos National Laboratory. Based [in part] on data collected at the Subaru Telescope and retrieved from the HSC data archive system, which is operated by Subaru Telescope and Astronomy Data Center at National Astronomical Observatory of Japan.

References

- R. Ahumada, C. A. Prieto, A. Almeida, et al. The 16th Data Release of the Sloan Digital Sky Surveys: First Release from the APOGEE-2 Southern Survey and Full Release of eBOSS Spectra. *Astrophysical Journal, Supplement*, 249(1):3, July 2020. doi: 10.3847/1538-4365/ab929e.
- T. T. Ananna, M. Salvato, S. LaMassa, et al. AGN Populations in Large-volume X-Ray Surveys: Photometric Redshifts and Population Types Found in the Stripe 82X Survey. *Astrophysical Journal*, 850(1):66, Nov. 2017. doi: 10.3847/1538-4357/aa937d.
- S. Arnouts, S. Cristiani, L. Moscardini, et al. Measuring and modelling the redshift evolution of clustering: the Hubble Deep Field North. *Monthly Notices of the RAS*, 310(2):540–556, Dec. 1999. doi: 10.1046/j.1365-8711.1999.02978.x.
- W. A. Baum. Photoelectric determinations of redshifts beyond 0.2 c. *Astronomical Journal*, 62:6–7, Feb. 1957. doi: 10.1086/107433.
- E. Bertin. Automated Morphometry with SExtractor and PSFEx. In I. N. Evans, A. Accomazzi, D. J. Mink, and A. H. Rots, editors, *Astronomical Data Analysis Software and Systems XX*, volume 442 of *Astronomical Society of the Pacific Conference Series*, page 435, July 2011.
- L. Bianchi. The Galaxy Evolution Explorer (GALEX). Its legacy of UV surveys, and science highlights. *Astrophysics and Space Science*, 354(1):103–112, Nov. 2014. doi: 10.1007/s10509-014-1935-6.
- L. Bianchi, B. Shiao, and D. Thilker. Revised Catalog of GALEX Ultraviolet Sources. I. The All-Sky Survey: GUVcat_AIS. *Astrophysical Journal, Supplement*, 230(2):24, June 2017. doi: 10.3847/1538-4365/aa7053.
- M. Bolzonella, J. M. Miralles, and R. Pelló. Photometric redshifts based on standard SED fitting procedures. *Astronomy and Astrophysics*, 363:476–492, Nov. 2000.
- J. Bosch, R. Armstrong, S. Bickerton, et al. The Hyper Suprime-Cam software pipeline. *Publications of the Astronomical Society of Japan*, 70:S5, Jan. 2018. doi: 10.1093/pasj/psx080.
- G. B. Brammer, P. G. van Dokkum, and P. Coppi. EAZY: A Fast, Public Photometric Redshift Code. *Astrophysics Source Code Library*, record ascl:1010.052, Oct. 2010.

- M. Brescia, M. Salvato, S. Caviuoti, et al. Photometric redshifts for X-ray-selected active galactic nuclei in the eROSITA era. *Monthly Notices of the RAS*, 489(1):663–680, Oct. 2019. doi: 10.1093/mnras/stz2159.
- M. J. I. Brown, K. J. Duncan, H. Landt, et al. The spectral energy distributions of active galactic nuclei. *Monthly Notices of the RAS*, 489(3):3351–3367, Nov. 2019. doi: 10.1093/mnras/stz2324.
- H. Brunner, T. Liu, G. Lamer, et al. The eROSITA Final Equatorial Depth Survey (eFEDS): The X-ray catalog. *arXiv e-prints*, art. arXiv:2106.14517, June 2021.
- D. Calzetti, L. Armus, R. C. Bohlin, et al. The Dust Content and Opacity of Actively Star-forming Galaxies. *Astrophysical Journal*, 533(2):682–695, Apr. 2000. doi: 10.1086/308692.
- G. Chabrier, I. Baraffe, F. Allard, and P. Hauschildt. Evolutionary Models for Very Low-Mass Stars and Brown Dwarfs with Dusty Atmospheres. *Astrophysical Journal*, 542(1):464–472, Oct. 2000. doi: 10.1086/309513.
- J. Comparat, A. Merloni, T. Dwelly, et al. The final SDSS-IV/SPIDERS X-ray point source spectroscopic catalogue. *Astronomy and Astrophysics*, 636:A97, Apr. 2020. doi: 10.1051/0004-6361/201937272.
- T. Dahlen, B. Mobasher, S. M. Faber, et al. A Critical Assessment of Photometric Redshift Methods: A CANDELS Investigation. *Astrophysical Journal*, 775(2):93, Oct. 2013. doi: 10.1088/0004-637X/775/2/93.
- R. S. de Jong, O. Agertz, A. A. Berbel, et al. 4MOST: Project overview and information for the First Call for Proposals. *The Messenger*, 175:3–11, Mar. 2019. doi: 10.18727/0722-6691/5117.
- S. P. Driver, J. Liske, L. J. M. Davies, et al. 4MOST Consortium Survey 7: Wide-Area VISTA Extragalactic Survey (WAVES). *The Messenger*, 175:46–49, Mar. 2019. doi: 10.18727/0722-6691/5126.
- R. Feldmann, C. M. Carollo, C. Porciani, et al. The Zurich Extragalactic Bayesian Redshift Analyzer and its first application: COSMOS. *Monthly Notices of the RAS*, 372(2):565–577, Oct. 2006. doi: 10.1111/j.1365-2966.2006.10930.x.
- A. Galametz, R. Saglia, S. Paltani, et al. SED-dependent galactic extinction prescription for Euclid and future cosmological surveys. *Astronomy and Astrophysics*, 598:A20, Feb. 2017. doi: 10.1051/0004-6361/201629333.
- J. D. Hunter. Matplotlib: A 2d graphics environment. *Computing in Science & Engineering*, 9(3):90–95, 2007. doi: 10.1109/MCSE.2007.55.
- O. Ilbert, S. Arnouts, H. J. McCracken, et al. Accurate photometric redshifts for the CFHT legacy survey calibrated using the VIMOS VLT deep survey. *Astronomy and Astrophysics*, 457(3):841–856, Oct. 2006. doi: 10.1051/0004-6361:20065138.
- O. Ilbert, P. Capak, M. Salvato, et al. Cosmos Photometric Redshifts with 30-Bands for 2-deg². *Astrophysical Journal*, 690(2):1236–1249, Jan. 2009. doi: 10.1088/0004-637X/690/2/1236.

- M. J. Irwin, J. Lewis, S. Hodgkin, et al. VISTA data flow system: pipeline processing for WFCAM and VISTA. In P. J. Quinn and A. Bridger, editors, *Optimizing Scientific Return for Astronomy through Information Technologies*, volume 5493 of *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, pages 411–422, Sept. 2004. doi: 10.1117/12.551449.
- D. C. Koo. Overview - Photometric Redshifts: A Perspective from an Old-Timer[!] on their Past, Present, and Potential. In R. Weymann, L. Storrie-Lombardi, M. Sawicki, and R. Brunner, editors, *Photometric Redshifts and the Detection of High Redshift Galaxies*, volume 191 of *Astronomical Society of the Pacific Conference Series*, page 3, Jan. 1999.
- J. Kormendy and D. Richstone. Inward Bound—The Search For Supermassive Black Holes In Galactic Nuclei. *Annual Review of Astronomy and Astrophysics*, 33:581, Jan. 1995. doi: 10.1146/annurev.aa.33.090195.003053.
- R. G. Kron. Photometry of a complete sample of faint galaxies. *Astrophysical Journal, Supplement*, 43:305–325, June 1980. doi: 10.1086/190669.
- K. Kuijken, C. Heymans, A. Dvornik, et al. The fourth data release of the Kilo-Degree Survey: ugri imaging and nine-band optical-IR photometry over 1000 square degrees. *Astronomy and Astrophysics*, 625:A2, May 2019. doi: 10.1051/0004-6361/201834918.
- D. Lang, D. W. Hogg, and D. Mykytyn. The Tractor: Probabilistic astronomical source detection and measurement, Apr. 2016.
- D. Le Borgne, M. Fioc, A. Lançon, et al. PÉGASE-HR: Stellar Population Synthesis at High Resolution Spectra. Astrophysics Source Code Library, record ascl:1108.008, Aug. 2011.
- O. Le Fèvre, G. Vettolani, B. Garilli, et al. The VIMOS VLT deep survey. First epoch VVDS-deep survey: 11 564 spectra with $17.5 \leq \text{IAB} \leq 24$, and the redshift distribution over $0 \leq z \leq 5$. *Astronomy and Astrophysics*, 439(3):845–862, Sept. 2005. doi: 10.1051/0004-6361:20041960.
- C. Maraston. Evolutionary population synthesis: models, analysis of the ingredients and application to high- z galaxies. *Monthly Notices of the RAS*, 362(3):799–825, Sept. 2005. doi: 10.1111/j.1365-2966.2005.09270.x.
- D. Masters, P. Capak, D. Stern, et al. Mapping the Galaxy Color-Redshift Relation: Optimal Photometric Redshift Calibration Strategies for Cosmology Surveys. *Astrophysical Journal*, 813(1):53, Nov. 2015. doi: 10.1088/0004-637X/813/1/53.
- R. G. McMahon, M. Banerji, E. Gonzalez, et al. First Scientific Results from the VISTA Hemisphere Survey (VHS). *The Messenger*, 154:35–37, Dec. 2013.
- M. L. Menzel, A. Merloni, A. Georgakakis, et al. A spectroscopic survey of X-ray-selected AGNs in the northern XMM-XXL field. *Monthly Notices of the RAS*, 457(1):110–132, Mar. 2016. doi: 10.1093/mnras/stv2749.
- S. Miyazaki, Y. Komiyama, S. Kawanomoto, et al. Hyper Suprime-Cam: System design and verification of image quality. *Publications of the Astronomical Society of Japan*, 70:S1, Jan. 2018. doi: 10.1093/pasj/psx063.
- P. Morrissey, T. Conrow, T. A. Barlow, et al. The Calibration and Data Products of GALEX. *Astrophysical Journal, Supplement*, 173(2):682–697, Dec. 2007. doi: 10.1086/520512.

- J. A. Newman, A. Abate, F. B. Abdalla, et al. Spectroscopic needs for imaging dark energy experiments. *Astroparticle Physics*, 63:81–100, Mar. 2015. doi: 10.1016/j.astropartphys.2014.06.007.
- M. S. Peeples, J. K. Werk, J. Tumlinson, et al. A Budget and Accounting of Metals at $z \sim 0$: Results from the COS-Halos Survey. *Astrophysical Journal*, 786(1):54, May 2014. doi: 10.1088/0004-637X/786/1/54.
- C. Péroux and J. C. Howk. The Cosmic Baryon and Metal Cycles. *Annual Review of Astronomy and Astrophysics*, 58:363–406, Aug. 2020a. doi: 10.1146/annurev-astro-021820-120014.
- C. Péroux and J. C. Howk. The Cosmic Baryon and Metal Cycles. *Annual Review of Astronomy and Astrophysics*, 58:363–406, Aug. 2020b. doi: 10.1146/annurev-astro-021820-120014.
- V. Petrosian. Surface Brightness and Evolution of Galaxies. *Astrophysical Journal*, 210:L53, Dec. 1976. doi: 10.1086/182301.
- M. Pettini. Element Abundances at High Redshifts. In J. R. Walsh and M. R. Rosa, editors, *Chemical Evolution from Zero to High Redshift*, page 233, Jan. 1999.
- A. J. Pickles. A Stellar Spectral Flux Library: 1150-25000 Å. *Publications of the Astronomical Society of the Pacific*, 110(749):863–878, July 1998. doi: 10.1086/316197.
- M. Polletta, M. Tajer, L. Maraschi, et al. Spectral Energy Distributions of Hard X-Ray Selected Active Galactic Nuclei in the XMM-Newton Medium Deep Survey. *Astrophysical Journal*, 663(1):81–102, July 2007. doi: 10.1086/518113.
- P. Predehl, R. Andritschke, V. Arefiev, et al. The eROSITA X-ray telescope on SRG. *Astronomy and Astrophysics*, 647:A1, Mar. 2021. doi: 10.1051/0004-6361/202039313.
- M. L. Prevot, J. Lequeux, E. Maurice, et al. The typical interstellar extinction in the Small Magellanic Cloud. *Astronomy and Astrophysics*, 132:389–392, Mar. 1984.
- J. J. Puschell, F. N. Owen, and R. A. Laing. Near-infrared photometry of distant radio galaxies - Spectral flux distributions and redshift estimates. *Astrophysical Journal*, 257: L57–L61, June 1982. doi: 10.1086/183808.
- J. Richard, J. P. Kneib, C. Blake, et al. 4MOST Consortium Survey 8: Cosmology Redshift Survey (CRS). *The Messenger*, 175:50–53, Mar. 2019. doi: 10.18727/0722-6691/5127.
- M. Salvato, G. Hasinger, O. Ilbert, et al. Photometric Redshift and Classification for the XMM-COSMOS Sources. *Astrophysical Journal*, 690(2):1250–1263, Jan. 2009. doi: 10.1088/0004-637X/690/2/1250.
- M. Salvato, O. Ilbert, and B. Hoyle. The many flavours of photometric redshifts. *Nature Astronomy*, 3:212–222, June 2019. doi: 10.1038/s41550-018-0478-0.
- M. Salvato, J. Wolf, T. Dwelly, et al. The eROSITA Final Equatorial-Depth Survey (eFEDS): Identification and characterization of the counterparts to the point-like sources. *arXiv e-prints*, art. arXiv:2106.14520, June 2021.
- D. Schlegel, A. Dey, D. Herrera, et al. DESI Legacy Imaging Surveys Data Release 9. In *American Astronomical Society Meeting Abstracts*, volume 53 of *American Astronomical Society Meeting Abstracts*, page 235.03, Jan. 2021.

- M. Schmidt. 3C 273 : A Star-Like Object with Large Red-Shift. *Nature*, 197(4872):1040, Mar. 1963. doi: 10.1038/1971040a0.
- G. A. Shields. A Brief History of Active Galactic Nuclei. *Publications of the Astronomical Society of the Pacific*, 111(760):661–678, June 1999. doi: 10.1086/316378.
- Y. Shu, S. E. Koposov, N. W. Evans, et al. Catalogues of active galactic nuclei from Gaia and unWISE data. *Monthly Notices of the RAS*, 489(4):4741–4759, Nov. 2019. doi: 10.1093/mnras/stz2487.
- M. B. Taylor. STILTS - A Package for Command-Line Processing of Tabular Data. In C. Gabriel, C. Arviset, D. Ponz, and S. Enrique, editors, *Astronomical Data Analysis Software and Systems XV*, volume 351 of *Astronomical Society of the Pacific Conference Series*, page 666, July 2006.
- J. Tumlinson, M. S. Peebles, and J. K. Werk. The Circumgalactic Medium. *Annual Review of Astronomy and Astrophysics*, 55(1):389–432, Aug. 2017. doi: 10.1146/annurev-astro-091916-055240.
- M. L. Turner, J. Schaye, R. A. Crain, et al. A comparison of observed and simulated absorption from H I, C IV, and Si IV around $z \approx 2$ star-forming galaxies suggests redshift-space distortions are due to inflows. *Monthly Notices of the RAS*, 471(1):690–705, Oct. 2017. doi: 10.1093/mnras/stx1616.
- C. M. Urry and P. Padovani. Unified Schemes for Radio-Loud Active Galactic Nuclei. *Publications of the Astronomical Society of the Pacific*, 107:803, Sept. 1995. doi: 10.1086/133630.
- G. van Rossum. Python tutorial. Technical Report CS-R9526, Centrum voor Wiskunde en Informatica (CWI), Amsterdam, May 1995.
- Wes McKinney. Data Structures for Statistical Computing in Python. In Stéfan van der Walt and Jarrod Millman, editors, *Proceedings of the 9th Python in Science Conference*, pages 56 – 61, 2010. doi: 10.25080/Majora-92bf1922-00a.
- M. Wold, M. S. Brotherton, and Z. Shang. The dependence of quasar variability on black hole mass. *Monthly Notices of the RAS*, 375(3):989–999, Mar. 2007. doi: 10.1111/j.1365-2966.2006.11364.x.