LOR for the LePhare PZ Estimator

Contributors

Stephane Arnouts[1], Johann Cohen-Tanugi[2], Olivier Ilbert[1], Eric Nuss[2] (DESC members), Mara Salvato[3] (AGN-SC member)

1. Laboratoire d'Astrophysique de Marseille, France, INSU
2. Laboratoire Univers et Particules de Montpellier, France, IN2P3
3. Max Planck Institute for extraterrestrial Physics, Garching

Co-signers: , Clotilde Laigle (IAP)

0. Summary Statement

We propose to use the template-fitting code LePhare for the LSST data. This public code has been extensively used for two decades, and was first described in Arnouts et al. (2002). Numerous improvements have been added over time (Ilbert et al. 2006, Ilbert et al. 2009) and the code was recently rewritten in c++ (gitlab.lam.fr/Galaxies/LEPHARE). This code is flexible : various sets of templates, attenuation curves, and additional priors can be applied, and easily added. Two different recipes have been developed to include the contribution of emission lines. To optimize the relative calibration of the different filters, a method to train the zero-points based on a subset of spectroscopic redshifts is implemented. A prior based on the N(z) could be activated (similar to Benitez 2000, with the distribution updated using the spec-z from VVDS).

We propose a standard configuration of LePhare, described in Ilbert et al. (2013) for the COSMOS survey. We successfully applied this configuration to the deep HSC-CLAUDS survey (UgrizY filters down to r~27th mag; Sawicki et al., 2019) which mimics in filterset and depth the LSST/Vera Rubin observations.

1. Scientific Utility

LePhare is a generic PZ estimator which has been used in numerous deep imaging surveys like CFHTLS (Coupon et al. 2009), COSMOS (Ilbert et al. 2009, 2013, Laigle et al. 2016) or HSC-CLAUDS (Desprez et al., in prep, Picouet et al., in prep). Numerous public PZ catalogues produced by LePhare are widely used in the astronomical community. We describe below the use of the code depending on the scientific objectives.

Galaxy evolution : this code is primarily developed for this community. LePhare produces point-like estimates of the redshift and associated uncertainties, as well as likelihood and posterior redshift probability distribution for each galaxy. More than just the photo-z, the code can also deliver the physical parameters (galaxy stellar mass, SFR, sSFR, etc). The code is based on a simple procedure of template-fitting, which allows us to deliver the PZ in the regions of the magnitude/color/redshift space not well covered in spectroscopy. This code doesn't need an extensive spectroscopic follow-up to produce robust redshifts. The precision and the low fraction of failures reached usually in deep surveys is sufficient to measure one-point statistics like mass function (Ilbert et al., 2013), or luminosity function (Moutard et al., 2020) and 2-point statistics like angular correlation functions (Coupon et al., 2015). The precision obtained with template-fitting code depends on the sensitivity of the imaging data. By limiting the sample to galaxies with a high signal-to-noise, photometric redshifts delivered by LePhare could be sufficiently accurate to reconstruct the cosmic web and perform studies linked to the environment (Laigle et al. 2018). We also tested the capability of LePhare code to estimate the redshifts and the physical parameters using the cosmological simulation Horizon-AGN and we demonstrated the good behavior of LePhare for LSST and Euclid simulated catalogues (Laigle et al. 2019).

Cosmology: The accuracy and the failure rate is sufficiently low to be used in defining the tomographic bins for the weak lensing technique. The biases on the PZ produced by template-fitting code remain too high to measure the mean redshift for weak lensing experiments. We can't directly measure the mean redshift with a bias below 0.2% using template-fitting, without additional calibration (Euclid, Ilbert et al. 2021). This code has been tested among other codes in Euclid (Euclid, Deprez et al. 2020) with such conclusions.

Stars:.For each source, LePhare can run stellar and galaxy libraries in parallel, thus serving also as a classifier. In several catalogues (e.g. Coupon et al. 2009, Laigle et al. 2016), we use this information to separate the galaxies from the stars. We are not aware of any application aiming to study stars. An output of LePhare would be the best-fit template, its normalization and the associated chi2 for the star.

AGN and QSO: The current version of Lephare is already versatile and priors and libraries can be tuned to reliably compute photo-z for AGN and QSO (e.g., Salvato et al 2009,2011, 2021, Ananna et al 2017). Recently an LSST in-kind contribution (PI: Salvato) was approved for a further development of the code to better account for the needs of the AGN SC community. Before LSST starts operations, LePhare will be able to run in parallel with independent star/galaxy/AGN settings, providing tuned photo-z for galaxies and AGN at once while also identifying stars.

2. Outputs

On the photo-z side, the code provides the redshift point estimate and associated uncertainties (68%, 90%, 99%) with two kinds of methods: using the best-fit template which minimizes the chi2 distribution, or using the median of the likelihood (or of the posterior if a prior is used). In output, the user gets the redshift posterior probability distribution function for each source (hereafter zPDF). The zPDF is obtained by summing the likelihood of each point in the library at a given redshift. If no prior is activated in LePhare, the zPDF corresponds to the likelihood.

On the physical parameter side, the code produces the point-estimates and associated 68% uncertainties. Different physical parameters can be obtained (stellar mass, SFR, sSFR, E(B-V), dust luminosity, absolute magnitudes). We usually don't advise using the same templates for physical parameters as the ones used for photo-z (large libraries of synthetized stellar populations could degrade the accuracy of the photo-z). In order to derive the point-estimate and the associated uncertainties, we derive a likelihood for each parameter. This likelihood is not a standard output of LePhare to avoid producing too massive catalogues. However, such output could be easily extracted if needed.

3. Performance

The paper from Laigle et al. (2019) forecasts the quality of the photo-z, stellar masses and SFRs that we expect to recover with the LSST data, as well as LSST+Euclid. Based on the hydro-dynamical simulation Horizon-AGN, we produce mock catalogues. We add noise to the predicted fluxes according to the expected depth of LSST after 10 years. The advantage of this simulation is that 500 000 galaxies are generated over 1 deg^2 up to high redshift. The complex star-formation and metal enrichment histories are peculiar for each simulated galaxy. We applied LePhare using the standard configuration. Table. 10 and Fig. 4 from Laigle+19 shows that we can expect a precision of sigma~0.037(1+z) for galaxies as faint as i~25 and a failure rate lower than 5%. Figure. 9 shows the expected precision on the stellar masses.

We also explore the performances of LePhare with the HSC-CLAUDS deep fields that will best match the LSST observations. With the same configuration as Ilbert et al. (2013), we get a precision of sigma~0.044(1+z) and a failure rate eta=15% for secure spectroscopic redshift sources with $24<i<25$, over the redshift range $0<z<2$ (Desprez et al., in prep).

4. Technical Aspects

Scalability - Will meet. We can process 300 sources per second using our default configuration, with a step of dz=0.01 and using 50 CPUs with a standard linux computer (using the openMP parallelized mode). However, considerable gain can be obtained by degrading the resolution of the library. For instance, by using a redshift resolution of dz=0.03 and twice less step in E(B-V), we gain a factor 10 in computational time. So, we meet the requirement of 1millisecond per source.

Inputs and Outputs - Will meet. LePhare requires the fluxes (or magnitudes) and their associated uncertainties as inputs, which will already exist in the LSST Object catalog. The output point estimates, errors, and binned PDFs are all consistent with the PZ-related Object elements defined in the DPDD.

Storage Constraints - Will probably meet. LePhare produces catalogues in ascii format. Additional tools need to be added to automatically convert these catalogues in e.g. fits format. In its current form, the output for a catalogue with one million sources, with the zPDF stored in ascii for each source from z=0 to z=10 with a step of dz=0.01 represents around 1 Gb of data (when zipped).

External Data Sets - Will meet. LePhare does not require a large amount of spectroscopic data to train the zero-points. Existing dataset will be sufficient, and we need only the value of the spec-z, not the spectra.

Estimator Training and Iterative Development - Will meet. The training of the photometric zero-points is often useful to improve the quality of the photo-z. However, this step doesn't require a large and representative training sample and the training is automatic, without the need of tuning.

Computational Processing Constraints - Will meet. LePhare does not require that a large amount of data is held in memory at any given time.

Implementation Language - Will meet. The code is written in C++11, with no extra dependency beyond OpenMP. It has been tested on recent MACOSX and Linux Ubuntu releases.

5. Références

LePhare code in C++ https://gitlab.lam.fr/Galaxies/LEPHARE

Original LePhare code in fortran http://cesam.lam.fr/lephare/lephare.html

- Arnouts et al., 2002, MNRAS, 329, 355, *Measuring the redshift evolution of clustering: the Hubble Deep Field South*
- Coupon J. et al., 2009, A&A, 500, 981, *Photometric redshifts for the CFHTLS T0004 deep and wide fields*
- Coupon, J. et al., 2015, MNRAS, 449, 1352, *The galaxy-halo connection from a joint lensing, clustering and abundance analysis in the CFHTLenS/VIPERS field*
- Euclid Collaboration, Ilbert O., 2021, A&A, 647, 117, *Euclid preparation. XI. Mean redshift determination from galaxy redshift probabilities for cosmic shear tomography*
- Euclid Collaboration and Desprez, G. et al., 2020, A&A, 644, 31*, Euclid preparation. X. The Euclid photometric-redshift challenge*
- Ilbert O. et al., 2006, A&A, 457, 841, *Accurate photometric redshifts for the CFHT Legacy Survey calibrated using the VIMOS VLT Deep Survey*
- Ilbert O. et al., 2009, ApJ, 690.1236, *Cosmos Photometric Redshifts with 30-Bands for 2-deg2*
- Ilbert O. et al., 2013, A&A, 556, 51, *Mass assembly in quiescent and star-forming galaxies since z ≃ 4 from UltraVISTA*
- Laigle C. et al. 2019, MNRAS, 486, 5104, *Horizon-AGN virtual observatory - 1. SED-fitting performance and forecasts for future imaging surveys*
- Laigle C. et al., 2018, MNRAS, 474, 5437, *COSMOS2015 photometric redshifts probe the impact of filaments on galaxy properties*
- Salvato M. et al., 2009, ApJ, 690.1250, *Photometric Redshift and Classification for the XMM-COSMOS Sources*

- *Salvato M. et al., 2011, ApJ, 742, 61, Dissecting Photometric redshift for Active Galactic Nuclei using XMM- and Chandra-COSMOS samples*
- Sawicki M., et al., 2019, MNRAS, 489, 5202, *The CFHT large area U-band deep survey (CLAUDS)*