Transactions IAU, Volume XXXA
Reports on Astronomy 2015-2018
Piero Benvenuti, ed.

DOI: 00.0000/X000000000000000X

Inter-Division D-G-H-J / Commission WG-2
The Shortest Way from Databases to Galaxy SED Fitting

IAU Link: https://www.iau.org/science/scientific_bodies/working_groups/308

CHAIRS
Samir Salim (Indiana University, USA)
Anja Schroeder (SAAO, South Africa)

MEMBERS
Sudhanshu Barway (SAAO, South Africa)
Alessandro Boselli (Laboratoire d'Astrophysique de Marseille, France)
Denis Burgarella (Laboratoire d'Astrophysique de Marseille, France)
Dongwei Fan (NAO, China)
Dmitry Makarov (Special Astrophysical Observatory, Russia)
Joseph Mazzarella (IPAC, USA)
Cristina Popescu (University of Central Lancashire)
Philippe Prugniel (University of Lyon, France)
Yannick Roehlly (Sussex University, UK)
Yue Shen (University of Illinois, USA)
Ani Thakar (Johns Hopkins University, USA)
Mattia Vaccari (University of Western Cape, South Africa)
Kevin Vinsen (ICRAR, Australia)

INITIAL REPORT

1. Introduction

The working group "The Shortest Way from Databases to Galaxy SED Fitting" (DB2SED) was formed in April 2017. The original main task of the working group was to propose ways of optimizing the use of databases containing galaxy information in order to feed the codes for the analysis of galaxy spectral energy distributions (SEDs). A large number of galaxy databases are publicly available, catering to diverse users and containing a wide range of some with huge amount of data. The number of databases data, from small to very large amounts. Their content will only grow in the future. Their impact could be further enhanced by certain practices to be developed or recommended by the WG. Possible deliverables included:

1) tools (i.e., mechanisms) that allow the user to collect homogenized data (e.g. units, aperture, ...) from a database

2) alternatively, provide recommendations on how such tools or mechanisms could be implemented by database curators

2. Identified issues

Up to now, the activities of the working group consisted in identifying:

1) the information needed to perform the SED analysis (SED fitting)
2) existing SED fitting tools/codes

3) an existing data model that can be utilized for the current goal
4) a database that may already provide the required tools

Based on the above a variety of issues have been recognized. Recommendations regarding the use of databases for galaxy SED fitting need to recognize the diversity of databases in terms of scope and the existing level of data (and sample) homogeneity and also the different forms of SED fitting.

2.1. *SED fitting*

Our focus is on the SED fitting that uses homogeneous (more on what that means later) broad-band (or medium-band) photometry to compare it with stellar population synthesis models with the goal of deriving galaxy physical parameters such as:

- stellar mass
- current star formation rate (SFR), averaged on some timescales, typically on UV/IR (ultraviolet/infrared) timescale of 100 Myr
- star formation history, e.g., SFRs at different times or timescales in the past (1 Gyr, 3 Gyr)
- star formation mode, e.g., the fraction of mass produced in bursty (non-continuous) star formation (SF) over some time period
- stellar population age (e.g., mass or luminosity weighted)
- continuum dust attenuation in various bands
- total luminosity absorbed (or emitted) by the dust
- dust mass
- stellar metallicity
- rest-frame absolute magnitudes; rest-frame colors and spectral indices (e.g., D4000)
- photometric redshifts

The basic components of SED fitting include 1) modeling of galaxy SEDs based on the combination of simple stellar populations produced by some stellar population synthesis code (e.g., Bruzual and Charlot 2003, Maraston et al. 2005, Conroy et al. 2010) according to the parameters that specify the star formation history, the IMF and the stellar metallicity; 2) attenuation of model SEDs according to some (or a range of) dust attenuation laws; 3) calculating model broad-band fluxes in filters that correspond to observations and at certain redshifts that correspond to the range of redshifts contained in the dataset; 4) comparison of observed and model SED points; 5) identification of the best fit model as well as the probability distribution functions of the physical parameters from which their nominal values and errors can be derived.

The SED fitting can involve only the stellar and nebular continuum emission (i.e., rest-frame UV to near-IR ($\lambda \lesssim 5\,\mu$m)) or may also include the dust emission in the IR. Furthermore, the SED models may include contribution of line spectra and contribution of the AGN. Consequently, the wavelength range of the SED fitting considered here extends from far-UV to sub-mm wavelengths.

Over the years many SED fitting tools, codes and packages have been made available to the public. Some of them have precomputed SEDs from which model photometry can be extracted for user specified filters and redshifts, while other allow model SEDs to be computed according to the specified parameters (e.g., regarding the star-formation history or the dust). Also some codes only provide the models, while others also perform the SED fitting and report the resulting physical parameters. Some of the best known codes currently in use are: MAGPHYS, CIGALE, FAST, Le Phare and BEAGLE.

The above, physical SED fitting may also be based on spectrophotometric data (e.g., codes such as GandALF) or a combination of photometry and spectrophotometry. However, such efforts are currently still limited, and require special efforts to homogenize the data (e.g., photometric vs. spectroscopic apertures) that go beyond what can be expected from most databases. Also, the scope of our work is to provide recommendations regarding the SED fitting to integrated light of galaxies. In principle, the SED fitting can be performed to different components of galaxies (through structural decomposition), or to individual regions or individual picture elements ("pixels"). Again, such data is expected to be specially curated and will not be part of most databases.

The focus of our efforts is not on *analytical* SED fitting, where the galaxy SED is fit by analytical functions or templates. E.g., by black body curves to determine the dust temperature or total luminosity in the IR or by power laws to determine spectral indices in the radio range. Several tools to perform analytical SED fitting exist, such as Iris SED Analysis Tool produced by USVAO, ASDC SED Builder", by ASDC (Italy) or VOSED SED Builder by SVO (Spain). These tools are usually geared towards the analysis and visual presentation of SEDs of individual galaxies (typically nearby galaxies with extensive SED coverage spanning from X-rays to radio) and communicate directly with general-purpose databases like NED at IPAC or local data repositories (e.g., ASDC). While this type of SED fitting would also benefit from being able to asses the homogeneity of the data, such aspect is less critical because the analytical SED fitting tends to be more crude. Furthermore, the challenges of homogenizing the data from such diverse spectral regions are likely insurmountable.

2.2. *Required data*

Returning to the SED fitting that uses a suite of models, one needs photometric data points (fluxes, errors and knowledge of bandpasses), distances (redshifts) and the estimate of Galactic reddening (E(B-V)). There are several aspects of homogenization and information specification that are relevant for carrying out the SED fitting.

1) Filter response. To be most useful for the purposes of SED fitting, the specification of the bandpass (filter) in the database needs to be such that the appropriate bandpass can be identified and specified in the SED fitting code.

2) Photometry aperture. SED fitting produces most reliable results with consistent photometry fluxes. What is meant by "consistent" is not necessarily clear. Typically, this is assumed to mean "using the same aperture", where aperture could be circular or elliptical. However, this will in some cases not be possible or recommended. Consider the SED fitting that involves UV fluxes from GALEX images with the PSF of 5" in combination with optical photometry with photometry having the PSF of 1". Adopting e.g., 3" would end up missing significant UV flux, whereas adopting the 10" aperture required to capture the UV flux would end up introducing additional noise and/or contamination in optical fluxes.

3) Extrapolated or total photometry. Photometry is often reported based on model extrapolations, e.g., by fitting a certain profile (e.g., a Sersic fit) and then integrating the profile to get the total flux. Whether this is preferred to aperture photometry will depend on the science question. Some galaxy catalogs use a combination of aperture and total photometry. E.g., they measure photometry in a small aperture and report aperture correction in one band. Of course, this assumes that there are no color gradients and that the seeing in all bands is the same.

4) Deblending. Simple aperture photometry may not be adequate if angular sizes of galaxies are large and the there is contaminating flux from foreground stars. There are

different ways of deblending, depending on science goals. Also, some profile-fitting photometry automatically does the deblending by fitting multiple objects simultaneously.

5) Photometry error. For the purposes of SED fitting it is important to understand if the quoted error is just the formal statistical uncertainty, or whether it also included the systematic error (e.g., from uncertainties in flux fielding). There may also be additional errors when going from relative flux measurements (i.e., the colors) to absolute fluxes.

6) Galactic reddening. It is important to understand if reported fluxes have already been corrected for Galactic redenning, and if so, using what extinction curve or which reddening coefficients.

7) Image resolution (PSF). To understand how to interpret points 1-5 it is also good to have information on the image PSF/FWHM.

8) Image depth. While not critical for the SED fitting itself, understanding the depth (e.g., the limiting flux) of photometry in a given bandpass is desirable.

9) Units. It is preferable if the units are the same across the bandpasses. Also, because of the upper limits or weak detections, it is better to have fluxes and flux errors than magnitudes.

10) Matching. Are fluxes in different bands measured based around some joint position (or position from some band) or do they come from independent measurements that are matched according to some criteria. For example, GALEX pipeline offers the measurement of FUV flux at NUV position and vice versa. While the effort was made to match FUV and NUV detections, they are often split in two entries.

2.3. *Homogenization and standardization of database outputs and SED code inputs*

While it may never be possible to use the photometry from databases "blindly", it would be good if the above aspects are specified in some, hopefully more standardized and easily accessible way. By easily accessible we mean not having to dig through original papers, documentation or "readme" files. Also, the specification should be easily interpreted by scripts.

Our eventual recommendations for DB2SED would need to take into account what type of database we are talking about, and to what extent they already contain the information that is needed to understand the data or to be able to extract "homogeneous" data. One should distinguish between meta-databases like NED and SIMBAD, which host data from a range of sources, from large surveys such as SDSS or GALEX to data from individual papers. By their very nature such databases are least homogenous simply because they contain inhomogeneous samples, and because they contain data from heterogeneous sources, including what may be considered duplicate measurements. NED, for example, already performs homogenization of units and contains comments that indicate the size of the aperture. However, if different teams have measured, for example, the NUV flux of some galaxy using different methods, or using different images (e.g., shallow vs. deep) or the same data from different data releases (e.g, GALEX GR1 vs. GR6), they will report all of them and it is up to the user to distinguish the differences. This is probably the reason why meta databases are almost never used for the SED fitting. The overarching question is what can be done to improve this, or should be instead focus on databases that are, by nature, more homogeneous.

Of such databases there are the ones that focus on multiple larger surveys. For example, IRSA hosts catalogs from some 30 projects, including WISE, IRAS, Akari, 2MASS, etc. as well as certain surveys from Spitzer and Herschel. Similarly, MAST hosts GALEX, HST, FUSE, etc. Both IRSA and MAST also host user-contributed surveys. Then there are databases dedicated to single large surveys, such as SDSS. And finally, there are project sites of various sizes, like H-ATLAS, HELP, etc. Generally, the level of homogeneity increases as one goes from general purpose databases to project databases, many of which have been specifically built for the purposes of SED fitting (e.g., GAMA performs multi-wavelength photometry in matched apertures).

3. Outstanding tasks

First, we may recommend best practices when it comes to how to perform the photometry in the first place (for future projects). Input from WG1 (RELIGAS) would be valuable for this task.

Second, possibly the main task is to propose a standard for photometry specification (i.e., photometry metadata) from the standpoint of SED fitting, so that the user can have the information needed to make a decision on what photometry to use. This may be guided to some extent by prior efforts from various Virtual Observatories, but it should not be overly intricate and comprehensive, but should focus on what is essential for broad-band SED fitting. The implementation of such standard information should be relatively easy for project-level databases, but may eventually find its way to larger databases with multiple project data. We may try to implement the standardization specification to some database that is curated by a WG2 (DB2SED) member. We should also asses to what extend do existing databases (of various scopes) readily provide the information that is needed to create such specification.

Finally, we can propose the format (units, columns specification) that the databases should use to export, *and* SED fitting software can use to import, photometry tables. Currently, different SED fitting software uses somewhat different formats for input data. This sort of standardization should be relatively simple compared to the standardization (or homogenization) of photometry itself.

The plan is to provide these recommendations in the final report for the IAU General Assembly.

Samir Salim Anja Schröder *Chairs of DB2SED*