Obserwatorium Astronomiczne Uniwersytetu Warszawskiego

Al. Ujazdowskie 4,
00-478 Warszawa, Poland
tel. +48 22-5530507

Prof. dr. hab. Łukasz Wyrzykowski
Astronomical Observatory
University of Warsaw
Al. Ujazdowskie 4
00-478 Warsaw
Poland

Warsaw, 1 July 2023

Review of the doctoral thesis of
Gabriele Riccio
National Centre for Nuclear Research
entitled:
The star formation activity of galaxies: multi-wavelength constraints based on Spectral Energy Distribution fitting
supervisor:
Prof. dr hab Katarzyna Małek

The thesis of Gabriele Roccio tackles the studies of galaxies' physical characteristics through the analysis of multiwavelength observations. In particular, the Star Formation Rate (SFR), is investigated as one of the most important parameters of galaxies, transforming interstellar gas into stars and creating metallic elements. The work uses existing observations of large samples of galaxies as well as generates simulated data as expected for the forthcoming instruments such as LSST, JWST and eROSITA. The work also includes the study of X-ray binaries as they contain imprints of the star formation history.

The work comprises 154 pages and contains five chapters, including an extended introduction followed by chapters based on articles led by the PhD candidate, published or submitted to high-profile astrophysical journals and the final chapter with a summary. The English in the text is of excellent quality and easy to read.

In the Introduction chapter, the Author describes the astrophysical background for the work. The main properties of galaxies such as Star formation rate, stellar mass and dust content are described in detail. The application of the Spectral Energy Distribution (SED) fitting method for galaxy parameter determination is also included, however, I miss a bit more lengthy (separate subsection?) description of the CIGALE code, which was used in the thesis. Fig. 1.3 misses the legend for the colours.

Chapter 2 focuses on studies of the properties of galaxies using data expected from the forthcoming LSST survey at the Vera Rubin Observatory. A simulation was conducted on approximately 50,000 galaxies to determine if LSST data will be able to accurately measure the parameters of galaxies without ambiguity. The sample of galaxies was built based on real archival data from the HELP project, containing multi-wavelength observations from *Herschel* observatory covering about 1200 sq.deg. In particular, two fields were selected, ELAIS N1 and COSMOS which were then supplemented with data from other surveys in order to cover a very broad range of wavelengths, from u-band up to 500 um. The total samples were built containing about 39000 and 15000 sources in each of the fields, respectively. The multiwavelength data were then modelled using CIGALE, a Bayesian SED fitting code available publicly. Galaxies for further studies were constrained up to redshift $z<2.5$. The sample was divided into starburst, passive and normal (i.e. main sequence) galaxies using Star Formation Rate (SFR) parameter from CIGALE. AGN and other outliers were removed, leaving a high-confidence MS sample of galaxies for both fields.

Next, the LSST observations were simulated taking care of the realistic error properties of the data. The SED fitting to LSST data alone was conducted and the resulting galaxy parameters were compared with those derived earlier, revealing a strong (factor of 6) overestimation of SFR in LSST-only data for low redshifts. Interestingly, for the $1.5<z<2.5$ redshift bin, LSST data is sufficiently good, while the comparison data (black in Fig. 2.6) is worse, most likely due to the faintness of the objects in that redshift range. The split into clearly two clusters of points in Fig.6 is seen already at $z>1$, also for LSST data - what is the reason for that?

LSST data were then supplemented with UV and Far infrared data, using GALEX and SPIRE data. This step has improved the estimation of the parameters significantly and these were used to derive an empirical formula for correcting the SFR value for LSST data used alone. Finally, the dust attenuation relation with the mass of the stellar component was derived and used to correct SFR derivation from LSST data alone.

It is not clear to me from the text, what was the main reason that only MS galaxies were selected for the LSST test. It would be interesting to see if LSST data can be used to distinguish between the types of galaxies (probably not) and what is the confusion between the types. As a consequence of the confusion, the main parameters could be derived even more incorrectly than shown in this work.

What is the cross-matching accuracy and how does it impact the study? Surveys like DECam can see faint stars located close to studied galaxies, what is the risk of using the wrong target? How can this be mitigated or assessed? Gaia was used under the assumption that Gaia sees only stars, which is not always the case, particularly in the case of compact and distant galaxies. Moreover, Gaia reaches only about G~20 mag, hence many of the faint sources were not in Gaia.

What forthcoming or planned facilities can help address the issue of missing parts of the SED on UV and mid-IR parts? ULTRASAT? What else? Dedicated targeted observations?

Fig. 2.4 would benefit from a larger legend. Also, the y-axis (magnitude error) shown in the log scale would make the figure easier to read. The simulated magnitude of LSST data goes up to 13 mag, however, the typical bright-end limit for LSST is going to be around 17 mag.

Minor typoes:

- GAIA should be written as Gaia (e.g. p.34)
- incorrect section number in 2.5.1, probably left of the paper (4.3)

Chapter 3 is based on a refereed publication, published as Ricio et al. (2022). It contains a detailed case study of the properties of globular clusters (GCs) in the extended region around a nearby Fornax galaxy cluster. GCs were identified within a 1 sq.deg around the centre of the cluster using data from a 2.6m large field of view telescope VST located at ESO's Paranal Observatory in Chile. The data overlapped with Chandra X-ray coverage of the same region. The catalogue of clusters used contained about 5000 GCs obtained after a series of cuts applied in order to minimize the contamination from objects not related to the studied cluster. I noticed that the cut on g-i and u-r colour is not explained in the text, but I guess the cut on the blue side was added to remove potential background quasars. Nevertheless, the colour histogram in Fig. 3.3 shows some excess on the blue side of the blue population - could that be related to QSOs from the background?

Fig. 3.1 is hard to read due to crowding in the very centre. It would be good to see an additional plot with a zoom-in on the central part. I'm wondering how close were individual GCs there and if they are well resolved.

The sample of GCs was divided into red and blue populations by fitting two Gaussians to the colour histogram. However, the split on two populations used a hard cut on g-i = 1 mag, meaning there was some mixture between populations hence contamination in about 30% of objects. What would change in the results if the red and blue populations were separated more conservatively by ignoring the mixing part of the g-i plot (e.g., blue: g-i<0.9 mag, red: g-i>1.1 mag)?

The X-ray data were cross-matched with the optical data in order to identify X-ray emitting sources and to study the population of LMXBs in these GCs. 168 objects were identified. These were then divided into host-galaxy GCs and infra-cluster GCs, depending on their projected distance from the nearest galaxy in the cluster. The Author notes the choice of the border is arbitrary, but the number 6 times the effective radius was chosen, which in my view is a good conservative choice, in particular given the fact that the observed distances are measured in projection.

Fig. 3.6 could also benefit from a zoom on the central part, similar to Fig.3.1. Fig.3.8 is hard to read, its labels are too small and there are many small grey points difficult to disentangle, somewhat in similar fashion as in Fig. 3.13.

The properties of GCs were discussed in Section 3.5. The clusters with LMXBs were identified among both blue and red populations of GCs, however, there were significantly more LMXB-hosting GCs in the red sample. It was found that the fraction of LMXB depends on the distance from a galaxy, which is a resolution of an ongoing debate. The X-ray luminosity function of LMXBs was also investigated for different populations of LMXBs after correcting for completeness. The slope of the LF was measured and while the host-galaxy LF slope agrees with previous studies, the intra-cluster sample seems to show a steeper LF. The shapes of LF were also studied for the blue-red cross-section of the sample and it was noted that the blue-intra-cluster sample might be too small for firm conclusions. I'm wondering about the shape of its LF in Fig.3.12 as it seems there were some issues with the completeness correction.

X-ray spectral properties of LMXBs were then studied and it was found that blue GCs have harder spectra than the red ones. This interesting observation was further investigated if it is real or if it is some observational bias or instrumental effect and it was concluded tentatively that this is indeed a real effect, possibly connected with the dependence on the metallicity of the GC.

Minor typo:

Sec. 3.5.1, page 66, in the text Fig. reference number is missing.

Chapter 4 continues the topic of the investigation of the derivation of crucial astrophysical parameters such as Star Formation Rate (SFR) or Stellar Mass (M_*) of galaxies. This time the study takes a new perspective by employing X-ray observations provided by the eROSITA space mission. This chapter is based on the publication published as Riccio et al. (2023).

In Section 4.3 the data selection is described. X-ray data from eROSITA verification observations were combined with available ultraviolet, optical and infrared observations in order to obtain as wide spectral coverage as possible. The cuts were very conservative, as only objects with at least one observation in each band were processed. I wonder how large the final sample would become if this early criterion was less strict, for example requiring only a combination of most necessary bands, e.g. g, I, H, K and WISE or u,r,z, J,K plus WISE. This very first cut removed 97% of the X-ray sample and would be worth investigating. On the other hand, the requirements for SNR>2 for WISE data are actually quite a low threshold, which could lead to very noisy data. WISE data which is the main discriminator against AGNs, and I think this very low-level cut could be partially the reason for the contamination of the sample seen later.

The next section describes the procedure of SED fitting, which uses CIGALE code, which extends its models up to the X-ray range. AGNs were removed using UV and X-ray fluxes and WISE data following a well-known procedure from Assef (2013).

Section 4.5 analyses the quality of the results of the SED fitting and the SFR estimates. The mock catalogue was generated and modelled, and the results were found to be consistent, with a slight overestimation of the SFR values at their low end. I am wondering if the Author has any ideas of the reason for this overestimation. The number of objects at the low SFR is low, could that be the reason?

The SFR was estimated again by supplementing the SED with Far Infrared data from HELP. As shown in Fig.4.5, this changed the value of SFR significantly in a non-negligible number of sources. I wonder why this data was not added at an earlier stage? The values of SFR were updated also using archival SDSS spectra. The spectral analysis on the BPT diagram also revealed new cases of AGNs, which passed through previously used criteria. I wonder if it was possible to perform the cross-match with SDSS and other archival spectra databases at an earlier stage of the work to prevent contamination.

Next, for the remaining sample of galaxies, the contributing elements to the X-ray emission were identified in order to limit the X-ray measurements to X-ray binaries only. It includes coronally active binaries, cataclysmic variables, and hot gas contribution. These components were removed using K-band luminosities, separately for normal and star-forming galaxies. Finally, the AGN contribution was removed in the sample of normal galaxies and the selection of the final sample was conducted again.

The relations between X-ray luminosity and SFR and sSFR were derived for the remaining (small) sample of galaxies. It was found it does not follow the relation seen by other authors and the completeness bias was identified as the main reason for such difference. Overall, I think the entire selection procedure is very detailed and hence complex. It would help if this procedure was summarised in a table or as a graph. I am also curious what would be the impact on the final outcome if some changes up-stream in the procedure were applied which would increase the available sample. Would the observed relation Lx-SFR shift be more pronounced or less? Would that help overcome the completeness bias? Would the use of photo-z help?

Minor problems in Chapter 4:

Sec.4.3 page 80 in sentence Considering the large PSF (~16)… - units are missing.

Sec.4.5.2. page 87, when referring to Fig.4.5, there is no bottom panel on that figure.

Sec. 4.5.2 and Sec. 4.5.3 on page 87: 1 matching radius - units are missing

Fig.4.8 caption, typo ar -> are

SUMMARY

Summarising, I think the PhD candidate has done an enormous research work. The three main research chapters each contain a wealth of studies on the topic, in many places confirming previous results from the literature and in some finding new interesting relations and dependencies. The Author demonstrated their proficiency in research through the completion of numerous complex tasks in observational data processing, astrophysical analysis, modelling, and interpretation throughout the thesis.

In my view, Mr Gabriele Riccio's work meets all the necessary requirements for a PhD thesis, and I, therefore, recommend it for further consideration.

Warsaw, 1 July 2023

prof. dr. hab. Łukasz Wyrzykowski