

First Year Progression Review

2017/2018 — Giorgio Manzoni

Abstract

In this document I report all the work done in my first year of the CDT PhD programme, and the work that I plan to do in the next upcoming years, under the supervision of Peder Norberg and Carlton Baugh. I will first give a short introduction (Section 1) meant to be an overview describing the big picture of my project. I will then explain the projects which I have already completed, in Section 2. This includes the two months team project for P&G and the workshop in Teruel (Spain) about Emission Line Galaxies (ELG). In Section 3 I will describe the projects that I plan to finish in a short time-scale. In particular the Galform simulation run and the completion of my VIPERS draft. I conclude with Section 4 describing the long-term project that I have started in the past year and that I will keep working on it for the rest of my PhD. These include the PAU survey validation (Section 4.1), PAUS observations (Section 4.2) and the test/improvement of an innovative group finder able to work in the redshift space (Section 4.3).

Contents

1	Introduction	1
2	Completed projects	2
2.1	Two months team project with Procter & Gamble (P&G)	2
2.2	Workshop on Emission-Line Galaxies (Teruel - Spain)	2
3	Short-term projects	2
3.1	GALFORM semi-analytic models	2
4	Long-term and future projects	3
4.1	PAUS Validation	3
4.2	PAUS observations	3
4.3	Group finder	3

1 Introduction

The main project of my PhD makes use of the data collected by the Probe of the Accelerating Universe Survey (PAUS). PAUS is an ongoing photometric survey with the advantage of having 40 narrow band photometric filters which make the redshift precision comparable to a spectroscopic survey. In particular the long-term aim of my project is to test an innovative group finder, based on the Markov CLustering (MCL) technique, on the PAUS data. Running this group finder on the PAUS data, require first to have a deep understanding of the survey. This has been the direction of my efforts in my first year. I have been part of the validation process, analysing the redshift accuracy and choosing the radius to be used for the forced photometry in order to better choose the parent spectroscopic survey. Additionally I have built an empirical relation for the BPT diagram to be used as a prior for the SED fitting used to retrieve photometric redshifts. In order to deeply understand the nature of the data, last March I have been observing for 7 nights as part of the PAUS observational run which took place at the William Herschel Telescope (WHT) in La Palma.

As part of the CDT programme I have been involved in a two month team project with Procter & Gamble, between May and June. The project aimed to optimise the density of the laundry powder as a function of the numerous ingredients. We achieved this optimising an already known statistical tool named LASSO. The tools we developed can be easily used for a generic function with a high number of parameters. As a CDT I had to take part to some schools relating to Machine Learning (ML) and Artificial Intelligence (AI), the main ones being in Cardiff and London UCL.

In early September I took part to a workshop in Teruel (Spain) related to Emission Line Galaxies (ELG). In this occasion, I had the opportunity to present the results of my master project, a work that made used of the VIPERS data, a spectroscopic survey with partial overlapping with PAUS. The good feedback I received pushed me to put some effort in finalising an advance draft of a VIPERS paper (that I attach to this document). What

this paper needs is a fair comparison with simulations. To do this, my second supervisor Carlton is training me to use GALFORM and this has double benefits since the skills I am acquiring will be needed to create some mock catalogues to test the MCL group finder.

2 Completed projects

2.1 Two months team project with Procter & Gamble (P&G)

As part of the CDT Data Intensive activities, I have been involved in a two months team project in collaboration with Procter & Gamble. The final aim of the project was to help the P&G team to find a more reliable analytic form of the function which describe the density of the laundry powder. The P&G team has previously identified a library of *base functions* suggested by chemistry processes. Each base function might depend on 1 up to 13 of the available ingredients ($f_i = f_i(x_1, \dots, x_{13})$). They made up the density function as a linear combination of those base functions and empirically they have chosen the combination that works better ($d = \sum_i^N a_i f_i(x_j)$). A more quantitative approach would be to perform a χ^2 minimisation but the problem is that the resulting best fit makes use of all the available base functions even in the case that the density function is physically just a linear combination of few of them. That's why we have chosen to use the Least Absolute Shrinkage and Selection Operator (LASSO) which basically add a *penalty term* to the usual χ^2 in order to penalise the sum of the parameters, i.e. favouring a fit with a small value of $\sum_{i=1}^N a_i$. This helps to discard the base functions that are not extremely relevant to the fit because a linear combination with a lot of parameter set to zero will have a smaller penalty term. However, we realised that the traditional LASSO is not perfectly calibrated to work with all the possible scenarios as it minimises the sum of the values of the parameters and not the number of them (small a_i is different from $a_i = 0$). So we developed a new *modified LASSO* to deal with that problem, however we agreed with the P&G team not to disclose the analytic form of the new penalty term. To test our new implementation of LASSO, we created increasingly difficult known functions (rising the number of variables and combining polynomials with exponentials terms) and adding various kinds of noise with different amplitudes we tried to recover the original functions. Since the penalty term depends on a free parameter λ , our case studies can be used in the future to train our algorithm to identify the best choice of λ .

2.2 Workshop on Emission-Line Galaxies (Teruel - Spain)

Last September I took part to the workshop named "Understanding Emission-line galaxies for the next generation of cosmological survey". In that occasion I presented, in a 15 minutes talk, the work done in my master project which studied the redshift evolution of the colour-magnitude relation in galaxies with the aim of getting insights on the quenching of star-formation activity. For this project I have used data from the VIMOS Public Extragalactic Redshift Survey (VIPERS), a spectroscopic survey which covers 23.5 deg^2 at intermediate redshifts¹. Specifically I studied the evolution of the bright-edge of the colour-magnitude developing an algorithm to quantify consistently this edge. Then, I have used the PEGASE 2 code to generate stellar population synthesis models to try and reproduce the evolution of the bright-edge, testing different star formation histories with and without a quenching event occurring. For this project I have an advance draft of a paper (attached to this document) which just need a comparison with simulations (see short term projects) in order to be complete.

3 Short-term projects

3.1 GALFORM semi-analytic models

As mentioned in Section 2.2, the VIPERS draft paper need some comparison with simulations in order to be complete. One possible way to accomplish this is to use GALFORM semi-analytic models. I firstly used the Gonzalez-Perez 2014 (GP14) version to check that the evolution of the colour-magnitude is similar to the one observed in VIPERS. With GP14 we can also check the behaviour of the progenitor galaxies in the colour magnitude (still under investigation). One other test to check which quenching process is potentially more important in affecting the colour-magnitude diagram is to identify the population of galaxies close to the bright-edge. Using GP14, we realised that central galaxies are dominant with respect to satellites. That means that the quenching in satellites (usually ram pressure stripping) would have just a minor effect in the location of the colour-magnitude. Instead AGN feedback (main quenching process for centrals in GALFORM) would play a crucial role in the evolution of the bright-edge in the colour-magnitude plane. For this reason, the only thing needed to complete this project is to run again GALFORM, switching on and off the AGN feedback and see which scenario better represents VIPERS observations.

¹The second (and complete) Public Data Release (PDR-2) is now available at <http://vipers.inaf.it/>.

4 Long-term and future projects

4.1 PAUS Validation

Excluding lectures and CDT events, I invested the most of my first year PhD getting involved in the PAUS collaboration, in particular in the validation of the survey. PAUS is a photometric redshift survey with the main feature being the 40 contiguous narrow band filters. Since it is an ongoing survey, one of the task was to make sure that observations have been carried out with all of the filters. Because of the survey strategy and the format of the CCD, we don't observe each galaxy with all the filters in a row. Instead it is more convenient to move the CCD in a nearly spiral way (observing in dithers rather than single exposures) letting galaxies fall in the 8 different filters that are part of each of the 5 trays. That is why it becomes important to count the number of times that a galaxy have been observed in one particular filter. The number of filters is just one of many potential biases that might arise in an ongoing survey and my task is to check that instead there are no bias both at the observation stage and at the processing stage (nightly, memba and photo-z pipelines). I did all of this tests in the COSMOS area observed by PAUS, with the possibility of comparing PAUS photometric redshifts with COSMOS spectroscopic redshifts. Now the survey has collected enough data in the W3 field so that in the close future I am ready to do all the tests in this area as well. The others two fields that PAUS is planning to cover are W1 and W4. For those fields the advantage is the overlapping with the VIPERS survey (which has been the main focus of my master project) which offers spectroscopic redshift to be used for comparison with the photometric ones.

Before processing the W3 data, an important choice about the parent survey had to be done. Since we are dealing with forced photometry, what drove this choice was the half-light radius. In particular we needed to find a survey which overlaps with W3 and that has a similar definition of half-light radius as the one in the COSMOS survey (because we want consistency among all the PAUS fields). So, in this scenario I made both statistical and (where possible) object-by-object comparison among the COSMOS, CFHTLS and CFHTLenS surveys. I finally worked out that the CFHTLenS survey has a radius consistent with the COSMOS data and so suitable for the PAUS/W3 field.

One further analysis was aimed to improve the photo-z capability prediction through SED fitting. Specifically the current SED fitting procedure used by the PAUS team is based on a set of fixed emission-line ratio assumptions. A more physical scenario would allow the emission line ratios to vary following specific BPT diagrams. For this reason I have investigated in the literature various BPT diagrams for surveys at similar redshifts with the aim of building an empirical relations. The PAUS collaboration has approved the new scheme of emission line relations and they are going to use it in the next run of the photo-z pipeline. Of course, once they have done it, my next step is to check how much this variation contributes to the estimate of photometric redshifts, comparing the different productions.

In a very long term view, the experience gained in the PAU survey validation might be useful for the upcoming Multi-Object Optical and Near-infrared Spectrograph (MOONS) survey.

4.2 PAUS observations

In order to understand the nature of the data and to build experience within the survey, last March I took part to the PAUS observational run. The PAU survey is using the William Herschel Telescope (WHT), a 4.2 m diameter telescope based in La Palma (Spain) combined with a large field camera (PAUCam) built appositely for the survey. I have been observing for 7 nights sided by expert observers who taught me how to practically set up the telescope in the afternoon and how to carry out observations during the night. The set up consisted of taking flat fields and bias images while the observation were aimed in covering part of the W3 and W1 fields accordingly to the spiral observing strategy. Galaxies observed with a seeing worse than 2.0 and transparency worse than 0.5 are flagged as "bad exposure" and the following night they are selected to be observed again. This observational experience was meant to be a training for the next PAUS observations. In fact, I have been assigned 4 nights for next December observational run.

4.3 Group finder

As mentioned in the Introduction, the main purpose of my PhD is to test and (if needed) to improve a group finder algorithm based on the Markov CLustering (MCL) technique. This technique seems to be more reliable (in terms of completeness, purity and variation of information, VI) than a simple Friend of Friend (FoF) technique which preferentially consider large structures as a single group even if they are barely connected. Until now the MCL has been tested only on mock catalogues and it looks promising for a real survey like PAUS. What makes different a real survey from a simulation is the fact that we are limited to work in the redshift space instead of the simulated real space and this has an impact in the MCL code. Running successfully this group finder on the PAUS data might provide some new insights relative to the clustering of galaxies.