PROGRAMME AND ABSTRACTS BOOKLET

V1.0 (30oct18)

XXX Canary Islands Winter School of Astrophysics

“BIG DATA ANALYSIS IN ASTRONOMY”

Instituto de Astrofísica de Canarias
La Laguna, Tenerife, Spain

4-10 November 2018

OUTLINE OF THE SCHOOL

The primary aim of this XXX Winter School is to educate the next generation of astronomers in analysis techniques that can be used to digest the gigantic amounts of data that will be produced by the next generation of telescopes and surveys. It will include lectures but also tutorials on advanced statistics and machine learning techniques. We intent to review the current state of the topic and educate the students by providing an overview of the main groups of machine learning techniques as well as the most used software. By the end of the school, the students will come back to their home institutions armed with a set of skills and tools that will allow them to make in-depth use of particular set of algorithms that fit best to their research.

Given its broad scope, the School is addressed to students in any research field in astronomy, and using different kinds of data, observational and theoretical.

Renowned specialists in the field will lecture on the following topics:

- General overview on the use of machine learning techniques in astronomy: past, present and perspectives. (Prof. S. George Djorgovski)
- Data challenges and solutions in forthcoming surveys. (Prof. Mario Juric)
- Machine learning methods for non-supervised classification and dimension reduction techniques. (Dalya Baron.)
- Supervised learning: classification and regression. (Prof. Michael Biehl)
- Deep learning. (Dr. Marc Huertas-Company)

Lectures will be complemented by tutorials consisting in implementing small projects with a design and scope based on what the students will learn during the machine learning lectures. **Participants of the Winter School will have the opportunity to display their current work by presenting a poster.**

The Winter School will take place in San Cristóbal de La Laguna (Tenerife, Canary Islands, Spain) **from Sunday 4th to Saturday 10th November 2018**.The lectures, which will take place from Monday to Friday, will be delivered in English. Speakers will present their topics in a series of four lectures. Additionally, there will be two tutorial sessions. Links to the material used by the lecturers as well as videos of the lectures will be made available on the website of the XXX Winter School. Visits to the IAC's Headquarter in La Laguna, the Teide Observatory in Tenerife and the Roque de los Muchachos Observatory in La Palma will be scheduled as part of the activities.

ORGANIZING COMMITTEE

Ana Monreal Ibero
E-mail: amonreal -at- iac.es — Tel: + 34-922-605200 ext. 5742

Jorge Sánchez Almeida
E-mail: jos -at- iac.es — Tel: + 34-922-605200 ext. 5233

Johan Knapen
E-mail: jhk -at- iac.es — Tel: + 34-922-605200 ext. 5420

Ignacio González Martínez-País
Head IAC Graduate Studies Division
E-mail: corens -at- iac.es — Tel: + 34-922-605200 ext. 5419

Lourdes González Pérez
IAC, Secretariat of the Graduate Studies Division & Winter School
E-mail: secens -at- iac.es — Tel: + 34-922-605200 ext. 5271

Rafael Rebolo López
IAC Director E-mail: director -at- iac.es

Press Room: uc3 -at- iac.es

LECTURERS AND SCIENTIFIC PROGRAMME

General overview on the use of machine learning techniques in astronomy

Prof. S. George Djorgovski
Caltech, Division of Physics, Mathematics and Astronomy

Data challenges and solutions in forthcoming surveys

Prof. Mario Juric
University of Washington

Machine learning methods for non-supervised classification and dimension reduction techniques

Dalya Baron
School of Physics and Astronomy, Tel-Aviv University

Supervised learning: classification and regression

Prof. Michael Biehl
Johann Bernoulli Institute for Mathematics and Computer Science, University of Groningen

Deep learning

Dr. Marc Huertas-Company
Université Paris-Diderot - Observatoire de Paris

TIMETABLE

* Last update: October 24, 2018

Time	Mon 5	Tue 6	Wed 7	Thu 8	Fri 9	Sat 10
8:45-9:00	Welcome			----		Visit ORM (back not earlier than 20:30)
9:00-10:00	Djorgovski	Djorgovski	Djorgovski	Djorgovski	---	
10:00-11:00	Juric	Juric	Juric	Juric	Huertas-Company	
11:00-11:30	Coffee break					
11:30-12:30	Biehl	Baron	Baron	Tutorial Huertas-Company	Tutorial Huertas-Company	
12:30-13:30	Biehl	Baron	Lunch	Biehl	Biehl	
13:30-14:00	Lunch			Lunch		
14:00-15:00			Visit OT			
15:00-16:00	Baron	Huertas-Company		Tutorial Baron	Tutorial Biehl	
16:00-16:30	Coffee break			Coffe break		
16:30-17:30	Huertas-Company	Huertas-Company		Tutorial Baron	Tutorial Biehl	
17:30-18:30	Poster session (01-18)	Poster session (19-36)				
18:30-20:00	---	---		Visit IAC	Public Talk Sánchez-Almeida	
20:30-22:30				School Dinner	---	

LIST OF PARTICIPANTS

1. **Alonso, Arias;** IAC, Spain
2. **Armstrong, John;** University of Glasgow, United Kingdom
3. **Battaglia, Giuseppina;** IAC, Spain
4. **Becker, Ignacio;** Pontificia Universidad Católica de Chile, Chile
5. **Blancato, Kirsten;** Columbia University, United States of America
6. **Bricman, Katja;** University of Nova Gorica, Slovenia
7. **Calderon, Victor;** Vanderbilt University,United States of America
8. **Camilo, Jiménez;** IAC, Spain
9. **Canducci, Marc;** University of Birmingham, United Kingdom
10. **Carnerero Martin, Maria Isabel;** INAF-Osservatorio Astrofisico di Torino, Italy
11. **Casasayas Barris, Núria;** Instituto de Astrofísicas de Canarias, Spain
12. **Chamba, Nushkia;** IAC, Spain
13. **Chen, Fanyao**; Lund Observatory, Sweden
14. **Cicuendez, Luis;** Instituto de Astrofisica de Canarias, Spain
15. **da Silva Santos, João Manuel;** Stockholm University, Sweden
16. **Díaz-García, Simón;** Instituto de Astrofísica de Canarias, Spain
17. **de la Vieuville, Geoffroy;** IRAP, France
18. **de Vicente, Angel;** Instituto de Astrofísica de Canarias, Spain
19. **del Moral-Castro, Ignacio;** Instituto de Astrofísica de Canarias, Spain
20. **Esteban Gutiérrez, Ana;** Instituto de Astrofísica de Canarias (IAC), Spain
21. **Favole, Ginevra;** ESA, Spain
22. **Fernandez, Joseph;** Liverpool John Moores University, United Kingdom
23. **Fernando, Buitrago;** Instituto de Astrofísica e Ciências do Espaço, Portugal
24. **Giri, Sambit Kumar;** Department of Astronomy, Stockholm University, Sweden
25. **González, Lucía;** IAC - Instituto de Astrofísica de Canarias, Spain
26. **Gran, Felipe;** PUC/MAS, Chile
27. **Haigh, Caroline;** University of Groningen, Netherlands
28. **Hu, Lei;** Purple Mountain Observatory, CAS, China
29. **Ider Chitham, Jacob;** Max Planck Institute for Extraterrestrial Physics, Germany
30. **Infante, Raul;** IAC, Spain
31. **Izquierdo, Paula;** Instituto de Astrofísica de Canarias, Spain
32. **Jaume Bestard, Jaume**; IAC, Spain
33. **Kelly, Brandon;** Liverpool John Moores University, United Kingdom
34. **Kim, Youngjoo;** University of Groningen, Netherlands
35. **Klemen, Čotar;** Faculty of Mathematics and Physics, University of Ljubljana, Slovenia
36. **Knapen, Johan;** IAC Tenerife,Spain
37. **Kruuse, Maarja;** Tartu Observatory, University of Tartu, Estonia
38. **LAVIE, Baptiste;** University of Geneva, Switzerland
39. **Lorenzo, Zanisi;** University of southampton, United Kingdom
40. **Luna, Manuel;** Instituto de Astrofísica de Canarias, Spain
41. **Luque, Rafael;** IAC, Spain
42. **Makrygianni, Lydia;** The University of Sheffield, United Kingdom
43. **Malyali, Adam;** Max Planck Institute for Extraterrestial Physics, Germany
44. **Martínez-Lombilla, Cristina;** Instituto de Astrofísica de Canarias (IAC), Spain
45. **Martínez-Vázquez, Clara;** Cerro Tololo Inter-American Observatory (CTIO), Chile
46. **Mastropietro, Michele;** Ghent University, Belgium
47. **Monreal-Ibero, Ana;** Instituto de Astrofísica de Canarias, Spain
48. **Montes solís, María;** IAC, Spain

49. **Morawski, Filip;** Nicolaus Copernicus Astronomical Center of the Polish Academy of Sciences, Poland
50. **Muller, Tomas;** University of Southampton, United Kingdom
51. **Nguyen, Xuan Thanh;** ESIEE Paris, Paris-Est, France
52. **Nolte, Aleke;** University of Groningen, Netherlands
53. **Pérez Prieto, Jorge A;** Instituto de Astrofísica de Canarias, Spain
54. **Pérez-Mesa, Víctor;** Instituto de Astrofísica de Canarias, Spain
55. **Panos, Brandon;** The University of Geneva and Fachhochschule Nordwestschweiz, Switzerland
56. **Pantaleoni, Michelangelo;** Centro de Astrobiología, Spain
57. **Pennock, Clara Marie;** Keele University, United Kingdom
58. **Peralta, Raphaël;** Institut d'Astrophysique Spatial, France
59. **Poidevin, Frédérick;** IAC, Spain
60. **Prieto Arranz, Jorge;** IAC, Spain
61. **Putko, Joseph;** IAC, Spain
62. **Raj, Maria Angela;** INAF-Observatory of Naples, Italy
63. **Rezaei, Samira;** PhD at RUG, Netherlands
64. **Rizos, Juan Luis;** IAC, Spain
65. **Román, Javier;** PhD student at IAC, Spain
66. **Rosado Belza, David;** IAC Tenerife, Spain
67. **Sökmen, Efşan;** PhD student, Turkey
68. **Sanchez, Paula;** Universidad de Chile / ESO, Chile
69. **Sanchez Almeida, Jorge;** Instituto de Astrofísica de Canarias, Spain
70. **Scott, Bryan;** University of California, Riverside, United States of America
71. **Su, Hung-Shuo (Alan);** University of Oulu, Finland
72. **Taibi, Salvatore;** Instituto de Astrofisica de Canarias, Spain
73. **Tarrío Alonso, Paula;** CEA-Saclay, France
74. **Unruh, Sandra;** Argelander-Institut fuer Astronomie, Germany
75. **Vazquez Acosta, Monica;** IAC, Spain
76. **Vioque, Miguel;** University of Leeds, United Kingdom
77. **Walo Martín, Daniel;** Instituto de Astrofísica de Canarias, Spain
78. **Weaver, Zachary;** Boston University, United States of America
79. **Wells, Mark;** PSU & Villanova, United States of America
80. **Whitten, Devin;** University of Notre Dame - Department of Physics, United States of America
81. **Williamson, Marc;** New York University, United States of America
82. **Woldeyes, Dejene Zewdie;** Ethiopian Space Science and Technology Institute, Ethiopia
83. **Woldeyes, Betelehem Bilata;** Graduate assistant, Ethiopia
84. **You, Jiwoo;** University of Groningen, Netherlands

POSTER CONTRIBUTIONS

1- X-ray transient classification for eROSITA
Presenting author: Adam Malyali

eROSITA on-board the SRG satellite will perform the next large X-ray all-sky survey. With its 30-fold increased sensitivity relative to its predecessor ROSAT and its multi-visit, multi-cadence survey strategy, eROSITA will provide a new and deeper look into X-ray time domain astrophysics. This holds the potential for the discovery of new populations of exotic, interesting and unexpected transient sources. However, these must first be detected amongst the millions of AGN, galaxy clusters and stars that will dominate the detected objects. Based on end-to-end simulations, we present the expected detection rate of white dwarf - black hole tidal disruption events with eROSITA. Furthermore, we discuss the challenges of developing machine learning algorithms for classification of X-ray transients for eROSITA, where the only training datasets available are non-representative and biased; as well as the results of an unsupervised approach to transient classification and outlier detection.

2- Prototype-based Analysis of GAMA Galaxy Catalogue Data.
Presenting author: Aleke Nolte

We present a prototype-based machine learning analysis of labeled galaxy catalogue data containing parameters from the Galaxy and Mass Assembly (GAMA) survey. Using both an unsupervised and supervised method, the Self-Organizing Map and Generalized Relevance Matrix Learning Vector Quantization, we find that the data does not fully support the popular visual-inspection-based galaxy classification scheme employed to categorize the galaxies. In particular, only one class, the Little Blue Spheroids, is consistently separable from the other classes. In a proof-of-concept experiment, we present the galaxy parameters that are most discriminative for this class.

3- Low Surface Brightness Segmentation Methods
Presenting author: Brandon Kelly

In this study the efficacy of three pieces of segmentation software will be compared to one another with respect to their low surface brightness (LSB) detection capabilities. Source Extractor NoiseChisel and ProFound will be judged based on object definition, noise separation, flexibility, ease of use and efficiency. Each tool created a segmentation map from an SDSS Stripe 82 image of the Abell 267 cluster in the r-deep band, and the identified objects were subtracted. A preliminary analysis suggests that NoiseChisel may be the most effective tool for LSB segmentation due to its additional segmentation of diffuse LSB components from sky.

4- Do all flares share similar chromospheric physics?
Presenting author: Brandon Panos

IRIS performs solar observations over a large range of atmospheric heights, including the chromosphere where the majority of flare energy is dissipated. The strong Mg II h&k spectral lines are capable of providing excellent atmospheric diagnostics, but have not been fully utilized for flaring atmospheres. We aim to investigate whether the physics of the chromosphere is identical for all flare observations by analyzing if there are certain spectra that occur in all flares. To achieve this, we automatically analyze hundreds of

thousands of Mg II h&k line profiles from a set of 33 flares, and use a machine learning technique which we call supervised hierarchical k-means, to cluster all profile shapes. We identify a single peaked Mg II profile, in contrast to the double-peaked quiet Sun profiles, appearing in every flare. Additionally, we find extremely broad profiles with characteristic blue shifted central reversals appearing at the front of fast-moving flare ribbons. These profiles occur during the impulsive phase of the flare, and we present results of their temporal and spatial correlation with non-thermal hard X-ray signatures, suggesting that flare-accelerated electrons play an important role in the formation of these profiles. The ratio of the integrated Mg II h&k lines can also serve as an opacity diagnostic, and we find higher opacities during each flare maximum. Our study shows that machine learning is a powerful tool for large scale statistical solar analyses.

5- RR Lyrae stars: hints about the early chemical evolution of the Universe.
Presenting author: C. E. Martínez-Vázquez

RR Lyrae stars are the best, unambiguous tracers of old (> 10 Gyr) stellar populations. They were in place at early times when most of the halo mass assembly occurred. Therefore, they provide a direct observable to trace the early star formation of the host galaxy, and constraints into galaxy formation and evolution models under the Cold Dark Matter scenario. This poster presents a series of results focused on the properties of RR Lyrae stars in Local Group galaxies, as tracers of the early evolution of the Milky Way and the M31 systems. On one side, the MW and M31 halo contain a population of relatively metal-rich RRL stars that is not observed in dwarf galaxies, with the exception of the most massive ones. This occurrence sets strong constraints on the building blocks of both halos, which were most likely formed by accretion of massive dwarf galaxies. On the other side, the pulsation properties of the RR Lyrae stars can be used to obtain individual metallicities, and derive the metallicity distribution of the purely old population, thus the early chemical evolution of the host galaxy. Using the Sculptor dwarf spheroidal galaxy as a test case, we compared the metallicity distribution obtained with RR Lyrae and RGB stars, and we found that i) the star formation in the center of this galaxy lasted substantially longer than in the outer parts, thus constraining the timescales for the outside-in evolution of this galaxy; ii) the RR Lyrae population has an intrinsic metallicity spread and presents a clear spatial gradient, which therefore was in place at a very early epoch. Finally, a similar analysis is applied to several Local Group galaxies, to show the potential of this approach.

6- Evaluating source extraction tools
Presenting author: Caroline Haigh

With the growth of the scale of imaging surveys and astronomical data, there is an increased need for automated detection and extraction of astronomical sources from images, with a high degree of accuracy. We present a comparison of several tools which have been developed to perform this task, including SourceExtractor, ProFound, NoiseChisel, and MTObjects.

In particular, we focus on evaluating performance in situations which present challenges for detection - for example, highly noisy images; faint and diffuse galaxies; extended structures, such as streams; and objects close to bright sources.

The tools will be evaluated on simulated data in order to establish a performance baseline in terms of the limits of detection. They will additionally be compared using labelled images

from astronomical surveys, in order to assess how well they perform on real-world data, as well as how well they generalise to different datasets.

7- Mining for exotica among galaxies behind the Magellanic Clouds
Presenting author: Clara M. Pennock

The nearby Magellanic Cloud system (MCs) covers tens of square degrees on the sky. Much of it has been mapped across the electromagnetic spectrum at high angular resolution and sensitivity - X-rays (XMM), optical (SMASH), IR (VISTA, WISE, Spitzer, Herschel), radio (ATCA, ASKAP). This provides a superb dataset to explore the galaxy populations in the background of the MCs. Our project focusses not on establishing sequences and distributions of common types of galaxies and active galactic nuclei (AGN), but seeks to identify extreme examples, building on the recent accidental discovery of a unique AGN. We will apply artificial intelligence to mining this vast multi-dimensional space (x,y,z,t,lambda) through a combination of critical source extraction, association and characterisation with machine learning search algorithms, within a robust statistical framework. Thus we also build readiness to fully exploit future generations of all-sky surveys (LSST, Euclid) when the Big Data challenge is destined to hit home hard.

8- PROPERTIES OF LOCAL GALAXIES WITH THE HIGHEST GROWTH RATE
Presenting author: Dejene Zewdie

In this work, we went a step further in studying the local galaxies with the highest growth rate, with the main aim to understand better their properties and contribute to better understanding of galaxy growth and morphological transformation. We report an analysis of 179 636 galaxies from the SDSS DR8 survey, using both photometric and spectroscopic data. Taking into account the results of Peréz et al. (2013), obtained using the CALIFA data, we selected those galaxies that have stellar mass in the range 10.73 - 11.03 solar masses being suggested to be good candidates of galaxies with the highest growth rate. We studied the distributions of selected sources in terms of their stellar mass, SFR, specific SFR and luminosity; we classified all galaxies into star-forming, composites, and AGN (Sy2 and LINERs); and studied morphology using two different classification catalogues. We analysed the location of all selected galaxies in relation to the main-sequence of star formation and on the colour-stellar mass diagram. We found that galaxies with the highest growth rate have different spectroscopic properties, most of them being either AGN (LINERs + Sy2) or composite, but not star-forming as have been known for the entire SDSS population at $z<0.1$.

9- Chronographic and Structural Analysis of the Milky Way Stellar Halo
Presenting author: Devin D. Whitten

Samples of Blue Horizontal-Branch (BHB) stars selected from Pan-STARRS DR1, GALEX GUVCat, and The Dark Energy Survey, are used to model the density and age profile of the Milky Way Stellar Halo. The selection of function of BHB stars from foreground contaminants is made possible by new applications of classification-based artificial neural networks in color-color space, while density models of varying complexity are explored with a Bayesian Markov Chain Monte Carlo methodology to identify a break radius in the Milky Way Halo. The implications of this break radius on the origin of the two-component Stellar Halo are discussed, as well as the impact on cosmological models of Milky Way formation.

10- Unsupervised Machine Learning on the Selected Near Infrared Color Magnitude Diagrams of the Low Latitude Southern Galactic Plane
Presenting author: Efşan Sökmen

Most of the stars in our galaxy is contained within the disk, of which the maximum size and the detailed structure is still unknown. Using the near infrared bands sheds light on the complex stellar distribution of the galactic thin disk since it is less affected by the scatter and obscureness due to the dust. Using The ESO near-infrared public survey VISTA variables in the Vía Lactea (VVV) from which we had obtained the deep photometric data for the disk in J and Ks filters reaching 19 mag in Ks, we apply the principal component analysis (PCA) to the color magnitude diagrams (CMDs) and investigate the significance of these results by comparing them to those that are applied on a set of synthetic populations with extinction. We found that the principal components correlate with extinction on the small scale, while on the large scale it can be used to find anomalies.

11- Unleashing the inner Galaxy: the proper motion revolution
Presenting author: Felipe Gran

The Galactic bulge is usually avoided by wide-field surveys due to the stellar crowding and the dust and gas in the line of sight that covers the heart of the Milky Way. The dust and gas also prevent us in the past to observe in more detail this feature of the Galaxy. Within this framework, and taking advantage of the near-IR multi-epoch observations of the VVV survey towards the Galactic bulge, we aim to characterize in full detail the innermost part of the Galaxy (|b| < 3 deg, |l| < 10 deg, in Galactic coordinates) in terms of its stellar populations, globular clusters, variable stars, and most importantly the individual proper motion of the stars. Because of the optical nature of Gaia, most of the sources on the Galactic plane are missing, being VVV the perfect complement to study the Galaxy as a whole.

Here we show the first glimpse of the results of this study, in which we are able to recover the entire sample of known globular clusters within the analyzed area and also find new ones due to the derived proper motion information. We also explore the variable stars present in the area, specifically, RR Lyrae stars which are perfect distance indicators, to give clues about its cluster membership.

12- Application of machine learning in gravitational waves survey
Presenting author: Filip Morawski

The Galactic bulge is usually avoided by wide-field surveys due to the stellar crowding and the dust and gas in the line of sight that covers the heart of the Milky Way. The dust and gas also prevent us in the past to observe in more detail this feature of the Galaxy. Within this framework, and taking advantage of the near-IR multi-epoch observations of the VVV survey towards the Galactic bulge, we aim to characterize in full detail the innermost part of the Galaxy (|b| < 3 deg, |l| < 10 deg, in Galactic coordinates) in terms of its stellar populations, globular clusters, variable stars, and most importantly the individual proper motion of the stars. Because of the optical nature of Gaia, most of the sources on the Galactic plane are missing, being VVV the perfect complement to study the Galaxy as a whole.

Here we show the first glimpse of the results of this study, in which we are able to recover the entire sample of known globular clusters within the analyzed area and also find new ones due to the derived proper motion information. We also explore the variable stars

present in the area, specifically, RR Lyrae stars which are perfect distance indicators, to give clues about its cluster membership.

13- Classification of Sodium MRI Data of Cartilage Using Machine Learning
Presenting author: Frédérick Poidevin

Purpose: To assess the possible utility of machine learning for classifying subjects with and subjects without osteoarthritis using sodium magnetic resonance imaging data. Theory: support vector machine, k-nearest neighbors, naive Bayes, discriminant analysis, linear regression, logistic regression, neural networks, decision tree, and tree bagging were tested. Methods: Sodium magnetic resonance imaging with and without fluid suppression by inversion recovery was acquired on the knee cartilage of 19 controls and 28 osteoarthritis patients. Sodium concentrations were measured in regions of interests in the knee for both acquisitions. Mean (MEAN) and standard deviation (STD) of these concentrations were measured in each regions of interest, and the minimum, maximum, and mean of these two measurements were calculated over all regions of interests for each subject. The resulting 12 variables per subject were used as predictors for classification. Results: Either Min [STD] alone, or in combination with Mean [MEAN] or Min [MEAN], all from fluid suppressed data, were the best predictors with an accuracy $>74\% $, mainly with linear logistic regression and linear support vector machine. Other good classifiers include discriminant analysis, linear regression, and naive Bayes. Conclusion: Machine learning is a promising technique for classifying osteoarthritis patients and controls from sodium magnetic resonance imaging data.

14- Probing the faint end of the LAE Luminosity Function in lensing clusters with MUSE
Presenting author: G. de La Vieuville

In our current understanding of the reionization era, the sources responsible for the transition of the universe from a neutral state to an ionized state are likely faint, low mass and star-forming galaxies. One way to study this galaxy population is to study the Luminosity Function (LF) of Lyman-alpha emitters. However the current studies are limited by the depth of the surveys, and the faint end of the LF remains poorly constrained, leaving a large uncertainty on the contribution of the LAE population to reionization.
In an attempt to reach more definitive conclusions, we are working in lensing clusters with MUSE, a large field of view Integral Field Unit (IFU) at the VLT. This instrument is ideal to work on the LF, as it allows a complete LAE selection in a wide redshift range ($2.9 < z < 6.9$), without any photometric prior. To the cost of a significant increase in complexity and a lower volume of universe explored, we are able to select a large population of LAE ($~160$) that are typically 10 - 100 fainter than in blank field surveys.
For this work, we developed new methods to precisely measure volumes in spectroscopic cubes behind lensing cluster and to do a completeness estimation of the sample using the real source profile. This poster will present the methods we used, the results obtained on the study of the LAE LF and a discussion on the possible implications for reionization.

15- Modeling the galaxy clustering properties of WISP/3D-HST Halpha emitters in preparation to Euclid
Presenting author: Ginevra Favole

Among star-forming galaxies there is a particular population whose optical spectra exhibit strong nebular emission lines. These galaxies will be the preferred targets of new-generation spectroscopic surveys as Euclid, the Dark Energy Spectroscopic Instrument,

the 4-metre Multi-Object Spectroscopic Telescope, and the Subaru Prime Focus Spectrograph. All these surveys will observe [OII] and Hα emission-line galaxies up to redshift z~2 to trace star formation and to measure the baryon acoustic oscillations as standard ruler for distances, in the attempt to unveil the nature of dark energy and probe the large scale structure of the Universe. Therefore, it is crucial to understand how to measure and precisely model the clustering properties of such galaxies and how they form, evolve and distribute within their dark matter halos. We address these issues using state-of-the-art data from the WISP and 3D-HST infrared surveys and large-volume high-resolution cosmological simulations to prepare the clustering prospects for the new generation of optical and near-IR experiments.

16- Deep cadence-invariant classification of variable stars.
Presenting author: Ignacio Becker

During the last decade, considerable effort has been made to perform automatic classification of variable stars using machine learning techniques. Traditionally, light curves are represented as a vector of statistical descriptors or features used as input for many algorithms. Features are computationally expensive, cannot be updated quickly and for large datasets can take up to days. Its computation cannot scale for large-scale datasets, which is expected to obtain from the LSST. Previous work has been done to develop unsupervised feature extraction algorithms for light curves. In this work, we propose an algorithm to automatically learn a representation of light curves that is
invariant to the cadence of the observations. We propose a series of architectures based on Recurrent Neural Networks and test them in automatic classification scenarios. Our method uses minimal preprocessing, can be updated with a low computational cost for new data and can scale for more massive datasets. We test our method in two surveys: OGLE-III and the cross-match with VVV. We obtain accuracies +75% in the majority of subclasses and +95% in classes. We compare our results with the Random Forest algorithm in variable stars classification.

17- Searching in IFS Surveys almost identical-twin galaxies differing in nuclear activity
Presenting author: Ignacio del Moral-Castro

All massive galaxies seems to harbour a supermassive black hole at the centre, suggesting that all galaxies experience nuclear activity for at least some part of their evolution. Therefore, unveil the mechanisms that trigger to active galactic nuclei (AGN) is crucial for understanding the formation and evolution of galaxies.

Legacy integral field surveys (e.g. CALIFA or MANGA) provide the opportunity of selecting large scale almost-identical pairs of isolated galaxies differing only in nuclear activity. This poster will present the search methodologies followed to identify, among this large number of galaxies, these pairs matched in galactic properties such as redshift, inclination, mass and magnitude.

18- Automated Galaxy Cluster Identification for eROSITA in the Era of Big Data
Presenting author: Jacob Ider Chitham

The presence of galaxy clusters can be inferred through observations of extended X-ray emission originating from hot intracluster gas trapped in their gravitational potential well. Approximately $100,000$ galaxy clusters are forecasted to be detected with eROSITA (extended ROentgen Survey with an Imaging Telescope Array; Merloni et al. 2012). In

order to confirm such clusters and derive accurate information about other observable properties, X-ray detections are often complemented with optical follow up methods that provide a powerful way to study the dynamics of their member galaxies via spectroscopy. In order to improve the completeness of cluster membership assignment at high redshift, the Pan-STARRS component of CODEX (COnstraining Dark Energy with X-rays; Finoguenov et al. in prep) aims to extend the capabilities of cluster identification over three quarters of the sky. Our automated Galaxy cluster identification pipeline for Pan-STARRS is based on the Multi-component Matched Filter Cluster Confirmation Tool for eROSITA developed by Klein et al (2017). The algorithm considers the spatial clustering of galaxies relative to the optical cluster centre and combines photometric information from a range of colours to determine the discrepancy or "colour-distance" for each potential member galaxy with respect to models of the expected red-sequence galaxy population. This is then combined with supervised machine learning techniques based in order to optimise spectroscopic target selection.

19- Ultra-deep imaging killers: scatter light and galactic cirri
Presenting author: Javier Román

Two are the main systematics effects preventing the full exploitation of ultra-deep imaging: the scatter light produced by the point spread function (PSF) and the curtains of dust also known as galactic cirri. Without addressing these phenomena, our efforts for getting ultra-deep imaging of the sky result fruitless. Here we present how to produce star-free imaging of the sky by careful characterization of the PSF and the scatter light associated. We also show a detailed analysis of the photometric properties of the galactic cirri. Having such characterization done, we describe how to discern between the galactic cirri and the background galaxies. We will show how these new techniques are fundamental if we want to fully exploit the science contained in ultra-deep data imaging.

20- Scalable max-tree and alpha-tree algorithm for high resolution, multispectral, and extreme dynamic range images
Presenting author: Jiwoo You

Max-tree and alpha-tree are hierarchical representations of image which are highly efficient in the detection of objects and segmentation in many application fields. However, the advance of image acquisition techniques has lead to a larger database of images with higher resolution, dimension and dynamic range, which makes the construction and the interpretation of max-tree and alpha-tree of those images less feasible. Here, we introduce a scalable max-tree and alpha-tree algorithm, which constructs a scale space of trees with respect to the resolution, the dimension and the dynamic range. The proposed method creates a scale space of the image that is easy to interpret, and provides a straightforward way for efficient parallel implementation since coarser scales could split a tree into independent subtrees that can be used in parallel construction of finer scales. The proposed method is tested on FDS data and remote sensing data to show the construction time and the resulting scale space.

21- Learning how to remove atmospheric seeing from ground-based solar observations
Presenting author: John Armstrong

Atmospheric seeing has been a problem for ground-based astronomy throughout history. Atmospheric seeing produces distortions in images due to varying density and temperature structure of the Earth's atmosphere. Bad seeing can be accounted for in part

by adaptive optics built into ground-based instruments. However, with the newer generation of higher resolution ground-based instruments AO systems cannot act quickly enough to remove the worst seeing from images. As a result, we propose a generative adversarial network (GAN) which will learn how to remove blur and distortions simulated onto space-based data from SOT. The goal of this network is generate solar images indistinguishable from ground-truth solar images. With the ability to generate these images, the model can then be applied to data with real seeing and they can be reconstructed with high accuracy to be included in our datasets for data analysis. The results are that spectroscopic and spectropolarimetric line profiles are successfully reconstructed by our network and so are feasible to be used for further data analysis.

22- The prospects of observing Tidal Disruption Events with the Large Synoptic Survey Telescope (LSST)
Presenting author: Katja Bricman

The Large Synoptic Survey Telescope (LSST) is an upcoming ground-based survey telescope in Chile, which will produce a multi-color survey of 18000 square degrees of the Southern Sky during its 10 years of observations. Due to its large coverage of the sky and its observing strategy it will be a perfect tool in search for transient astrophysical events, including rare Tidal Disruption Events (TDEs). TDEs occur when a star passes close by a supermassive black hole (SMBH) in a center of a galaxy and gets disrupted by its tidal forces. These events emit a bright flare of light, which can be observed to cosmological distances. Using an end-to-end LSST simulations framework we have simulated light curves of TDEs using different proposed LSST cadences. We present the preliminary results on the estimated absolute rates of TDE detection in 10 years of LSST operations and on the effect of different cadences on the observed light curves, especially in their early phase.

23- The GALAH survey: Solar twins and their multiplicity
Presenting author: Klemen Čotar

We used a set of more than 500k high resolution spectra acquired in the scope of the GALAH survey to search for stars whose spectrum resembles the Solar spectrum. Direct spectral comparison between observed and Solar spectrum was performed for every observed object. During this comparison, Gaussian process analysis was performed to model spectral noise in the observed spectrum and reduce its influence on the used spectral similarity metrices. Spectroscopic results, combined with the Gaia DR2 distance and magnitudes, show that some of the best matching candidates exhibit excess luminosity. To determine the number of stellar components in those systems, we modeled observed spectroscopic and photometric signatures using data driven models that were trained on the real observations.

24- Image Subtraction in Fourier Space for Transient Detection
Presenting author: Lei Hu

Image subtraction is a fundamental method for transient detection in time domain astronomy. Since point spread function (PSF) of images from ground-based telescopes is generally varying, mainly due to non-constant seeing condition, thus image subtraction is non-trivial for effective transient detection, for example in SN survey. Some convolution-based algorithm (like Alard & Lupton (1998) and Miller (2008)) has long been applied to solve such problems, and also new statistics-based optimal subtraction algorithm (Zackay, Ofek (2016)) was suggested to do it with better transient candidates identification.

Here we developed an algorithm of image subtraction in Fourier Space, where GPUs would be naturally able to speed up the underlying FFTs processes and make it to be a robust method for BIG astronomical time-domain data. It shows excellent performance especially for crowded field like LMC center images, and we have applied the algorithm for latest DECam SN-survey data.

25- Understanding galaxy sizes in the light of the galaxy-halo connection
Presenting author: Lorenzo Zanisi

The physical processes at play in shaping the sizes of galaxies are a matter of intense debate and current models of galaxy evolution struggle to reproduce both the observed size evolution and size distribution of galaxies. However, several studies have now shown that some of the properties of galaxies and those of their dark matter haloes are closely intertwined. We implement this idea in a simple semi-empirical model in which galaxy sizes are assigned according to their dark matter halo size. We compare our model results to data from CANDELS and SDSS and we show that this simple assumption is enough to reproduce both the size evolution and the size distribution of galaxies, provided that the relation between galaxy and halo size has an intrinsic scatter that varies with galaxy stellar mass as well as galaxy morphology. We give a physical interpretation of this result.

26- Stripped-Envelope Core-Collapse Supernovae: Principal Component Analysis
Presenting author: Marc Williamson

In the new era of time-domain astronomy, it will become increasingly important to have rigorous, data driven models for classifying transients, including exploding stars (SNe). We present the first application of Principal Component Analysis (PCA) to stripped-envelope core-collapse supernovae (SESNe). Previous studies of SNe types Ib, IIb, Ic and broad-line Ic (Ic-BL) focus only on specific spectral features, while our PCA algorithm uses all of the information contained in each spectrum. We use one of the largest compiled datasets of SESNe, containing over 150 SNe, each with spectra taken at multiple phases. Our work focuses on spectra taken 15\pm 5 days after maximum V-band light where better distinctions can be made between SNe type Ib and Ic spectra. We find that spectra of SNe type IIb and Ic BL are separable from the other types in PCA space, indicating that PCA is a promising option for developing a purely data driven model for SESNe classification.

27- A blazar science case for the Large Synoptic Survey Telescope (LSST)
Presenting author: Maria Isabel Carnerero Martin

In the next future (2022-2032) the Large Synoptic Survey Telescope (LSST; a 8.4 meter telescope located in Chile) will carry out a survey of the southern hemisphere sky (at least 18,000 square degrees). Each ~9 square degree observing field will be revisited every few days with a different filter (ugrizy).. This is known as the Wide-Fast-Deep survey (WFD, or "Main Survey"). Moreover, a fraction of time between 10 and 20% will be devoted to special surveys on selected sky areas: MiniSurveys and Deep Drilling Fields (DDFs), for which there is an upcoming call for proposal to select sky areas and define cadences.

The Italian Istituto Nazionale di Astrofisica (INAF) financially supports a Blazar group affiliated to the LSST Science Collaboration “Transients and Variable Stars” (TVS). The group is actively working to give inputs on the blazar science case. The collaboration with the LSST-TVS at this time is mainly based on:

- the definition of a best observational strategy for blazars, with particular interest in proposing an additional DDF covering a region of the sky containing a variety of blazars with different characteristics (bright/faint BL Lacs and FSRQs);

- the simulation of the LSST MAIN SURVEY and DDF cadences to investigate their effect on blazar studies;

- the test of different methods of AGN selection on blazars to identify strategies to discover new blazars.

We present preliminary results of the above analyses.

28- Gaia Study on the Formation of Intermediate Mass Stars
Presenting author: Miguel Vioque

The intermediate mass Herbig Ae/Be stars are young stars approaching the Main Sequence and are key to understanding the differences in formation mechanisms between magnetic low mass stars and the non-magnetic high mass stars. Almost all known Herbig Ae/Be stars have Gaia parallaxes which allowed us to derive their luminosities and locations in the HR diagram.

In combination with other catalogues, we studied the characteristics of the infrared excesses, of the photometric variability and of the Halpha emission line typical of these Pre-Main Sequence objects; resulting in the largest homogeneous analysis of these sources to date and a step forward to achieve a major increase of known objects of the class. Gaia Data Release 2 has improved and greatly increased the number of parallaxes available, which also allows us to search and identify new Herbig Ae/Be stars using the HR diagram. We present our plan to discover new Herbig Ae/Be stars by first constructing HR diagrams with all the sources within different star forming regions of interest and then combining the position in the HR diagram with the previously described characteristics. Different statistical learning techniques have to be applied to deal with such a large amount of data and find the region in the parameter space populated just by Herbig Ae/Be stars and without other objects. This in turn will allow us to study the Pre-Main Sequence evolution as a function of mass, age and location in the galaxy to an unprecedented precision.

29- Selection of AGN candidates through AGN optical variability
Presenting author: P. Sánchez

In preparation for the Large Synoptic Survey Telescope (LSST), which will revolutionize time-domain astronomy, it is critical that we understand variable phenomena as deeply as possible. Variability is arguably the defining feature of Active Galactic Nuclei (AGN), and is observed in every waveband, so variability studies are fundamental to understanding the extreme physical conditions of accretion disks near supermassive black holes. We are using the QUEST camera on the ESO-Schmidt telescope to obtain well sampled optical light curves of AGN in well-studied extragalactic fields that already have multiwavelength observations. The survey uses a broadband filter, the Q-band, similar to the union of the g and the r filters. We have a total of ~ 500.000 light curves in our five fields. In this poster, I will present our results on variability-based AGN selection. We have selected AGNs candidates in the more problematic region of the colour-colour parameter space, where high-redshift AGN resemble stars, and we have spectroscopically confirmed thirty of them achieving a 90% success rate.

30- Identifying reionization-epoch galaxies with extreme levels of Lyman continuum leakage in a JWST survey
Presenting author: Sambit K. Giri

The NIRSpec instrument on-board the James Webb Space Telescope (JWST) will allow spectroscopic studies of galaxies from the epoch of reionization (EoR). Recent observations of lensed high redshift galaxies suggest that JWST will provide a large sample of galaxy spectra from this era. The fraction of hydrogen-ionizing photons which do not escape into the intergalactic medium will leave an imprint on the spectra in the form of the spectral slope and strength of nebular emission lines.
Here we present a machine learning framework to predict the escape fraction of the hydrogen-ionizing radiation (Lyman continuum) photons using the spectra of high redshift ($z\gtrsim 6$) galaxies. We simulate JWST/NIRSpec observations of galaxies at $z=$6--10 by matching the fluxes of galaxies observed in the Frontier Fields observations of galaxy cluster Abell-2744 and MACS-J0416 for two exposure times with NIRSpec (10 h and 1.5 h). We find that the mean absolute error in predicting escape fractions $\Delta f_{esc}\approx$ 0.12 for the 10 h exposure time and three times worse for a 1.5 h exposure time. The framework also estimates the redshift of those galaxies with a relative error of 0.0128.

31- Shear-Ratio Test - A tool against systematics (in weak gravitational lensing)
Presenting author: Sandra Unruh

Weak gravitational lensing describes the effect of distortions in the galaxy shapes due to the intervening matter. Since the effect is rather small, observations are dominated by systematics that we have to carefully remove. A way to check consistency is to apply null-tests to the extracted data.

We will present the shear-ratio test as such a null-test. By using ray-tracing data through the Millennium simulation, we can quantify its result. Furthermore, we introduce a way to optimize the null-test strategy.

32- Cosmology with VEILS: Building an Infrared SN Ia Hubble Diagram
Presenting author: Tomás Müller

Type Ia Supernovae (SNe Ia) have been studied for many years as standardisable candles for cosmological distance measurement. However, most of this research has been done at optical wavelengths, where extinction highly affects the SN light, perhaps setting a fundamental limit in the cosmological precision of these objects. Near-Infrared (NIR) wavelengths are less affected by extinction, in principle making these bands better suited for cosmology. The VISTA Extragalactic Infrared Legacy Survey (VEILS) will obtain NIR light curves of around 300 SNe Ia out to a redshift of 0.6 over the next three years. I will provide an overview of the VEILS project, and a summary of the current status of my research.

33- Prediction of Cluster and Group Mass via a Machine Learning Approach
Presenting author: Victor Calderon

We make use of machine learning (ML) algorithms to try to correctly predict the masses of galaxy groups and clusters of galaxies from the Sloan Digital Sky Survey (SDSS). We use synthetic catalogues of galaxies that aim to represent the clustering and other properties of

galaxies in the real Universe. We use Random Forest, XGBoost, and Neural Network to predict the masses of these galaxy groups, and compare our results with the more traditional methods, such as halo abundance matching (HAM) and dynamical masses. We find that ML algorithms tend to predict the group mass to within 10% of the true group masses. However, HAM esti- mates tend to be slightly more robust in determining high masses. We find that the luminosity of galaxies and (g – r) color are the most important features used when determining group and cluster masses.

34- Visual Analytics for High-Dimensional Astronomical Data Using Dimension Reduction
Presenting author: Youngjoo Kim

Recent astronomical surveys like Gaia and GALAH provide scientists with an abundance of data, in terms of not only the number of observations, but also the number of dimensions. When visually analyzing such high dimensional datasets, it is crucial to reduce the number of dimensions to two or three, so that we can directly visually present the data for exploration purposes. Only after then we are able to visually search for underlying patterns or outliers in the data. In this poster, we propose the first steps towards a dimension reduction method that is scalable in terms of computational speed, feasible for cluster separation, and predictable for end-users. The high-dimensional data is first preprocessed using a technique called mean shift, where points within each potential cluster in the high-dimensional space move towards the center of the cluster. Then ISOMAP is used to perform dimension reduction. The proposed method is tested using both synthetic and real astronomical datasets including Gaia DR2 and GALAH's recent data release, and its performance and power to depict separate clusters in the projection is compared with that of t-SNE, a dimension reduction method with high cluster separation.

35- Identifying Blazar Variability with Machine Learning
Presenting author: Zachary R Weaver

One of the most common characteristics of blazars is the presence of rapid variability at almost all wavelengths of the electromagnetic spectrum. Thus, variability studies are one of the most powerful tools for understanding the physical processes of blazars. There have been several statistical methods to detect variability in light curves of blazars, such as the C-test, the F-test, the χ^2 test, and the analysis of variance (ANOVA) test. However, all such statistical tests require well-sampled light curves in order to determine variability. Machine learning has the capability to also detect variability for poorly-sampled light curves after being trained with well-sampled light curves. Here we present preliminary results in the use of machine learning to identify variability in the light curve of blazars, and an analysis of which light curve features are most helpful in determining variability.