1. Benjamin Joachimi (email: b.joachimi@ucl.ac.uk)
Title: Probabilistic cosmological analysis of large galaxy surveys

Project description:

The most promising route to shedding light on the fundamental open problems of dark matter and dark energy is to study the cosmic large-scale structure as probed by large galaxy surveys, such as the forthcoming ESA Euclid and LSST projects. The key physical signals are the weak gravitational lensing effect and the clustering of galaxies, extracted from up to a billion individual, very low signal-to-noise measurements. Traditionally, the cosmological analysis of these measurements is formulated as a very large, highly non-linear, multi-step, inverse problem, relying on many simplifying assumptions in order to obtain viable solutions. The analyses rely on compressing the huge amount of noisy raw data into a small number of derived measurements (like power spectra), which is computationally convenient but causes a number of flaws in the analysis process, for instance: spatially varying effects are averaged over and thus require an a posteriori correction; noisy observables and model predictions are conflated in the inference process; and there is mounting evidence that the ubiquitous Gaussian likelihood assumption for derived measurements is inaccurate (Sellentin et al. 2017). These problems can be avoided by resorting to a fully probabilistic forward-modelling approach via a Bayesian hierarchical model (BHM). The model consists of a network of probability distributions with which the inference process from observed data to fundamental parameters can be efficiently formulated and computed. All inter-dependencies between signals are accounted for, complex selection effects and spatial dependencies are naturally incorporated, and statistical uncertainties are fully propagated to final results. The key step towards robustness of clustering and weak lensing studies is their joint analysis in the same cosmic volume, including cross-correlations of the signals. This enables the self-consistent joint modelling of the astrophysical systematics together with the cosmology, as first quantified by Joachimi & Bridle (2010). As a bonus, degeneracies among cosmological parameters, present in the analysis of individual probes, can also be effectively lifted (e.g., Kirk et al. 2015a). We will therefore develop a BHM that simultaneously models weak lensing measurements ('shapes'), the angular clustering of galaxies ('counts'), and the galaxies' multi-band photometry ('colours') in order to jointly infer the fundamental model parameters of cosmology and the galaxy population. Recent implementations of part of this network have already shown promising results (Sanchez & Bernstein 2018; Alsing et al. 2016). With our new framework we will pioneer consistent and accurate cosmological inference on state of the art data from the KiDS and LSST surveys.

2. Supervisor: Jay Farihi (email: j.farihi@ucl.ac.uk)
Title: Spectra classification using Machine Learning techniques

Project Description:

The next generation of Large Spectroscpic Surveys will revolutionize Astrophysics in a multi-disciplinary manner. From cosmology to the Milky Way and local volume, there are a wide range of science goals for which millions of spectra will be obtained by surveys such as DESI, WEAVE, 4MOST, etc. At present, spectra are classified based on traditional techniques such as cross-correlation with spectral templates or standards, but this fails for new objects classes, or subtle variations within known object classes. This project will implement machine learning techniques to optimize the rapid classification of optical (galaxy and stellar) that are expected to be gathered in the millions within the next few years. The aims will include dramatically increasing the speed, efficiency and accuracy of object identification, and also explore the ability to distinguish hybrid and new classes of spectra.

3. Supervisor: Felipe Abdalla (email: filipe.abdalla@gmail.com)
Title: The next generation of optical and radio surveys in Cosmology

Project Description:

As we look forward, the next generation of optical and radio surveys will face serious data related challenges which will have to be solved in the areas of Cosmology. If we are not able to tackle these challenged we will not be able to extract all the necessary science from future surveys. This PhD studentship will focus on very specific areas where we need scaling solutions to be ready and prepared for the time we will have data analysis problems which are currently unsolved with current techniques.

In the area of Radio Cosmology, being able to tackle the vast amounts of data that the future SKA will be able to produce will be a challenge. Our group has developed techniques to fit data from these future surveys using Bayesian techniques which will allow for science such as weak lensing and intensity mapping to be produced with radio telescopes. These studies have shown that the use of high end GPUs and methods such as Hamiltonian Monte Carlo Sampling are necessary to tackle the problem appropriately. However these same studies show that a scalable implementation and arrangement of such prototypes is needed to fully implement these techniques with the data volumes which will be available when future telescopes are in place. The aim of this first part of the studentship will be to focus on these techniques for scaling to be achieved. These will be also implemented on CUDA cards, we intend to developed links with NVidia where possible.

Cosmology in optical has become a multi faceted approach where cross correlations will be at the heart of future discoveries. However obtaining theoretical forecasts and covariance's in order to make appropriate forecasts and inferences from data given the amount of data to be cross correlated will increase exponentially as the volume of data increases. The astronomy group at UCL has already put a significant amount of effort in developing techniques to speed up the and perform several types of cross correlation win optical data with improved techniques to speed up the MCMC approaches which will be used to perform this inference. This studentship will improve on these by implemented machine learning techniques to enhance the way parameters will be sampled and explored. We will use current datasets such as the dataset available by UCL researchers and benchmark these techniques. We envisage the student to spend around one year of research time in this project.
The DTC is a 4 year degree, we proposed above three projects in two distinct areas which should cover the time dedicated to research. The project as a whole requires cutting edge research software implementation with CUDA, MPI, machine learning and Monte Carlo methods. The primary supervisor has extensive experience in these area having supervised PDRA's in the above areas and the secondary supervisor will provide the crucial help with

optimisation of the choices from the scientific computing side where the primary supervisor has less experience.

4. Supervisor: Serena Viti (email: serena.viti@ucl.ac.uk)
Title: Determination of the most likely formation routes that lead to simple amino acids in space

Project Description:

Amino acids are important organic compounds that play a key role in the formation of proteins. In the protein production process, amino acids join together to form polymer chains, which represent the structural units of proteins. It is believed that the formation of amino acids may have occurred in the interstellar medium (ISM) since they have been firmly detected in meteorites. This is supported by laboratory experiments where amino acids are found to form on the surface of model dust grains by UV-photon and ion irradiation under astrophysically relevant conditions.

Despite great efforts towards their detection, a firm detection of amino acids in the ISM is still eluding us. Nevertheless, it is clear that such amino acids must form on the mantle of dust grains. It is therefore essential we determine the optimal conditions simple amino acids, such as glycine and alanine, can form on the dust icy mantles and are released back into the gas phase.

Key prerequisites for understanding how and where glycine and alanine form are the determination of the surface reactions that lead to their formation as a function of (i) temperature; (ii) gas density, (iii) structure of the gas, (iv) presence and abundance of other molecular species on the ices, (v) UV and cosmic ray ionization rates, among other factors. To date, no large-scale predictions of the formation routes of glycine and alanine have been made, as a function of the parameters above.

We propose an interdisciplinary project that seeks to address one of the fundamental problems in Astrobiology by making use of advanced and cutting edge techniques developed in the information sciences, machine learning and statistical disciplines, namely: the determination of the most likely surface reactions that lead to simple aminoacids, such as glycine, as a function of the environment. Specifically, the student will devise and apply a novel combination of statistical and neural networks techniques to perform large scale chemical models, involving large datasets of gas- and surface-phase chemical reactions to derive the physical and chemical conditions under which every surface chemical reaction is viable.

Recently (Holdship et al. 2018), we showed how, by using ab initio computational chemical models combined with Bayesian inference and MCMC sampling, one could a priori determine which reactions are key for the formation and destruction of a specific ice species.

The PhD student will couple Bayesian statistical techniques, Monte Carlo sampling methods and Neural Networks to solve the following challenge: simultaneously investigate the paths and efficiencies of formation and destruction of chemical reactions for glycine and alanine over a large physical (densities, temperatures, radiation fields, and cosmic ray ionization rates) and chemical (rate coefficients) parameter space, and discover if, where and under what conditions amino acids can form in space.

5. Supervisor: Serena Viti (email: serena.viti@ucl.ac.uk)
Title: Spectra line Identification in the ALMA era.

Project description:

The dramatic increase in sensitivity and frequency coverage in millimetre instruments in the last decade has made it possible for the first time to study the chemical complexity of the molecular interstellar medium (ISM) in galaxies beyond the Milky Way.

Thanks to ALMA molecular diagnostics are a powerful new tool to study the ISM of galaxies. These diagnostics are particularly important for the study of deeply dust-obscured young starbursts, such as those within luminous infrared galaxies.

However, before molecular diagnostics can be applied to distant galaxies or to large surveys, they need to be tested in local galaxies first and validated by physical and chemical models.

We were recently awarded a large program with ALMA, called ALCHEMI, to study the central molecular zone of the nearby starburst galaxy NGC 253. The proposed PhD project involves the development of AI tools for the identification, analysis and the modelling of the forest of molecular lines included in this survey with the aim of determining the physical conditions of the star forming gas in this galaxy. In particular the project will aim to design a neural network that can operate under flexible conditions required by the challenges that line identification at the high spectral resolution and sensitivity of ALMA. This may require the design of dependent neural network architecture inspired by those
used for sets.

6. Supervisor: Jason McEwen (email: jason.mcewen@ucl.ac.uk)
Title: Quantifying uncertainties in sparse imaging (compressive sensing) for radio interferometric and magnetic resonance imaging

Project Description:

In many fields high-dimensional inverse imaging problems are encountered. For example, imaging the raw data acquired by radio interferometric (RI) telescopes involves solving an ill-posed inverse problem to recover an image of the sky from noisy and incomplete Fourier measurements. Future telescopes, such as the Square Kilometre Array (SKA), will usher in a new big-data era for radio interferometry, with data rates comparable to world-wide internet traffic today. Magnetic resonance (MR) imaging (MRI) involves solving a very similar inverse problem to radio interferometric imaging. MRI also encounters a growing big-data problem as new imaging modalities are considered (e.g. 3D and diffusion MRI) and are pushed to higher resolutions.

Uncertainty quantification (e.g. error estimation) is an important missing component in both RI and MR imaging for quantitative imaging, inquiry, and decision-making. Moreover, since the imaging problems are often (severely) ill-posed, uncertainty quantification regarding reconstructed images becomes increasingly important. Nevertheless, uncertainty information is currently lacking in both RI and MR imaging techniques. In a recent series on companion articles, McEwen develop novel uncertainty quantification techniques for RI imaging, supporting the sparsity-promoting priors that have been shown in practice to be highly effective for both RI and MR imaging (Cai, Pereyra & McEwen 2017a, 2017b; arXiv:1711.04818, arXiv:1711.04819). Furthermore, McEwen also recently developed online algorithms for sparse radio interferometric imaging (Cai, Pratley & McEwen 2017; arXiv:1712.04462) but, at present, without uncertainty quantification. Online imaging is critical for big-data applications where imaging can begin as the data are acquired. In online imaging data can be assimilated as it is acquired and then discarded if the full data-set is so large that storage is not feasible (e.g. for the SKA).

In this project we will develop fast algorithms to apply the uncertainty quantification techniques presented in Cai, Pereyra & McEwen (2017a, 2017b) to big-data. Moreover, we will also develop online uncertainty quantification so that uncertainty quantification can be applied to extremely large data-sets that are too large to ever be stored. GPU implementations of the algorithms will be developed, as we have done already for the underlying sparse imaging methods (SOPT; http://baspgroup.github.io/sopt/). GPU implementations will be based on standard matrix algebra libraries so that ArrayFire can be used to deploy the code on various hardware (avoiding the need to write lower level GPU-specific code in, e.g., CUDA). All code will be made public and open-source. We will apply our fast, big-data uncertainty quantification methods to RI and MR data to demonstrate its effectiveness.

7. Supervisor: Nick Achilleos (email: nicholas.achilleos@ucl.ac.uk)
Title: Data Intensive Studies for Planetary Magnetospheres

Project Description:

- Modelling and observation – global circulation of atmospheres, magnetic field and plasma models based on MHD equilibrium
- Image processing
- Particle tracing in magnetic fields.

8. Supervisor: Daisuke Kawata (email: d.kawata@ucl.ac.uk)
Title: Revealing the Milky disk formation history with the Gaia data

Project Description: European Space Agency's Gaia mission (launched in Dec. 2013), which MSSL is heavily involved in, has made the second data release in April 2018, which provide the position and velocity measurements for more than one billion of stars in the Milky Way. These big data provide the information of kinematics of stars in the large fraction of the Milky Way disk for the first time. We are currently applying a Bayesian neural network model to measure the age and metallicity for the stars observed with the Gaia and ground-based spectroscopic survey. The combined information of the age, metallicity and kinematics for stars in the different region of the Milky Way disk must tell us the formation history of the

Milky Way disk. However, secular evolution mechanisms, such as radial migration due to the bar and spiral arms of the Milky Way, and kinematic heating by the bar, molecular clouds and satellite interactions, move the stars from their birth place, making the current stellar structure different from the initial structure when they were born. Hence, to decipher the Milky Way disk formation history from the observational data of the current Milky Way, this project will develop a data-driven Bayesian model of the Galactic disk formation, including the inside-out disk growth, radial migration of stars and heating mechanism. Then, we will fit the Gaia data with the model with Markov Chain Monte Carlo sampling, to reveal the formation history of the Milky Way disk and the significance of the radial migration and heating. Similar concept and strategies can be found in Frankel et al. (2018, ApJ, 865, 96), Kawata et al. (2018, MNRAS, 473, 867), Ting et al. (arXiv:1808.03278). The project will improve on these models, and combine it with our age estimate model.