Summary of technical skills for the ESO fellowship

Claudia Del P. Lagos - ESO fellowship

My technical skills are mainly computing-oriented, which I summarize in five broad points below.

Large data management

One of the main issues to deal with when running Cosmological simulations and galaxy formation models, is the amount of data generated and stored. In GALFORM we make extensive use of the Hierarchical Data Format, which was designed to store and organize large amounts of numerical data. This has given me the expertise to deal with such amounts of data and the most effective ways to store them. In the same line, as part of my research on the relation between supermassive black holes and their host galaxies, I made use of the Sloan Digital Sky Survey, Data Release 7, which represents one of the largest surveys currently available, in terms of the number of objects observed. This trained me in SQL and gave me more experience in the management of large amounts of data from the point of view of observations.

Visualization of simulated images

An important part of the theoretical work in galaxy formation is to construct softwares to help the visualization of simulated galaxies. Within this, I have been working on visualizing the atomic and molecular hydrogen, and the Carbon-monoxide emission of simulated galaxies. I have written my own software to visualize column densities of gas and brightness temperature intensities to predict how a galaxy field would be observed, for example, by ALMA, WALLABY, etc. I am also currently making use of the CASA software, and in particular, the ALMA simulator, to include instrumental errors in my catalog of simulated images.

Data Analysis

My main tool for data analysis is IDL, and secondarily C and fortran. I have been developing code in IDL for data analysis, oriented to statistical assessment, for 5 years. The typical amount of data I have to deal with is very large, e.g. billions of galaxies are output by GALFORM at different cosmic epochs. This has trained me in many statistical tools to describe the behavior of galaxy samples, such as the jackknife technique, bootstrapping, chi-square analysis, polynomial fitting, etc. Some of these tools are already part of the open IDL packaging, but most of them I had to developed myself on the IDL environment.

Scripting and programming languages

I have extensively developed code for the semi-analytic models of galaxy formation SAG and GALFORM, which are written in C and Fortran, respectively. This has given me an advanced knowledge on these two computing languages. I also have experience in running and modifying halo finder algorithms such as the Sub-Find algorithm, and the AMIGA algorithm, which are both written in C. I have also been formally trained in Message Passing Interface (MPI) by the Institute for Computational Cosmology and the Information Technology Service in Durham University, in the context of these two languages, C and Fortran.

Unix, Linux and high-performance computing

My working environment has been Unix since 2003. My research insemi-analytic models of galaxy formation has given me access to big supercomputers. I am currently working in the cluster COSMA4 at Durham which has a total of 2640 Intel X5650 cores and a total storage capacity approximately 612 Tera bytes. By working in this cluster, I have acquired a fruitful knowledge on bash, csh, queuing systems and high-performance computing. In the past, I worked in the cluster GERYON at Pontificia Universidad Catolica de Chile, composed by 512 cores and a storage capacity of 13 Tb, which allowed my first experiences with supercomputing.

I also have also included in my Curriculum Vitae general knowledge I have in plotting softwares and in special packages of image and spectra reduction in which I was formally trained (i.e. Coplot, supermongo, IRAF), but these do not constitute part of my daily work.