Astronomy and Computing 3–4 (2013) 1–12

Contents lists available at ScienceDirect

Astronomy and Computing

journal homepage: www.elsevier.com/locate/ascom

Full length article

A BOINC[1] based, citizen-science project for pixel spectral energy distribution fitting of resolved galaxies in multi-wavelength surveys

Kevin Vinsen[a,*], David Thilker[b]

[a] International Centre for Radio Astronomy Research (ICRAR), The University of Western Australia, M468, 35 Stirling Highway, Crawley, Perth, WA 6009, Australia

[b] Department of Physics & Astronomy, The Johns Hopkins University, 3400 N. Charles Street, Baltimore, MD 21218, USA

ARTICLE INFO

Article history:
Received 12 June 2013
Accepted 1 October 2013

Keywords:
BOINC
Pan-STARRS1
Cloud computing
Spectral energy distribution
Distributed client computing
MAGPHYS

ABSTRACT

In this work we present our experience from the first year of theSkyNet Pan-STARRS1 Optical Galaxy Survey (POGS) project. This citizen-scientist driven research project uses the Berkeley Open Infrastructure for Network Computing (BOINC) middleware and thousands of Internet-connected computers to measure the resolved galactic structural properties of ∼100,000 low redshift galaxies. We are combining the spectral coverage of GALEX, Pan-STARRS1, SDSS, and WISE to generate a value-added, multi-wavelength UV–optical–NIR galaxy atlas for the nearby Universe. Specifically, we are measuring physical parameters (such as local stellar mass, star formation rate, and first-order star formation history) on a resolved pixel-by-pixel basis using spectral energy distribution (SED) fitting techniques in a distributed computing mode.

1. Introduction

TheSkyNet (TheSkyNet, 2013) is a citizen science project that aims to raise the public profile of science, specifically astronomy and radio astronomy. At the same time it creates a research grade data processing resource for astronomers. TheSkyNet allows members of the public to donate their spare computing power to process astronomy data, with the software running quietly in the background of their personal machines.

In August 2012, The International Centre for Radio Astronomy Research (ICRAR) and The Johns Hopkins University (JHU) signed a Memorandum of Understanding for theSkyNet to process observations of galaxies observed by Pan-STARRS1 (PS1) (Kaiser et al., 2010). This dataset is complemented by multi-wavelength imaging from: the Sloan Digital Sky Survey (SDSS) (York et al., 2000), Galaxy Evolution Explorer (GALEX) (Martin et al., 2005) and Wide-field Infrared Survey Explorer (WISE) (Wright et al., 2010).

The Panoramic Survey Telescope and Rapid Response System (Pan-STARRS) will use gigapixel mosaic cameras on four modest aperture telescopes to survey the entire Hawaiian sky in visible and near-IR bands. The first Pan-STARRS prototype survey began in May 2010 using a single telescope system (PS1). The PS1 telescope has a 1.8-m $f/4$ design, producing a 3° diameter field of view. Images are recorded on a 1.4 gigapixel CCD camera (Tonry and Onaka, 2009), the world's largest instrument of its kind. All PS1 science goals are supported by several parallel surveys (Chambers, 2006). The 3π Steradian Survey covers the entire sky north of declination $-30°$ with a total of 60 epochs, twelve in each of 5 filters (*grizy*). The 3π survey is providing images of higher resolution (mode FWHM $= 1.2''$ for g and $< 1''$ for all other bands) and moderately higher sensitivity than SDSS, over a much larger area on the sky.

TheSkyNet Pan-STARRS1 Optical Galaxy Survey (POGS) project (POGS, 2013) was created to efficiently process and interpret the multi-wavelength data from the resolved galaxies detected by PS1. This large volume of data would take many tens of years to process on a large University computing cluster. We believe that this can be done in 2–3 years with theSkyNet using the BOINC middleware (BOINC, 2013a). For each catalogued galaxy meeting our selection criteria, the computers of our volunteers are being harnessed, via the Internet, to perform pixel-by-pixel spectral energy distribution (SED) fitting. This will provide us with physical parameters about the galaxy, such as the stellar mass, star formation rate (SFR), dust attenuation, and first-order star formation history (SFH). With the resultant pixel SED fitting output, we will then constrain parametric models of galaxy morphology in a more meaningful way than ordinarily achieved. In particular, we will fit multi-component (bulge, bar, disk) trial galaxy models, directly to the distribution of stellar mass rather than surface brightness in a single band, which is locally biased by star formation activity and dust obscuration.

To date (as at 30-Sep-2013) the project has 4663 active users with 11,629 computers and has processed 8348 galaxies or 53

* Corresponding author. Tel.: +61 435 164 060.
E-mail addresses: kevin.vinsen@uwa.edu.au, kevin.vinsen@icrar.org (K. Vinsen).

[1] Berkeley Open Infrastructure for Network Computing.

2213-1337/$ – see front matter
http://dx.doi.org/10.1016/j.ascom.2013.10.001

million pixels. We are presently a modestly sized BOINC project compared to the '@Home' projects, which have been operating for many years and have tens of thousands of volunteers, and in the case of SETI@Home, a million plus volunteers. We anticipate significant continued growth in our user base as we are currently growing at ∼50–60 computers a day.

The acronym POGS is a reference to a game played with discs, originating on Maui, Hawaii, in the 1920s (Wikipedia, 2013) and the fact that the Pan-STARRS PS1 telescope is situated on Mount Haleakala, Maui.

2. Spectral energy distribution fitting and MAGPHYS

The power of multi-wavelength data for understanding the substructure of nearby galaxies has only just begun to be exploited systematically. Various authors have described methods to extract the information content in a pixel using SED based methods. These include:

1. SED fitting (Crockett et al., 2011; Jeong et al., 2007);
2. pixel-z (Conti et al., 2003; Welikala et al., 2009; Wijesinghe et al., 2010);
3. PCA (Zibetti and Groves, 2011).

For this project, the pixel-by-pixel SED fitting is done using the 'Multi-wavelength Analysis of Galaxy Physical Properties' code (MAGPHYS) of da Cunha et al. (2008). We have adapted it to run within the BOINC distributed-computing framework. Other SED fitting codes are also being evaluated for our project. MAGPHYS was implemented first because it offers a physically motivated model to interpret stellar and dust emission at UV, optical, and infrared wavelengths in a self-consistent manner.

Evolution of the intrinsic starlight from stellar populations of a specified SFH are computed using the Bruzual and Charlot (2003) population synthesis code. The allowed SFHs are realistic in the sense that they are a combination of parameterised exponential declines plus a superposition of random bursts at later times.

MAGPHYS adopts the two-component model of Charlot and Fall (2000) in order to realistically predict the attenuation of starlight for stars of various ages, and the corresponding re-emission of this energy by dust in the interstellar medium (ISM). The two redshift-dependent model libraries consist of one library of 50,000 combinations of stellar population spectra, reflecting the SFH (UV-optical) model being used and a second library to describe the Infrared emission from dust, using the approach of da Cunha et al. (2008).

From these 1.54 GB files, MAGPHYS generates two models to be used for the SED fitting based on the filters used in the observations. The filter bands used in theSkyNet POGS at the time of writing are shown in Fig. 1. MAGPHYS allows the redshift of these model files to be within $z = 0.005$ of the redshift of the observed pixels. This means we do not have to generate model files for every redshift of every galaxy. Currently, the observed galaxies have varying redshifts up to a maximum $z = 0.1$, so we have created model files for redshifts of: 0.00, 0.01, 0.02, …, 0.09, 0.10.

The resultant model files are quite large and increase in size with the number of filter bands being processed. Using the filter bands listed in Fig. 1 the star formation model files (describing the UV-optical portion of the SED) are ∼15–20 MB, and the infrared model files (describing the dust properties) are ∼5–10 MB.

Using these two models, MAGPHYS takes the observed fluxes and the associated uncertainties through the various filters (both in Jy) and calculates the physical parameters shown in Table A.5. For each parameter the seven values listed in Table A.6 are recorded. The likelihood distributions for each of the physical parameters in Table A.5 are also recorded.

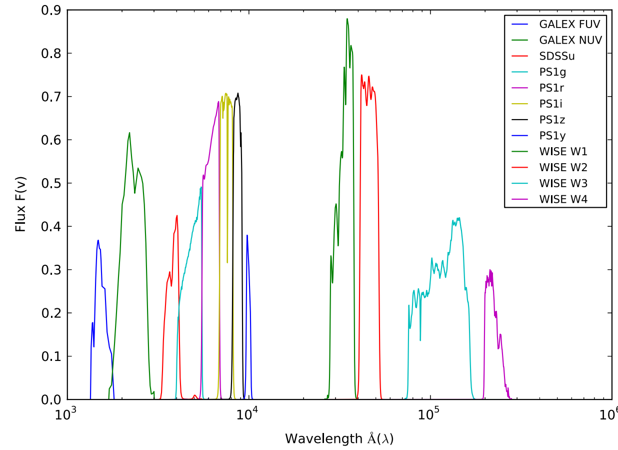

Fig. 1. Typical filter curves used in a MAGPHYS run with PS1 data.

3. BOINC distribution and processing

The BOINC middleware is well proven through projects like: SETI@Home (Korpela, 2012) and Einstein@Home (Allen et al., 2013). A detailed description of the BOINC middleware can be found in Korpela (2012) and Allen et al. (2013). We will only provide a brief overview here. The basic components of a BOINC server are:

1. the BOINC daemons, called the scheduler, feeder, transitioner, validator, assimilator and file deleter;
2. the Apache web server with two CGI programs;
3. the MySQL database server.

Only two programs must be provided by the project: the validator and the assimilator. The rest are provided 'out of the box' by the BOINC project.

The validator decides whether results returned by the clients are correct. It does this by comparing the results of runs from different computers. If the results match it assumes the results are valid. The MAGPHYS code rounds its output to only 3 decimal places, which means results tend to be consistent across different operating systems and hardware. Differences do occasionally occur, mainly when an old 32-bit machine compares results against a 64-bit machine or an Android device. This is discussed further in Section 9. TheSkyNet POGS validator performs a simple line by line comparison. It excludes the end of line characters because they are different for different operating systems. If the validator cannot get a matching result after five attempts it will abandon that particular work unit.

The assimilator processes completed jobs. It parses the text file output from MAGPHYS and only stores the 4 results needed to draw the images described in Section 7 into the database. The text file is then copied to a high availability file store for processing when all the pixels for a galaxy have been returned.

An overview of the process of getting work to the clients is given in Fig. 2. The steps are as follows:

1. The BOINC client requests a number of new work units (a work unit describes the computation to be performed) from the BOINC server.
2. The server checks the database for work units. BOINC has a number of different distribution methods depending on the nature of the project. POGS uses the simplest priority-order, which takes high priority tasks first (created when validation fails) and then tasks in the order they were created.
3. The database gives back details of work units to be assigned to the client.