Comparing theoretical models of our galaxy with observations

S. Sharma[1,a], J. Bland-Hawthorn[1], K.V. Johnston[2]
and J. Binney[3]

[1]*Sydney Institute for Astronomy, School of Physics, University of Sydney,
NSW 2006, Australia*
[2]*Department of Astronomy, Columbia University, New York, NY-10027, USA*
[3]*Rudolf Peierls Centre for Theoretical Physics, 1 Keble Rd, Oxford, OX1 3NP, UK*

Abstract. With the advent of large scale observational surveys to map out the stars in our galaxy, there is
a need for an efficient tool to compare theoretical models of our galaxy with observations. To this end, we
describe here the code *Galaxia*, which uses efficient and fast algorithms for creating a synthetic survey of
the Milky Way, and discuss its uses. Given one or more observational constraints like the color-magnitude
bounds, a survey size and geometry, *Galaxia* returns a catalog of stars in accordance with a given theoretical
model of the Milky Way. Both analytic and N-body models can be sampled by *Galaxia*. For N-body models,
we present a scheme that disperses the stars spawned by an N-body particle, in such a way that the phase
space density of the spawned stars is consistent with that of the N-body particles. The code is ideally
suited to generating synthetic data sets that mimic near future wide area surveys such as GAIA, LSST and
HERMES. In future, we plan to release the code publicly at http://galaxia.sourceforge.net. As an application
of the code, we study the prospect of identifying structures in the stellar halo with future surveys that will
have velocity information about the stars.

1. INTRODUCTION

Over the last decade there has been great progress both in theory and observations to understand the
formation of the Milky Way. Large photometric surveys, such as 2MASS and SDSS, have mapped out
a large portion of our galaxy. In addition, a number of spectroscopic surveys, such as GCS, SEGUE and
RAVE, have enabled us to also get radial velocity and spectra of a large number of stars. Even bigger
observational surveys like the LSST, GAIA, SkyMapper are being planned in the near future.

With the tremendous increase in the amount of data, in the form of number of stars and its span
across the sky, there is a need for an efficient tool to compare theoretical models of our galaxy with
observations. To do such a comparison, a theoretical model needs to be transferred into the observational
space and then convolved with appropriate observational uncertainties. *Galaxia* as a code is designed to
facilitate this. A schematic description of the framework of *Galaxia* is shown in Fig. 1.

The present state of the art codes for analytical modeling of the Milky Way are the Besançon
code by [1] and TRILEGAL by [2]. Here the disc is constructed from a set of isothermal populations
that are assumed to be in equilibrium. Analytic functions for density distributions, the age/metallicity
relation and the IMF are provided for each population. However, the current schemes of sampling
analytical models have a number of drawbacks. They were primarily designed for simulating stars along

---

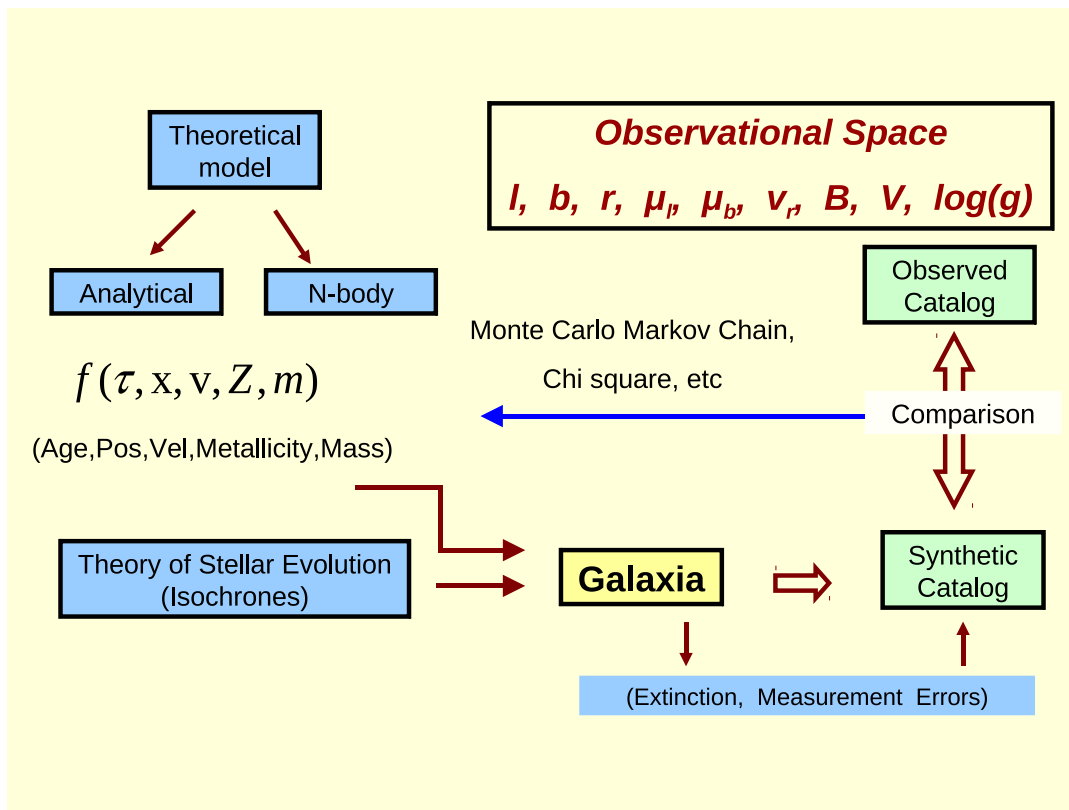[a]e-mail: `sanjib@physics.usyd.edu.au`

Figure 1. A schematic overview of comparing theoretical models of our galaxy with observational surveys and the role of *Galaxia*.

a particular line of sight and are hence, limited in their ability for generating wide area surveys. They typically require the user to set discrete step sizes for radial, and angular coordinates and results might differ depending upon the chosen step size. Moreover, they are computationally expensive for generating a large catalog of stars.

Additionally, analytical models cannot simulate substructures or incorporate features formed by tidal disruption of satellite galaxies. N-body models, employing collision-less particles, are ideally suited for this purpose. N-body hydrodynamical simulations, using cosmological initial conditions and incorporating star formation and feedback, are also being done to understand the formation of the Milky Way. However, as highlighted by [3] there are several unresolved issues related to sampling of an N-body model. In order to overcome these limitations, in *Galaxia* we use a novel approach to sampling theoretical models and this is described below.

2. METHODS

In order to sample an analytic model we use the von Neumann rejection technique which is ideally suited for continuous sampling over a multidimensional space. As the naive approach is computationally expensive, we split up the galaxy into a set of roughly equal mass nodes and apply the rejection sampling over them. Splitting up the galaxy into multiple nodes also allows us to further optimize the code for speed, e.g., depending upon the distance of the node one can choose a suitable lower limit on the mass of the star to be generated. For sampling the N-body models we use the multi-dimensional density

estimation code EnBiD [4], which gives us the phase space volume of each N-body particle. Stars spawned by each N-body particle are then dispersed over this phase space volume. The advantage of this approach is that the sampled stars obey the underlying phase space density of the N-body model.

Given a mass, age and metallicity of a star, a library of synthetic stellar isochrones by the Padova group [5, 6], is employed to generate its observational properties, e.g., colors, magnitudes and other stellar parameters. Finally, a 3d extinction model, consisting of a double exponential disc with warp and flare (scale length $h_R = 4.4$ kpc, and scale height $h_z = 0.088$ kpc), whose $E(B - V)$ at infinity matches the Schlegel maps is used to get the final apparent magnitude.

As a concrete example, we use the Besançon analytical model for the disc and for the stellar halo, we use the simulated N-body models of [7] that can reproduce the substructure in the halo. Further details on *Galaxia* are described in [8].

2.1 Computational performance

Since, the galaxy is simulated by splitting it up into a set of roughly equal mass nodes, the run time of the code is nearly linear with the mass of the galaxy being simulated. On a single 2.44 GHz CPU, *Galaxia* can generate stars at a speed of about 0.16 million stars per second. For shallower surveys the speed is about 3 times less. A $V < 20$ magnitude limited survey of the NGP, covering 10,000 square degrees and consisting of about 35 million stars, can be generated in 220 seconds. A $V < 20$ all sky GAIA like survey would require about 6 hours on a single CPU.

3. GALACTIC ARCHEOLOGY WITH THE STELLAR HALO

As an application, we study the ability of future surveys to detect structures in the stellar halo. In the currently favored ΛCDM paradigm of structure formation the stellar halo is thought to have been built up by accretion of satellite galaxies. N-body simulations of the hierarchical build up of the stellar halo were carried out by [7]. We employ these simulations in *Galaxia* to generate synthetic catalogs of stars corresponding to different types of surveys.

In order to objectively identify the structures we use the multi-dimensional group finding algorithm EnLink [9]. The advantage of using EnLink being that it can work in spaces of arbitrary dimensions. Moreover, it has an in built significance estimator such that only genuine groups that stand out above the Poisson noise are considered. Since, the N-body models are constrained by the number of dwarf galaxies in the Milky Way, we ignore the bound structures from our analysis and consider only the unbound structures.

Group finding studies in three dimensional (l, b, r) space, as mapped by photometric surveys, was carried out by [10] and it was shown that the number of structures in N-body models are roughly consistent with those seen in the 2MASS photometric survey. It was also shown that depending upon the stellar population used to map the stellar halo different surveys are sensitive to different types of accretion events. Here, we specifically address the question of what more can we learn if we also have velocity information, e.g., radial motion and/or proper motions? Since, observational uncertainties play a crucial role in our ability to detect structures we analyze surveys with different amounts of observational uncertainties.

We first generated a sample of 10^5 stars with a color range of $0.1 < g - r < 0.3$ and an apparent magnitude limit of $M_r < 27.5$. This typically samples the main sequence turnoff stars out to 400 kpc but could be used as a proxy for any other tracer population which samples the halo without any significant bias. Using this sample, we generated a set of surveys with different observational uncertainties in distance, radial velocity and proper motion. Group finder was then run on each of them in three dimensional (x, y, z) space, four dimensional (x, y, z, v_r) space and six dimensional (x, y, z, v_r, v_l, v_b) space. The number of groups recovered per halo (averaged over 11 stellar halo), as a function of distance (binned in steps of 10 kpc) are shown in Fig. 2. Note, a single accretion event might correspond to more
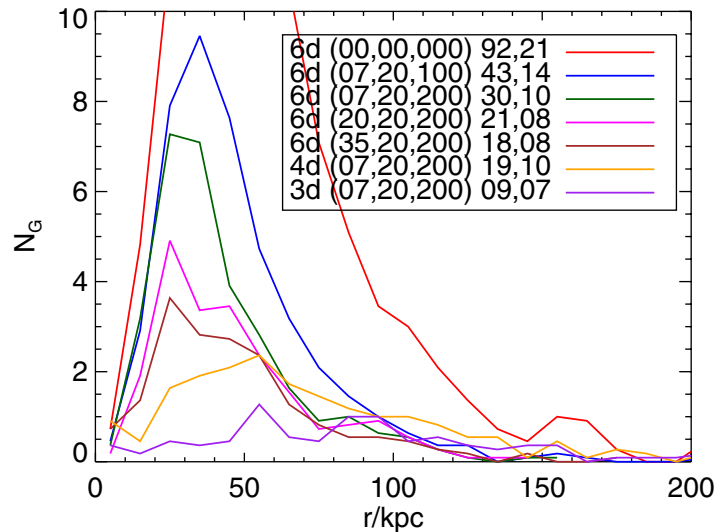
Figure 2. Comparison of the ability to detect structures in the stellar halos by different type of surveys. Surveys with and without velocity information and with varying amounts of uncertainties in radial distance, radial velocity and proper motion are compared. The plot shows the number of groups (in bins of 10 kpc each) that can be discovered as a function radius. The three quantities in parenthesis are the percentage radial distance uncertainty, the radial velocity uncertainty in km/s and the proper motion uncertainty in μas/yr. The last two quantities are the number of groups recovered and the number of unique accretion events they correspond to.

than one recovered group. Hence, in the figure we label both the number of groups and the unique accretion events they correspond to. It can be seen from the figure that as one adds radial velocity and proper motion information, we discover more groups. Also we sample a larger number of unique accretion events. Specifically in the distance range 20–70 kpc, when velocity information is added, there is a potential for discovery of a large number of new structures.

4. CONCLUSION

We have shown that the code *Galaxia* is ideally suited for comparing theoretical models of the Milky Way with the current and upcoming observational surveys. Its application to the stellar halo suggests that upcoming surveys with velocity information should be able to discover a lot of new structures in the distance interval of 20–70 kpc.

References

[1] A.C. Robin, C. Reylé, S. Derrière, S. Picaud, A&A **409**, 523 (2003)
[2] L. Girardi, M.A.T. Groenewegen, E. Hatziminaoglou, L. da Costa, A&A **436**, 895 (2005)
[3] A.G.A. Brown, H.M. Velázquez, L.A. Aguilar, MNRAS **359**, 1287 (2005)
[4] S. Sharma, M. Steinmetz, MNRAS **373**, 1293 (2006)
[5] P. Marigo, L. Girardi, A. Bressan, M.A.T. Groenewegen, L. Silva, G.L. Granato, A&A **482**, 883 (2008)
[6] G. Bertelli, A. Bressan, C. Chiosi, F. Fagotto, E. Nasi, A&AS **106**, 275 (1994)
[7] J.S. Bullock, K.V. Johnston, ApJ **635**, 931 (2005), arXiv:astro-ph/0506467
[8] S. Sharma, J. Bland-Hawthorn, K.V. Johnston, J. Binney, ApJ **730**, 3 (2011)
[9] S. Sharma, K.V. Johnston, ApJ **703**, 1061 (2009)
[10] S. Sharma, K.V. Johnston, S.R. Majewski, J. Bullock, R.R. Muñoz, ApJ **728**, 106 (2011)