A comprehensive tool for the statistical comparison of Large Surveys to Models of the Galaxy

Andreas Ritter[1]

[1]Graduate Department of Astronomy, National Central University,
300 Zhongda Rd., Jhongli City, 325 Taoyuan County, Taiwan,
email: `azuri.ritter@gmail.com`

Abstract. The advent of large spectroscopic surveys of the Galaxy offers the possibility to compare Galactic models to actual measurements for the first time. I have developed a tool for the comprehensive comparison of any large data set to the predictions made by models of the Galaxy using sophisticated statistical methods, and to visualise the results for any given direction. This enables us to point out systematic differences between the model and the measurements, as well as to identify new (sub-)structures in the Galaxy. These results can then be used to improve the models, which in turn will allow us to find even more substructures like stellar streams, moving groups, or clusters. In this paper I show the potential of this tool by applying it to the RAdial Velocity Experiment (RAVE, Steinmetz 2003) and the Besançon model of the Galaxy (Robin *et al.* 2003).

Keywords. methods: statistical, surveys, stars: statistics, Galaxy: kinematics and dynamics, Galaxy: stellar content, Galaxy: structure

1. Introduction

The current and future large spectroscopic surveys of the Galaxy offer the possibility for the statistical comparison to Galactic Models for the first time. These comparisons aim to answer questions like

- What are the selection effects and systematics of the survey?
- Where do we need to change the model, and why?
- What is the exact structure of the Galaxy?

In order to investigate these questions, a comprehensive tool for the comparison using sophisticated statistical methods is needed.

No survey can observe all stars in the Galaxy, instead only an (ideally) random sample of the underlying stellar populations can be observed. The main reasons for this are crowding and extinction. We therefore need to compare the full data set provided by a theoretical model of the Galaxy to a random sample observed by the survey. Using the standard method of plotting histograms for both observational and model data, naturally discrepancies between the data sets will show up. However, this method cannot answer the most important question: How significant is the discrepancy from a statistical point of view?

The approach presented here is to mimic the survey a number of times by randomly choosing stars from the model, creating multiple data sets which take into account the selection function(s) of the survey. For each parameter and data set the moments of the
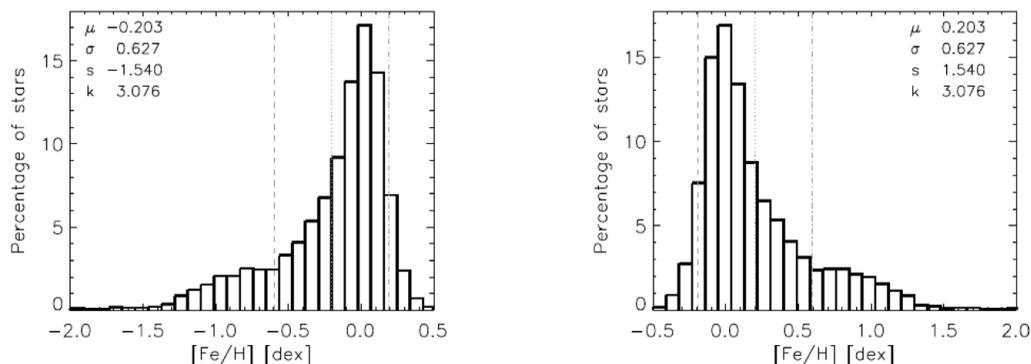
Figure 1. The figure shows the two parent distributions chosen to test how well we can reproduce the moments with random samples. Note that the second distribution is the same as the first one, but was flipped with respect to the zero axis. The vertical lines indicate the mean value plus/minus one standard deviation.

parameter distribution are calculated, and statistically compared to the moments from the survey data.

2. Justification of the Method

The first question which needs to get answered to justify the method presented here is: How well we can reproduce the moments of a parent distribution by choosing random samples, and how many stars and samples are needed to achieve reliable results? To answer this question one random sample from the Besançon Model of the Galaxy containing 10,000 stars was chosen. As the parent distribution we selected the metallicities from our random sample. As a control parent distribution with opposite mean value and skewness, but same standard deviation and kurtosis, the first distribution was flipped with respect to the zero axis (Fig. 1). For the first test 1,000 random samples containing 1,000 stars each (10% of the parent distribution) were taken, and the moments calculated. Fig. 2 shows the differences of the mean values of the moments of a certain number of random samples, and the moments of the parent distributions (zero axis). The horizontal lines indicate the standard deviations of each moment after 1,000 random samples. As expected, the mean values and standard deviations get reproduced very well even after a small number of samples. Note that the random samples are systematically less skewed and peaked than the parent distributions. However, the mean values of the skewness and kurtosis are safely within one standard deviation. The effect is small, but needs to be taken into account for a detailed analysis of survey data to models of the Galaxy. As a second test it was investigated how large the sample size needs to be to get reliable results. Different sizes of parent distributions were taken by randomly selecting stars from the distributions shown in Fig. 1, and the mean values of the moments for different sample sizes were calculated. The results of this test are shown in Fig. 3. The colour code shows how the mean values of 1,000 random samples compare to each moment of the parent distribution. For the mean values and standard deviations an excellent agreement for all possible combinations is found, while certain minimum sample sizes for the skewness and kurtosis are required for a reliable reproduction of the respective moment of the parent distribution.
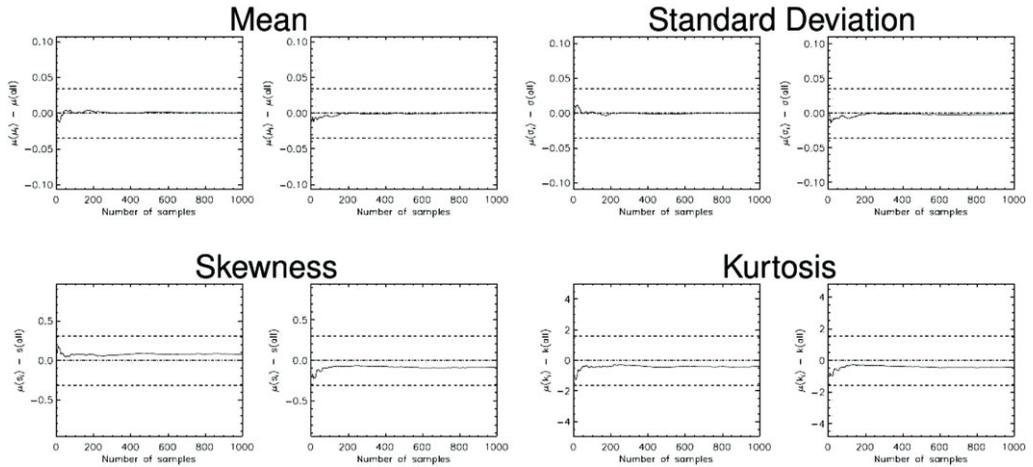
Figure 2. Plots comparing the mean values of the moments after a certain number of random samples to each moment of the parent distributions. The horizontal lines mark one standard deviation after 1,000 samples.
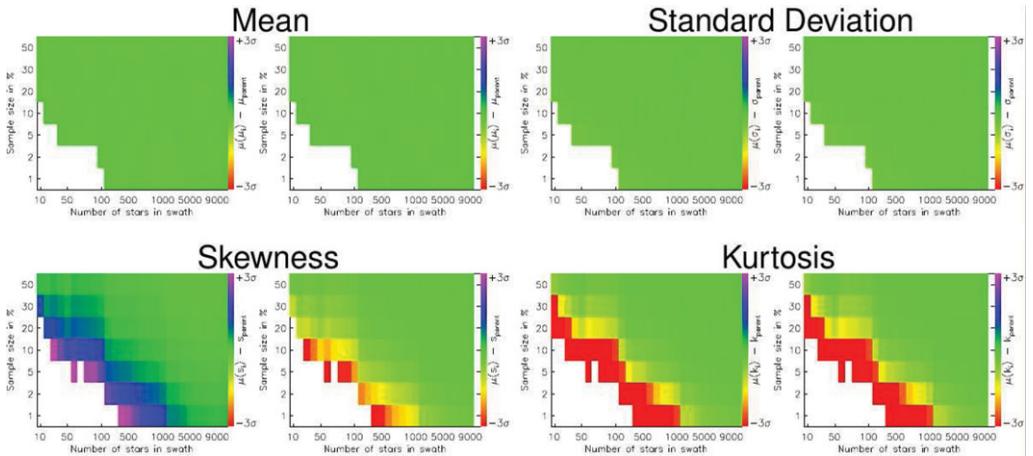
Figure 3. Plots comparing the mean values of the moments after 1,000 random samples to each moment of the parent distributions for different sample sizes with respect to the size of the parent distributions. (Coloured figures are available in the online version.)

3. Comparison RAVE – Besançon for the radial velocities of giant stars

Now that we have validated the statistical method for the comparison of large Galactic surveys to models of the Galaxy, we can apply the method to some real data. As example data sets the RAdial Velocity Experiment (RAVE, Steinmetz 2003) and the Besançon Model of the Galaxy (Robin *et al.* 2003) were chosen. The RAVE Survey observed ∼500,000 stars with $9 \leqslant I_C \leqslant 12$ mag in the whole southern hemisphere. It has been nearly completed for the brightest stars, while not all fainter stars from the input catalogue could be observed because of technical problems and time constraints. As a magnitude limited survey, the only selection function for observed RAVE stars taken from Input Catalogue 2 (IC2) is a different percentage of observed stars in the I_C-magnitude bins compared to the underlying stellar populations (nearly 100% for bright stars, a lot less for the faintest stars). For the comparison the internal data release 10 (IDR10) was
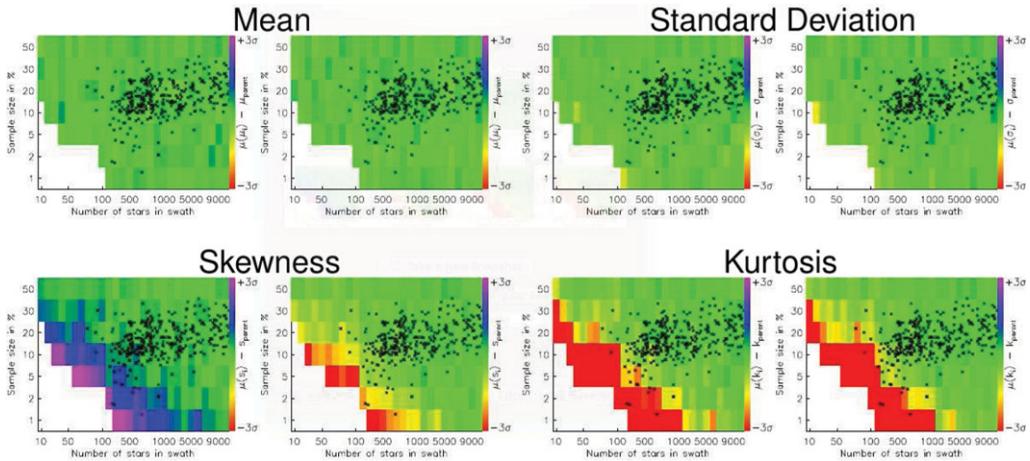
Figure 4. Same as Fig. 3, but for 30 random samples. Over-plotted are the sizes of the RAVE data with respect to the Besançon data for swaths on the sky which are 5×5 deg wide in l and b, respectively. (Coloured figures are available in the online version.)

used, containing 456,637 observations in total. IC1 suffered from subtle colour biases this means that the distribution of RAVE stars in the colour-magnitude diagram is not entirely representative of the underlying Galactic stellar population (Seabroke 2007). Therefore all RAVE stars which were present in IC1 but not in IC2 were removed from the data set, as well as all stars which were flagged as somehow peculiar, and multiple observations, leaving the one with the highest signal-to-noise ratio. This procedure left a total of 327,652 stars in the final sample. After convolving the Besançon I magnitudes with the 2MASS errors, and the radial velocities with the typical errors in the RAVE survey, we can feed both data sets into the comparison tool. To mimic the RAVE survey, the program fills up small bins in I_C observed by RAVE with randomly selected stars from the Besançon Model. This procedure is repeated for 30 random samples (30 to keep the computing time within reasonable limits). The justification for this relatively small number of random samples is shown in Fig. 4. This figure is the same as Fig. 3, but for 30 random samples, over-plotted with the sample sizes from the RAVE data with respect to the sizes of the Besançon data set for $\sim$1,000 5×5 deg (in l and b) swaths on the sky. About 95% of the swaths are in the area where 30 random samples can reproduce the moments of the parent distribution reliably.

For each swath histograms of the parameter distributions of the RAVE survey and the random samples drawn from the Besançon Model were plotted, and the moments calculated. Fig. 5 shows how the mean and standard deviation of the RAVE radial velocities and calibrated Metallicities (See Siebert *et al.* 2011 for the calibration procedure) compare to all Besançon stars as well as the 30 samples. Note that always two parameters are compared at the same time, e.g. radial velocity vs. metallicity or effective temperature vs. surface gravity. This allows for a more detailed later investigation of the individual swaths.

In Fig. 6 the results of the comparison of the mean radial velocities are shown for giant stars ($\log g \leqslant 3.5$ dex). The top images show the absolute values (same colour scale). The bottom left image shows the absolute differences, and the bottom right image the difference in terms of standard deviations of the mean radial velocities from the 30 random samples drawn from the Besançon Model. Note that for these plots the LSR from Coşkunoğlu *et al.* (2011), derived from the RAVE data, was adopted, as the LSR used in
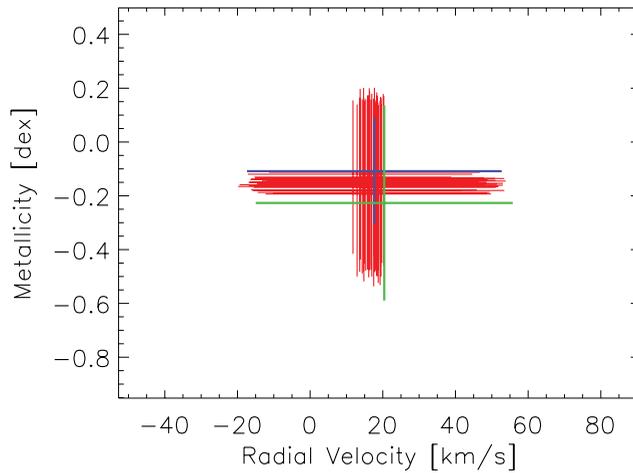
Figure 5. Plot showing the radial velocity and metallicity mean values and standard deviations of the RAVE stars (blue), all Besançon stars (green), and the random samples (red) for the swath $280° < l \leqslant 285°$, $25° < b \leqslant 30°$. (Coloured figures are available in the online version.)
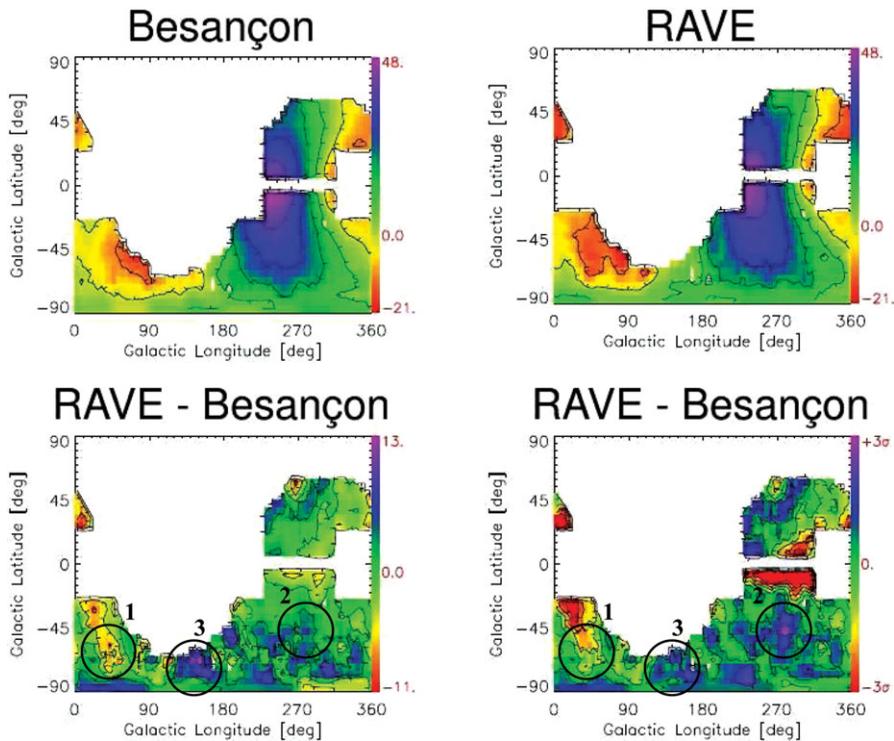
Figure 6. Final plots comparing the mean radial velocities for giant stars. Top images: absolute mean radial velocities for Besançon (left) and RAVE (right) with the same colour scale. Bottom left: absolute difference RAVE - Besançon. Bottom right: difference RAVE - Besançon in terms of standard deviations. All plots were smoothed once using a standard boxcar-smoothing algorithm (3 pixels wide) to reduce the noise. The circles mark individual structures as described in the text. (Coloured figures are available in the online version.)

the Besançon Model lead to a strong dipole in the previous comparison plots. Statistically significant structures in the bottom left plot show up much clearer in the bottom right plot. The yellow-orange structure in the lower left corner (circle 1) is the recently in the RAVE data discovered Stream of Aquarius (Williams *et al.* 2011). The blue-purple structure in the bottom right plot (circle 2), located at $l \approx 280°$, $b \approx -45°$ is the LMC, an example of a small absolute difference with high statistical significance. Other structures with large absolute differences turn out to be of low statistical significance. One example is the blue-purple structure in the lower left plot (circle 3) at $l \approx 145°$, $b \approx -65°$.

References

Coşkunoğlu, B., *et al.* 2011, *MNRAS*, 412, 1237
Robin, A. C., Reylé, C., Derriére, S., & Picaud, S. 2003, *A&A*, 409, 523
Seabroke, G. M. 2007, PhD thesis, University of Cambridge
Siebert, A., *et al.* 2011, *AJ*, 141, 187
Steinmetz, M. 2003, *ASP Conference Series*, 298, 381
Williams, M. E. K., *et al.* 2011, *ApJ*, 728, 102