DETERMINATION OF STAR FORMATION HISTORIES FROM GAIA-TYPE PHOTOMETRIC AND ASTROMETRIC SURVEY DATA

B.R. Jørgensen, L. Lindegren

Lund Observatory, Lund University, 22100 Lund, Sweden

ABSTRACT

To understand the evolution of the Milky Way Galaxy requires detailed knowledge of the star formation history of various populations. The vast amounts of photometric and astrometric data provided by the Gaia mission give unprecedented opportunities in this area. The relationships between the observed data and the ages of stars are however complex and highly non-linear and great care must be taken in analyzing the data. We describe a Bayesian approach to calculate the star formation rate (SFR) from astrophysical data, using a genetic algorithm to solve the basic integral equation. We present simulations showing that the method is capable of resolving structures in the SFR that cannot be seen from a distribution of the individually estimated stellar ages.

Key words: Stellar ages; Populations; Galaxy: evolution.

1. INTRODUCTION

Several different ways to determine the star formation rate (SFR) as a function of time have been used in the past. Three main groups of methods exist. One is to use the distribution of individual ages, e.g., Rocha-Pinto et al. (2000). Another is the parametric method, e.g., Bertelli & Nasi (2001), where the problem is reduced to determining certain parameters of the SFR, such as slope and level. The third method is to use colour-magnitude diagram (CMD) inversion, e.g., Dolphin (2002), or more general inverse methods, e.g., Hernandez et al. (1999, 2000); Vergely et al. (2002). This is the approach used here.

We assume that astrophysical parameters such as M_V, $\log T_{\rm eff}$ and [Me/H] are available for a sample of stars representing the population for which the SFR shall be derived. We also assume that stellar evolutionary and atmosphere models are available (e.g., in the form of isochrone sets with derived magnitudes) from which the astrophysical parameters can be predicted as a function of the fundamental stellar parameters: initial mass, initial chemical composition, and age (τ).

2. INDIVIDUAL STELLAR AGES

We start by calculating the relative posterior probability density function of the age, $G(\tau)$, for the individual stars. This is done by applying Bayes' theorem to the observed data and theoretical isochrones, combined with a flat prior density for the SFR and $\log Z$ distribution, and a standard initial mass function. The G function specifies the relative probability that a certain age τ can explain the observed data, in view of the assumed priors, and thus conveniently summarizes the available age information on that star. In particular, the location of the maximum of $G(\tau)$ is an estimate of the age of the star, and the width of the function indicates the uncertainty of that estimate. For a detailed discussion of the determination of individual stellar ages we refer to Jørgensen & Lindegren (submitted to A&A).

3. ESTIMATING THE SFR

3.1. The Distribution of Individual Age Estimates

The most straightforward way to estimate the SFR is to plot a histogram of the individual ages derived as explained above. It turns out, however, that this is not a good estimate of the SFR. The histogram is smeared by the uncertainties of the individual age estimates, which are often much greater than the desired resolution of the SFR. If, on the other hand, only ages with a small relative error are used, then the sample tends to become small and selection effects distort the resulting SFR. The distortion comes from the fact that good ages are more easily determined in certain age intervals due to the very non-linear mapping between observational data and age. For instance, for slightly evolved stars the isochrones between 4 and 5 Gyr cross several times, resulting in uncertain and/or biased age estimates whenever the data fall in that part of the HR diagram. As a result, ages in this interval tend to be under-represented when only the most precise ages are selected. It is clear that in order to take into account all the available information from the sample, we must use more advanced methods which directly estimate the SFR without necessarily assigning a specific age to each individual.

3.2. Combining G Functions

A better way to determine the SFR is to statistically combine all the individual G functions. Following Hernandez et al. (1999) we construct the likelihood L that a given star formation rate $R(\tau)$ can explain the complete set of observational data, as summarized by the G functions:

$$L = \prod_{i=1}^{n} \left(\int_0^{\tau_{\max}} R(\tau) G_i(\tau) \mathrm{d}\tau \right) \tag{1}$$

Here, i is the index of the star. The task is then to find the star formation history that maximizes this likelihood. Unfortunately this is a so-called ill-posed problem, which means that there is an infinite number of different solutions that are nearly equally good. To resolve this ambiguity we look for smooth solutions by maximizing the regularized expression:

$$\log L - \alpha \times \int_0^{\tau_{\max}} \left(\frac{d^2 R(\tau)}{d\tau^2} \right)^2 \mathrm{d}\tau \tag{2}$$

The parameter α determines the degree of smoothness. To find the maximum we use a genetic algorithm, which we find to be quite efficient.

4. EXAMPLES

To illustrate the power of this method we show here the results of trying to recover the star formation histories of two artificial samples, which look very alike in an HR diagram, but actually have very different underlying histories. Both samples have 1200 stars, but the first one (Figure 1) has a constant SFR over the past 10 Gyr; the second one (Figure 2) has three bursts distributed over the same period. Each burst lasts for 2 Gyr with constant SFR within the burst.

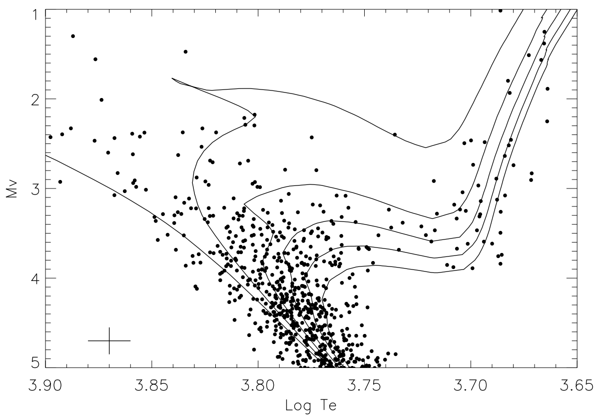

Figure 1. HR diagram for an artificial sample generated with a constant SFR over the last 10 Gyr.

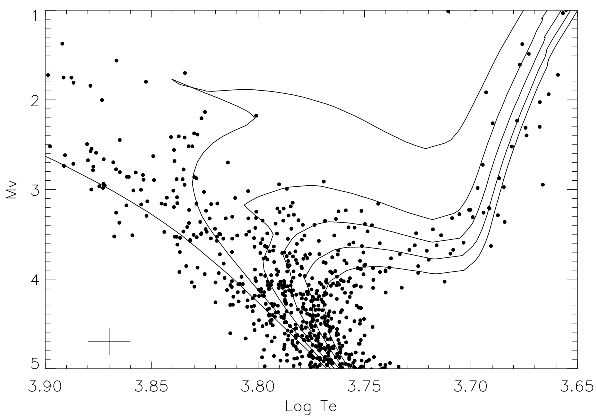

Figure 2. HR diagram for an artificial sample generated with an on-off SFR (constant SFR in the three age intervals 0–2, 4–6, 8–10 Gyr, zero in between).

4.1. Assumptions

We assume that relevant astrophysical quantities can be determined for individual stars from the astrometric, photometric and spectroscopic data. Specifically, we assume that $\log T_{\mathrm{eff}}$ can be determined to ± 0.01 dex, M_V to ± 0.1 mag and [Me/H] to ± 0.1 dex. We use Padova isochrones (Girardi et al. 2000) to establish the relationship between these data and the stellar evolutionary model parameters m (initial mass), Z (initial metallicity), and τ (age). Most generally, the problem is to estimate the joint distribution of (τ, Z, m) for a stellar population based on the set of observed data. However, in this contribution we focus on the distribution of ages, i.e., the star formation rate $R(\tau)$.

4.2. Results

Histograms of the individually estimated ages, shown in Figures 3 and 4, do not reveal any significant difference between the two samples. The histograms are fairly similar, mimicking the similarity of the two HR diagrams, although the sudden decrease at 2 Gyr in Figure 4 may be a real effect. However, in both diagrams there are significant distortions that may be explained in terms of the morphology of the isochrones, in particular the peak at 3.5 Gyr and depression around 4–5 Gyr.

However, when the star formation rates are estimated as described in Section 3.2, we find a very clear difference between the two samples (Figures 5 and 6). Over a wide range of values for the smoothing parameter α we recover star formation histories that are clearly distinct in the two cases and similar to the true histories. The choice of optimal value for α is non-trivial, but fortunately uncritical: in these examples we see a clear progression from 'oversolved' to 'oversmoothed' solutions over a range of four orders of magnitude in α. It it clear that the optimum must lie between these two extremes, but the precise choice will not affect the results much.

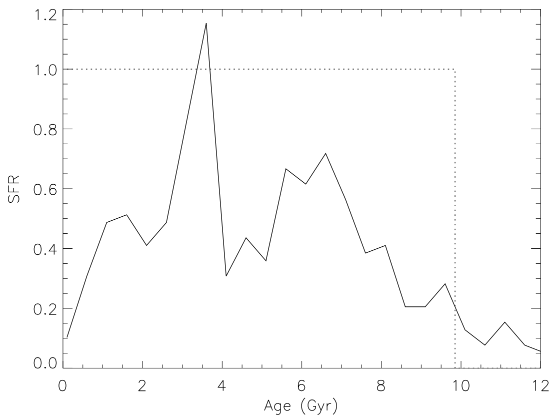

Figure 3. Histogram of individually estimated ages for the artificial sample in Figure 1. The dotted line is the actual SFR.

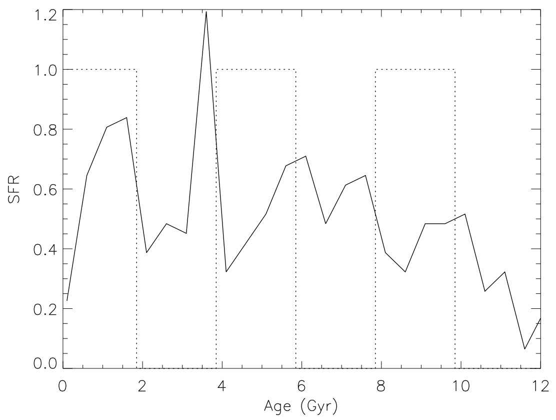

Figure 4. Histogram of individually estimated ages for the artificial sample in Figure 2. The dotted line is the true SFR.

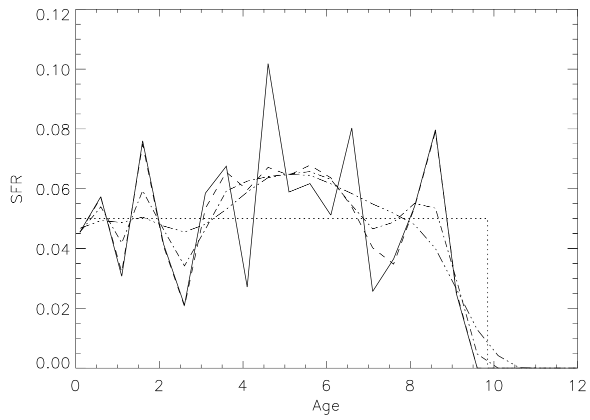

Figure 5. Recovered SFR with α varying from 10 (solid line) to 10 000 (dash-dot-dotted line), for the artificial sample in Figure 1.

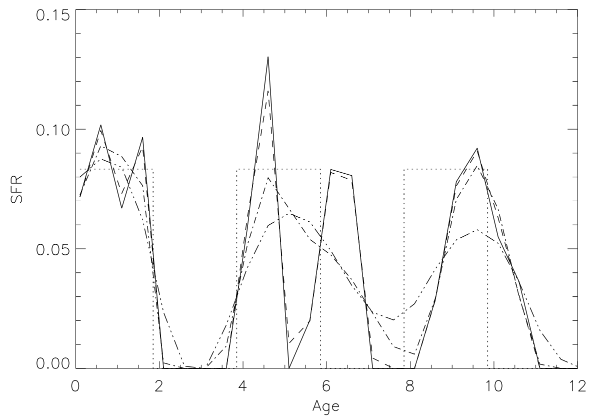

Figure 6. Same as Figure 5, but for the artificial sample in Figure 2.

5. CONCLUSIONS

For quite modest-sized samples ($\sim$ 1000 stars) it is possible, using the Bayesian technique with regularization, to resolve structures in the star formation history that are not easily revealed from an analysis of the individually determined ages. Inversion techniques are powerful to discover such structures, which may be difficult to find by other techniques such as CMD matching or the parametric approach. However, the complex morphology of stellar isochrones and a multitude of bias and selection effects could easily distort the results, which should therefore always be checked by careful simulation. The Bayesian approach should however be well suited for exploratory analysis of the large samples of astrometric and photometric data that Gaia will generate.

REFERENCES

Bertelli, B., & Nasi, E., 2001, AJ 121, 1013

Dolphin, A.E., 2002, MNRAS 332, 91

Girardi, L., Bressan, A., Bertelli, G., & Chiosi, C., 2000, A&AS 141, 371

Hernandez, X., Valls–Gabaud, D., & Gilmore, G., 1999, MNRAS 304, 705

Hernandez, X., Valls–Gabaud, D., & Gilmore, G., 2000, MNRAS 316, 605

Rocha-Pinto, H.J., Scalo, J., Maciel, W.J., & Flynn, C., 2000, A&A 358, 869

Vergely, J.-L., Köppen, J., Egret, D., & Bienaymé, O., 2002, A&A 390, 917