

# Determination of stellar ages from isochrones: Bayesian estimation versus isochrone fitting

B. R. Jørgensen and L. Lindegren

Lund Observatory, Lund University, Box 43, 221 00 Lund, Sweden  
e-mail: [bjarne;lennart]@astro.lu.se

Received 15 October 2004 / Accepted 19 February 2005

**Abstract.** We present a new method, using Bayesian estimation, to determine stellar ages and their uncertainties from observational data and theoretical isochrones. The result for an individual star is obtained as the relative posterior probability density as function of the age (“ $G$  function”). From this can be derived the most probable age and confidence intervals. The convoluted morphology of isochrones and strong non-linearities make the age determination by any method difficult and susceptible to statistical biases, and as a result age uncertainties have often been underestimated in the literature. From simulations we find that the  $G$  functions provide a general, robust and reliable way to quantify age information. Resulting age estimates are at least as accurate as those obtained with conventional isochrone fitting methods, and in some cases much better, especially when the observational uncertainties are large. We also find that undetected binaries, on the whole, have a surprisingly small effect on the age determinations. For a stellar sample, the individual  $G$  functions can be combined to derive the star formation history of the population; this will be developed in a forthcoming paper. For a coeval population the combination simplifies to computing the product of the individual  $G$  functions, and we apply that method to estimate the ages of the two open clusters IC 4651 and M 67, using Padova isochrones and photometric data from the literature. For IC 4651 we find an estimated age of  $1.56 \pm 0.03$  Gyr, assuming a true distance modulus of 9.80. For M 67 we find  $4.05 \pm 0.05$  Gyr for true distance modulus 9.48. The small formal errors of these age estimates do not include the (much larger) uncertainties from calibration and model errors, but illustrate the statistical power of combining  $G$  functions. Our statistical approach to the age determination problem is well suited for the mass treatment of data resulting from large-scale surveys such as the Gaia mission.

**Key words.** stars: fundamental parameters – stars: evolution – solar neighbourhood – methods: data analysis – methods: statistical

## 1. Introduction

Determination of stellar ages is crucial in observational studies aiming to understand stellar and galactic evolution (e.g., Fuhrmann 2004). The use of theoretical stellar evolutionary sequences, or equivalently isochrone fitting, is one of the few methods available for this purpose, and possibly the only one that can be applied on a large scale to the full range of stellar ages. When studying evolution on galactic timescales (e.g., the star formation history and age–metallicity relations of the solar neighbourhood), or the ages of individual field stars, the method is mainly applied to late A-type, F and G dwarfs, whose main-sequence life times range from below 1 Gyr up to the age of the universe. Examples of such studies include Edvardsson et al. (1993), Gómez et al. (1997), Ng & Bertelli (1998), Lachaume et al. (1999), Chen et al. (2000), Liu & Chaboyer (2000), Feltzing et al. (2001), Ibukiyama & Arimoto (2002), Lastennet & Valls-Gabaud (2002), Laws et al. (2003), Reddy et al. (2003), Pont & Eyer (2004), and Nordström et al. (2004).

In spite of the considerable practical importance of ages determined from isochrones, relatively little attention has been

given to quantifying their statistical reliability in view of the observational uncertainties. Although it is recognized (e.g., Lachaume et al. 1999) that the uncertainties of isochrone ages are often underestimated in the literature, there has been little systematic study of the reasons for this phenomenon, or of the possible statistical biases introduced by the estimation procedure. Aspects of the problem were recently discussed by Pont & Eyer (2004) from the viewpoint of Bayesian probability theory.

In the present paper we use Bayesian theory to derive stellar ages based on a comparison of observed data with theoretical isochrones. The basic assumptions are given in Sect. 2, while in Sect. 3 we derive the relative posterior probability density  $G$  as function of the age  $\tau$ . Under the given assumptions the  $G$  function<sup>1</sup> specifies the available age information on an individual star. From this it is possible to assign a unique (most probable) age to the star, as well as confidence intervals (CI) at any given confidence level, but for many applications the complete  $G$  function is required rather than a single value or interval for

<sup>1</sup> The “ $G$ ” notation was adopted from Hernandez et al. (2000), see Sect. 6.1.

the age. The effects of binaries on the age estimates are also studied (Sect. 3.8).

In Sect. 4 we contrast the Bayesian age estimates with those obtained using classical methods of isochrone fitting, where the uncertainties may be obtained by classical error propagation or Monte Carlo techniques. The properties of the estimators are studied by means of numerical simulations.

Section 5 describes the application of the method to the determination of the ages of nearby field stars, in particular the sample of more than 13 000 F and G dwarfs by Nordström et al. (2004). Moreover, we derive the ages of the two open clusters IC 4651 and M 67 by statistically combining the  $G$  functions for member stars. Finally, in Sect. 6, we discuss the present method in relation to some other methods described in the literature, as well as some of the limitations implied by the adopted model.

## 2. Problem formulation and assumptions

### 2.1. Model parameters and observational data

Determination of ages from isochrones depends on comparing observed data with theoretically computed values obtained from stellar models, where age ( $\tau$ ) is one of the free parameters, but not the only one: at least the initial stellar mass ( $m$ ) and initial chemical composition ( $Z$  and possibly more parameters) must also be considered. Quite generally we may summarize the relevant model parameters in a vector  $\mathbf{p}$ .

The observational data for a given object are similarly collected in a vector  $\mathbf{q}$  to which observational uncertainties are attached. Since isochrone fitting is traditionally linked to the HR or colour–magnitude diagram, the observational data are often taken to be quantities associated with the axes of the HR diagram, such as  $\log T_{\text{eff}}$  and  $M_V$ . These are of course derived from more basic photometric, spectroscopic and astrometric data and therefore, strictly speaking, not directly observed quantities; however, in the present context they can be regarded as such.

The theoretical models provide a mapping from the parameter space  $\mathbf{p}$  to the data space  $\mathbf{q}$ . Determination of stellar ages is a special case of the inverse problem, i.e., to find a mapping from  $\mathbf{q}$  to  $\mathbf{p}$ , which is possible if  $\dim(\mathbf{q}) \geq \dim(\mathbf{p})$  and the mapping is non-degenerate.

In the following, the model parameters are taken to be the age  $\tau$ , initial mass  $m$ , and initial metallicity, for which we introduce the parameter  $\zeta = \log(Z/Z_{\odot})$  (with  $Z_{\odot} = 0.019$ ). The helium content ( $Y$ ) and the mixture of heavier elements are regarded as known functions of  $Z$ , so that the stellar evolutionary models are uniquely labelled by the parameter vector  $\mathbf{p} = (\tau, \zeta, m)$ . Similarly, we usually take the observational data to be  $\mathbf{q} = ([\text{Me}/\text{H}], \log T_{\text{eff}}, M_V)$ , although an application using the colour index  $(V - I)_0$  instead of  $\log T_{\text{eff}}$  is given in Sect. 5.2.2. However, it should be noted that the estimation method described in Sect. 3 in no way is limited to this particular choice of variables, but is generally applicable as long as the model provides a unique mapping from  $\mathbf{p}$  to  $\mathbf{q}$ .

The age determination can thus be understood as the problem to invert the function  $\mathbf{q}(\tau, \zeta, m)$ , and in particular to calculate  $\tau(\mathbf{q})$ . The problem can basically be approached either by

direct numerical inversion (isochrone fitting), or using probabilistic methods (Bayesian estimation). We discuss variants of both approaches in Sects. 3 and 4.

### 2.2. Choice of stellar isochrone data

The function  $\mathbf{q}(\tau, \zeta, m)$  is in practice realized by numerical interpolation in a set of pre-computed stellar evolutionary tracks or isochrones. We have chosen to use the the Padova evolutionary models by Girardi et al. (2000), including convective overshooting, for all numerical examples and simulations in this paper. Although any of the recent sets of models (e.g., the Geneva models, Mowlavi et al. 1998; or the Yonsei-Yale models, Yi et al. 2003) could in principle equally well have been used to illustrate and compare age determination methods, our choice was mainly guided by convenience: the Padova models cover sufficient ranges in parameter space with a reasonably dense grid for realistic test cases, and broad-band magnitudes are provided for every model.

The Padova isochrones span the age interval  $0.063 \lesssim \tau \lesssim 17.8$  Gyr, the metallicity range  $-1.68 \lesssim \zeta \lesssim +0.20$ , and initial masses  $0.15 M_{\odot} \leq m \leq 7 M_{\odot}$ . The applications require that the observational quantities ( $\mathbf{q}$ ) can be computed for any valid combination of  $(\tau, \zeta, m)$  within these ranges.

### 2.3. Computing interpolated isochrones

A special interpolation method was devised to calculate  $\mathbf{q}(\tau, \zeta, m)$  while ensuring that no significant artifacts are introduced by the extremely non-linear relation between  $m$  and the quantities along any isochrone. Briefly, we defined a continuous parameter  $u$  along each isochrone by identifying a set of equivalent evolutionary points: the foot of the isochrone at  $m = 0.15 M_{\odot}$  ( $u = 0$ ), the core H exhaustion ( $u = 1$ ), the point of maximum  $T_{\text{eff}}$  before the RGB ( $u = 2$ ), the start of the RGB ( $u = 3$ ), etc. Between these points  $u$  was taken to be proportional to the curve length of the isochrone, using the (somewhat arbitrary) line element  $ds^2 = dM_V^2 + (20d \log T_{\text{eff}})^2$ . Linear interpolation was then carried out in the  $(\log \tau, \zeta, u)$  space with  $m$  regarded as a dependent variable like  $\log T_{\text{eff}}$  and  $M_V$ . This procedure ensures that the general morphology of the isochrones is preserved so that, for example, the interpolated isochrones do not intersect where they should not do so.

### 2.4. Metallicity parameter

The observational quantity  $[\text{Me}/\text{H}]$  is theoretically related to  $\zeta$  through the formula:

$$\begin{aligned} [\text{Me}/\text{H}] &= \log(Z/Z_{\odot}) - \log(X/X_{\odot}) \\ &= \zeta - \log\left[1 - k(10^{\zeta} - 1)\right] \end{aligned} \quad (1)$$

assuming that Me represents the same mixture of elements (weighted by atomic weight) as  $Z$ . Here  $k = (Z_{\odot}/X_{\odot})(1 + \Delta Y/\Delta Z) \simeq 0.094$  for  $X_{\odot} = 0.708$ ,  $Z_{\odot} = 0.019$ , and  $\Delta Y/\Delta Z = 2.5$  (e.g., Pagel & Portinari 1998). For  $|\zeta| \ll 1$  we have  $[\text{Me}/\text{H}] \simeq 1.094\zeta$  while for  $\zeta \lesssim -1$  there is an almost constant offset  $[\text{Me}/\text{H}] \simeq \zeta - 0.04$ . The actual relation between  $\zeta$

and the observational quantity is much more complicated because the relative abundances of elements in  $Z$  may vary (in particular the  $\alpha$  elements versus other elements). These complications are however beyond the scope of this paper and we simply assume

$$[\text{Me}/\text{H}] \simeq \zeta. \quad (2)$$

This is a reasonable approximation for the present purpose of exploring the methodology of age determination by isochrones. In real applications, considering that typical observational uncertainties in  $[\text{Me}/\text{H}]$  are (at least) of order  $\pm 0.1$  dex, this relation is still useful for stars of about solar metallicity, but should not be used for metal-poor stars.

### 2.5. Assumed observational errors

In simulations and for illustration purposes we usually assume the following standard errors, which are fairly representative of what can be achieved for F and G dwarfs in the solar neighbourhood (Nordström et al. 2004):  $\pm 0.1$  dex in  $[\text{Me}/\text{H}]$ ,  $\pm 0.01$  dex in  $\log T_{\text{eff}}$ , and  $\pm 0.1$  mag in  $M_V$ . We call these *nominal errors*. However, we shall sometimes consider observational errors that are half or twice the nominal ones.

## 3. Bayesian estimation of stellar ages

### 3.1. The posterior probability density function

In Bayesian estimation the parameters to be determined (in this case  $\tau$ ,  $\zeta$  and  $m$ ) are regarded as random variables and their (posterior) joint probability density function is given by

$$f(\tau, \zeta, m) \propto f_0(\tau, \zeta, m) L(\tau, \zeta, m) \quad (3)$$

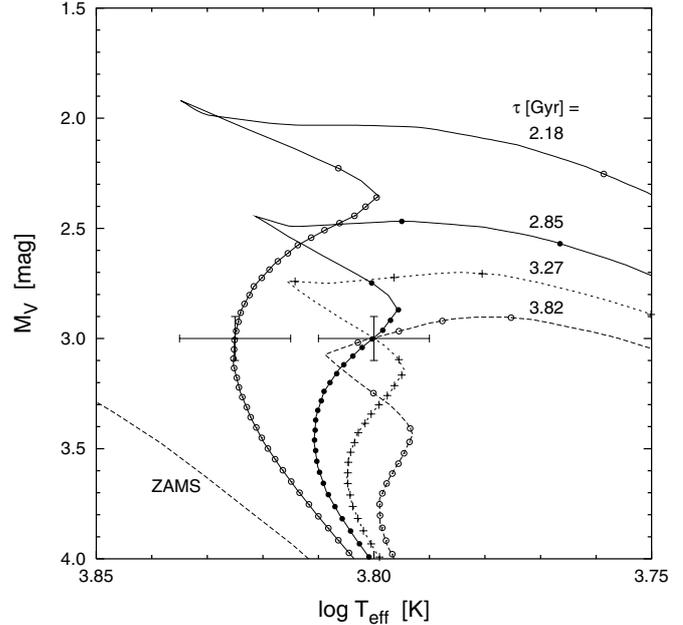
where  $f_0$  is the prior probability density of the parameters (Sect. 3.3) and  $L$  the likelihood function (Sect. 3.2). The probability density function (pdf)  $f$  is defined such that  $f(\tau, \zeta, m) d\tau d\zeta dm$  is the fraction of stars with ages between  $\tau$  and  $\tau + d\tau$ , metallicities between  $\zeta$  and  $\zeta + d\zeta$ , and initial masses between  $m$  and  $m + dm$ . The constant of proportionality in Eq. (3) must be chosen to make  $\int \int \int f(\tau, \zeta, m) d\tau d\zeta dm = 1$ .

Integrating  $f$  with respect to  $m$  gives  $f(\tau, \zeta)$ , which is the posterior joint pdf of  $\tau$  and  $\zeta$ . This function is of interest when studying the age–metallicity relation, since it summarizes the available information concerning these two parameters, given the observational data. Integrating once more with respect to  $\zeta$  gives  $f(\tau)$ , the posterior pdf of  $\tau$ , which similarly summarizes the available information concerning the age of the star.

In the next sections we discuss in some detail the two functions  $L$  and  $f_0$  that together define the posterior probability density.

### 3.2. The likelihood function

The likelihood function  $L$  equals the probability of getting the observed data  $\mathbf{q}$  for given parameters  $\mathbf{p}$ . Regarded as a function of  $\mathbf{p}$  it is not a pdf – for instance, its integral may be infinite. We assume independent Gaussian observational errors in each



**Fig. 1.** HR diagram showing the location of two hypothetical observations at  $(\log T_{\text{eff}}, M_V) = (3.825, 3.0)$  and  $(3.800, 3.0)$ , with nominal uncertainties as in Sect. 2.5 (error bars show  $\pm 1\sigma$ ). The zero-age main sequence (ZAMS) and selected isochrones for  $\log(Z/Z_{\odot}) = -0.2$  ( $\simeq [\text{Me}/\text{H}]$ ) are also shown. The symbols along the isochrones show where the initial mass is a multiple of  $0.01 M_{\odot}$ . See text for further explanation.

of the  $n = \dim(\mathbf{q})$  observed quantities, with standard errors  $\sigma_i$ . The likelihood function is then

$$L(\tau, \zeta, m) = \left( \prod_{i=1}^n \frac{1}{(2\pi)^{1/2} \sigma_i} \right) \times \exp(-\chi^2/2) \quad (4)$$

where

$$\chi^2 = \sum_{i=1}^n \left( \frac{q_i^{\text{obs}} - q_i(\tau, \zeta, m)}{\sigma_i} \right)^2. \quad (5)$$

A maximum-likelihood (ML) estimate of the stellar parameters  $(\tau, \zeta, m)$  may be obtained by finding the maximum of this function, which (in the case of Gaussian errors) is equivalent to minimizing  $\chi^2$ . ML estimators are in general good estimators and in many cases nearly optimal (Casella & Berger 1990), so why not simply adopt the ML age estimate?

The difficulty with the ML estimate in the present case has to do with the complex morphology of the isochrones, i.e., of the highly non-linear mapping from  $\mathbf{p}$  to  $\mathbf{q}$ . Effectively, it means that more information is needed to make a good estimate of the age than provided by the likelihood function alone. An illustration is given in Fig. 1, where we consider the hypothetical observation of two isolated field stars. The observed data are depicted together with a selection of isochrones. All isochrones and data are for  $[\text{Me}/\text{H}] \simeq \zeta = -0.2$ .

For the left data point ( $\log T_{\text{eff}} = 3.825$ ,  $M_V = 3.0$ ) there is only one isochrone going exactly through the observed point, namely  $\tau = 2.18$  Gyr. This is then the best-fitting age in terms of the  $\chi^2$  in Eq. (5) and consequently also the ML estimate of the age based on the given data. For the second

point ( $\log T_{\text{eff}} = 3.800$ ,  $M_V = 3.0$ ), there are *three* distinct isochrones going exactly through the data point:  $\tau = 2.85$ , 3.27 and 3.82 Gyr. There are thus three points in parameter space where  $\chi^2 = 0$  and where consequently  $L(\mathbf{p})$  attains its global maximum. Which of the three solutions should be chosen? *In the absence of any additional information*, a reasonably experienced astronomer, forced to make a choice, would probably select the youngest isochrone (2.85 Gyr) as the most reasonable solution. Why? Because at that age the star would still be in the slowly evolving stage of core hydrogen burning. If any of the higher ages were adopted, we would have to assume that the star is in a more advanced and rapid evolutionary stage, which is less probable for a (normal) star picked at random<sup>2</sup>. This is an example where the astronomer, consciously or not, makes use of prior information. The Bayesian approach offers a systematic way to quantify and make use of this information by means of the prior pdf  $f_0$  in Eq. (3).

### 3.3. Choice of prior distributions

The prior density of the model parameters in Eq. (3) can be written

$$f_0(\tau, \zeta, m) = \psi(\tau)\phi(\zeta|\tau)\xi(m|\zeta, \tau) \quad (6)$$

where  $\psi(\tau)$  is the a priori star-formation rate (SFR) history,  $\phi(\zeta|\tau)$  the a priori metallicity distribution as function of age, and  $\xi(m|\zeta, \tau)$  the a priori initial mass function (IMF) as function of metallicity and age.

To proceed further it is necessary to make specific assumptions about the functions  $\psi$ ,  $\phi$  and  $\xi$ . Inevitably, the end result will to some extent depend on these assumptions. However, if the observations are good enough there will only be a weak such dependence. On the other hand, if the data are poor, it is important to make only the most non-committal assumptions concerning the prior statistics. For example, it makes sense to assume that  $\tau$ ,  $\zeta$  and  $m$  are independent, so that

$$f_0(\tau, \zeta, m) = \psi(\tau)\phi(\zeta)\xi(m). \quad (7)$$

A contrary assumption, e.g. that there is a priori a correlation between  $\tau$  and  $\zeta$ , would clearly be unreasonable if the objective of the study is to determine a possible age–metallicity relation. For a similar reason, we assume that the prior SFR  $\psi(\tau)$  is flat.

The choice of the prior metallicity distribution  $\phi(\zeta)$  is less obvious. On one hand, the actual metallicity distribution of nearby field stars is known to be strongly peaked at around  $-0.1 \pm 0.3$  dex (e.g., Nordström et al. 2004), which might indeed be a useful a priori assumption e.g. in the absence of any metallicity measurement at all for a given (field) star. If the true metallicity happens to be much different from this assumption, it would of course result in a biased age estimate, but that may still be preferable to no estimate at all. On the other hand, if a metallicity measurement *is* available, one would probably not

want to prejudice the age determination by making an additional assumption about what the metallicity *should* have been. Thus, provided a metallicity measurement of adequate precision is at hand, we think that a flat prior is generally the most reasonable choice. It could still be debated whether it should be flat in the logarithmic variable  $\zeta$  or in the linear variable  $Z$ ; we have chosen the former alternative (cf. below). For a more extensive discussion of the role and importance of the prior distributions we refer to Pont & Eyer (2004).

For the IMF, finally, there can be no doubt that underlying star formation processes generate a decreasing function at least for the relevant mass range ( $m \gtrsim 1 M_\odot$ ), for which we assume a power-law

$$\xi(m) \propto m^{-a}. \quad (8)$$

An exponent  $a$  of order 2 to 3 would be representative for the empirical IMF at around  $1 M_\odot$  (Kroupa et al. 1993). For the simulations described below we normally use  $a = 2.7$ .

### 3.4. Calculating the G function

With the assumptions from Sect. 3.3 we find, after inserting Eq. (6) into Eq. (3) and integrating with respect to  $m$  and  $\zeta$ , that the posterior pdf of  $\tau$  can be written

$$f(\tau) \propto \psi(\tau)G(\tau) \quad (9)$$

where

$$G(\tau) \propto \int \int L(\tau, \zeta, m)\xi(m) dm d\zeta. \quad (10)$$

For the IMF  $\xi(m)$  we use Eq. (8). We normalize (10) such that  $G(\tau) = 1$  at its maximum. Comparison with Eq. (3) suggests that  $G(\tau)$  can be interpreted as the relative likelihood of  $\tau$  after elimination of  $m$  and  $\zeta$ .

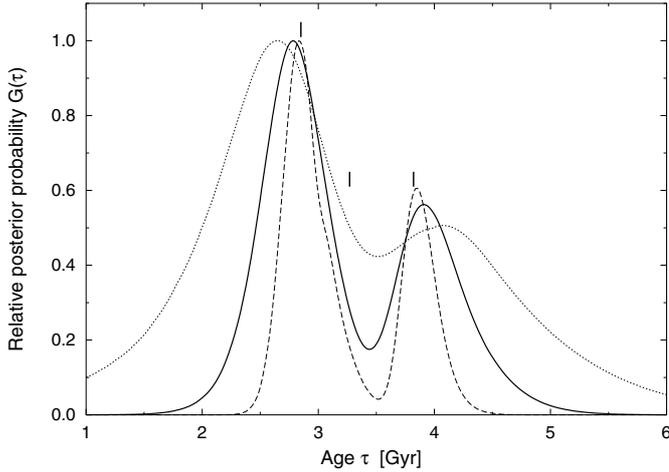
Numerically, we evaluate Eq. (10) for each age value ( $\tau_j$ ) as a double sum along a set of isochrones at the required age that are equidistant in metallicity ( $\zeta_k$ ). (In practice we use pre-computed isochrones for a step size of 0.04 dex in  $\zeta$ , and consider only those within  $\pm 3.5\sigma_{[\text{Me}/\text{H}]}$  of the observed metallicity.) Let  $m_{jkl}$  be the initial-mass values along each isochrone ( $\tau_j, \zeta_k$ ); then

$$G(\tau_j) \propto \sum_k \sum_\ell L(\tau_j, \zeta_k, m_{jkl})\xi(m_{jkl})(m_{jkl+1} - m_{jkl-1}). \quad (11)$$

Discretization and interpolation errors sometimes produce a slight raggedness of the computed  $G$  function which we eliminate by numerical smoothing. This is mainly needed in order to compute reliably the mode of the  $G$  function (Sect. 3.7).

The solid line in Fig. 2 shows the  $G$  function for the right data point in Fig. 1, assuming nominal observational errors (Sect. 2.5). The dashed and dotted curves show the  $G$  functions computed for the same data but assuming half or double the nominal errors. These functions have a global maximum around  $\tau = 2.85$  Gyr, corresponding to the most probable age according to the discussion of this example in Sect. 3.2. Around the alternative age 3.82 Gyr there are lower peaks, suggesting that this solution is less probable. Around 3.27 Gyr there is no visible peak at all.

<sup>2</sup> The assumption that no additional information is available is crucial: for example, if the star is located in a cluster whose HR diagram clearly outlines one of the older isochrones, one would not hesitate to assign the corresponding age also to this star.



**Fig. 2.** Plot of the relative posterior probability density,  $G(\tau)$ , for the observed data point ( $[\text{Me}/\text{H}] = -0.2$ ,  $\log T_{\text{eff}} = 3.800$ ,  $M_V = 3.0$ ). The three curves are for the nominal uncertainties (thick solid line), half the nominal values (dashed) and twice the nominal (dotted); see Sect. 2.5. The vertical bars indicate the ages for the three isochrones in Fig. 1 that go through the observed data point.

This behaviour of the  $G$  functions can be understood with reference to Fig. 1. Along each isochrone in that diagram we have marked off the models that are equidistant in initial mass ( $\Delta m = 0.01 M_{\odot}$ ). From the markings it is clear that the error ellipse around the observed point covers a much wider range of masses along the 2.85 Gyr isochrone than along the 3.82 Gyr isochrone, and a very small mass range for the intermediate isochrone. When performing the integration over  $m$  in Eq. (10) there is a correspondingly larger contribution to the  $G$  function at  $\tau = 2.85$  Gyr than at 3.82 Gyr, and a very small contribution at the intermediate age. The varying density of markings reflects the changing speed of stellar evolution. Thus, by performing the integration over  $m$  in Eq. (10) we automatically take into account the different evolutionary time scales.

Note that the assumed IMF has only a marginal impact on the shape of the  $G$  functions compared with the varying speed of stellar evolution. For example, changing the exponent  $a$  in Eq. (8) from 2.7 to 0 (i.e., a flat IMF) only reduces the secondary peak in Fig. 2 from 0.59 to 0.45 for the nominal errors. Similarly the assumed prior in  $\zeta$  is normally uncritical. For example, it could be argued that it is more reasonable to assume a flat prior density in  $Z = Z_{\odot} \times 10^{\zeta}$  than in the logarithmic parameter  $\zeta$ . Using a flat prior in  $Z$  implies  $\phi(\zeta) \propto 10^{\zeta}$ . Including this factor in the integrand of Eq. (10) would reduce the secondary peak in Fig. 2 from 0.59 to 0.51. The corresponding effects on the estimated age are on the 1% level for the nominal errors, but increase quickly with the observational errors.

### 3.5. Age determination from the $G$ function

Assuming a flat prior pdf in  $\tau$  we have  $f(\tau) \propto G(\tau)$  and the age determination can be entirely based on the  $G$  function. As discussed in Sect. 3.7 we will adopt the mode (maximum) of the  $G$  function,  $\hat{\tau}$ , as our “best” estimate of a stellar age, based exclusively on the given observational data.

It may happen that the observed data point falls well outside all the isochrones. In this case it is obviously not possible to infer anything about the age. Although the  $G$  function can still be computed and even may show a well-defined peak, it is clearly meaningless and should not be used. We adopt (somewhat arbitrarily) the criterion

$$\min_p \chi^2 > \chi_{0.99}^2 \quad (12)$$

for such a “no solution” condition. The left hand side is the minimum of Eq. (5) for *any* combination of model parameters, and  $\chi_{0.99}^2$  is the 99th percentile value of the chi-square distribution for  $\text{dim}(\mathbf{q})$  degrees of freedom – in our case  $\chi_{0.99}^2 = 11.3$  for 3 d.o.f. This means that there is less than 1% probability for a data point falling that far from the isochrones purely by chance, rather than for some technical or astrophysical reason (peculiar star, bad data, etc.).

### 3.6. Confidence intervals

In addition to the single age estimate  $\hat{\tau}$  it is often desirable to provide a confidence interval (CI) at some specified confidence level  $1 - \alpha$  (say, 68% or 95%). The meaning of the confidence level is that we expect the true age to be inside the CI in a fraction  $1 - \alpha$  of all the cases. The usual  $\pm 1\sigma$  is a CI at 68% confidence level if the errors are Gaussian.

Since  $G(\tau)$  is proportional to the posterior pdf of  $\tau$ , one could use the following equation to construct a confidence interval  $[\tau_1, \tau_2]$  for any given confidence level  $1 - \alpha$ :

$$\frac{\int_{\tau_1}^{\tau_2} G(\tau) d\tau}{\int_{\tau_{\min}}^{\tau_{\max}} G(\tau) d\tau} = 1 - \alpha. \quad (13)$$

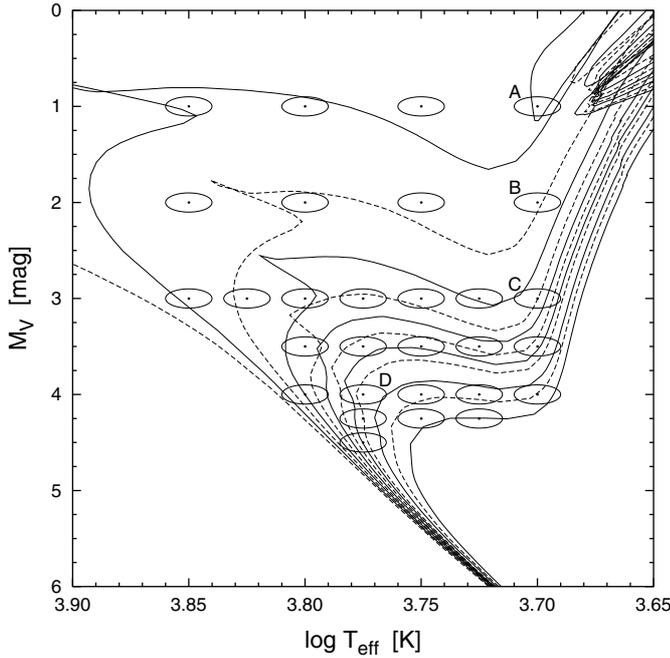
Here,  $\tau_{\min}$ ,  $\tau_{\max}$  are the extreme ages considered, e.g., as covered by the isochrones. The fractional area below  $\tau_1$  and above  $\tau_2$  need not be the same, but could be chosen to minimize the length of the confidence interval.

However, we have *not* adopted this method to compute a confidence interval, mainly for the following reason. Suppose that the data give very little information about the age so that the  $G$  function is essentially flat. Then any sub-interval of length  $(1 - \alpha)(\tau_{\max} - \tau_{\min})$  would be a formally correct confidence interval at the chosen confidence level, without carrying any useful information at all, but with an obvious risk of misinterpretation.

We have therefore adopted a different procedure to derive a CI from the  $G$  function. As suggested in connection with Eq. (10),  $G(\tau)$  can be regarded as the relative likelihood function with  $G(\hat{\tau}) = 1$  at the mode  $\hat{\tau}$ . For any other age,  $G(\tau)$  is therefore the likelihood ratio obtained by constraining one parameter ( $\tau$ ) to a given value. It is well known (e.g., Casella & Berger 1990) that the likelihood ratio, under certain regularity conditions, has a simple asymptotic distribution which in our case suggests that  $-2 \ln G(\tau)$  has a chi-square distribution with 1 degree of freedom. Thus,

$$1 - \alpha = \Pr[G(\tau) \geq G_{\text{lim}}] \approx 2\Phi\left(\sqrt{-2 \ln G_{\text{lim}}}\right) - 1 \quad (14)$$

where  $\Phi(x) = (2\pi)^{-1/2} \int_{-\infty}^x \exp(-t^2/2) dt$  is the standard normal distribution. The reasoning behind this expression is admittedly



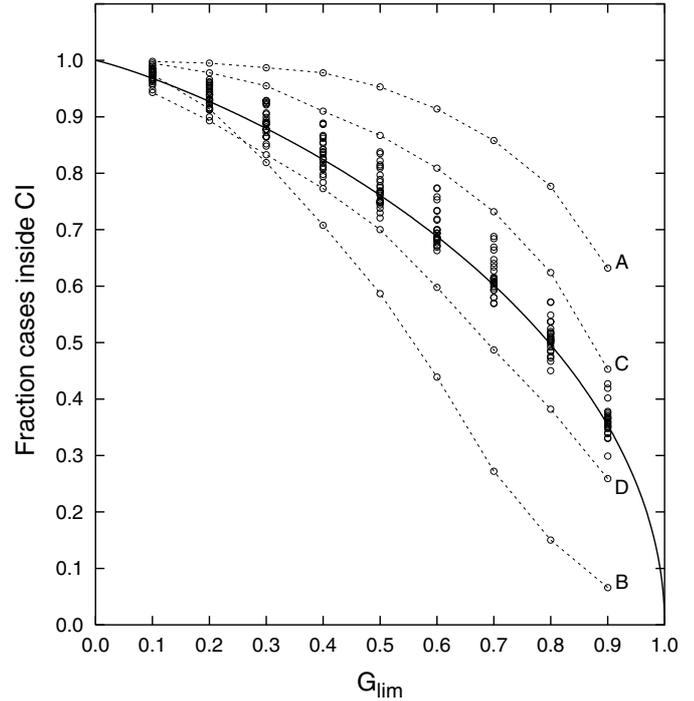
**Fig. 3.** Representative locations in the HR diagram used in the Monte Carlo experiments (Sect. 3.6) to test the validity of the approximation (14). The isochrones are for  $\zeta = -0.2$  and  $\tau = 0, 1, 2, \dots, 8, 10, 12$  and  $15$  Gyr. The points at the centres of the error ellipses indicate the true observational data, to which independent Gaussian errors are added with nominal errors. Note the behavior of the points marked A, B, C and D in Fig. 4.

weak, but we will substantiate the result below by means of numerical simulations.

Assuming that Eq. (14) holds, we compute for the given confidence level  $1 - \alpha$  the corresponding value  $G_{\text{lim}}$  and then define the CI  $[\tau_1, \tau_2]$  to be the shortest interval such that  $G(\tau) < G_{\text{lim}}$  outside the interval. (Since  $G(\tau)$  can have multiple maxima, it may happen that  $G(\tau)$  is locally below  $G_{\text{lim}}$  inside the interval as well.) The 68, 90 and 95% confidence levels correspond respectively to  $G_{\text{lim}} = 0.61, 0.26$  and  $0.15$ .

We have made extensive Monte Carlo simulations to test whether Eq. (14) holds to a reasonable accuracy in practical situations. For representative points in the HR diagram (Fig. 3) we generated “observed” data with nominal uncertainties (Sect. 2.5), computed the resulting  $G$  functions and derived the confidence intervals corresponding to  $G_{\text{lim}} = 0.1, 0.2, \dots, 0.9$ . We then counted the number of cases when the true age fell within the derived confidence interval. Figure 4 shows the fraction of such cases as function of  $G_{\text{lim}}$  for each set of 1000 experiments. The solid curve is the theoretically expected relation from Eq. (14). It is seen that the experiments follow the theoretical relation rather well for most of the selected points in the HR diagram.

Exceptions are the points marked A and B in Fig. 3, and to a smaller degree points C and D. The problem at points A and B can be understood as an effect of the multiply overlapping isochrones on the upper part of the giant branch, especially when also the uncertainty in metallicity is considered, while it is not so obvious what happens at C and D. Nevertheless, we conclude that the simple recipe of using a fixed threshold



**Fig. 4.** Results of the Monte Carlo experiments described in Sect. 3.6. Each circle represents the statistics from 1000 experiments. The ordinate is the fraction of experiments in which the true age falls within the confidence interval defined by  $G_{\text{lim}}$ . The solid curve is the relation expected if  $-2 \ln G$  follows the chi-square distribution with 1 d.o.f., Eq. (14). The experiments follow the theoretical relation rather well except those for the points marked A, B, C and D in Fig. 3.

value  $G_{\text{lim}}$  is in general quite reasonable and should produce realistic confidence intervals under the given assumptions.

In the following we always use a confidence level of 68% (corresponding to the usual  $\pm 1\sigma$  for Gaussian errors), or  $G_{\text{lim}} \approx 0.6$ .

It will often happen that  $G(\tau)$  exceeds  $G_{\text{lim}}$  at one (or both) of the extreme ages considered. In that case the confidence interval is correspondingly truncated. Thus, if  $G(\tau_{\text{min}}) > G_{\text{lim}}$  we set  $\tau_1 = \tau_{\text{min}}$ , while if  $G(\tau_{\text{max}}) > G_{\text{lim}}$  we set  $\tau_2 = \tau_{\text{max}}$ . Effectively, this means that only an upper or lower bound to the age can be set at the chosen confidence level, or possibly no bound at all. In contrast, when the confidence interval is entirely within the considered range (i.e., when  $\tau_{\text{min}} < \tau_1, \tau_2 < \tau_{\text{max}}$ ), we call our age estimate *well-defined* (Nordström et al. 2004).

Having defined a confidence interval  $[\tau_1, \tau_2]$  it is also useful to have a measure of the *relative precision* ( $\epsilon$ ) of the age estimate. We define

$$\epsilon = \sqrt{\tau_2/\tau_1} - 1. \quad (15)$$

This definition was chosen in order to cope with the whole range of confidence intervals encountered in practice. For small  $\epsilon$  it agrees with any sensible representation of relative precision, e.g.,  $[\tau_1, \tau_2] = [0.99, 1.01]$  gives  $\epsilon \approx 0.01$ . However, confidence intervals like  $[0.5, 2.0]$  are much more typical for age determinations, and it is then not obvious if the relative precision should be expressed relative to the lower limit, the

upper limit, or some mid-point of the confidence interval. In this particular example our definition leads to a relative precision of 100% ( $\epsilon = 1.0$ ).

It should be noted that a large value of the relative precision, e.g.  $\epsilon > 1$ , does not imply the absence of age information; for example,  $[\tau_1, \tau_2] = [0, 1]$  Gyr corresponds to infinite  $\epsilon$  but would still provide a very meaningful upper limit on the age.

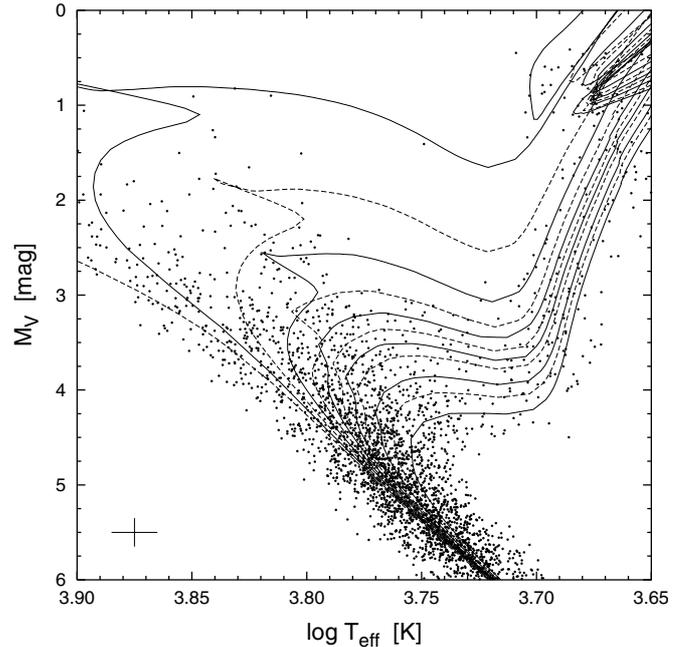
### 3.7. Defining the “best” age estimate from the $G$ function

As already emphasized, we need the complete  $G$  function to characterize the available age information on a given star. Compressing this to a single age estimate and confidence interval inevitably results in a loss of information. For example, in Fig. 2, knowing the location of the maximum and 68% confidence interval (i.e., where  $G > 0.6$ ) may be enough to characterize the main peak, but completely ignores the secondary maximum. When defining a “best” age estimate the goal should be to minimize the loss of relevant information in practical situations.

There are at least three natural choices for computing a unique age estimate from the  $G$  function: the mean value (centre of gravity of the area under the  $G$  curve), the median value (bisecting the area under the  $G$  curve), and the mode (the position of the global maximum in  $G$ ). We compared the behaviour of the three estimates through Monte Carlo simulations of (initially) 7080 single stars. A constant SFR was assumed, with a Kroupa et al. (1993) IMF in the range  $0.8\text{--}5.4 M_{\odot}$  and a constant metallicity of  $\zeta = -0.2$ . As it turns out, 4112 of the stars evolve out of the range of the isochrones, leaving a sample of 2968 stars. For these we generated observed data  $[\text{Me}/\text{H}]$ ,  $\log T_{\text{eff}}$  and  $M_V$  with errors of 0.1 dex, 0.01 dex and 0.15 mag, respectively; see Fig. 5. For each star we then calculated the  $G$  function and hence the age estimates corresponding to the mean, median and mode.

In Fig. 7 we compare the three age estimates with the true ages for the 2968 stars in the synthetic data set. It is seen that the mean and median behave similarly while the mode exhibits a different pattern. The difference between the mode and the other estimators is mainly due to stars with ages not well-defined in the sense of Sect. 3.6. These stars often have very broad  $G$  functions that are cut off at the minimum and/or maximum isochrone age, which tends to move the mean and median toward the center of the age interval. The mode, on the other hand, may be located anywhere in the age interval, but very often at the extreme ages (producing the “lines” at the top and bottom of the panel for the mode). Thus, while the mean and median superficially appear to be better estimates than the mode (in the sense that there are much fewer such points far away from the 1:1 line), they tend to be deviously biased towards the middle of the full age range.

Considering only the subset of 937 stars with well-defined ages, much of these differences between the three estimators disappear, and this is even more true for the 193 stars that in addition obtain a relative precision  $\epsilon \leq 0.2$ . The left panels in Figs. 8 and 9 show the mode versus true age for these



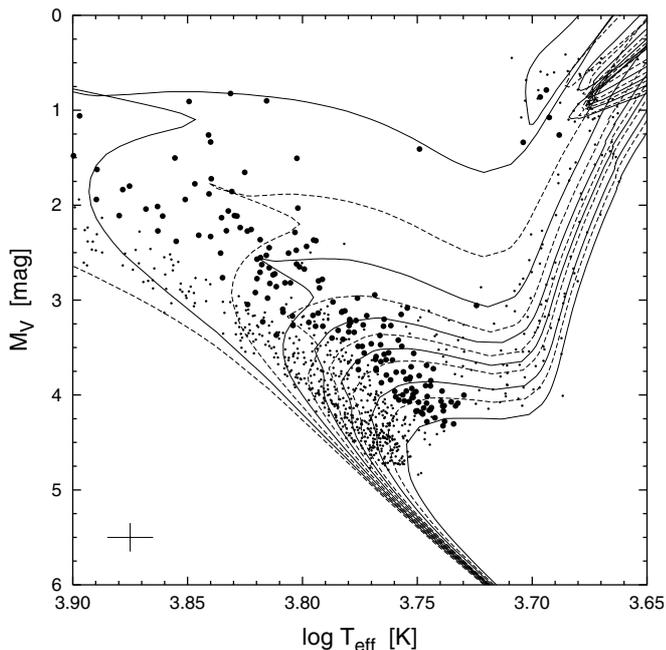
**Fig. 5.** HR diagram for the synthetic sample of 2968 stars generated as described in the text. The assumed observational uncertainties in  $\log T_{\text{eff}}$  and  $M_V$  are indicated in the lower-left corner (in addition there is a  $\pm 0.1$  dex uncertainty in metallicity). The isochrones are for 0, 1, 2, ..., 8, 10, 12 and 15 Gyr.

subsets; we do not show the rather similar plots for the mean and median.

Although it is not easily seen in the scatter plots, there is a slight bias towards lower ages in all the (sub-) samples and for all three estimators. To estimate the size of this bias we fitted a bisector to each scatter plot, i.e., a straight line through the origin dividing the stars into two equally large groups. For well-defined ages the bisector slope is between 0.93 and 0.95, depending on the estimator. In the  $\epsilon \leq 0.2$  case the bisector slope is between 0.95 and 0.98. This 2–7% bias towards lower ages is caused by the strongly asymmetric non-linearity of the mapping from  $(\tau, m)$  to  $(\log T_{\text{eff}}, M_V)$ . For any given position in the HR diagram above the main sequence, the younger isochrones in a region around the observed position cover a wider mass interval than the older isochrones on the opposite side (cf. Fig. 1). There are therefore more young  $(\tau, m)$  combinations that can “explain” the observed data than old combinations.

When a single value is required for the estimated age, we strongly recommend to use the mode rather than the mean or median, because it avoids the tendency of the latter to assign an age in the middle of the permitted range. By contrast, the mode tends to assign an extreme age in such cases, which clearly signals a bad estimate. Furthermore, when a confidence interval is computed as in Sect. 3.6, the mode will always be within that confidence interval, while this (highly desirable!) property cannot be guaranteed for the mean or median.

Figure 6 is an HR diagram for the 937 stars in the synthetic sample for which the ages are well-defined. Comparison with the full synthetic sample in Fig. 5 shows that well-defined ages are normally obtained for data points that are at least 1–2 standard deviations above the isochrones for the most extreme ages



**Fig. 6.** Same as Fig. 5 but showing only the data for the 937 stars with well-defined ages according to the criterion in Sect. 3.6. Large dots show the subset having a relative precision ( $\epsilon$ ) in age better than 20%.

considered. A relative precision better than 20% (large dots in Fig. 6) is mainly achieved in a band some 1–2 mag above the zero-age main sequence.

### 3.8. Influence of undetected binaries

Thus far we have always assumed that we are dealing with single stars. Binary (and multiple) stars are however the norm rather than the exception, and it is important to quantify how this affects the age determinations. Known binary systems can be removed from a sample or dealt with in other manners, but a certain fraction of undetected binary systems is unavoidable in most samples and could pose a problem.

An unresolved secondary affects the observed parameters of a star in different ways depending on the mass ratio ( $q = m_2/m_1 \leq 1$ ) of the components and their evolutionary status. On or near the main sequence, a small  $q$  means that the primary completely dominates the light and there will be no perceptible change in the observational data. For systems with comparable masses the effects may be quite large, and in the worst case, when  $q = 1$ , the  $M_V$  of a binary system is 0.75 mag brighter than the corresponding single star. This usually leads to a (sometimes considerable) overestimation of the age.

To examine the overall statistical effect of unrecognized binaries on the age determination of a stellar sample we used the synthetic sample from Sect. 3.7 and added a random secondary to each of the stars. Two cases were considered for the mass-ratio distribution: in the first case we assumed a decreasing pdf  $f(q) = 1.5 - q$ , somewhat similar to the empirical mass-ratio distribution found by Duquennoy & Mayor (1991) for field F, G and K dwarfs. In the second (worst-case) scenario every star was given an equal-mass companion. For each

system we calculated the total  $M_V$  and obtained an effective  $T_{\text{eff}}$  from the luminosity-weighted mean of  $T_{\text{eff}}^{-4}$  for the two components. New age estimates were then computed for each system as for a single star.

The middle and right panels in Figs. 8–9 show the results of the age determinations (using the mode) for the two cases of a decreasing mass-ratio distribution and equal-mass components. Figure 10 is a histogram of the shift in the estimated age caused by the presence of the secondary. As expected, the addition of secondary components usually causes the ages to be overestimated, but the overall statistical effect is surprisingly small, in particular for the decreasing mass-ratio distribution (note the logarithmic scale of Fig. 10!). This can perhaps be attributed to the fact that most of the stars for which useful ages can be determined are located near the main-sequence turn-off point (cf. Fig. 6), where the age estimate is relatively insensitive to a shift in  $M_V$ . In fact, for the high-precision sample ( $\epsilon \leq 0.2$ , Fig. 9, right panel, and the big dots in Fig. 6) the binarity even tends to make the ages seem a bit too small. The bisector slope in the middle and right panels of Fig. 9 is 0.97 and 1.13, respectively.

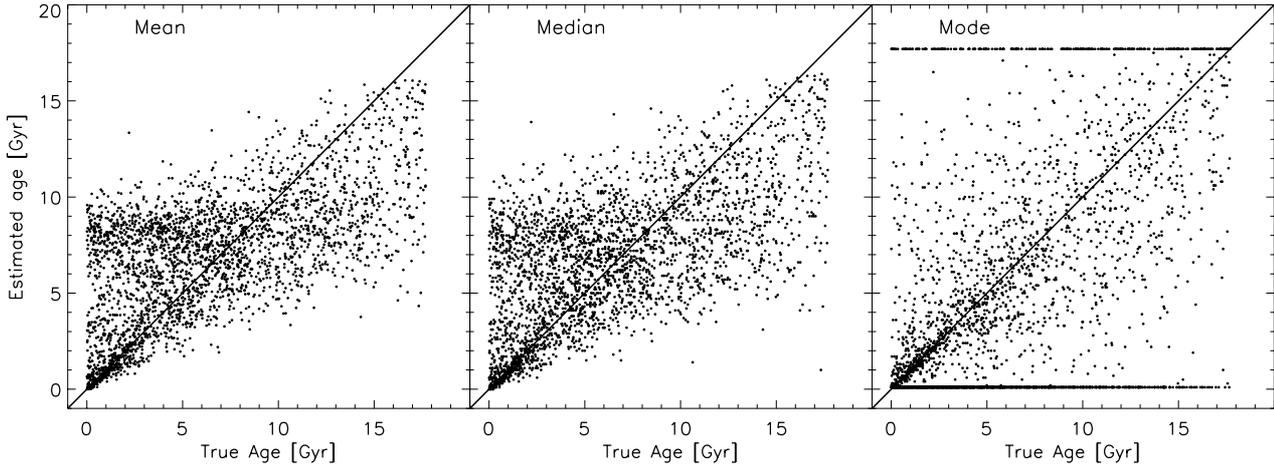
We conclude that the overall effect of (undetected) binaries in a large stellar sample is rather modest, compared with the typical statistical uncertainties of isochrone ages, unless there is a high fraction of equal-mass binaries in the sample. However, any one single age determination may be very strongly affected by the duplicity of that star.

## 4. Comparison with age estimates from conventional isochrone fitting

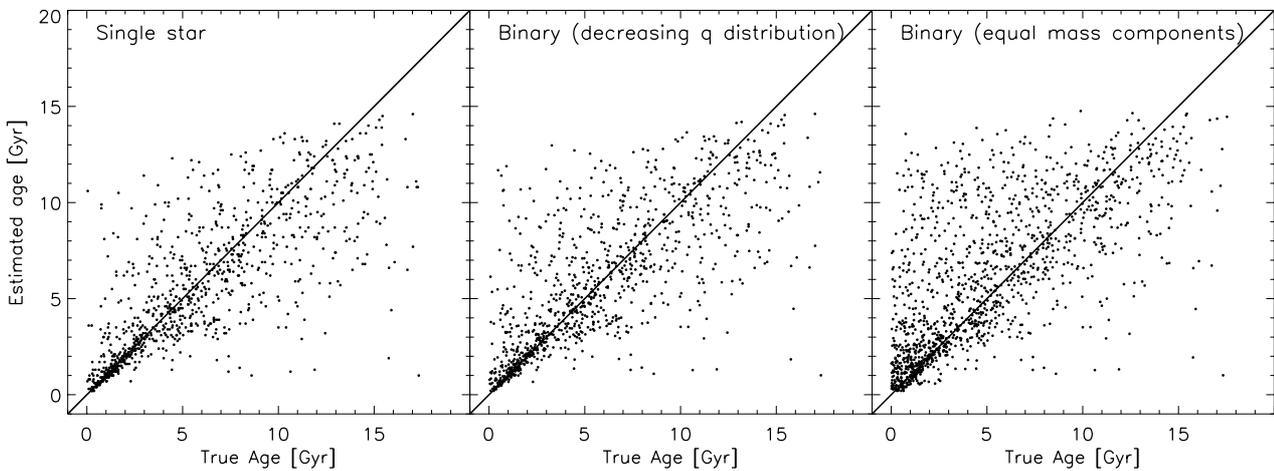
In this section we compare the Bayesian method of Sect. 3 with more “conventional” methods, which are based on a direct comparison of the observed data with theoretical isochrones. A range of techniques may be used for the isochrone fitting, from simple visual inspection to highly elaborated automatic procedures, and we therefore start by describing a typical such procedure and then discuss how the errors are estimated.

For relevant sections of the HR diagram, stellar ages are conventionally estimated by some variant of the following procedure. From the set of isochrones for the  $\zeta$  most closely matching the observed  $[\text{Me}/\text{H}]$  (or two isochrone sets bracketing the observed value), the isochrone passing through the observed point ( $T \log_{\text{eff}}, M_V$ ) is found by visual or numerical interpolation, and the corresponding age is assigned to the star. If two isochrone sets are used to bracket the metallicity, a further interpolation to the observed  $[\text{Me}/\text{H}]$  is required. The procedure is a simple practical recipe for computing  $\tau(q)$ . As discussed in Sect. 3.2, the procedure is formally equivalent to minimizing the  $\chi^2$  in Eq. (5) or (under the assumption of Gaussian errors) to ML estimation. If the fitting is made visually it is sometimes called “chi-by-eye”.

The age ambiguity resulting from overlapping isochrones can be resolved on the grounds discussed in Sect. 3.2, viz., by selecting the younger isochrone. For stars with solar-metallicity this occurs roughly for  $\log T_{\text{eff}} > 3.8$  in a band some 1.5–2 mag above the ZAMS; for low-metallicity stars it occurs at higher temperatures. (We do not consider here evolutionary stages on and beyond the giant branch, where the overlapping



**Fig. 7.** Estimated ages for the synthetic sample of 2968 single stars (same as in Fig. 5) compared with the true values. *From left to right*, the mean, median or mode of the  $G$  function was taken as an estimate of the stellar age.



**Fig. 8.** Estimated ages (using the mode of  $G$ ) for the synthetic sample of 937 stars with well-defined ages, compared with the true ages. The left panel is for single stars (same sample as in Fig. 6); for the middle and right panels secondary components have been added with two different mass-ratio distributions (see Sect. 3.8 for details).

becomes much more severe.) It is thus possible to define a unique inverse function  $\tau(\mathbf{q})$ . A more tricky question is how to obtain an estimate of the uncertainty of the resulting age.

#### 4.1. Error estimates: The “step method”

If the uncertainties in the observed data are known, it is possible to derive the uncertainty in the resulting age by examining how sensitive it is to changes in the observed data. The method used by Ng & Bertelli (1998, as described in Lachaume et al. 1999) may be representative: from the observed data point  $\mathbf{q}^{\text{obs}}$  we take a step of one standard deviation in each direction along each of the three axes; including the original point this gives a set of seven points in the three-dimensional data space for which ages are obtained through isochrone fitting. The mean value of the age estimates is adopted as the final age, and the scatter among the ages is taken to be the uncertainty of the mean value. (If some of the seven points fall outside the region covered by the isochrones, the resulting age estimate may be considered less reliable.) We refer to this particular method

as the “step method”. Superficially it looks like a reasonable method, but actually it severely underestimates the uncertainty, since only the error along one axis is considered at a time.

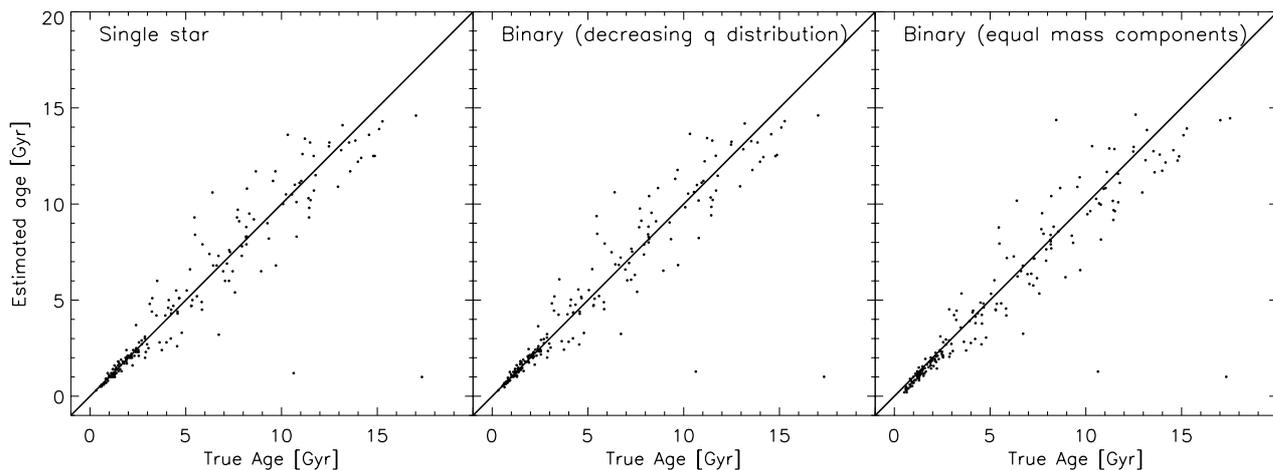
The effect is readily demonstrated by considering classical error propagation based on the linearization

$$\tau(\mathbf{q}^{\text{obs}} + \Delta\mathbf{q}) = \tau(\mathbf{q}^{\text{obs}}) + \Delta\tau \simeq \tau(\mathbf{q}^{\text{obs}}) + \sum_{i=1}^3 \frac{\partial\tau}{\partial q_i} \Delta q_i. \quad (16)$$

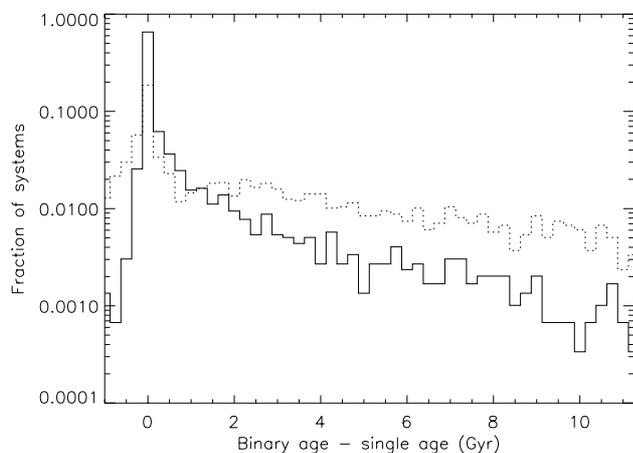
Assuming uncorrelated, centred errors with standard deviations  $\sigma_i$ , i.e.,  $\langle \Delta q_i \Delta q_j \rangle = \delta_{ij} \sigma_i^2$ , we have

$$\sigma_\tau^2 = \langle \Delta\tau^2 \rangle \simeq \sum_{i=1}^n \left( \frac{\partial\tau}{\partial q_i} \right)^2 \sigma_i^2. \quad (17)$$

On the other hand, also from Eq. (16), we have for the seven selected points  $\Delta\tau = 0, \pm(\partial\tau/\partial q_1)\sigma_1, \pm(\partial\tau/\partial q_2)\sigma_2,$  and  $\pm(\partial\tau/\partial q_3)\sigma_3$ ; their scatter (sample standard deviation) is exactly  $\sigma_\tau/\sqrt{3}$ . Thus, interpreting this scatter as the uncertainty of  $\tau$  underestimates the standard error by a factor  $\sqrt{3}$ , compared with what is obtained from the classical linear error



**Fig. 9.** Same as Fig. 8 but for the subsample with relative uncertainty  $\epsilon \leq 0.2$ .



**Fig. 10.** Histogram of the age error for binary systems, when their ages are determined from the integrated photometry as for single stars. The full line is for a decreasing mass-ratio distribution (see text), the dotted for equal-mass binaries. The bin size is 0.25 Gyr.

propagation formula. (The number 3 comes from the dimensionality of  $\mathbf{q}$ .) A simple expedient to remedy this effect is to multiply the computed error by  $\sqrt{3}$ . We call this the “corrected step method”.

Other factors may however contribute to making the error estimate unreliable. The strong non-linearity of  $\tau(\mathbf{q})$  means that there is in general a neglected contribution to the variance of  $\tau$  from truncated terms in the Taylor expansion Eq. (16). Moreover, in the parts of the HR diagram where the isochrones overlap, the required choice of  $\tau(\mathbf{q})$  among the multiple possible solutions could statistically lead to an underestimation of the ages and of their uncertainties.

#### 4.2. Error estimates: The Monte Carlo method

Monte Carlo (MC) simulation (Press et al. 1992) is often prescribed as a reliable and robust tool for error estimation. Indeed, it should effectively remove much of the difficulties caused by local non-linearity (if not by overlapping isochrones). In the present context the MC technique can be

used as follows. A large number ( $N$ ) of hypothetical data points  $\mathbf{q} = \mathbf{q}^{\text{obs}} + \Delta\mathbf{q}$  are generated for independent Gaussian errors  $\Delta\mathbf{q}_i$  with the assumed standard deviations  $\sigma_i$ . An age estimate  $\tau(\mathbf{q})$  is computed for each point by isochrone fitting and the sample standard deviation is adopted as the uncertainty of the mean estimated age. (Alternatively, confidence intervals can be estimated.) In contrast to the step method, errors in all dimensions of the data space are considered simultaneously, and non-linearities of  $\tau(\mathbf{q})$  are taken into account because data points several standard deviations from the observed values are also included in the statistics. Note that some  $N \sim 50$  experiments are required for a 10% precision of the error estimate, and some 5000 experiments if 1% precision is aimed at.

#### 4.3. Numerical examples

To compare the performance of our Bayesian estimation with the conventional isochrone fitting techniques described above, we have made extensive numerical simulations for the two representative cases illustrated in Fig. 1.

In Case 1 we assume the following *true* stellar model parameters:  $m = 1.34 M_{\odot}$ ,  $\zeta = -0.2$ ,  $\tau = 2.18$  Gyr. The corresponding true values of the observables are  $[\text{Me}/\text{H}] = -0.2$ ,  $\log T_{\text{eff}} = 3.825$ ,  $M_V = 3.0$ , corresponding to the left point in Fig. 1. The location of the star near the turn-off point should make the age determination relatively straightforward.

In Case 2 the true parameters are:  $m = 1.33 M_{\odot}$ ,  $\zeta = -0.2$ ,  $\tau = 2.85$  Gyr, corresponding to  $[\text{Me}/\text{H}] = -0.2$ ,  $\log T_{\text{eff}} = 3.800$ ,  $M_V = 3.0$ . This is a region in the HR diagram (right point in Fig. 1) where isochrone overlap may be a problem.

For each case we generated 1000 synthetic sets of observational data with the nominal errors (Sect. 2.5). The actual locations of the data points in the HR diagram are therefore scattered about the points indicated in Fig. 1 according to the error bars. An equal number of synthetic data sets were generated with half and twice the nominal errors. We then applied the various age determination methods to each data set. Error estimates were obtained as  $\pm 1\sigma$ , as described in Sects. 4.1 and 4.2,

**Table 1.** Results of numerical experiments comparing age determination methods using the step method, the corrected step method, the Monte Carlo method, and Bayesian estimation. Each line gives summary statistics from 1000 independent sets of synthetic data generated with the uncertainties specified in the first three columns. All ages are in Gyr. See Sect. 4.3 for further explanation.

Assumed $\sigma$ in			Median $\tau_{\text{est}}$ , $\pm$ median $\sigma_{\tau}$ , (% within CI), [median absolute deviation]			
[Me/H]	$\log T_{\text{eff}}$	$M_V$	Step method	Corrected	Monte Carlo	Bayesian
Case 1: True age $\tau_{\text{true}} = 2.18$ Gyr						
0.05	0.005	0.05	$2.19 \pm 0.11$ (44%) [0.13]	$\pm 0.19$ (67%)	$2.16 \pm 0.20$ (71%) [0.13]	$2.18^{+0.18}_{-0.20}$ (66%) [0.14]
0.10	0.010	0.10	$2.15 \pm 0.24$ (42%) [0.27]	$\pm 0.41$ (73%)	$2.12 \pm 0.51$ (75%) [0.32]	$2.20^{+0.36}_{-0.42}$ (68%) [0.25]
0.20	0.020	0.20	$2.05 \pm 0.61$ (50%) [0.65]	$\pm 1.05$ (77%)	$2.13 \pm 1.01$ (70%) [0.68]	$2.14^{+0.63}_{-0.88}$ (76%) [0.40]
Case 2: True age $\tau_{\text{true}} = 2.85$ Gyr						
0.05	0.005	0.05	$2.98 \pm 0.13$ (48%) [0.23]	$\pm 0.22$ (72%)	$3.10 \pm 0.49$ (68%) [0.27]	$2.83^{+0.21}_{-0.14}$ (72%) [0.09]
0.10	0.010	0.10	$3.13 \pm 0.44$ (41%) [0.46]	$\pm 0.76$ (64%)	$3.17 \pm 0.65$ (64%) [0.47]	$2.76^{+0.36}_{-0.27}$ (71%) [0.21]
0.20	0.020	0.20	$3.05 \pm 0.66$ (44%) [0.69]	$\pm 1.14$ (70%)	$3.21 \pm 1.10$ (69%) [0.70]	$2.60^{+0.61}_{-0.52}$ (70%) [0.37]

while for the Bayesian estimate the 68% CI was computed as in Sect. 3.6.

Results are summarized in Table 1 in the form of the following statistics:

1. The median<sup>3</sup> of the estimated ages,  $\tau_{\text{est}}$ . This should ideally equal the true age,  $\tau_{\text{true}} = 2.18$  Gyr (Case 1) or 2.85 Gyr (Case 2). The deviation from these values is the bias of the estimator.
2. The median of the error estimate  $\sigma_{\tau}$  (or error limits according to the CI), indicating the formal precision of the estimator. In the table this statistic is indicated by a  $\pm$  sign.
3. The percentage of data points for which the true age falls within the computed error limits. This number, which ideally should be 68%, is given within parentheses in the table. It indicates whether the preceding error limits are statistically realistic.
4. The median of the absolute deviation from the true age,  $|\tau_{\text{est}} - \tau_{\text{true}}|$ . This indicates the overall accuracy of the estimator, including bias, and is given in square brackets. For unbiased Gaussian errors, the median absolute deviation is  $0.675\sigma$ .

The results in Table 1 show some interesting trends:

- From the median  $\tau_{\text{est}}$  (item 1) we see that the bias is predominantly negative in Case 1 for all estimators, although quite small (at most  $-6\%$ ). This is consistent with the general bias discussed in Sect. 3.7. In Case 2, however, the step and MC methods produce a positive bias, generally increasing with the observational errors (up to  $+13\%$ ), while the Bayesian estimate goes in the opposite direction (up to  $-9\%$ ).
- The percentages in parentheses (item 3) show that all the error estimates are fairly realistic, except for the uncorrected

step method, which (as expected from Sect. 4.1) severely underestimates the errors.

- The corrected step method and MC method perform about equally well, but the Bayesian method nearly always gives smaller formal (item 2) and actual (item 4) errors. In the limit of vanishing observational errors the three methods should converge, as they clearly do at least in Case 1.

Although it may be hazardous to draw general conclusions from such a limited comparison, the results support our contention, based on the theoretical arguments in Sect. 3, that Bayesian estimation provides a more robust and accurate method for the determination of stellar ages than conventional isochrone fitting methods. This appears to be the case especially when the observational uncertainties are relatively large.

## 5. Application of the Bayesian age estimator

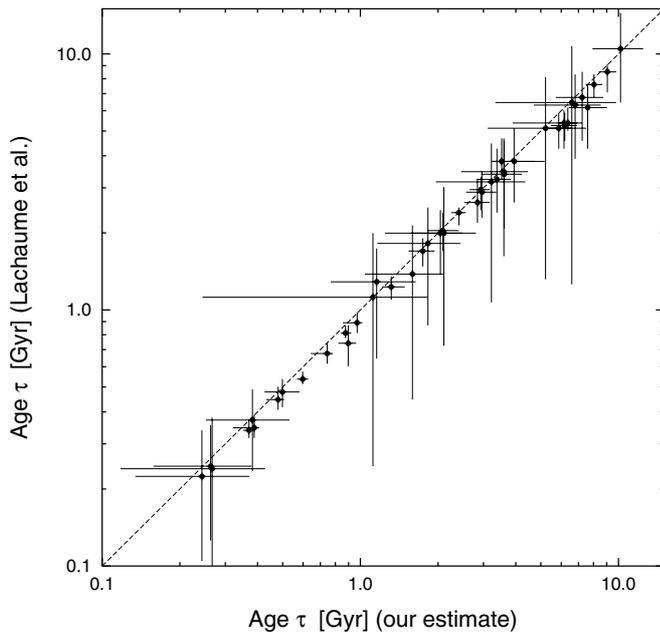
In Sect. 3 we described a general method to estimate the age of an individual star from the relative posterior probability density function  $G(\tau)$ , as well as several tests based on simulated data. In this section we present results obtained by applying the method to real data, including a limited comparison with other age estimates obtained by isochrone fitting. The statistical combination of  $G$  functions is briefly discussed in Sect. 5.2, where the method is applied to the two open clusters IC 4651 and M 67.

### 5.1. Ages of individual field stars

#### 5.1.1. The Geneva–Copenhagen survey of the Solar neighbourhood

The most extensive application to date of our method has been for the Geneva–Copenhagen survey of F and G dwarfs in the Solar neighbourhood (Nordström et al. 2004). From the total sample of 16 682 stars, age estimates were obtained for 13 679 stars, of which 11 468 are well-defined according to our criterion in Sect. 3.6. We refer to that paper for an extensive discussion of the astrophysical biases that may affect these age

<sup>3</sup> We consistently use the median for the statistics, as it is much more robust than the mean value for the sometimes extremely skew error distributions. The median also has the important advantage of being invariant with respect to the arbitrary choice of variable considered, viz.,  $\tau$  rather than  $\log \tau$ .



**Fig. 11.** A comparison of our age estimates and confidence intervals with those of Lachaume et al. (1999) for a sample of nearby stars. Only the stars with well-defined ages from both sources have been included.

determinations, in addition to the statistical effects addressed above.

It should be noted that Nordström et al. (2004) found it necessary to make an empirical adjustment of the temperature scale for the metal-deficient models ( $[\text{Fe}/\text{H}] < -0.3$ ), in order to obtain agreement with the observed lower main sequence. No such adjustment has been used in the present paper, which focuses on the computational techniques.

As an external check, Nordström et al. (2004) compared the ages derived by the present method with those by Edvardsson et al. (1993). For the 160 stars in common where our method gives well-defined ages according to the criterion in Sect. 3.6, the ages agree well for the oldest stars ( $\sim 10$  Gyr), while for the younger stars Edvardsson et al. (1993) obtained slightly higher ages (by about 20% at 2 Gyr). The scatter about the 1:1 line in a  $\log \tau_{\text{our}} - \log \tau_{\text{Edv93}}$  diagram (see their Fig. 8) is 0.12 dex ( $\approx 32\%$ ), scarcely more than the  $\pm 0.1$  dex uncertainty estimated by Edvardsson et al. (1993) for their data. Given the many improvements in stellar models, calibrations and distances since the earlier study, the overall agreement is quite gratifying.

### 5.1.2. Comparison with Lachaume et al. (1999)

Apart from the limited comparison with Edvardsson et al. (1993) described above we have made a more detailed comparison for the small sample of nearby stars in Table A1 of Lachaume et al. (1999). These authors explicitly give for each star the observational data with assumed uncertainties, as well as the resulting ages and confidence intervals, which makes a comparison relatively straightforward. Adopting their observational data we find that 44 stars have ages that are well-defined in both sets. The overall agreement is very good (Fig. 11) in

spite of the fact that Lachaume et al. (1999) used the older isochrone set by Bertelli et al. (1994). An unweighted linear regression gives

$$\log \tau_{\text{Lach99}} = (-0.027 \pm 0.005) + (1.002 \pm 0.009) \log \tau_{\text{our}} \quad (18)$$

(where  $\tau$  is in Gyr), with a scatter of 0.028 dex (7%). Thus our ages are in the mean about 6% higher than those of Lachaume et al. In addition, we find that our confidence intervals are systematically about 10% narrower. Both differences can perhaps be explained by the same effect as was seen for the Monte Carlo estimator in Case 1 of Table 1, namely a tendency to underestimate the age and overestimate the uncertainty. Equation (2) in Lachaume et al. (1999) is clearly the analytical equivalent of the Monte Carlo method of Sect. 4.2, which should therefore give comparable results. However, the situation is probably more complex, as illustrated by Case 2 in our Table 1, where the biases go in the opposite direction. In any case the very small scatter shows that the two methods give quite similar results for this particular sample with rather small observational uncertainties. This is consistent with the results in Sect. 4.3.

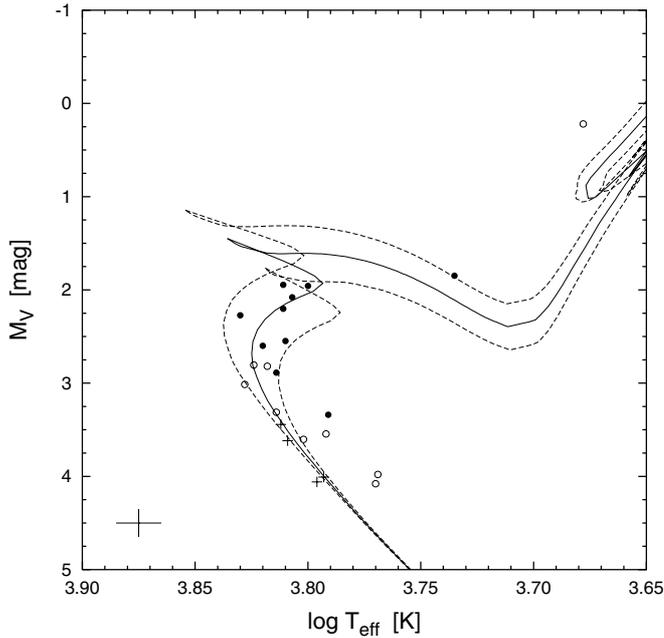
## 5.2. Ages of open clusters: The combination of $G$ functions

In an open cluster the population is coeval and the age estimates of the individual member stars should all be consistent with a single age. This can be used to test the consistency of age estimates obtained with our method. However, the prior information that the stars are coeval can be used much more powerfully in the Bayesian context, namely to estimate the age of the cluster. Recall that the relative posterior probability density for the age of an individual star ( $i$ ) is given by its  $G$  function  $G_i(\tau)$ . Assuming that the  $n$  stars in a cluster are all single, members of the cluster, and coeval, and that their observational data are independent, it follows that the relative posterior probability density for the cluster age is simply given by the product  $\prod_i G_i(\tau)$ . The cluster age and a confidence interval can be derived from this product as for an individual star.

Determining the age of a cluster can be regarded as a special case of the more general (and much more difficult) problem to determine the star formation history of a stellar population. In the same sense, taking the product of the  $G$  functions solves a special case of the general integral equation for the SFR as formulated e.g. in Hernandez et al. (2000). While the more general problem will be addressed in a future paper, we here apply the simple multiplication method to determine the ages of IC 4651 and M 67, using data from the literature. In these applications we neglect differential reddening (which is small in these clusters) as well as the radial extent of the clusters (contributing  $\lesssim 0.03$  mag to the scatter in  $M_V$ ), i.e., we assume that the same reddening and distance modulus apply to all stars in the cluster.

### 5.2.1. The age of the open cluster IC 4651

Meibom (2000) and Meibom et al. (2002) studied the intermediate-age open cluster IC 4651 in great detail, providing Strömgren photometry from which individual estimates

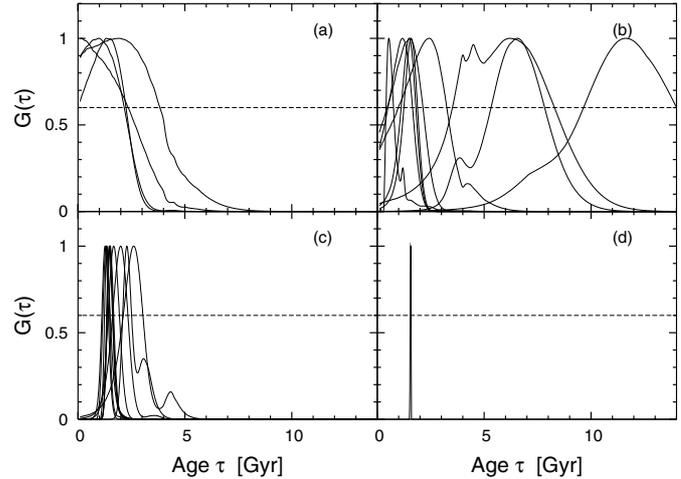


**Fig. 12.** HR diagram for 23 stars in IC 4651, together with isochrones at  $[\text{Me}/\text{H}] = +0.1$  for the best-fitting age 1.56 Gyr (solid line), and for 1.3 and 1.9 Gyr (dashed). The symbols are: filled circles – well-defined ages with  $\epsilon \leq 0.2$  (as in panel c of Fig. 13); open circles – well-defined ages with  $\epsilon > 0.2$  (panel b); crosses – not well-defined ages (panel a).

of  $\log T_{\text{eff}}$  and  $[\text{Me}/\text{H}]$  can be obtained, as well as diagnostics for duplicity  $[P(\chi^2)]$  and membership  $[P(RV)]$  based on multi-epoch radial velocities. From Table 1 in Meibom et al. (2002) we found 33 stars with  $P(\chi^2) \geq 0.1$  and  $P(RV) \geq 0.1$ , which sample should therefore be relatively clean from (spectroscopic) binaries and non-members. Using the calibrations in Nordström et al. (2004), estimates of  $\log T_{\text{eff}}$  and  $[\text{Me}/\text{H}]$  could be obtained for 23 of these stars (Holmberg, private comm.), assuming a reddening of  $E_{b-y} = 0.071$  mag (Meibom et al. 2002). The mean metallicity was found to be  $[\text{Me}/\text{H}] = 0.11 \pm 0.05$ , in good agreement with previous determinations, e.g., Meibom et al. (2002) who find  $[\text{Fe}/\text{H}] = 0.13 \pm 0.05$ .

Meibom et al. (2002) also found a true distance modulus of 10.03 mag to IC 4651 based on a direct fit to the Hyades main sequence. In comparison with theoretical isochrones they found however a better fit for an assumed distance modulus of 9.72–9.80 mag. For calculating  $M_V$  we assume a true distance modulus of 9.80 (corresponding to an apparent distance modulus of 10.105 mag), which we find gives more consistent age estimates than using 10.03 mag. Figure 12 shows the resulting HR diagram for the final sample of 23 stars.

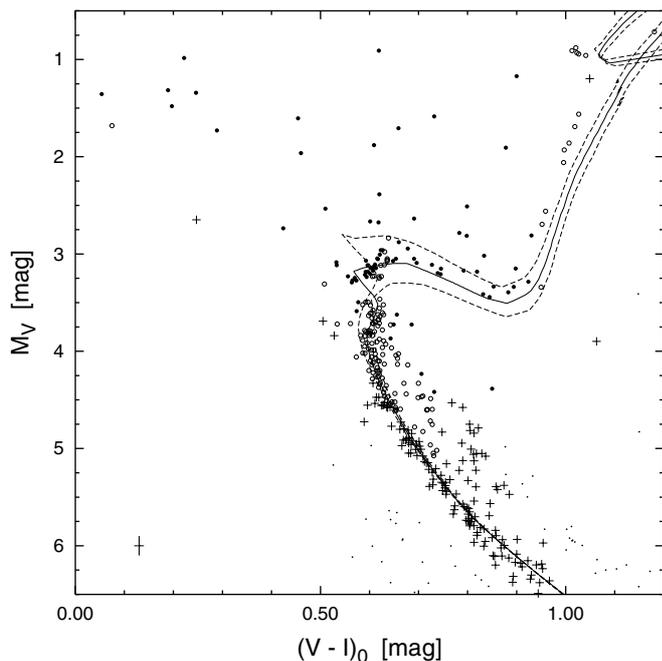
Figure 13 shows the  $G$  functions for the 23 probable single members of IC 4651. The dashed line shows the threshold  $G_{\text{lim}} = 0.6$  used to compute confidence intervals. The curves are grouped in panels (a)–(c) according to whether the ages are well-defined and their relative precisions  $\epsilon$ . In the HR diagram these are marked by different symbols. The mean value of the 10 best age estimates in panel (c) is  $1.70 \pm 0.15$  Gyr, identical to the best estimate for the age of IC 4651 by Meibom et al. (2002).



**Fig. 13.** Panels a)–c) show the  $G$  functions for the 23 stars in Fig. 12, subdivided according to the quality of the age estimates: a) – not well-defined; b) – well-defined with  $\epsilon > 0.2$ ; c) – well-defined with  $\epsilon \leq 0.2$ . Panel d) shows the product of all 23  $G$  functions, normalized to  $G = 1$  at the mode. This combined  $G$  function gives an estimated cluster age of  $1.56 \pm 0.03$  Gyr.

The fourth panel (d) in Fig. 13 shows the re-normalized product of all 23  $G$  functions. As previously explained, this product can under certain assumptions be interpreted as the posterior pdf of the cluster age, and leads to an estimated cluster age of  $1.56 \pm 0.03$  Gyr. However, this result (and in particular the very small error limits) should not be over-interpreted – for example, the assumption of independent astrophysical data for the different stars is clearly violated since we have to assume a common distance modulus for the cluster. Of the 23 stars, 13 (56%) have confidence intervals for their individual age estimates that include the value 1.56 Gyr, while the expected number is 68% or  $16 \pm 4$ . The individual age estimates are thus consistent with the derived cluster age.

In Fig. 12 we also plot the isochrones for 1.30, 1.56 and 1.90 Gyr, corresponding to a logarithmic uncertainty interval of  $\epsilon = 0.2$  around the estimated cluster age. As expected, the 10 stars with the best individual age estimates (filled circles) are mostly located around the turn-off point. However, one of them (MEI 9745) is conspicuously located  $\approx 0.7$  mag above the main sequence (a binary?), resulting in the highest individual age estimate ( $2.6 \pm 0.5$  Gyr) in the sub-sample. Eliminating this star gives a mean estimated age of  $1.60 \pm 0.12$  Gyr for the remaining 9 stars. The situation illustrates a potential source of bias when combining individual age estimates, namely that the stars with the most precise age estimates tend to be located above the main sequence (cf. Fig. 6) and therefore often associated with estimated ages that are too high (viz., if they are scattered from an actual location closer to the main sequence). The use of  $G$  functions helps to eliminate this effect: for instance, in Fig. 13 the combined  $G$  function rigorously takes into account also the four stars where the isochrones only give an upper limit to the age (panel (a) and the crosses in Fig. 12).

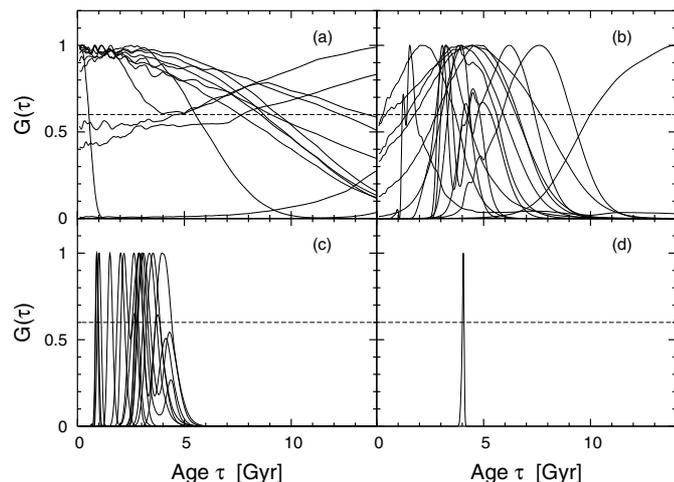


**Fig. 14.** HR diagram for 412 stars in the M67 region, together with isochrones at  $[\text{Me}/\text{H}] = 0.02$  for 4.05 Gyr (solid line), 3.4 and 4.9 Gyr (dashed line). The symbols are: filled circles – well-defined ages with  $\epsilon \leq 0.2$  (as in panel c of Fig. 15); open circles – well-defined ages with  $\epsilon > 0.2$  (panel b); crosses – not well-defined ages (panel a); small dots – no solution by Eq. (12).

### 5.2.2. The age of the open cluster M67

The data used for IC 4651 are unusually clean, and could therefore be considered an “easy” test case. A more challenging test is offered by M67. This old ( $\sim 4$  Gyr), well-studied open cluster (see Sandquist 2004, and references therein) is known to have a large binary population as well as a number of blue stragglers (Fan et al. 1996), and is at an age where the morphology of the isochrones near the turn-off point is particularly complex and sensitive to model assumptions (e.g., convective core overshooting).

Since the Padova isochrones give  $M_I$  (Cousins) for each theoretical model, we chose to compute the  $G$  functions directly with  $q_2 = (V-I)_0$  replacing  $q_2 = \log T_{\text{eff}}$  in Eq. (5), rather than transforming colour indices to effective temperature. We took  $V$  and  $V-I$  for 412 stars (some of them probably field stars) from Laugalys et al. (2004), which appear to be the most accurate such data currently available for M67. (That study gives  $V$  in the Vilnius system, which is equivalent to Johnson  $V$  for the range of colours considered; see Straižys 1973.) The  $V-I$  are on the same scale as in Sandquist (2004), and we adopt as in that paper the metallicity  $[\text{Fe}/\text{H}] = 0.02$ , total interstellar extinction  $A_V = 0.12$ , and colour excess  $E_{V-I} = 0.05$ . Following Sandquist (see his Fig. 14) we also apply an empirical correction of  $-0.05$  mag to the synthetic  $(V-I)_0$  in the Padova isochrones. In order to get a good fit to the main sequence a few magnitudes below the turn-off point, a distance modulus of  $(m-M)_0 = 9.48$  mag was assumed. This is intermediate between the value 9.42 given by Laugalys et al. (2004) and 9.60 from Sandquist (2004), but consistent with



**Fig. 15.** A random selection of some 50  $G$  functions for the M67 data, subdivided according to the quality of the age estimates as in Fig. 13. The combined  $G$  function in panel (d) gives an estimated cluster age of  $4.05 \pm 0.05$  Gyr.

both. Figure 14 shows the HR diagram derived with these assumptions.

$G$  functions could be computed for 364 of the 412 stars (assuming standard errors of 0.1 dex, 0.01 mag and 0.1 mag in  $[\text{Me}/\text{H}]$ ,  $(V-I)_0$  and  $M_V$ , respectively); a representative selection of them is shown in Fig. 15. A combined  $G$  function was computed by multiplying together the 364 individual  $G$  functions. To avoid numerical underflow and stabilize the result a constant 0.01 was added to each  $G$  function before multiplication<sup>4</sup>. The resulting combined  $G$  function, shown in panel (d) of Fig. 15, is very sharp and corresponds to an age estimate of  $4.05 \pm 0.05$  Gyr, in good agreement with Fan et al. (1996), who find  $4.0 \pm 0.5$  Gyr and Sandquist (2004), who also find an age around 4 Gyr. The very small formal error in our estimate naturally does not include the much bigger uncertainties from the colour transformation and intrinsic model errors, but illustrates the statistical power of the method. In contrast, many of the individual age estimates are too small, mainly for the blue straggler population, and as a result the mean value of the individual high-precision ( $\epsilon < 0.2$ ) age estimates is only 3.3 Gyr. Clearly the combined  $G$  function provides an extremely robust estimate of the cluster age even in the presence of all these complications.

## 6. Discussion

### 6.1. The Bayesian method in relation to some other methods described in the literature

There are age determination methods described in the literature that more or less resemble our method using  $G$  functions. Here, we briefly discuss how our method differs from these. We do not consider the several variants of isochrone interpolation

<sup>4</sup> Such an ad hoc procedure can be theoretically motivated by the finite probability (of order  $\sim 0.01$ ) that the star is a non-member or outlier in some other sense. We have verified that the resulting estimate of the cluster age is quite insensitive to the value of this constant.

(including  $\chi^2$  fitting, Sect. 4) used, e.g., by Twarog (1980), Edvardsson et al. (1993), Pols et al. (1997), Ng & Bertelli (1998), Chen et al. (2000), Liu & Chaboyer (2000), Ibukiyama & Arimoto (2002), and Lastennet & Valls-Gabaud (2002).

The Lachaume et al. (1999) method is in principle equivalent to the Monte Carlo (MC) described in Sect. 2.5. Lachaume et al. (1999) assume, as we have done, uncorrelated Gaussian errors in the three observed variables  $T_{\text{eff}}$ ,  $M_V$ , and  $[\text{Fe}/\text{H}]$ , from which they derive the probability density of the age,  $\mathcal{P}(t)$ , somewhat similar to our  $G$  function.

Note, however, that their Eq. (2) gives  $\mathcal{P}(t)$  as an integral over the *data space* ( $M_V, T_{\text{eff}}$ ), neglecting the uncertainty in  $[\text{Fe}/\text{H}]$ , whereas our Eq. (10) gives  $G$  as an integral over the *parameter space* ( $\zeta, m$ ). The difference is crucial, because it is only in parameter space that the prior densities (in this case of  $m$  and  $\zeta$ ) can be properly defined, due to the non-unique inverse transformation from data to parameters. The presence of the (non-unique) function  $t(T_{\text{eff}}, M_V)$  in their Eq. (2) clearly suggests a problem with their choice of integration variables. Consequently, their  $\mathcal{P}(t)$  does not take into account the different prior probabilities of finding a star at different locations in the HR diagram as a result of the varying rate of stellar evolution.

The method used by Reddy et al. (2003) to obtain stellar ages superficially resembles that of Lachaume et al. (1999), although they use  $\log g$  as one of the observables instead of  $M_V$ . However, in contrast to Lachaume et al. they perform the integration in the parameter space ( $M_i, Z$ ), where  $M_i$  (our  $m$ ) is the initial mass (their Eq. (4)). They consequently avoid the non-uniqueness problem and effectively compute the posterior probability density of the age assuming flat priors for  $M_i$  and  $Z$ . However, in this calculation they only include the observed metallicity by selecting isochrones with a metallicity within 0.25 dex of the observed value.

Hernandez et al. (1999, 2000) developed a method to reconstruct the star formation history through inversion of the HR diagram. This is a related, but from a statistical viewpoint quite separate problem from the determination of individual isochrone ages, and one that we will consider in a forthcoming paper. However, in their second paper, the authors introduce an expression (their Eq. (2)) involving a function  $G_i(t)$  of the age ( $t$ ) of each observed star ( $i$ ). This is equivalent to our Eq. (10), except that the latter is generalized to include any number of observables (our Eqs. (4) and (5)). Although Hernandez et al. (2000) do not discuss the determination of individual stellar ages, the possibility is clearly implied by their observation that  $G_i(t)$  “represents the probability that a given star,  $i$ , was actually formed at time  $t$  with any mass”, and consequently we have adopted our “ $G$ ” notation from that paper.

Pont & Eyer (2004) describe a Bayesian approach to the age determination problem which has much in common with our method. Their formulation leads to an expression for the posterior pdf of the age, their Eq. (4), which (apart from minor differences in the choice of variables) differs from our Eqs. (9) and (10) only in their inclusion of the Jacobian  $J(\mathbf{X})$  under the double integral. They motivate this factor by the change of variables in the likelihood function from the observables to the model parameters. We believe that is incorrect, since the likelihood is by definition a function of the model parameters (for

given observed data). Nevertheless, our computation of posterior densities (or  $G$  functions) agrees well with that of Pont & Eyer for selected stars from the Edvardsson et al. (1993) sample (their Fig. 8), indicating that their Monte Carlo evaluation leads to results that are practically equivalent to ours. Pont & Eyer (2004) do not explicitly discuss the choice of age estimate or the assignment of confidence intervals based on the posterior density.

## 6.2. Limitations of the method

The calculation of  $G$  functions, on which the age estimation rests, depends on a complex modelling of the observations in several steps, each with its associated theoretical and practical uncertainties. In a Bayesian sense, the adoption of a specific grid of theoretical models – such as the Padova isochrones used in this paper – corresponds to the a priori choice of a particular hypothesis about the nature of the observed stars. We attempt to obtain age determinations that are as good as possible within the framework of that particular hypothesis, but it is clear that the use of an alternative set of physical assumptions might have led to significantly different age estimates. We have chosen to neglect most of these problems in order to focus on the methodology. In applications to real observations, however, several such issues require more careful discussion than has been possible in this paper. Specifically for the application to the Geneva–Copenhagen survey (Sect. 5.1.1), we refer to the extensive discussion in Nordström et al. (2004). Here, we shall only add some general remarks concerning the various limitations of the method.

In spite of very considerable progress in the theoretical modelling of stellar interiors and evolution, standard model grids still employ highly simplified physical descriptions of the transport processes (convection, overshooting, diffusion, etc.) and mass loss, and suffer from some uncertainties in element abundances. Although the physical simplifications are often masked by good formal fits to the observed data obtained by adjusting ad hoc model parameters (such as the mixing-length parameter), the uncertainties in the underlying physics remain. For example, although the importance of convective core overshooting for the evolution of stars with masses  $\geq 1.2 M_{\odot}$  has been known for a long time (e.g., Maeder & Mermillion 1981), the extent of the overshooting, e.g. as function of mass, is still quite uncertain (VandenBerg & Stetson 2004). Diffusive processes, which affect both the evolutionary course and the surface abundances, have only recently become possible to treat in a self-consistent way (Michaud et al. 2004) and are not yet included in standard model grids. On the upper main sequence, mass loss and rotation add to the uncertainties; for example, accounting for the rotation of O- and early B-type stars could lead to ages larger by about 25% compared with the use of non-rotating models (Maeder & Meynet 2000).

As a result of the physical inadequacies normally present even in solar-type models, Young & Arnett (2005) estimate that the predictive accuracy of such models is limited to some 5–10% in the main physical parameters (luminosity, radius and effective temperature) of stars other than the Sun. In terms of

the position of the main sequence turn-off point, such errors in the effective temperature would translate into (systematic) age errors of order 25–100%. Of course, the situation is not quite so bad for stars around  $1 M_{\odot}$  and solar metallicity, since all models are carefully tuned to agree with the accurately known parameters of the Sun, including its age. However, for increasingly non-solar masses and metallicities, or more evolved stars, part of these uncertainties and systematic errors certainly come into play.

In a similar fashion, inadequacies in classical (1D, LTE) model atmosphere calculations (which provide both the outer boundary conditions for stellar interior models and the necessary transformations from theoretical to observable quantities such as colour indices) may be masked by semi-empirical corrections to the theoretical transformations or the use of empirical colour–temperature relations. The next generation of stellar atmosphere models (see, for example, Asplund 2003; Hauschildt 2003, and references therein), will no doubt eliminate some of these uncertainties, but do not yet cover more than a limited range of the stellar parameters.

In the present context, the most embarrassing discrepancies between theory and observations concern the location of the subdwarf sequence (cf. Sect. 5.1.1) and the location and slope of the red giant branch (cf. Fig. 14). In Nordström et al. (2004) the theoretical isochrones for subdwarfs were corrected by up to  $\delta T_{\text{eff}} = -0.022$  (at  $[\text{Fe}/\text{H}] = -1$ ). If left uncorrected, this would translate into an over-estimation of the ages (in terms of the location of the turn-off point) by up to 50%. By empirically correcting the temperature scale, one can hope that the resulting systematic error in the age determinations is substantially reduced, but this cannot be further quantified since the reasons for the discrepancy are largely unknown (Lebreton 2001).

Concerning the location of the red giant branch (RGB), different models predict differences in  $\log T_{\text{eff}}$  (for the same age, metallicity and luminosity) by up to  $\approx 0.015$  dex, while uncertainties in the transformation to  $V - I$  (for example) may be  $\sim 0.05$  mag (Salaris 2002); both correspond to an error in age by up to a factor 2–3 in either direction. Given these uncertainties, and the sensitivity of its location to metallicity, the RGB is mainly useful for relative age determination.

Of the many observational effects that influence the age determinations, in addition to those discussed in previous sections, we briefly mention photometric variability and interstellar extinction. Unless repeated measurements are taken, the photometric data on a variable star may not be representative of its average properties. A total amplitude exceeding 0.1 mag may add significant uncertainty to the age determination, especially if the different colour bands are not observed simultaneously. Based on the survey part of the Hipparcos and Tycho Catalogues (ESA 1997), we find that about 5% of the bright field stars are variable with a peak-to-valley amplitude  $> 0.1$  mag; about half of these are very red giants ( $V - I \gtrsim 2.0$ ) for which reliable ages cannot be estimated anyway. Of the remaining, more than half are main-sequence or sub-giant stars with  $V - I \lesssim 0.7$ , most relevant for age determination. Thus of the order 2% of the stars may be affected, including early-type variables ( $V - I \lesssim 0.0$ ),  $\delta$  Scuti stars ( $V - I \approx 0.30\text{--}0.45$ ) and eclipsing binaries (any  $V - I$ ).

Interstellar reddening has a very direct impact on the ages determined for stars near the turn-off point. Moreover, from photometry alone it is difficult to determine an accurate reddening for individual late-type stars because of its near-degeneracy with effective temperature. Roughly speaking, an error in  $V - I$  of 0.01 mag corresponds to an age error of the order of 10%. In clusters, differential reddening and a non-standard extinction law (e.g.,  $R$  greater than the standard value 3.1) may contribute to uncertainties in both colour and absolute magnitude.

In summary, several uncertainties of the theoretical models and calibrations, as well as observational effects, limit the systematic accuracy of the age determination, in addition to the statistical uncertainties and biases discussed in previous sections. Except for stars that are close to “solar-type” in all its main parameters, systematic errors caused by model inadequacies can easily reach 10% and more.

Other limitations of the present age determination method depend on specific implementation details that could, at least in principle, easily be modified. For example, as was pointed out in Sect. 2.1, the use of ( $[\text{Me}/\text{H}], \log T_{\text{eff}}, M_V$ ) as “observables” is to some extent arbitrary and mainly motivated by present uncertainties in the colour transformations. Given a better understanding of the latter, which may be achieved with the new generation of stellar atmosphere models, it will be straightforward to modify the method to use instead the observed colour indices in any suitable multi-colour system. This will, in addition, give a better handle on the correlations among the observables, which can then be explicitly introduced in the likelihood function by generalization of Eqs. (4) and (5). In the meantime, the use of different colour systems may help to quantify the systematic errors introduced by current model inadequacies, since a correct model should of course give consistent age determinations in the different systems.

Another limitation of the present implementation is the range of stellar parameters and evolutionary stages covered by the Padova isochrones. This can be overcome by the use of other (existing or future) data sets; in principle the modification is straightforward but in practice each set has its own particularities which require a separate interface to be coded. Among the several possible extensions, the inclusion of an additional parameter for the  $\alpha$ -element enhancement is very desirable for a better treatment of metal-deficient stars. To cope with young populations, it may be desirable to include the pre-main-sequence phase (the Padova isochrones start at the zero-age main sequence). These limitations are not important for the applications discussed in this paper, but in a future implementation the incorporation of, e.g., the  $Y^2$  isochrones (Yi et al. 2003) would add that flexibility to the method.

## 7. Conclusions

Conventional methods to determine stellar ages by isochrone interpolation have a number of weaknesses, most of them related to the strongly non-linear mapping from stellar model parameters (including age) to the observable data, and sometimes manifested in severely underestimated (or overestimated) error limits. Our alternative approach, founded on Bayesian techniques, avoids or at least takes these weaknesses into account

by systematically exploring the relevant regions of parameter space. In practical terms it leads to the calculation of a  $G$  function summarizing the posterior knowledge of an individual stellar age, from which age estimates and confidence intervals can be derived.

While undetected duplicity may lead to drastically wrong ages for individual objects, we have found that the statistical effects of binaries are rather modest, compared with the many other uncertainties of the age determination, when considering a stellar sample with a realistic binary frequency and mass ratio distribution.

Current and planned large-scale observational programmes in galactic astrophysics, such as the Gaia mission (Perryman et al. 2001), require highly automated, efficient and robust methods to classify and characterize large samples of stars, including the provision of age estimates. A main conclusion from the present analysis is that the complex and highly non-linear morphology of isochrones makes it difficult to assign unique age estimates, or even to define reasonable confidence intervals, for given observational data except in the most favourable cases and small observational errors. Moreover, selecting stars based on the inferred quality of the age estimates will introduce strong biases towards the more “favourable” age intervals.

The  $G$  functions are useful to characterize the age information independent of such considerations, and moreover provide a starting point for estimating the statistical properties of stellar samples, such as the history of star formation rate. Simple applications are given in Sect. 5.2, where the ages of two open clusters are robustly derived by combining the  $G$  functions for the individual stars. In a subsequent paper we explore a method to derive more generally the star formation rate of a population from the  $G$  functions of a stellar sample.

*Acknowledgements.* We are grateful to B. Nordström for giving the impetus to this study in the context of the Geneva–Copenhagen survey. We thank J. Holmberg for providing the data for IC 4651, J. Andersen, B. Nordström and S. Feltzing for many valuable comments on the manuscript, and C. Tout and A. Kučinskas for stimulating discussions on various aspects of the problem. Comments by the Editor encouraged us to expand substantially the discussion of possible errors. This work is supported by the Swedish National Space Board.

## References

- Asplund, M. 2003, in *CNO in the Universe*, ed. C. Charbonnel, D. Schaerer, & G. Meynet, ASP Conf. Ser., 304, 275
- Bertelli, G., Bressan, A., Chiosi, C., Fagotto, F., & Nasi, E. 1994, *A&AS*, 106, 275
- Casella, G., & Berger, R. L. 1990, *Statistical inference* (Belmont: Duxbury Press)
- Chen, Y. Q., Nissen, P. E., Zhao, G., Zhang, H. W., & Benoni, T. 2000, *A&AS*, 141, 491
- Duquennoy, A., & Mayor, M. 1991, *A&A*, 248, 485
- Edvardsson, B., Andersen, J., Gustafsson, B., et al. 1993, *A&A*, 275, 101
- ESA 1997, *The Hipparcos and Tycho Catalogues*, ESA SP-1200
- Fan, X., Burstein, D., Chen, J.-S., et al. 1996, *AJ*, 112, 628
- Feltzing, S., Holmberg, J., & Hurley, J. R. 2001, *A&A*, 377, 911
- Fuhrmann, K. 2004, *AN*, 325, 3
- Girardi, L., Bressan, A., Bertelli, G., & Chiosi, C. 2000, *A&AS*, 141, 371
- Gómez, A. E., Grenier, S., Udry, S., et al. 1997, *Proc. Hipparcos – Venice*, ESA SP-402, 621
- Hauschildt, P. H., Allard, F., Baron, E., Aufdenberg, J., & Schweitzer, A. 2003, in *GAIA Spectroscopy, Science and Technology*, ed. U. Munari, ASP Conf. Ser., 298, 179
- Hernandez, X., Valls-Gabaud, D., & Gilmore, G. 1999, *MNRAS*, 304, 705
- Hernandez, X., Valls-Gabaud, D., & Gilmore, G. 2000, *MNRAS*, 316, 605
- Ibukiyama, A., & Arimoto, N. 2002, *A&A*, 394, 927
- Kroupa, P., Tout, C. A., & Gilmore, G. 1993, *MNRAS*, 262, 545
- Lachaume, R., Dominik, C., Lanz, T., & Habing, H. J. 1999, *A&A*, 348, 897
- Lastennet, E., & Valls-Gabaud, D. 2002, *A&A*, 396, 551
- Laugalys, V., Kazalaukas, A., Boyle, R. P., et al. 2004, *Baltic Astron.*, 13, 1
- Laws, C., Gonzalez, G., Walker, K. M., et al. 2003, *AJ*, 125, 2664
- Lebreton, Y. 2001, *ARA&A*, 38, 35
- Liu, W. H., & Chaboyer, B. 2000, *ApJ*, 544, 818
- Maeder, A., & Mermilliod, J. C. 1981, *A&A*, 93, 136
- Maeder, A., & Meynet, G. 2000, *ARA&A*, 38, 143
- Meibom, S. 2000, *A&A*, 361, 929
- Meibom, S., Andersen, J., & Nordström, B. 2002, *A&A*, 386, 187
- Michaud, G., Richard, O., Richer, J., & Vandenberg, D. 2004, *ApJ*, 606, 452
- Mowlavi, N., Schaerer, D., Meynet, G., et al. 1998, *A&AS*, 128, 471
- Ng, Y. K., & Bertelli, G. 1998, *A&A*, 329, 943
- Nordström, B., Mayor, M., Andersen, J., et al. 2004, *A&A*, 418, 989
- Pagel, B. E. J., & Portinari, L. 1998, *MNRAS*, 298, 747
- Perryman, M. A. C., de Boer, K. S., Gilmore, G., et al. 2001, *A&A*, 369, 339
- Pols, O. R., Tout, C. A., Schröder, K.-P., Eggleton, P. P., & Manners, J. 1997, *MNRAS*, 289, 869
- Pont, F., & Eyer, L. 2004, *MNRAS*, 351, 487
- Press, W. H., Teukolsky, S. A., Vetterling, W. T., & Flannery, B. P. 1992, *Numerical Recipes*, 2nd ed. (Cambridge: Cambridge Univ. Press)
- Reddy, B. E., Tomkin, J., Lambert, D. L., & Allende Prieto, C. 2003, *MNRAS*, 340, 304
- Salaris, M. 2002, in *Observed HR diagrams and stellar evolution*, ed. T. Lejeune, & J. Fernandes, ASP Conf. Ser., 274, 50
- Sandquist, E. L. 2004, *MNRAS*, 347, 101
- Straizys, V. 1973, *A&A*, 28, 354
- Twarog, B. A. 1980, *ApJ*, 242, 242
- Vandenberg, D., & Stetson, P. B. 2004, *PASP*, 116, 997
- Yi, S. K., Kim, Y.-C., & Demarque, P. 2003, *ApJS*, 144, 259
- Young, P. A., & Arnett, D. 2005, *ApJ*, 618, 908