

ECOLE DOCTORALE 127  
ASTRONOMIE ET ASTROPHYSIQUE D'ILE DE FRANCE

OBSERVATOIRE DE PARIS

Doctorat  
ASTRONOMIE ET ASTROPHYSIQUE

Présentée et soutenue par

Céline GUÉDÉ

Détermination de l'âge des étoiles  
dans le cadre de la mission  
spatiale Gaia

Thèse dirigée par Yveline LEBRETON  
préparée à l'Observatoire de Paris, GEPI  
soutenue le 11 octobre 2013

**Jury :**

Carine BABUSIAUX	<i>Examinatrice</i>
Santi CASSISI	<i>Examineur</i>
João FERNANDES	<i>Examineur</i>
Marie-Jo GOUPIL	<i>Présidente du Jury</i>
Misha HAYWOOD	<i>Examineur</i>
Yveline LEBRETON	<i>Directrice de thèse</i>
Annie ROBIN	<i>Rapporteur</i>
Frédéric THÉVENIN	<i>Rapporteur</i>



# Résumé

---

Afin de dater les étoiles de la future mission Gaia, nous utilisons une méthode d'inversion basée sur les modèles stellaires. Cette méthode consiste à ajuster la position de l'étoile dans le diagramme HR par des modèles d'évolution stellaire (trajets évolutifs ou isochrones). Les données d'observation que nous utilisons sont : la température effective, la magnitude absolue et la métallicité de l'étoile.

Dans un premier temps, nous avons développé une méthode d'inférence bayésienne et nous avons effectué des tests afin de choisir les priors les plus adéquats. Nous avons choisi d'utiliser les trajets évolutifs de BaSTI (Pietrinferni et al. 2004). Le choix de la résolution de la grille est un compromis entre le temps de calcul et la finesse de cette grille. En étudiant les fonctions de densité de probabilité *a posteriori* découlant du processus de détermination de l'âge, nous avons identifié huit morphologies différentes. Nous avons ensuite mis au point un traitement automatique de ces fonctions afin de ne garder que celles qui sont bien définies. Cette méthode est également appliquée pour déterminer la masse et la métallicité des étoiles.

Afin d'effectuer tous ces tests nous avons créé un catalogue simulé basé sur les trajets évolutifs de BaSTI et sur les spécifications de Gaia. Il y a 69 % des âges qui sont bien déterminés. Les étoiles avec un âge mal déterminé sont localisées dans des régions dégénérées où la vitesse d'évolution des étoiles est très rapide ou lente.

Ensuite nous avons comparé plusieurs méthodes d'inférence bayésienne de la littérature afin d'identifier les différences entre celles-ci. Nous avons également comparé notre méthode avec une minimisation du  $\chi^2$ .

Afin d'améliorer la détermination des âges nous avons ajouté des contraintes qui proviennent d'observations complémentaires. Nous avons utilisé les paramètres sismiques et le paramètre de transit exoplanétaire mesurés par exemple par les missions CoRoT et Kepler et la gravité de surface provenant des observations spectroscopiques.

Enfin nous avons évalué les âges déterminés avec notre méthode pour des catalogues simulés à des distances, des directions d'observation et des erreurs observationnelles différentes. Nous avons également appliqué notre méthode sur des données réelles pour dater des amas.

**Mots clés :** âge des étoiles, inférence bayésienne, diagramme Hertzsprung-Russell, Gaia

---





# Abstract

---

We used an inversion method to date the stars observed by the Gaia mission. This method consists in estimating the age of an observed star with measured effective temperature, absolute magnitude and metallicity by looking for a theoretical stellar model (evolutionary tracks or isochrones) that fits the observed parameters.

Firstly, we developed a bayesian estimation and we realized several tests to adopt the most appropriate priors. We used the evolutionary tracks of BaSTI and we realized a compromise between computed time and thinness of the grid to select the appropriate resolution. We studied the shape of each probability density function resulting of the age-dating and we found eight main categories of shape. We implemented a method to analyze automatically the shape of the probability density function to select the ill-defined function and reject the associated star. This method is also applied to determine the mass and the metallicity of stars.

To test the capability of the age-dating method to return the true age of stars we built a simulated catalogue on the basis of the specifications of Gaia and the evolutionary tracks of BaSTI. We found 69 per cents of stars with a well determined age. The other 31 per cents are located in the degeneracy region where the evolution speed is either too rapid or slow.

Then we compared the several Bayesian estimation of the literature to analyze the differences with our method. We also implemented two  $\chi^2$ -minimizations to compare with our Bayesian estimation.

We added constraints -provided from complementary observations- to improve the age-dating. We used the asteroseismic and the exoplanetary transit parameters observed by the CoRoT and Kepler missions, and the surface gravity obtained by the spectroscopic observations.

Finally we determined the ages -with our Bayesian estimation- of simulated catalogues at several distances, galactic coordinates and observational errors. We also applied our method to date some clusters.

**Keywords** : stellar age, Bayesian estimation, Hertzsprung-Russell diagram, Gaia

---



---

# Remerciements

---

Ce manuscrit est le résultat de trois années travail pendant lesquelles de nombreuses personnes m'ont aidé. Ces quelques lignes sont l'occasion pour moi de les remercier comme il se doit, en espérant n'oublier personne.

Je remercie tout d'abord ma directrice de thèse Yveline Lebreton pour m'avoir acceptée puis encadrée pendant ces trois dernières années et pour m'avoir donné le goût de la recherche. Elle m'a permis de devenir plus autonome tout au long de ce travail de recherche.

Je voudrais ensuite remercier tous les membres du jury d'avoir accepté d'être présent lors de ma soutenance de thèse et pour m'avoir fait l'honneur d'évaluer mon travail. Je remercie tout particulièrement mes deux rapporteurs Annie Robin et Frédéric Thévenin qui ont relu et corrigé mon manuscrit. Leurs remarques constructives m'ont permis de voir mon travail sous un autre angle. Je remercie également Marie Jo Goupil qui a accepté de présider le jury, et João Fernandes et Santi Cassisi pour l'intérêt qu'ils ont portés à mon travail. Je remercie Misha Haywood et Carine Babusiaux pour notre collaboration sur ce travail, leurs remarques et l'aide qu'ils m'ont apporté.

Je voudrais maintenant remercier les personnes de l'Observatoire de Paris qui m'ont aidées et soutenues pendant ces trois années. Didier Pelat qui a pris beaucoup de temps pour répondre à mes diverses questions sur l'inférence bayésienne et sur la minimisation du  $\chi^2$ . David Katz, Frédéric Arenou, Frédéric Royer, Rosine Lallement, Monique Spite, François Spite, Piercarlo Bonifacio et Benoit Mosser qui ont également pris le temps de répondre à mes questions. Malek Mekkas et Sylvestre Taburet qui ont conçu la page web, ainsi que toute l'équipe informatique du GEPI. Les administratifs du GEPI qui m'ont aidés pour mes missions à Rennes, notamment Olga Martins pour sa gentillesse. Noël Robichon et Caroline Barban qui m'ont aidé à réaliser mon monitorat pendant 2 ans. Enfin je remercie toutes les personnes du bâtiment Copernic pour leur accueil et leur convivialité.

Je remercie également les autres doctorants et mes amies de l'Observatoire notamment Lucky, Hui-Chen, Cesar, Chloé, Yannick, Anne-Lise et Thierry avec qui j'ai partagé des déjeuners et des pauses thés dans la bonne humeur. Thank you again Hui-Chen, Cesar and Lucky for helping me to improve my english.

Pour mon pôt de thèse je voudrais remercier Irmelin, Monique, David, Jonathan et Georges ainsi que Frédéric Arenou pour m'avoir emmener chercher les courses. Un grand bravo à Irmelin et Jonathan pour leurs desserts succulents qui ont fait sensation.

Une thèse est impossible sans le soutien affectif dont j'ai eu besoin car j'ai réalisé ma thèse à distance entre Rennes et Meudon, c'est pourquoi je remercie particulièrement David, ma mère et Jonathan sans qui je n'aurais certainement pas tenu durant ces années.

Ces remerciements ne seraient pas complets sans une pensée pour mes parents, mon frère et ma soeur qui ont cru en moi et sans qui je ne serais pas là aujourd'hui. Je remercie également ma belle-famille pour son soutien.

Je souhaite remercier très chaleureusement Jonathan qui a commencé cette aventure doctoral en même temps que moi, on s'est donc écouté, aidé, soutenu et débloqué dans notre travail. Merci également pour ces super soirées d'observation à la coupole Arago à Paris.

Un merci spécial pour mes amis qui m'ont écouté et qui m'ont changé les idées quand j'en avais besoin, Johanna pour ces soirées entre filles devant de bons petits films, Brice et Mathieu pour ces soirées ciné-resto à Paris, Jennifer, Sophie, Cynthia, les deux Xavier, Cédric, Linda et Virginie pour ces soirées crêperies. Merci aussi à mon chat Gandalf avoir chassé les araignées de la terre du Milieu.

Je tenais à préciser que j'ai réalisé cette thèse sans café, sans thé et sans coca mais avec beaucoup de sucre, de chocolat et surtout de Kinder.

Pour terminer, le plus grand remerciement revient à David, mon compagnon, qui a été ma bulle d'oxygène pendant ces trois ans. Il a été d'un soutien indéfectible, il m'a également soutenu, épaulé, écouté, préparé de bon repas et surtout supporté tout le temps bien que lui aussi soit en thèse.

May the force be with you.

---

# Table des matières

Liste des figures	ix
Listes des tableaux	xiii
<b>1 Introduction</b>	<b>1</b>
1.1 Importance de l'âge des étoiles . . . . .	1
1.2 Mission spatiale Gaia . . . . .	2
1.2.1 Hipparcos . . . . .	2
1.2.2 Caractéristiques . . . . .	4
1.2.3 Objectifs scientifiques . . . . .	6
1.3 Détermination de l'âge des étoiles . . . . .	8
1.3.1 Différentes méthodes . . . . .	8
1.3.2 Diagramme de Hertzsprung-Russell . . . . .	9
1.3.3 Ajustement par des modèles stellaires . . . . .	11
1.4 Contenu général . . . . .	12
<b>I Présentation du formalisme des différentes méthodes</b>	<b>15</b>
<b>2 Inférence bayésienne</b>	<b>17</b>
2.1 Formalisme . . . . .	17
2.2 Choix des <i>a priori</i> . . . . .	20
2.2.1 Fonction de distribution de la métallicité . . . . .	21
2.2.2 Fonction de masse initiale . . . . .	23
2.2.3 Taux de formation stellaire . . . . .	28
2.3 Modèles d'évolution stellaire . . . . .	28
2.3.1 Résolution de la grille des trajets évolutifs . . . . .	30
2.3.2 Isochrones par rapport aux trajets évolutifs . . . . .	36
2.4 Morphologies de la fonction $\psi(\tau x)$ . . . . .	36
2.5 Détermination de l'incertitude sur l'âge . . . . .	39
2.5.1 Intervalle de confiance bayésien . . . . .	39
2.5.2 Intervalle de confiance de Jørgensen & Lindegren . . . . .	40
2.5.3 Vérification de l'intervalle de confiance . . . . .	40
2.6 Détermination des autres paramètres . . . . .	43
2.6.1 Masse . . . . .	43
2.6.2 Métallicité . . . . .	44

<b>3</b>	<b>Travaux antérieurs</b>	<b>47</b>
3.1	Inférence bayésienne . . . . .	47
3.1.1	Jørgensen & Lindegren . . . . .	47
3.1.2	da Silva et al. . . . .	48
3.1.3	Casagrande et al. . . . .	48
3.2	Minimisation du $\chi^2$ . . . . .	49
3.2.1	Ng & Bertelli . . . . .	50
3.2.2	$\chi^2$ standard . . . . .	50
3.2.3	Détermination de l'incertitude sur l'âge . . . . .	51
<b>4</b>	<b>Catalogues simulés de Gaia</b>	<b>53</b>
<b>II</b>	<b>Résultats</b>	<b>59</b>
<b>5</b>	<b>Résultats dans le HRD avec [M/H]</b>	<b>61</b>
5.1	Comparaison de l'âge réel simulé et de l'âge déterminé par inférence bayésienne . . . . .	61
5.2	Etudes des fonctions $\psi(\tau x)$ . . . . .	64
5.3	Impact de différents choix dans l'inférence bayésienne . . . . .	66
5.3.1	Densité d'étoiles le long des trajets évolutifs . . . . .	66
5.3.2	Métallicités . . . . .	67
5.3.3	Pas en âge ou $\log(\text{âge})$ . . . . .	68
5.4	Détermination de la masse . . . . .	69
5.5	Comparaisons des méthodes . . . . .	72
5.6	Conclusions et perspectives . . . . .	84
<b>6</b>	<b>Ajouts de contraintes observationnelles</b>	<b>85</b>
6.1	Astérosismologie . . . . .	87
6.2	Transits d'exoplanètes . . . . .	92
6.3	Gravité de surface . . . . .	93
6.4	Conclusions et perspectives . . . . .	96
<b>III</b>	<b>Applications</b>	<b>97</b>
<b>7</b>	<b>Estimation des résultats obtenus par Gaia</b>	<b>99</b>
7.1	Catalogues simulés . . . . .	99
7.2	Âges et masses . . . . .	100
7.3	Conclusions et perspectives . . . . .	106

---

<b>8 Applications à des catalogues d'observation</b>	<b>107</b>
8.1 Catalogue d'observation . . . . .	107
8.1.1 Geneva Copenhagen Survey . . . . .	107
8.1.2 Catalogue de Setiawan et al. . . . .	114
8.2 Comparaison avec les travaux antérieurs . . . . .	115
8.3 [M/H] vs. [Fe/H] . . . . .	117
8.4 Datation des amas . . . . .	118
8.5 Conclusions et perspectives . . . . .	124
<b>9 Conclusions générales et perspectives</b>	<b>125</b>
<b>IV Annexes</b>	<b>129</b>
<b>A Tables des résultats</b>	<b>131</b>
<b>B Interpolation des trajets évolutifs et des isochrones</b>	<b>141</b>
<b>C Compléments mathématiques</b>	<b>145</b>
C.1 Méthode de Simpson . . . . .	145
C.2 Méthode de Box-Muller . . . . .	145
C.3 Fonction de répartition de la loi normale . . . . .	146
C.4 Table de la distribution du $\chi^2$ . . . . .	148
<b>D Page web</b>	<b>151</b>
<b>E Publications</b>	<b>153</b>
<b>F Autres publications</b>	<b>187</b>
<b>G Nouveau chapitre de la thèse</b>	<b>207</b>





# Liste des figures

1.1	Structures de la Galaxie . . . . .	2
1.2	Mesure de la parallaxe . . . . .	3
1.3	Loi de balayage du satellite Gaia . . . . .	4
1.4	Bandes spectrales du photomètre de Gaia. . . . .	5
1.5	Capteur de Gaia. . . . .	6
1.6	Objectifs de la mission Gaia . . . . .	7
1.7	Diagramme HR . . . . .	9
1.8	Ajustement dans le diagramme HR . . . . .	12
2.1	Fonctions $\psi(\tau x)$ avec le mode, la moyenne et la médiane. . . .	21
2.2	Histogramme des métallicités . . . . .	22
2.3	Résultats des âges déterminées avec et sans MDF . . . . .	22
2.4	Représentation des IMFs . . . . .	27
2.5	Résultats des âges avec différentes IMFs. . . . .	27
2.6	Résultats de la résolution de la grille. . . . .	31
2.7	Comparaison des âges avec différentes résolution en métallicité	33
2.8	Comparaison des âges avec différentes résolution . . . . .	34
2.9	Comparaison des âges avec différentes résolution . . . . .	35
2.10	Morphologies des fonctions $\psi(\tau x)$ . . . . .	37
2.11	Lissage de la fonction $\psi(\tau x)$ . . . . .	38
2.12	Diagramme HR des étoiles de référence du Monte-Carlo . . . .	41
2.13	Simulation de Monte-Carlo pour une étoile . . . . .	42
2.14	Morphologie de la fonction $\psi(m x)$ . . . . .	44
2.15	Morphologie de la fonction $\psi([M/H] x)$ . . . . .	45
4.1	Diagramme HR du catalogue simulé de référence de Gaia . . .	54
4.2	Histogrammes des paramètres du catalogue simulé de référence	55
4.3	Histogrammes de la magnitude apparente G et des erreurs ob- servationnelles . . . . .	56
4.4	Diagramme HR du catalogue simulé de Gaia avec CESAM . .	57
4.5	Diagramme HR du catalogue simulé de Gaia avec Padoue . . .	57
5.1	Diagramme HR des étoiles non datées . . . . .	62
5.2	Diagrammes HR des résultats l'inférence bayésienne de notre travail . . . . .	63
5.3	Distribution des âges simulés et des âges déterminés . . . . .	63
5.4	Positions dans le diagramme HR des différentes morphologies de la fonction $\psi(\tau x)$ . . . . .	65

5.5	Diagrammes HR des âges déterminés avec la densité . . . . .	67
5.6	Résultats du nombre de métallicités . . . . .	68
5.7	Comparaison des âges déterminés entre un pas en âge et en log(âge) constant . . . . .	69
5.8	Diagrammes HR des résultats de la détermination des masses	69
5.9	$\log \tau$ - $\log(m/m_{\odot})$ . . . . .	70
5.10	Diagrammes HR des résultats de la détermination des masses avec K02 . . . . .	71
5.11	Histogramme des masses déterminées . . . . .	72
5.12	Résultats de la comparaison des méthodes . . . . .	73
5.13	Diagrammes HR des résultats des inférences bayésiennes . . .	74
5.14	Diagrammes HR des résultats des minimisations du $\chi^2$ . . . .	75
5.15	Résultats de la comparaison croisée $\chi^2$ -bayésien . . . . .	77
5.16	Diagrammes HR des résultats avec l'inférence bayésienne pour le catalogue simulé avec BaSTI. . . . .	78
5.17	Diagrammes HR des résultats avec la minimisation du $\chi^2$ pour le catalogue simulé avec BaSTI. . . . .	79
5.18	Diagrammes HR des résultats avec l'inférence bayésienne pour le catalogue simulé avec CESAM. . . . .	80
5.19	Diagrammes HR des résultats avec la minimisation du $\chi^2$ pour le catalogue simulé avec CESAM. . . . .	81
5.20	Diagrammes HR des résultats avec l'inférence bayésienne pour le catalogue simulé avec Padoue. . . . .	82
5.21	Diagrammes HR des résultats avec la minimisation du $\chi^2$ pour le catalogue simulé avec Padoue. . . . .	83
6.1	Spectres de fréquences . . . . .	86
6.2	Oscillations des étoiles dans le diagramme HR . . . . .	88
6.3	Résultats des âges avec l'ajout de contraintes. . . . .	90
6.4	Diagrammes HR des résultats avec les paramètres sismiques. .	91
6.5	Transit d'exoplanète . . . . .	92
6.6	Diagrammes HR des résultats avec la contrainte des transits d'exoplanètes. . . . .	93
6.7	Diagrammes HR des résultats avec la contrainte $\log g$ . . . . .	94
6.8	$\log \tau$ - $\log g$ . . . . .	95
7.1	Diagrammes HR simulés à différentes distances en direction du centre galactique . . . . .	101
7.2	Diagrammes HR simulés à différentes distances dans le plan galactique . . . . .	101

---

7.3	Diagrammes HR simulés à différentes distances au dessus du plan galactique . . . . .	102
7.4	Résultats à différentes distances, directions d'observations et $\sigma_{[M/H]} = 0.1$ dex . . . . .	103
7.5	Résultats à différentes distances, directions d'observations et $\sigma_{[M/H]} = 0.2$ dex . . . . .	104
7.6	Résultats à différentes distances, directions d'observations et $\sigma_{[M/H]} = 0.3$ dex . . . . .	105
7.7	Résultats de la détermination de la masse à différentes distances, directions d'observations et $\sigma_{[M/H]} = 0.2$ dex . . . . .	105
8.1	Diagramme HR du GCS de Casagrande et al. . . . .	113
8.2	Comparaison du GCS III et du GCS CSA11 . . . . .	114
8.3	Diagramme HR des étoiles de Setiawan et al. (2004) . . . . .	115
8.4	Comparaison des âges sur les catalogues réels. . . . .	116
8.5	Comparaison entre les métallicités $[Fe/H]$ et $[M/H]$ . . . . .	117
8.6	Comparaison entre les âges déterminées avec $[M/H]$ et $[Fe/H]$ . . . . .	118
8.7	Diagrammes HR des résultats entre les âges déterminées avec $[M/H]$ et $[Fe/H]$ . . . . .	119
8.8	Fonctions sommées pour la détermination de l'âge d'un amas . . . . .	120
8.9	Diagramme HR des Hyades et de Basel 4 . . . . .	120
8.10	Comparaison des âges déterminés des amas . . . . .	121
B.1	Diagrammes HR avec les PPEs de Basti et Padoue . . . . .	143
C.1	Méthode de Simpson . . . . .	146
C.2	Fonction de répartition de la loi normale . . . . .	148
D.1	Page web . . . . .	152



# Liste des tableaux

1.1	Types spectraux . . . . .	10
1.2	Classes de luminosité . . . . .	10
2.1	Résolution de la grille des modèles stellaires . . . . .	30
2.2	Temps de calcul et nombre de fonctions mal-définies pour les différentes résolutions de la grille . . . . .	31
5.1	Pourcentage d'étoiles des différents morphologies de la fonction $\psi(\tau x)$ . . . . .	64
6.1	Résultats des âges avec les contraintes sismiques selon les régions du diagramme HR . . . . .	90
7.1	Absorption en fonction de la distance et des coordonnées galactique . . . . .	100
8.1	Indice de la calibration de Casagrande et al. . . . .	113
8.2	Caractéristiques des amas . . . . .	122
8.3	Agés des amas . . . . .	123
A.1	Table des résultats de la détermination des âges . . . . .	132
A.1	Suite . . . . .	133
A.1	Suite . . . . .	134
A.1	Suite . . . . .	135
A.1	Suite . . . . .	136
A.1	Suite . . . . .	137
A.2	Table des résultats de la détermination de la masse . . . . .	138
A.3	Table des résultats de la détermination de la métallicité . . . . .	139
B.1	PPEs de BaSTI . . . . .	142
B.2	PPEs de Padoue . . . . .	144
C.1	Fonction de répartition de la loi normale . . . . .	147
C.2	Distribution du $\chi^2$ . . . . .	149



# Introduction

## Sommaire

<b>1.1</b>	<b>Importance de l'âge des étoiles</b>	<b>1</b>
<b>1.2</b>	<b>Mission spatiale Gaia</b>	<b>2</b>
1.2.1	Hipparcos	2
1.2.2	Caractéristiques	4
1.2.3	Objectifs scientifiques	6
<b>1.3</b>	<b>Détermination de l'âge des étoiles</b>	<b>8</b>
1.3.1	Différentes méthodes	8
1.3.2	Diagramme de Hertzsprung-Russell	9
1.3.3	Ajustement par des modèles stellaires	11
<b>1.4</b>	<b>Contenu général</b>	<b>12</b>

## 1.1 Importance de l'âge des étoiles

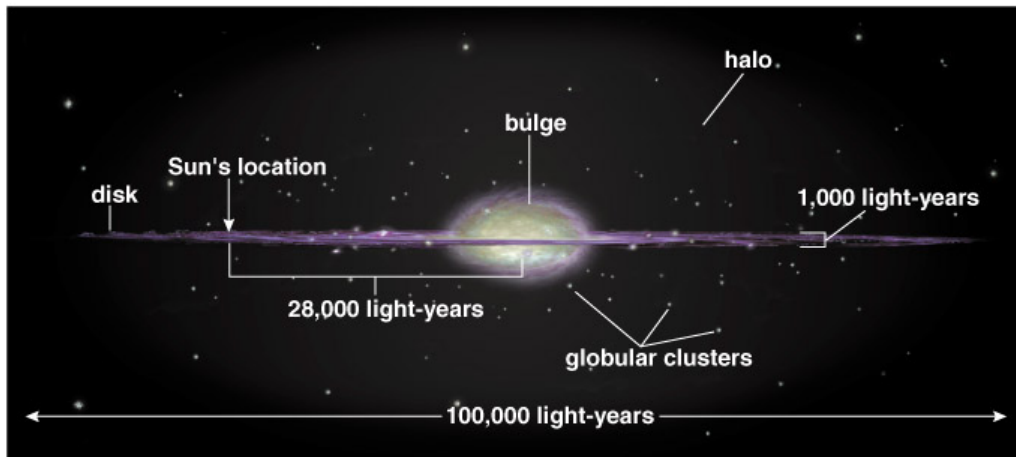
L'âge des étoiles est un paramètre important dans plusieurs domaines de l'astrophysique tels que la physique stellaire, l'étude des exoplanètes, le physique galactique et extragalactique et la cosmologie.

La composition chimique des étoiles évolue avec le temps, pour les caractériser il est nécessaire de connaître leur âge. D'autres paramètres comme la rotation (Kiraga & Stepien 2007) et le champ magnétique (Lakatos et al. 2005) dépendent de l'âge. La connaissance des relations entre ces différents paramètres permet de mieux comprendre l'évolution stellaire.

Actuellement, l'exoplanétologie est en développement avec les récentes découvertes de nombreuses planètes extra-solaires. Les propriétés des planètes évoluent avec le temps, il est donc indispensable de connaître précisément l'âge de l'étoile hôte afin de caractériser les systèmes exoplanétaires.

Pour caractériser notre Galaxie, la Voie Lactée, il faut également connaître l'âge des étoiles appartenant aux différentes structures qui la composent. Notre Galaxie est composée de quatre structures (figure 1.1) : le bulbe, le disque mince, le disque épais et le halo. Le bulbe sphéroïdal, de l'ordre de 2 000 parsecs de diamètre, est dense et il est composé principalement d'étoiles vieilles.

Le disque galactique de l'ordre de 50 000 parsecs de diamètre entoure le bulbe central. Ce disque est divisé en deux parties : le disque mince contenant des étoiles vieilles est situé dans le plan galactique et le disque épais est situé dans le plan galactique et s'étend au dessus du plan jusqu'à une distance de l'ordre de 1 000 à 5 000 parsecs. Ces structures sont entourées par le halo sphéroïdal qui est composé d'étoiles vieilles et d'amas globulaires.



(b)  
Copyright © Addison Wesley

FIGURE 1.1 – Schéma représentant les différentes structures de notre Galaxie (note de cours d'Addison Wesley)

Il est également important de connaître l'âge des plus vieilles étoiles de notre Galaxie qui sont situées dans le halo. Cet âge permet de donner une limite inférieure à l'âge de l'Univers. Cette valeur peut être comparée aux mesures obtenues avec le taux d'expansion de l'Univers (Watson 1998).

La détermination de la masse des étoiles est également importante car elle permet de mieux caractériser les étoiles, comprendre l'évolution et la formation stellaire.

## 1.2 Mission spatiale Gaia

### 1.2.1 Hipparcos

La mission spatiale Gaia (Global Astrometric Interferometer for Astrophysics), qui sera lancée en novembre 2013, observera plus d'un milliard d'étoiles de la Galaxie. Cette mission succède au satellite Hipparcos (High Precision PARallax COLlecting Satellite) dans l'observation astrométrique. Hipparcos (Perryman 1989), lancé en août 1989, était en orbite elliptique. Il a mesuré



de façon précise (milliseconde d'arc) la position, la parallaxe et les mouvements propres d'environ 120 000 étoiles du voisinage solaire de magnitude de 12.5 mag pendant 4 ans. Plus d'un million d'autres étoiles ont également été observées mais avec une précision moindre (20 - 30 millisecondes d'arc) et composent les deux catalogues de Tycho.

La position d'une étoile est définie par la mesure de l'ascension droite, de la déclinaison et de la distance de l'étoile au Soleil. Les mouvements propres de l'étoile représentent son déplacement apparent par rapport au système solaire. La parallaxe permet de déterminer la distance d'une étoile. Elle représente le déplacement de l'étoile sur la voûte céleste lorsque l'observateur se déplace (figure 1.2). Ce phénomène est visible pour les étoiles proches du Soleil. Par exemple, à six mois d'intervalle, vue depuis la Terre, la position de l'étoile a changé par rapport aux étoiles lointaines qui sont considérées comme fixes sur le ciel. Ce léger décalage angulaire peut être mesuré et c'est ce qu'on appelle la parallaxe. Au cours de l'année, l'étoile décrit un cercle apparent sur le ciel, appelé ellipse parallactique, qui correspond à la projection du mouvement de la Terre autour du Soleil. Le demi-grand axe de l'ellipse parallactique correspond au rayon  $R$  de l'orbite de la Terre (figure 1.2) et l'angle sous lequel l'observateur voit le demi grand axe de l'ellipse parallactique est la parallaxe  $\pi$ . Elle s'exprime en milliseconde d'arc et elle est liée à la distance avec la relation suivante :  $\pi = R/r$ , avec  $r$  la distance en parsecs de l'étoile au Soleil. Le rayon  $R$  s'exprime en unités astronomiques (U.A.) et vaut 1.0 car il est par définition la distance Terre-Soleil.

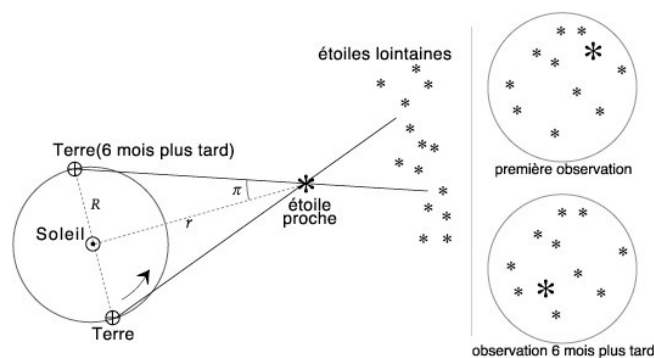


FIGURE 1.2 – Schéma représentant l'effet de parallaxe  $\pi$ , avec  $r$  la distance du Soleil à l'étoile et  $R$  la distance de la Terre au Soleil. Les étoiles lointaines sont considérées comme fixes sur le ciel.

Hipparcos observait simultanément dans deux directions d'observation situées à  $58^\circ$  l'une de l'autre. Un mélangeur de faisceau rassemblait les deux images sur les capteurs CCDs ce qui permettait à Hipparcos de mesurer la parallaxe des étoiles et les autres paramètres astrométriques.

Une description en 3D du voisinage solaire a été réalisée grâce aux données obtenues. Cela a permis des études détaillées dans plusieurs domaines de l'astrophysique : en physique galactique, en physique stellaire et dans l'étude du système solaire.

### 1.2.2 Caractéristiques

Le satellite Gaia (Perryman et al. 2001) sera placé au point de Lagrange L2 et observera dans deux directions d'observation (figure 1.3) comme son prédécesseur Hipparcos. Le satellite tournera sur lui même avec une légère précession en gardant le même angle avec la direction du Soleil, ce qui lui permettra d'observer tout le ciel. Pendant 5 ans, Gaia observera en moyenne 70 fois chaque étoile. Cette technique permettra d'obtenir des mesures précises en astrométrie pour plus d'un milliard d'étoiles. Pour les étoiles de magnitude apparente inférieure à 10 mag, la parallaxe sera mesurée avec une précision de 7 millièmes de seconde d'arc, pour les étoiles de magnitude apparente comprise entre 10 et 15 mag la précision sera de 12 à 24 micros seconde d'arc et pour les étoiles de magnitude apparente comprise entre 15 et 20 mag la précision sera de 100 à 300 micros seconde d'arc.

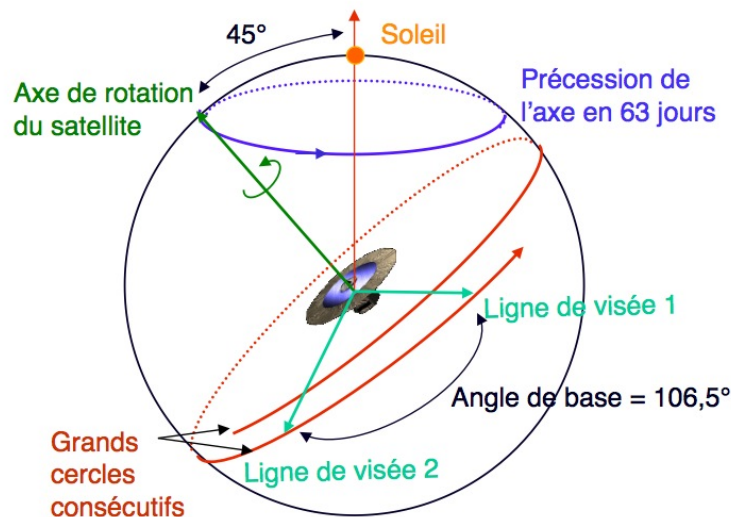


FIGURE 1.3 – Schéma représentant la loi de balayage du satellite Gaia (Turon & Arenou 2009).

Le satellite Gaia fera également des observations en photométrie et en spectroscopie. Gaia dispose d'un photomètre (figure 1.4) constitué de deux prismes en silice fondue. Le premier, nommé BP pour Blue Photometer, couvre le spectre visible de 330 à 680 nm. Le second, nommé RP pour Red Photometer, couvre la gamme de 640 à 1050 nm. Le réponse instrumentale astrométrique donne la magnitude G. A partir des deux bandes photométriques (BP et RP), de la magnitude G et de la parallaxe, les sources seront classifiées. Pour les étoiles simples, la température effective, la gravité de surface et la métallicité pourront être estimées.

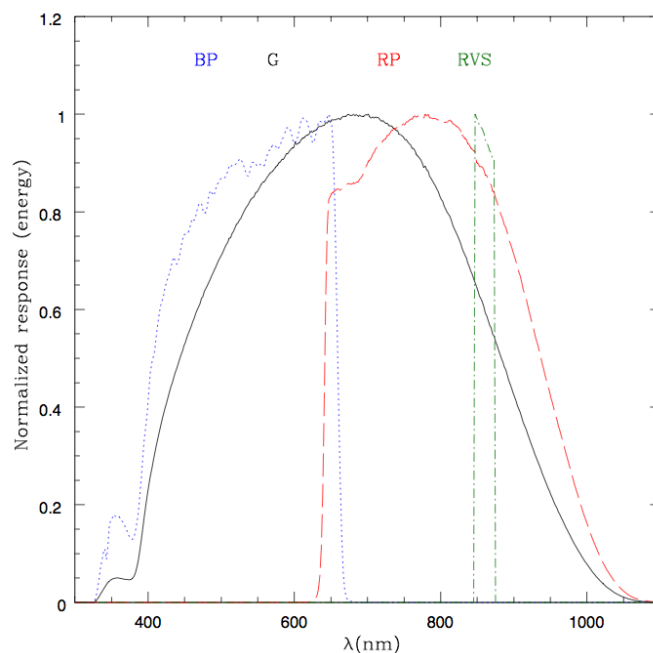


FIGURE 1.4 – Bandes spectrales du photomètre de Gaia (Jordi et al. 2010), avec le filtre astrométrique G, BP la bande spectrale dans le bleu, RP la bande spectrale dans la rouge et RVS la bande spectrale du spectroscopie haute résolution.

Le spectromètre haute résolution (RVS) observera dans une petite bande spectrale couvrant les longueurs d'onde de 847 à 871 nm, c'est le domaine des raies du triplet du calcium une fois ionisé. La spectroscopie permettra d'obtenir la vitesse radiale, c'est-à-dire la vitesse de l'étoile le long de la ligne de visée de l'observateur, par effet Doppler-Fizeau grâce au décalage des raies dans le spectre. Le RVS permettra également d'avoir des indications d'abondances individuelles, d'activité et de duplicité.

Pour effectuer toutes ces acquisitions, Gaia possède un capteur constitué de 106 CCDs (Charge-Coupled Device) de  $4500 \times 1966$  pixels (figure 1.5) situé

dans le plan focal. Quatre CCDs sont utilisés pour contrôler la stabilité du satellite et la qualité optique. Les images des deux lignes de visées arrivent tout d'abord sur le repéreur de sources constitué de 14 CCDs et qui doit confirmer la présence d'objets ponctuels à mesurer. Si le repéreur de sources confirme la présence d'une source ponctuelle, les deux images se déplacent sur les 62 CCDs dédiés aux mesures astrométriques (figure 1.5). Ensuite les images de la source détectée se déplacent sur les sept CCDs du photomètre bleu puis sur les sept CCDs du photomètre rouge. Enfin les images des étoiles sont lues en sortie des CCDs dédiés au spectromètre.

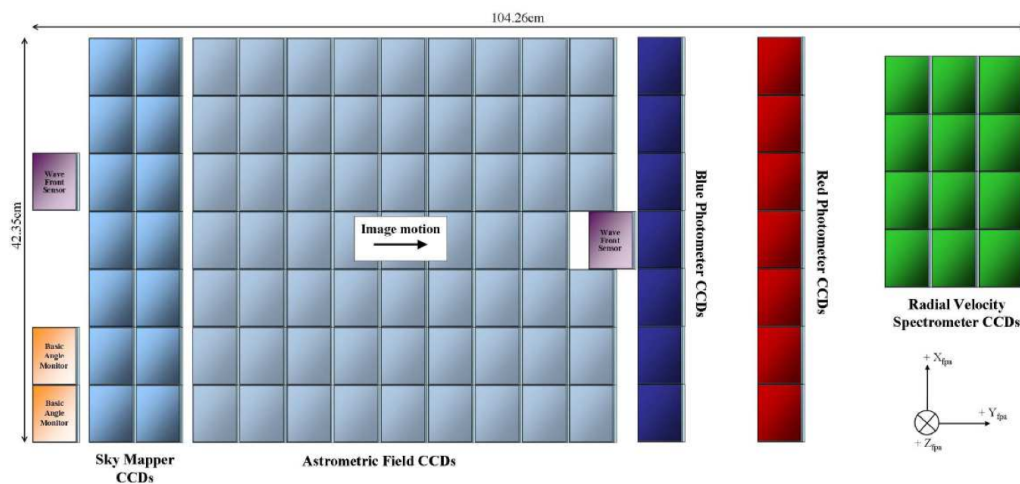


FIGURE 1.5 – Capteur de Gaia constitué de 106 CCDs (Jordi et al. 2010). Les images des étoiles se déplacent du repéreur d'étoiles (sky mappers) jusqu'aux CCDs dédiés au spectromètre (radial velocity spectrometer).

### 1.2.3 Objectifs scientifiques

La mission spatiale Gaia a plusieurs objectifs. Le premier étant d'observer tous les objets de magnitude apparente  $G$  inférieure à 20 mag et de déterminer de façon très précise la parallaxe, les mouvements propres et la luminosité de ces objets. Le second objectif est d'obtenir un vaste échantillon d'étoiles de toutes luminosités afin de construire un catalogue d'étoiles non biaisé.

Avec ces performances, Gaia va apporter des contributions dans plusieurs domaines de l'astrophysique (figure 1.6).

- **Physique galactique** : la position et la cinématique de toutes les étoiles des populations de la Galaxie et des amas permettraient d'établir des scénarios de formation et d'évolution de la Galaxie. De plus, la connaissance précise des positions et des mouvements propres des autres ga-

- laxies du groupe local permettra d'améliorer la compréhension de la dynamique au sein du groupe local.
- **Etude du système solaire** : l'observation des centaines de milliers de petits objets du système solaire permettra d'affiner les modèles de formation de ces objets.
  - **Etude des systèmes exoplanétaires** : l'observation de plusieurs milliers d'exoplanètes de la masse de Jupiter à une distance inférieure à 200 pc apportera une contribution à l'étude des exoplanètes.
  - **Physique stellaire** : grâce à la grande statistique de l'échantillon observé, des étoiles et des supernovæ de tous types seront observées, cela permettra de mieux comprendre l'évolution stellaire et la structure interne des étoiles.
  - **Cosmologie** : l'observation de Céphéides, des RR Lyrae et des supernovæ extragalactiques permettront de déterminer de façon plus précise les distances des galaxies lointaines et une meilleure détermination de la constante cosmologique. De plus, l'observation de nombreuses galaxies permettra de mieux les comprendre et de les classifier. Enfin le grand nombre de mesures permettra également d'identifier les traces des perturbations gravitationnelles dues à la matière noire.
  - **Physique fondamentale** : la mesure de la position des étoiles autour de Jupiter permettra d'affiner les valeurs des paramètres de la relativité générale.

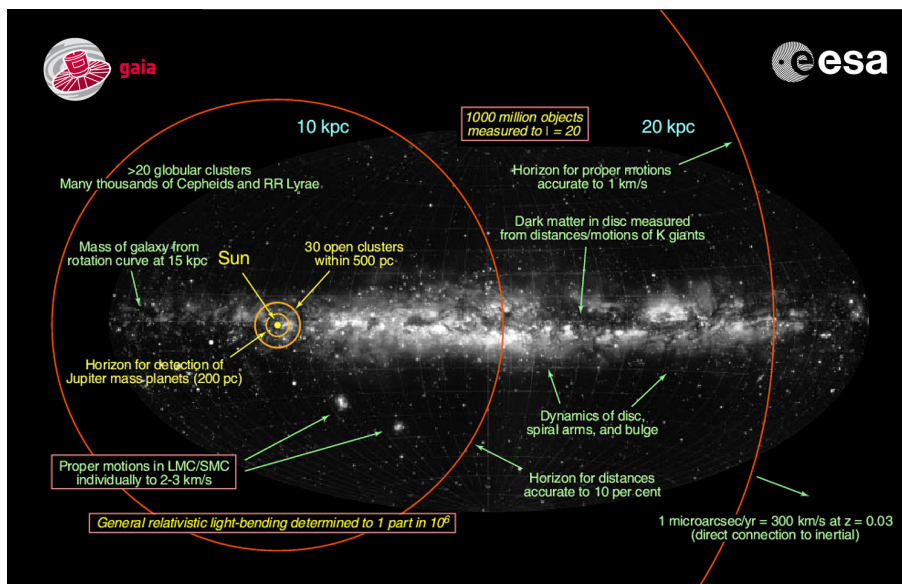


FIGURE 1.6 – Schéma représentant les différents objectifs de la mission spatiale Gaia. Site de l'ESA <http://sci.esa.int/gaia>.

Gaia observant plus d'un milliard d'étoiles il est nécessaire d'automatiser tous les processus d'analyse et de traitement des données C'est dans le cadre de la mission spatiale Gaia que nous avons développé une méthode de détermination des âges des étoiles.

## 1.3 Détermination de l'âge des étoiles

### 1.3.1 Différentes méthodes

L'âge des étoiles ne pouvant être mesuré directement, il est donc déterminé à partir d'estimations et d'inférences. Il existe plusieurs méthodes de datation selon le type de l'étoile. Soderblom (2010) les classifie en quatre catégories : les méthodes fondamentales et semi-fondamentales, les méthodes empiriques, les méthodes statistiques et les méthodes basées sur les modèles d'évolution stellaire.

Les méthodes fondamentales et semi-fondamentales sont basées sur des observations reliées à des phénomènes physiques bien connus. Par exemple, la nucléochronométrie (Cowan et al. 1991) mesure la décroissance radioactive des isotopes du thorium et de l'uranium. Comme cette décroissance est bien connue, il est possible remonter à l'âge des étoiles à partir de l'observation de ces isotopes. Toutefois, les incertitudes sur les données atomiques limitent cette méthode de datation.

Les méthodes empiriques sont basées sur des relations entre les âges et les paramètres observés. La quantité de lithium à la surface de l'étoile (Soderblom et al. 1993) diminue au cours de leur vie. Comme cette diminution est supposée suivre une loi de puissance, l'âge de l'étoile peut être déterminé en connaissant la quantité de lithium initiale. L'activité de la chromosphère<sup>1</sup> permet également de dater les étoiles, elle est reliée par une loi de puissance à son âge (Skumanich 1972). La gyrochronologie (Barnes 2009; Irwin & Bouvier 2009; Mamajek 2009) permet de d'estimer l'âge des étoiles à partir de leur couleur et de leur période de rotation.

Les méthodes statistiques sont basées sur des tendances observées dans la Galaxie, comme la relation âge-métallicité (Feltzing & Bensby 2009) ou les processus dynamiques (Wielen 1977). Ces méthodes restent aujourd'hui imprécises.

Enfin, les méthodes basées sur les modèles stellaires permettent de déterminer les âges par inversion. Les observations en astérosismologie permettent

---

1. La chromosphère est une région de l'atmosphère des étoiles située au dessus de la photosphère. La chromosphère solaire est épaisse de 10 000 km et sa température varie de 4 500 K à 10 000 K.

de sonder l'intérieur des étoiles. Oti Floranes et al. (2005); Mazumdar (2005); Lebreton & Montalbán (2009) déduisent le stade d'évolution et l'âge des étoiles à partir des observations et des modèles d'évolution stellaire. La méthode par ajustement des modèles d'évolution stellaire permet également de dater les étoiles (Edvardsson et al. 1993) dans le diagramme de Hertzsprung-Russell.

### 1.3.2 Diagramme de Hertzsprung-Russell

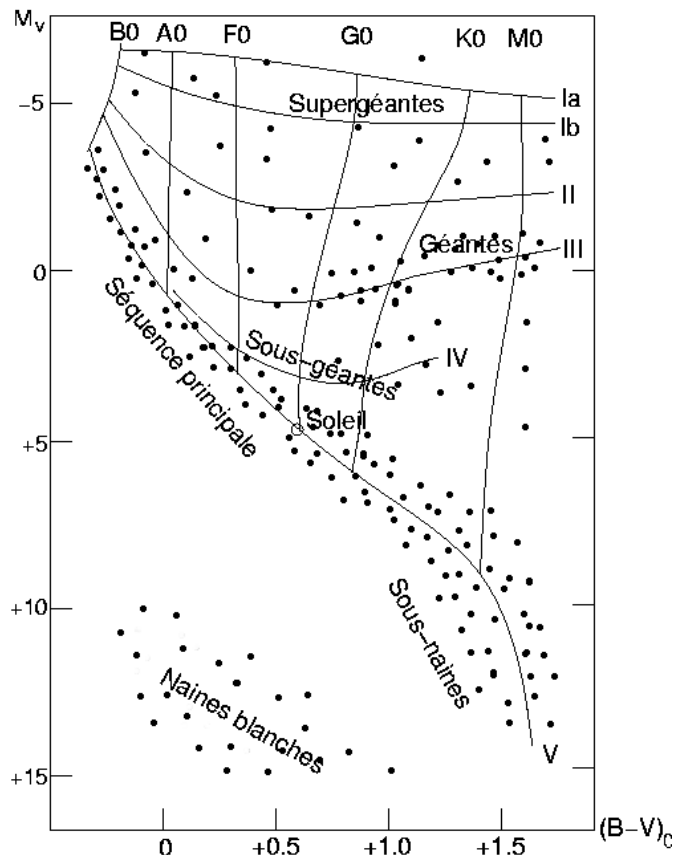


FIGURE 1.7 – Diagramme HR avec les différentes classes de luminosité et les types spectraux. Crédit : Wikipédia

Le diagramme de Hertzsprung-Russell a été inventé en 1910 par l'astronome danois Ejnar Hertzsprung et par l'astronome américain Henry Norris Russell. Ce diagramme peut être construit avec la luminosité des étoiles en fonction de la température effective. Il est appelé diagramme couleur-magnitude lorsqu'il représente la magnitude absolue ou apparente en fonction d'un indice de couleur. Nous les appellerons diagramme HR par la suite.

La classification spectrale des étoiles est visible sur un diagramme HR (figure 1.7), elle est définie par le type spectral (axe horizontal sur la figure 1.7) et par la classe de luminosité (axe vertical sur la figure 1.7). Le type spectral permet de classer les étoiles en fonction de leur température. En effet, les raies d'absorption présentes dans le spectre des étoiles varient en fonction de la température, de la gravité de surface et de la composition chimique de l'atmosphère stellaire. Les sept types spectraux sont présentés dans la table 1.1 selon la classification moderne de Keenan & Morgan (1951).

La classe de luminosité permet de distinguer les étoiles en fonction de leur luminosité, elle est également basée sur les raies spectrales. Les neuf classes de luminosité sont décrites dans la table 1.2

Ce type de diagramme a permis d'étudier les populations d'étoiles et d'établir la théorie de l'évolution stellaire.

TABLE 1.1 – Caractéristiques des types spectraux.

Types	Température	Couleur
O	> 25 000 K	bleue
B	10 000 à 25 000 K	bleue-blanche
A	7 500 à 10 000 K	blanche
F	6 000 à 7 500 K	jaune-blanche
G	5 000 à 6 000 K	jaune
K	3 500 à 5 000 K	jaune-orange
M	< 3 500 K	rouge

TABLE 1.2 – Présentation des différentes classes de luminosité.

Classe	Description
0	Hypergéantes
Ia	Supergéantes très lumineuses
Ib	Supergéantes moins lumineuses
II	Géantes lumineuses
III	Géantes
IV	Sous-géantes
V	Naines (ou étoiles de la séquence principale)
VI	Sous-naines
VII	Naines blanches



### 1.3.3 Ajustement par des modèles stellaires

Nous utilisons la méthode de l'ajustement des modèles d'évolution stellaire car elle permet de dater un grand échantillon d'étoiles et surtout parce que nous ne disposons pas d'observables qui nous permettraient d'utiliser d'autres méthodes comme l'astérosismologie. Cette méthode consiste à estimer l'âge à partir des modèles d'évolution stellaire en ajustant les paramètres observés des étoiles dans le diagramme HR.

Les modèles d'évolution stellaire les plus utilisés sont les trajets évolutifs et les isochrones. Les trajets évolutifs sont calculés à partir des codes d'évolution stellaire et ils représentent l'évolution temporelle d'une étoile pour une métallicité  $[M/H]$ , une fraction de masse de l'hélium et une masse initiale données. Les isochrones sont déterminées à partir des trajets évolutifs et elles représentent les différents stades d'évolution, c'est-à-dire pour différentes masses, à un âge et une métallicité donnés.

Edvardsson et al. (1993) détermine par interpolation graphique l'âge des étoiles en comparant la position de l'étoile dans le diagramme magnitude absolue-température effective avec des isochrones ayant la métallicité de l'étoile. Ng & Bertelli (1998) utilisent une minimisation du  $\chi^2$  afin de sélectionner l'isochrone la plus proche de l'étoile. Dans certaines régions du diagramme HR, il est difficile de sélectionner l'isochrone la plus proche car plusieurs d'entre elles peuvent ajuster correctement l'étoile. En effet sur la figure 1.8, l'étoile de droite est ajustée par trois isochrones qui sont dans des stades d'évolution différents. L'isochrone à 2.85 Gyr ajuste l'étoile sur la séquence principale, celle à 3.27 Gyr au turn-off et celle à 3.82 Gyr sur la branche des sous géantes. Cette dégénérescence est due à la vitesse d'évolution de l'étoile qui est trop rapide dans cette région. Le problème est de savoir comment sélectionner l'isochrone correspondant à l'étoile sans avoir d'information sur le stade évolutif de l'étoile. Pont & Eyer (2004) puis Jørgensen & Lindegren (2005); da Silva et al. (2006); Takeda et al. (2007); Hernandez & Valls-Gabaud (2008); Casagrande et al. (2010) utilisent l'inférence bayésienne afin de déterminer l'âge le plus probable de l'étoile à partir de trois observables : la température effective  $T_{\text{eff}}$ , la magnitude absolue  $M_V$  et la métallicité  $[M/H]$  de l'étoile. L'inférence bayésienne permet de prendre en compte des *a priori* qui sont basés sur les connaissances de la physique stellaire et galactique. C'est la méthode que nous avons choisi d'utiliser pour déterminer l'âge, la masse et la métallicité des étoiles.

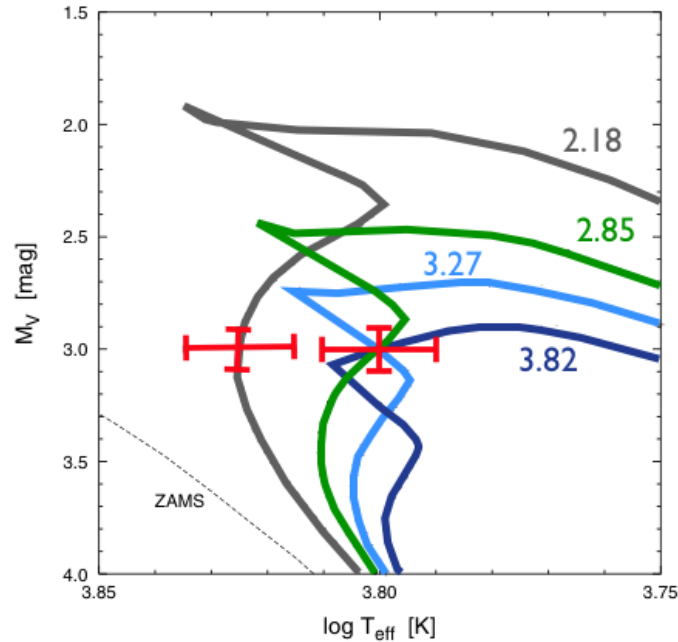


FIGURE 1.8 – Ajustement de deux étoiles par des isochrones dans le diagramme HR (Jørgensen & Lindegren 2005).

## 1.4 Contenu général

Ce manuscrit détaille notre travail réalisé sur la détermination de l'âge, de la masse et de la métallicité des étoiles par inversion. Il s'articule en trois parties :

1. Dans un premier temps, nous décrivons le formalisme de l'inférence bayésienne que nous avons développé ainsi que les choix des différents *a priori*, de la forme et de la résolution de la grille des modèles d'évolution stellaire et le traitement que nous avons développé sur les fonctions résultantes de l'inférence bayésienne. Nous décrivons également le formalisme des travaux antérieurs sur l'inférence bayésienne et la minimisation du  $\chi^2$ . Enfin, nous décrivons la construction et les spécificités des catalogues simulés que nous utilisons pour les études.
2. Dans la deuxième partie, nous présentons les résultats de la détermination de l'âge, de la masse et de la métallicité par l'inférence bayésienne sur un catalogue simulé. Nous comparons ensuite notre méthode avec les travaux antérieurs sur l'inférence bayésienne et la minimisation du  $\chi^2$ . Enfin, nous présentons notre étude sur l'ajout de contraintes d'ob-

servations, issues d'observations complémentaires, afin de réduire les dégénérescences dans la détermination de l'âge.

3. Dans la troisième partie, nous appliquons notre méthode de datation afin d'estimer les résultats obtenus avec la mission spatiale Gaia. Pour cela nous déterminons l'âge et la masse des étoiles des catalogues simulés à différentes distances, dans différentes directions d'observation et pour différentes erreurs observationnelles. Nous appliquons également notre méthode à des catalogues observées notamment pour dater les amas.

Pour ce travail, nous avons effectué 168 calculs d'âge, de masse ou de métallicité sur des catalogues simulés de 10 000 étoiles ou sur des catalogues observés, cela représente plus de 100 jours de calculs sur un serveur. Nous avons résumé les résultats dans les tables présentées en annexe A.



## Première partie

### Présentation du formalisme des différentes méthodes



# Inférence bayésienne

---

## Sommaire

---

<b>2.1</b>	<b>Formalisme</b>	<b>17</b>
<b>2.2</b>	<b>Choix des <i>a priori</i></b>	<b>20</b>
2.2.1	Fonction de distribution de la métallicité	21
2.2.2	Fonction de masse initiale	23
2.2.3	Taux de formation stellaire	28
<b>2.3</b>	<b>Modèles d'évolution stellaire</b>	<b>28</b>
2.3.1	Résolution de la grille des trajets évolutifs	30
2.3.2	Isochrones par rapport aux trajets évolutifs	36
<b>2.4</b>	<b>Morphologies de la fonction <math>\psi(\tau x)</math></b>	<b>36</b>
<b>2.5</b>	<b>Détermination de l'incertitude sur l'âge</b>	<b>39</b>
2.5.1	Intervalle de confiance bayésien	39
2.5.2	Intervalle de confiance de Jørgensen & Lindegren	40
2.5.3	Vérification de l'intervalle de confiance	40
<b>2.6</b>	<b>Détermination des autres paramètres</b>	<b>43</b>
2.6.1	Masse	43
2.6.2	Métallicité	44

---

## 2.1 Formalisme

Soit deux événements  $A$  et  $B$  appartenant à l'espace des épreuves  $\Omega$ . Si l'événement  $B$  est connu, la probabilité conditionnelle de  $A$  sachant  $B$  peut être calculée telle que  $P(A|B) = P(AB)/P(B)$  si  $P(B) \neq 0$ . De même  $P(AB) = P(A|B)P(B) = P(B|A)P(A)$  si  $P(A) \neq 0$ . La formule de Bayes s'écrit donc :

$$P(B|A) = \frac{P(A|B)P(B)}{P(A)}. \quad (2.1)$$

Cette formule peut être généralisée pour des fonctions continues, où  $\theta$  est l'espace des paramètres et  $x$  l'espace des observables. La fonction de densité

de probabilité *a posteriori*  $\psi(\theta|x)$ , qui représente la probabilité d'obtenir les paramètres  $\theta$  sachant les observations  $x$ , s'écrit :

$$\psi(\theta|x) = \frac{\psi(x|\theta)\pi(\theta)}{\int \psi(x|\theta)\pi(\theta)d\theta}. \quad (2.2)$$

Avec la vraisemblance  $\psi(x|\theta)$  qui est la fonction de densité de probabilité des observations  $x$  sachant les paramètres  $\theta$ , notée usuellement  $\mathcal{L}(x|\theta)$ . La fonction de densité de probabilité *a priori*  $\pi(\theta)$  représente le degré de connaissance sur les paramètres  $\theta$ . La distribution marginale  $m(x) = \int \psi(x|\theta)\pi(\theta)d\theta$  est un facteur de normalisation qui est indépendant des hypothèses. Dans notre méthode de datation nous avons choisi de ne pas inclure ce terme (Robert 2007).

Pour la détermination de l'âge des étoiles, l'espace des observables  $x$  est souvent basé sur des observations photométriques. Il est composé de la température effective  $\widetilde{\log T_{\text{eff}}}$ , de la magnitude absolue  $\widetilde{M_V}$  et de la métallicité  $\widetilde{[M/H]}$  de l'étoile. La magnitude absolue correspond à la magnitude de l'étoile si elle était observée à une distance de 10 pc du Soleil. Nous utilisons la magnitude absolue dans le filtre  $V$  du système photométrique de Johnson (Johnson & Morgan 1953). La métallicité désigne la proportion d'atomes plus lourds que l'hélium dans l'étoile, elle est calculée en la comparant au Soleil :  $[M/H] = \log(\sum_{Z \geq 3} N_Z/N_H) - \log(\sum_{Z \geq 3} N_Z/N_H)_\odot$ .  $N_H$  est le nombre d'atomes d'hydrogène par unité de volume,  $\sum_{Z \geq 3} N_Z$  est la somme du nombre d'atomes de tous les éléments plus lourds que l'hélium et  $Z$  est le numéro atomique des éléments. Nous notons les observables avec un "~" afin de les différencier des paramètres. Chacune de ces observables est affectée d'une erreur observationnelle  $\sigma$ .

L'espace des paramètres est défini à partir des modèles d'évolution stellaire. Il comprend la métallicité  $[M/H]$ , la masse  $m$  et l'âge de l'étoile  $\tau$ . La fonction de densité de probabilité *a posteriori*  $\psi([M/H], m, \tau | \widetilde{\log T_{\text{eff}}}, \widetilde{M_V}, \widetilde{[M/H]})$  s'écrit :

$$\begin{aligned} \psi([M/H], m, \tau | \widetilde{\log T_{\text{eff}}}, \widetilde{M_V}, \widetilde{[M/H]}) &\propto \mathcal{L}(\widetilde{\log T_{\text{eff}}}, \widetilde{M_V}, \widetilde{[M/H]} | [M/H], m, \tau) \\ &\times \pi([M/H], m, \tau). \end{aligned} \quad (2.3)$$

Nous définissons  $\pi([M/H], m, \tau) = \phi([M/H]) \times \xi(m) \times \lambda(t)$  où  $\phi([M/H])$  est la fonction de distribution de la métallicité,  $\xi(m)$  est la fonction de masse initiale et  $\lambda(t)$  est le taux de formation stellaire depuis la formation de l'Univers. La fonction de densité de probabilité *a priori* dépend de trois fonctions marginales car nous supposons que ces fonctions sont statistiquement indépendantes. Nous faisons cette supposition afin de ne pas introduire de biais



dans la détermination de l'âge. Nous pouvons considérer que les erreurs observationnelles suivent une loi normale  $\mathcal{N}(x, \sigma^2)$ , la vraisemblance est alors définie par :

$$\mathcal{L}(\widetilde{\log T_{\text{eff}}}, \widetilde{M_V}, \widetilde{[M/H]} | [M/H], m, \tau) = \left( \prod_{i=1}^n \frac{1}{(2\pi)^{\frac{1}{2}} \sigma_{x_i}} \right) \exp\left(\frac{-\chi^2}{2}\right) \quad (2.4)$$

$$\text{avec } \chi^2 = \sum_{i=1}^n \left( \frac{x_i - \tilde{x}_i}{\sigma_{x_i}} \right)^2. \quad (2.5)$$

Avec  $n$  le nombre d'observables,  $\tilde{x}_i$  la  $i^{\text{ème}}$  observable,  $\sigma_{x_i}$  son erreur observationnelle et  $x_i$  qui est le  $i^{\text{ème}}$  point des isochrones ou des trajets évolutifs qui correspond aux paramètres  $\theta$ .

Pour déterminer la fonction de densité de probabilité *a posteriori* en fonction de l'âge, notée  $\psi(\tau|x)$ , nous intégrons l'équation 2.3 en fonction de la métallicité et de la masse de la façon suivante :

$$\begin{aligned} \psi(\tau|x) &= \lambda(t) \int_{[M/H]} \left[ \phi([M/H]) \int_m \xi(m) \right. \\ &\quad \left. \times \mathcal{L}(\widetilde{\log T_{\text{eff}}}, \widetilde{M_V}, \widetilde{[M/H]} | [M/H], m, \tau) dm \right] d[M/H]. \end{aligned} \quad (2.6)$$

Nous pouvons également déterminer la masse et la métallicité en intégrant l'équation 2.3 en fonction des deux autres paramètres. Les déterminations de la masse et de la métallicité sont décrites dans la partie 2.6.

Notre méthode d'inférence bayésienne est basée sur le travail de Jørgensen & Lindegren (2005) et de da Silva et al. (2006). Les deux intégrales peuvent être simplifiées comme une double somme en fonction de la métallicité et de la masse sur tous les trajets évolutifs ou les isochrones. Nous obtenons alors :

$$\begin{aligned} \psi(\tau_k|x) &= \lambda(t_k) \times \sum_i \left[ \int_{[M/H]_i^{ijk}}^{[M/H]_i^{ij(k+1)}} \phi([M/H]_i) d[M/H] \right] \sum_j \\ &\quad \times \left[ \int_{m_j^{ijk}}^{m_j^{ij(k+1)}} \xi(m_j) dm \right] \mathcal{L}(\widetilde{\log T_{\text{eff}}}, \widetilde{M_V}, \widetilde{[M/H]} | [M/H], m, \tau). \end{aligned} \quad (2.7)$$

Les indices  $ijk$  représentent le point de l'isochrone ou du trajet évolutif de métallicité  $i$ , de masse initiale  $j$  et d'âge  $k$ . L'équation 2.7 est pour les métallicités des trajets évolutifs ou des isochrones dans l'intervalle  $[M/H] \pm 3\sigma_{[M/H]}$ . Nous utilisons cette simplification afin de gagner du temps de calcul tout en prenant en compte entièrement l'erreur observationnelle sur la métallicité.

Pour ne pas calculer l'âge d'une étoile qui est trop éloignée des isochrones ou des trajets évolutifs nous utilisons une condition de rejet sur

le  $\chi^2$  (Jørgensen & Lindegren 2005). L'âge d'une étoile n'est pas calculé si  $\min(\chi^2) > \chi_{\max}^2 = 11.34$ .  $\chi_{\max}^2$  est défini en respectant la distribution du  $\chi^2$  pour trois paramètres avec un niveau de confiance de 0.01 (table C.2).

Pour les *a priori*, nous prenons la fonction de masse initiale de Kroupa (2002), un taux de formation stellaire constant et tronqué et une fonction de distribution de la métallicité constante. Ces choix sont décrits dans la partie 2.2. Dans la littérature, la majorité des méthodes utilisent les isochrones ; nous avons choisi de travailler avec les trajets évolutifs (partie 2.3). Nous utilisons les trajets évolutifs de BaSTI (Pietrinferni et al. 2004) avec la résolution de la grille décrite dans la partie 2.3. Enfin, nous appliquons un traitement sur la fonction  $\psi(\tau|x)$  pour corriger la méthode de certain biais (partie 2.4).

Pour l'implémentation avec les trajets évolutifs, un tableau contenant les âges pour lesquels nous souhaitons calculer la fonction  $\psi(\tau|x)$  est créé. Ensuite nous interpolons dans les trajets évolutifs en fonction de ce tableau afin d'obtenir les paramètres  $x_i([M/H], m, \tau)$ .

La fonction  $\psi(\tau|x)$  obtenue donne la probabilité que l'âge  $\tau$  corresponde à l'âge réel de l'étoile. Elle est normalisée par son maximum pour déterminer l'incertitude sur l'âge. Pour déterminer l'âge le plus probable  $\hat{\tau}$  de la fonction  $\psi(\tau|x)$ , la moyenne, la médiane ou le mode de degré 1 peuvent être utilisés. La valeur moyenne correspond au centre de gravité de l'aire sous la fonction  $\psi(\tau|x)$ . La valeur médiane correspond à la bissectrice de l'aire sous la fonction  $\psi(\tau|x)$ , et enfin le mode de degré 1 correspond au maximum de la fonction  $\psi(\tau|x)$ . La figure 2.1 montre les différences entre le mode, la moyenne et la médiane pour quatre étoiles pour lesquelles les fonctions  $\psi(\tau|x)$  ont une morphologie différente (partie 2.4). Pour chaque étoile, l'âge est mieux déterminé avec le mode. C'est pourquoi dans notre méthode, le mode de la fonction  $\psi(\tau|x)$ , noté  $\psi(\hat{\tau}|x)$ , est utilisé afin de déterminer l'âge le plus probable des étoiles.

## 2.2 Choix des *a priori*

Pour tester le choix des *a priori*, nous utilisons un catalogue simulé afin de vérifier l'accord entre les âges réels simulés et les âges déterminés avec l'inférence bayésienne présentée dans la partie précédente. Ce catalogue est construit à partir des trajets évolutifs de BaSTI (Pietrinferni et al. 2004), de la fonction de masse initiale de Kroupa (2002), d'un taux de formation stellaire constant et tronqué (partie 2.2.3) et d'une fonction de distribution de la métallicité (partie 2.2.1) basée sur un catalogue de données d'observation (Casagrande et al. 2011). La construction et les spécificités de ce catalogue sont décrites dans la partie 4.

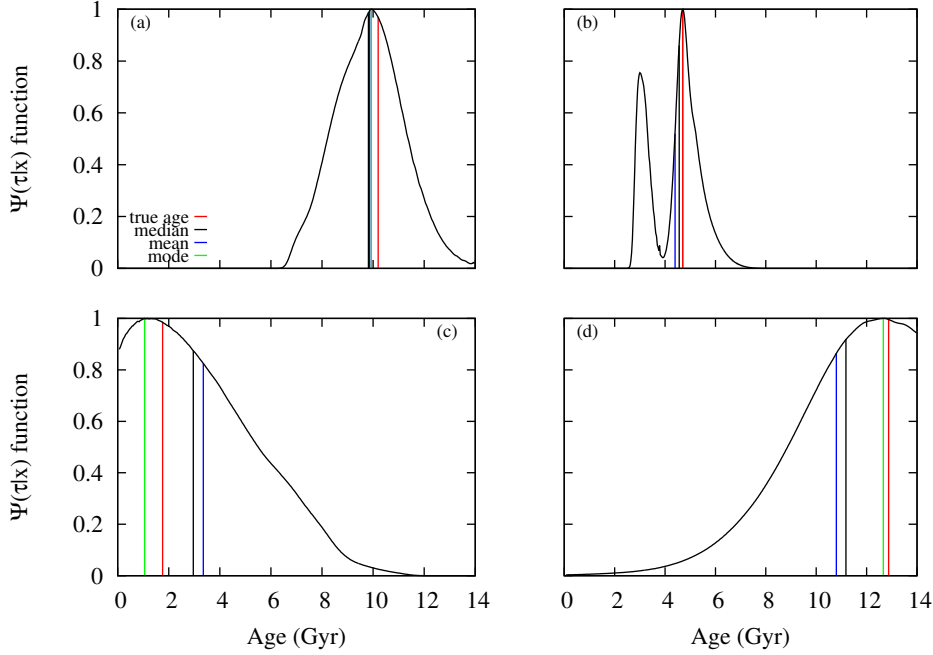


FIGURE 2.1 – Figure représentant l’âge réel simulé de l’étoile ainsi que le mode, la moyenne et la médiane pour quatre morphologies de fonctions  $\psi(\tau|x)$  (partie 2.4). Dans le cas de la figure (b) l’âge réel simulé de l’étoile et le mode sont confondus.

### 2.2.1 Fonction de distribution de la métallicité

Les étoiles de métallicité solaire sont les plus courantes dans le voisinage solaire et la prise en compte d’une fonction de distribution de la métallicité  $\phi([M/H])$  (MDF, metallicity distribution function) permet de favoriser ces métallicités. Nous utilisons la distribution des métallicités des étoiles du Geneva Copenhagen Survey (GCS, Casagrande et al. 2011, plus de détails sur ce catalogue sont donnés dans la partie 8.1.1) afin de construire la MDF. C’est un relevé de 16 682 étoiles du voisinage solaire.

L’histogramme des métallicités, représenté sur la figure 2.2, a deux composantes : une correspondant au disque mince (avec un pic à 0 dex) et une au disque épais (pic à -0.4 dex). Nous négligeons la composante du disque épais pour construire la MDF car elle représente une faible contribution d’étoiles du voisinage solaire. En ajustant la composante du disque mince par une gaussienne (en rouge, Fig. 2.2) nous avons obtenu la MDF suivante :

$$\phi([M/H]) = \frac{1}{0.15 \times \sqrt{2\pi}} \exp \left[ -\frac{1}{2} \left( \frac{[M/H]}{0.15} \right)^2 \right]. \quad (2.8)$$

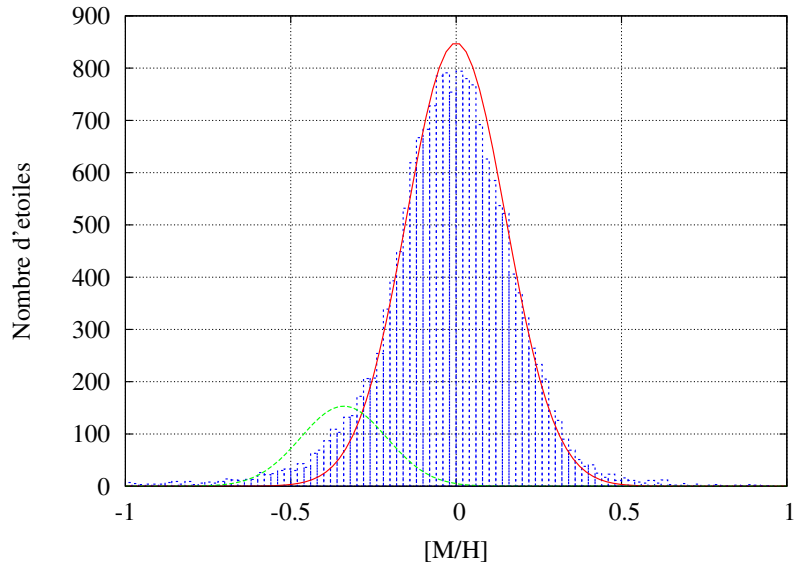


FIGURE 2.2 – Histogramme des métallicités des étoiles du Geneva Copenhagen Survey de Casagrande et al. (2011).

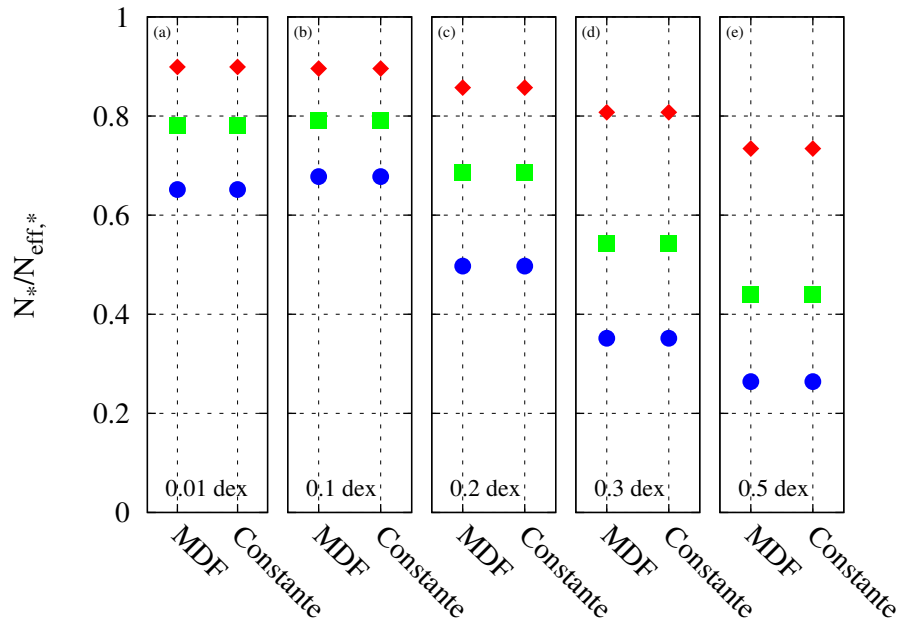


FIGURE 2.3 – Résultats de la datation des étoiles du catalogue simulé de référence pour différentes MDF.  $N_{eff,*}$  représente le nombre total d'étoiles d'âge déterminé.  $N_*$  est le nombre d'étoiles avec  $\Delta\tau/\tau < 10\%$  (cercle bleu),  $< 20\%$  (carré vert) et  $< 50\%$  (losange rouge).

Nous avons déterminé l'âge des étoiles, avec la MDF choisie et une MDF constante, pour les étoiles du catalogue simulé mais avec différentes erreurs observationnelles sur la métallicité (0.01, 0.1, 0.2, 0.3 et 0.5 dex). Nous trouvons que la MDF n'a aucun effet quelle que soit l'erreur observationnelle sur la métallicité. Nous avons choisi d'inclure une MDF constante pour la détermination de l'âge car celle-ci a peu d'effet sur les âges.

### 2.2.2 Fonction de masse initiale

Dans notre Galaxie, les étoiles de faibles masses sont beaucoup plus courantes que les étoiles massives, la fonction de masse initiale  $\xi(m)$  (IMF, Initial Mass Function) permet de décrire cette tendance. La fonction de masse initiale donne le nombre d'étoiles  $dn$  par unité de masse  $dm$  en fonction de la masse initiale  $m$ , elle est notée  $\xi(m) = dn/dm$ .

Nous souhaitons ajouter une IMF dans la détermination de l'âge afin de donner un poids plus grand aux trajets évolutifs les moins massifs. Il existe plusieurs IMFs dans la littérature, nous en avons considéré sept. Certaines d'entre elles sont exprimées en fonction du logarithme de la masse  $\log m$  mais elles peuvent être reliées à l'IMF en fonction de la masse  $m$  grâce à la relation suivante  $\xi(m) = \xi(\log m)/(m \ln 10)$ . Les IMFs sont généralement exprimées en loi de puissance telle que  $\xi(m) \propto m^{-\alpha}$  et  $\xi(\log m) \propto m^{-x}$  avec  $x = 1 - \alpha$ . Pour la description des IMFs, nous avons gardé les notations des articles mais pour les utiliser nous les exprimons toutes en fonction de  $\xi(m)$ .

#### 1. Salpeter (1955, S55)

S55 est le premier à avoir introduit l'IMF, elle est déterminée grâce à la fonction de luminosité observée  $\Phi_t(M_V) = dn/dM_V$  qui donne le nombre d'étoiles  $dn$  pour un intervalle de magnitude absolue  $dM_V$  donné à un temps  $t$ . La fonction de luminosité est reliée à la masse grâce à la relation masse-luminosité :  $L \propto m^3$ . S55 calcule la fonction de luminosité pour les étoiles de la séquence principale. S55 définit également la fonction de luminosité initiale :  $\Psi(M_V) = \xi(m)d(\log_{10} m)/dM_V$  qui dépend de l'IMF à une magnitude absolue fixée. Elle donne la fonction de luminosité pour la masse initiale des étoiles. Ces deux fonctions de luminosité peuvent être reliées, cela permet d'obtenir l'IMF. S55 établit son IMF entre 0.4 et 10  $M_\odot$  telle que

$$\xi_{\text{S55}}(\log m) = 0.03 \left( \frac{m}{M_\odot} \right)^{-1.35} \quad (2.9)$$

## 2. Miller & Scalo (1979, MS79)

La méthode utilisée par MS79 est différente. La fonction de masse actuelle  $\Xi_{\text{ms}}(\log m)$  (PDMF, Present day Mass Function) est déterminée empiriquement pour les étoiles de la séquence principale. La PDMF donne le nombre d'étoiles  $dn$  par unité de masse actuelle. La PDMF est reliée à la fonction de luminosité observée par la relation  $\Xi_{\text{ms}}(\log m) = \Phi(M_V) \times \left| \frac{dM_V}{d \log m} \right| \times 2H(M_V) \times f_{\text{ms}}(M_V)$ . La pente de la relation magnitude absolue-masse  $\frac{dM_V}{d \log m}$  permet de convertir la fonction de luminosité en fonction de masse.  $2H(M_V)$  est le résultat de l'intégration de la fonction de luminosité perpendiculairement au plan de la Galaxie.  $f_{\text{ms}}(M_V)$  donne la fraction d'étoiles de la séquence principale pour une magnitude absolue donnée.

L'IMF est reliée à la PDMF par la relation  $b(T_0)$  tel que  $\xi(m) = \Xi_{\text{ms}}(\log m)/T_{\text{ms}} \times T_0/b(T_0)$ , où  $b(T_0)$  est le taux de formation d'étoiles,  $T_{\text{ms}}$  est le temps passé sur la séquence principale et  $T_0$  l'âge de l'univers. MS79 prennent plusieurs taux de formation d'étoiles. Nous utilisons l'IMF avec  $b(T_0)$  de forme exponentielle d'où  $b(T_0) = B \exp(T_0/\tau)$  où  $B = T_0/(\tau \times [\exp(T_0/\tau) - 1])$  et  $\tau$  une constante de temps. Pour obtenir l'IMF, MS79 ajustent  $b(T_0)$  par un filtre en loi de puissance, un filtre quadratique et un filtre semi-gaussien. L'IMF obtenue dépend donc du filtre utilisé, de l'âge de la Galaxie  $T_0$  et de la forme du taux de naissance d'étoiles. Nous avons choisi de travailler avec un filtre en loi de puissance,  $T_0 = 12$  Gyr et  $B$  constant. Dans ce cas  $\tau = 7.5$  et l'IMF est définie par :

$$\xi_{\text{MS79}}(\log m) \begin{cases} 42 \times m^{-0.4} & \text{si } 0.1 M_{\odot} \leq m \leq 1.0 M_{\odot}, \\ 42 \times m^{-1.5} & \text{si } 1.0 M_{\odot} < m \leq 10.0 M_{\odot}, \\ 240 \times m^{-2.3} & \text{si } m > 10.0 M_{\odot}. \end{cases} \quad (2.10)$$

## 3. Kroupa et al. (1993, KTG93)

La méthode utilisée par KTG93 est semblable à celle de S55. A partir d'un échantillon d'étoiles du voisinage solaire, KTG93 déterminent la fonction de luminosité observée. La masse de chaque étoile est déterminée grâce à une relation empirique masse-magnitude absolue. De façon indépendante, KTG93 déterminent plusieurs types de relation masse-magnitude absolue théorique grâce à un générateur de masses aléatoires. KTG93 comparent ces relations avec la relation masse-magnitude absolue empirique ainsi la relation la plus probable est obtenue. Les paramètres de l'IMF sont déterminés à partir de cette relation et d'une PDMF pour compenser l'évolution stellaire (basée sur MS79). L'IMF

obtenue est définie par :

$$\xi_{\text{KTG93}}(m) = \begin{cases} 0.035 \times m^{-1.3} & \text{si } 0.08 M_{\odot} \leq m < 0.5 M_{\odot}, \\ 0.019 \times m^{-2.2} & \text{si } 0.5 M_{\odot} \leq m < 1.0 M_{\odot}, \\ 0.019 \times m^{-2.7} & \text{si } m \geq 1.0 M_{\odot}. \end{cases} \quad (2.11)$$

#### 4. Scalo (1998, S98)

S98 représente le nombre d'étoiles  $n$  en fonction de la masse  $m$  pour 75 amas. Ensuite S98 ajuste la pente de cette fonction afin d'obtenir  $x = 1 - \alpha = d \log n / d \log m$ . Cette méthode est appelée le  $\alpha$ -plot. L'IMF est définie selon une loi de puissance  $m^x$  avec  $x = d \log \xi(\log m) / d \log m$ . S98 effectue des ajustements différents selon l'intervalle de masse :

$$\xi_{\text{S98}}(\log m) \propto \begin{cases} m^{-0.2 \pm 0.3} & \text{si } 0.1 M_{\odot} \leq m < 1.0 M_{\odot}, \\ m^{-1.7 \pm 0.5} & \text{si } 1.0 M_{\odot} \leq m < 10.0 M_{\odot}, \\ m^{-1.3 \pm 0.5} & \text{si } 10.0 M_{\odot} \leq m < 100.0 M_{\odot}. \end{cases} \quad (2.12)$$

#### 5. Kroupa (2001, K01)

K01 utilise également la méthode du  $\alpha$ -plot de S98 mais pour un échantillon d'étoiles du voisinage solaire plus important. L'IMF obtenue est définie par :

$$\xi_{\text{K01}}(m) \propto \begin{cases} m^{-0.3 \pm 0.7} & \text{si } 0.01 M_{\odot} \leq m < 0.08 M_{\odot}, \\ m^{-1.8 \pm 0.5} & \text{si } 0.08 M_{\odot} \leq m < 0.5 M_{\odot}, \\ m^{-2.7 \pm 0.3} & \text{si } 0.5 M_{\odot} \leq m < 1.0 M_{\odot}, \\ m^{-2.3 \pm 0.7} & \text{si } m \geq 1.0 M_{\odot}. \end{cases} \quad (2.13)$$

#### 6. Kroupa (2002, K02)

La méthode utilisée est semblable à celle de KTG93 mais elle est appliquée à un échantillon d'étoiles plus grand. La PDFM  $\Xi(m)$  est également construite différemment. Le nombre d'étoiles  $dN$  par  $pc^3$  pour un intervalle de masse  $m, m + dm$  est supposé égal à  $dN = \Xi(m)dm$ . Le nombre d'étoiles par  $pc^3$  pour un intervalle de magnitude absolue  $M_V, M_V + dM_V$  est supposé égal à  $dN = -\Phi(M_V)dM_V$ , où  $\Phi$  est la fonction de luminosité. Donc  $\Xi(m) = -\Phi(M_V)(dm/dM_V)^{-1}$  peut être calculé grâce à la relation masse-luminosité. Le nombre d'étoiles  $dN$  par  $pc^3$ , pour un intervalle de masse  $m, m + dm$ , formé dans un intervalle de temps  $t, t + dt$  est  $dN = \xi(m, t)dm \times b(t)dt$ . Ici  $b(t)$  représente l'histoire de la formation stellaire (SFH, stellar formation history), avec  $(1/\tau_G) \int_0^{\tau_G} b(t)dt = 1$ , et où  $\tau_G$  est l'âge de la Galaxie. La PDMF s'écrit donc :

$$\Xi(m) = \xi(m) \frac{1}{\tau_G} \times \begin{cases} \int_{\tau_G - \tau(m)}^{\tau_G} b(t)dt & \text{si } \tau(m) < \tau_G, \\ \int_0^{\tau_G} b(t)dt & \text{si } \tau(m) \geq \tau_G. \end{cases} \quad (2.14)$$

Dans cette expression,  $\tau(m)$  est la durée de vie sur la séquence principale. Pour les étoiles de faible masse  $\Xi(m) = \xi(m)$  car elles n'ont pas quitté la séquence principale. Pour un SFH constant  $b(t) = 1$ , l'IMF obtenue est :

$$\xi_{\text{K02}}(m) = 0.877 \pm 0.045 \times \quad (2.15)$$

$$\left\{ \begin{array}{l} \left(\frac{m}{m_1}\right)^{-\alpha_0} \quad \text{si } 0.01 M_{\odot} \leq m < 0.08 M_{\odot}, \\ \left(\frac{m}{m_1}\right)^{-\alpha_1} \quad \text{si } 0.08 M_{\odot} \leq m < 0.5 M_{\odot}, \\ \prod_{i=2}^{n>2} \left[ \left(\frac{m_i}{m_{i-1}}\right)^{-\alpha_{i-1}} \right] \times \left(\frac{m}{m_n}\right)^{-\alpha_n}, \end{array} \right.$$

avec  $\alpha_0 = 0.3 \pm 0.7$ ,  $\alpha_1 = 1.3 \pm 0.5$ ,  $m_0 = 0.01 M_{\odot}$ , et  $m_1 = 0.08 M_{\odot}$ ,  
pour  $n = 2$ ,  $\alpha_2 = 2.3 \pm 0.3$  si  $m_2 = 0.5 M_{\odot} \leq m < 1.0 M_{\odot}$ ,  
pour  $n = 3$ ,  $\alpha_3 = 2.7 \pm 0.3$  si  $m_3 = 1.0 M_{\odot} \leq m$ .

### 7. Chabrier (2003, C03)

C03 utilise la même méthode mais en prenant une forme différente pour l'IMF et en prenant une PDMF basée sur la théorie de formation stellaire. L'IMF est déterminée par C03 pour le disque galactique, le bulbe et le halo. Dans notre cas nous avons choisi de prendre l'IMF du disque galactique car elle correspond au voisinage solaire.

$$\xi_{\text{C03}}(\log m) = \begin{cases} 0.158 \times \exp \left[ \frac{-(\log m - \log(0.079))^2}{2 \times 0.69^2} \right] & \text{si } m \leq 1.0 M_{\odot}, \\ 4.43 \cdot 10^{-2} \times m^{-1.3} & \text{si } m > 1.0 M_{\odot}. \end{cases} \quad (2.16)$$

Les IMFs sont représentées sur la figure 2.4, elles sont toutes normalisées à 1 à 1 masse solaire afin de les visualiser sur la même figure. Nous utilisons l'IMF de K02 pour le calcul de l'âge et pour la construction du catalogue simulé car elle ajuste correctement les observations du voisinage solaire (Kroupa 2002).

Afin d'analyser les effets sur les âges, nous déterminons l'âge des étoiles du catalogue simulé avec toutes les IMFs de la littérature ainsi qu'une IMF constante. Pour comparer les résultats nous utilisons l'erreur relative entre l'âge déterminé et l'âge réel simulé de l'étoile telle que  $\Delta\tau/\tau = (\tau_{\text{simulé}} - \tau_{\text{calculé}})/\tau_{\text{simulé}}$ . Les résultats de la comparaison des différentes IMFs sont présentés sur la figure 2.5. Nous constatons que la détermination de l'âge est meilleure avec une IMF de la littérature qu'avec l'IMF constante. En effet nous trouvons 40 % des étoiles avec  $\Delta\tau/\tau < 20$  % pour une IMF constante alors qu'en moyenne il y a 65 % des étoiles avec  $\Delta\tau/\tau < 20$  % pour une IMF de



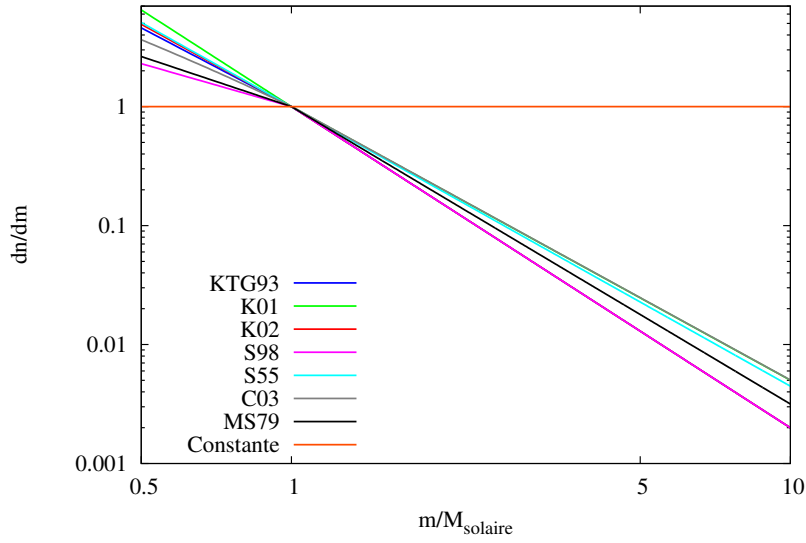


FIGURE 2.4 – Représentation de sept IMFs étudiées et de l’IMF constante. Les IMFs sont normalisées à 1 à 1 masse solaire afin de toutes les visualiser sur la même figure.

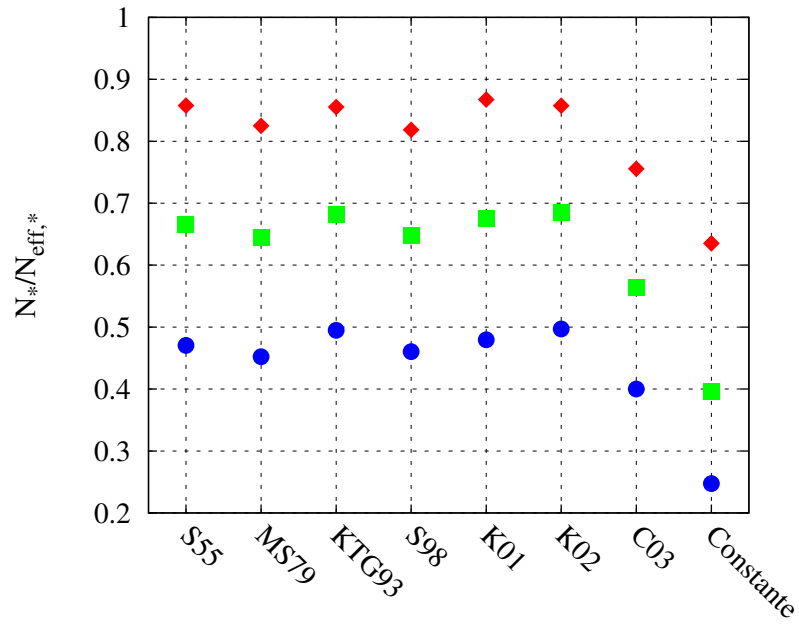


FIGURE 2.5 – Résultats de la datation des étoiles du catalogue simulé de référence pour différentes IMFs. Les couleurs sont les mêmes que celles de la figure 2.3.

la littérature. L'inclusion d'une IMF est donc préférable pour la détermination des âges.

Nous constatons également que les IMFs de KTG93 et K01, qui sont similaires à celle du catalogue simulé (K02), donnent des résultats similaires. Nous trouvons 68 % des étoiles avec  $\Delta\tau/\tau < 20\%$ . Par contre lorsque l'IMF utilisée est légèrement différente de celle de K02, nous obtenons une moins bonne détermination des âges. Avec l'IMF de C03, il y a 56 % des étoiles avec  $\Delta\tau/\tau < 20\%$ , car l'IMF de C03 est différente de celle de K02. En effet, sur la figure 2.4, nous constatons que l'IMF de C03 favorise moins les petites masses et plus les grandes masses que celle de K02.

Les étoiles qui ont des résultats différents sont situées dans les mêmes régions du diagramme HR (partie 5.1). Le choix de l'IMF est donc important pour la détermination des âges, dans notre méthode nous utilisons l'IMF de K02.

### 2.2.3 Taux de formation stellaire

Le taux de formation stellaire  $\lambda(t)$  (SFR, Stellar Formation Rate) décrit le nombre d'étoiles  $n$  formées en fonction du temps  $t$  qui s'est écoulé depuis la formation de l'Univers. Pour la détermination de l'âge, nous incluons un taux de formation stellaire, ceci permet de ne pas attribuer d'âge supérieur à celui de l'Univers. En effet les âges des trajets évolutifs vont de 0 à 22 Gyr (pour les trajets évolutifs des étoiles les moins massives). Le SFR est défini par :

$$\lambda(t) = \begin{cases} 1 & \text{si } t_1 = 0 \leq t \leq t_2 = 14 \text{ Gyr} , \\ 0 & \text{si } t > 14 \text{ Gyr} \end{cases} \quad (2.17)$$

Avec  $t_1$  et  $t_2$  les âges limites du SFR.

Nous déterminons l'âge des étoiles du catalogue simulé avec et sans SFR afin de vérifier son impact sur la datation. Les âges déterminés sont similaires dans les deux cas, il y a moins de 15 étoiles qui ont un âge différent. Les âges déterminés avec le SFR sont légèrement meilleurs. Nous gardons le SFR constant et tronqué en *a priori* dans la datation des étoiles car il correspond à une limite physique de l'âge de l'univers.

## 2.3 Modèles d'évolution stellaire

Les codes d'évolution stellaire (Kippenhahn et al. 2013) calculent les modèles d'évolution stellaire pour une fraction de masse initiale d'éléments lourds  $Z_0$  et une masse initiale  $m_0$ . Les codes d'évolution stellaire déterminent, par pas de temps successifs, l'évolution de l'étoile en calculant les variations de

la composition chimique puis les paramètres stellaires (rayon, température effective, gravité, luminosité, ...). Pour cela les codes d'évolution stellaire ont besoin de connaître le paramètre de convection<sup>1</sup>  $\alpha_{\text{conv}}$  ainsi que la fraction de masse initiale de l'hélium  $Y_0$ . Des relations permettent de relier la métallicité  $[M/H]$  aux fractions de masse de l'hydrogène  $X$ , de l'hélium  $Y$  et des éléments lourds  $Z$  :

$$[M/H] = \log \left( \frac{Z}{X} \right) - \log \left( \frac{Z}{X} \right)_{\odot}, \quad (2.18)$$

$$X + Y + Z = 1, \quad (2.19)$$

$$Y = Y_p + \frac{\Delta Y}{\Delta Z} Z, \quad (2.20)$$

où  $\Delta Y/\Delta Z$  est l'enrichissement relatif moyen de l'hélium par rapport aux éléments lourds de la Galaxie, il est calibré sur le Soleil.  $Y_p$  (Peimbert et al. 2007) est l'abondance d'hélium primordiale et  $(Z/X)_{\odot}$  (Grevesse & Noels 1993) est la proportion des éléments lourds solaire par rapport à l'hydrogène solaire.  $Y$  étant la fraction de masse initiale de l'hélium de l'étoile.

Les modèles d'évolution stellaire de BaSTI (Pietrinferni et al. 2004) sont utilisés dans notre travail comme modèles de référence pour la détermination de l'âge des étoiles. Les fractions de masse des éléments lourds  $Z$  de la grille sont comprises entre 0.0001 et 0.04. La fraction de masse des métaux par rapport à l'hélium est de  $\Delta Y/\Delta Z = 1.4$  et l'abondance d'hélium primordiale est de  $Y_p = 0.245$  (Cassisi et al. 2003). Pietrinferni et al. (2004) ont pris un mélange des éléments lourds solaire avec  $(Z/X)_{\odot} = 0.0245$  (Grevesse & Noels 1993). Ces modèles ne prennent pas en compte la diffusion microscopique et sont proposés dans deux versions l'une avec et l'autre sans overshooting des cœurs convectifs. Nous utilisons ici les modèles d'évolution stellaire sans overshooting. Les masses sont comprises entre 0.75 et 10  $M_{\odot}$  avec deux régimes de perte de masse  $\eta$  :  $\eta = 0.2$  pour  $0.75 \leq m < 2.6 M_{\odot}$  et  $\eta = 0.4$  pour  $2.6 \leq m < 10 M_{\odot}$  (Reimers 1975). Les trajets évolutifs couvrent les stades d'évolution de la phase d'allumage de la combustion de l'hydrogène central jusqu'à la fin de la combustion de l'hélium sur la branche horizontale.

Nous utilisons aussi les modèles d'évolution stellaire de CESAM (Morel & Lebreton 2008) pour effectuer des comparaisons. Ces modèles stellaires ne prennent pas en compte l'overshooting des cœurs convectifs, mais la diffusion microscopique de l'hélium et des éléments lourds est incluse. Les valeurs solaires initiales adoptées sont  $Y_{\odot} = 0.2674$ ,  $Z_{\odot} = 0.0175$  et  $(\Delta Y/\Delta Z)_{\odot} = 1.9$  (Lebreton et al. 2001). Les trajets évolutifs couvrent les stades d'évolution du début de la phase d'allumage de la combustion de l'hydrogène central jusqu'au

1. Longueur de mélange (Böhm-Vitense 1958).

milieu de la branche des géantes rouges. Nous disposons de cinq métallicités  $[M/H]$  allant de -0.2 à 0.2 dex et des masses comprises entre 0.8 et  $10 M_{\odot}$ . Dans chacun de ces trajets évolutifs il y a 81 âges.

Les modèles d'évolution stellaire de Padoue (Bertelli et al. 1994; Girardi et al. 1996, 2000) prennent en compte l'overshooting et le mélange des éléments par diffusion microscopique. Les valeurs solaires adoptées sont  $X_{\odot} = 0.708$ ,  $Y_{\odot} = 0.273$  et  $Z_{\odot} = 0.019$ . L'abondance en hélium primordiale  $Y_p$  est de 0.23 et  $(\Delta Y/\Delta Z)_{\odot} = 2.25$ . Les isochrones de Padoue couvrent les stades d'évolution de la phase d'allumage de la combustion de l'hydrogène central jusqu'à la fin de la combustion de l'hélium sur la branche horizontale. La grille disponible sur internet<sup>1</sup> va de la métallicité  $Z = 0.0004$  à  $Z = 0.3$  pour des masses comprises entre 0.15 et  $7.0 M_{\odot}$ .

### 2.3.1 Résolution de la grille des trajets évolutifs

La grille des trajets évolutifs de BaSTI, disponible sur internet<sup>2</sup>, comprend 11 métallicités, 34 masses et 2250 points (âges). Cette grille, que nous appelons grille 0, n'est pas assez fine pour pouvoir déterminer correctement les âges. En effet sur la figure 2.6.a, nous constatons qu'il y a seulement 28 % des étoiles avec un âge bien déterminé ( $\Delta\tau/\tau < 20$  %) avec cette grille. Une grille plus fine est alors construite à partir d'interpolations linéaires en échelle  $\log - \log$  (annexe B) afin d'obtenir la grille 1 du tableau 2.1. Nous obtenons maintenant 67 % des étoiles qui ont  $\Delta\tau/\tau < 20$  %.

TABLE 2.1 – Résolution des deux grilles obtenues par interpolation linéaire (annexe B).  $\Delta X$  est le pas entre deux valeurs du paramètre  $X$ . La grille 2 est adoptée pour la détermination de l'âge.

$X$	Intervalle	Grille 1 $\Delta X$	Grille 2 $\Delta X$
$[M/H]$ (dex)	-2.26 à 0.39	0.03	0.05
$\tau$ (Gyr)	0.063 à 22	0.0075	0.03
$m$ ( $M_{\odot}$ )	0.75 à 1	0.001	0.005
	1.0 à 1.3	0.001	0.003
	1.3 à 10.0	0.001	0.002

Le calcul avec cette grille est très long (tableau 2.2), il faut plus de quatre jours pour déterminer l'âge de 10 000 étoiles. Les calculs sont réalisés sur un

1. Padoue : <http://pleiadi.oapd.inaf.it/>

2. BaSTI : <http://193.204.1.62/index.html>

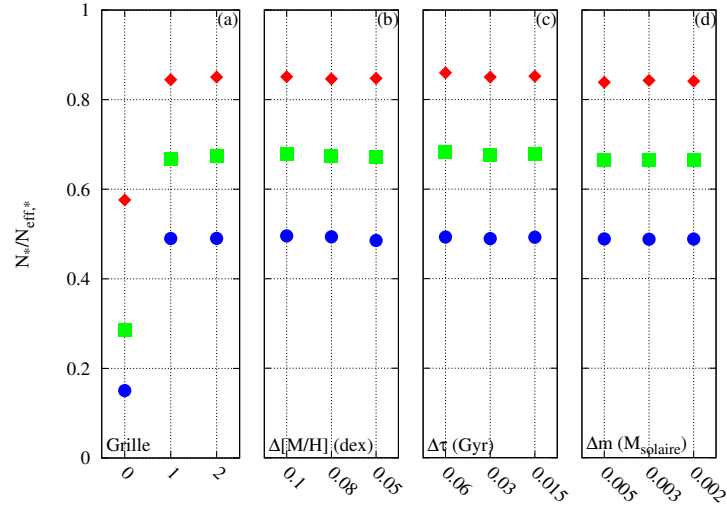


FIGURE 2.6 – Résultats de la datation des étoiles du catalogue simulé de référence pour différentes résolutions de la grille des trajets évolutifs de BaSTI (figure a). Les figures b, c et d présentent les résultats pour différentes résolutions en métallicité, âge et masse. Les couleurs sont les mêmes que celles de la figure 2.3.

TABLE 2.2 – Temps de calcul (en heures) et pourcentage de fonctions mal-définies pour différentes résolutions de la grille. Ces calculs sont réalisés sur les 10 000 étoiles du catalogue simulé. Pour les résolutions en fonction de  $\Delta[M/H]$ ,  $\Delta m$  et  $\Delta\tau$  les autres paramètres sont fixés à la résolution de la grille 1. Tous ces calculs ont été réalisés sur le même serveur.

Résolution	Temps de calcul	Fonctions mal-définies
Grille 0	0.23 h	40 %
Grille 1	100 h	35 %
Grille 2	10 h	34 %
$\Delta[M/H] = 0.1$ dex	37 h	44 %
$\Delta[M/H] = 0.08$ dex	44 h	41 %
$\Delta[M/H] = 0.05$ dex	75 h	36 %
$\Delta m = 0.005 M_{\odot}$	23 h	34 %
$\Delta m = 0.003 M_{\odot}$	39 h	34 %
$\Delta m = 0.002 M_{\odot}$	61 h	33 %
$\Delta\tau = 0.06$ Gyr	14 h	34 %
$\Delta\tau = 0.03$ Gyr	34 h	34 %
$\Delta\tau = 0.015$ Gyr	50 h	35 %

serveur qui possède deux processeurs Intel à 2.33 GHz et chacun de ces processeurs contient quatre cœurs et 16 Go de RAM. Afin de gagner du temps de calcul et de la taille mémoire, nous souhaitons réduire la finesse de cette grille. Pour cela nous effectuons plusieurs déterminations d'âges avec différentes résolutions en masse, en âge et en métallicité. Lorsque nous testons un paramètre les autres sont fixés à la résolution de la grille 1.

On utilise une grille qui va seulement jusqu'à  $2.6 M_{\odot}$  car la grille étendue à  $10 M_{\odot}$  a été calculée après ce travail.

### 2.3.1.1 Résolution en métallicité

Nous étudions, dans un premier temps, la résolution de la grille en métallicité en s'intéressant à trois autres résolutions possibles :  $\Delta[M/H] = 0.05$  dex,  $0.08$  dex et  $0.1$  dex. Nous comparons les âges déterminés avec ces résolutions par rapport à la résolution de la grille 1 ( $\Delta[M/H] = 0.03$  dex). Nous constatons, sur la figure 2.7, que plus le pas en métallicité est grand plus la dispersion est importante entre les âges déterminés avec la résolution de la grille 1 et ceux avec des pas plus grands. Par contre, nous obtenons des résultats similaires (figure 2.6.b) sur le pourcentage d'étoiles avec un âge bien déterminé, soit  $N_{\star}/N_{eff,\star} = 67\%$  avec  $\Delta\tau/\tau < 20\%$ .

$N_{eff,\star}$  correspond au nombre total d'étoiles dont l'âge est déterminé. Il y a deux cas pour lesquels l'étoile ne peut être datée : soit l'étoile ne passe pas le test de rejet sur le  $\chi^2$ , soit la fonction  $\psi(\tau|x)$  est mal-définie (partie 2.4). Le tableau 2.2 donne le pourcentage d'étoiles avec une fonction  $\psi(\tau|x)$  mal-définie et le temps de calcul pour chaque résolution de grille. Plus le pas de la résolution en métallicité est grand, plus il y a de fonctions  $\psi(\tau|x)$  mal-définies.

Pour la datation des étoiles, nous avons choisi de prendre un pas en métallicité égal à  $0.05$  dex car nous obtenons un résultat très similaire à celui de la grille 1 et parce que le pourcentage d'étoiles avec une fonction  $\psi(\tau|x)$  mal-définie est semblable à celui de la grille 1. En prenant ce pas en métallicité, le temps de calcul est réduit de  $25\%$ .

### 2.3.1.2 Résolution en âge

Nous considérons trois pas en âge  $\Delta\tau$  différents :  $0.015$  Gyr,  $0.03$  Gyr et  $0.06$  Gyr. Nous les comparons avec la résolution de la grille 1 ( $0.0075$  Gyr). Comme pour la résolution en métallicité les résultats sont similaires. Par contre, plus nous augmentons le pas de la résolution, moins il y a d'étoiles avec une fonction  $\psi(\tau|x)$  mal-définie (tableau 2.2) et plus la dispersion des âges augmente (figure 2.8). C'est pourquoi nous avons choisi de prendre un pas en âge de  $0.03$  Gyr, ce qui nous permet de gagner  $66\%$  du temps de calcul.

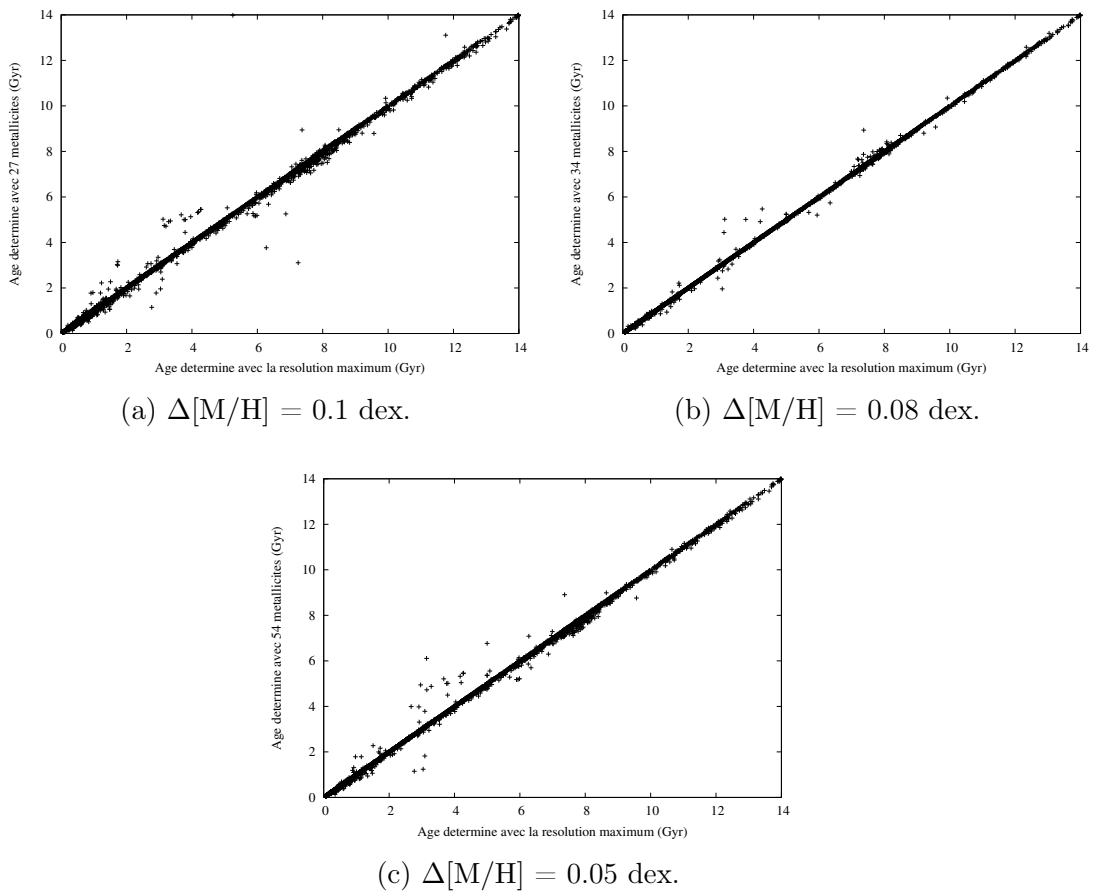


FIGURE 2.7 – Comparaison des différentes résolutions en métallicité  $[M/H]$  par rapport à la résolution de la grille 1 ( $\Delta[M/H] = 0.03$  dex, table 2.1).

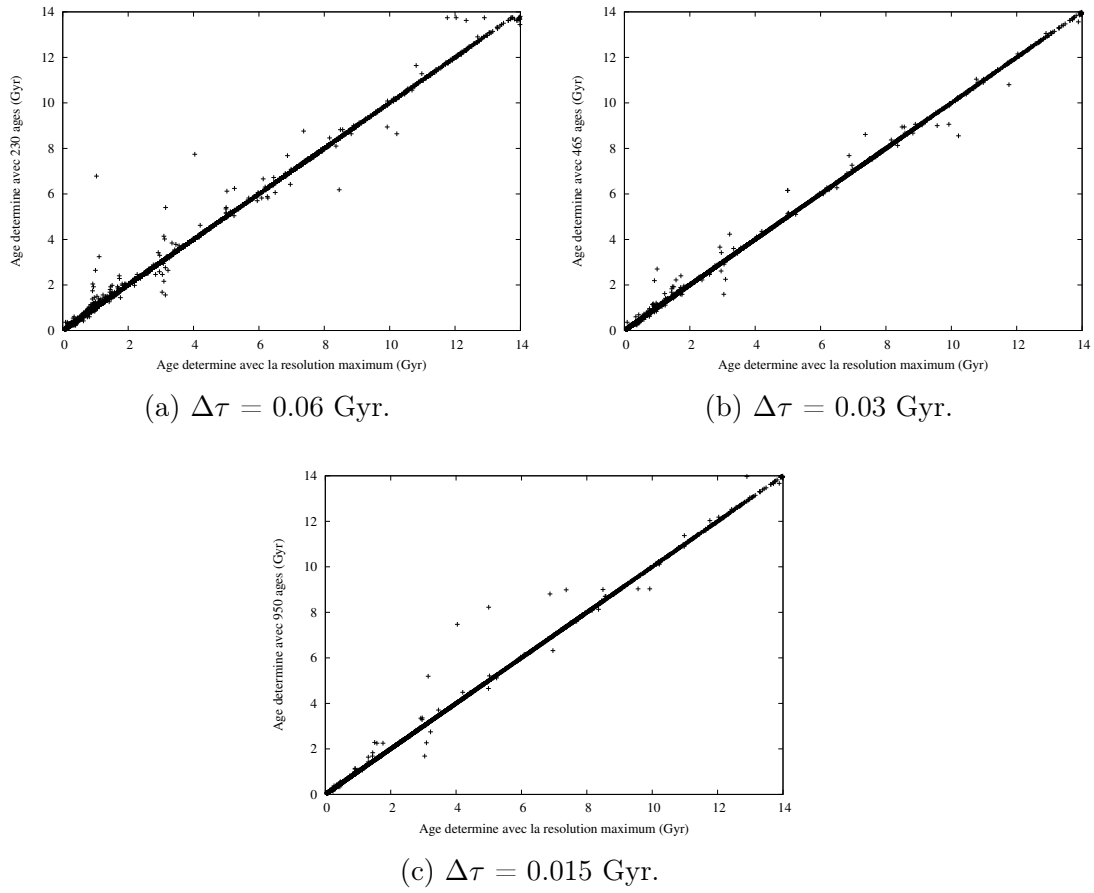


FIGURE 2.8 – Comparaison des différentes résolutions en âge par rapport à la résolution de la grille 1 ( $\Delta\tau = 0.0075$  Gyr, table 2.1).



### 2.3.1.3 Résolution en masse

Nous testons trois résolutions en masse :  $0.002 M_{\odot}$ ,  $0.003 M_{\odot}$  et  $0.005 M_{\odot}$ . Nous constatons, sur la figure 2.9 que plus le pas en masse  $\Delta m$  est grand, plus les âges déterminés sont dispersés par rapport à ceux déterminés avec le pas de la grille 1 ( $0.001 M_{\odot}$ ). De plus, quand le pas est plus grand il y a légèrement plus d'étoiles avec des fonctions mal-définies et le pourcentage d'étoiles avec un âge bien déterminé diminue.

Nous avons besoin que la grille soit fine pour les grandes masses mais le temps de calcul est très long dans ces cas là. C'est pourquoi nous prenons trois pas en masse. Pour les petites masses, inférieures à  $1 M_{\odot}$  nous prenons un pas de  $0.005 M_{\odot}$ , pour des masses entre  $1$  et  $1.3 M_{\odot}$  nous prenons  $0.003 M_{\odot}$  et enfin pour des masses supérieures à  $1.3$  nous prenons  $0.002 M_{\odot}$ .

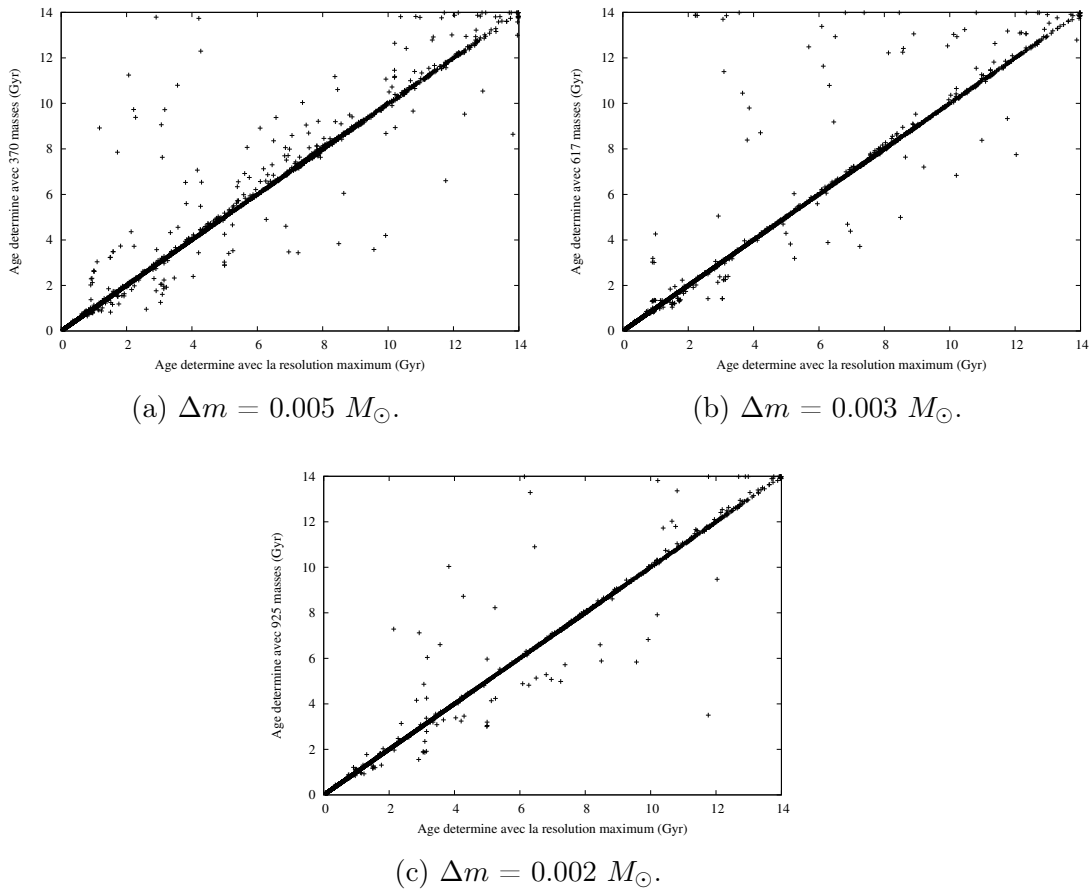


FIGURE 2.9 – Comparaison des différentes résolutions en masse par rapport à la résolution de la grille 1 ( $\Delta m = 0.003 M_{\odot}$ , table 2.1).

La grille que nous adoptons, et qui est décrite dans le tableau 2.1, permet de réduire le temps de calcul de 90 % en ayant des résultats très similaires. Nous utilisons cette grille pour les trajets évolutifs de BaSTI, de CESAM et pour les isochrones de Padoue.

### 2.3.2 Isochrones par rapport aux trajets évolutifs

Nous avons choisi d’adopter les trajets évolutifs plutôt que les isochrones. En effet les isochrones sont construites à partir des trajets évolutifs qui sont issus des codes d’évolution stellaire. En utilisant les trajets évolutifs il y a moins d’interpolations et les erreurs numériques sont réduites. Pour tester l’impact des modèles d’évolution stellaire sur la détermination des âges, il faudrait déterminer les âges avec les isochrones et les trajets évolutifs.

## 2.4 Morphologies de la fonction $\psi(\tau|x)$

L’âge le plus probable  $\hat{\tau}$  correspond au mode de degré 1 de la fonction  $\psi(\tau|x)$  (partie 2.1). Les fonctions  $\psi(\tau|x)$  peuvent prendre plusieurs formes (Jørgensen & Lindegren 2005; Takeda et al. 2007). Nous les avons classifiées en huit catégories (figure 2.10) :

1. un pic,
2. plusieurs pics,
3. oscillations,
4. tronquée aux petits âges sans plateau,
5. tronquée aux petits âges avec plateau,
6. tronquée aux âges élevés sans plateau,
7. tronquée aux âges élevés avec plateau,
8. non datée (strictement croissante ou décroissante)

Nous considérons que les fonctions  $\psi(\tau|x)$  non datées et les fonctions tronquées aux petits âges et aux âges élevés avec plateau sont mal-définies. Nous ne donnons pas d’âge pour les étoiles qui ont ces fonctions car nous ne pouvons pas extraire d’information sur l’âge. Les fonctions  $\psi(\tau|x)$  avec un plateau ont une équiprobabilité des âges sur un grand intervalle, nous ne pouvons donc pas extraire de maximum dans ce cas. Les fonctions  $\psi(\tau|x)$  non datées sont strictement croissantes (ou décroissantes) et le premier (dernier) point correspond au premier (dernier) âge disponible. Les autres fonctions  $\psi(\tau|x)$  sont dites bien-définies.

Le type de morphologie d’une fonction est recherché automatiquement grâce à un traitement numérique.

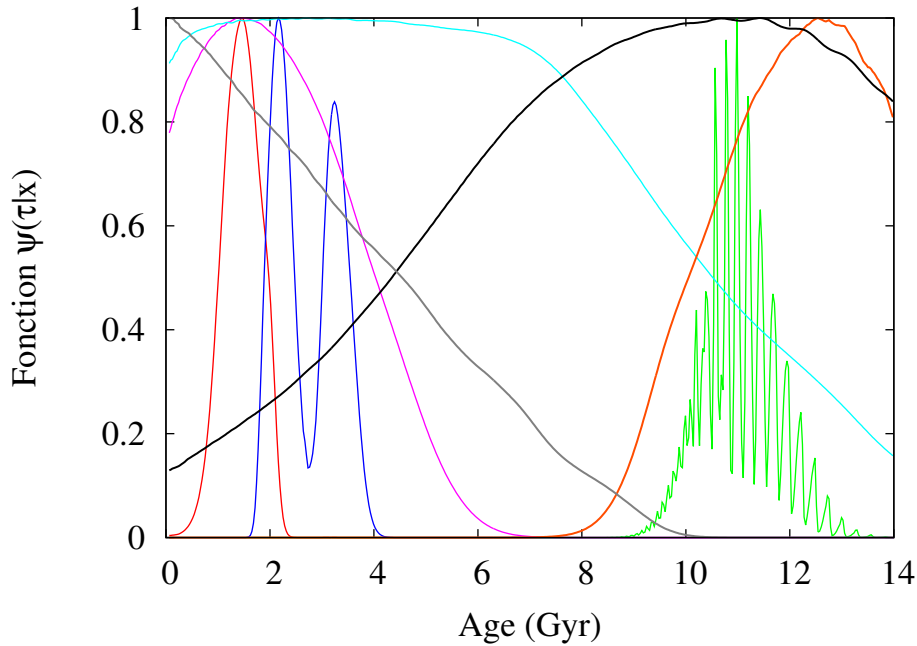


FIGURE 2.10 – Représentation des huit morphologies de la fonction  $\psi(\tau|x)$  : un pic (rouge), plusieurs pics (bleu), oscillations (verte), non daté (gris), tronquée aux petits âges avec plateau (turquoise), tronquée aux petits âges sans plateau (rose), tronquée aux âges élevés (orange) et tronquée aux âges élevés avec plateau (noir).

1. Dans un premier temps, nous vérifions la présence d'oscillations dans la fonction  $\psi(\tau|x)$  grâce au nombre de pics. Pour cela le nombre de changements de signe de la pente est déterminé, entre deux points, sur toute la fonction  $\psi(\tau|x)$ . Les petites fluctuations sont négligées. Si le nombre de pics est supérieur à trois nous considérons que la fonction  $\psi(\tau|x)$  possède des oscillations et nous effectuons un lissage. Nous avons considéré deux approches : la moyenne glissante et la convolution avec un filtre gaussien. Nous étudions dans un premier temps la moyenne glissante. Cette méthode consiste à faire une moyenne sur  $N$  points afin d'obtenir la moyenne d'un point  $n$ . Elle est dite glissante car elle parcourt toute la courbe.

Pour choisir le nombre de points  $N$  utilisés pour effectuer la moyenne, nous effectuons plusieurs tests sur plusieurs fonctions  $\psi(\tau|x)$ . La figure 2.11 montre, pour une fonction  $\psi(\tau|x)$ , la moyenne glissante avec  $N = 5, 7$  et  $9$  points. La moyenne glissante à cinq points ne lisse pas assez la fonction  $\psi(\tau|x)$ , les pics sont toujours présents mais de plus faible amplitude. Avec neuf points, les pics sont inversés par rapport à

la fonction  $\psi(\tau|x)$ , car  $N$  est supérieur au nombre de points décrivant un pic. La moyenne glissante sur sept points donne un bon lissage. Le problème de cette méthode est le choix du nombre de points  $N$  car le lissage en dépend énormément ainsi que du nombre de points décrivant un pic. Comme le nombre de points peut varier d'une étoile à l'autre nous avons choisi de ne pas adopter cette méthode.

Nous utilisons donc la convolution avec un filtre gaussien. Pour un point  $n$  de la fonction à lisser, la convolution est définie par :

$$(\psi \star g)(n) = \sum_{m=-\infty}^{\infty} \psi(n-m)g(m) \quad (2.21)$$

avec  $m$  qui correspond aux points de la fonction à lisser,  $\psi$  qui est la fonction  $\psi(\tau|x)$  et  $g$  qui est le filtre que l'on a pris gaussien, centré sur 0 et d'écart type 0.5 Gyr avec 15 points. Lorsque nous effectuons la convolution il faut faire attention aux conditions de bord. Nous constatons sur la figure 2.11 que le lissage fonctionne bien mais qu'il peut rester quelques petits artéfacts qui sont négligeables devant l'amplitude de la fonction  $\psi(\tau|x)$ . Grâce au lissage le mode de la fonction  $\psi(\tau|x)$  est mieux déterminé.

Si après un premier lissage il reste plus de trois pics nous lissons à nouveau. Ensuite la fonction obtenue est normalisée par le maximum de la fonction  $\psi(\tau|x)$ . Cette fonction ainsi obtenue sera utilisée par la suite.

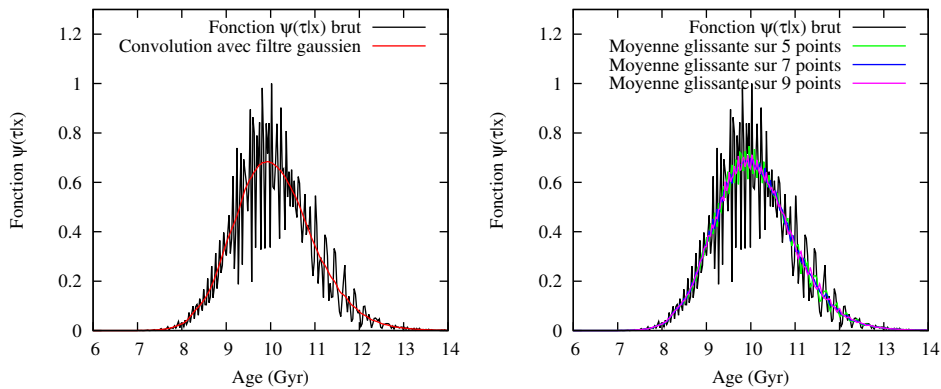


FIGURE 2.11 – Comparaison du lissage sur une fonction  $\psi(\tau|x)$  par convolution avec un filtre gaussien (gauche) et par la moyenne glissante (gauche) avec 5,7 et 9 points.

2. Ensuite nous étudions les fonctions tronquées, avec les deux largeurs à mi-hauteur ( $FWHM_{droite}$  et  $FWHM_{gauche}$ )<sup>1</sup> qui encadrent l'âge le plus probable  $\hat{\tau}$ . Nous considérons que la fonction est tronquée aux petits âges si  $\left| \frac{\tau_{FWHM,gauche} - \hat{\tau}}{\tau_{FWHM,droite} - \hat{\tau}} \right| > 1.5$  et aux âges élevés si  $\left| \frac{\tau_{FWHM,droite} - \hat{\tau}}{\tau_{FWHM,gauche} - \hat{\tau}} \right| > 1.5$ . Il y a trois types de fonctions tronquées :
- **Les fonctions non datées.** Le premier point de ces fonctions correspond à l'âge minimum  $\tau_1$  (ou le dernier point correspond à l'âge maximum  $\tau_2$ ). Elles ont également une largeur à mi-hauteur supérieure à 3 Gyr. Ces fonctions sont considérées comme mal-définies et nous n'évaluons pas l'âge. Si la largeur à mi-hauteur est inférieure à 3 Gyr la fonction est considérée comme tronquée.
  - **Les fonctions tronquées avec plateau** ont une largeur supérieure à 3 Gyr à  $\psi(\tau|x) = 0.95$ . Elles sont considérées comme mal-définies car la présence du plateau montre une équiprobabilité des âges et le mode de la fonction  $\psi(\tau|x)$  est trouvé à cause des fluctuations dans la fonction  $\psi(\tau|x)$ .
  - **Les fonctions tronquées sans plateau** représentent les étoiles qui ne sont pas dans les deux cas précédents. Elles sont considérées comme bien-définies et l'âge de l'étoile est évalué.
3. Les fonctions restantes possèdent un ou plusieurs pics. Afin de différencier les deux cas, nous utilisons le nombre de pics qui est déterminé dans la première étape. Pour les fonctions  $\psi(\tau|x)$  à plusieurs pics, nous vérifions que ce ne sont pas des fonctions non datées avec de grandes fluctuations. Nous cherchons le minimum situé entre les pics puis nous vérifions que cette valeur est présente des deux autres côtés des pics. Si ce n'est pas le cas nous considérons que la fonction  $\psi(\tau|x)$  est mal-définie.

## 2.5 Détermination de l'incertitude sur l'âge

Pour déterminer l'erreur sur l'âge des étoiles, nous avons considéré deux méthodes : l'une évalue l'intervalle de confiance bayésien et l'autre l'intervalle de confiance basé sur Jørgensen & Lindegren (2005). Dans les deux cas l'intervalle de confiance est calculé pour  $1 - \alpha = 68\%$ , cela correspond à une erreur de  $1\sigma$ .

### 2.5.1 Intervalle de confiance bayésien

L'intervalle de confiance bayésien  $1 - \alpha$  permet de déterminer l'intervalle entre deux âges pour lequel il y a une probabilité  $1 - \alpha$  d'avoir l'âge de l'étoile.

---

1. Full Width at Half Maximum

Cette intervalle est déterminé à partir de la fonction  $\psi(\tau|x)$  ce qui permet de prendre en compte les *a priori* dans la détermination de l'incertitude. Nous avons :

$$1 - \alpha = \frac{\int_{\tau_{min}}^{\tau_{max}} \psi(\tau | x) d\tau}{\int_{\tau_1}^{\tau_2} \psi(\tau | x) d\tau}. \quad (2.22)$$

Où  $\tau_{min}$  et  $\tau_{max}$  sont les âges correspondant aux bornes de l'intervalle de confiance et  $\tau_1$  et  $\tau_2$  sont les âges limites pris en compte dans la détermination de l'âge. Le terme  $\int_{\tau_1}^{\tau_2} \psi(\tau | x) d\tau$  permet de normaliser par l'aire totale sous la fonction  $\psi(\tau|x)$ .

$\tau_{max}$  est déterminé par itération, à partir de l'âge  $\hat{\tau}$  correspondant au mode de la fonction  $\psi(\tau|x)$ . L'intégrale  $\int_{\hat{\tau}}^{\tau} \psi(\tau | x) d\tau$  est calculée jusqu'à obtenir  $\int_{\hat{\tau}}^{\tau} \psi(\tau | x) d\tau = \frac{1-\alpha}{2} \times \int_{\tau_1}^{\tau_2} \psi(\tau | x) d\tau$ . Quand cette condition est réalisée nous avons  $\tau_{max} = \tau$ . L'intégrale est calculée grâce à la méthode de Simpson décrite en annexe C.1. La même méthode est utilisée pour obtenir  $\tau_{min}$ .

### 2.5.2 Intervalle de confiance de Jørgensen & Lindegren

Jørgensen & Lindegren (2005, JL05) n'adoptent pas l'intervalle de confiance bayésien car JL05 supposent que si la fonction  $\psi(\tau|x)$  est monotone cela peut entraîner une mauvaise interprétation de l'erreur dans leur méthode. Cependant JL05 ne précisent pas le traitement effectué sur les fonctions  $\psi(\tau|x)$ . JL05 supposent que la fonction  $\psi(\tau|x)$  peut être considérée comme une vraisemblance qui est assimilée à une distribution du  $\chi^2$  à un degré de liberté. Cette méthode ne prend pas en compte les *a priori* dans l'estimation de l'erreur. L'intervalle de confiance est :

$$1 - \alpha = 2\phi(\sqrt{-2 \ln \psi_{lim}}), \quad (2.23)$$

où  $\psi_{lim}$  correspond à la valeur limite pour laquelle nous avons  $\psi(\tau|x) > \psi_{lim}$  pour intervalle  $1 - \alpha$  donné.  $\phi(\sqrt{-2 \ln \psi_{lim}})$  est la fonction de répartition de la loi normale telle que  $\phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x \exp(-t^2/2) dt$ . Pour calculer  $\psi_{lim}$ , la fonction de répartition de la loi normale est évaluée (annexe C.4). Pour  $1 - \alpha = 68 \%$  nous avons  $\psi_{lim} = 0.61$ .

Enfin nous recherchons à partir de  $\psi(\hat{\tau}|x)$  les deux limites telles que  $\psi(\tau|x) > \psi_{lim}$ . Ensuite nous déterminons le  $\tau_{min}$  et  $\tau_{max}$  correspondants.

### 2.5.3 Vérification de l'intervalle de confiance

Pour vérifier si les intervalles de confiance ainsi évalués sont corrects, nous effectuons une simulation de Monte-Carlo (MC) autour d'une étoile de référence du catalogue simulé. Nous simulons 30 étoiles (magnitude absolue

et température effective) pour une métallicité donnée et nous répétons cette étape pour 30 métallicités simulées. Nous obtenons alors un échantillon de 900 étoiles simulées autour de l'étoile de référence. Les étoiles sont simulées avec un tirage aléatoire basé sur la méthode de Box-Muller (annexe C.2) en prenant les incertitudes de l'étoile de référence. Les erreurs observationnelles de ces étoiles sont les mêmes que celles de l'étoile de référence. La simulation de MC pour une étoile de la branche des géantes est représentée sur la figure 2.13.

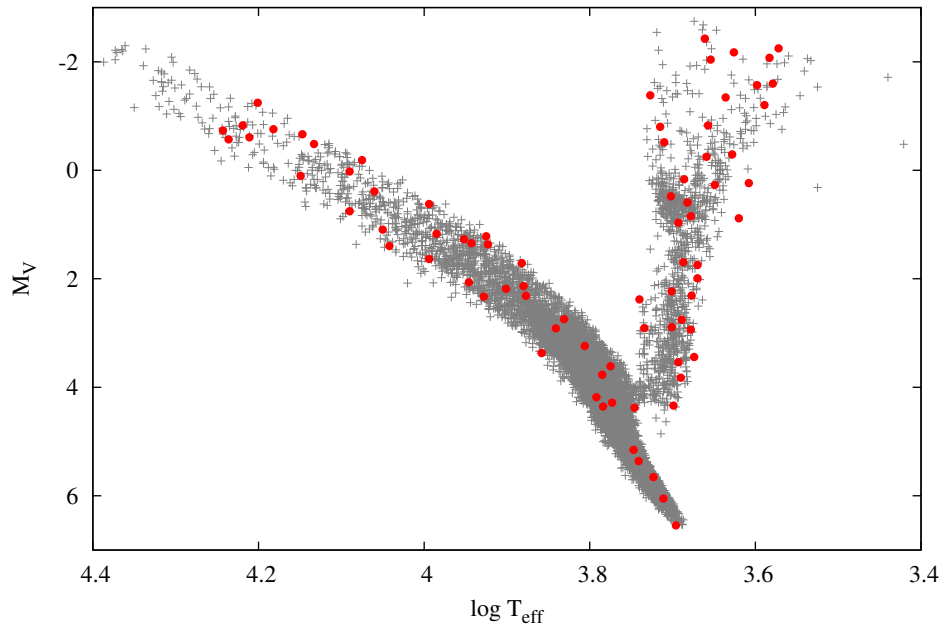


FIGURE 2.12 – Position des étoiles de référence (rouge) dans le diagramme HR du catalogue simulé (noir). Ces étoiles sont utilisées pour vérifier la détermination de l'intervalle de confiance.

Cette simulation est réalisée pour 86 étoiles réparties dans le diagramme HR (figure 2.12). L'âge de toutes les étoiles est calculé puis nous vérifions que pour une simulation de MC il y a au minimum 68 % des étoiles<sup>1</sup> qui ont leurs âges compris dans l'intervalle de confiance de l'étoile de référence.

Il y a 20 étoiles de référence qui ont plus de 100 étoiles de MC d'âge non-déterminé, soit parce que la fonction  $\psi(\tau|x)$  est mal-définie ou parce que le test de rejet du  $\chi^2_{\text{max}}$  n'est pas satisfait. Pour la suite de l'analyse, nous ne les prenons pas en compte. La vérification se fait donc sur 64 étoiles. Il y a 62 % des étoiles de référence avec un intervalle de confiance bayésien correct

1. Nous souhaitons en effet obtenir 68 % d'étoiles car cela correspond à la valeur qui est utilisée pour définir l'intervalle de confiance.

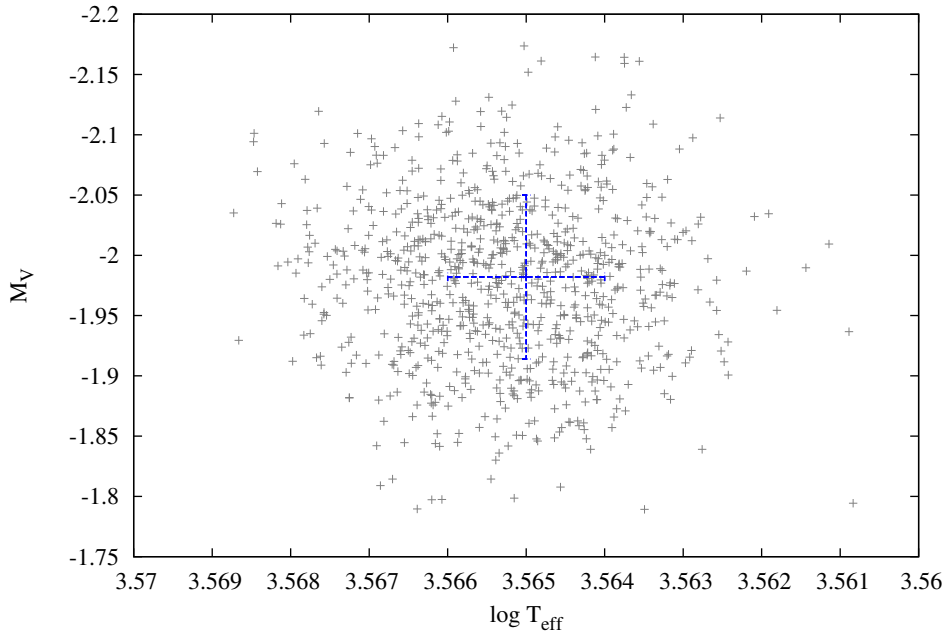


FIGURE 2.13 – Simulation de 900 étoiles avec la méthode de Monte-Carlo autour d’une étoile de référence (bleue) avec ses erreurs observationnelles.

et seulement 17 % avec l’intervalle de confiance de Jørgensen & Lindegren (2005). Nous avons donc choisi d’adopter l’intervalle de confiance bayésien pour la détermination des erreurs. La différence entre les deux méthodes vient de la prise en compte ou non des *a priori*.

Les étoiles avec un intervalle de confiance bayésien qui ne fonctionne pas sont localisées dans les régions de dégénérescence du diagramme HR (partie 5.1). Il y a quatre types d’étoiles avec un intervalle de confiance qui n’est pas correct :

- Les étoiles avec un seul pic fin (20 % des étoiles). Le pic est tellement fin que l’intervalle de confiance est inférieur à 0.1 Gyr. Il est trop étroit pour contenir toutes les étoiles de MC.Ö
- Les étoiles de la branche des géantes (16 %). Leur vitesse d’évolution est grande et, pour une petite distance dans le diagramme HR, la différence en âge est grande.
- Les étoiles proches de la séquence principale d’âge zéro (7 %). Pour ces étoiles c’est le contraire, elles évoluent très lentement. Pour la même position dans le diagramme HR à des métallicités différentes les âges diffèrent beaucoup.
- Les étoiles proches du turn-off (3 %). Ces étoiles sont situées dans une zone de dégénérescence où les âges sont mal restitués (partie 5.1).



## 2.6 Détermination des autres paramètres

L'inférence bayésienne décrite dans ce chapitre permet également de déterminer la masse et la métallicité de l'étoile à partir de l'équation suivante :

$$\begin{aligned} \psi([M/H], m, \tau | \widetilde{\log T_{\text{eff}}}, \widetilde{M_V}, \widetilde{[M/H]}) &\propto \mathcal{L}(\widetilde{\log T_{\text{eff}}}, \widetilde{M_V}, \widetilde{[M/H]} | [M/H], m, \tau) \\ &\times \pi([M/H], m, \tau). \end{aligned} \quad (2.24)$$

Nous intégrons cette équation en fonction de la masse et de la métallicité pour obtenir l'âge. De la même façon, pour obtenir la masse et la métallicité il suffit d'intégrer l'équation 2.24 en fonction des deux autres paramètres.

### 2.6.1 Masse

Pour obtenir la masse, nous intégrons l'équation 2.24 en fonction de la métallicité et de l'âge, d'où :

$$\begin{aligned} \psi(m|x) &= \xi(m) \int_{[M/H]} \left[ \phi([M/H]) \int_{\tau} \lambda(t) \right. \\ &\quad \left. \times \mathcal{L}(\widetilde{\log T_{\text{eff}}}, \widetilde{M_V}, \widetilde{[M/H]} | [M/H], m, \tau) d\tau \right] d[M/H]. \end{aligned} \quad (2.25)$$

La même simplification, que dans la partie 2.1, est utilisée pour les intégrales, elles peuvent être considérées comme une double somme le long des trajets évolutifs :

$$\begin{aligned} \psi(m_j|x) &= \xi(m_j) \times \sum_i \left[ \left( \int_{[M/H]_j^{ijk}}^{[M/H]_j^{ij(k+1)}} \phi([M/H]) d[M/H] \right) \right. \\ &\quad \left. \times \sum_k \left( \int_{\tau_j^{ijk}}^{\tau_j^{ij(k+1)}} \lambda(t_k) d\tau \right) \mathcal{L}(\widetilde{\log T_{\text{eff}}}, \widetilde{M_V}, \widetilde{[M/H]} | [M/H], m, \tau) \right]. \end{aligned} \quad (2.26)$$

Pour le choix des *a priori*, nous utilisons la même MDF et le même SFR que ceux décrits dans les parties 2.2.1 et 2.2.3, c'est à dire une MDF constante et un SFR constant et tronqué. L'IMF que nous prenons est constante afin de ne pas introduire de biais dans la détermination de la masse. Nous utilisons la même résolution de grille que pour l'âge et le même traitement que sur la fonction  $\psi(\tau|x)$ .

La masse la plus probable correspond au mode de la fonction  $\psi(m|x)$ . Pour la détermination de la masse, il n'y aucune fonction  $\psi(m|x)$  mal-définie. Il y a des fonctions  $\psi(m|x)$  avec un ou plusieurs pics (figure 2.14) et des fonctions qui ont subi un lissage (figure 2.14). Après le lissage ces fonctions ont un ou plusieurs pics.

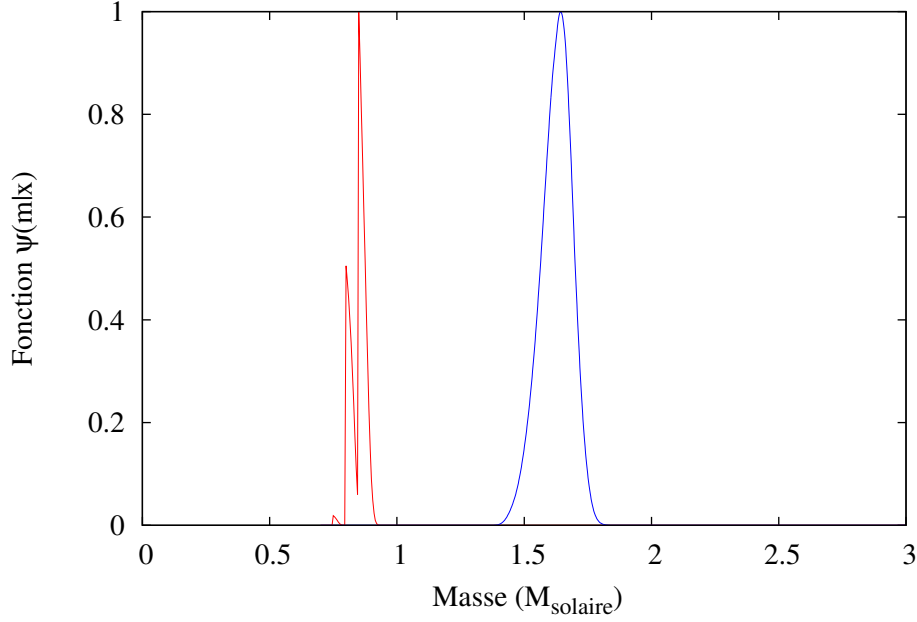


FIGURE 2.14 – Représentation des fonctions  $\psi(m|x)$  avec un pic (bleu) ou plusieurs pics (rouge). Les fonctions qui ont subi un lissage ne sont pas représentées car après le lissage elles possèdent un ou plusieurs pics.

### 2.6.2 Métallicité

Pour la détermination de la métallicité, nous obtenons l'équation suivante à partir de l'équation 2.24 :

$$\begin{aligned} \psi([M/H]|x) &= \phi([M/H]) \int_m \left[ \xi(m) \int_\tau \lambda(t) \right. \\ &\quad \left. \times \mathcal{L}(\widetilde{\log T_{\text{eff}}}, \widetilde{M_V}, [\widetilde{M/H}]|[M/H], m, \tau) d\tau \right] dm. \end{aligned} \quad (2.27)$$

Avec la même simplification pour les intégrales, qui peuvent être considérées comme une double somme le long des trajets évolutifs, nous obtenons :

$$\begin{aligned} \psi([M/H]_i|x) &= \phi([M/H]) \times \sum_j \left[ \left( \int_{m_j^{ijk}}^{m_j^{ij(k+1)}} \xi(m_j) dm \right) \right. \\ &\quad \left. \times \sum_k \left( \int_{\tau_j^{ij(k+1)}}^{\tau_j^{ijk}} \lambda(t_k) d\tau \right) \mathcal{L}(\widetilde{\log T_{\text{eff}}}, \widetilde{M_V}, [\widetilde{M/H}]|[M/H], m, \tau) \right]. \end{aligned} \quad (2.28)$$

La métallicité la plus probable correspond au mode de cette fonction  $\psi([M/H]|x)$ . Il n'y a que des fonctions  $\psi([M/H]|x)$  avec un ou plusieurs pics

(figure 2.15). Nous utilisons la même résolution de grille, le même traitement sur la fonction  $\psi(\tau|x)$  et les mêmes *a priori*.

La détermination de la métallicité peut être intéressante si nous utilisons des trajets évolutifs prenant en compte la diffusion microscopique. En effet les abondances de surface sont affectées par la diffusion et ne reflètent plus les abondances initiales.

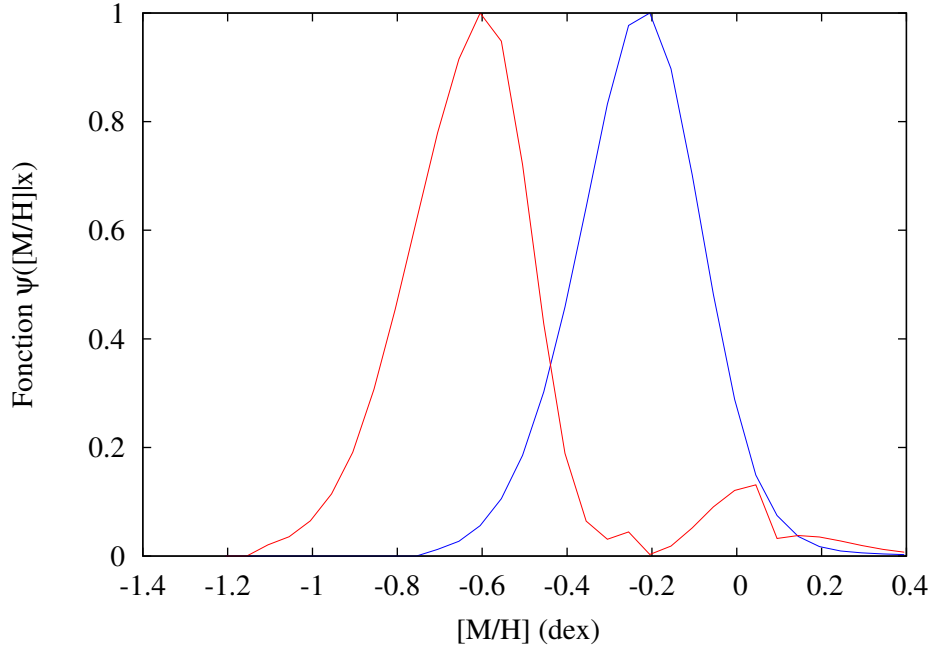


FIGURE 2.15 – Représentation des fonctions  $\psi([M/H]|x)$  avec un pic (bleu) ou plusieurs pics (rouge).



# Travaux antérieurs

---

## Sommaire

---

<b>3.1</b>	<b>Inférence bayésienne . . . . .</b>	<b>47</b>
3.1.1	Jørgensen & Lindegren . . . . .	47
3.1.2	da Silva et al. . . . .	48
3.1.3	Casagrande et al. . . . .	48
<b>3.2</b>	<b>Minimisation du <math>\chi^2</math> . . . . .</b>	<b>49</b>
3.2.1	Ng & Bertelli . . . . .	50
3.2.2	$\chi^2$ standard . . . . .	50
3.2.3	Détermination de l'incertitude sur l'âge . . . . .	51

---

## 3.1 Inférence bayésienne

Nous avons souhaité comparer les résultats de notre méthode d'inférence bayésienne avec ceux de travaux antérieurs, c'est pourquoi nous avons implémenté les méthodes d'inférence bayésienne de Jørgensen & Lindegren (2005, JL05), da Silva et al. (2006, dSGP06) et Casagrande et al. (2011, CSA11). Nous avons utilisé les mêmes *a priori* et la même grille de trajets évolutifs (partie 2.1). Pour ces méthodes le traitement sur la fonction  $\psi(\tau|x)$  et le test de rejet sur le  $\chi^2$  ont également été ajoutés. L'erreur sur la détermination de l'âge est estimée à partir de l'intervalle de confiance défini dans la partie 2.5.1.

### 3.1.1 Jørgensen & Lindegren

Jørgensen & Lindegren (2005) évaluent l'équation 2.7 comme une double somme le long des trajets évolutifs, la probabilité *a posteriori*  $\psi_{\text{JL05}}(\tau|x)$  s'écrit :

$$\begin{aligned} \psi_{\text{JL05}}(\tau_k | x) &= \lambda(t_k) \times \sum_i \sum_j \phi([M/H]_i) \xi(m_j) \Delta m \\ &\times \mathcal{L}(\widetilde{M_V}, \widetilde{\log T_{\text{eff}}}, \widetilde{[M/H]} | [M/H]_i, m_j, \tau_k), \end{aligned} \quad (3.1)$$

où  $\Delta m$  correspond à la discrétisation de la masse. Ici nous avons choisi de prendre  $\Delta m = (m_{j+1} - m_{j-1})/2$ . La principale différence avec notre méthode est l'implémentation des *a priori* car dans notre méthode nous intégrons l'IMF et la MDF contrairement à JL05.

### 3.1.2 da Silva et al.

da Silva et al. (2006) n'incluent pas la métallicité  $\widetilde{[M/H]}$  dans la vraisemblance parce que c'est à la fois un paramètre et une observable. La vraisemblance dépend alors seulement de la température effective et de la magnitude absolue. dSGP06 pondèrent la vraisemblance par un terme dépendant de la métallicité et de son erreur pour ne pas négliger l'erreur observationnelle sur la métallicité. Comme l'erreur sur la métallicité est supposée gaussienne, le pondérateur s'écrit :

$$f(\widetilde{[M/H]}) = \frac{1}{\sigma_{[M/H]}\sqrt{2\pi}} \exp \left[ -\frac{1}{2} \left( \frac{[M/H]_i - \widetilde{[M/H]}}{\sigma_{[M/H]}} \right)^2 \right], \quad (3.2)$$

où  $[M/H]_i$  est la métallicité du trajet évolutif et  $\widetilde{[M/H]}$  la métallicité observée. La probabilité *a posteriori*  $\psi_{\text{dSGP06}}$  est donc :

$$\begin{aligned} \psi_{\text{dSGP06}}(t_k|x) &= \lambda(\tau_k) \sum_i \left[ \int_{[M/H]_i^{ij^k}}^{[M/H]_i^{ij^{(k+1)}}} \phi([M/H]_i) d[M/H] \right] f(\widetilde{[M/H]}) \\ &\times \sum_j \left[ \int_{m_j^{ij^k}}^{m_j^{ij^{(k+1)}}} \xi(m_j) dm \right] \exp(-\chi_{\log T_{\text{eff}}, M_V}^2). \end{aligned} \quad (3.3)$$

La formulation des méthodes est différente de la nôtre à cause du traitement de la métallicité et de l'expression simplifiée de la vraisemblance. Dans l'implémentation cela revient au même d'inclure la métallicité dans la vraisemblance ou de la prendre en tant que pondérateur, car nous supposons dans les deux cas que les erreurs observationnelles sur la métallicité sont gaussiennes.

### 3.1.3 Casagrande et al.

Casagrande et al. (2011) ajoutent l'expression de la parallaxe  $\pi$  dans la détermination de l'âge afin de prendre en compte l'erreur observationnelle de celle-ci. Nous avons donc :

$$\begin{aligned} \psi_{\text{CSA11}}(\tau_k|x) &= \sum_i \sum_j P_{M_V}(M_{V,ijk} | \widetilde{M_V}, \sigma_{\text{pr}}) \times G(\log T_{\text{eff},ijk} - \widetilde{\log T_{\text{eff}}}, \sigma_{T_{\text{eff}}}) \\ &\times G([M/H]_{ijk} - \widetilde{[M/H]}, \sigma_{[M/H]}) \\ &\times \xi(m_j) \lambda(t_k) \phi([M/H]_i) d_m d_\tau d_{[M/H]}, \end{aligned} \quad (3.4)$$

$$\text{avec } P_{M_V}(M_{V,ijk}|\widetilde{M}_V, \sigma_\pi) = 10^{0.2\Delta M_V} \exp \left[ -\frac{(10^{0.2\Delta M_V} - 1)^2}{2\sigma_{\text{pr}}^2} \right]. \quad (3.5)$$

$G(x - x_0, \sigma_x)$  est une fonction gaussienne de moyenne  $x_0$  et d'écart-type  $\sigma_x$ . La quantité  $\sigma_{\text{pr}} = \sigma_\pi/\widetilde{\pi}$  est l'erreur relative sur la parallaxe et  $\Delta M_V = M_{V,ijk} - \widetilde{M}_V$ . La valeur  $d_m d_\tau d_{[M/H]}$  correspond au volume effectif que couvre un point dans l'espace des trajets évolutifs. CSA11 supposent que les grandes métallicités situées dans la queue de la MDF sont surestimées et au contraire les petites métallicités situées dans la queue de la MDF sont sous-estimées. Cet effet modifie donc l'âge déterminé de l'étoile c'est pourquoi CSA11 incluent une correction sur la métallicité observée. Nous avons choisi de ne pas utiliser cette correction dans la détermination de l'âge pour pouvoir comparer les méthodes entre elles.

La méthode de CSA11 exprime la vraisemblance de façon différente de la nôtre. CSA11 incluent différemment le volume couvert par les trajets évolutifs qui correspond à la discrétisation de l'intégration. La principale différence avec notre méthode est le traitement de la parallaxe que CSA11 inclut pour prendre en compte son erreur observationnelle supposée gaussienne. CSA11 négligent l'erreur observationnelle sur la magnitude apparente car elle est négligeable devant celle sur la parallaxe.

Dans le cadre de la mission spatiale Gaia, la détermination des âges est réalisée par le groupe FLAME (Final Luminosity Age and Mass Estimator) qui fournit la luminosité. La majorité des travaux de la littérature déterminent également les âges avec les trois observables :  $L, \log T_{\text{eff}}, [M/H]$ .

Le travail de CSA11 sur le traitement de la magnitude absolue et sur la parallaxe est plus rigoureux car l'erreur observationnelle sur la magnitude absolue n'est pas forcément symétrique à cause de la dispersion qu'entraîne l'erreur observationnelle sur la parallaxe. Dans notre méthode, nous supposons que l'erreur observationnelle sur la magnitude absolue est gaussienne et symétrique. Pour étudier ces effets, il faudrait construire un catalogue qui prendrait en compte une erreur observationnelle gaussienne sur la parallaxe et une erreur observationnelle non symétrique sur la magnitude absolue. Il faudrait également étudier l'expression la plus appropriée de la vraisemblance pour prendre en compte ce phénomène.

## 3.2 Minimisation du $\chi^2$

Nous avons également implémenté deux méthodes de minimisation du  $\chi^2$  pour les comparer avec l'inférence bayésienne de notre travail. La première est basée sur Ng & Bertelli (1998, NB98) et la seconde sur un  $\chi^2$  standard. La minimisation du  $\chi^2$  est effectuée à partir des trajets évolutifs de BaSTI. L'âge

le plus probable  $\hat{\tau}$  correspond au  $\chi^2$  minimum. Nous en déduisons également la masse et la métallicité la plus probable. Cette méthode est dite conjointe contrairement à l'inférence bayésienne qui est dite marginale car elle détermine l'âge quelles que soient la masse et la métallicité.

Pour les minimisations du  $\chi^2$  nous utilisons la même grille de trajets évolutifs et nous ajoutons le test de rejet sur le  $\chi^2$  (partie 2.1).

### 3.2.1 Ng & Bertelli

Ng & Bertelli (1998, NB98) déterminent l'âge pour une seule métallicité, la plus proche de celle de l'étoile. NB98 minimisent la distance entre le rayon de l'étoile et celui issu des trajets évolutifs. Ce rayon est calculé à partir de la loi de Stefan Boltzmann. La minimisation du  $\chi^2$  s'écrit donc :

$$\chi_{NB98}^2 = \frac{\sqrt{\widetilde{M}_V - M_{V,ijk})^2 + 16(\log \widetilde{T}_{\text{eff}} - \log T_{\text{eff},ijk})^2}}{\sigma_{M_V}}. \quad (3.6)$$

NB98 négligent l'erreur observationnelle sur la température effective car ils la considèrent petite par rapport à l'erreur observationnelle sur la magnitude absolue. Le  $\chi^2$  est calculé pour tous les points des trajets évolutifs qui ont une magnitude absolue dans l'intervalle  $\widetilde{M}_V \pm 3\sigma_{M_V}$  et une température effective dans l'intervalle  $\log \widetilde{T}_{\text{eff}} \pm 3\sigma_{\log T_{\text{eff}}}$ .

### 3.2.2 $\chi^2$ standard

Le  $\chi^2$  standard utilise la même sélection sur la température effective et la magnitude absolue que NB98. Par contre il est calculé pour tous les trajets évolutifs qui ont une métallicité dans l'intervalle  $[\widetilde{\text{M/H}}] \pm 3\sigma_{[\text{M/H}]}$  :

$$\begin{aligned} \chi_{ST}^2 = & \left( \frac{\log \widetilde{T}_{\text{eff}} - \log T_{\text{eff},ijk}}{\sigma_{\log T_{\text{eff}}}} \right)^2 + \left( \frac{\widetilde{M}_V - M_{V,ijk}}{\sigma_{M_V}} \right)^2 \\ & + \left( \frac{[\widetilde{\text{M/H}}] - [\text{M/H}]_{ijk}}{\sigma_{[\text{M/H}]}} \right)^2. \end{aligned} \quad (3.7)$$

Les différences entre les deux minimisations du  $\chi^2$  concernent le traitement de la métallicité et l'erreur observationnelle sur la température effective.



### 3.2.3 Détermination de l'incertitude sur l'âge

Pour déterminer l'incertitude sur l'âge avec la minimisation du  $\chi^2$  nous utilisons la propagation des incertitudes :

$$\begin{pmatrix} \sigma_m^2 & \rho_{12}\sigma_m\sigma_\tau \\ \rho_{12}\sigma_m\sigma_\tau & \sigma_\tau^2 \end{pmatrix} = D\Sigma^x D^T, \quad (3.8)$$

où  $D$  est la matrice design telle que :

$$D = \begin{pmatrix} \frac{\delta m}{\delta \log T_{\text{eff}}} & \frac{\delta m}{\delta M_V} & \frac{\delta m}{\delta [M/H]} \\ \frac{\delta \tau}{\delta \log T_{\text{eff}}} & \frac{\delta \tau}{\delta M_V} & \frac{\delta \tau}{\delta [M/H]} \end{pmatrix}, \quad (3.9)$$

et  $\Sigma^x$  la matrice de variance-covariance des erreurs observationnelles :

$$\Sigma^x = \begin{pmatrix} \sigma_{\log T_{\text{eff}}}^2 & 0 & 0 \\ 0 & \sigma_{M_V}^2 & 0 \\ 0 & 0 & \sigma_{[M/H]}^2 \end{pmatrix}. \quad (3.10)$$

Pour calculer la matrice design nous calculons l'âge et la masse pour six cas supplémentaires :  $\widetilde{\log T_{\text{eff}}} \pm \sigma_{\log T_{\text{eff}}}$ ,  $\widetilde{M_V} \pm \sigma_{M_V}$  et  $\widetilde{[M/H]} \pm \sigma_{[M/H]}$ . Cette méthode permet simultanément de déterminer l'erreur sur l'âge et sur la masse.

Pour tester l'incertitude sur l'âge nous utilisons la même méthode que pour l'intervalle de confiance de l'inférence bayésienne (partie 2.5.3) avec le même catalogue simulé de Monte Carlo. Nous trouvons 62 % des étoiles de référence ayant une incertitude correcte sur l'âge . Les autres étoiles sont localisées dans la branche des géantes rouges et sur la séquence principale à proximité de la séquence principale d'âge zéro.



# Catalogues simulés de Gaia

---

Pour vérifier l'accord entre l'âge déterminé et l'âge réel de l'étoile nous construisons un catalogue simulé selon les spécificités de la mission spatiale Gaia (Perryman et al. 2001). Le catalogue simulé est calculé pour une distance fixée et pour une direction d'observation fixée. Pour le catalogue simulé de référence, sur lequel nous allons réaliser les différents tests et les comparaisons, nous prenons une distance de 1 000 pc en direction du centre galactique (longitude  $0^\circ$ , latitude  $0^\circ$ ). Ce catalogue est construit à partir des trajets évolutifs de BaSTI (Pietrinferni et al. 2004).

Tout d'abord nous tirons aléatoirement, selon la loi de Box Muller (annexe C.2), une métallicité en fonction de la MDF basée sur le GCS (équation 2.8) et une masse qui respecte l'IMF de K02. Le GCS (Casagrande et al. 2011) est un relevé de 16 682 étoiles du voisinage solaire, plus de détails sur ce catalogue sont donnés dans la partie 8.1.1.4. Pour la masse, nous devons inverser l'IMF afin de tirer aléatoirement un nombre entre 0 et 1 qui correspond à une valeur du taux de l'IMF et qui nous donne donc une masse.

Ces deux tirages permettent de sélectionner quatre trajets évolutifs : il y a deux trajets évolutifs pour les deux métallicités qui encadrent celle sélectionnée et cela pour les deux masses qui encadrent la masse choisie. Enfin nous tirons aléatoirement un âge selon le SFR (équation 2.17), cet âge doit être inclus dans les quatre trajets évolutifs. Pour obtenir les différents paramètres de l'étoile (magnitudes apparentes et absolues, luminosité, température effective, couleurs et masse actuelle) nous effectuons des interpolations linéaires en métallicité, en masse et en âge entre les huit points répartis sur les quatre trajets évolutifs.

Afin de respecter les spécificités de la mission spatiale Gaia<sup>1</sup> nous ne gardons que les étoiles qui ont une magnitude apparente  $G > 6$  car sinon elles sont trop brillantes pour être observées par le satellite. La magnitude apparente  $G$  est calculée, à partir de la magnitude apparente  $V$  et de la couleur  $(B - V)$  de l'étoile, grâce à la calibration de Jordi et al. (2010) :

$$G = V - 0.0424 - 0.0851(B - V) - 0.3348(B - V)^2 + 0.0205(B - V)^3. \quad (4.1)$$

---

1. Version d'août 2012 de [http://www.rssd.esa.int/index.php?project=GAIA&page=Science\\_Performance](http://www.rssd.esa.int/index.php?project=GAIA&page=Science_Performance)

Nous avons choisi de ne garder que les étoiles dont la précision sur la parallaxe est  $\sigma_\pi/\pi < 10\%$ . L'incertitude de la parallaxe est déterminée à partir des spécificités données sur le site de l'ESA<sup>1</sup> :

$$\sigma_\pi = (9.3 + 658.1z + 4.568z^2)^{1/2} \times [0.986 + (1 - 0.986)(V - I)] \quad (4.2)$$

$$\text{avec } z = \text{MAX}[10^{0.4(12-15)}, 10^{0.4(G-15)}]. \quad (4.3)$$

Nous attribuons une erreur observationnelle à chaque paramètre ( $M_V$ ,  $\log T_{\text{eff}}$  et  $[M/H]$ ) selon les prévisions de la mission Gaia (Bailer-Jones 2010; Liu et al. 2012). Pour la température effective nous prenons une erreur observationnelle de 1 % de  $T_{\text{eff}}$  et nous prenons une erreur observationnelle constante de 0.2 dex sur la métallicité. L'erreur sur la magnitude absolue est déterminée à partir de la relation magnitude absolue-magnitude apparente :

$$\sigma_{M_V} = \sqrt{\sigma_V^2 + \left(\frac{5}{\ln 10} \frac{\sigma_\pi}{\pi}\right)^2 + \sigma_{A_V}^2}. \quad (4.4)$$

Avec  $A_V$  qui est l'extinction et qui dépend de la distance et des coordonnées galactiques de l'étoile. Selon Drimmel et al. (2003), pour une distance de 1 000 pc en direction du centre galactique,  $A_V$  est égale à 0.7 mag. Les spécificités de la mission Gaia nous donnent  $\sigma_{A_V} = 0.1A_V$  et  $\sigma_G = 10^{-3}(0.02076z^2 + 2.7224z + 0.004352)^{1/2}$ , nous avons supposé que  $\sigma_V = \sigma_G$ . Nous calculons le terme  $\frac{\sigma d}{d}$  à partir de l'erreur sur la parallaxe.

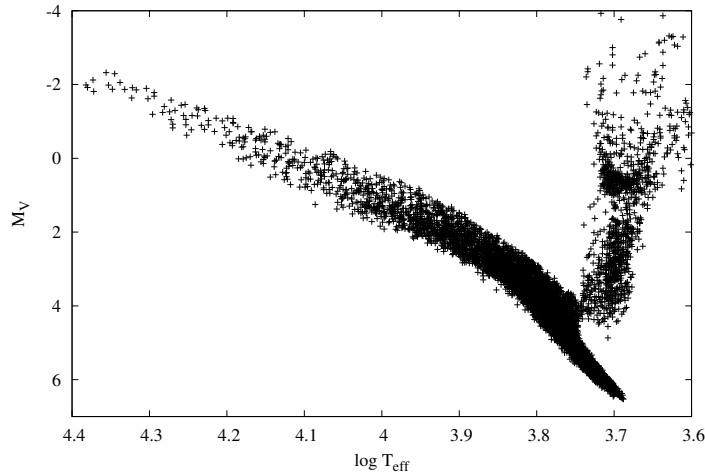


FIGURE 4.1 – Diagramme HR du catalogue simulé de référence de 10 000 étoiles situées à 1 000 pc. Il est construit à partir des spécificités de la mission Gaia et des trajets évolutifs de BaSTI.

Pour calculer l'erreur observationnelle sur la parallaxe nous utilisons la transformation de Vandenberg & Clem (2003) pour déterminer la couleur ( $V-I$ ) à partir de la température effective et de la gravité de surface. Dans une partie de la région des sous-géantes la transformation de Vandenberg & Clem (2003) n'est pas définie ( $\log g$  hors grille). Il en résulte une coupure nette au niveau de la séquence principale et un manque de sous géantes dans le catalogue simulé. Dans l'avenir il faudra traiter ce problème.

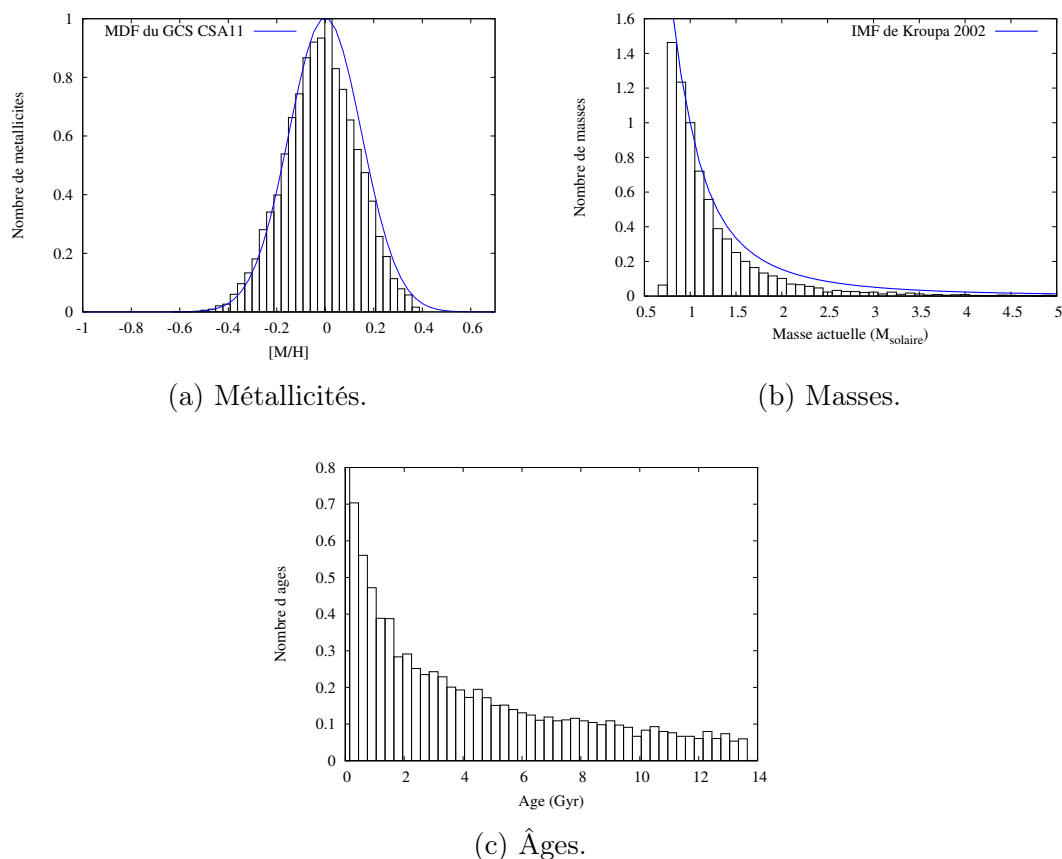


FIGURE 4.2 – Histogrammes des paramètres du catalogue simulé de référence.

Sur les figures 4.2 sont représentés les histogrammes des paramètres tirés aléatoirement ( $[M/H]$ ,  $m$ ,  $\tau$ ) pour le catalogue simulé de référence (figure 4.1). Les métallicités du catalogue simulé de référence sont bien incluses dans la MDF du GCS CSA11. Les masses suivent l'IMF de KR02 sauf pour les petites masses qui ne sont pas assez nombreuses car les trajets évolutifs de BaSTI commencent à des masses comprises entre  $0.65$  et  $0.8 M_{\odot}$  selon la métallicité. Par contre, la distribution des âges n'est pas monotone ceci est dû à la sélection des paramètres. Il y a moins de grands âges car les étoiles massives évoluent

plus rapidement et donc pour ces étoiles il y a seulement des petits âges de disponibles.

Sur la figure 4.3 est représenté l’histogramme des magnitudes apparentes  $G$  et les erreurs observationnelles sur la magnitude absolue et sur la température effective en fonction de la magnitude apparente  $G$ . Les étoiles avec une magnitude  $G$  inférieure à 10 mag correspondent à des étoiles brillantes de la branche des géantes rouges ou à des étoiles massives de la séquence principale. L’erreur sur la température est plus petite pour  $G < 8$  mag car se sont des géantes rouges qui ont une température effective plus faible que les étoiles massives de la séquence principale (entre 8 et 10 mag). Le pic à  $G = 11$  mag correspond au clump<sup>2</sup>, ces étoiles sont plus froides que les étoiles massives de la séquence principale. L’erreur observationnelles sur la température effective de ces étoiles est plus petite. Pour  $G > 12$  mag ce sont les étoiles de la séquence principale de petites masses. L’erreur sur la magnitude absolue augmente car l’erreur observationnelle sur  $G$  augmente et l’erreur observationnelle sur la température effective diminue car la température effective diminue.

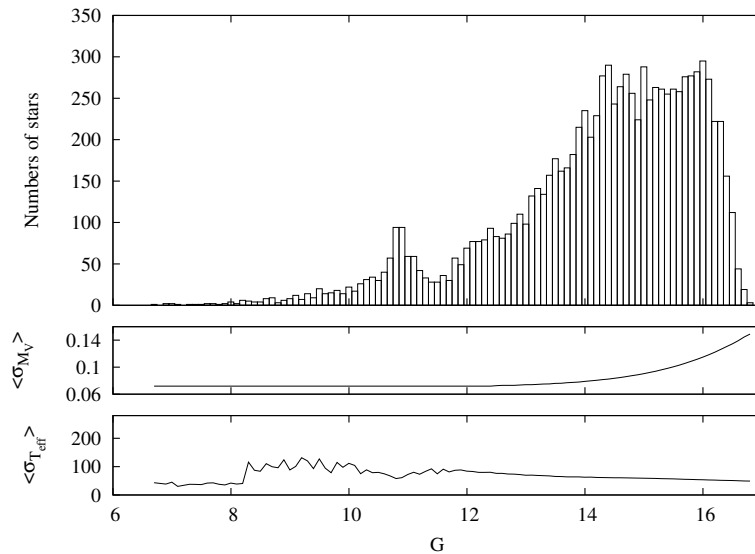


FIGURE 4.3 – Distribution des magnitudes apparentes  $G$  (figure du haut) pour les étoiles du catalogue simulé de référence. Les variations des erreurs observationnelles de la température effective  $T_{\text{eff}}$  et de la magnitude absolue  $M_V$  en fonction de  $G$  sont représentées sur les deux figures du bas.

Nous construisons également des catalogues simulés à d’autres distances (100 pc, 500 pc, 2 500 pc et 5 000 pc, 7 500 pc et 10 000 pc) et pour différentes

2. Le clump est la région du diagramme HR où les étoiles brûlent l’hélium dans le cœur de l’étoile.

directions d'observations : le centre galactique (longitude  $0^\circ$ , latitude  $0^\circ$ ), le plan galactique (longitude  $50^\circ$ , latitude  $0^\circ$ ) et au dessus du plan galactique (longitude  $0^\circ$ , latitude  $15^\circ$ ). Ces catalogues sont présentés dans la partie 7.

Nous construisons un catalogue avec les trajets évolutifs de CESAM (figure 4.4) et de Padoue (figure 4.5). Il n'y a aucune étoile dans le clump car ce stade d'évolution ne sont pas couverts par les trajets évolutifs de CESAM.

Chaque catalogue contient 10 000 étoiles simulées, ceci ne représente pas le nombre d'étoiles dans le voisinage solaire. En effet selon Robin et al. (2012) il y a environ  $10^6$  étoiles à 1 000 pc. Pour l'étude des résultats sur les âges nous ne pouvons pas prendre autant d'étoiles car les temps de calcul seraient trop longs. C'est pourquoi nous prenons 10 000 étoiles ce qui est satisfaisant pour réaliser de bonnes statistiques.

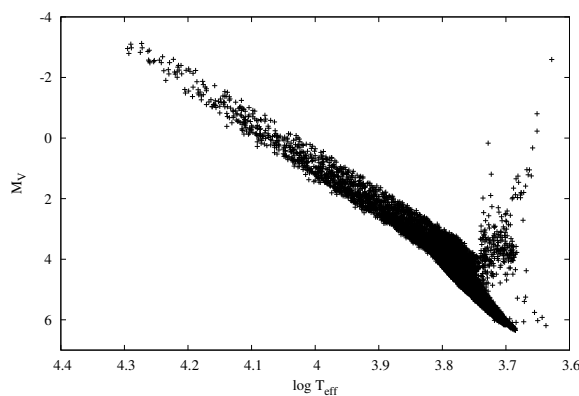


FIGURE 4.4 – Diagramme HR du catalogue simulé avec les trajets évolutifs de CESAM, pour 10 000 étoiles à 1 kpc.

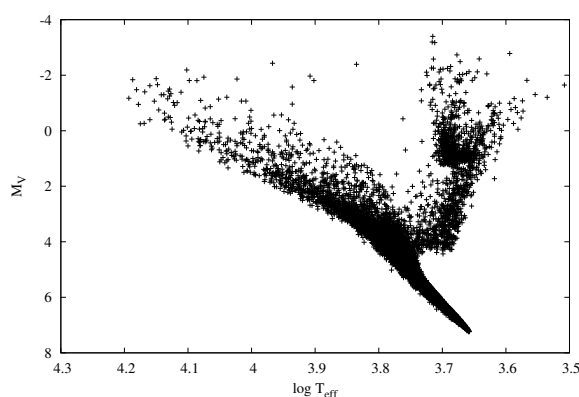


FIGURE 4.5 – Diagramme HR du catalogue simulé avec les trajets évolutifs de Padoue, pour 10 000 étoiles à 1 kpc.





Deuxième partie

Résultats



# Résultats dans le HRD avec $[M/H]$

---

## Sommaire

---

<b>5.1</b>	<b>Comparaison de l'âge réel simulé et de l'âge déterminé par inférence bayésienne . . . . .</b>	<b>61</b>
<b>5.2</b>	<b>Etudes des fonctions <math>\psi(\tau x)</math> . . . . .</b>	<b>64</b>
<b>5.3</b>	<b>Impact de différents choix dans l'inférence bayésienne</b>	<b>66</b>
5.3.1	Densité d'étoiles le long des trajets évolutifs . . . . .	66
5.3.2	Métallicités . . . . .	67
5.3.3	Pas en âge ou $\log(\hat{\text{âge}})$ . . . . .	68
<b>5.4</b>	<b>Détermination de la masse . . . . .</b>	<b>69</b>
<b>5.5</b>	<b>Comparaisons des méthodes . . . . .</b>	<b>72</b>
<b>5.6</b>	<b>Conclusions et perspectives . . . . .</b>	<b>84</b>

---

Les résultats, que nous présentons dans ce chapitre, sont valables pour les 10 000 étoiles du catalogue simulé de référence (partie 4), une analyse plus approfondie des résultats en fonction de la distance des étoiles, de la direction d'observation et des erreurs observationnelles est présentée dans la partie 7.

## 5.1 Comparaison de l'âge réel simulé et de l'âge déterminé par inférence bayésienne

L'inférence bayésienne ne permet pas de dater 31 % des étoiles du catalogue simulé, (figure 5.1). Parmi celles-ci, 1 % ne passent pas le test de rejet sur le  $\chi^2$  (partie 2.1). Ces étoiles sont localisées dans le haut de la séquence principale ( $m > 3.5 M_{\odot}$ ) car la résolution en masse de la grille des trajets évolutifs n'est pas assez fine dans cette région. Les autres étoiles (30 %) ont des fonctions  $\psi(\tau|x)$  mal-définies, ce cas est discuté en détail dans la partie suivante (5.2).

Dans la suite de la comparaison, on considère que les 6 900 étoiles qui ont un âge déterminé. L'erreur relative  $\Delta\tau/\tau = (\tau_{\text{simulé}} - \tau_{\text{calculé}})/\tau_{\text{simulé}}$  entre l'âge réel simulé et l'âge déterminé est représentée pour toutes les étoiles sur la figure 5.2. Nous trouvons 69 % (respectivement 50 %) des étoiles qui ont une

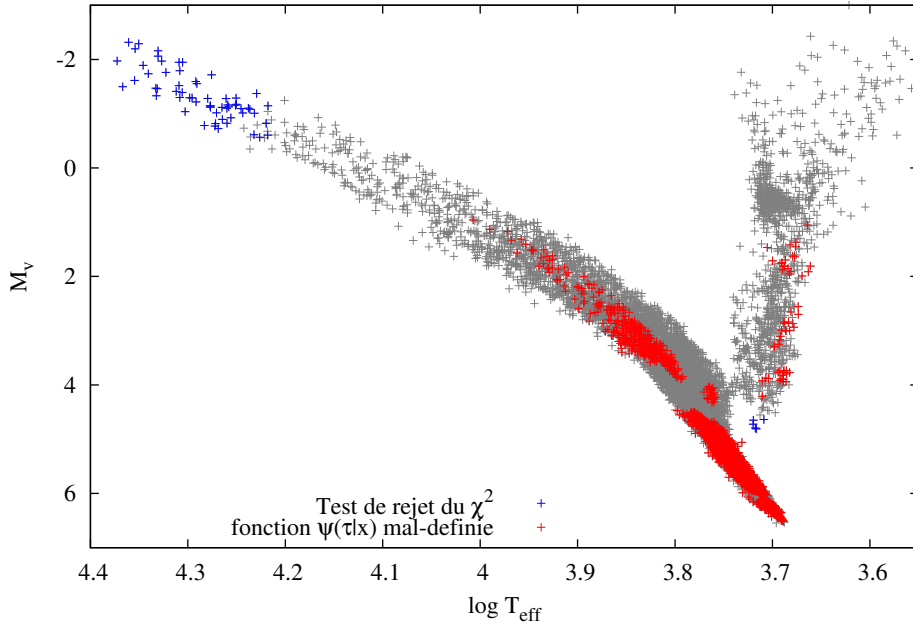


FIGURE 5.1 – Diagramme HR présentant les étoiles non datées par l’inférence bayésienne. En bleu les étoiles qui ne passent pas le test de rejet du  $\chi^2$  et en rouge celles qui ont une fonction  $\psi(\tau|x)$  mal-définie.

erreur relative  $\Delta\tau/\tau < 20\%$  ( $10\%$ ). Nous considérons qu’elles ont un âge bien-déterminé. Elles sont réparties dans tout le diagramme HR mais surtout au turn-off<sup>1</sup>, sur la branche des sous géantes et pour les masses intermédiaires ( $1.0 M_\odot < m < 2.0 M_\odot$ ) sur la séquence principale. Il y a  $14\%$  d’étoiles avec  $\Delta\tau/\tau > 50\%$ , nous considérons que l’âge est mal-déterminé. Elles sont localisées dans la branche des géantes rouges, à proximité de la séquence principale d’âge zéro et dans la zone des étoiles de faible masse ( $m < 1.0 M_\odot$ ) sur la séquence principale. Ces dégénérescences sont dues à la vitesse d’évolution des étoiles qui est lente sur la séquence principale d’âge zéro et au contraire très rapide sur la branche des géantes rouges. Les distributions des âges réels simulés et des âges déterminés (figure 5.3) sont similaires. Il y a plus d’étoiles avec un âge déterminé inférieur à  $0.3$  Gyr car la détermination de l’âge donne des petits âges pour les étoiles qui sont à proximité de la séquence principale d’âge zéro ou dans le bas de la séquence principale.

1. Le turn-off correspond à la fin de la combustion de l’hydrogène dans le noyau.

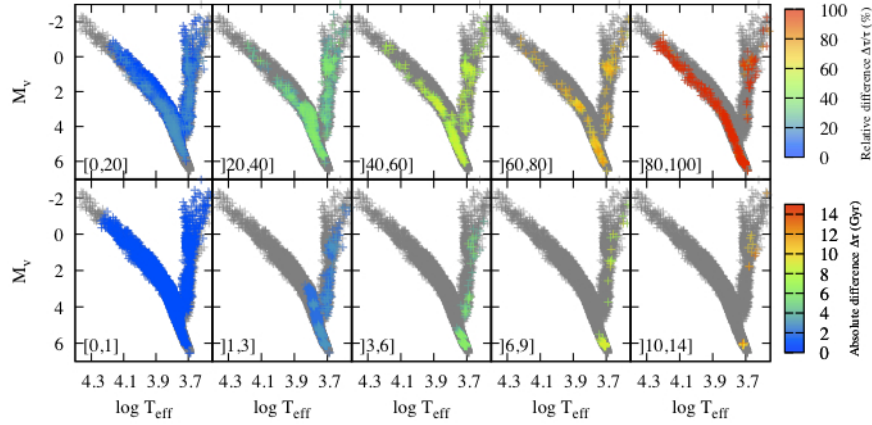


FIGURE 5.2 – Diagrammes HR montrant la comparaison entre l'âge réel simulé et l'âge déterminé par l'inférence bayésienne en fonction de l'erreur relative  $\Delta\tau/\tau$  (figures du haut) et de l'erreur absolue  $\Delta\tau$  (figures du bas). Les âges sont déterminés pour les étoiles du catalogue simulé de référence. La couleur représente l'erreur relative  $\Delta\tau/\tau$ , les étoiles en bleu ont un âge bien déterminé alors que celles en rouge ont un âge mal déterminé. Sur les figures du bas, la couleur représente l'erreur absolue  $\Delta\tau$ , les étoiles en bleu ont un âge déterminé proche de celui simulé alors que celles en rouge ont une grande différence entre les deux âges.

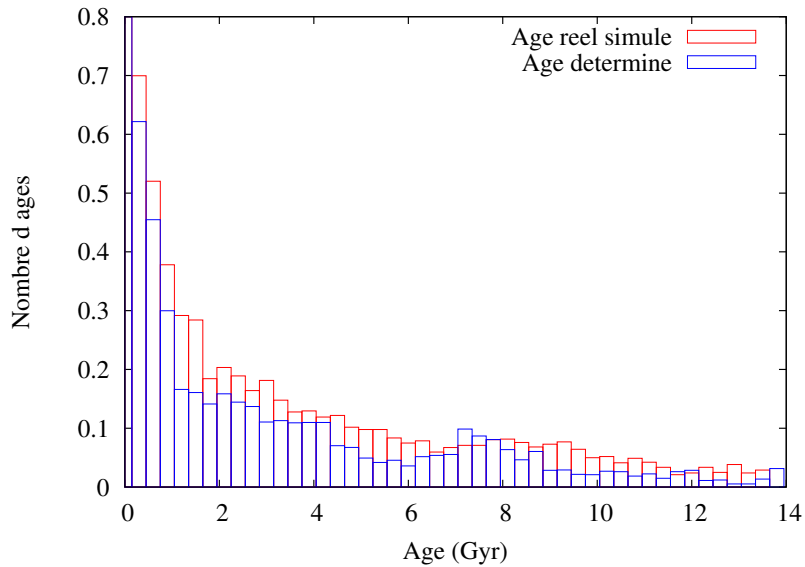


FIGURE 5.3 – Distribution des âges réels simulés du catalogue simulé de référence et des âges déterminés par l'inférence bayésienne.

## 5.2 Etudes des fonctions $\psi(\tau|x)$

Nous étudions maintenant les résultats en fonction de la morphologie de la fonction  $\psi(\tau|x)$ , ces résultats sont présentés dans le tableau 5.1 et sur la figure 5.4. Il y a 16 % des étoiles qui ont une fonction bien-définie avec un seul pic, l'âge de ces étoiles est bien restitué dans 88 % des cas (avec  $\Delta\tau/\tau < 20$  %). Les étoiles présentant des oscillations de la fonction  $\psi(\tau|x)$  sont localisées dans la branche des géantes rouges et sur la séquence principale proche du turn-off. Ces oscillations sont dues à la vitesse d'évolution des étoiles qui est très rapide dans ces régions. Les âges sont correctement déterminés pour 53 % des étoiles.

TABLE 5.1 – Pourcentage d'étoiles des différents types de morphologie de la fonction  $\psi(\tau|x)$  (partie 2.4). Les fonctions  $\psi(\tau|x)$  qui sont en gras sont mal définies et les étoiles correspondantes ne sont pas datées. Le pourcentage d'étoiles avec  $\Delta\tau/\tau < 20$  % est donné pour chaque cas. Dans ce tableau nous n'avons pas inclut les étoiles qui ne passent pas le test de rejet sur le  $\chi^2$  (1 % des étoiles).

Type de fonctions $\psi(\tau x)$	Etoiles	$\Delta\tau/\tau < 20\%$
Un pic	16 %	88 %
Plusieurs pics	18 %	84 %
Oscillations	8 %	53 %
Tronquée aux petits âges		
<i>sans plateau</i>	26 %	47 %
<i>avec plateau</i>	20 %	19 %
Tronquée aux âges élevés		
<i>sans plateau</i>	1 %	100 %
<i>avec plateau</i>	4 %	81 %
<b>Non datée</b>	6 %	62 %

Les fonctions  $\psi(\tau|x)$  avec plusieurs pics sont localisées dans le clump et sur la séquence principale pour des masses comprises entre 0.9 et 1.4  $M_{\odot}$ . La présence de plusieurs pics est causée par la morphologie des trajets évolutifs. En effet sur la séquence principale, apparait à environ 1.2  $M_{\odot}$  (cela dépend de la métallicité de l'étoile) un crochet à la fin de la séquence principale. Il apparait quand les étoiles, qui ont eu un cœur convectif, ont fini de brûler l'hydrogène central. Dans le diagramme HR la température effective et la magnitude absolue varient peu au niveau du crochet, mais comme l'étoile

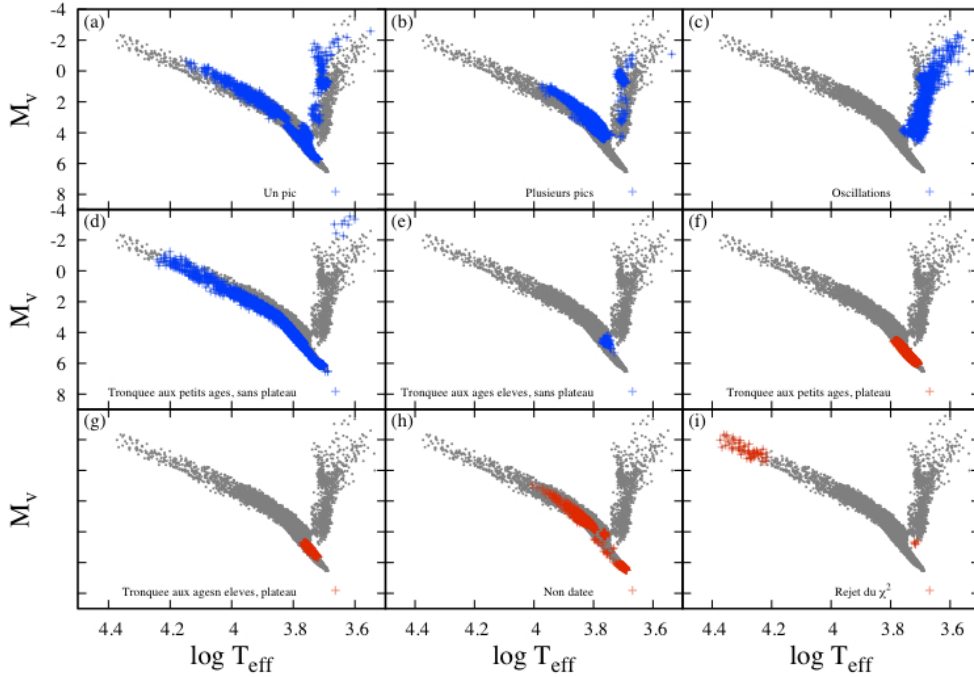


FIGURE 5.4 – Localisation des différentes morphologies de la fonction  $\psi(\tau|x)$  dans le diagramme HR. Les âges sont déterminés pour le catalogue simulé de référence (étoile en gris).

évolue rapidement l'âge est incertain. C'est le même problème dans le clump qui est en forme de crochet. L'âge des étoiles avec une fonction  $\psi(\tau|x)$  avec plusieurs pics est bien restitué pour 84 % d'entre elles ( $\Delta\tau/\tau < 20\%$ ).

Il y a 46 % des étoiles avec une fonction  $\psi(\tau|x)$  tronquée aux petits âges. Seulement 19 % des étoiles dont les fonctions  $\psi(\tau|x)$  sont mal-définies (avec plateau) ont un âge bien déterminé. Cela confirme que les âges de ces fonctions  $\psi(\tau|x)$  sont mal-déterminés, et que l'on ne doit pas les dater. Ces étoiles sont localisées sur la séquence principale plus particulièrement dans le bas ( $m < 1 M_{\odot}$ ). C'est une des régions avec beaucoup de dégénérescence. En revanche, nous trouvons 47 % des étoiles avec une fonction  $\psi(\tau|x)$  tronquée aux petits âges sans plateau qui ont un âge bien restitué.

Enfin, il y a très peu d'étoiles dont la fonction  $\psi(\tau|x)$  est tronquée aux âges élevés. Nous ne datons que celles correspondant à des fonctions  $\psi(\tau|x)$  bien définies sans plateau. Ces étoiles sont localisées dans le bas de la séquence principale et à proximité de la séquence principale d'âge zéro (figure 5.2).

## 5.3 Impact de différents choix dans l'inférence bayésienne

### 5.3.1 Densité d'étoiles le long des trajets évolutifs

La vitesse d'évolution des étoiles le long des trajets évolutifs n'est pas constante, en effet le temps passé sur la séquence principale est beaucoup plus long que le temps passé au turn-off, sur la branche des sous géantes ou des géantes rouges. Hernandez & Valls-Gabaud (2008, HVG08) préconisent de favoriser les régions où les étoiles passent le plus de temps, pour cela il faut inclure dans le calcul de l'âge une densité d'étoiles le long des isochrones. Elle s'écrit :

$$\frac{dN}{ds} = - \left( \frac{dN}{dm} \right) \left( \frac{\partial m}{\partial t} \Big|_s \right) \left( \frac{\partial t}{\partial s} \Big|_m \right), \quad (5.1)$$

avec  $N$  le nombre d'étoiles et  $s = s(m, t)$  la coordonnée curviligne sur l'isochrone. Le premier terme est relié à l'IMF, le second est constant et le dernier est similaire à une vitesse d'évolution. Nous avons modifié cette équation afin de l'adapter aux trajets évolutifs, nous avons alors  $\frac{dN}{ds} = \frac{\partial s}{\partial t}$ . Ce terme correspond à l'inverse de la vitesse d'évolution donc plus l'étoile évolue lentement plus la densité d'étoiles le long des trajets évolutifs est grande. On ne prend pas en compte l'IMF dans le calcul de la densité car elle est déjà incluse dans la détermination de l'âge en tant que *a priori*. La densité d'étoiles le long des trajets évolutifs est calculée pour tous les points de chaque trajet évolutif, entre deux points, elle s'écrit :

$$\frac{dN}{ds} = \frac{\Delta\tau}{\sqrt{(\Delta \log T_{\text{eff}})^2 + (\Delta M_V)^2}}, \quad (5.2)$$

où  $\Delta$  correspond à la différence entre deux points.

Nous déterminons l'âge des étoiles du catalogue simulé de référence avec cette densité (figure 5.5). Nous trouvons 22 % des étoiles avec  $\Delta\tau/\tau < 20$  %. La densité dégrade la restitution des âges surtout pour les étoiles situées à proximité de la séquence principale d'âge zéro et dans la branche des géantes rouges. La densité favorise donc trop les petits âges et cela entraîne un biais dans la détermination des âges. Nous avons choisi de ne pas l'intégrer dans la datation des étoiles. De plus, la densité d'étoiles le long des trajets évolutifs est incluse implicitement avec le nombre de points contenu dans les différentes régions.



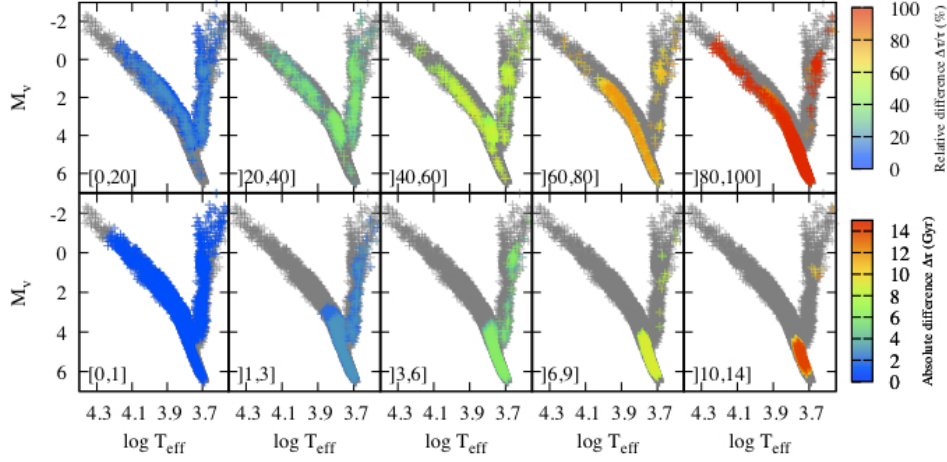


FIGURE 5.5 – Diagrammes HR montrant la comparaison entre l'âge réel simulé et celui déterminé par l'inférence bayésienne et la densité d'étoiles le long des trajets évolutifs, en fonction de l'erreur relative  $\Delta\tau/\tau$  (figures du haut) et l'erreur absolue  $\Delta\tau$  (figures du bas). Les couleurs sont les mêmes que celles de la figure 5.2.

### 5.3.2 Métallicités

Le nombre de métallicités des trajets évolutifs qui sont prises en compte dans la détermination de l'âge dépend de l'erreur observationnelle sur la métallicité et de l'intervalle de métallicité que nous prenons en compte.

Dans notre méthode, nous prenons tous les trajets évolutifs dont les métallicités sont incluses dans l'intervalle  $[M/H] \pm 3\sigma_{[M/H]}$ . Nous déterminons l'âge des étoiles du catalogue simulé de référence avec deux autres intervalles :  $1\sigma_{[M/H]}$  et  $2\sigma_{[M/H]}$  (figure 5.6.a). Les âges sont mieux restitués avec l'intervalle  $1\sigma_{[M/H]}$  car il y a moins de trajets évolutifs pris en compte dans le calcul. Les calculs avec 2 et  $3\sigma_{[M/H]}$  donnent des résultats similaires. Dans notre méthode de datation nous utilisons l'intervalle de  $3\sigma_{[M/H]}$  afin de prendre en compte entièrement l'erreur observationnelle sur la métallicité.

De la même façon lorsque nous modifions l'erreur observationnelle sur la métallicité nous modifions le nombre de métallicités des trajets évolutifs qui sont utilisés pour déterminer l'âge. La figure 5.6.b présente les résultats de la datation des étoiles du catalogue simulé de référence avec cinq erreurs observationnelles sur la métallicité (0.01, 0.1, 0.2, 0.3 et 0.5 dex). Nous constatons que plus l'erreur observationnelle est petite, mieux les âges sont restitués. Pour avoir des âges bien déterminés nous avons besoin de déterminer la métallicité de façon précise.

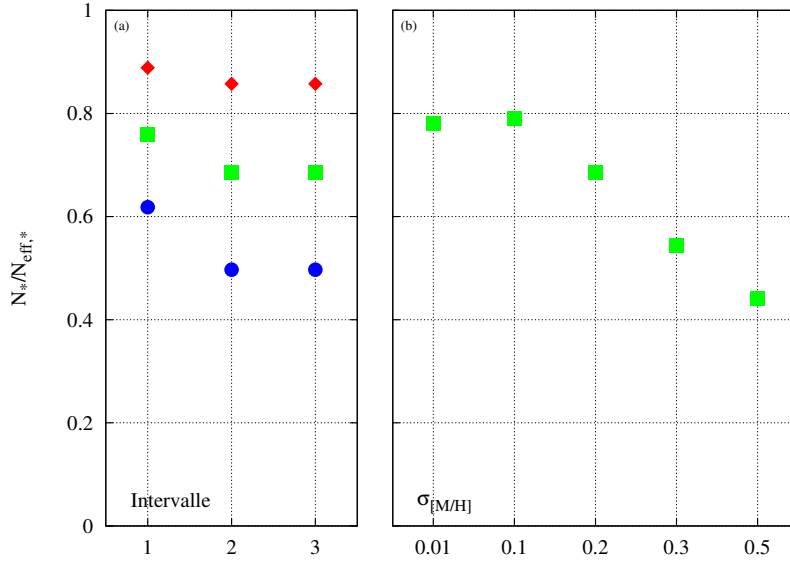


FIGURE 5.6 – Résultats de la comparaison entre l’âge réel simulé et l’âge déterminé pour les étoiles du catalogue simulé de référence. On utilise différents intervalles  $\widetilde{[M/H]} \pm n\sigma_{[M/H]}$  (figure a) et différentes erreurs observationnelles sur la métallicité (figure b). Les couleurs sont les mêmes que celles de la figure 2.3

### 5.3.3 Pas en âge ou $\log(\text{âge})$

Pour déterminer les âges, nous travaillons avec un pas en âge constant, mais certaines grilles de trajets évolutifs donnent les âges en  $\log(\text{âge})$ . Pour analyser l’impact sur les âges, nous déterminons les âges avec un pas en  $\log(\text{âge})$  constant. Nous constatons qu’il y a peu de différence entre les deux approches (figure 5.7). Il y a des différences entre les âges obtenues pour les étoiles qui sont localisées dans les zones de dégénérescence c’est à dire à proximité la séquence principale d’âge zéro et pour les petites masses sur la séquence principale ( $m < 1.0 M_{\odot}$ ).

Ici nous avons choisi de prendre un pas en âge constant. Notons que la répartition des points est différente dans les deux cas et qu’avec un pas en  $\log(\text{âge})$  constant, plus de points sont introduits aux petits âges.

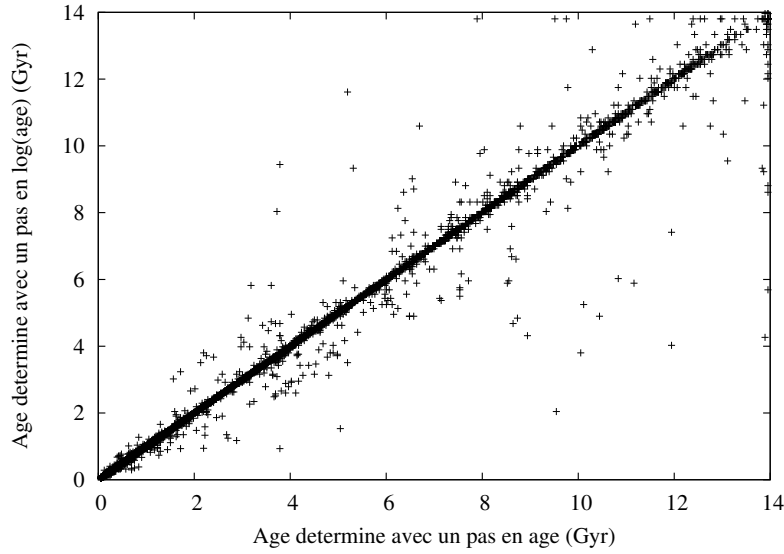


FIGURE 5.7 – Comparaison entre les âges déterminés avec un pas en âge constant et ceux déterminés avec un pas en  $\log(\text{âge})$  constant.

## 5.4 Détermination de la masse

Nous déterminons la masse des étoiles du catalogue simulé de référence par l'inférence bayésienne décrite dans la partie 2.6.1. Il y a 98 % des étoiles avec une masse bien déterminée avec  $\Delta m/m < 20\%$  (figure 5.8). Les quelques étoiles qui ont une masse mal-déterminée sont localisées dans la branche des géantes rouges (figure 5.8).

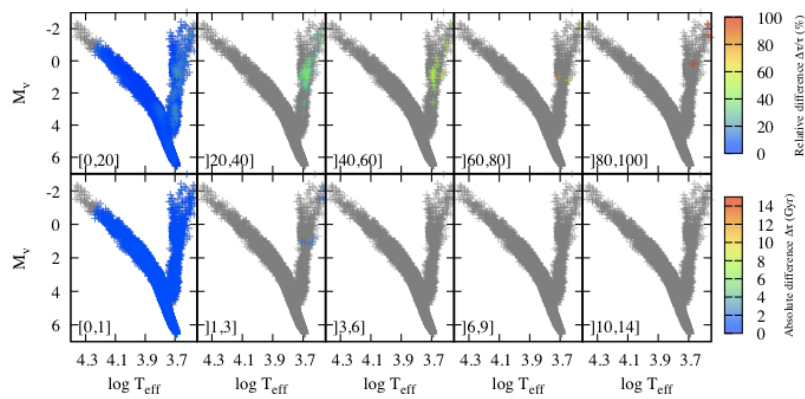


FIGURE 5.8 – Comparaison entre la masse réelle simulée et celle déterminée pour les étoiles du catalogue simulé de référence. Les masses sont déterminées par l'inférence bayésienne décrite dans la partie 2.6.1. Les couleurs sont les mêmes que celles de la figure 5.2.

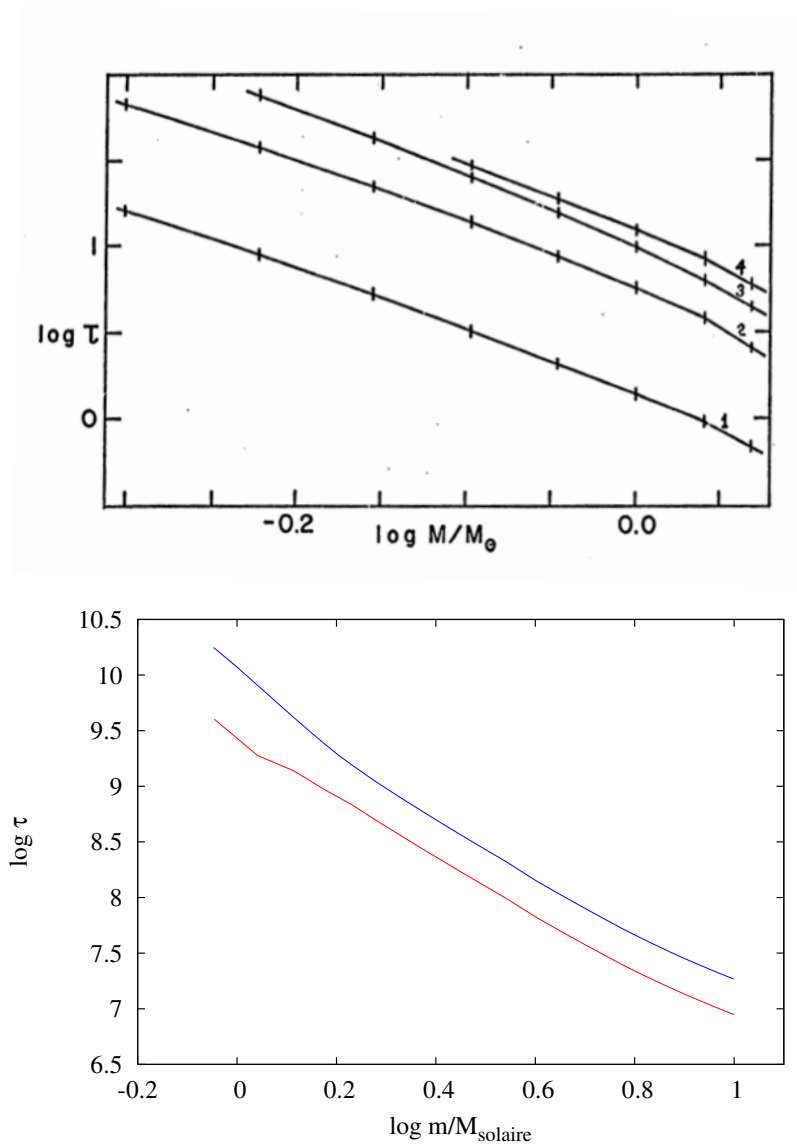


FIGURE 5.9 – *Figure du haut* : relation entre  $\log \tau$ - $\log(m/M_{\odot})$  pour les isochrones de Prather (1976). Chaque courbe correspond à un stade d'évolution : la courbe 1 correspond à la séquence principale d'âge zéro, la courbe 2 correspond au moment où le cœur de l'étoile a une fraction de masse de l'hydrogène égale à 0.30 et pour la courbe 3 la fraction de masse de l'hydrogène du cœur de l'étoile est de 0.0002. La courbe 4 correspond au moment où la masse du cœur de l'étoile est de  $0.150 M_{\odot}$ . *Figure du bas* : relation entre  $\log \tau$ - $\log(m/M_{\odot})$  pour les trajets évolutifs de BaSTI à une métallicité solaire. La courbe rouge correspond à la séquence principale et la bleue à la branche des géantes rouges.

La masse est mieux restituée que l'âge car la masse et l'âge ne varient pas de la même façon dans le diagramme HR. Prather (1976) montre que la variation entre le log de l'âge et log de la masse est linéaire, en moyenne  $d\tau/\tau = -4 dm/m$  (figure 5.9, haut). Avec les trajets de CESAM nous trouvons une variation de  $d\tau/\tau = -3 dm/m$  en moyenne, et  $d\tau/\tau = -2 dm/m$  pour les trajets évolutifs de BaSTI (figure 5.9, bas). Par exemple dans le bas de la séquence principale les masses disponibles dans une petite région du diagramme HR vont de 1 à  $1.2 M_{\odot}$ , alors que les âges vont de 0.03 Gyr à 9 Gyr dans la même région.

La masse est déterminée sans IMF afin de ne pas introduire de biais en faveur des petites masses. Pour vérifier cela, nous avons déterminé la masse avec l'IMF de Kroupa (2002). Nous trouvons 78 % des étoiles avec  $\Delta m/m < 20$  % (figure 5.10). L'histogramme des masses réelles et des masses déterminées (figure 5.11) montre que les masses déterminées avec l'IMF de K02 ne respectent pas la distribution des masses contrairement aux masses déterminées avec une IMF constante.

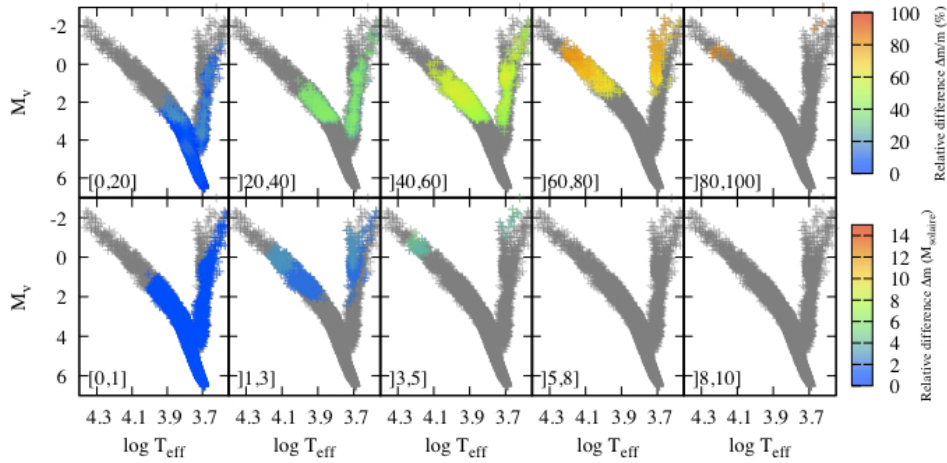


FIGURE 5.10 – Comparaison entre la masse réelle simulée et celle déterminée pour les étoiles du catalogue simulé de référence. Les masses sont déterminées par l'inférence bayésienne et l'IMF de K02 décrite dans cette partie. Les couleurs sont les mêmes que celles de la figure 5.2.

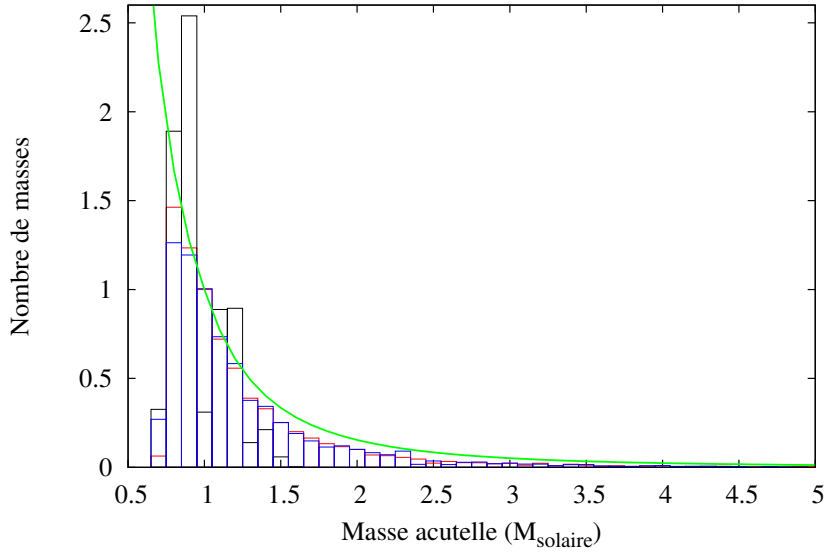


FIGURE 5.11 – Histogramme des masses réelles simulées en rouge, en bleu celui des masses déterminées par l’inférence bayésienne et une IMF monotone et en noir celui des masses déterminées avec l’inférence bayésienne et l’IMF de K02. L’IMF de K02 que nous avons utilisée pour construire le catalogue simulé de référence est représentée en vert.

## 5.5 Comparaisons des méthodes

Nous comparons les méthodes d’inférence bayésienne et de minimisation du  $\chi^2$  présentées dans le chapitre 3. Nous déterminons l’âge des étoiles du catalogue simulé de référence avec toutes les méthodes. Dans un premier temps, nous comparons les inférences bayésiennes (figures 5.12 et 5.13). L’inférence bayésienne de notre travail et les méthodes de JL05 et dSGP06 donnent des résultats similaires, soit environ 69 % des âges déterminés avec  $\Delta\tau/\tau < 20\%$ . Les différences entre ces trois méthodes ont peu d’effets sur la datation des étoiles. En revanche, pour la méthode de CSA11 nous obtenons des résultats différents, il y a 20 % d’étoiles avec  $\Delta\tau/\tau < 20\%$ . La différence est due au traitement de la parallaxe. Dans notre catalogue simulé, nous avons déterminé l’erreur observationnelle sur la magnitude absolue en supposant qu’elle était gaussienne et en prenant en compte l’extinction, la parallaxe et la magnitude apparente. Dans leur méthode CSA11 supposent que l’erreur observationnelle sur la parallaxe est gaussienne. Nous ne pouvons pas conclure sur les résultats de CSA11 par rapport aux autres méthodes car nous ne traitons pas les erreurs de la même façon.

La comparaison entre les deux minimisations du  $\chi^2$  (figures 5.12 et 5.14)

montre que les résultats sont similaires. La minimisation du  $\chi^2$  de NB98 donne des résultats légèrement meilleurs car elle ne prend en compte qu'une seule métallicité. Pour déterminer l'âge de façon rigoureuse, il faut utiliser toutes les métallicités qui sont incluses dans la barre d'erreur de la métallicité observée. Nous utilisons le  $\chi^2$  standard pour effectuer la comparaison avec l'inférence bayésienne de notre travail.

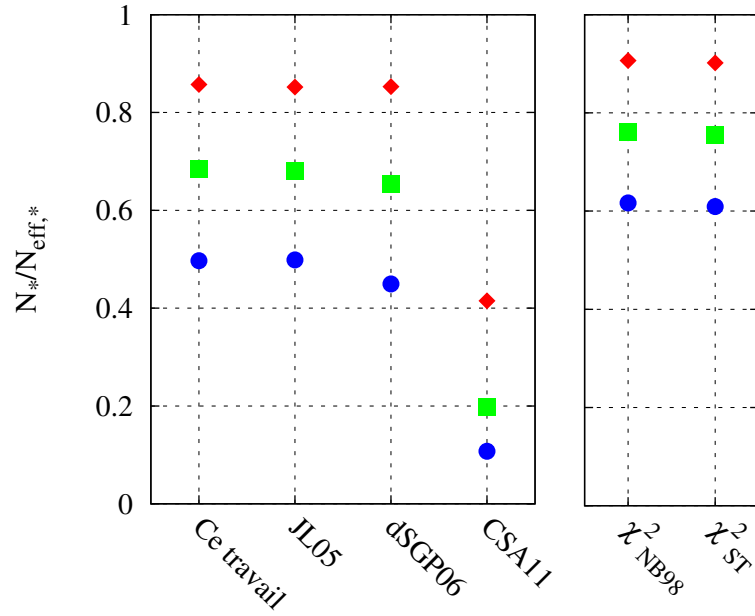


FIGURE 5.12 – Histogramme de la comparaison des inférences bayésiennes et des minimisations du  $\chi^2$  pour des âges déterminés avec le catalogue simulé de référence. Les couleurs sont les mêmes que celles de la figure 2.3.

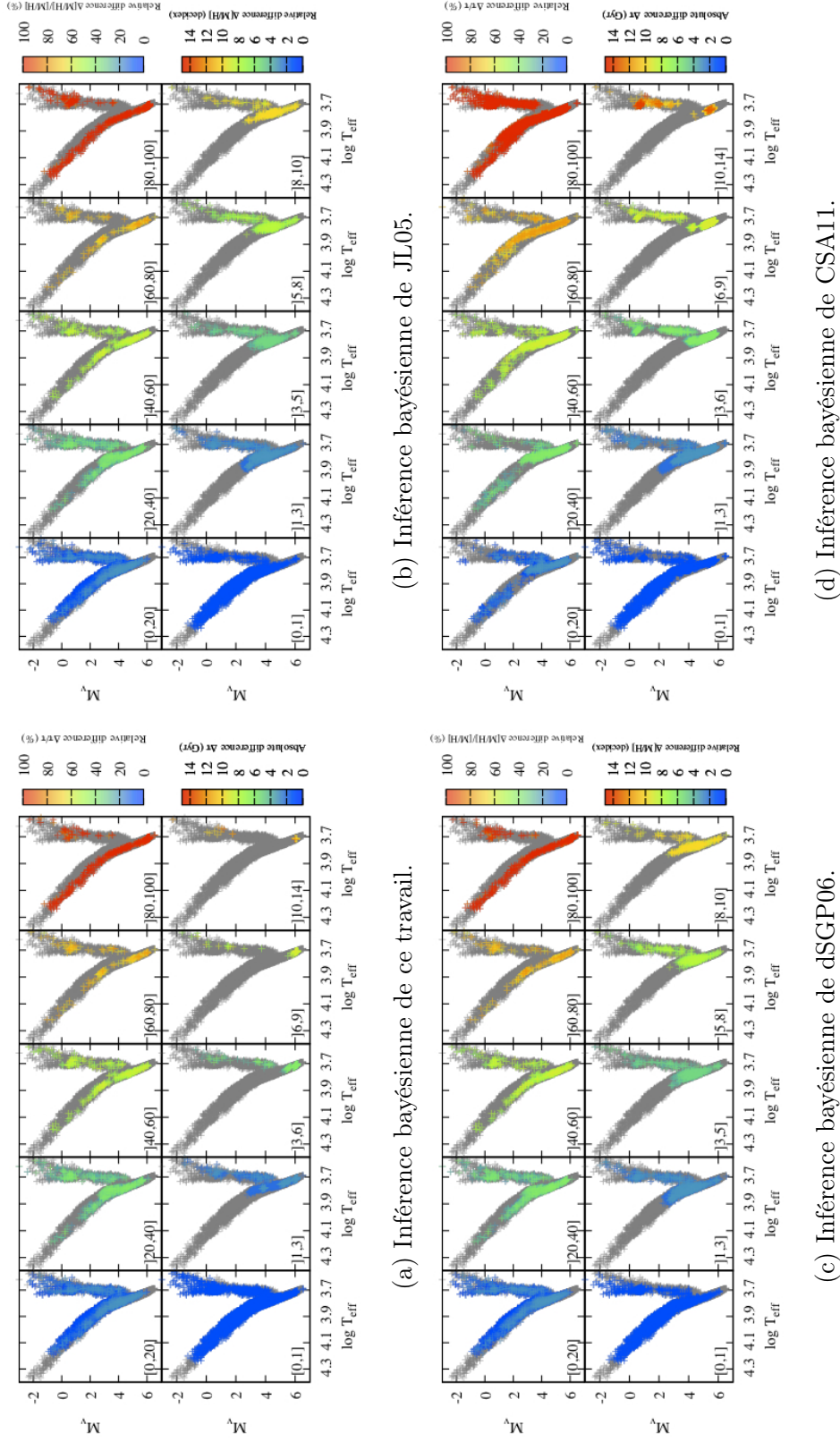


FIGURE 5.13 – Comparaison entre l'âge réel simulé et l'âge déterminé pour les étoiles du catalogue simulé de référence. Les âges sont déterminés par les quatre méthodes d'inference bayésienne. Les couleurs sont les mêmes que celles de la figure 5.2.



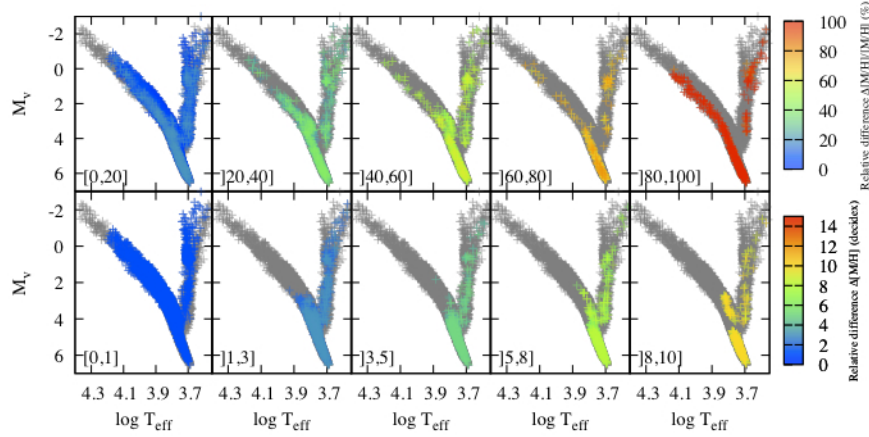
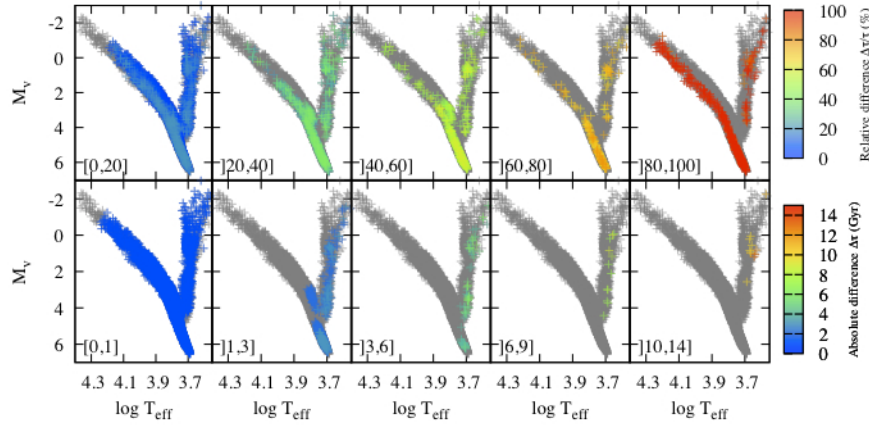
(a) Minimisation du  $\chi^2$  de NB98.(b) Minimisation du  $\chi^2$  standard.

FIGURE 5.14 – Comparaison entre l'âge réel simulé et l'âge déterminé pour les étoiles du catalogue simulé de référence. Les âges sont déterminés par les deux méthodes de minimisation du  $\chi^2$ . Les couleurs sont les mêmes que celles de la figure 5.2.

Pour faire la comparaison entre l'inférence bayésienne et la minimisation du  $\chi^2$ , nous travaillons avec trois catalogues simulés : celui de référence où les étoiles sont sélectionnées dans les trajets évolutifs de BaSTI (figure 4.1), le catalogue simulé à partir des trajets évolutifs de CESAM (figure 4.4) et le catalogue simulé à partir des trajets évolutifs de Padoue (figure 4.5). Nous calculons l'âge des étoiles des trois catalogues avec les trajets évolutifs de BaSTI, de CESAM et de Padoue. Les résultats sont présentés sur les figures 5.15, 5.16, 5.17, 5.18, 5.19, 5.20 et 5.21.

Lorsque nous utilisons un catalogue simulé qui est basé sur les mêmes trajets évolutifs que la détermination de l'âge, nous trouvons que la minimisa-

tion du  $\chi^2$  donne de meilleurs résultats que l'inférence bayésienne. En revanche lorsque les trajets évolutifs utilisés sont différents, l'inférence bayésienne donne de meilleurs résultats. La minimisation du  $\chi^2$  est donc plus sensible aux grilles d'évolution stellaire que l'inférence bayésienne. En effet nous avons déterminé l'âge de 10 000 points exacts des trajets évolutifs de BaSTI et la minimisation du  $\chi^2$  permet de restituer correctement 98 % des âges avec  $\Delta\tau/\tau < 20\%$ .

Il est difficile de comparer l'inférence bayésienne avec la minimisation du  $\chi^2$  car les deux méthodes sont différentes. La minimisation du  $\chi^2$  détermine les paramètres de façon conjointe alors que l'inférence bayésienne les détermine de façon marginale.

Nous avons déterminé de façon conjointe l'âge par l'inférence bayésienne. Pour cela au lieu de faire la somme sur la métallicité et la masse, nous recherchons la valeur la plus petite de l'expression de la fonction densité de probabilité *a posteriori* de l'équation 2.7. Nous trouvons 58 % des étoiles avec  $\Delta\tau/\tau < 20\%$ , cela reste moins bon que la minimisation du  $\chi^2$  et que l'inférence bayésienne.

L'erreur observationnelle sur la parallaxe peut entraîner une dispersion supplémentaire et faire que l'erreur observationnelle sur la magnitude absolue ne soit pas symétrique. Or dans notre méthode, nous supposons que l'erreur observationnelle sur la magnitude absolue est gaussienne et symétrique. Afin de vérifier que les différences entre la minimisation du  $\chi^2$  et de l'inférence proviennent de cet effet, il faudrait modifier la vraisemblance afin de prendre en compte cet effet, comme dans l'inférence bayésienne de Casagrande et al. (2011).

Actuellement, les connaissances en structure interne et en physique stellaire ne permettent pas d'avoir des modèles stellaires qui reproduisent parfaitement la réalité. Le cas où le calcul de l'âge et le catalogue simulé n'utilisent pas les mêmes trajets évolutifs semble donc mieux correspondre à la réalité. C'est pourquoi, pour l'instant nous recommandons d'utiliser l'inférence bayésienne qui permet de minimiser les effets de la différence entre les modèles stellaires grâce aux priors.

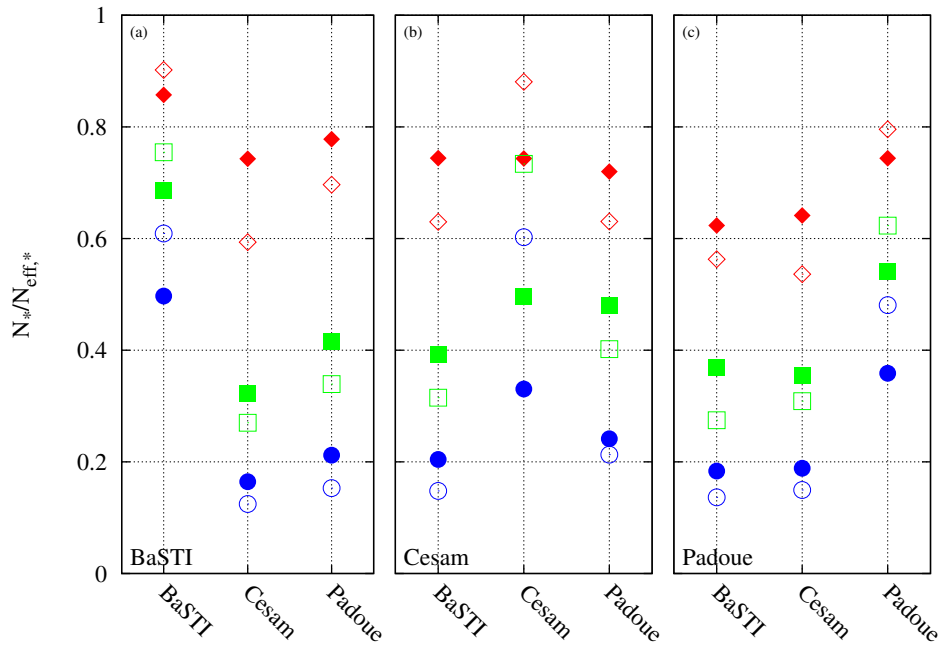


FIGURE 5.15 – Histogramme de la comparaison entre l’inférence bayésienne utilisée et de la minimisation du  $\chi^2$  standard pour trois types de trajets évolutifs (BaSTI et CESAM). La comparaison est faite pour trois catalogues simulés à partir des trajets évolutifs de BaSTI (figure a), de CESAM (figure b) et de Padoue (figure c).  $N_{\text{eff},*}$  représente le nombre total d’étoiles d’âge déterminé.  $N_*$  est le nombre d’étoiles avec  $\Delta\tau/\tau < 10\%$  (cercle bleu),  $< 20\%$  (carré vert) et  $< 50\%$  (losange rouge). L’inférence bayésienne est représentée par les symboles au fond coloré et la minimisation du  $\chi^2$  par les symboles au fond blanc.

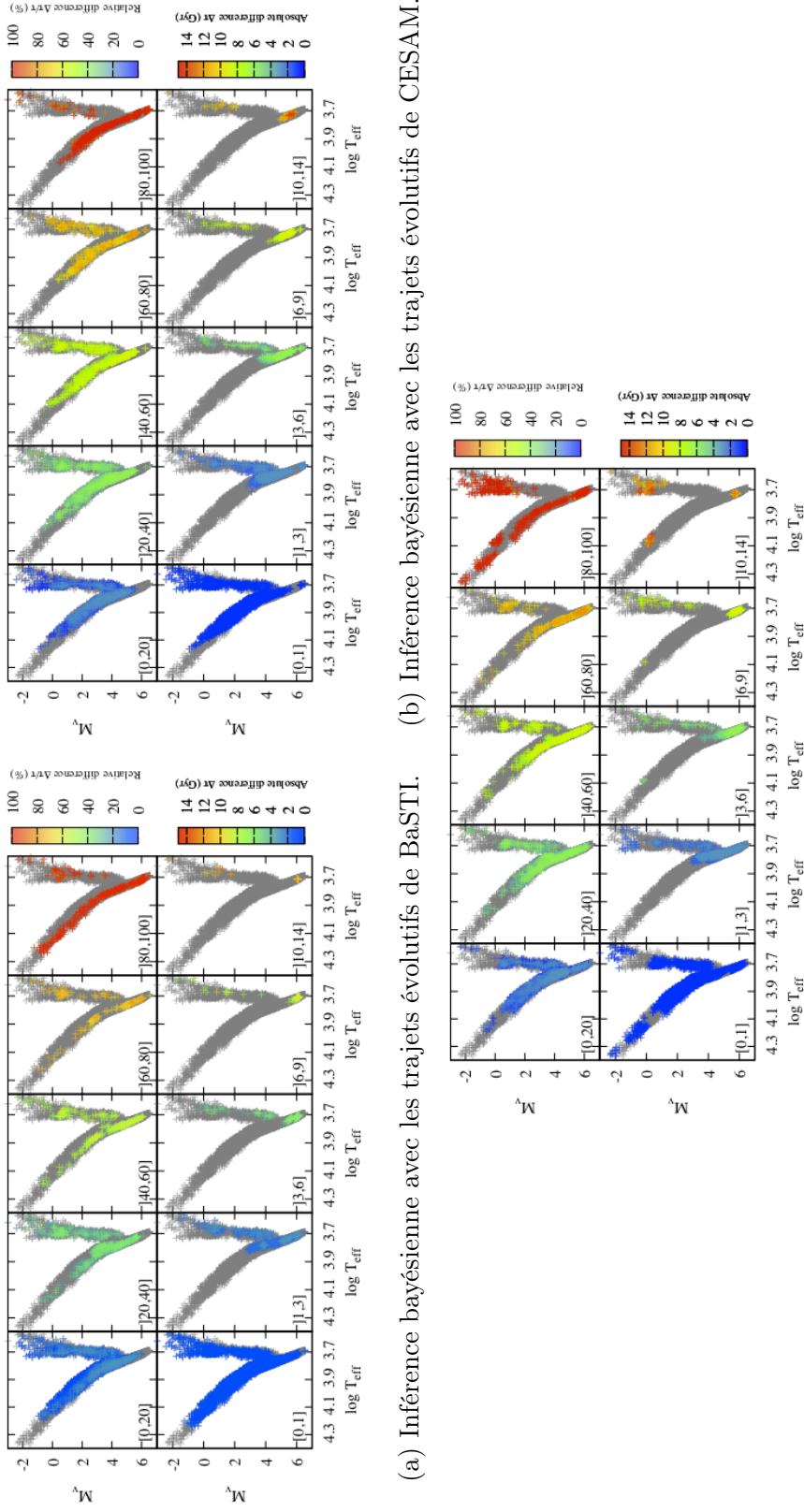
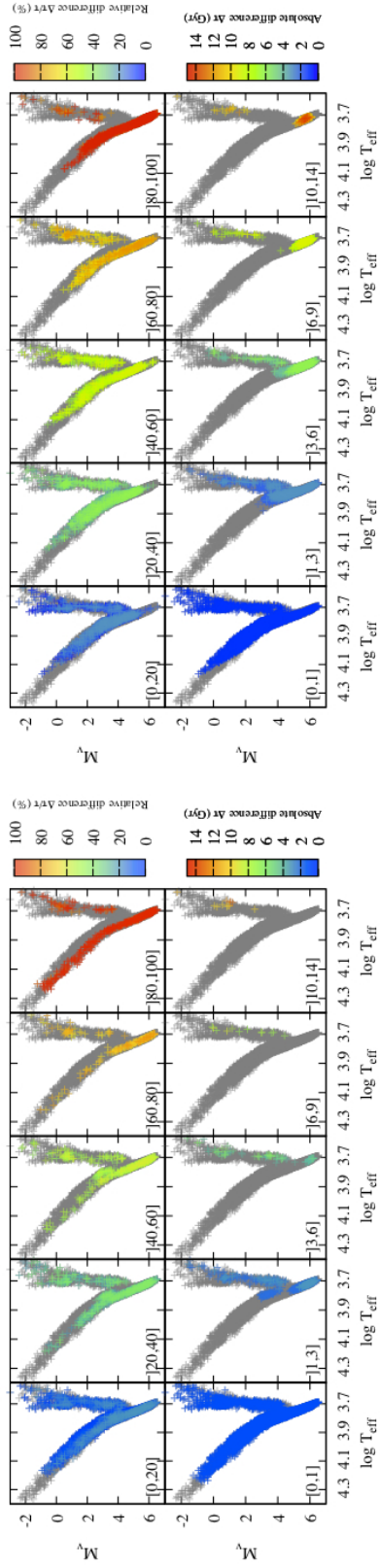
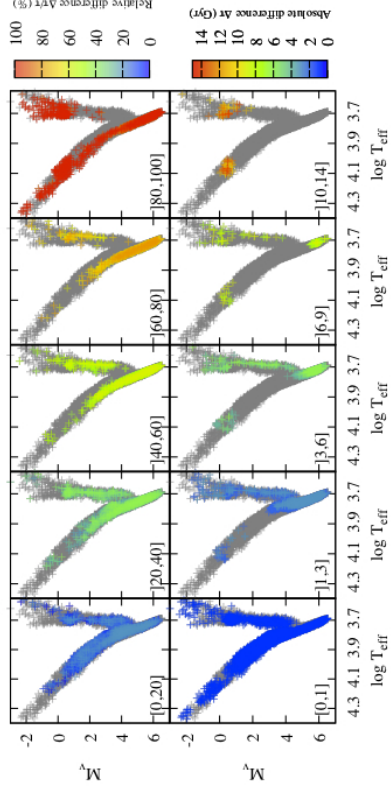


FIGURE 5.16 – Comparaison entre l'âge réel simulé et l'âge déterminé pour les étoiles du catalogue simulé de référence avec les trajets évolutifs de BaSTI. Les âges sont déterminés par l'inférence bayésienne de ce travail avec les trajets évolutifs de BaSTI, de CESAM et de Padoue. Les couleurs sont les mêmes que celles de la figure 5.2.



(a) Minimisation du  $\chi^2$  avec les trajets évolutifs de BaSTI. (b) Minimisation du  $\chi^2$  avec les trajets évolutifs de CESAM.



(c) Minimisation du  $\chi^2$  avec les trajets évolutifs de Padoue.

FIGURE 5.17 – Comparaison entre l'âge réel simulé et l'âge déterminé pour les étoiles du catalogue simulé de référence avec les trajets évolutifs de BaSTI. Les âges sont déterminés par la minimisation du  $\chi^2$  standard avec les trajets évolutifs de BaSTI, de CESAM et de Padoue. Les couleurs sont les mêmes que celles de la figure 5.2.

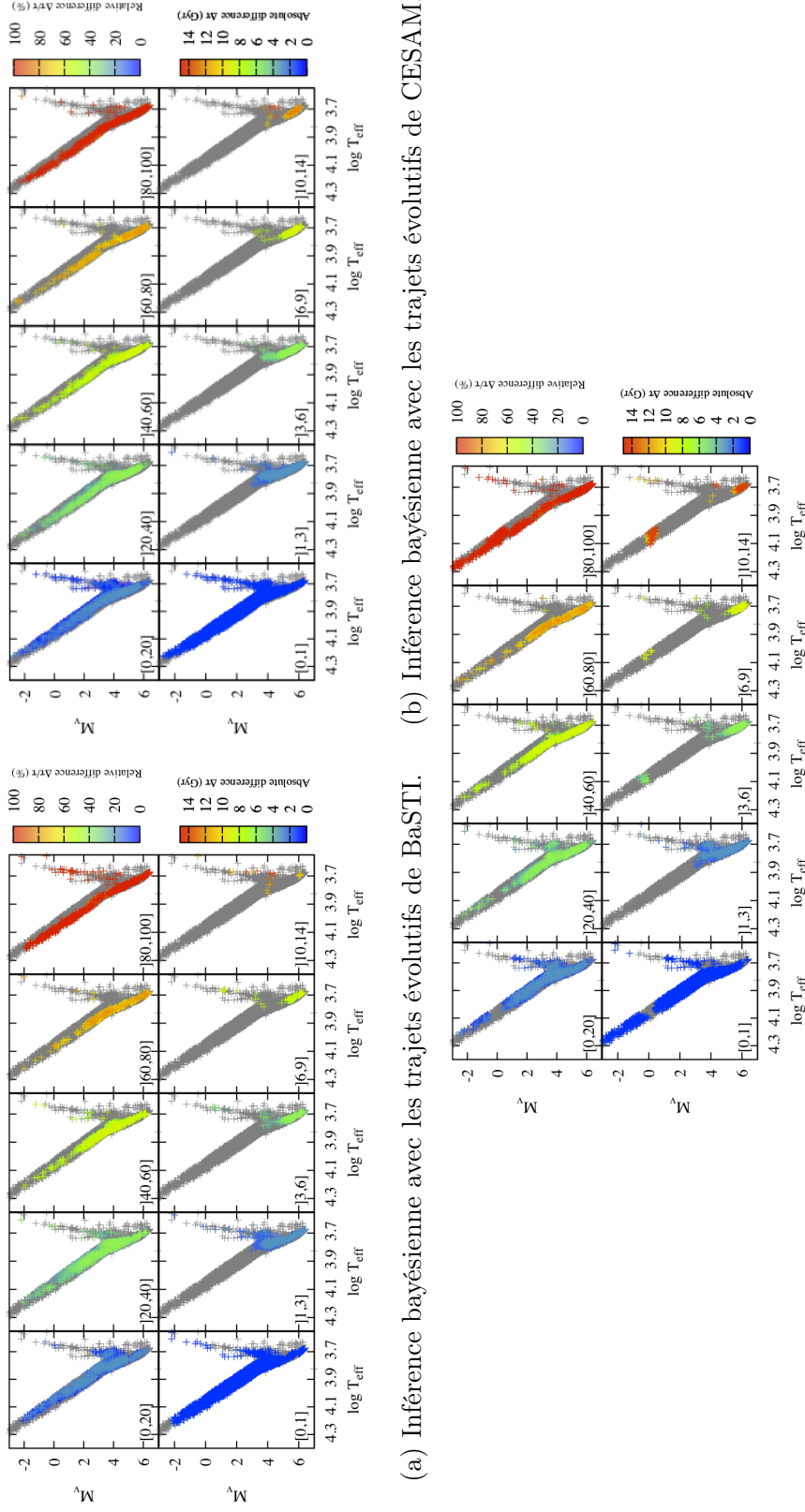
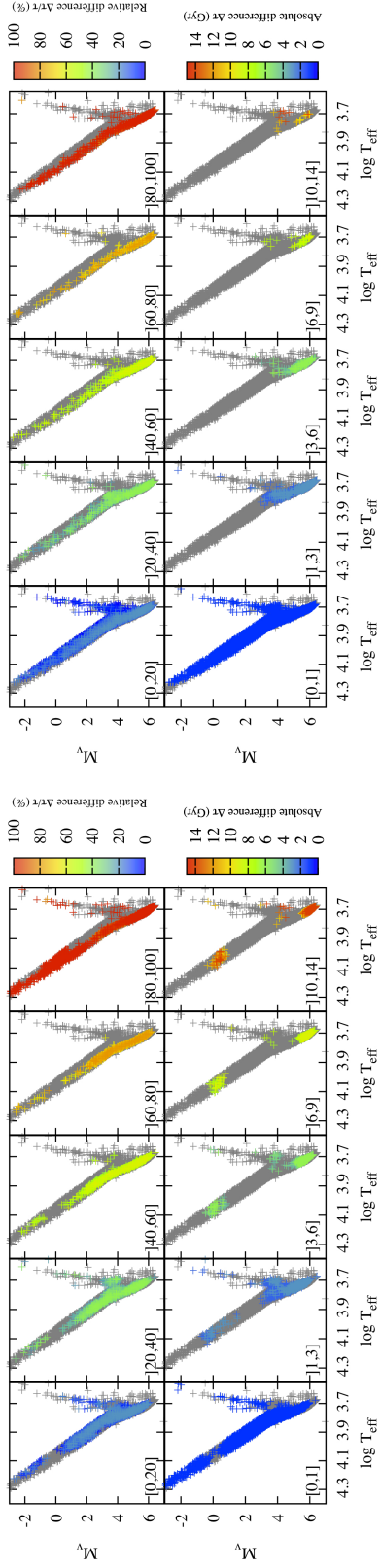
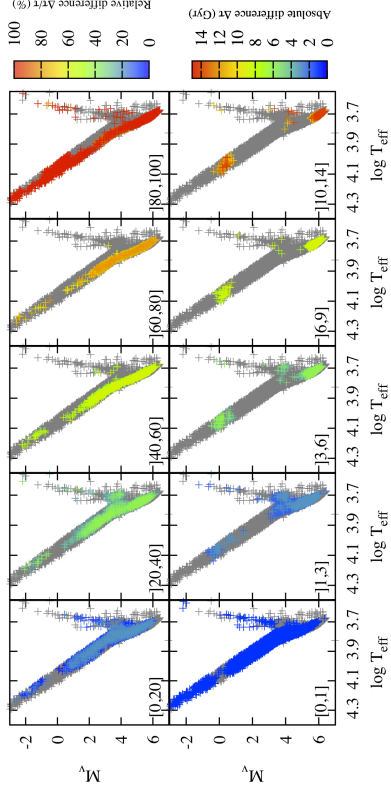


FIGURE 5.18 – Comparaison entre l'âge réel simulé et l'âge déterminé pour les étoiles du catalogue simulé de référence avec les trajets évolutifs de CESAM. Les âges sont déterminés par l'inférence bayésienne de ce travail avec les trajets évolutifs de BaSTI, de CESAM et de Padoue. Les couleurs sont les mêmes que celles de la figure 5.2.

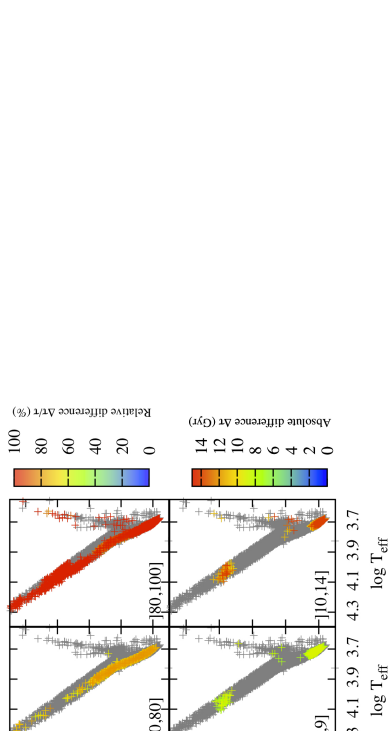




(a) Minimisation du  $\chi^2$  avec les trajets évolutifs de BaSTI.



(b) Minimisation du  $\chi^2$  avec les trajets évolutifs de CESAM.



(c) Minimisation du  $\chi^2$  avec les trajets évolutifs de Padoue.

FIGURE 5.19 – Comparaison entre l'âge réel simulé et l'âge déterminé pour les étoiles du catalogue simulé de référence avec les trajets évolutifs de CESAM. Les âges sont déterminés par la minimisation du  $\chi^2$  standard avec les trajets évolutifs de BaSTI, de CESAM et de Padoue. Les couleurs sont les mêmes que celles de la figure 5.2.

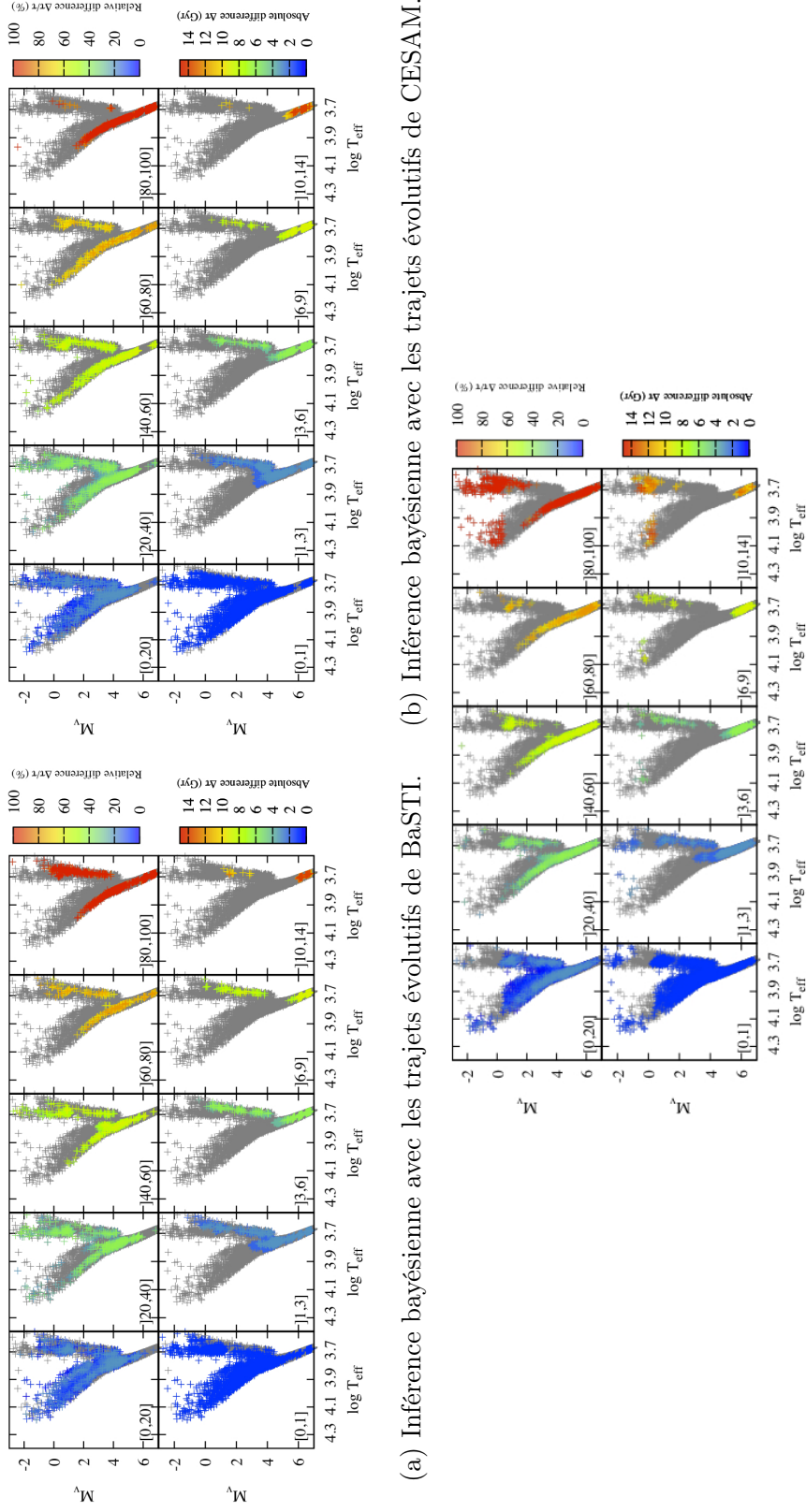


FIGURE 5.20 – Comparaison entre l'âge réel simulé et l'âge déterminé pour les étoiles du catalogue simulé de référence avec les trajets évolutifs de Padoue. Les âges sont déterminés par l'inférence bayésienne de ce travail avec les trajets évolutifs de BaSTI, de CESAM et de Padoue. Les couleurs sont les mêmes que celles de la figure 5.2.



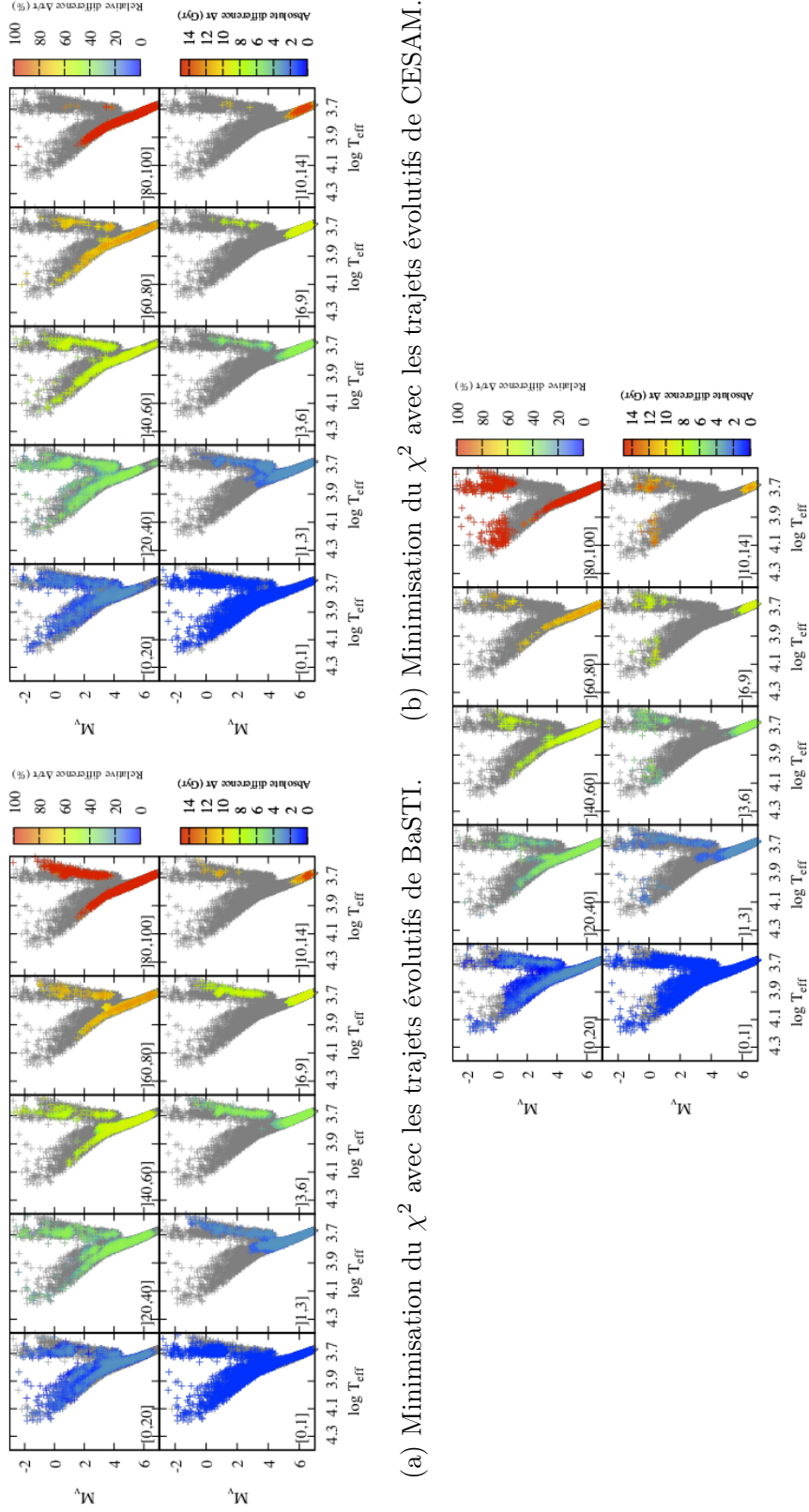


FIGURE 5.21 – Comparaison entre l’âge réel simulé et l’âge déterminé pour les étoiles du catalogue simulé de référence avec les trajets évolutifs de Padoue. Les âges sont déterminés par la minimisation du  $\chi^2$  standard avec les trajets évolutifs de BaSTI, de CESAM et de Padoue. Les couleurs sont les mêmes que celles de la figure 5.2.

## 5.6 Conclusions et perspectives

La méthode de détermination des âges que nous avons développé permet de restituer correctement l'âge de 69 % des étoiles. Par contre, il y a de nombreuses dégénérescences dans certaines régions du diagramme HR : à proximité de la séquence principale d'âge zéro et dans la branche des géantes rouges. Le traitement que nous avons réalisé sur les fonctions  $\psi(\tau|x)$  nous permet de ne pas dater des étoiles dont on sait que l'âge sera faux. Nous constatons aussi que ces étoiles sont localisées dans les régions de dégénérescence.

L'analyse des paramètres de la méthode nous montre qu'il existe des biais sur la détermination des âges, c'est pourquoi nous avons choisi de ne pas inclure la densité d'étoiles le long des trajets évolutifs et de garder un pas en âge constant. Nous constatons aussi que le nombre de métallicités pris en compte dans la détermination des âges influe beaucoup sur les résultats, c'est pourquoi nous avons besoin d'avoir une bonne précision sur la métallicité. Le mieux serait d'utiliser des métallicités déterminées par analyse spectroscopique.

L'âge et la métallicité ont une variation moyenne semblable, nous trouvons donc les mêmes dégénérescences dans la détermination de l'âge et de la métallicité. En revanche, la masse est mieux restituée que les âges car les masses ont une variation moyenne plus faible que l'âge et la métallicité.

Dans cette partie, nous avons également comparé plusieurs méthodes de datation. Dans un premier temps, nous avons étudié les différences entre les méthodes d'inférence bayésienne. Nous constatons que les faibles différences entre notre méthode et celles de JL05 et dSGP06 n'ont aucun effet sur la détermination de l'âge. En revanche, la méthode de CSA11 donne des résultats différents car les erreurs observationnelles dans le catalogue simulé ne sont pas traitées de la même façon que dans la méthode de CSA11. Pour étudier les résultats de cette méthode il faudrait construire un catalogue avec le même type d'erreurs observationnelles.

Nous avons ensuite effectué une comparaison croisée entre la minimisation du  $\chi^2$  et l'inférence bayésienne avec trois grilles de modèles d'évolution stellaire. Nous avons constaté que la minimisation du  $\chi^2$  est plus sensible aux modèles d'évolution stellaire. Pour l'instant il vaut mieux favoriser l'inférence bayésienne pour déterminer les âges car les grilles d'évolution stellaire ne reproduisent pas parfaitement l'évolution des étoiles.

Il y a d'autres avantages à utiliser l'inférence bayésienne notamment avec les fonctions  $\psi(\tau|x)$ . Celles ci permettent de ne pas donner d'âge aux étoiles quand nous savons qu'ils seront faux, grâce au repérage des fonctions  $\psi(\tau|x)$  mal définies. Les fonctions  $\psi(\tau|x)$  apportent des informations supplémentaires comme par exemple la levée de l'ambiguïté entre plusieurs âges. Elles peuvent être utilisées afin de dater les amas, ceci est plus détaillé dans la partie 8.4.

# Ajouts de contraintes observationnelles

---

## Sommaire

<b>6.1</b>	<b>Astérosismologie . . . . .</b>	<b>87</b>
<b>6.2</b>	<b>Transits d'exoplanètes . . . . .</b>	<b>92</b>
<b>6.3</b>	<b>Gravité de surface . . . . .</b>	<b>93</b>
<b>6.4</b>	<b>Conclusions et perspectives . . . . .</b>	<b>96</b>

---

Afin d'essayer de réduire les dégénérescences dans la détermination des âges, nous incluons des paramètres supplémentaires. Ces contraintes sont ajoutées dans la vraisemblance (équation 2.4), qui dépend normalement de trois observables ( $\widetilde{\log T_{\text{eff}}}$ ,  $\widetilde{M_V}$  et  $\widetilde{[M/H]}$ ). Lorsque nous ajoutons des contraintes dans la détermination des âges nous devons changer la valeur du  $\chi_{\text{max}}^2$  pour le test de rejet sur le  $\chi^2$ , qui vérifie que l'étoile est bien incluse dans les trajets évolutifs<sup>1</sup>.

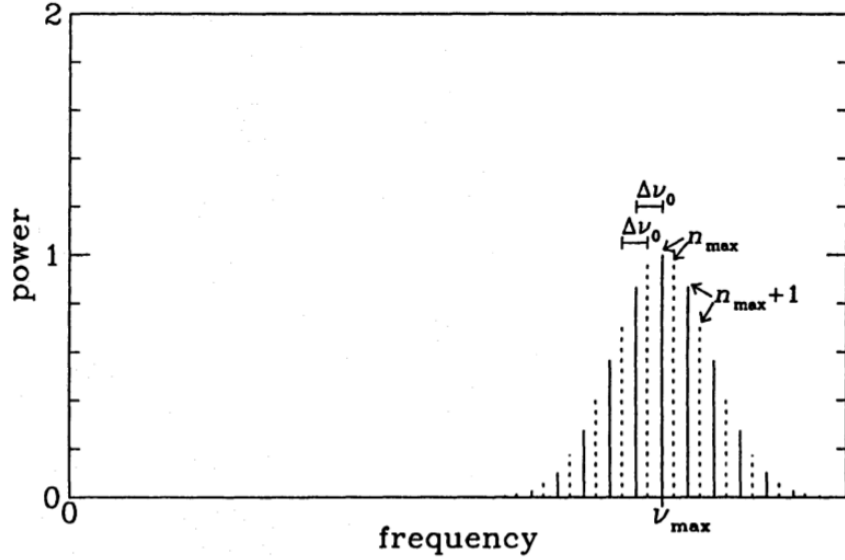
Les contraintes proviennent d'observations complémentaires, nous considérons les quantités obtenues par les missions spatiales CoRoT (Baglin et al. 2006) et Kepler (Borucki et al. 2004). La mission spatiale CoRoT (Baglin et al. 2006, CONvection, ROTation et Transits planétaires) du CNES, lancée le 27 décembre 2006, a réalisé de la photométrie stellaire de très grande précision sur quatre capteurs dont deux étaient dédiés à l'astérosismologie et les deux autres à la recherche des planètes extrasolaires. CoRoT est le premier télescope spatial destiné à la recherche d'exoplanètes. Le satellite a observé alternativement pendant six mois deux régions du ciel : en direction du centre et de l'anti-centre de la Galaxie.

Le satellite Kepler (Borucki et al. 2004) de la NASA observe une seule région du ciel depuis son lancement le 7 mars 2009. Kepler réalise également de la photométrie stellaire de très grande précision afin d'étudier la sismologie des étoiles et de rechercher les planètes extrasolaires.

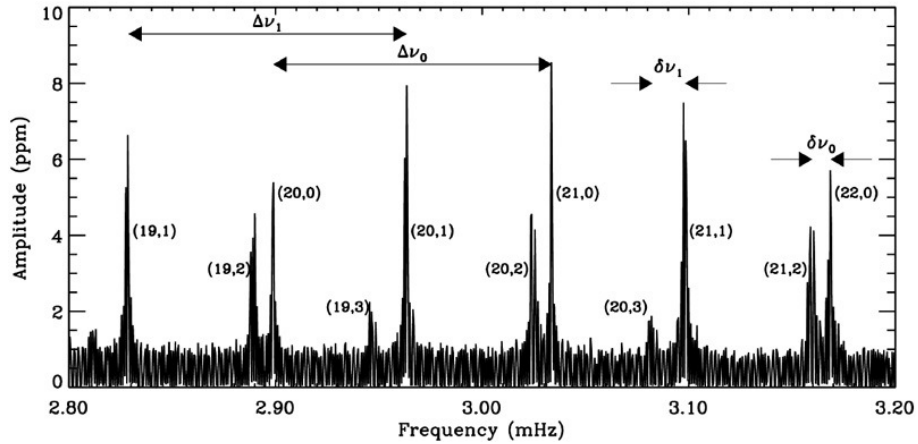
---

1. D'après la distribution du  $\chi^2$  décrite dans l'annexe C.4, pour un niveau de confiance de 0.01, nous prenons  $\chi_{\text{max}}^2 = 13.28$  pour quatre observables,  $\chi_{\text{max}}^2 = 15.09$  pour cinq et  $\chi_{\text{max}}^2 = 16.81$  pour six.

Dans ce chapitre, nous utilisons toujours le catalogue simulé de référence de 10 000 étoiles auquel nous ajoutons les contraintes considérées.



(a) Spectre de puissance entier.



(b) Zoom sur les oscillations de type solaire.

FIGURE 6.1 – Spectre de fréquences (figure a, Bedding & Kjeldsen 2003) avec un zoom (figure b, Kjeldsen & Bedding 1995) sur les oscillations de type solaire. Chaque pic caractérise un mode d'oscillation de fréquence  $\nu_{n,l}$  où  $n$  est l'ordre radial et  $l$  le degré du mode.  $\nu_{\max}$  est la fréquence au maximum de puissance. Les grandes séparations  $\Delta\nu_l$  correspondent à  $\Delta\nu_l = \nu_{n,l} - \nu_{n-1,l}$ . Les petites séparations  $\delta\nu_l$  sont telles que  $\delta\nu_l = \nu_{n,l} - \nu_{n-1,l+2}$ .

## 6.1 Astérosismologie

Les oscillations stellaires sont des ondes de pression ou de gravité se propageant à travers l'étoile et qui sont caractéristiques de ses propriétés internes. Elles dépendent ainsi des modes de vibrations propres de l'étoile. Ces modes permettent de sonder l'intérieur de l'étoile et de la caractériser. Ces modes sont définis par des degrés qui caractérisent la structure horizontale du mode et des ordres qui caractérisent la structure radiale.

Les fréquences des oscillations des étoiles peuvent être extraites des courbes de lumière observées par les missions spatiales CoRoT (Baglin et al. 2006) et Kepler (Borucki et al. 2004). Ceci permet d'obtenir le spectre de fréquences des étoiles (figure 6.1.a).

Les oscillations de type solaire sont observées sur une partie du spectre de fréquences, on les appelle ainsi car elles ont été détectées pour la première fois dans le Soleil. Ces oscillations sont dues à des mouvements stochastiques dans la zone convective externe. Un zoom sur le spectre de fréquence sur des oscillations de type solaire est représenté sur la figure 6.1.b. Les oscillations de type solaire sont observées dans les étoiles de faible masse sur la séquence principale et les étoiles de la branche des géantes rouges. Ces deux régions sont représentées sur la figure 6.2.

A partir du spectre de fréquences (figure 6.1.b), trois paramètres sismiques peuvent être déterminés : la grande séparation  $\Delta\nu$ , la fréquence du maximum de puissance  $\nu_{\max}$  et le maximum d'autocorrélation du signal  $A_{\max}$ . Ce dernier correspond à l'amplitude du signal à la fréquence  $\nu_{\max}$ . La grande séparation  $\Delta\nu_l$  correspond à la distance entre deux pics de même degré et d'ordre différent d'une unité. Généralement, nous utilisons la moyenne des grandes séparations déterminée sur tout le spectre pour tout les degrés, elle est notée  $\langle \Delta\nu \rangle$ .

Nous incluons ces trois contraintes dans la détermination de l'âge. Nous les calculons à partir des relations d'échelle établies par Kjeldsen & Bedding (1995); Huber et al. (2011). En effet l'étude des oscillations de type solaire permet de déterminer les propriétés internes des étoiles, les relations d'échelle relient les paramètres sismiques aux paramètres stellaires :

$$\frac{\langle \Delta\nu \rangle}{\langle \Delta\nu_{\odot} \rangle} = \left( \frac{m}{M_{\odot}} \right)^{\frac{1}{2}} \left( \frac{R}{R_{\odot}} \right)^{-\frac{3}{2}} \propto \left( \frac{\rho_{\text{moy}}}{\rho_{\text{moy},\odot}} \right)^{\frac{1}{2}}, \quad (6.1)$$

$$\frac{\nu_{\max}}{\nu_{\max,\odot}} = \left( \frac{m}{M_{\odot}} \right) \left( \frac{R}{R_{\odot}} \right)^{-2} \left( \frac{T_{\text{eff}}}{T_{\text{eff},\odot}} \right)^{-\frac{1}{2}} \propto \frac{g}{g_{\odot}} \left( \frac{T_{\text{eff}}}{T_{\text{eff},\odot}} \right)^{-\frac{1}{2}}, \quad (6.2)$$

$$\frac{A_{\max}}{A_{\max,\odot}} = \left( \frac{L}{L_{\odot}} \right)^{0.838} \left( \frac{m}{M_{\odot}} \right)^{-1.32} \left( \frac{T_{\text{eff}}}{T_{\text{eff},\odot}} \right)^{-2}. \quad (6.3)$$

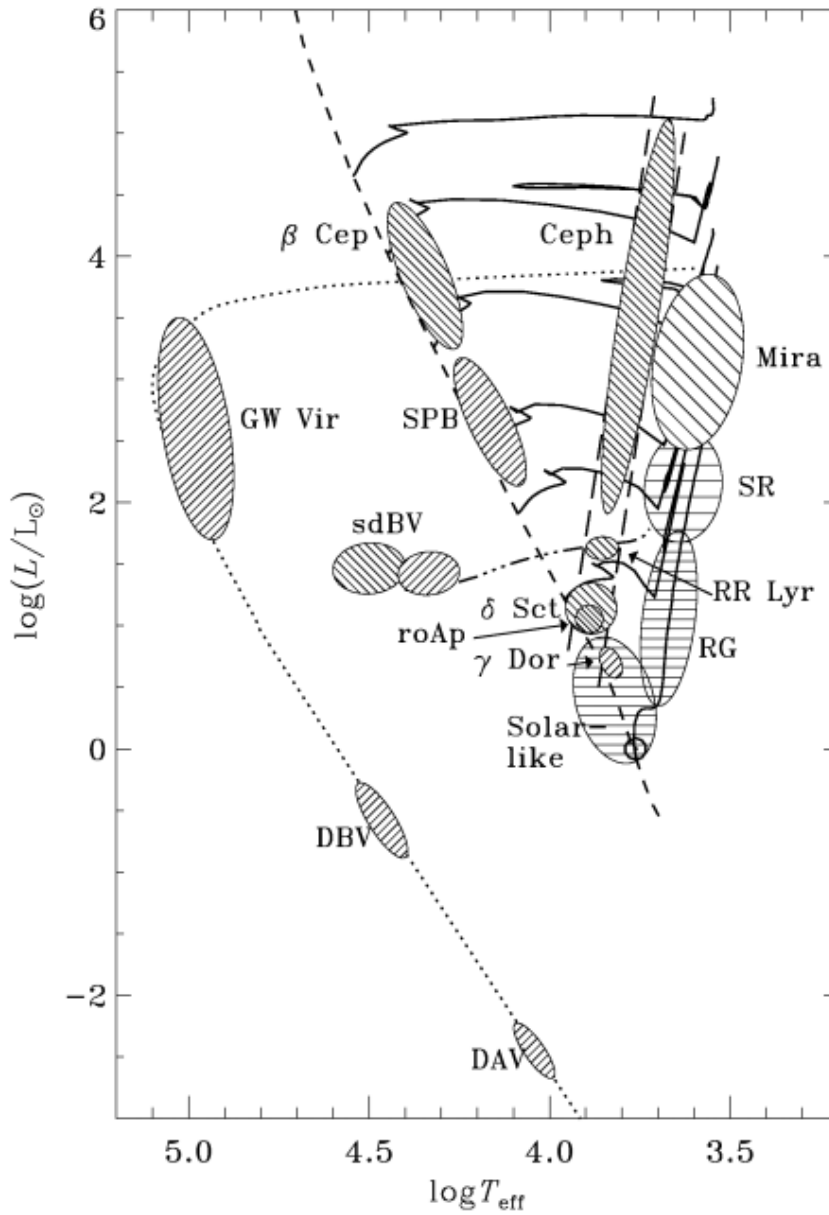


FIGURE 6.2 – Diagramme HR illustrant la position des différents oscillateurs (Christensen-Dalsgaard 2012). La courbe en tiretés représente la séquence principale d'âge zéro, la courbe en pointillés correspond à l'évolution des naines blanches lors du refroidissement et les courbes continues sont des trajets évolutifs de 1, 2, 3, 4, 7, 12 et 20  $M_{\odot}$ . Les étoiles avec des oscillations de type solaire (solar-like) sont localisées sur la séquence principale et dans la branche des géantes rouges.

Avec  $g$  la gravité de surface de l'étoile et  $\rho_{\text{moy}}$  sa densité moyenne. Les constantes solaires utilisées sont :  $\langle \Delta\nu_{\odot} \rangle = 134.9 \mu\text{Hz}$ ,  $\nu_{\text{max},\odot} = 3050.0 \mu\text{Hz}$  (Kjeldsen et al. 2008),  $A_{\text{max},\odot} = 2.5 \text{ ppm}$  (Huber et al. 2011),  $T_{\text{eff},\odot} = 5777.54 \text{ K}$ ,  $R_{\odot} = 6.9599 \cdot 10^5 \text{ km}$ ,  $M_{\odot} = 1.98919 \cdot 10^{30} \text{ kg}$  et  $L_{\odot} = 3.842 \cdot 10^{26} \text{ W}$  (Lebreton et al. 2008).

Nous devons également attribuer une erreur observationnelle à chacune des contraintes. Nous prenons  $\sigma_{A_{\text{max}}} = 0.098 A_{\text{max}}$  d'après Verner et al. (2011). Pour les étoiles de la séquence principale, nous prenons  $\sigma_{\langle \Delta\nu \rangle} = 0.018 \langle \Delta\nu \rangle$  et  $\sigma_{\nu_{\text{max}}} = 0.038 \nu_{\text{max}}$  (Verner et al. 2011) et pour la branche des géantes rouges et des sous géantes, nous utilisons  $\sigma_{\langle \Delta\nu \rangle} = 0.4 \mu\text{Hz}$ ,  $\sigma_{\nu_{\text{max}}} = 2.5 \mu\text{Hz}$  si  $\nu_{\text{max}} \leq 80 \mu\text{Hz}$  et  $\sigma_{\nu_{\text{max}}} = 1.0 \mu\text{Hz}$  si  $\nu_{\text{max}} > 80 \mu\text{Hz}$  (Hekker et al. 2011).

Attention ici l'âge est déterminé pour toutes les étoiles du diagramme HR bien que les oscillations de type solaire ne soient observées que dans le bas de la séquence principale et dans la branche des géantes rouges (figure 6.2). Les âges ne sont bien sûr exploitable que dans ces régions.

Dans un premier temps, nous déterminons l'âge des étoiles du catalogue simulé avec les trois contraintes sismiques (figure 6.3.b). Nous trouvons 84 % des étoiles avec  $\Delta\tau/\tau < 20 \%$ . Les informations supplémentaires apportées par les paramètres sismiques permettent de réduire les dégénérescences à proximité de la séquence principale d'âge zéro et dans la branche des géantes rouges.

Nous regardons maintenant les résultats avec l'ajout d'une seule contrainte sismique (figures 6.3 et 6.4). La détermination des âges avec  $\langle \Delta\nu \rangle$  est moins bonne qu'avec  $\nu_{\text{max}}$  ou  $A_{\text{max}}$ . Les résultats en fonction des régions dans le diagramme HR (table 6.1) montrent que ce sont surtout les étoiles de la branche des géantes rouges qui ont des âges déterminés moins bien restitués avec  $\langle \Delta\nu \rangle$ . La densité moyenne dans cette région, proportionnelle à  $\Delta\nu$ , varie très peu.

C'est pourquoi, les âges sont mieux restitués avec le couple de contraintes  $\nu_{\text{max}} - A_{\text{max}}$  car  $\langle \Delta\nu \rangle$  (figure 6.4.d).

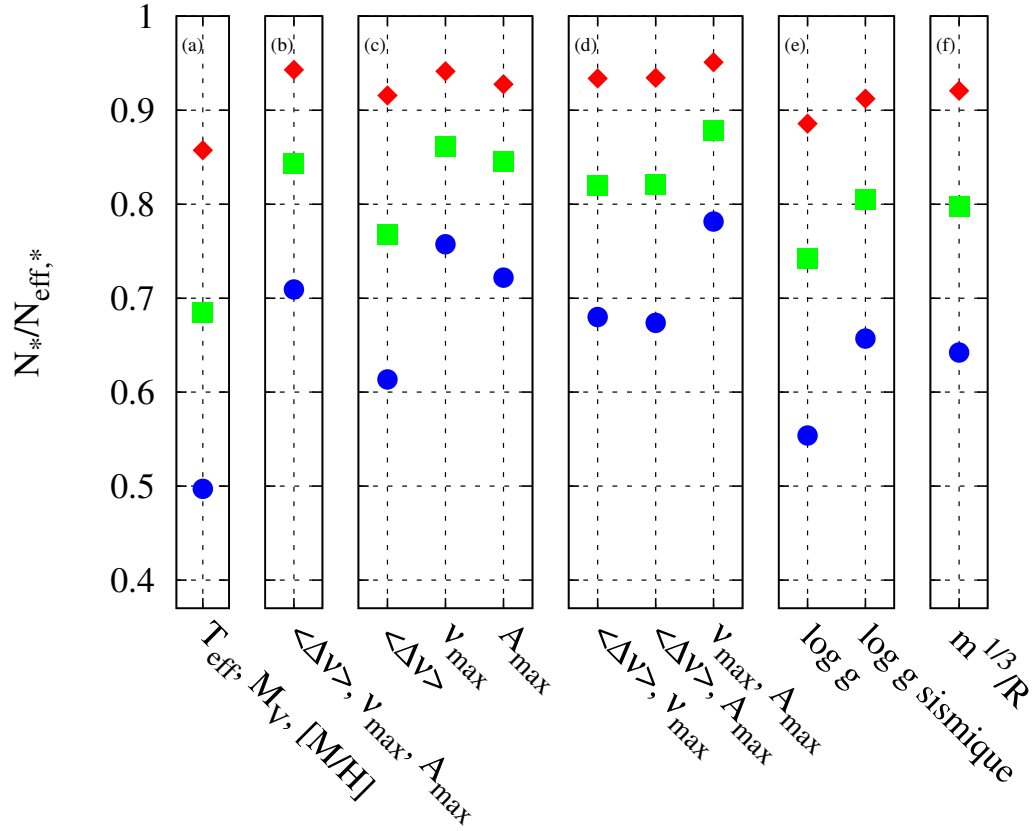


FIGURE 6.3 – Résultats de la comparaison des âges déterminés avec différentes contraintes. Les âges sont déterminés par l’inférence bayésienne décrite dans le chapitre 2, pour les étoiles du catalogue simulé de référence. Les couleurs sont les mêmes que celles de la figure 2.3.

TABLE 6.1 – Pourcentage d’étoiles avec  $\Delta\tau/\tau < 20\%$  selon la région du diagramme HR. Les âges sont déterminés avec différents paramètres astérosismiques. Les étoiles de la séquence principale de masse inférieure à  $2 M_{\odot}$  sont dans la région 1 et celles de masse supérieure à  $2 M_{\odot}$  sont dans la région 2. La région 3 comprend les étoiles de la branche des sous géantes et la région 4 les étoiles de la branche des géantes rouges.

Contrainte	Région 1	Région 2	Région 3	Région 4
$\langle \Delta \nu \rangle$	78	77	76	60
$\nu_{\max}$	88	78	75	75
$A_{\max}$	85	77	85	81



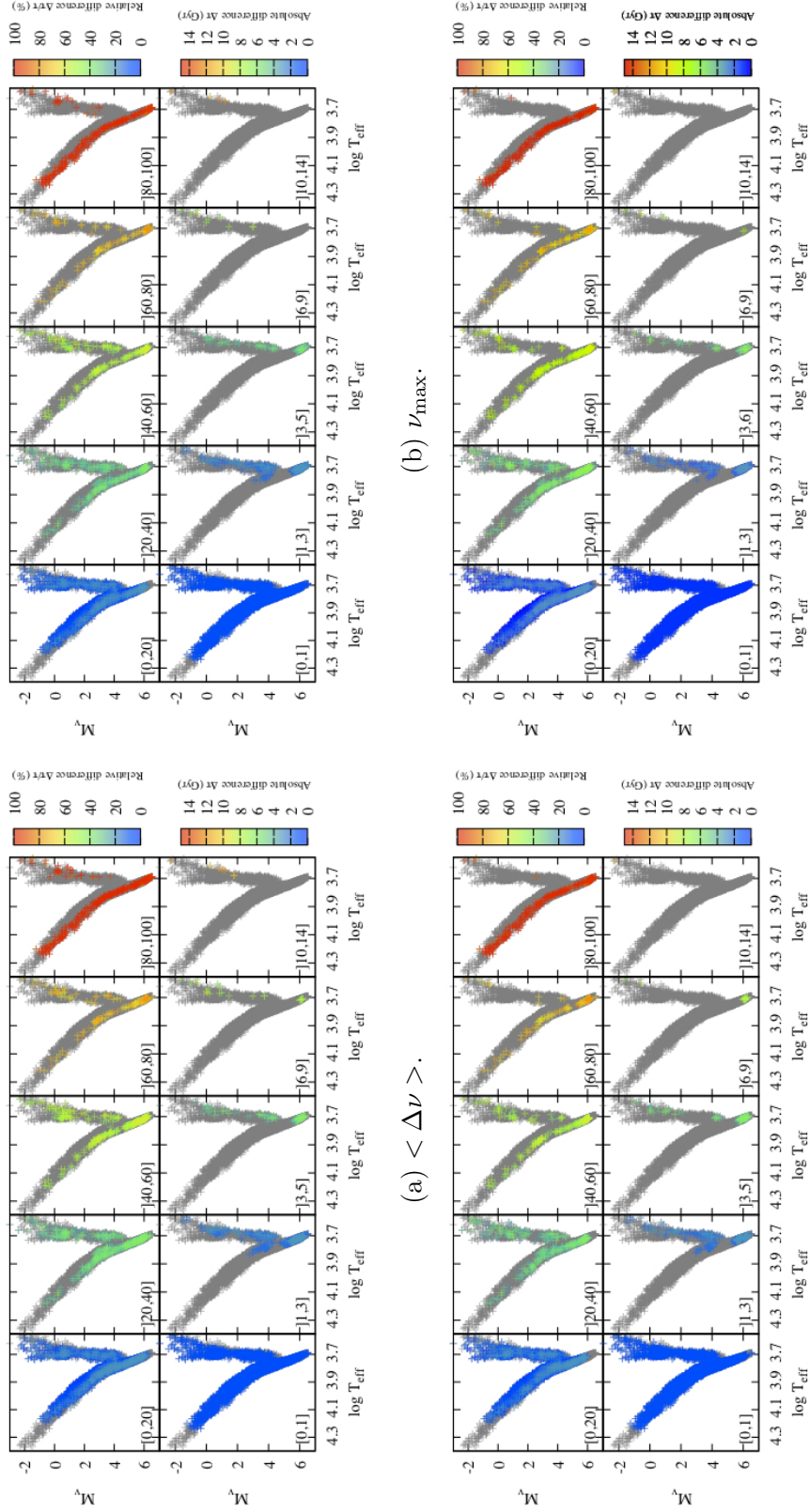


FIGURE 6.4 – Comparaison entre l'âge réel simulé et l'âge déterminé pour les étoiles du catalogue simulé avec l'ajout de paramètres sismiques. Les âges sont déterminés par l'inférence bayésienne. Les couleurs sont les mêmes que celles de la figure 5.2.

## 6.2 Transits d'exoplanètes

CoRoT et Kepler recherchent également les transits de planètes devant leur étoile hôte, comme représenté schématiquement sur la figure 6.5. L'exoplanète est détectée grâce à la légère diminution de la luminosité de l'étoile lors de son transit (Southworth 2008; Torres et al. 2008). La durée du transit  $d$  est de :

$$d = \frac{PR}{\pi a} \sqrt{\left(1 + \frac{r}{R}\right)^2 - \left(\frac{a}{R} \cos i\right)^2}, \quad (6.4)$$

$$= 1.8 \sqrt{(1 - b^2)} \frac{R}{m^{1/3}} P^{1/3}. \quad (6.5)$$

$P$  est la période orbitale de la planète en jours,  $R$  est le rayon de l'étoile et  $r$  celui de la planète,  $a$  est le rayon orbital de la planète en unités astronomiques,  $b$  est la distance projetée de la planète au centre de l'étoile et  $i$  est l'angle d'inclinaison de son orbite. On peut donc en déduire la contrainte  $m^{1/3}/R$  proportionnelle à la densité moyenne de l'étoile :

$$\frac{m^{1/3}}{R} = \frac{1.8 \sqrt{(1 - b^2)} P^{1/3}}{d} \propto \rho_{moy}^{1/3} \quad (6.6)$$

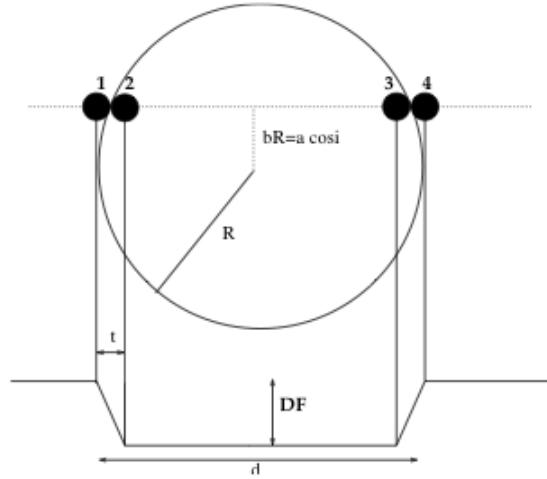


FIGURE 6.5 – Schéma d'un transit planétaire (Moutou & Pont 2006), avec  $DF$  qui correspond à la diminution de la luminosité de l'étoile lors du transit.  $R$  est le rayon de l'étoile,  $d$  est la durée du transit,  $t$  est le temps nécessaire pour que la planète aille du premier (troisième) au deuxième (quatrième) point de contact,  $a$  correspond au rayon orbital,  $bR$  est la distance projetée entre l'équateur de l'étoile et le rayon orbital et  $i$  est l'angle d'inclinaison de l'orbite de la planète.

A partir de l'équation 6.6, nous pouvons ajouter la contrainte  $m^{1/3}/R$  dans la détermination de l'âge et dans le catalogue simulé de référence. Nous prenons la moyenne des erreurs observationnelles trouvées dans la littérature pour la mission CoRoT :  $\sigma_{m^{1/3}/R} = 0.0224 m^{1/3}/R$ , afin de l'ajouter au catalogue simulé.

Les résultats de la datation des étoiles du catalogue simulé avec la contrainte  $m^{1/3}/R$  sont présentés sur les figures 6.3.c et 6.6. La détermination des âges est améliorée par ces informations supplémentaires et les dégénérescences sont réduites. La densité moyenne  $\rho_{\text{moy}}$  est proportionnelle au paramètre des transits d'exoplanètes  $m^{1/3}/R$  (équation 6.6) et la grande séparation  $\Delta\nu$  (équation 6.2). Nous obtenons des résultats similaires avec ces deux contraintes car les erreurs observationnelles sont du même ordre de grandeur.

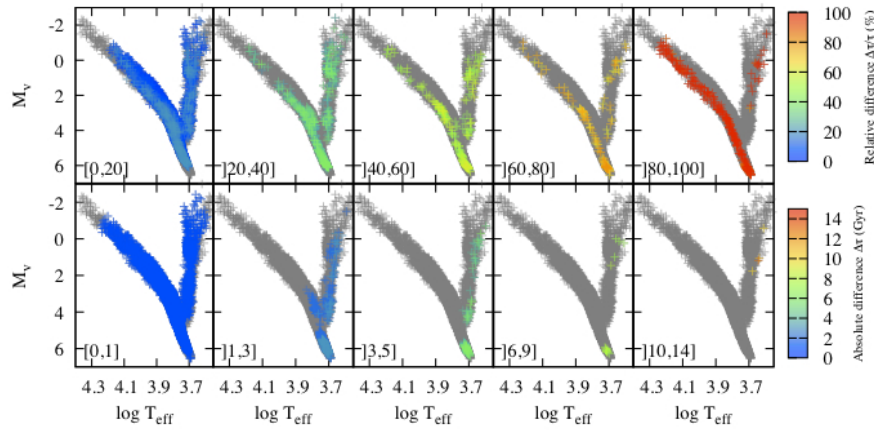


FIGURE 6.6 – Comparaison entre l'âge réel simulé et l'âge déterminé pour les étoiles du catalogue simulé de référence avec l'ajout de la contrainte  $m^{1/3}/R$ . Les couleurs sont les mêmes que celles de la figure 5.2.

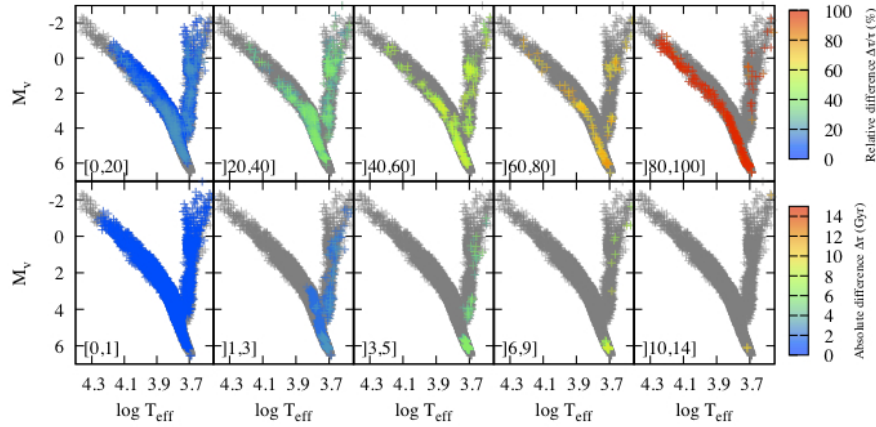
### 6.3 Gravité de surface

Les paramètres stellaires de l'étoile sont déterminés en comparant le spectre synthétique, issu de modèles d'atmosphère, qui ajustent le mieux le spectre de l'étoile (Cuisinier et al. 1994). Les paramètres déterminés sont : la vitesse radiale, la température effective, la gravité de surface, la métallicité et les abondances chimiques.

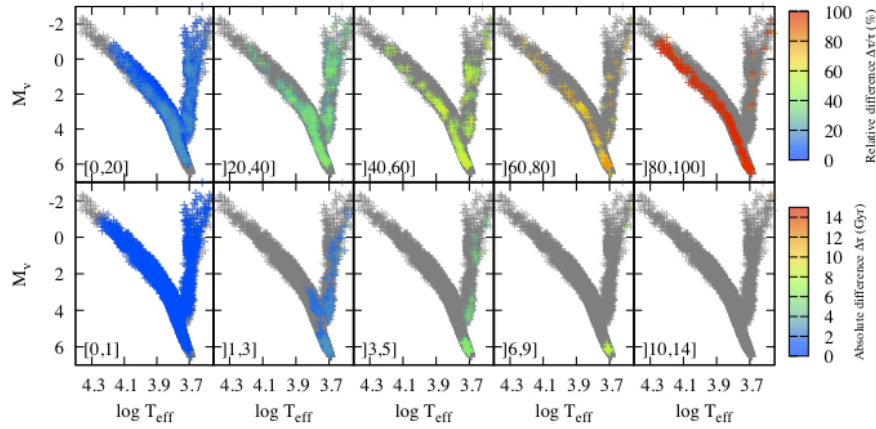
Nous ajoutons la gravité de surface  $\log g$  dans la détermination de l'âge des étoiles ; elle s'écrit  $g = GM/R^2$  où  $G = 6.671.10^{-11} m^3.kg^{-1}.s^{-1}$ . L'erreur observationnelle  $\sigma_{\log g} = 0.2$  dex est incluse dans le catalogue simulé, elle

provient des spécifications de Gaia<sup>1</sup>.

Les résultats ne montrent pas d'amélioration dans la détermination des âges avec la gravité de surface (figures 6.3 et 6.7), soit parce que cette contrainte n'apporte pas d'informations supplémentaires ou parce que l'erreur observationnelle est trop grande.



(a)  $\log g$  spectroscopique avec  $\sigma_{\log g} = 0.2$  dex.



(b)  $\log g$  sismique avec  $\sigma_{\log g} = 0.05$  dex.

FIGURE 6.7 – Comparaison entre l'âge réel simulé et l'âge déterminé pour les étoiles du catalogue simulé avec l'ajout de la contrainte  $\log g$  pour deux erreurs observationnelles. Les couleurs sont les mêmes que celles de la figure 5.2.

1. [http://www.rssd.esa.int/index.php?project=GAIA&page=Science\\_Performance](http://www.rssd.esa.int/index.php?project=GAIA&page=Science_Performance).

Pour vérifier, nous testons la détermination des âges avec une gravité de surface issue de la sismologie (Creevey et al. 2012) avec :

$$\frac{g}{g_{\odot}} = \left( \frac{\nu_{\max}}{\nu_{\max,\odot}} \right) \left( \frac{T_{\text{eff}}}{T_{\text{eff},\odot}} \right)^{1/2}. \quad (6.7)$$

L'erreur observationnelle sur la gravité de surface est de 0.05 dex.

Les âges déterminés avec la gravité de surface sismique (figures 6.3 et 6.7) sont mieux restitués sur la branche des géantes rouges qu'avec une gravité de surface spectroscopique. En revanche, il n'y a pas d'amélioration sur la séquence principale.

Sur la figure 6.8, nous avons représenté la relation entre l'âge et la gravité de surface, pour les trajets évolutifs de BaSTI de métallicité solaire. Sur la séquence principale, la variation relative entre l'âge et la gravité de surface est en loi de puissance et en moyenne de  $d\tau/\tau = -0.8 d\log g/\log g$ . Les variations de la gravité de surface n'apportent pas d'information supplémentaire, c'est pourquoi les résultats ne sont pas améliorés pour les étoiles dans cette région. Par contre dans la branche des géantes rouges, où les âges sont mieux restitués, la variation relative est beaucoup plus grande et est en moyenne de  $d\tau/\tau = 25 d\log g/\log g$ .

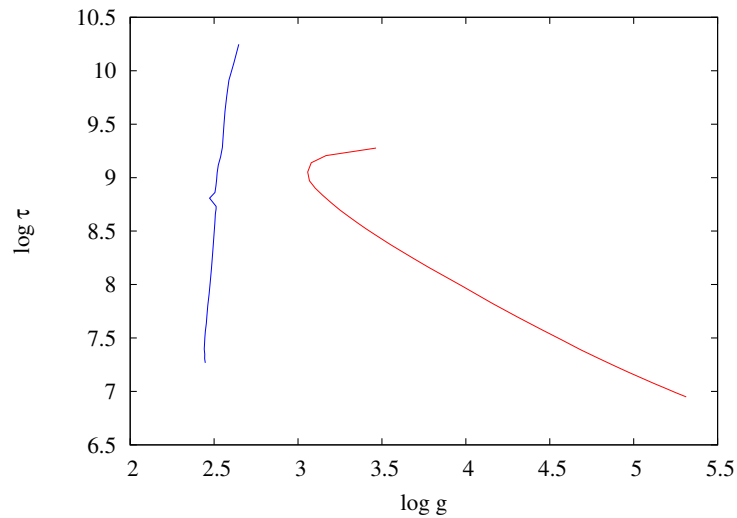


FIGURE 6.8 – Relation entre  $\log \tau$  et  $\log g$  pour les trajets évolutifs de BaSTI compris entre  $0.75$  et  $10 M_{\odot}$ , pour une métallicité solaire. La courbe rouge correspond à la séquence principale et la courbe bleue à la branche des géantes rouges.

## 6.4 Conclusions et perspectives

L'ajout de contraintes dans la détermination des âges permet de réduire les dégénérescences dans la branche des géantes et à proximité de la séquence principale d'âge zéro. Une contrainte est dite "bonne" si la restitution des âges est améliorée. Pour cela, les variations de la contrainte en fonction de l'âge doivent être grandes. En effet, la prise en compte de la gravité de surface, qui varie peu en fonction de l'âge sur la séquence principale, n'améliore pas la datation. Au contraire, la datation est améliorée dans la branche des géantes rouges par la gravité de surface car celle-ci varie significativement en fonction de l'âge.

Pour avoir une amélioration des résultats il faut également que l'erreur observationnelle sur les contraintes soit petite. En effet, la gravité de surface peut être déterminée par la spectroscopie ou les paramètres sismiques, ces derniers diminuent d'un facteur 5 l'erreur observationnelle. La gravité de surface est également proportionnelle à la fréquence du maximum de puissance  $\nu_{\max}$ . Nous constatons une amélioration de la détermination des âges lorsque l'erreur observationnelle est petite. Pour déterminer les âges avec la gravité de surface, il faut donc favoriser la fréquence du maximum de puissance  $\nu_{\max}$  qui a une erreur observationnelle plus petite.

Pour préciser les résultats sur l'effet de l'erreur observationnelle, il faudrait effectuer une étude plus approfondie en testant plusieurs erreurs observationnelles sur chaque contrainte. Il serait également intéressant d'analyser l'effet d'autres paramètres stellaires comme la masse issue de l'observation des binaires ou le rayon spectroscopique.

Il faudrait également intégrer l'enrichissement en éléments  $\alpha$  (Takeda et al. 2007). Pour les étoiles qui ont un enrichissement en éléments  $\alpha$  différent de zéro, les trajets évolutifs ne correspondent donc pas à la composition de l'étoile. Pour prendre en compte l'enrichissement en éléments  $\alpha$  il faut donc utiliser une grille de trajets évolutifs avec plusieurs enrichissements en éléments  $\alpha$ . Cela nécessite une grille beaucoup plus fine et donc un temps de calcul et une taille mémoire plus importante pour la détermination des âges.

---

1. Les éléments  $\alpha$  (oxygène, magnésium, silicium, calcium ...) sont formés par captures de particule  $\alpha$  (noyau d'hélium) lors de la fusion du silicium des étoiles massives avant l'explosion en supernova de type SNI.

Troisième partie

Applications





# Estimation des résultats obtenus par Gaia

---

## Sommaire

<b>7.1</b>	<b>Catalogues simulés . . . . .</b>	<b>99</b>
<b>7.2</b>	<b>Ages et masses . . . . .</b>	<b>100</b>
<b>7.3</b>	<b>Conclusions et perspectives . . . . .</b>	<b>106</b>

---

## 7.1 Catalogues simulés

La mission spatiale Gaia va observer des étoiles jusqu'à une magnitude apparente  $G$  de 20 mag. Dans la partie 4, nous avons indiqué que les erreurs observationnelles sur la magnitude absolue  $M_V$  et sur la température effective  $\log T_{\text{eff}}$ , obtenues par Gaia, dépendent de la magnitude apparente  $G$  et de la distance  $d$ . L'erreur observationnelle sur la magnitude absolue dépend également de l'absorption  $A_V$ . La magnitude apparente d'une étoile dépend du milieu que traverse la lumière de l'étoile avant de nous arriver. L'absorption correspond donc à la diminution de la magnitude due aux poussières ou aux nuages interstellaires. L'absorption dépend donc de la direction d'observation dans le ciel et de la distance de l'étoile.

Afin d'étudier les effets de la distance et de la direction d'observation sur les âges, nous avons construit plusieurs catalogues simulés à différentes distances (100 pc, 500 pc, 1 000 pc, 2 500 pc et 5 000 pc) et cela dans trois directions d'observation : le centre galactique (longitude  $0^\circ$ , latitude  $0^\circ$ ), le plan galactique (longitude  $50^\circ$ , latitude  $0^\circ$ ) et au dessus du plan galactique (longitude  $0^\circ$ , latitude  $15^\circ$ ).

Les absorptions  $A_V$  ont été calculées d'après les travaux de Drimmel et al. (2003)<sup>1</sup>; elles dépendent de la distance de l'étoile au Soleil et de sa position dans le ciel. Les absorptions  $A_V$  que l'on utilise sont décrites dans la table 7.1.

---

1. Programme disponible sur <ftp://ftp.to.astro.it/astrometria/extinction/>

TABLE 7.1 – Absorption  $A_V$  (mag) d’après le modèle de Drimmel et al. (2003) en fonction de la distance  $d$  et de la direction d’observation, où  $l$  représente la longitude et  $b$  la latitude galactique.

$d$ (pc)	$l = 0^\circ, b = 0^\circ$	$l = 50^\circ, b = 0^\circ$	$l = 0^\circ, b = 15^\circ$
100	0.058	0.058	0.05
500	0.32	0.31	0.26
1 000	0.70	0.66	0.39
2 500	2.90	1.98	0.43
5 000	11.35	4.99	0.43

Pour chacune des différentes distances nous avons construit plusieurs catalogues de 10 000 étoiles. Ce nombre ne représente pas le nombre d’étoiles réellement présentes aux distances considérées. Selon Robin et al. (2012) il y a  $10^6$  étoiles à 1 000 pc,  $10^7$  étoiles à 2 500 pc et  $2 \cdot 10^7$  à 5 000 pc. Nous avons choisi de garder 10 000 étoiles afin d’avoir une statistique correcte sur les âges obtenus.

Les catalogues simulés obtenus sont représentés sur les figures 7.1, 7.2 et 7.3. Les catalogues à 100 pc contiennent peu d’étoiles de masse supérieure à  $2.7 M_\odot$  sur la séquence principale et aucune étoile dans la branche des géantes rouges. Les étoiles massives et les géantes rouges sont trop brillantes pour être observées par Gaia qui est limité à des magnitudes apparentes  $G$  supérieures à 6 mag. Les catalogues à 2 500 et 5 000 pc n’ont pas d’étoiles de faibles masses dans le bas du diagramme HR. Ces étoiles ne sont pas assez brillantes pour être observées par Gaia ou elles ont une erreur relative sur la parallaxe  $\sigma_\pi/\pi$  qui est supérieure à 10 % (partie 4). Il n’y a pas de catalogue simulé à une distance de 5 000 pc en direction du centre galactique et dans le plan galactique car l’absorption y est trop grande.

Dans la partie 5.3.2 nous avons montré que l’erreur observationnelle sur la métallicité modifie les âges déterminés. Afin de prendre en compte cette erreur, nous construisons les catalogues pour trois erreurs observationnelles sur la métallicité : 0.1, 0.2 et 0.3 dex.

## 7.2 Ages et masses

Les résultats de la détermination des âges de tous les catalogues simulés sont présentés sur les figures 7.4, 7.5 et 7.6. Les âges sont mieux restitués avec une erreur observationnelle de 0.1 dex sur la métallicité quelle que soit la direction d’observation et la distance des étoiles.

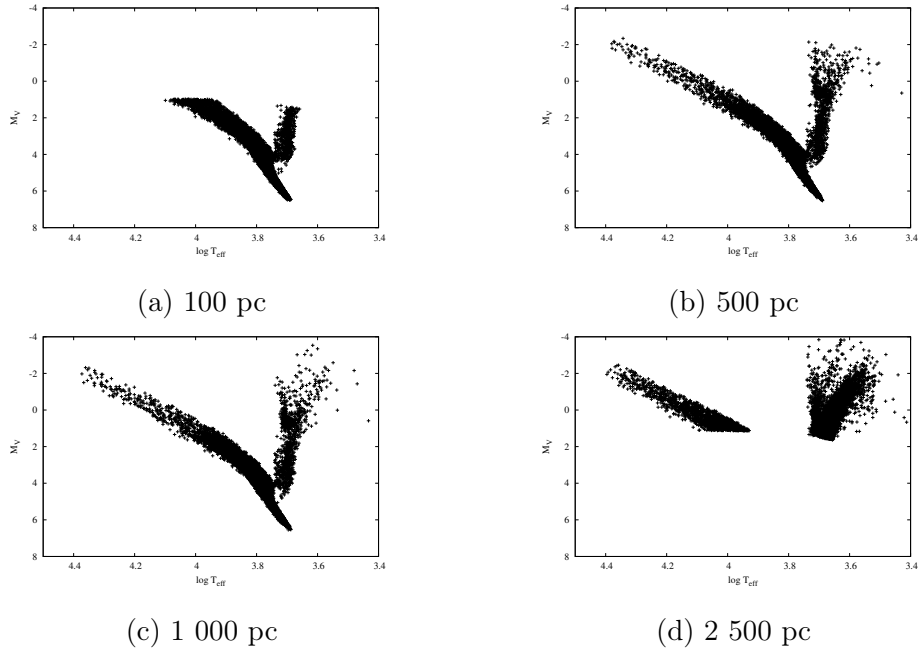


FIGURE 7.1 – Diagrammes HR simulés à différentes distances en direction du centre galactique (longitude  $0^\circ$ , latitude  $0^\circ$ ).

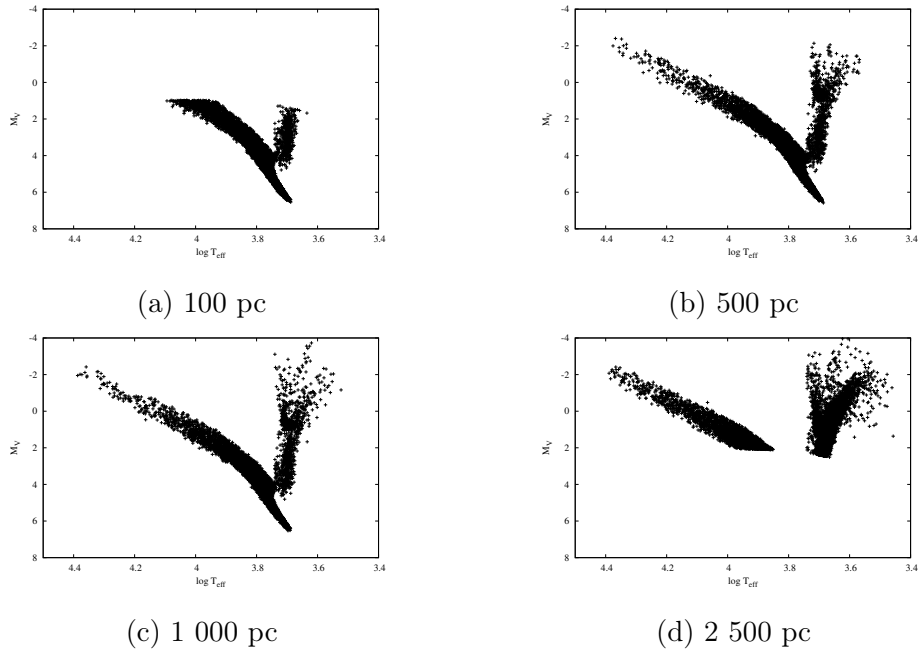


FIGURE 7.2 – Diagrammes HR simulés à différentes distances pour une direction d'observation dans le plan galactique (longitude  $50^\circ$ , latitude  $0^\circ$ ).

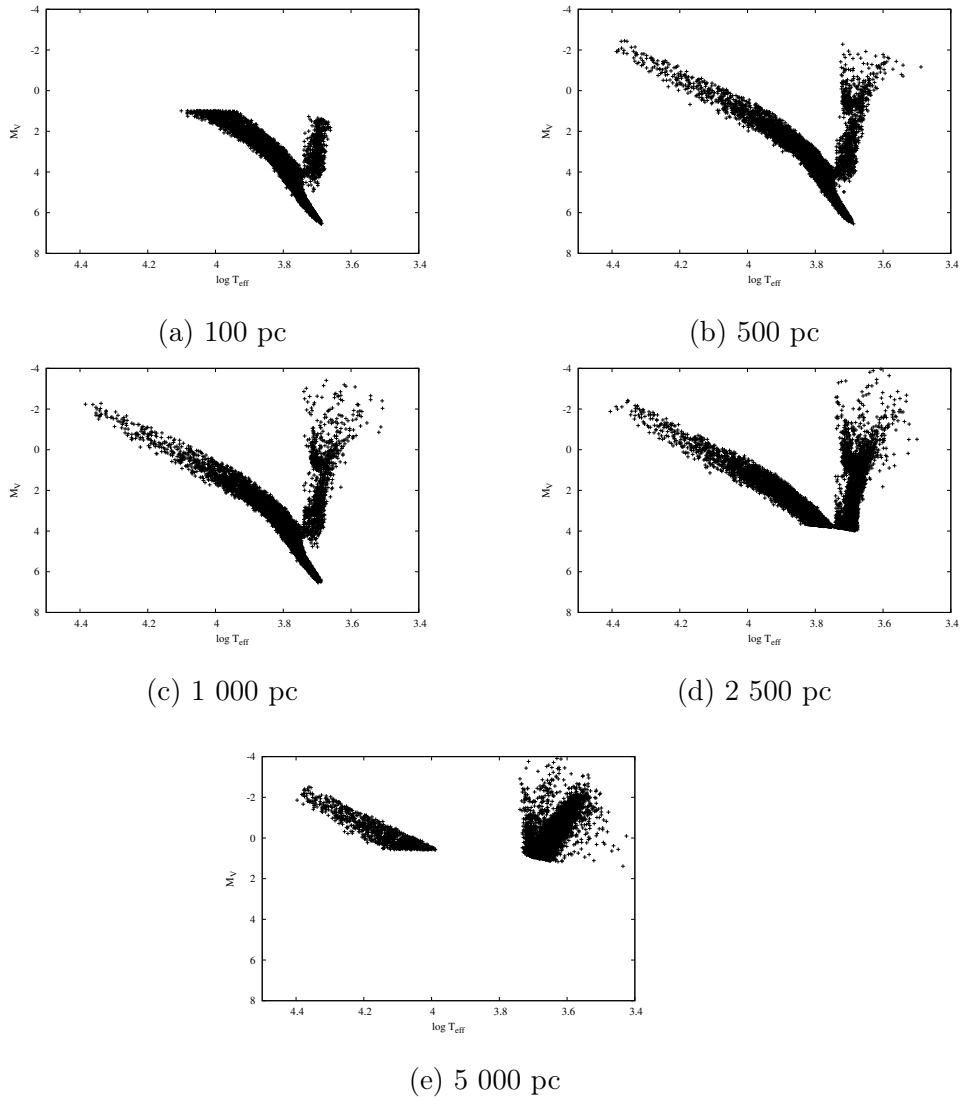


FIGURE 7.3 – Diagrammes HR simulés à différentes distances pour une direction d’observation au dessus du centre galactique (longitude  $0^\circ$ , latitude  $15^\circ$ ).

Le pourcentage des âges bien déterminés pour les étoiles des catalogues à 100 pc, 500 pc et 1 000 pc est similaire quelle que soit la direction d’observation car l’absorption y est équivalente dans les trois directions d’observation et pour les trois distances (table 7.1). L’absorption varie peu à ces distances et a donc peu d’effet sur la détermination des âges. A 100 pc, les âges sont légèrement mieux restitués car l’absorption est très faible.

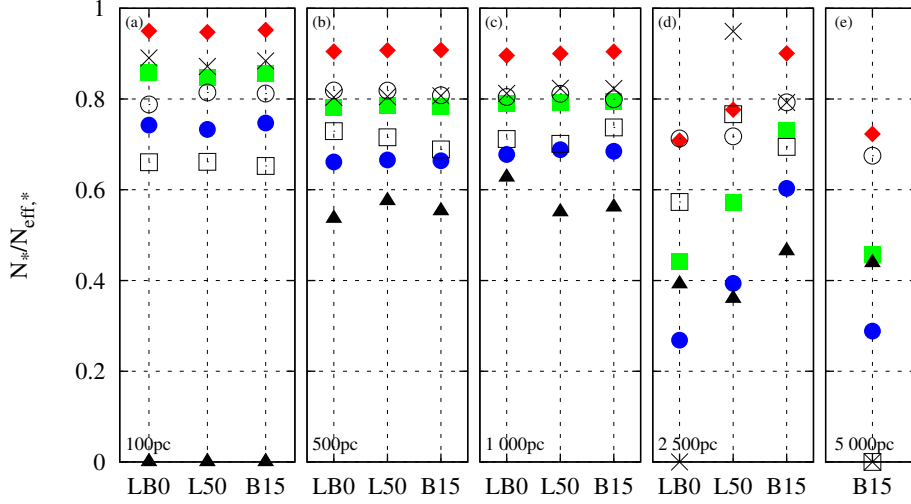


FIGURE 7.4 – Résultats de la datation des étoiles des catalogues simulés avec  $\sigma_{[M/H]} = 0.1$  dex et à différentes distances : 100 pc, 500 pc, 1 kpc, 2.5 kpc et 5 kpc. Les catalogues sont construits selon trois directions d’observation : le centre galactique (LB0), le plan galactique (L50) et au dessus du plan galactique (B15).  $N_*$  est le nombre d’étoiles avec  $\Delta\tau/\tau < 10\%$  (cercle bleu),  $< 20\%$  (carré vert) et  $< 50\%$  (losange rouge). Les résultats sont également donnés pour  $\Delta\tau/\tau < 20\%$  en fonction de la région du diagramme HR : le rond noir pour la séquence principale et  $m < 2.0 M_\odot$ , la croix noire pour la séquence principale et  $m > 2.0 M_\odot$ , le cercle noir pour la branche des sous géantes et le triangle noir pour la branche des géantes rouges.

A 2 500 pc, l’absorption varie beaucoup selon les directions d’observations : en direction du centre galactique  $A_V$  est de 2.9 mag, dans le plan galactique  $A_V$  est de 1.98 mag et  $A_V$  est de 0.43 mag au dessus du plan galactique. Plus l’absorption est importante, plus la magnitude apparente  $G$  est grande et donc plus l’erreur observationnelle sur la magnitude absolue augmente. Les âges sont donc moins bien déterminés.

Au dessus du plan galactique avec  $\sigma_{[M/H]} = 0.2$  et 0.3 dex, les âges semblent mieux restitués à 2 500 pc qu’à 1 000 pc. A 2 500 pc les étoiles de petites masses sur la séquence principale ne sont pas dans le diagramme HR. Ces étoiles sont situées dans une zone de dégénérescence où les âges sont moins bien estimés. C’est pour cela que les résultats semblent meilleurs à 2 500 pc.

A 5 000 pc seule la région au dessus du plan galactique peut donner lieu à

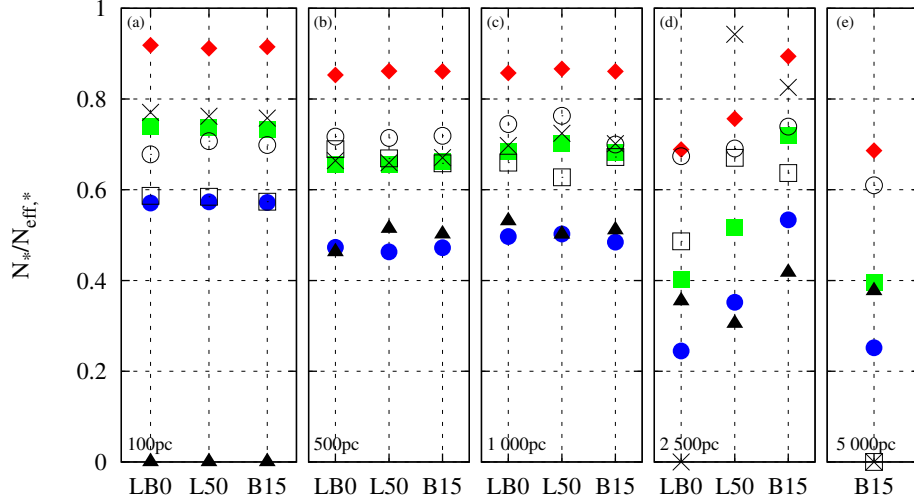


FIGURE 7.5 – Résultats de la datation des étoiles des catalogues simulés avec  $\sigma_{[M/H]} = 0.2$  dex, à différentes distances et pour différentes directions d’observation sur le ciel. Les couleurs et les directions d’observation sont les mêmes que celles de la figure 7.4.

un catalogue simulé. Les âges  $y$  sont moins bien évalués à cette distance car il n’y a que des étoiles massives et des géantes rouges. Il y a peu de différences entre les trois erreurs observationnelles sur la métallicité : ce sont les erreurs observationnelles sur la magnitude absolue et sur la température effective qui dominent.

Sur les figures 7.4, 7.5 et 7.6, nous avons également représenté les résultats en fonction de quatre régions du diagramme HR. A 100 pc il n’y a pas d’étoiles dans la branche des géantes rouges et très peu sur la séquence principale avec  $m > 2 M_{\odot}$ . Les âges sont moins bien restitués dans la branche des sous géantes.

A 500 pc et à 1 000 pc les résultats sont similaires selon les directions d’observation. A 2 500 pc dans le plan galactique, les âges sont bien déterminés pour les petites masses de la séquence principale car il y en a très peu dans l’échantillon d’étoiles.

Nous avons réalisé la même étude pour la détermination de la masse mais seulement avec  $\sigma_{[M/H]} = 0.2$  dex (figure 7.7). Les masses sont très bien déterminées jusqu’à 1 000 pc quelle que soit la direction d’observation. La masse des géantes rouges est moins bien estimée par rapport à celles des autres étoiles. A 2 500 et 5 000 pc, l’absorption devient plus grande et la masse est moins bien restituée.

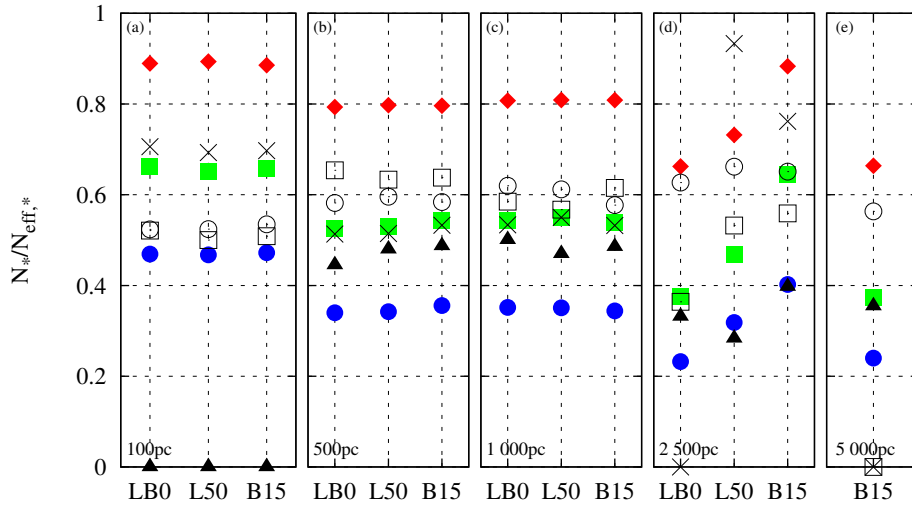


FIGURE 7.6 – Résultats de la datation des étoiles des catalogues simulés avec  $\sigma_{[M/H]} = 0.3$  dex, à différentes distances et pour différentes directions d’observation sur le ciel. Les couleurs et les directions d’observation sont les mêmes que celles de la figure 7.4.

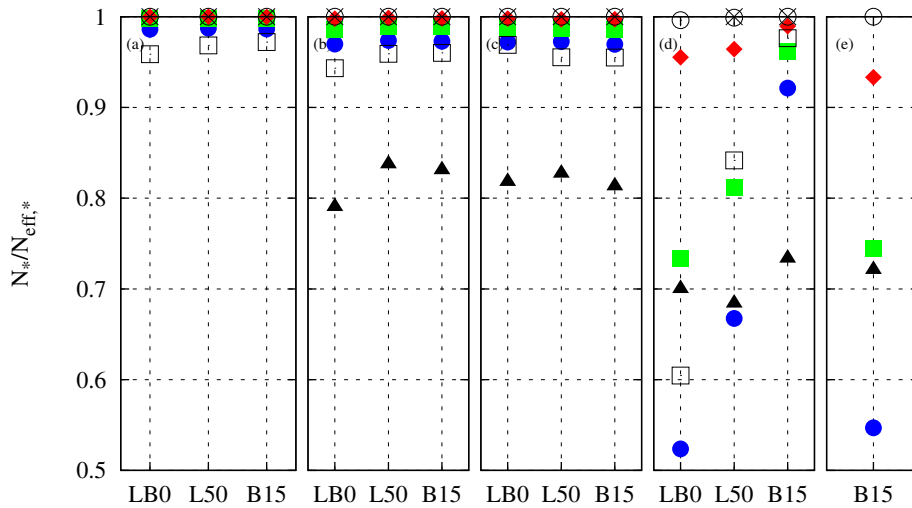


FIGURE 7.7 – Résultats de la détermination de la masse des étoiles des catalogues simulés avec  $\sigma_{[M/H]} = 0.2$  dex, à différentes distances et pour différentes directions d’observation sur le ciel. Les couleurs et les directions d’observation sont les mêmes que celles de la figure 7.4.

### 7.3 Conclusions et perspectives

Plus l'étoile est éloignée du Soleil, plus l'absorption augmente et plus les erreurs observationnelles sur la magnitude absolue et sur la température augmentent. Les âges des étoiles sont bien restitués jusqu'à une distance de 1 000 pc pour les étoiles de la séquence principale et de la branche des sous géantes. En revanche à partir de 2 500 pc, quand l'absorption est trop grande il devient difficile de dater correctement les étoiles. Il faut donc à ces distances privilégier des observations précises de la métallicité afin d'avoir une faible erreur observationnelle.

Les masses sont bien évaluées jusqu'à 1 000 pc, ensuite l'absorption affecte la détermination de la masse.

A de grandes distances seulement les étoiles massives et les géantes rouges sont observées avec une bonne précision sur la parallaxe. Comme ces étoiles sont difficiles à dater, la restitution des âges ne sera correcte que pour 40 % d'entre elles.

L'absorption a également un fort impact sur la détermination des âges : quand elle devient importante les âges sont moins bien restitués. Pour la détermination des âges il est important d'avoir une connaissance précise des absorptions quelles que soient la distance et la direction d'observation.

Pour Gaia, il est donc important de réaliser cette étude pour d'autres directions d'observation et pour d'autres de distances afin de pouvoir estimer l'exactitude des âges et des masses déterminés avec l'inférence bayésienne.



# Applications à des catalogues d'observation

## Sommaire

<b>8.1</b>	<b>Catalogue d'observation</b>	<b>107</b>
8.1.1	Geneva Copenhagen Survey	107
8.1.2	Catalogue de Setiawan et al.	114
<b>8.2</b>	<b>Comparaison avec les travaux antérieurs</b>	<b>115</b>
<b>8.3</b>	<b>[M/H] vs. [Fe/H]</b>	<b>117</b>
<b>8.4</b>	<b>Datation des amas</b>	<b>118</b>
<b>8.5</b>	<b>Conclusions et perspectives</b>	<b>124</b>

## 8.1 Catalogue d'observation

### 8.1.1 Geneva Copenhagen Survey

Le Geneva Copenhagen Survey (GCS) est construit à partir d'un relevé photométrique d'Olsen (1983, 1993, 1994a,b) dans le système de Strömgen  $wby\beta$  (Strömgen 1965). Ce relevé contient 30 465 étoiles du voisinage solaire qui sont à une distance inférieure à 40 pc. Les étoiles de type A5 à G0 ont une magnitude apparente  $V$  inférieure à 8.3, les étoiles G0 ont  $8.3 \leq V \leq 8.4$  et les étoiles G ont  $V < 8.5$ . Nordström et al. (2004, NMA04) effectuent des sélections afin de garder seulement les étoiles qui respectent toutes les conditions suivantes :

1. Les étoiles respectant les conditions des calibrations de Crawford (1975) et d'Olsen (1988). Ces calibrations permettent de déterminer la couleur intrinsèque  $(b - y)_0$  à partir de la magnitude  $\beta$ . La calibration de Crawford (1975) s'écrit :

$$\begin{aligned}
 (b - y)_0 &= 0.222 + 1.11\Delta\beta + 2.7(\Delta\beta)^2 \\
 &- 0.05\delta c_1 - (0.1 + 3.6\Delta\beta)\delta m_1,
 \end{aligned} \tag{8.1}$$

où  $\Delta\beta = 2.720 - \beta$ ,  $\delta m_1 = m_1(\text{standard}) - m_1(\text{observé})$  avec  $m_1(\text{observé}) = (v - b) - (b - y)$  et  $\delta c_1 = c_1(\text{standard}) - c_1(\text{observé})$  avec  $c_1(\text{observé}) = (u - v) - (v - b)$ .  $m_1(\text{standard})$  et  $c_1(\text{standard})$  sont déterminés à partir des tables de Crawford (1975) pour un  $\beta$  donné. Cette calibration est appliquée pour les étoiles avec  $2.720 < \beta < 2.590$  et pour  $\delta c_1 < 0.280$  si  $\beta > 2.630$  ou  $\delta c_1 < 0.200$  si  $\beta < 2.630$ .

La calibration d'Olsen (1988) s'écrit :

$$(b - y)_0 = 0.217 + 1.34\Delta\beta + 1.6(\Delta\beta)^2 + C\delta c_1 - D, \quad (8.2)$$

où  $C = 4.9\Delta\beta + 32.2\delta m_0 - 262(\delta m_0)^2 - 1.31$  et  $D = (0.16 + 4.5\delta m_0 + 3.5\Delta\beta)\delta m_0$ . Si  $\delta m_0 < 0.060$ , si  $\delta m_0 \geq 0.060$  on a  $D = 0.24\delta m_0 + 0.035$ . Elle est utilisée pour  $1.6\Delta\beta \geq C > 0.13$  et  $\delta m_1 \leq 0.080$ .

2. Les étoiles sans valeurs de  $\beta$  et avec  $0.240 \leq (b - y) \leq 0.460$ ,  $m_1 \geq 0.120$ ,  $\delta c_1 \leq 0.400$  et  $V \leq 9.600$ .
3. Les étoiles avec  $0.205 \leq (b - y) \leq 0.240$ ,  $m_1 \geq 0.120$ ,  $\delta c_1 \leq 0.400$  et  $V \leq 9.600$ .
4. Les étoiles de type GK respectant les conditions de la calibration d'Olsen (1984) sur la magnitude absolue  $M_V$  :

$$\begin{aligned} M_V &= M_V(ZAMS, (b - y)) - f\delta c_1 \times (b - y) \\ &+ 3.2\delta m_1 \times (b - y) - 0.07, \end{aligned} \quad (8.3)$$

avec  $f = 10 - 80[(b - y) - 0.380]$  et  $M_V(ZAMS, (b - y))$  qui est une correction sur la magnitude absolue  $M_V$  pour un  $(b - y)$  donné.

Le GCS contient 16 682 étoiles après ces sélections. Les vitesses radiales et celles de rotation, les parallaxes et les mouvements propres sont observés pour chaque étoile. A partir des ces observations, plusieurs études ont permis de déterminer les paramètres stellaires, c'est pourquoi il y a plusieurs versions de ce catalogue qui dépendent des méthodes utilisées. Nous décrivons maintenant quatre de ces versions.

### 8.1.1.1 GCS I

Le GCS I (NMA04) est la première version du catalogue. Le rougissement interstellaire est déterminé avec  $E(b - y) = (b - y) - (b - y)_0$  où  $(b - y)$  est obtenu par les observations et  $(b - y)_0$  est déterminé avec la calibration d'Olsen (1988). La magnitude absolue et la parallaxe proviennent des observations d'Hipparcos (ESA 1997). La température effective  $T_{\text{eff}}$  est déterminée avec la calibration de Alonso et al. (1996) :

$$\begin{aligned} \theta_{\text{eff}} &= 0.537 + 0.854(b - y) + 0.196(b - y)^2 - 0.198(b - y)c_1 \\ &- 0.026(b - y)[\text{Fe}/\text{H}] - 0.014[\text{Fe}/\text{H}] - 0.009[\text{Fe}/\text{H}]^2, \end{aligned} \quad (8.4)$$

où  $T_{\text{eff}} = 5040/\theta_{\text{eff}}$ . Cette calibration est utilisée pour  $-3 \leq [\text{Fe}/\text{H}] \leq 0.5$ .

NMA04 utilisent plusieurs calibrations pour déterminer la métallicité  $[\text{Fe}/\text{H}]$ . Pour la majorité des étoiles de type FG, NMA04 prennent les calibrations de Schuster & Nissen (1989) quand elles sont applicables. Pour celles de type F avec  $0.22 \leq (b - y) \leq 0.38$ ,  $0.17 \leq c_1 \leq 0.58$ ,  $0.03 \leq m_1 \leq 0.21$  et avec  $-3.5 \leq [\text{Fe}/\text{H}] \leq 0.2$ , la calibration s'écrit :

$$\begin{aligned} [\text{Fe}/\text{H}] &= 1.052 - 73.21m_1 + 280.9m_1(b - y) - 333.95m_1^2(b - y) & (8.5) \\ &- 595.5m_1(b - y)^2 + [5.486 - 41.62m_1 - 7.963(b - y)] \log(m_1 - c_3), \\ \text{avec } c_3 &= 0.6322 - 3.58(b - y) + 5.20(b - y)^2. \end{aligned}$$

Pour les étoiles de type G avec  $0.37 \leq (b - y) \leq 0.59$ ,  $0.10 \leq c_1 \leq 0.47$ ,  $0.03 \leq m_1 \leq 0.57$  et avec  $-2.6 \leq [\text{Fe}/\text{H}] \leq 0.4$ , la calibration de Schuster & Nissen (1989) donne :

$$\begin{aligned} [\text{Fe}/\text{H}] &= -0.0965 - 22.45m_1 - 53.8m_1^2 - 62.04m_1(b - y) & (8.6) \\ &+ 145.5m_1^2(b - y) + [85.1m_1 - 13.8c_1 - 137.2m_1^2]c_1. \end{aligned}$$

Pour environ 2 100 étoiles de type GK qui ont un  $(b - y) \geq 0.44$ , NMA04 calibrent une relation basée sur les termes de Schuster & Nissen (1989) ainsi que des études de spectroscopie haute résolution (Flynn & Morell 1997; Tomkin & Lambert 1999; Thorén & Feltzing 2000; Santos et al. 2001) :

$$\begin{aligned} [\text{Fe}/\text{H}] &= -2.06 + 24.56m_1 - 31.61m_1^2 - 53.64m_1(b - y) \\ &+ 73.50m_1^2(b - y) + [26.34m_1 - 0.46c_1 - 17.76m_1^2]c_1. & (8.7) \end{aligned}$$

Il y a également 2 400 étoiles avec des grandes températures et des faibles densités pour lesquelles les calibrations ne sont pas applicables. Dans ce cas NMA04 utilisent la calibration de Edvardsson et al. (1993) :

$$[\text{Fe}/\text{H}] = \frac{-[10.5 + 50(\beta - 2.626)]\delta m_1 + 0.12 - 0.030}{1.049}, \quad (8.8)$$

pour  $2.58 < \beta < 2.72$ ,  $0.04 < (\delta c_1/0.5\delta m_1) < 0.16$  et  $-1.0 < [\text{M}/\text{H}] < 0.3$ .

NMA04 construisent une calibration pour les 342 étoiles restantes qui ont  $0.18 \leq (b - y) \leq 0.38$ ,  $0.07 \leq m_1 \leq 0.26$ ,  $0.21 \leq c_1 \leq 0.86$  et  $-1.5 \leq [\text{Fe}/\text{H}] \leq 0.8$ . La calibration est déterminée à partir d'étoiles spectroscopiques (Burkhardt & Coupry 1991; Glaspey et al. 1994; Taylor 2003) :

$$\begin{aligned} [\text{Fe}/\text{H}] &= 9.60 - 61.16m_1 + 81.25m_1(b - y) + 224.65m_1^2(b - y) & (8.9) \\ &- 153.18m_1(b - y)^2 + [12.23 - 90.23m_1 + 30.70(b - y)] \log(m_1 - c_3), \\ \text{avec } c_3 &= 0.45 - 3.98(b - y) + 5.08(b - y)^2. \end{aligned}$$

NMA04 déterminent les âges avec la méthode d'inférence bayésienne de Jørgensen & Lindegren (2005) qui est décrite dans la partie 3.1.1. Pour cela ils utilisent les isochrones de Padoue (Girardi et al. 2000) et de Geneve (Mowlavi et al. 1998). Avant de déterminer les âges, ils corrigent la métallicité observée de l'enrichissement en éléments  $\alpha$ . L'enrichissement en éléments  $\alpha$  modifie la composition chimique de l'étoile, l'opacité, les réactions nucléaires et l'équation d'états des modèles stellaires. Selon Salaris et al. (1993) l'enrichissement en éléments  $\alpha$  dans l'étoile équivaut à modifier sa métallicité  $[M/H]$ . NMA04 utilisent la calibration de Vandenberg et al. (2000) qui permet de modifier la métallicité de l'étoile en fonction de l'enrichissement en éléments  $\alpha$ . Cela permet de prendre les trajets évolutifs qui ont une métallicité qui correspond à l'étoile :

$$\begin{aligned}
 [\text{Fe}/\text{H}]_{\text{cor}} &= [\text{Fe}/\text{H}] + 0.75 & (8.10) \\
 &\times \begin{cases} 0.0 & \text{si } 0. \leq [\text{Fe}/\text{H}], \\ -0.25 \times [\text{Fe}/\text{H}] & \text{si } -1.6 \leq [\text{Fe}/\text{H}] < 0., \\ 0.4 & \text{si } -1.6 \geq [\text{Fe}/\text{H}]. \end{cases}
 \end{aligned}$$

Ils corrigent également la température effective des modèles stellaires pour diminuer l'écart entre les modèles et les observations et que les modèles ajustent bien les étoiles :

$$\begin{aligned}
 T_{\text{eff,cor}} &= T_{\text{eff}} & (8.11) \\
 &+ \begin{cases} 0.0 & \text{si } -0.3 \leq [\text{Fe}/\text{H}], \\ 0.03 [\text{Fe}/\text{H}] + 0.009 & \text{si } -1.0 \leq [\text{Fe}/\text{H}] < -0.3, \\ -0.014 [\text{Fe}/\text{H}] - 0.036 & \text{si } -1.5 \leq [\text{Fe}/\text{H}] < -1.0, \\ -0.015 & \text{si } -1.5 \geq [\text{Fe}/\text{H}]. \end{cases}
 \end{aligned}$$

### 8.1.1.2 GCS II

Dans cette version du GCS, Holmberg et al. (2007, HNA07) utilisent des calibrations différentes pour la température effective et la métallicité. Pour la température effective elles sont basées sur la calibration de di Benedetto (1998) qui utilise la couleur  $(V - K)$ . HNA07 ont modifié cette calibration afin d'utiliser la couleur  $(b - y)$ . Trois intervalles de couleurs sont considérés. Pour  $0.20 < (b - y) < 0.33$  :

$$\begin{aligned}
 \theta_{\text{eff}} &= 0.649 + 0.095(b - y) + 0.034[\text{Fe}/\text{H}] + 1.207(b - y)^2 \\
 &\quad - 0.005[\text{Fe}/\text{H}]^2 - 0.181(b - y)[\text{Fe}/\text{H}]. & (8.12)
 \end{aligned}$$

Pour  $0.33 < (b - y) < 0.50$  :

$$\begin{aligned}
 \theta_{\text{eff}} &= 0.754 - 0.365(b - y) - 0.001[\text{Fe}/\text{H}] + 1.635(b - y)^2 \\
 &\quad - 0.011[\text{Fe}/\text{H}]^2 - 0.091(b - y)[\text{Fe}/\text{H}]. & (8.13)
 \end{aligned}$$

Pour  $0.50 < (b - y) < 0.60$  :

$$\begin{aligned} \theta_{\text{eff}} &= 0.255 + 1.656(b - y) + 0.018[\text{Fe}/\text{H}] - 0.397(b - y)^2 \\ &- 0.011[\text{Fe}/\text{H}]^2 - 0.101(b - y)[\text{Fe}/\text{H}]. \end{aligned} \quad (8.14)$$

Où  $\theta_{\text{eff}} = 5040/T_{\text{eff}}$ .

HNA07 utilisent certaines calibrations de NMA04 pour la métallicité (équations 8.7 et 8.9) en fonction de leurs intervalles respectifs. Enfin pour  $0.30 < (b - y) < 0.46$  NMA04 utilisent la calibration suivante :

$$\begin{aligned} [\text{Fe}/\text{H}] &= -2.19 - 1.02(b - y) + 7.34m_1 - 0.27c_1 + 5.86(b - y)^2 \quad (8.15) \\ &- 43.74m_1^2 - 0.14c_1^2 + 25.03(b - y)m_1 + 5.29(b - y)c_1 \\ &+ 25.95m_1c_1 - 31.10(b - y)^3 + 46.19m_1^3 - 3.86c_1^3 \\ &+ 4.54(b - y)^2m_1 + 19.31(b - y)^2c_1 - 17.46m_1^2(b - y) \\ &+ 18.36m_1^2c_1 - 9.99c_1^2(b - y) + 6.60c_1^2m_1 - 59.65(b - y)m_1c_1. \end{aligned}$$

Dans cette version du catalogue, les âges sont également déterminés avec l'inférence bayésienne de Jørgensen & Lindegren (2005) avec les corrections sur la métallicité (équation 8.10) et sur la température effective (équation 8.11).

### 8.1.1.3 GCS III

Le travail de van Leeuwen (2007), sur le catalogue Hipparcos, réduit les erreurs sur les parallaxes d'un facteur 1.5. Les distances et les magnitudes absolues sont donc modifiées. Dans cette version du GCS, Holmberg et al. (2009, HNA09) modifient donc les magnitudes absolues puis HNA09 déterminent à nouveau les âges avec la méthode de Jørgensen & Lindegren (2005).

### 8.1.1.4 GCS CSA11

Casagrande et al. (2011, CSA11) déterminent les paramètres stellaires du catalogue à partir des observations du GCS I. L'enrichissement en éléments  $\alpha$  est déterminé par la spectroscopie haute résolution. CSA11 déterminent la métallicité  $[\text{Fe}/\text{H}]$  à partir de deux calibrations qu'ils ont réalisées. L'une pour les étoiles avec  $0.23 \leq (b - y) \leq 0.63$ ,  $0.05 \leq m_1 \leq 0.68$  et  $0.13 \leq c_1 \leq 0.60$  :

$$\begin{aligned} [\text{Fe}/\text{H}] &= 3.927 \log m_1 - 14.459m_1^3 - 5.394(b - y) \log m_1 \quad (8.16) \\ &+ 36.069(b - y)m_1^3 + 3.537c_1 \log m_1 - 3.500m_1^3c_1 \\ &+ 11.034(b - y) - 22.780(b - y)^2 + 10.684c_1 - 6.759c_1^2 - 1.584. \end{aligned}$$

Pour les étoiles rouges avec  $0.43 \leq (b - y) \leq 0.63$ ,  $0.07 \leq m_1 \leq 0.68$  et  $0.16 \leq c_1 \leq 0.49$ , ils utilisent :

$$\begin{aligned} [\text{Fe}/\text{H}] &= -0.116c_1 - 1.624c_1^2 + 8.955(b - y)c_1 & (8.17) \\ &+ 42.008(b - y) - 99.596(b - y)^2 + 64.245(b - y)^3 \\ &+ 8.928c_1m_1 + 17.275m_1 - 48.106m_1^2, \end{aligned}$$

La métallicité  $[M/\text{H}]$  est déterminée avec des calibrations similaires à celles-ci mais qui prennent en compte l'enrichissement en éléments  $\alpha$ .

CSA11 utilisent deux méthodes pour déterminer la température effective : la méthode du flux infrarouge (IRFM InfraRed Flux Method) et la calibration photométrique (CLBR). La méthode IRFM (Casagrande et al. 2010) est utilisée pour 6670 étoiles qui ont de bonnes mesures photométriques dans toutes les bandes. Ces étoiles sont également supposées non variables et non multiples. CSA11 ont conservé les étoiles avec  $\sigma_{B_T} + \sigma_{V_T} < 0.10$  où  $B_T$  et  $V_T$  sont les magnitudes dans le système de Tycho2 (Høg et al. 2000) et  $\sigma_{J_S} + \sigma_{H_S} + \sigma_{K_S} < 0.10$  où  $J_S$ ,  $H_S$  et  $K_S$  sont les magnitudes dans le système 2MASS (Skrutskie et al. 2006). La méthode IRFM utilise le flux bolométrique  $F_{bol}$  et le flux bolométrique en infrarouge  $F_{\lambda_{IR}}$  qui sont obtenus à partir des observations. CSA11 obtiennent une première mesure de la température effective  $T_{\text{eff}}$  à partir du flux bolométrique, de la métallicité  $[\text{Fe}/\text{H}]$  et de la gravité de surface  $\log g$ . Celle-ci est déterminée avec  $\log(g/g_{\odot}) = \log(M/M_{\odot}) + 4 \log(\log T_{\text{eff}}/T_{\text{eff},\odot}) - \log(L/L_{\odot})$  où la luminosité bolométrique  $L$  est déterminée à partir du flux. Cela permet donc de sélectionner un spectre synthétique dans la bande infrarouge. En comparant le flux bolométrique en infrarouge observé avec celui du spectre synthétique, ils extraient le rayon de l'étoile. La température effective est déduite grâce à la loi de Stefan-Boltzmann  $L = 4\pi\sigma R^2 T_{\text{eff}}^4$  où  $\sigma$  est la constante de Boltzmann. CSA11 itèrent cette méthode jusqu'à converger vers une température effective.

Pour les autres étoiles CSA11 appliquent la calibration photométrique de Casagrande et al. (2010) avec  $\theta_{\text{eff}} = a_0 + a_1X + a_2X^2 + a_3X[\text{Fe}/\text{H}] + a_4[\text{Fe}/\text{H}] + a_5[\text{Fe}/\text{H}]^2$ ,  $X$  est la couleur et  $a_i$  les coefficients (tableau 8.1). La température effective de l'étoile correspond à la moyenne de celles calculées avec les couleurs  $(B_T - V_T)$ ,  $(V_T - J_S)$ ,  $(V_T - H_S)$ ,  $(V_T - K_S)$  et  $(b - y)$ .

La détermination de l'âge est basée sur une inférence bayésienne qui est décrite dans la partie 3.1.3.

Pour les applications on utilise cette version du catalogue. Le diagramme HR des étoiles est représenté sur la figure 8.1.

TABLE 8.1 – Coefficients  $a_i$  pour la calibration de la température effective selon la couleur  $X$  et la métallicité  $[\text{Fe}/\text{H}]$  de l'étoile (Casagrande et al. 2010). Les intervalles pour lesquels la calibration est applicable ( $[\text{Fe}/\text{H}]$  et couleur) sont donnés pour chaque couleur.

$X$	$\Delta[\text{Fe}/\text{H}]$	$\Delta X$	$a_0$	$a_1$	$a_2$	$a_3$	$a_4$	$a_5$
$(B_T - V_T)$	[-2.7,0.4]	[0.19,1.49]	0.5839	0.4000	-0.0067	-0.0282	-0.0346	-0.0087
$(V_T - J_S)$	[-2.7,0.4]	[0.77,2.56]	0.4525	0.3797	-0.0357	-0.0082	0.0123	-0.0009
$(V_T - H_S)$	[-2.7,0.4]	[0.77,3.16]	0.5286	0.2354	-0.0073	-0.0182	0.0401	0.0021
$(V_T - K_S)$	[-2.7,0.4]	[0.99,3.29]	0.4892	0.2634	-0.0165	-0.0121	0.0249	-0.0001
$(b - y)$	[-3.7,0.5]	[0.18,0.72]	0.5796	0.4812	0.5747	-0.0633	0.0042	-0.0055

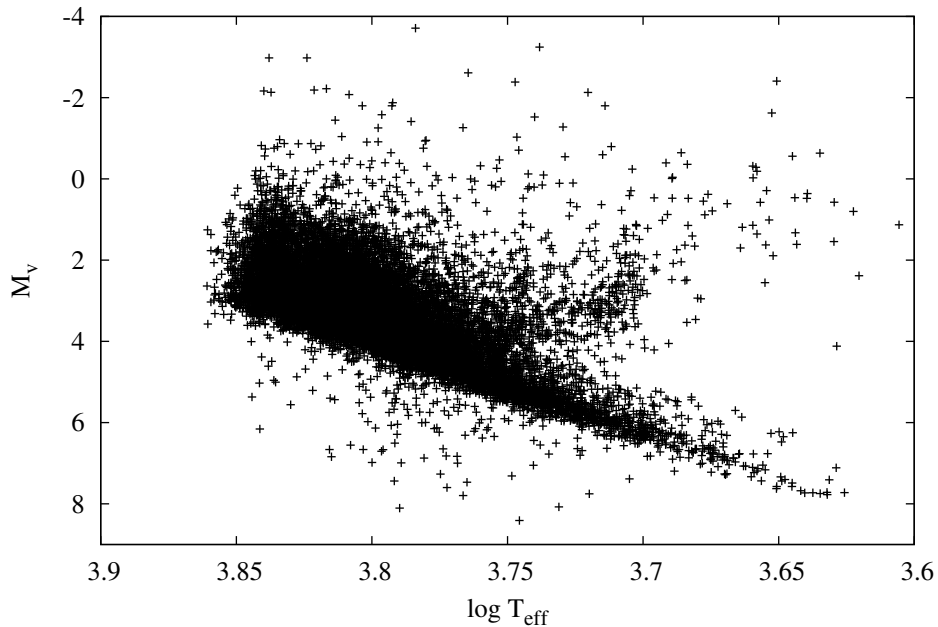


FIGURE 8.1 – Diagramme HR du GCS de CSA11 (Casagrande et al. 2010). Ils contient 16 682 étoiles de type FGK situées à une distance inférieure à 40 pc.

### 8.1.1.5 Comparaison du GCS III et du GCS CSA11

Les comparaisons des températures effectives, des magnitudes absolues et des métallicités  $[\text{Fe}/\text{H}]$  entre les deux versions sont représentées sur les figures 8.2). Nous constatons qu'il y a des différences sur les trois paramètres. Les différences sur les métallicités et les températures effectives proviennent des calibrations qui sont différentes. Les différences sur les magnitudes absolues

sont dus à la parallaxe est différente et aux transformations de couleurs entre les différents filtres.

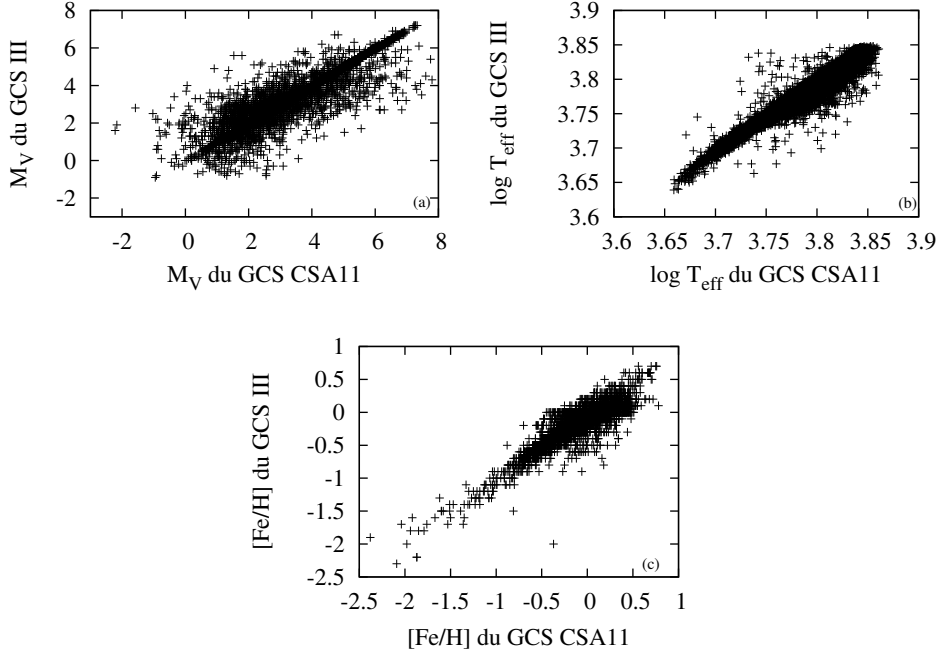


FIGURE 8.2 – Comparaison des magnitudes absolues, des températures effectives et des métallicités entre le GCS III Holmberg et al. (2009) et celui de Casagrande et al. (2010).

### 8.1.2 Catalogue de Setiawan et al.

Setiawan et al. (2004) ont réalisé des observations avec le spectrographe FEROS sur 83 étoiles géantes FK pendant 2 ans et demi. Ces observations ont permis de déterminer la vitesse radiale, la magnitude apparente  $V$  et la couleur ( $B - V$ ) de ces étoiles. Pour la détermination des autres paramètres, nous utilisons ceux déterminés par da Silva et al. (2006, dSGP06) pour 72 étoiles. Dans un premier temps, dSGP06 calculent la température effective avec la calibration d'Alonso et al. (1999) à partir de la couleur ( $B - V$ ) :

$$\begin{aligned} \theta_{\text{eff}} = & 0.6177 + 0.4354(B - V) - 4.025 \cdot 10^{-3}(B - V)^2 \\ & - 5.204 \cdot 10^{-2}(B - V)[\text{Fe}/\text{H}] - 0.1127[\text{Fe}/\text{H}] - 1.385 \cdot 10^{-2}[\text{Fe}/\text{H}], \end{aligned} \quad (8.18)$$

où  $\theta_{\text{eff}} = 5040/T_{\text{eff}}$ .

La métallicité  $[\text{Fe}/\text{H}]$  est déterminée à partir de l'abondance en Fe I. La gravité de surface  $\log g$  est obtenue par inversion dans les trajets évolutifs



de Girardi et al. (2000), à partir de la température effective et de la magnitude absolue  $M_V$ . Les magnitudes absolues sont calculées grâce aux parallaxes d’Hipparcos. dSGP06 déterminent les âges avec une inférence bayésienne qui est décrite dans la partie 3.1.2.

Le diagramme HR de ces étoiles est représenté sur la figure 8.3.

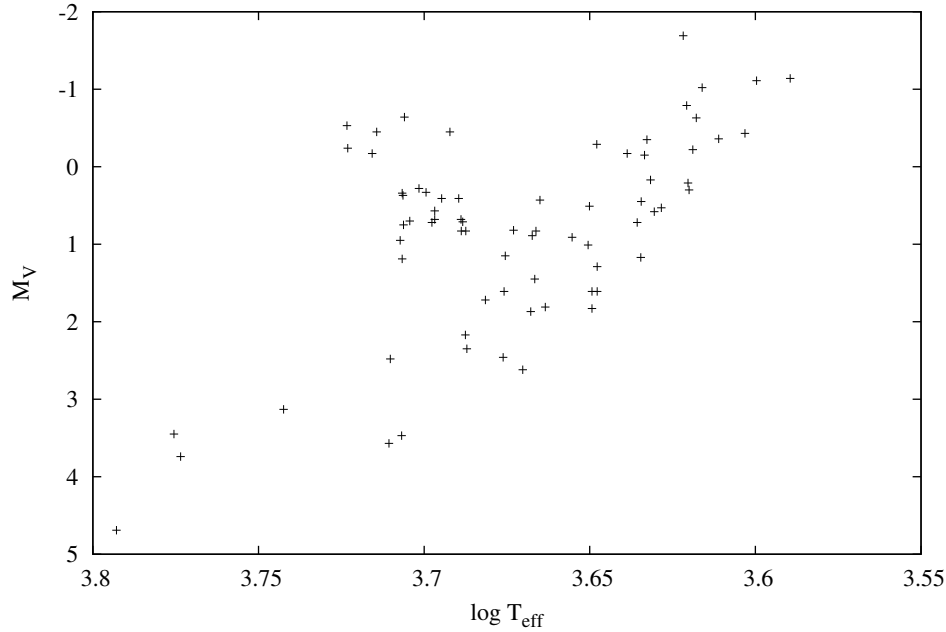


FIGURE 8.3 – Diagramme HR des 72 géantes FK observées par Setiawan et al. (2004) dont les paramètres ont été déterminés par da Silva et al. (2006).

## 8.2 Comparaison avec les travaux antérieurs

Dans la partie précédente, nous avons présenté trois catalogues observationnels, le GCS III, le GCS CSA11 et le catalogue de Setiawan et al. (2004). Les âges des étoiles de ces catalogues sont déterminés par des inférences bayésiennes présentées dans la partie 3.1 et que nous avons également implémentées. Nous comparons l’âge des étoiles des catalogues avec l’âge de ces mêmes étoiles mais déterminé par notre implémentation de leurs méthodes. Pour effectuer cette comparaison nous avons adopté les mêmes *a priori* que leurs méthodes et nous n’appliquons pas de traitement sur les fonctions  $\psi(\tau|x)$ . Pour les modèles stellaires, nous utilisons les trajets évolutifs de BaSTI avec la résolution de la grille présentée dans le tableau 2.3.2. Les résultats sont présentés sur la figure 8.4.

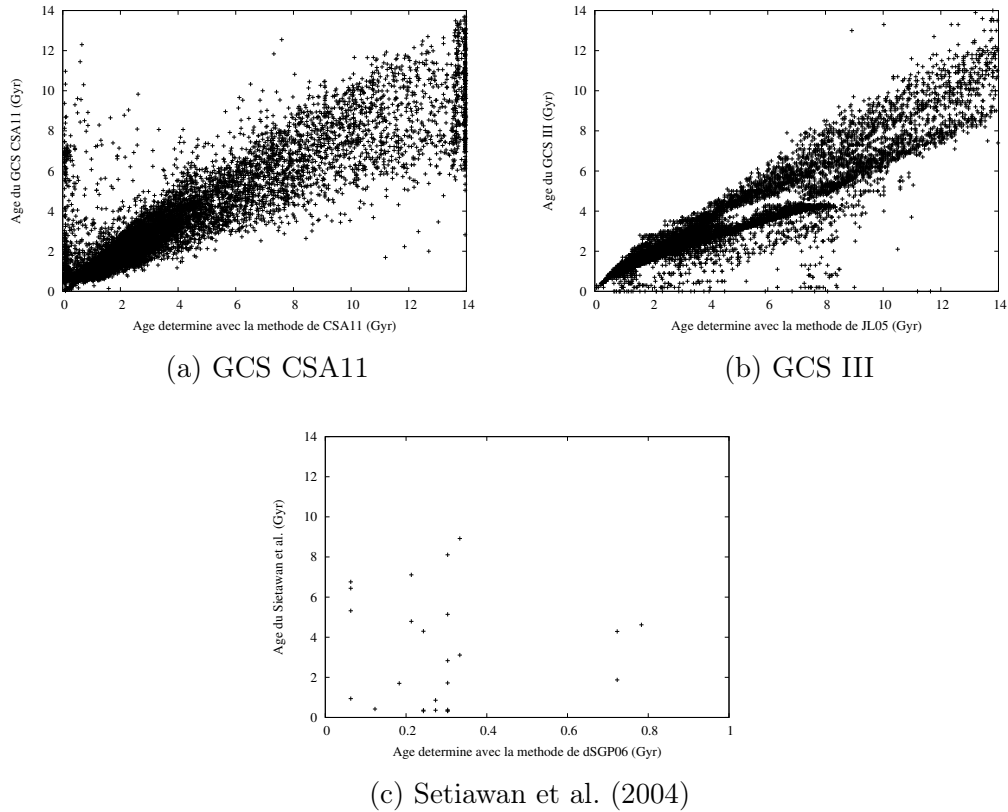


FIGURE 8.4 – Comparaison entre les âges déterminés avec la méthode de JL05 pour le GCS III, avec la méthode dSGP06 pour le catalogue de Setiawan et al. (2004) et avec la méthode de CSA 11 pour le GCS CSA 11.

Parmi les 72 géantes rouges du catalogue de Setiawan et al. 29 ont passés le test de rejet du  $\chi^2$  ; pour ces étoiles nous obtenons des âges inférieurs à ceux obtenus par dSGP06. Ces étoiles sont situées dans une zone de dégénérescence où les âges sont difficiles à restituer. C'est également dans cette région que nous obtenons le plus grand écart entre l'âge déterminé et celui de l'étoile (figure 5.2). Pour calculer l'âge de ces étoiles, nous avons utilisé l'IMF de K02 car dSGP06 ne précisent pas l'IMF utilisée. Par contre, le calcul par la méthode de dSGP06, qui peut être effectué sur internet<sup>1</sup>, détermine les âges avec les IMFs de Kroupa et al. (1993), Scalo (1998) et Chabrier (2003). La différence entre ces deux IMFs induit un léger effet dans la détermination des âges comme nous l'avons montré dans la partie 2.2.2. Enfin dSGP06 utilisent les isochrones de Padoue dont l'âge maximum est de 12 Gyr. Pour toutes ces raisons il est difficile de conclure sur la comparaison entre les âges déterminés avec notre implémentation et ceux de dSGP06. Pour les étoiles du GCS III,

1. PARAM : <http://cgi-bin/param>

nous obtenons également des âges supérieurs à ceux de JL05. Ces différences peuvent être expliquées par les modèles stellaires : nous utilisons les trajets évolutifs de BaSTI alors que JL05 prennent les isochrones de Padoue.

Enfin pour les étoiles du GCS CSA11, nous obtenons des différences parfois importantes entre les âges bien que les deux méthodes utilisent les modèles stellaires de BaSTI<sup>1</sup>. La méthode de CSA11 ne donne aucun âge très petit ou très grand, contrairement à la nôtre ; cela peut être dû à l'inclusion d'un traitement sur les fonctions  $\psi(\tau|x)$ .

### 8.3 [M/H] vs. [Fe/H]

Les modèles d'évolution stellaire sont calculés pour une métallicité [M/H], alors que les observations donnent la métallicité [Fe/H]<sup>2</sup>. Ainsi pour déterminer les âges, nous supposons que ces deux métallicités sont égales. Le catalogue GCS CSA11 fournit les métallicités [M/H] et [Fe/H] représentées sur la figure 8.5. Il y a un décalage entre les deux métallicités, lorsque [Fe/H] est inférieure à zéro, [M/H] est supérieure à [Fe/H]. Au contraire, pour [Fe/H] supérieure à zéro [M/H] est inférieure à [Fe/H].

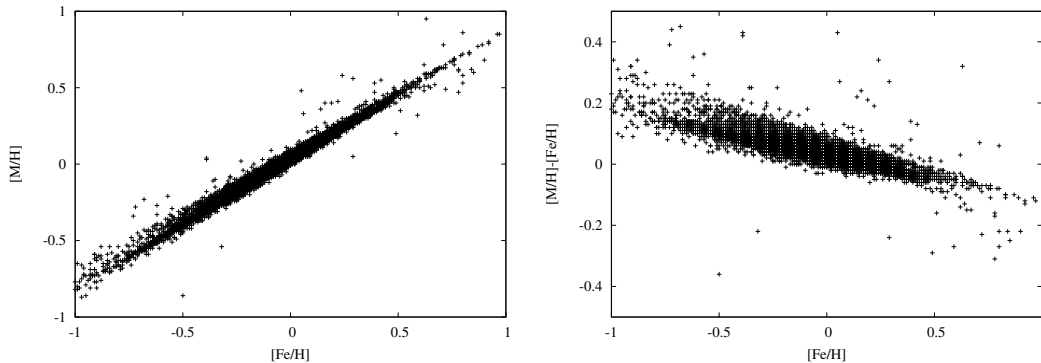


FIGURE 8.5 – Comparaison entre les métallicités (figure de gauche) [M/H] et [Fe/H] pour les étoiles du GCS CSA11. La différence entre les métallicités en fonction de la métallicité [Fe/H] est représentée sur la figure de droite.

Nous déterminons les âges avec ces deux quantités afin d'en analyser les effets ; les résultats sont présentés sur la figure 8.6. Nous avons identifié deux régions (en carrés rouges et en ronds verts) où le comportement des âges est

1. Notons que CSA11 utilisent une grille calculée spécialement par l'équipe de BaSTI et de résolution beaucoup plus fine que la grille 0 disponible sur internet dont nous sommes partis.

2.  $[Fe/H] = \log(N_{Fe}/N_H) - \log(N_{Fe}/N_H)_{\odot}$ .  $N_H$  est le nombre d'atomes d'hydrogène par unité de volume et  $N_{Fe}$  est le nombre d'atomes de fer.

particulier. Les 291 étoiles en rouge forment une seconde séquence et les 31 étoiles en vert ont des âges plus élevés avec la métallicité  $[\text{Fe}/\text{H}]$ . Ces deux régions sont également identifiées dans le diagramme HR et dans le diagramme  $\log T_{\text{eff}} - [\text{M}/\text{H}]$  (figure 8.7). Ces étoiles sont localisées sur la séquence principale et dans la branche des sous-géantes. Leurs fonctions  $\psi(\tau|x)$  ont soit plusieurs pics, soit ont subi un lissage, ce qui montre une incertitude dans la détermination de l'âge.

Pour la détermination des âges, supposer que  $[\text{M}/\text{H}] = [\text{Fe}/\text{H}]$  crée un léger biais pour environ 5 % des étoiles.

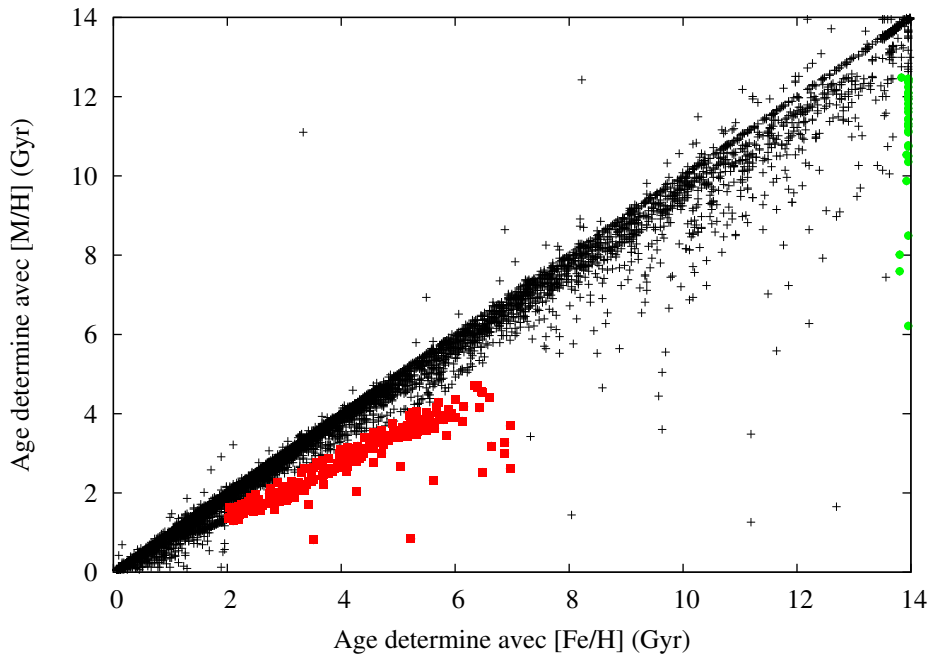


FIGURE 8.6 – Comparaison entre les âges déterminées avec  $[\text{M}/\text{H}]$  et  $[\text{Fe}/\text{H}]$  pour les étoiles du GCS CSA11. Les étoiles en rouge et en vert sont identifiées pour vérifier leur position dans le diagramme HR et le diagramme  $\log T_{\text{eff}} - [\text{M}/\text{H}]$  (figure 8.7).

## 8.4 Datation des amas

Il existe deux types d'amas : les amas globulaires et les amas ouverts. Ces derniers, situés dans le disque de la Galaxie, sont composés de quelques dizaines à quelques milliers d'étoiles plutôt jeunes. Les amas globulaires sont composés de dizaines de milliers à plusieurs millions d'étoiles âgées et sont situés dans le halo galactique.

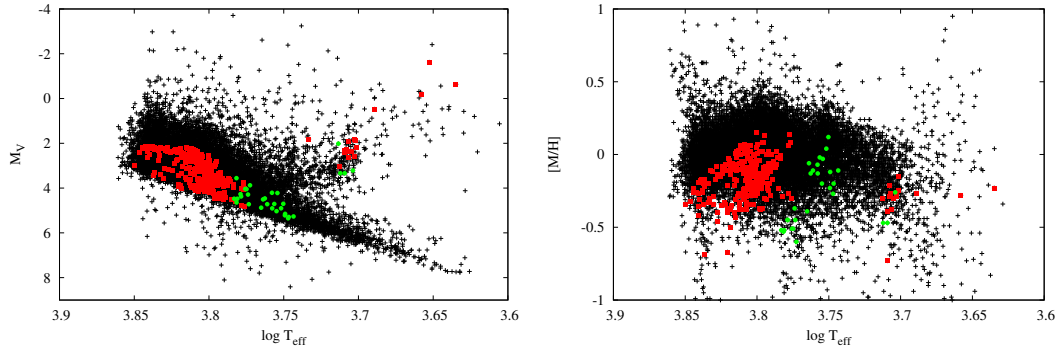


FIGURE 8.7 – Position des étoiles de la figure 8.6 dans le diagramme HR (gauche) et dans le diagramme  $\log T_{\text{eff}} - [M/H]$  (droite), pour les étoiles du GCS CSA11.

Les amas sont composés d'étoiles qui se sont formées dans le même nuage moléculaire. si l'amas n'est pas trop jeune on peut considérer que toutes ces étoiles ont le même âge car le temps de formation stellaire peut être négligé par rapport à leur durée de vie. Elles ont la composition du nuage moléculaire dont elles sont issues, on considère donc qu'elles ont toutes la même métallicité. Enfin toutes les étoiles sont liées gravitationnellement, elles ont des vitesses radiales semblables et leurs distances au soleil sont identiques. En revanche, étant donné que la masse initiale des étoiles diffère, elles sont dans des stades d'évolution différents.

Pour dater les amas, la méthode la plus courante consiste à ajuster les étoiles de l'amas par des isochrones de la métallicité égale à celle des étoiles de l'amas. L'isochrone qui ajuste le mieux les étoiles donne l'âge de l'amas. Habituellement, cet ajustement est fait au niveau du turn-off.

Nous pouvons dater un amas, en déterminant l'âge des étoiles qui le compose par notre méthode d'inférence bayésienne, par les trois méthodes suivantes :

- en réalisant la moyenne des âges obtenus,
- en sommant les fonctions  $\psi(\tau|x)$  bien définie,
- en sommant toutes les fonctions  $\psi(\tau|x)$  sauf celles non datées.

Dans les deux derniers cas, l'âge de l'amas correspond au mode de cette fonction. En prenant en compte les fonctions tronquées avec plateau nous incluons une information supplémentaire dans la datation de l'amas car elles donnent une limite inférieure ou supérieure à l'âge de l'amas. La figure 8.8 représente, pour l'amas Berkeley 70, l'estimation des âges obtenus par ces trois méthodes. Les âges déterminés à partir des deux sommes sur les fonctions  $\psi(\tau|x)$  sont similaires pour cet amas. Nous constatons également que le mode de la fonction

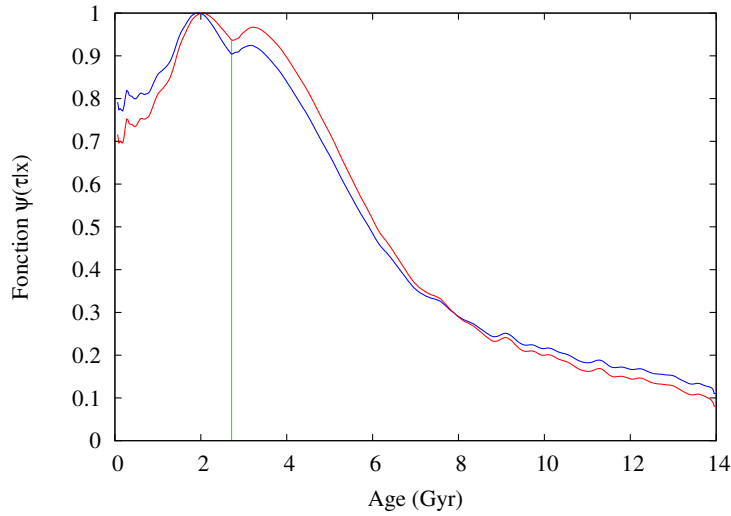


FIGURE 8.8 – Détermination de l'âge de l'amas Berkeley70 avec la moyenne des modes des fonctions  $\psi(\tau|x)$  en vert, avec la fonction sommée sur toutes les fonctions  $\psi(\tau|x)$  sauf celles sans âge en bleu et la fonction sommée sur toutes les fonctions bien définies en rouge.

sommée ne correspond pas forcément à la moyenne des âges car les fonctions  $\psi(\tau|x)$  donnent des informations supplémentaires.

Nous avons daté 18 amas listés dans la table 8.2. Les Hyades (de Bruijne et al. 2001) est l'amas ouvert le plus proche du soleil à environ 46 pc. Nous étudions également 14 amas ouverts observés dans l'anti-centre de la Galaxie par Hasegawa et al. (2004), M35 (Deliyannis & Steinhauer 2001) et deux amas ouverts observés par Yadav & Sagar (2004). Les Hyades et Basel 4 sont représentés sur la figure 8.9. Nous avons choisi de dater ces amas car nous disposons pour chacun d'entre eux des observables nécessaires à la datation.

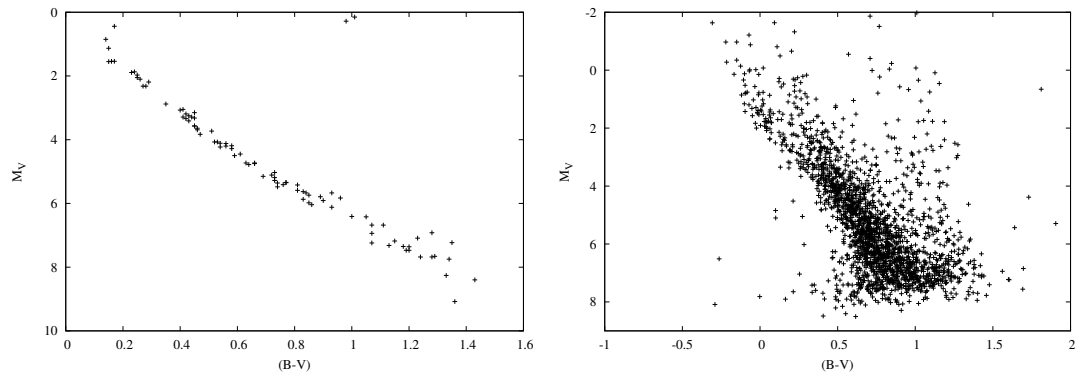


FIGURE 8.9 – Diagramme HR des Hyades (gauche) et de Basel 4 (droite).

Pour les Hyades nous disposons de la couleur intrinsèque  $(B - V)_0$ , de la magnitude absolue  $M_V$  et de la métallicité  $[\text{Fe}/\text{H}]$ . En revanche, pour les autres amas, nous disposons de la couleur observée  $(B - V)$ , de la magnitude apparente  $V$ , de la distance  $d$ , de l'extinction  $E(B - V)$  et de la métallicité  $[\text{Fe}/\text{H}]$ . Nous calculons la couleur intrinsèque avec la relation suivante :  $(B - V)_0 = (B - V) - E(B - V)$ . La magnitude absolue  $M_V$  est déterminée à partir de la distance, de la magnitude apparente  $V$  et de l'absorption  $A_V$  telle que  $A_V = 3.1 E(B - V)$  (Ryter 1996).

Les erreurs observationnelles de la couleur intrinsèque  $(B - V)_0$  et de la magnitude absolue  $M_V$  sont calculées à partir des erreurs observationnelles de l'extinction  $E(B - V)$ , de la distance  $d$ , de la couleur observée  $(B - V)$  et de la magnitude apparente  $V$ . Pour les erreurs observationnelles sur la distance  $d$  et sur l'extinction  $E(B - V)$ , nous avons adopté celles de Yadav & Sagar (2004) et nous prenons 0.1 dex pour la métallicité.

Pour chacun de ces amas, l'âge a été déterminé par des études antérieures d'ajustement par des isochrones (table 8.2). Les résultats de la datation sont présentés dans la table 8.3 et sur la figure 8.10.

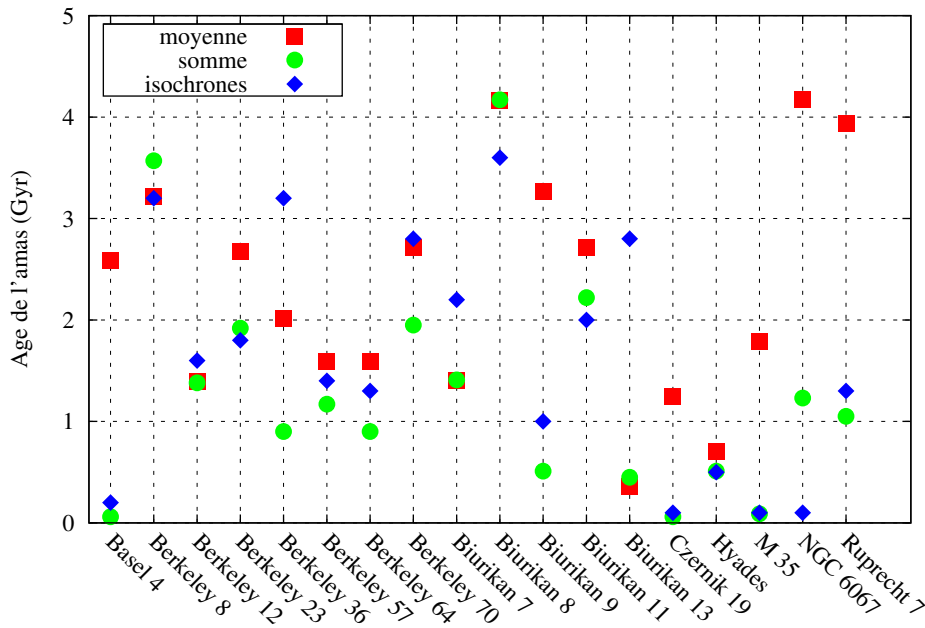


FIGURE 8.10 – Comparaison de l'âge des amas déterminés avec la moyenne des âges des étoiles de l'amas, avec la somme des fonctions  $\psi(\tau|x)$  et avec l'ajustement par des isochrones.

TABLE 8.2 – Caractéristiques des amas datés.

Noms	étoiles	[M/H] (dex)	$d$ (kpc)	$E(B - V)$ (mag)	Références
Basel 4	2056	$-0.35 \pm 0.1$	$3.0 \pm 0.2$	$0.45 \pm 0.05$	Yadav & Sagar (2004)
Berkeley 8	80	$-0.35 \pm 0.1$	$3.2 \pm 0.2$	$0.75 \pm 0.05$	Hasegawa et al. (2004)
Berkeley 12	91	$-0.35 \pm 0.1$	$3.8 \pm 0.2$	$0.8 \pm 0.05$	Hasegawa et al. (2004)
Berkeley 23	151	$-0.66 \pm 0.1$	$5.8 \pm 0.2$	$0.3 \pm 0.05$	Hasegawa et al. (2004)
Berkeley 36	91	$0.0 \pm 0.1$	$6.1 \pm 0.2$	$0.4 \pm 0.05$	Hasegawa et al. (2004)
Berkeley 57	179	$-0.66 \pm 0.1$	$4.1 \pm 0.2$	$0.75 \pm 0.05$	Hasegawa et al. (2004)
Berkeley 64	69	$-0.66 \pm 0.1$	$4.3 \pm 0.2$	$0.95 \pm 0.05$	Hasegawa et al. (2004)
Berkeley 70	159	$-0.66 \pm 0.1$	$4.1 \pm 0.2$	$0.5 \pm 0.05$	Hasegawa et al. (2004)
Biurikan 7	184	$-0.35 \pm 0.1$	$9.2 \pm 0.2$	$0.15 \pm 0.05$	Hasegawa et al. (2004)
Biurikan 8	268	$-0.35 \pm 0.1$	$3.7 \pm 0.2$	$0.1 \pm 0.05$	Hasegawa et al. (2004)
Biurikan 9	151	$-0.66 \pm 0.1$	$3.4 \pm 0.2$	$0.4 \pm 0.05$	Hasegawa et al. (2004)
Biurikan 11	151	$0.25 \pm 0.1$	$5.0 \pm 0.2$	$0.05 \pm 0.05$	Hasegawa et al. (2004)
Biurikan 13	107	$0.0 \pm 0.1$	$7.2 \pm 0.2$	$0.75 \pm 0.05$	Hasegawa et al. (2004)
Czerwik 19	62	$-0.35 \pm 0.1$	$2.5 \pm 0.2$	$0.85 \pm 0.05$	Hasegawa et al. (2004)
Hyades	92	$0.14 \pm 0.05$	$0.045 \pm 0.003$	$0.003 \pm 0.002$	de Bruijne et al. (2001); Lebreton et al. (2001)
M 35	1150	$-0.21 \pm 0.1$	$0.83 \pm 0.4$	$0.255 \pm 0.024$	Deliyannis & Steinhauser (2001)
NGC 6067	1502	$0.0 \pm 0.1$	$3.6 \pm 0.2$	$0.75 \pm 0.05$	Yadav & Sagar (2004)
Ruprecht 7	158	$-0.35 \pm 0.1$	$3.9 \pm 0.2$	$0.3 \pm 0.05$	Hasegawa et al. (2004)



TABLE 8.3 – Ages des différents amas obtenus avec la moyenne des fonctions des modes, avec la somme des fonctions  $\psi(\tau|x)$ , la somme des fonctions  $\psi(\tau|x)$  bien définies et avec l'ajustement des isochrones donnés par des travaux antérieurs.

Noms de l'amas	Moyenne	Sommes	Sommes, bien-définies	Isochrones
Basel 4	2.59	0.06	0.06	0.2
Berkeley 8	3.22	3.57	3.54	3.2
Berkeley 12	1.39	1.38	1.38	1.6
Berkeley 23	2.67	1.92	1.95	1.8
Berkeley 36	2.01	0.90	0.87	3.2
Berkeley 57	1.59	1.17	1.17	1.4
Berkeley 64	1.59	0.90	0.87	1.3
Berkeley 70	2.71	1.95	2.04	2.8
Biurikan 7	1.40	1.41	1.41	2.2
Biurikan 8	4.16	4.17	3.60	3.6
Biurikan 9	3.27	0.51	0.54	1.0
Biurikan 11	2.71	2.22	2.16	2.0
Biurikan 13	0.36	0.45	0.45	2.8
Czernik 19	1.25	0.06	0.06	0.1
Hyades	0.70	0.51	0.51	0.5-0.65
M 35	1.79	0.09	0.12	0.1
NGC 6067	4.17	1.23	1.23	0.1
Ruprecht 7	3.94	1.05	1.08	1.3

Les amas Berkeley 8, Berkeley 12, Biurikan 7 et Biurikan 8 ont peu de fonctions  $\psi(\tau|x)$  mal définies, donc les âges obtenus par la moyenne et par la somme des fonctions  $\psi(\tau|x)$  sont similaires. Pour les autres amas, les âges obtenus par la moyenne et la somme des fonctions  $\psi(\tau|x)$  sont semblables. Les âges déterminés avec la somme des fonctions  $\psi(\tau|x)$  et la somme des fonctions  $\psi(\tau|x)$  bien définies sont similaires. Les fonctions tronquées avec plateau confirment l'information donnée par les autres fonctions  $\psi(\tau|x)$ .

Dans la majorité des cas, l'âge déterminé avec la somme des fonctions  $\psi(\tau|x)$  est plus proche de l'âge déterminé par l'ajustement des isochrones car la fonction  $\psi(\tau|x)$  permet de prendre en compte toute l'information sur la détermination des âges.

Pour dater les amas, nous avons utilisés les fonctions  $\psi(\tau|x)$  qui sont normalisées à leur maximum, or pour déterminer l'âge des amas il faut utiliser des fonctions  $\psi(\tau|x)$  telles que  $\int \psi(\tau|x)d\tau = 1$ . Cette différence de normalisation devrait avoir peu d'effets sur la détermination de l'âge des amas.

## 8.5 Conclusions et perspectives

Pour les applications sur les catalogues réels, l'hypothèse que la métallicité  $[M/H]$  est égale à la métallicité  $[Fe/H]$  entraîne un biais dans la détermination des âges pour 5 % des étoiles. Ce biais affecte les régions dégénérées du diagramme HR.

La datation des amas ouverts à partir des fonctions  $\psi(\tau|x)$  permet de restituer l'âge dans la majorité des cas étudiés. Nous avons également essayé de déterminer l'âge d'amas globulaires comme M3 et M13 situés à des distances respectives de 10.4 et 7.5 kpc. Cependant, les erreurs observationnelles étant grandes nous n'arrivons pas à déterminer l'âge de l'amas. Avec la mission spatiale Gaia, la détermination des paramètres sera plus précise ce qui permettra donc de dater les amas globulaires.

Cette méthode peut être utilisée pour des amas où l'on suppose qu'il y a deux périodes de formation stellaire. En utilisant la somme des fonctions  $\psi(\tau|x)$  un deuxième pic dans la fonction sommée devrait être présent. L'amas NGC 6791 a peut-être deux périodes de formation, mais comme sa métallicité est de 0.4 dex nous n'avons pas pu le dater car cette métallicité n'est pas disponible dans notre grille de trajets évolutifs.

# Conclusions générales et perspectives

---

La mission spatiale Gaia, qui va être lancée en novembre 2013, observera un milliard d'étoiles de la Galaxie. Dans ce contexte, il a été nécessaire de développer un programme automatique qui date cet échantillon. C'est dans ce cadre que cette thèse a été réalisée. Ce travail a permis de réaliser une étude détaillée sur la datation des étoiles par inférence bayésienne.

Afin de pouvoir étudier et analyser les résultats, nous avons construit des catalogues simulés, basés sur les spécifications de la mission spatiale Gaia. Ces catalogues contiennent 10 000 étoiles, ce qui n'est pas représentatif du nombre d'étoiles dans le voisinage solaire mais cela permet d'avoir une statistique suffisante.

Nous avons tout d'abord développé une méthode qui permet de déterminer l'âge et la masse par inférence bayésienne. Pour choisir les *a priori* nous avons réalisé de nombreux tests, en particulier sur la fonction de masse initiale qui influe beaucoup sur la restitution des âges. En effet, lorsque la fonction de masse initiale diffère de celle utilisée pour construire le catalogue simulé la restitution des âges est moins bonne. C'est pourquoi, pour la mission spatiale Gaia, les âges et les masses devront être déterminés par itération. Les connaissances de la fonction de masse initiale et du taux de formation stellaire pourront être affinées à partir d'une première détermination des masses et des âges. Ainsi des *a priori* plus représentatifs de la Galaxie pourront être ajoutés dans la détermination de la masse et de l'âge des étoiles. Il faudra itérer plusieurs fois afin de converger vers les âges et les masses les plus probables.

La méthode que nous avons développé permet de dater de façon satisfaisante 69 % des étoiles de notre catalogue simulé, avec une erreur relative entre l'âge réel et l'âge déterminé inférieure à 20 %. Les étoiles qui ont un âge mal déterminé sont localisées dans des régions particulières du diagramme HR : la branche des géantes rouges, à proximité de la séquence principale d'âge zéro et les étoiles avec une masse inférieure à une masse solaire sur la séquence principale. Ces dégénérescences sont causées par la vitesse d'évolution qui est lente sur la séquence principale et au contraire très rapide dans la branche des géantes rouges.

Ces résultats dépendent de la distance de l'étoile et de la direction d'observation. En effet, l'erreur observationnelle sur la magnitude apparente dépend de la distance de l'étoile, de l'extinction et de l'absorption qui dépend également de la distance et de la direction d'observation. Par contre, l'erreur observationnelle sur la température effective ne dépend que de la distance et de la température effective de l'étoile. Il est donc très difficile de prévoir si une étoile sera datée correctement. Nous avons réalisé une étude en choisissant quelques distances et quelques directions d'observation pour essayer de donner un indicateur sur la restitution des âges et des masses. Comme nous pouvons l'attendre les âges et les masses déterminés sont moins bons à de grandes distances et pour de grandes valeurs de l'absorption.

La détermination des âges dépend également de l'erreur observationnelle sur la métallicité. En effet et comme nous pouvons l'attendre plus l'erreur observationnelle sur la métallicité est petite, mieux les âges sont restitués. Afin de déterminer correctement l'âge des étoiles observées par la mission spatiale Gaia, il faut favoriser les métallicités obtenues par les relevés spectroscopiques complémentaires, comme le GAIA-ESO survey (Gilmore et al. 2012) qui donneront des métallicités plus précises.

L'une des originalités de ce travail a été d'ajouter des contraintes à notre méthode afin de réduire les dégénérescences. Les âges sont mieux restitués lorsque la contrainte apporte de nouvelles informations. Comme pour les autres observables, l'erreur observationnelle sur la contrainte doit être suffisamment petite. Nous avons testé l'ajout de contraintes dans la vraisemblance avec les paramètres astérosismiques, le paramètre des transits exoplanétaires et la gravité de surface. Nous obtenons une amélioration des résultats pour 20 % des étoiles en moyenne. Cette méthode est limitée à une minorité d'étoiles car nous avons besoin d'observations complémentaires pour l'appliquer. Il faut donc que l'étoile soit observée avec plusieurs instruments.

Ce travail a également permis de comparer l'inférence bayésienne avec la minimisation du  $\chi^2$ . Les résultats obtenus sont en apparence contre intuitifs. En effet quand la détermination des âges et le catalogue simulé utilisent les mêmes codes d'évolution stellaire la minimisation du  $\chi^2$  donne de meilleurs résultats que l'inférence bayésienne. En revanche lorsque les codes d'évolution stellaire sont différents l'inférence bayésienne donne de meilleurs résultats que la minimisation du  $\chi^2$ . Cela montre que la minimisation du  $\chi^2$  est plus sensible au codes d'évolution stellaire et que les *a priori* permettent de réduire les différences entre les codes d'évolution stellaire. Pour la mission spatiale Gaia il vaut mieux délivrer aussi bien les âges obtenus par l'inférence bayésienne qu'avec la minimisation du  $\chi^2$ .

Il serait également intéressant de fournir la forme de la fonction  $\psi(\tau|x)$  qui contient énormément d'information. En effet la forme de la fonction  $\psi(\tau|x)$

---

est associée à des régions particulières du diagramme HR, cela nous permet d'obtenir une information sur la vitesse d'évolution de l'étoile et son stade évolutif. Le traitement automatique des fonctions  $\psi(\tau|x)$  permet également d'identifier les différentes morphologies et ainsi de ne pas dater les étoiles qui ont une fonction mal-définie.

Néanmoins, il reste des études supplémentaires à faire afin de compléter ce travail :

- Pour la mission spatiale Gaia, il faudrait réaliser une étude plus complète en étudiant davantage de directions d'observations et de distances. Ainsi nous pourrions établir des statistiques plus précises sur la restitution des âges et des masses en fonction de la position des étoiles dans le diagramme HR, et ainsi donner une indication de l'exactitude des âges et des masses déterminés.
- Afin d'approfondir le travail réalisé sur l'ajout de contraintes dans l'inférence bayésienne, il faudrait une analyse plus complète des erreurs observationnelles sur les contraintes. Nous pourrions également utiliser d'autres contraintes comme les masses issues de l'observation d'étoiles binaires, les rayons spectroscopiques, la rotation, la fraction de masse en hélium  $Y$  et l'enrichissement en éléments  $\alpha$ .
- Il faut également essayer de bien comprendre les différences entre les âges déterminés par l'inférence bayésienne et par la minimisation du  $\chi^2$ . Pour vérifier que ces différences ne proviennent pas du traitement de l'erreur observationnelle sur la magnitude absolue, il faudrait inclure l'erreur observationnelle sur la parallaxe dans la vraisemblance (Casagrande et al. 2011). Pour cela, il faut également modifier le catalogue simulé afin d'inclure l'erreur observationnelle gaussienne sur la parallaxe.
- Nous avons déterminé les âges avec trois codes d'évolution stellaire et nous constatons que cela influe sur la restitution des âges. Il faudrait étudier les effets des différences entre les codes d'évolution stellaire afin de quantifier ces effets sur la restitution des âges.



# Quatrième partie

## Annexes





ANNEXE A

# Tables des résultats

---

TABLE A.1 – Table des résultats de tous les calculs réalisés pour la détermination de l'âge.

Calcul	Nombre d'étoiles avec $\min(\chi^2) > \chi_{lim}^2$		Nombre d'étoiles avec fonctions mal-définies		Pourcentage d'étoiles avec $\Delta\tau/\tau < 10\%$		Pourcentage d'étoiles avec $\Delta\tau/\tau < 50\%$	
	67	3131			49.7	68.5	85.7	
<b>Inférence bayésienne</b>								
<b>Comparaisons des priors</b>								
Avec MDF, $\sigma_{[M/H]} = 0.01$ dex	137	2233			65.2	78.1	89.9	
Avec MDF, $\sigma_{[M/H]} = 0.1$ dex	89	2447			67.8	79.0	89.6	
Avec MDF, $\sigma_{[M/H]} = 0.2$ dex	67	3131			49.7	68.5	85.7	
Avec MDF, $\sigma_{[M/H]} = 0.3$ dex	43	4113			35.1	54.3	80.7	
Avec MDF, $\sigma_{[M/H]} = 0.5$ dex	26	4792			26.4	44.1	73.4	
MDF monotone, $\sigma_{[M/H]} = 0.01$ dex	137	2233			65.2	78.1	89.9	
MDF monotone, $\sigma_{[M/H]} = 0.1$ dex	89	2447			67.8	79.0	89.6	
MDF monotone, $\sigma_{[M/H]} = 0.2$ dex	67	3131			49.7	68.5	85.7	
MDF monotone, $\sigma_{[M/H]} = 0.3$ dex	43	4113			35.2	54.3	80.7	
MDF monotone, $\sigma_{[M/H]} = 0.5$ dex	26	4792			26.4	44.1	73.4	
IMF de S55	67	3130			47.0	66.6	85.8	
IMF de MS79	67	2897			45.2	64.5	82.5	
IMF de S98	67	2831			46.0	64.8	81.8	
IMF de KTG93	67	3124			49.5	68.2	86.7	
IMF de K01	67	3206			47.9	67.5	86.7	
IMF de C03	67	2401			40.0	56.3	75.5	
IMF Monotone	67	3198			24.7	39.6	63.5	
Sans SFR	67	3136			49.7	68.5	85.7	
<b>Résolution de la grille</b>								
Grille 0	604	3936			15.1	28.5	57.6	

TABLE A.1 – Suite.

Calcul	Nombre d'étoiles avec		Pourcentage d'étoiles avec		
	$\min(\chi^2) > \chi_{lim}^2$	fonctions mal-définies	$\Delta\tau/\tau < 10\%$	$\Delta\tau/\tau < 20\%$	$\Delta\tau/\tau < 50\%$
Grille 1	290	3299	49.0	66.8	84.4
Grille 2	297	3151	49.0	67.4	85.0
$\Delta\tau = 0.015$ Gyr	286	3211	49.2	67.8	85.2
$\Delta\tau = 0.03$ Gyr	288	3152	49.0	67.5	85.0
$\Delta\tau = 0.06$ Gyr	294	3133	49.3	68.2	86.5
$\Delta M = 0.002 M_{\odot}$	290	2943	48.9	66.5	84.1
$\Delta M = 0.003 M_{\odot}$	291	2979	48.8	66.4	84.2
$\Delta M = 0.005 M_{\odot}$	296	3017	37.4	66.5	83.9
$\Delta[M/H] = 0.05$ dex	287	3403	49.3	67.3	84.6
$\Delta[M/H] = 0.08$ dex	288	3466	48.5	67.1	84.8
$\Delta[M/H] = 0.1$ dex	291	3464	49.6	67.8	85.1
<b>Isochrones</b>	76	3506	39.4	59.4	84.8
<b>Sans traitement sur <math>\psi(\tau x)</math></b>	67	0	39.6	57.9	77.0
<b>Effets sur les âges</b>					
Densité d'étoiles	67	615	11.2	17.9	29.5
$[M/H] \pm 1\sigma_{[M/H]}$	90	2652	61.8	76.0	88.9
$[M/H] \pm 2\sigma_{[M/H]}$	67	3130	49.9	68.7	85.9
Pas en $\log(age)$	67	3131	46.3	65.7	83.4
<b>Comparaisons des méthodes</b>					
JL05	69	3138	49.9	68.0	85.2
dSGP06	60	3009	45.0	65.4	85.3
CSA11	68	2916	10.7	19.8	41.5

TABLE A.1 – Suite.

Calcul	Nombre d'étoiles avec $\min(\chi^2) > \chi_{lim}^2$		Pourcentage d'étoiles avec		
	$\Delta\tau/\tau < 10\%$	$\Delta\tau/\tau < 20\%$	$\Delta\tau/\tau < 10\%$	$\Delta\tau/\tau < 20\%$	$\Delta\tau/\tau < 50\%$
$\chi_{NG98}^2$	116	0	61.7	76.2	90.7
$\chi_{ST}^2$	68	0	60.9	75.5	90.2
<b>Age conjoint</b>	67	0	38.7	52.7	72.5
<b>Catalogue simulé avec BaSTI</b>					
Bayes tr. de CESAM	225	2927	16.4	32.3	74.3
Bayes tr. de Padoue	33	2417	21.2	41.6	77.8
$\chi_{ST}^2$ tr. de CESAM	223	0	12.4	27.0	59.3
$\chi_{ST}^2$ tr. de Padoue	33	0	15.3	33.9	69.7
<b>Catalogue simulé avec CESAM</b>					
Bayes tr. de BaSTI	130	3312	20.4	39.2	74.4
Bayes tr. de CESAM	122	2290	33.3	49.5	74.3
Bayes tr. de Padoue	77	2320	24.1	47.9	72.0
$\chi_{ST}^2$ tr. de BaSTI	131	0	14.8	31.5	63.0
$\chi_{ST}^2$ tr. de CESAM	121	0	60.2	73.4	88.1
$\chi_{ST}^2$ tr. de Padoue	80	0	21.3	40.2	63.1
<b>Catalogue simulé avec Padoue</b>					
Bayes tr. de BaSTI	367	3294	18.3	36.9	62.3
Bayes tr. de CESAM	419	2453	18.9	35.5	64.1
Bayes tr. de Padoue	0	3179	35.9	54.0	74.4
$\chi_{ST}^2$ tr. de BaSTI	332	0	13.6	24.5	56.3
$\chi_{ST}^2$ tr. de CESAM	405	0	15.0	30.9	53.6
$\chi_{ST}^2$ tr. de Padoue	1	0	48.1	62.3	79.6

TABLE A.1 – Suite.

Calcul	Nombre d'étoiles avec		Pourcentage d'étoiles avec		
	$\min(\chi^2) > \chi_{lim}^2$	fonctions mal-définies	$\Delta\tau/\tau < 10\%$	$\Delta\tau/\tau < 20\%$	$\Delta\tau/\tau < 50\%$
<b>Ajout de contraintes</b>					
$\Delta\nu$	65	286	61.4	76.8	91.6
$\nu_{\max}$	68	56	75.7	86.1	94.1
$A_{\max}$	69	841	72.2	84.6	92.8
$\Delta\nu - \nu_{\max}$	70	130	68.0	82.0	93.4
$\Delta\nu - A_{\max}$	73	141	67.4	82.1	93.4
$\nu_{\max} - A_{\max}$	75	109	78.1	87.8	95.1
$\Delta\nu - \nu_{\max} - A_{\max}$	74	84	70.9	84.4	94.3
$\log g, \sigma_{\log g}=0.1$ dex	63	2648	55.4	74.2	88.6
$\log g, \sigma_{\log g}=0.05$ dex	66	1537	65.7	80.5	91.2
$m^{1/3}/R$	99	583	64.2	79.8	92.1
<b>(<math>l=0^\circ, b=0^\circ</math>), <math>\sigma_{[M/H]} = 0.1</math> dex</b>					
100 pc	8	3610	74.3	85.9	95.0
500 pc	95	2051	66.1	78.2	90.5
1 000 pc	89	2447	67.8	79.0	89.6
2 500 pc	202	492	26.9	44.2	70.9
<b>(<math>l=0^\circ, b=0^\circ</math>), <math>\sigma_{[M/H]} = 0.2</math> dex</b>					
100 pc	6	4610	57.1	74.0	91.9
500 pc	63	3045	47.3	65.6	85.3
1 000 pc	67	3131	49.7	68.5	85.7
2 500 pc	63	653	24.5	40.3	68.9
<b>(<math>l=0^\circ, b=0^\circ</math>), <math>\sigma_{[M/H]} = 0.3</math> dex</b>					

TABLE A.1 – Suite.

Calcul	Nombre d'étoiles avec $\min(\chi^2) > \chi_{lim}^2$		Pourcentage d'étoiles avec	
	$\Delta\tau/\tau < 10\%$	$\Delta\tau/\tau < 20\%$	$\Delta\tau/\tau < 10\%$	$\Delta\tau/\tau < 20\%$
100 pc	4	5874	46.9	66.2
500 pc	41	4138	34.0	52.5
1 000 pc	43	4113	35.2	54.3
2 500 pc	94	582	23.2	37.5
<b>(<math>l=50^\circ, b=0^\circ</math>), <math>\sigma_{[M/H]} = 0.1</math> dex</b>				
100 pc	12	3599	73.3	84.8
500 pc	81	2002	66.6	78.6
1 000 pc	71	2366	68.8	79.3
2 500 pc	180	656	39.4	57.3
<b>(<math>l=50^\circ, b=0^\circ</math>), <math>\sigma_{[M/H]} = 0.2</math> dex</b>				
100 pc	11	4655	57.4	73.8
500 pc	45	3027	46.3	65.5
1 000 pc	50	3111	50.3	70.2
2 500 pc	122	813	35.2	51.6
<b>(<math>l=50^\circ, b=0^\circ</math>), <math>\sigma_{[M/H]} = 0.3</math> dex</b>				
100 pc	11	5851	46.7	65.1
500 pc	31	4226	34.2	52.9
1 000 pc	32	4115	35.1	54.9
2 500 pc	74	882	31.8	46.8
<b>(<math>l=0^\circ, b=15^\circ</math>), <math>\sigma_{[M/H]} = 0.1</math> dex</b>				
100 pc	13	3608	74.7	85.7
500 pc	99	1981	66.4	78.3

TABLE A.1 – Suite.

Calcul	Nombre d'étoiles avec		Pourcentage d'étoiles avec		
	$\min(\chi^2) > \chi_{lim}^2$	fonctions mal-définies	$\Delta\tau/\tau < 10\%$	$\Delta\tau/\tau < 20\%$	$\Delta\tau/\tau < 50\%$
1 000 pc	92	2217	68.5	79.5	90.4
2 500 pc	130	382	60.3	73.0	90.1
5 000 pc	271	29	28.8	45.7	72.3
<b>(<math>l=0^\circ, b=15^\circ</math>), <math>\sigma_{[M/H]} = 0.2</math> dex</b>					
100 pc	11	4639	57.2	73.3	91.5
500 pc	70	3048	47.2	66.2	86.1
1 000 pc	62	3028	48.5	63.3	86.1
2 500 pc	89	859	53.4	72.0	89.4
5 000 pc	193	41	25.2	39.6	68.6
<b>(<math>l=0^\circ, b=15^\circ</math>), <math>\sigma_{[M/H]} = 0.3</math> dex</b>					
100 pc	10	5807	47.2	65.7	88.5
500 pc	50	4218	35.6	54.3	79.6
1 000 pc	63	1893	40.2	64.5	88.3
2 500 pc	94	582	23.2	37.5	66.2
5 000 pc	140	71	24.0	37.3	66.4

TABLE A.2 – Table des résultats de tous les calculs réalisés pour la détermination des masses.

Calcul	Nombre d'étoiles avec		Pourcentage d'étoiles avec		
	$\min(\chi^2) > \chi_{lim}^2$	fonction mal-définies	$\Delta m/m < 10\%$	$\Delta m/m < 20\%$	$\Delta m/m < 50\%$
<b>Inférence bayésienne</b>	67	0	97.2	98.8	99.8
<b>Comparaisons des priors</b>					
Avec IMF de K02	67	0	74.8	77.7	90.4
<b>Comparaisons des méthodes</b>					
$\chi_{ST}^2$	68	0	98.1	99.0	99.8
Masse conjointe	67	0	95.9	98.0	99.8
<b>(l=0°, b=0°), <math>\sigma_{[M/H]} = 0.2</math> dex</b>					
100 pc	6	0	98.6	99.8	100
500 pc	63	0	97.0	98.6	99.8
1 000 pc	67	0	97.2	98.8	99.8
2 500 pc	148	0	52.4	73.3	95.5
<b>(l=50°, b=0°), <math>\sigma_{[M/H]} = 0.2</math> dex</b>					
100 pc	11	0	98.7	99.8	100
500 pc	43	0	97.4	98.9	99.9
1 000 pc	50	0	97.3	98.7	99.7
2 500 pc	122	0	66.7	81.2	96.4
<b>(l=0°, b=15°), <math>\sigma_{[M/H]} = 0.2</math> dex</b>					
100 pc	11	0	98.6	99.8	100
500 pc	70	0	97.3	98.9	99.8
1 000 pc	63	0	97.0	98.7	99.7
2 500 pc	89	0	92.1	96.1	99.0
5 000 pc	193	0	54.7	74.4	93.3



TABLE A.3 – Table des résultats de tous les calculs réalisés pour la détermination des métallicités.

Calcul	Nombre d'étoiles avec		Pourcentage d'étoiles avec		
	$\min(\chi^2) > \chi_{lim}^2$	fonction mal-définies	$\Delta m/m < 10\%$	$\Delta m/m < 20\%$	$\Delta m/m < 50\%$
<b>Inférence bayésienne</b>	67	0	57.2	63.4	76.0
$\chi_{ST}^2$	68	0	71.5	82.3	90.7
Métallicité conjointe	67	0	63.6	71.7	83.3



# Interpolation des trajets évolutifs et des isochrones

---

Nous travaillons avec les trajets évolutifs et les isochrones de BaSTI (Pietrinferni et al. 2004), les trajets évolutifs de CESAM (Morel & Lebreton 2008) et les isochrones de Padoue (Bertelli et al. 1994; Girardi et al. 1996, 2000). Les grilles de ces différents modèles d'évolution stellaire sont disponibles sur internet mais elles ne sont pas assez fines pour déterminer l'âge. Nous devons donc effectuer une interpolation pour obtenir la résolution de la grille que l'on souhaite (table 2.1).

Dans un premier temps, nous décrivons la méthode avec les modèles d'évolution stellaire de BaSTI qui sont les plus simples d'utilisation. Pour effectuer l'interpolation des trajets évolutifs ou des isochrones, nous utilisons les Points Particuliers Équivalents (PPEs). Ce sont des points caractéristiques repérables dans tous les modèles d'évolution stellaire (Prather 1976). Les trajets évolutifs de BaSTI contiennent le même nombre de lignes pour chaque trajet évolutif ou isochrone : 2250 pour les trajets évolutifs et 2000 pour les isochrones. Les PPEs sont situés au même numéro de ligne pour tous les trajets évolutifs et les isochrones. Le tableau B.1 décrit les phases d'évolution stellaire et la position des 17 PPEs, ils sont également représentés sur la figure B.1.a pour trois masses différentes.

A partir des PPEs, nous pouvons effectuer facilement une interpolation entre deux isochrones ou deux trajets évolutifs. Par contre il faut garder le même nombre de points entre chaque PPE pour avoir des isochrones ou des trajets évolutifs homogènes. Pour les trajets évolutifs nous effectuons d'abord une interpolation linéaire en log de la masse pour chaque métallicité  $[M/H]$  afin d'obtenir la résolution en masse que l'on souhaite. Ensuite nous effectuons une interpolation linéaire entre les métallicités. Pour les isochrones nous procédons de la même façon mais nous faisons d'abord une interpolation en âge pour toutes les métallicités puis nous faisons l'interpolation entre les métallicités. Pour les trajets évolutifs de CESAM nous utilisons la même méthode car chaque trajet évolutif contient 81 points qui correspondent à des PPEs.

Pour les isochrones de Padoue c'est un peu plus compliqué car les fichiers ne contiennent pas le même nombre de lignes. Les différents PPEs sont donnés

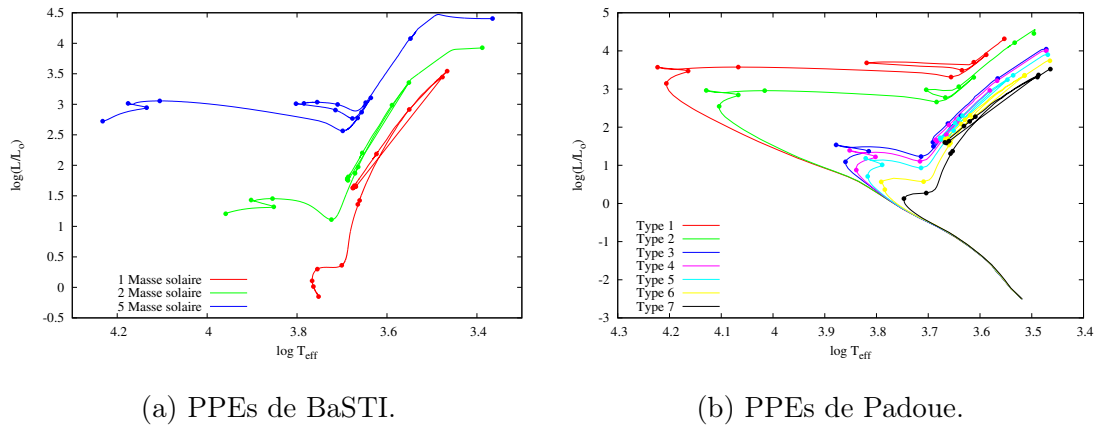
## 142 Annexe B. Interpolation des trajets évolutifs et des isochrones

TABLE B.1 – Tableau décrivant les points particuliers équivalents (PPEs) des modèles d'évolution stellaire de BaSTI (Pietrinferni et al. 2004). Avec TO qui est le Turn-Off, RGB (Red Giant Branch) qui est la branche des géantes rouges et AGB (Asymptotic Giant Branch) qui est la branche asymptotique des géantes.  $X_c$  et  $Y_c$  sont les concentrations d'hydrogène et d'hélium au cœur de l'étoile.

PPE	Ligne	Phase d'évolution
1	1	Démarrage de la combustion de l'hydrogène central
2	200	Premier minimum de $\log T_{\text{eff}}$ pour les grandes masses avec $X_c = 0.30$ pour les petites masses
3	260	Maximum de $\log T_{\text{eff}}$ sur la séquence principale. TO
4	320	Maximum de $\log(L/L_{\odot})$ pour les grandes masses avec $X_c = 0.0$ pour les petites masses
5	390	Minimum de $\log(L/L_{\odot})$ pour les grandes masses ou début de la RGB pour les petites masses
6	760	Maximum de $\log(L/L_{\odot})$ sur le "bump" de la RGB
7	790	Minimum de $\log(L/L_{\odot})$ sur le "bump" de la RGB
8	1190	Fin de la RGB
9	1200	Début de la combustion de l'hélium central
10	1350	$Y_c = 0.55$
11	1450	$Y_c = 0.50$
12	1550	$Y_c = 0.40$
13	1630	$Y_c = 0.20$
14	1710	$Y_c = 0.10$
15	1850	$Y_c = 0.00$
16	2000	L'énergie produite par le cycle CNO est plus grande que celle produite par la combustion de l'hélium sur l'AGB
17	2250	Fin de la combustion de l'hélium central sur l'AGB (trajets évolutifs seulement)

dans un autre fichier et ils ne correspondent pas à la même ligne dans toutes les isochrones. De plus, le nombre de PPEs varie selon le type de l'isochrone, les PPEs sont décrits dans le tableau B.2. Il y a sept types d'isochrones qui dépendent de l'âge et de la métallicité, elles sont représentées sur la figure B.1.b. Pour faire l'interpolation des isochrones nous devons d'abord rechercher tous les PPEs dans chaque isochrone et nous générons des PPEs intermédiaires s'ils n'en existent pas afin d'obtenir 17 PPEs dans chaque isochrone. Ensuite

pour l'interpolation nous utilisons la même méthode que pour les isochrones de BaSTI sauf qu'il faut d'abord effectuer un interpolation linéaire entre les masses pour obtenir des isochrones avec le même nombre de points.



(a) PPEs de BaSTI.

(b) PPEs de Padoue.

FIGURE B.1 – Position des points particuliers équivalents dans le diagramme HR, pour trois trajectes évolutifs de BaSTI (figure a) de masses différentes et de métallicité solaire. Les sept types d'isochrones de Padoue (Bertelli et al. 1994; Girardi et al. 1996) sont représentés sur la figure b.

## 144 Annexe B. Interpolation des trajets évolutifs et des isochrones

TABLE B.2 – Tableau décrivant les points particuliers équivalents (PPEs) de Padoue (Bertelli et al. 1994; Girardi et al. 1996) selon le type d’isochrone. Où SGB (Sub Giant Branch) est la branche des sous géantes et RGB (Red Giant Branch) est la branche des géantes rouges.

PPE	Phase d’évolution	Type d’isochrone						
		1	2	3	4	5	6	7
TO	Turn-Off	1	1	1	1	1	1	1
Te-m	Minimum de $\log T_{\text{eff}}$ sur la SGB	2	2	2	2	2		
Te-M	Maximum de $\log T_{\text{eff}}$ sur la SGB	3	3	3	3	3	2	
L-M	Maximum de $\log(L/L_{\odot})$ sur la SGB	4	4					
RGBb	Début de la RGB	5	5	4	4	4	3	2
L-M	Maximum de $\log(L/L_{\odot})$ sur la RGB				5	5	4	3
L-m	Minimum de $\log(L/L_{\odot})$ sur la RGB				6	6	5	4
RGBt	Fin de la RGB	6	6	5	7	7	6	5
BHeb	Début de la combustion de l’hélium central	7	7	6	8	8	7	6
Te-m	Minimum de $\log T_{\text{eff}}$				9	9	8	7
Te-M	Maximum de $\log T_{\text{eff}}$	8	8	7	10	10	9	8
EHeb	Fin de la combustion de l’hélium central	9	9	8	11	11	10	9
L-M	Maximum de $\log(L/L_{\odot})$					12	11	10
L-m	Minimum de $\log(L/L_{\odot})$					13	12	11
1TP	Début de la AGB		10	9	12	14	13	12
AGBt	Fin de la AGB		11	10	13	15	14	13
Cb	Début de la combustion du carbone central	10						

# Compléments mathématiques

---

## Sommaire

---

C.1	Méthode de Simpson . . . . .	145
C.2	Méthode de Box-Muller . . . . .	145
C.3	Fonction de répartition de la loi normale . . . . .	146
C.4	Table de la distribution du $\chi^2$ . . . . .	148

---

## C.1 Méthode de Simpson

La méthode de Simpson permet de calculer numériquement l'intégrale d'une fonction  $f(x)$  entre deux points  $a$  et  $b$  (figure C.1). L'intégrale est définie par :

$$\int_a^b f(x) = \frac{b-a}{6} \left[ f(a) + 4f\left(\frac{a+b}{2}\right) + f(b) \right] \quad (\text{C.1})$$

## C.2 Méthode de Box-Muller

La méthode de Box-Muller permet de générer un couple  $(X_1$  et  $X_2)$  de variables aléatoires indépendantes selon une loi normale  $\mathcal{N}(\mu, \sigma)$  centrée et réduite. Soient  $u_1$  et  $u_2$  deux variables aléatoires, indépendantes et uniformément distribuées entre 0 et 1.

$$T_1 = \sqrt{-2 \ln u_1} \cos(2\pi u_2), \quad (\text{C.2})$$

$$\text{et } T_2 = \sqrt{-2 \ln u_1} \sin(2\pi u_2). \quad (\text{C.3})$$

$X_1$  et  $X_2$  sont définis par :

$$X_1 = \mu + \sigma T_1, \quad (\text{C.4})$$

$$\text{et } X_2 = \mu + \sigma T_2. \quad (\text{C.5})$$

Avec  $\mu$  la moyenne et  $\sigma$  l'écart type de la loi normale  $\mathcal{N}(\mu, \sigma)$ .

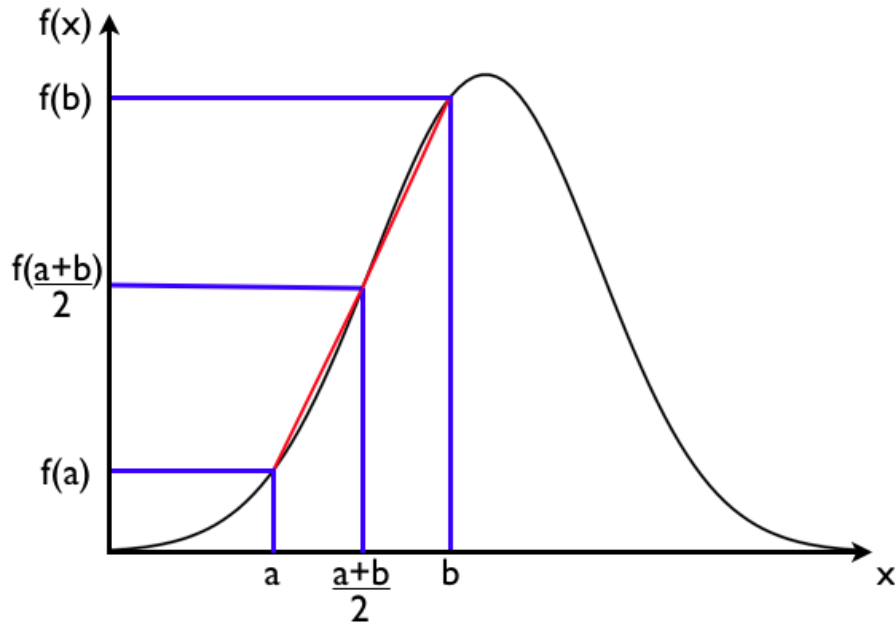


FIGURE C.1 – Figure représentant les paramètres utilisés pour le calcul de l'intégrale avec la méthode de Simpson.

### C.3 Fonction de répartition de la loi normale

Le fonction de répartition permet de caractériser la loi de probabilité d'une variable aléatoire réelle, c'est à dire que cette fonction permet de décrire le comportement de la variable. Pour une loi normale la fonction de répartition s'écrit :

$$\phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x \exp(-t^2/2) dt. \quad (\text{C.6})$$

Elle donne la probabilité qu'une variable aléatoire de loi normale appartienne à l'intervalle  $P(X \in [a, b]) = \phi(b) - \phi(a)$ , où  $X$  suit une loi normale  $\mathcal{N}(0, 1)$ . Pour la calculer on utilise la table C.1 qui donne les valeurs de la fonction de répartition pour  $\phi(x) = [X \leq x]$ . La figure C.2 représente la fonction de répartition d'une loi normale pour  $x = 1.25$ .

Sur la table C.1, les valeurs de la première colonne donnent la première partie de la variable et les valeurs de la première ligne donnent la deuxième partie. Par exemple pour avoir la fonction de répartition de  $x=1.25$  (C.2), on prend la case de la 14ème ligne et de la 7ème colonne, on a donc  $\phi(1.25) = 0.3944$



TABLE C.1 – Table de la fonction de répartition de la loi normale  $\phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x \exp(-t^2/2) dt$ .

$\phi(x)$	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.0000	0.0040	0.0080	0.0120	0.0160	0.0199	0.0239	0.0279	0.0319	0.0359
0.1	0.0398	0.0438	0.0478	0.0517	0.0557	0.0596	0.0636	0.0675	0.0714	0.0753
0.2	0.0793	0.0832	0.0871	0.0910	0.0948	0.0987	0.1026	0.1064	0.1103	0.1141
0.3	0.1179	0.1217	0.1255	0.1293	0.1331	0.1368	0.1406	0.1443	0.1480	0.1517
0.4	0.1554	0.1591	0.1628	0.1664	0.1700	0.1736	0.1772	0.1808	0.1844	0.1879
0.5	0.1915	0.1950	0.1985	0.2019	0.2054	0.2088	0.2123	0.2157	0.2190	0.2224
0.6	0.2257	0.2291	0.2324	0.2357	0.2389	0.2422	0.2454	0.2486	0.2517	0.2549
0.7	0.2580	0.2611	0.2642	0.2673	0.2704	0.2734	0.2764	0.2794	0.2823	0.2852
0.8	0.2881	0.2910	0.2939	0.2967	0.2995	0.3023	0.3051	0.3078	0.3106	0.3133
0.9	0.3159	0.3186	0.3212	0.3238	0.3264	0.3289	0.3315	0.3340	0.3365	0.3389
1.0	0.3413	0.3438	0.3461	0.3485	0.3508	0.3531	0.3554	0.3577	0.3599	0.3621
1.1	0.3643	0.3665	0.3686	0.3708	0.3729	0.3749	0.3770	0.3790	0.3810	0.3830
1.2	0.3849	0.3869	0.3888	0.3907	0.3925	0.3944	0.3962	0.3980	0.3997	0.4015
1.3	0.4032	0.4049	0.4066	0.4082	0.4099	0.4115	0.4131	0.4147	0.4162	0.4177
1.4	0.4192	0.4207	0.4222	0.4236	0.4251	0.4265	0.4279	0.4292	0.4306	0.4319
1.5	0.4332	0.4345	0.4357	0.4370	0.4382	0.4394	0.4406	0.4418	0.4429	0.4441
1.6	0.4452	0.4463	0.4474	0.4484	0.4495	0.4505	0.4515	0.4525	0.4535	0.4545
1.7	0.4554	0.4564	0.4573	0.4582	0.4591	0.4599	0.4608	0.4616	0.4625	0.4633
1.8	0.4641	0.4649	0.4656	0.4664	0.4671	0.4678	0.4686	0.4693	0.4699	0.4706
1.9	0.4713	0.4719	0.4726	0.4732	0.4738	0.4744	0.4750	0.4756	0.4761	0.4767
2.0	0.4772	0.4778	0.4783	0.4788	0.4793	0.4798	0.4803	0.4808	0.4812	0.4817
2.1	0.4821	0.4826	0.4830	0.4834	0.4838	0.4842	0.4846	0.4850	0.4854	0.4857
2.2	0.4861	0.4864	0.4868	0.4871	0.4875	0.4878	0.4881	0.4884	0.4887	0.4890
2.3	0.4893	0.4896	0.4898	0.4901	0.4904	0.4906	0.4909	0.4911	0.4913	0.4916
2.4	0.4918	0.4920	0.4922	0.4925	0.4927	0.4929	0.4931	0.4932	0.4934	0.4936
2.5	0.4938	0.4940	0.4941	0.4943	0.4945	0.4946	0.4948	0.4949	0.4951	0.4952
2.6	0.4953	0.4955	0.4956	0.4957	0.4959	0.4960	0.4961	0.4962	0.4963	0.4964
2.7	0.4965	0.4966	0.4967	0.4968	0.4969	0.4970	0.4971	0.4972	0.4973	0.4974
2.8	0.4974	0.4975	0.4976	0.4977	0.4977	0.4978	0.4979	0.4979	0.4980	0.4981
2.9	0.4981	0.4982	0.4982	0.4983	0.4984	0.4984	0.4985	0.4985	0.4986	0.4986
3.0	0.4987	0.4987	0.4987	0.4988	0.4988	0.4989	0.4989	0.4989	0.4990	0.4990

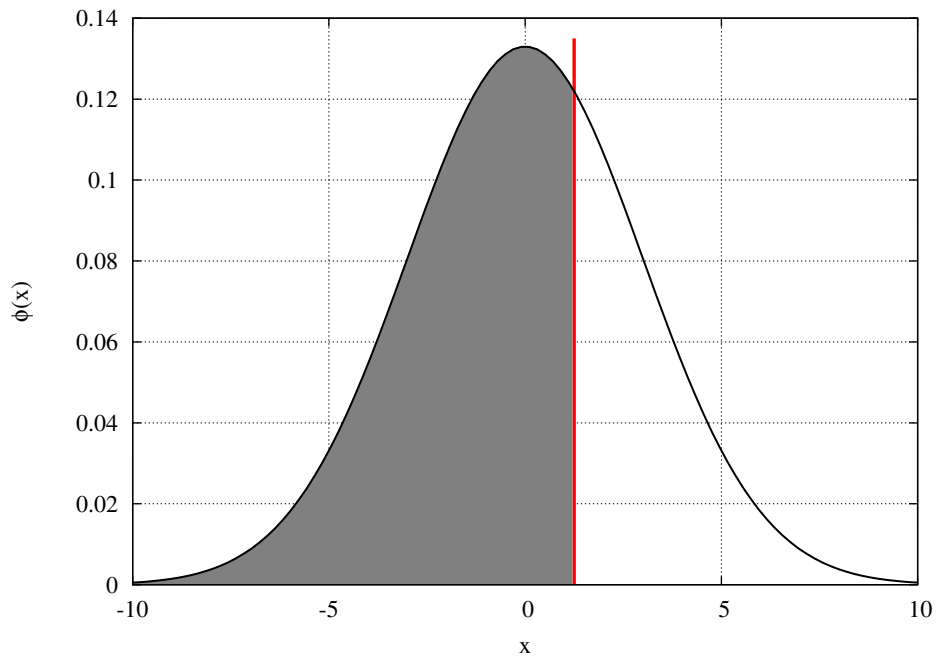


FIGURE C.2 – Courbe représentant la loi normale avec une fonction de répartition (gris) de la loi normale pour  $\phi(x) = [X \leq x]$ , où  $x = 1.25$ .

#### C.4 Table de la distribution du $\chi^2$

TABLE C.2 – Table de la distribution du  $\chi^2$  qui donne le seuil du  $\chi^2$  pour le degré de liberté (dl) et pour différents niveaux de confiance.

dl	0.995	0.975	0.20	0.10	0.05	0.025	0.02	0.01	0.005	0.002	0.001
1	0.0000393	0.000982	1.642	2.706	3.841	5.024	5.412	6.635	7.879	9.550	10.828
2	0.0100	0.0506	3.219	4.605	5.991	7.378	7.824	9.210	10.597	12.429	13.816
3	0.0717	0.216	4.642	6.251	7.815	9.348	9.837	11.345	12.838	14.796	16.266
4	0.207	0.484	5.989	7.779	9.488	11.143	11.668	13.277	14.860	16.924	18.467
5	0.412	0.831	7.289	9.236	11.070	12.833	13.388	15.086	16.750	18.907	20.515
6	0.676	1.237	8.558	10.645	12.592	14.449	15.033	16.812	18.548	20.791	22.458
7	0.989	1.690	9.803	12.017	14.067	16.013	16.622	18.475	20.278	22.601	24.322
8	1.344	2.180	11.030	13.362	15.507	17.535	18.168	20.090	21.955	24.352	26.124
9	1.735	2.700	12.242	14.684	16.919	19.023	19.679	21.666	23.589	26.056	27.877
10	2.156	3.247	13.442	15.987	18.307	20.483	21.161	23.209	25.188	27.722	29.588
11	2.603	3.816	14.631	17.275	19.675	21.920	22.618	24.725	26.757	29.354	31.264
12	3.074	4.404	15.812	18.549	21.026	23.337	24.054	26.217	28.300	30.957	32.909
13	3.565	5.009	16.985	19.812	22.362	24.736	25.472	27.688	29.819	32.535	34.528
14	4.075	5.629	18.151	21.064	23.685	26.119	26.873	29.141	31.319	34.091	36.123
15	4.601	6.262	19.311	22.307	24.996	27.488	28.259	30.578	32.801	35.628	37.697
16	5.142	6.908	20.465	23.542	26.296	28.845	29.633	32.000	34.267	37.146	39.252
17	5.697	7.564	21.615	24.769	27.587	30.191	30.995	33.409	35.718	38.648	40.790
18	6.265	8.231	22.760	25.989	28.869	31.526	32.346	34.805	37.156	40.136	42.312
19	6.844	8.907	23.900	27.204	30.144	32.852	33.687	36.191	38.582	41.610	43.820
20	7.434	9.591	25.038	28.412	31.410	34.170	35.020	37.566	39.997	43.072	45.315
21	8.034	10.283	26.171	29.615	32.671	35.479	36.343	38.932	41.401	44.522	46.797
22	8.643	10.982	27.301	30.813	33.924	36.781	37.659	40.289	42.796	45.962	48.268
23	9.260	11.689	28.429	32.007	35.172	38.076	38.968	41.638	44.181	47.391	49.728
24	9.886	12.401	29.553	33.196	36.415	39.364	40.270	42.980	45.559	48.812	51.179
25	10.520	13.120	30.675	34.382	37.652	40.646	41.566	44.314	46.928	50.223	52.620



## ANNEXE D

# Page web

---

Nous avons développé une page web permettant de déterminer l'âge des étoiles avec l'inférence bayésienne présentée dans la partie 2 et la minimisation standard du  $\chi^2$  décrite dans la partie 3.2.2. Les âges sont déterminés à partir des trajets évolutifs de BaSTI. Pour déterminer l'âge d'une étoile, il faut fournir la température effective  $\log T_{\text{eff}}$ , la métallicité  $[M/H]$  et la magnitude absolue  $M_V$  ou la magnitude apparente  $V$  et la parallaxe  $\pi$ . Pour chacun de ces paramètres, l'erreur observationnelle doit être donnée.

Cette page a été développée en html avec du java script par Malek Mekkas et Sylvestre Taburet. Un script PERL permet de récupérer les paramètres sur la page web et de les transférer au programme de datation que nous avons développé en Fortran. C'est le script PERL qui exécute le programme de détermination des âges et qui transmet les résultats à la page web.

La page web, qui est hébergée par le GEPI, est à l'adresse suivante : <http://stellarage.obspm.fr/index.html>.



**Program for the age-dating of stars.** Possible choice of methods ([Guédé Lebreton & al. in prep](#)) Bayesian estimation or  $\chi^2$ -minimization.

**Inputs:** effective temperature  $\log T_{\text{eff}}$ , absolute magnitude  $M_v$  (or apparent magnitude  $V$  and parallax  $\pi$ ) and metallicity  $[M/H]$ .

**Methods:**

- Bayesian estimation

- based on [Jørgensen and Lindegren \(2005\)](#) and [da Silva, Girardi & al \(2006\)](#)
- flat metallicity distribution function
- initial mass function of [Kroupa \(2002\)](#)
- flat stellar formation rate between 0 and 14 Gyr and 0 elsewhere
- BaSTI [Pietrinferni, Cassisi & al. 2004](#) stellar evolutionary tracks with enhanced resolution
- errors bars determined with a credible interval at 68 %.

-  $\chi^2$ -minimization.

- based on a standard  $\chi^2$
- BaSTI [Pietrinferni, Cassisi & al. 2004](#) stellar evolutionary tracks with enhanced resolution
- errors bars determined according to the propagation of uncertainty.

Enter  $\log T_{\text{eff}}$ ,  $[M/H]$  and either  $M_v$  or  $V$  and  $\pi$  :

log $T_{\text{eff}}$ =	<input style="width: 50px;" type="text"/>	K	, $\sigma_{\log T_{\text{eff}}} =$	<input style="width: 50px;" type="text"/>	K
<input type="radio"/> $M_v =$	<input style="width: 50px;" type="text"/>	mag	, $\sigma_{M_v} =$	<input style="width: 50px;" type="text"/>	mag
<input checked="" type="radio"/> $V =$	<input style="width: 50px;" type="text"/>	mag	, $\sigma_V =$	<input style="width: 50px;" type="text"/>	mag
$\Pi =$	<input style="width: 50px;" type="text"/>	mas	, $\sigma_{\Pi} =$	<input style="width: 50px;" type="text"/>	mas
$[M/H] =$	<input style="width: 50px;" type="text"/>	dex	, $\sigma_{[M/H]} =$	<input style="width: 50px;" type="text"/>	dex

**References:**

- [da Silva, L., Girardi, L., Pasquini, L., et al. 2006, A&A, 458, 609](#)  
[Guédé C., Lebreton, Y & al in prep](#)  
[Jørgensen and Lindegren 2005, A&A, 436, 127](#)  
[Kroupa, P. 2002, Science, 295, 82](#)  
[Pietrinferni, A., Cassisi, S., Salaris, M., & Castellì, F. 2004, ApJ, 612, 168](#)



gaia

contact address: [Yveline Lebreton](#) or [Céline Guédé](#)  
 webmaster : [Malck Mekkas](#)



FIGURE D.1 – Aperçu de notre page web qui permet de déterminer l'âge des étoiles

ANNEXE E

# Publications

---

Article soumis à *Astronomy and Astrophysics*.

# Age-dating stars in the context of the Gaia mission

## I. Evaluation and comparison of the methods

C. Guédé,<sup>1</sup> Y. Lebreton<sup>1,2</sup>, C. Babusiaux<sup>1</sup>, M. Haywood<sup>1</sup>, D. Pelat<sup>3</sup>

<sup>1</sup> Observatoire de Paris, GEPI, CNRS UMR 8111, 92195, Meudon, France

<sup>2</sup> Institut de Physique de Rennes, Université de Rennes 1, CNRS UMR 6251, 35042, Rennes, France

<sup>3</sup> Observatoire de Paris, LUTH, CNRS UMR 8102, 92195, Meudon, France ,  
e-mail: celine.guede@obspm.fr

### ABSTRACT

**Context.** The Gaia ESA astrometric, photometric and spectroscopic mission to be launched this October will survey one billion stars in the Galaxy and will revolutionize our vision of the formation and evolution of the Galaxy. However, to take full benefit of Gaia data, it will be crucial to determine the age of the observed stars with optimal accuracy.

**Aims.** We aimed at evaluating and quantifying the issues and uncertainties that are related to the stellar age determination process. We focused on so-called isochrone placement methods that use evolutionary stellar models to age-date stars. Such methods will be well suited to age-date the large stellar samples expected from Gaia.

**Methods.** To infer the age and mass of a given star, we developed a programme based on a Bayesian approach. Age and mass are inferred by comparing observational data (position in the Hertzsprung–Russell diagram and metallicity) with stellar models predictions. We built simulated Gaia catalogues to test the ability of the programme to return the correct ages and masses of stars. We examined the impact of the priors and inputs on the age-dating process (initial mass function, metallicity distribution function, stellar formation rate, stellar models grid). We implemented other age-dating methods described in the literature (Bayesian approaches and  $\chi^2$  minimization) for comparison.

**Results.** We found that for a simulated catalogue of 10,000 stars observed by Gaia at 1 kpc, the Bayesian approach allows to retrieve the age of 70 per cents of stars with an accuracy better than 20 per cents for  $M_v < 5$ , and 30 per cents for  $M_v > 5$ . The age-dating results are impacted by the choice of the IMF -which should not be too different from the actual one and of the MDF -a flat MDF is preferred- but not much by the choice of the evolutionary models. On the other hand, the  $\chi^2$  minimization method is quite sensitive to the choice of the stellar models. The mass is well retrieved for 95 per cents of the stars in the Bayesian approach. This method is made available via an interactive web page (<http://stellarage.obspm.fr/index.html>).

**Key words.** Stars: fundamental parameters - Methods: statistical - (Stars:) Hertzsprung-Russell and C-M diagrams

## 1. Introduction

Stellar ages are crucial input data in many astrophysical studies. For instance, the knowledge of the stellar formation rate and age-metallicity relation (Gilmore 1999; Hernandez et al. 2000) are important inputs to understand the formation and evolution of our Galaxy (Freeman 1993). Also, it is fundamental to have precise ages of the oldest galactic stars, located in the halo, to set a lower limit to the age of the Universe (Watson 1998). Moreover, the huge harvest of newly discovered exoplanets now calls for accurate and precise determination of the ages of exoplanetary host-stars, a crucial parameter in the characterization of planetary interiors (Havel et al. 2011). Furthermore, we are now close to the launch of the Gaia ESA mission – scheduled October 2013 – which will observe one billion stars belonging to all the populations of the Galaxy. In this context, it is fundamental to design and provide tools able to estimate stellar ages for large stellar samples.

Soderblom (2010) reviewed the various methods that can be used to determine the age of a star. Fundamental and semi-fundamental methods such as nucleocosmochronometry (Cowan et al. 1991) or the study of the kinematics (e.g. the age velocity dispersion relation, Blaauw 1964) rely on observations and well-known physical processes. Empirical methods are based on relations linking age and observational parameters : pe-

riod (gyrochronology, Barnes 2009; Irwin & Bouvier 2009; Mamajek 2009), activity (Skumanich 1972), surface lithium depletion (Soderblom et al. 1993). Statistical methods consider broad galactic trends: age-metallicity relation (Feltzing & Bensby 2009) or dynamical processes (Wielen 1977). Finally, stellar evolution models allow to infer stellar ages. The isochrone placement method has been used for long (Edvardsson et al. 1993) and is very powerful to age-date large samples of stars. Very recently, with the advent of high precision photometric missions CoRoT (Baglin et al. 2002) and Kepler (Koch et al. 2010), asteroseismology has proved to be able to provide other independent and very powerful tools to probe stellar interiors and infer their evolutionary state and age (see e.g. Oti Floranes et al. 2005; Mazumdar 2005; Lebreton & Montalbán 2009).

Gaia will perform astrometric, photometric and spectroscopic observations which will provide stellar global parameters for large samples of stars located everywhere in our Galaxy (disc, bulge, halo). In this context, we are interested in evaluating and refining the isochrone placement method, well-suited to age-date such large samples of stars. The method consists in estimating the age (and mass) of an observed star with measured effective temperature, absolute magnitude and metallicity by looking for a theoretical model isochrone that fits the observed stellar parameters (see e.g. Edvardsson et al. 1993; Ng & Bertelli



1998, hereafter NB). The adjustment can be performed in different ways. NB (1998) selected the appropriate isochrone by a  $\chi^2$ -minimization (i.e. by searching the minimum distance between the star and the isochrones in the parameter space). However, the selection is difficult in some regions of the Hertzsprung-Russell diagram (hereafter HRD) where the isochrones have a complex shape. In such regions, the evolutionary state of the star cannot be determined unambiguously and isochrones (of different ages) can equally fit the stellar position. To choose the most probable age Pont & Eyer (2004) proposed a Bayesian approach which permits to add prior information on the stellar and galactic properties. In the literature, there are several papers that use Bayesian estimation methods to date the stars, as Pont & Eyer (2004), Jørgensen & Lindegren (2005), da Silva et al. (2006), Takeda et al. (2007), Hernandez & Valls-Gabaud (2008) and Casagrande et al. (2011).

In the present study we aim at refining Bayesian age-dating and mass determination and at carefully estimating the associated sources of errors with the objective to provide robust tools to interpret future Gaia observations. We have implemented a Bayesian method including several priors (initial mass function, stellar formation rate and metallicity distribution function). The method is inspired from previous works by da Silva et al. (2006, hereafter dSGP(2006)) and Jørgensen & Lindegren (2005, hereafter JL(2005)). As for the stellar models, we have chosen to use the BaSTI models (Pietrinferni et al. 2004) in the form of evolutionary tracks rather than isochrones. We built a Gaia simulated catalogue and evaluated the consistency between the age inferred from HRD inversion and the simulated “true” age. The catalogue is built from the BaSTI evolutionary tracks and Gaia mission specifications. This catalogue is built at constant distance of 1 kpc. According to Robin et al. (2012) at this distance they are  $10^6$  stars, we take only 10,000 stars, it is enough to realize a statistical study on the results.

We carefully studied the effects of the choice of the priors, of the stellar models grid resolution and of the observational errors on the metallicity. For purpose of comparison and evaluation, we implemented the Bayesian methods of JL (2005), dSGP (2006) and Casagrande et al. (2011, hereafter CSA(2011)) with the same priors and same stellar models as well as two  $\chi^2$ -methods (a standard  $\chi^2$ -approach and the  $\chi^2$ -approach by NB (1998)).

In Sec. 2 we present the tools we used for age-dating: the Bayesian inversion method and its priors, the BaSTI stellar models and the adopted method to estimate the error on age. The age-dating results and the effects of the observational errors on the metallicity are presented in Sect. 3. These studies are based on our Gaia simulated catalog described in Appendix A. In Sect. 4 we describe other methods of age-dating and compare them. The determination of the mass is detailed in Sect. 5. Section 6 is devoted to discussion and conclusions.

## 2. Age-dating method

### 2.1. Bayesian method adopted

The Bayesian inference gives the posterior probability density function (PDF) of the vector of parameters  $\theta$  taking into account the vector of the observed data  $x$ . The posterior PDF is written:

$$\psi(\theta | x) = \frac{L(x | \theta)\pi(\theta)}{\psi(x)}, \quad (1)$$

where  $\pi(\theta)$  is the prior PDF,  $L(x | \theta)$  is the likelihood and  $\psi(x)$  is the marginal likelihood. In this work we neglected the latter

which is a normalization of the posterior PDF (see e.g. Robert 2007). In the case of age-dating the vector of parameters corresponds to the parameters that define the stellar evolutionary tracks, i.e. the metallicity  $[M/H]$  (logarithm of the relative number ratio of metals to hydrogen), the mass  $\mathfrak{M}$  and the age  $\tau$ . As for the observables, in the initial approach by Pont & Eyer (2004) three observables were considered: the metallicity which is therefore both an observable and a model parameter, the absolute magnitude  $M_V$ , the effective temperature  $\log T_{\text{eff}}$  and their error bars. In the notations, we hereafter distinguish the observed metallicity  $[\widetilde{M}/H]$  from the model one  $[M/H]$ . In the Bayesian approach of age determination, the posterior PDF then reads:

$$\psi([M/H], \mathfrak{M}, \tau | M_V, \log T_{\text{eff}}, [\widetilde{M}/H]) \propto L(M_V, \log T_{\text{eff}}, [\widetilde{M}/H] | [M/H], \mathfrak{M}, \tau) \times \pi([M/H], \mathfrak{M}, \tau), \quad (2)$$

where  $\pi([M/H], \mathfrak{M}, \tau) = \xi(\mathfrak{M}) \times \lambda(\tau) \times \phi([M/H])$  where  $\xi(\mathfrak{M})$  is the initial mass function (IMF),  $\lambda(\tau)$  is the stellar formation rate (SFR) and  $\phi([M/H])$  is the metallicity distribution function (MDF). The prior PDF depends of three marginal density because they are statistically independant. The likelihood  $L(M_V, \log T_{\text{eff}}, [\widetilde{M}/H] | [M/H], \mathfrak{M}, \tau)$  for a Gaussian distribution of  $M_V, \log T_{\text{eff}}, [\widetilde{M}/H]$  reads:

$$L(M_V, \log T_{\text{eff}}, [\widetilde{M}/H] | [M/H], \mathfrak{M}, \tau) = \prod_{i=1}^n \left( \frac{1}{\sqrt{2\pi}\sigma_i} \right) \exp\left(-\frac{\chi^2}{2}\right), \quad (3)$$

$$\text{with } \chi^2 = \sum_{i=1}^n \left( \frac{x_i^{\text{mod}}([M/H], \mathfrak{M}, \tau) - \widetilde{x}_i}{\sigma_i} \right)^2, \quad (4)$$

where  $\widetilde{x}_i$  is the  $i^{\text{th}}$  observed parameter of the star,  $x_i^{\text{mod}}([M/H], \mathfrak{M}, \tau)$  is the value of the corresponding  $i^{\text{th}}$  parameter on the isochrones or evolutionary tracks and  $\sigma_i$  is the observational error on the  $i^{\text{th}}$  parameter.

The posterior PDF  $\psi([M/H], \mathfrak{M}, \tau | M_V, \log T_{\text{eff}}, [\widetilde{M}/H])$  can be integrated with respect to the metallicity and mass to obtain the posterior PDF -hereafter denoted as  $\psi(\tau | x)$ - that only depends on the age. It can be written

$$\psi(\tau | x) = \lambda(\tau) \int_{[M/H]} \phi([M/H]) \int_{\mathfrak{M}} \xi(\mathfrak{M}) L(M_V, \log T_{\text{eff}}, [\widetilde{M}/H] | [M/H], \mathfrak{M}, \tau) d\mathfrak{M} d[M/H] \quad (5)$$

Following Pont & Eyer (2004), different approaches for the Bayesian age estimation method have been proposed. Here we based our method on the works by JL (2005) and dSGP (2006).

The posterior PDF  $\psi(\tau | x)$  can be written

$$\begin{aligned} \psi(\tau_k | x) &= \sum_i \left[ \int_{[M/H]_{ijk}^j}^{[M/H]_{ijk}^{i(k+1)}} \phi([M/H]_i) d[M/H] \right] \\ &\times \sum_j \left[ \int_{\mathfrak{M}_j^{ijk}}^{\mathfrak{M}_j^{i(k+1)}} \xi(\mathfrak{M}_j) d\mathfrak{M} \right] \\ &\times L(M_V, \log T_{\text{eff}}, [\widetilde{M}/H] | [M/H]_i, \mathfrak{M}_j, \tau_k) \lambda(\tau_k), \end{aligned} \quad (6)$$

The  $ijk$  triple index corresponds to the point on the evolutionary track of metallicity  $i$ , initial mass  $j$  and age  $k$ .

Concerning the prior we have chosen to use the IMF of Kroupa (2002) (see Section 2.2.1), we took a flat truncated SFR

(described in Section 2.2.3), a flat MDF (see Section 2.2.2) and we did not take into account any age-metallicity relation. As for the stellar models, we used the BaSTI models (Pietrinferni et al. 2004) and based the age-dating on the evolutionary tracks rather than on the isochrones (see Sect. 2.3). The choice of the resolution of the grid of evolutionary tracks is described in Sect. 2.3.

We considered a  $3\sigma$ -error bar on metallicity and we evaluated Eq. 6 for  $[M/H]_i$  in the restricted range  $[M/H] \pm 3\sigma_{[M/H]}$ . Also, we followed JL (2005) who adopted a restrictive condition on the  $\chi^2$ -value to reject the stars lying too far outside the zone covered by the evolutionary tracks: we rejected a star when the minimum value of the  $\chi^2$  calculated according to Eq. 4 exceeds a maximum value  $\chi_{\max}^2$  chosen according to the  $\chi^2$ -distribution law at a confidence level of 0.01 (JL (2005)), i.e.  $\chi_{\max}^2 = 11.34$  for three parameters.

The  $\psi(\tau | x)$  function (Eq. 6) gives for each age the probability that it corresponds to the true age. JL (2005) established that to determine the most probable age  $\hat{\tau}$  of the star it was better to use the maximum of the function  $\psi(\hat{\tau} | x)$  rather than the mean or the median. The differences between the true age, the mean, the median and the mode are represented in Fig. 1 for four types of  $\psi(\tau | x)$  functions. In all cases the ages are better determined with the mode. Here we also used the mode of  $\psi(\tau | x)$ . We systematically -and automatically- analysed the shape of the  $\psi(\tau | x)$ -function of each star. We found six main categories of  $\psi(\tau | x)$  functions described that are in Appendix B, some of them lead to an ill-defined age and a rejection of the associated star.

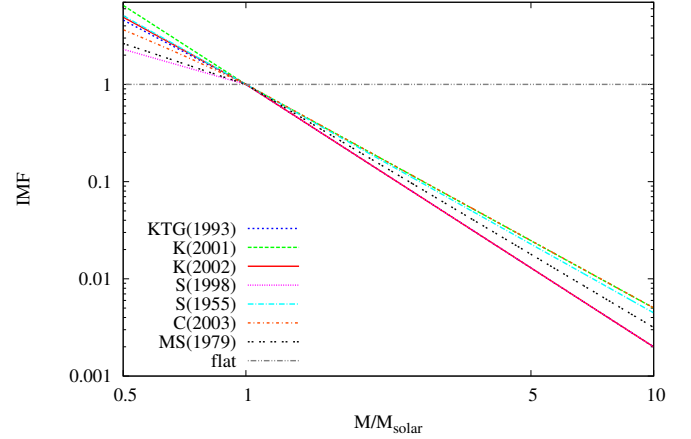
## 2.2. Choice of the priors

In order to evaluate the pertinence and effects of IMF, SFR and MDF priors on the age-dating process, we built a simulated stellar catalogue according to the specifications of Gaia. As described below, the catalogue was constructed with the IMF of Kroupa (2002), a flat truncated SFR and a MDF corresponding to that of the solar neighbourhood. The catalogue properties and building are described in Appendix A.

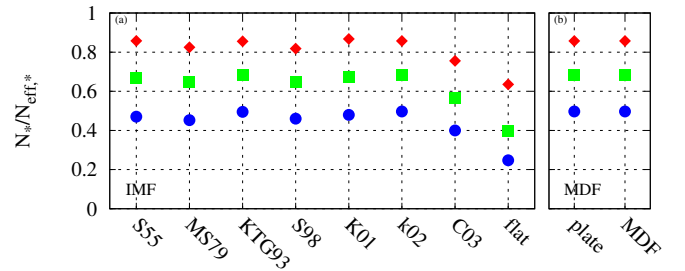
### 2.2.1. Initial mass function

It is interesting to include an IMF prior in the age-dating process to favour the lower mass stars, more frequent in the Galaxy. Figure 2 shows different IMFs available in the literature by: Salpeter (1955, hereafter S(1955)), Miller & Scalo (1979, MS(1979)), Kroupa et al. (1993, KTG(1993)), Scalo (1998, S(1998)), Kroupa (2001, K(2001)), Kroupa (2002, K(2002)) and Chabrier (2003, C(2003)). We considered one of the most recent ones, the K(2002) IMF as our reference and used it to build the Gaia simulated catalogue (Appendix A).

We compared the ages obtained with the K(2002) IMF to the ages obtained with alternative, recent IMFs (S(1998); KTG(1993); C(2003)) and to the ages evaluated with a flat IMF prior (no IMF). We observed that the IMF changes the determined age of the stars located in the degeneracy regions. The relative differences between the inferred age and the simulated “true” age in the catalogue,  $\Delta\tau/\tau = (\tau_{\text{simulated}} - \tau_{\text{inferred}})/\tau_{\text{simulated}}$  are presented in Fig. 3, Panel (a), for different choices of the IMF. The comparison of the ages based on the C(2003) (or S(1998)) IMF with those based on the reference K(2002) IMF clearly illustrates how it would be a problem if the chosen IMF prior was very different from the actual one.



**Fig. 2.** Some IMFs available in the literature. We normalized the IMF to a value of 1. at  $1 M_{\odot}$  to have them all on the same plot.



**Fig. 3.** Simulated age restitution for different priors as labeled in abscissa. In the ratio  $N_*/N_{\text{eff},*}$ ,  $N_{\text{eff},*}$  is the number of stars in the Gaia simulated catalogue minus the simulated stars for which the age cannot be determined (see Sect. 2.1) and  $N_*$  is the number of stars with an age retrieved to 10% or better (blue circle), 20% or better (green square) or 50% or better (red diamond). *Panel a:* Values for several IMF priors or a flat IMF prior. *Panel b:* Values for the MDF derived from the GCS histogram (see Fig. 4) or a flat MDF (no MDF prior) and for an observational error on the metallicity of 0.2 dex.

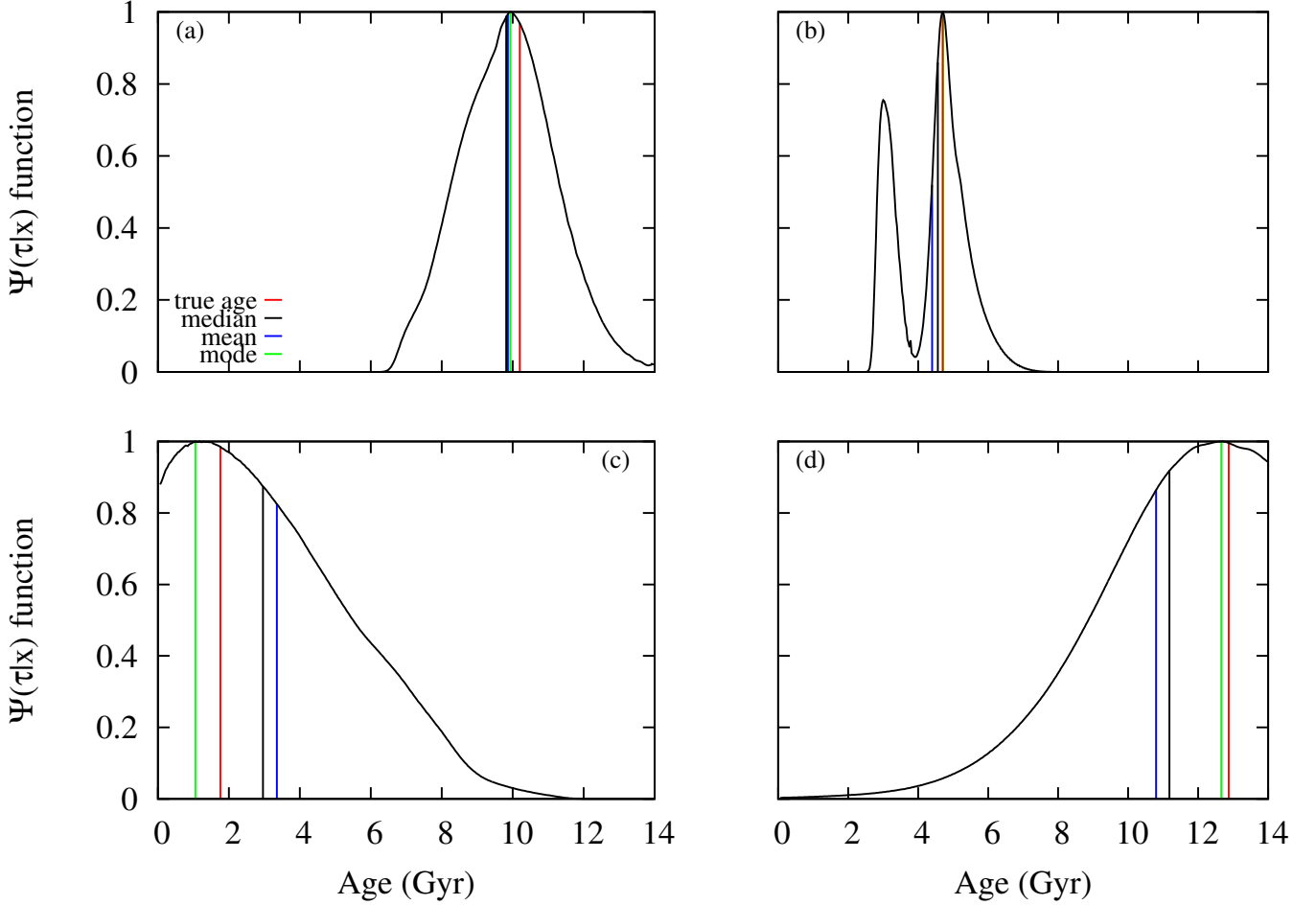
### 2.2.2. Metallicity distribution function

In the galactic disc, the distribution of metallicity of the stars peaks close to the solar metallicity. Introducing a prior on the MDF therefore allows to favour those metallicities. Figure 4 presents the histogram of the metallicities of the Geneva Copenhagen survey (CGS) of 16,682 stars in the solar neighbourhood (CSA but see also Nordström et al. 2004). The distribution evidences two main components, one for the thin disc -peaked at  $[M/H] \sim 0.0$  dex- and one for the thick disc -peaked at  $[M/H] \sim -0.4$  dex-. From the GCS histogram and neglecting the thick disc component which represents a tiny contribution in the solar neighbourhood, we derived the following MDF:

$$\phi([M/H]) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{([M/H] - \mu)^2}{2\sigma^2}\right), \quad (7)$$

with  $\mu = 0.0$  and  $\sigma = 0.15$ .

We determined the ages of the stars in the Gaia simulated catalogue adopting the MDF of Eq. 7 as a prior which we compared with ages estimated without prior (Fig. 3, Panel b) for an observational error on the metallicity of 0.2 dex. We also considered other possible values of the observational error on the metallicity (0.01, 0.1, 0.2, 0.3 and 0.5 dex). As expected, the chosen MDF



**Fig. 1.** Representation of the true age, and the mode, the mean and the median for four types of  $\psi(\tau | x)$  functions (see the morphology of  $\psi(\tau | x)$  functions in Appendix B). In the Panel b the true age and the mode are superimposed.

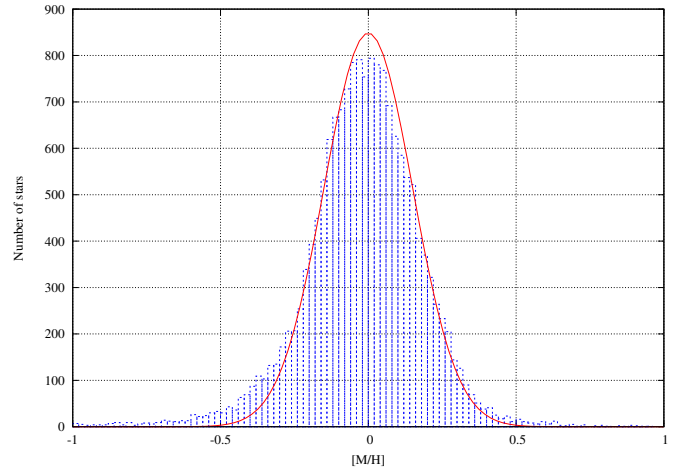
promotes the metallicities at the centre of the distribution but this introduces a bias for the metallicities in the tail of the MDF. The larger the observational error on metallicity, the larger the bias. Indeed we found that the chosen MDF prior has no effect on the age-dating if  $\sigma_{[M/H]} \leq 0.2$  dex while when  $\sigma_{[M/H]} > 0.2$  dex the chosen MDF severely deteriorates the age restitution. We consider that  $[M/H]$  are well-determined so we do not need to include an MDF that have no effect in these case. Under these considerations, in the following we have chosen to include a flat MDF prior (no MDF) in the age-dating process.

### 2.2.3. Stellar formation rate

The ages of the BaSTI stellar evolutionary tracks are in the range 0-22 Gyr. We restricted our working age domain to an upper limit of 14 Gyr, i.e. roughly the age of the Universe (Watson 1998) and used the following flat truncated SFR:

$$\lambda(\tau) = \begin{cases} 1 & \text{for } \tau_1 = 0 \leq \tau \leq \tau_2 = 14 \text{ Gyr} , \\ 0 & \text{elsewhere.} \end{cases} \quad (8)$$

We evaluated the ages of the stars in the simulated Gaia catalogue in the case with and without a SFR prior. Although the age differences are small (lower than 0.15 per cents) we took a flat SFR prior in the following, but truncated at 14 Gyr which corresponds to the physical limit at the age of the Universe.



**Fig. 4.** Histogram of the metallicities for all stars in the GCS catalogue (Casagrande et al. 2011). The MDF (continuous red line) is adjusted to the components of the thin disc only, the components of the thick disc being neglected.

### 2.3. Stellar models choice and handling

We used the stellar models grids of BaSTI as reference (Pietrinferni et al. 2004). The grid covers a mass range 0.75 to  $10 M_{\odot}$ . The heavy elements mass fraction  $Z$  is in the range

**Table 1.** Resolution of the two grids that we obtained by interpolation of the BaSTI evolutionary tracks (Pietrinferni et al. 2004). Grid 2 is the optimal one adopted here to determine the ages. Note that we interpolated on the whole range of parameters covered by the model grids, although some of them are never reached by the studied stars (for instance high ages because of the cut-off at 14 Gyr).

$X^{(1)}$	Range	Grid 1 $\Delta X^{(2)}$	Grid 2 $\Delta X^{(2)}$
[M/H]	-2.26 to 0.39	0.03	0.05
Age (gyr)	0.063 to 22	0.01	0.03
$\mathfrak{M} (M_{\odot})$	0.75 to 1	0.003	0.005
	1.0 to 1.3	0.003	0.003
	1.3 to 10.0	0.003	0.002

**Notes.** <sup>(1)</sup> Parameters. <sup>(2)</sup> Step in the grid between two values of the X parameter

0.0001 to 0.04. The helium abundance in mass fraction follows the relation  $\Delta Y/\Delta Z = 1.45$  with a primordial helium abundance  $Y_p = 0.245$  (Cassisi et al. 2003). The heavy elements mixture is solar with  $(Z/X)_{\odot} = 0.0245$  (Grevesse & Noels 1993). The BaSTI models do not take into account microscopic diffusion neither overshooting of convective cores. The models include mass-loss with two regimes of efficiency:  $\eta = 0.2$  for  $0.75 M_{\odot} \leq \mathfrak{M} \leq 2.6 M_{\odot}$  and  $\eta = 0.4$  for  $2.6 M_{\odot} < \mathfrak{M} \leq 10.0 M_{\odot}$  (Reimers 1975). Each evolutionary track covers the evolution stages from the onset of the central H-burning phase to the end of He-burning on the horizontal branch. We point out that the absolute  $M_V$  magnitude is provided in the BaSTI tracks, via the bolometric corrections of Castelli (1998, 1999) and Bessell et al. (1998).

The original BaSTI grid of evolutionary tracks corresponds to 11 values of the metallicity,  $\sim 34$  values of the mass and 2250 age points. From this grid we built, by interpolation, thinner grids (grids 1 and 2 in Table 1) with higher resolution. The spacings in mass, metallicity and age in grid 1 have been chosen following Takeda et al. (2007). Like Takeda et al. (2007) we chose constant steps in age (see e.g. Maeder 1974). We checked that using constant steps in logarithm of the age ( $\Delta\tau = 0.005$  dex) lead to similar results. In order to be able to treat large stellar samples and save computing time, we investigated different grid resolutions and obtained grid 2 as an optimal choice. From grid 1 to grid 2, we saved half the computing time and obtained ages that do not differ by much than 0.5 per cent.

In the literature, the ages are usually derived from model isochrones provided by modelers. However since the BaSTI evolutionary tracks are available, we decided to use them rather than isochrones ; this saves computing time and reduces the numerical errors generated in the construction of isochrones.

For comparison tests (see Sect. 4.1) we also simulated a Gaia catalogue based on a grid of Cesam models (Morel & Lebreton 2008) computed by one of us (Y. Lebreton) with the input physics described in ?. The grid covers the evolution of stars with masses in the range  $0.8 - 10 M_{\odot}$  from the main sequence to the middle of the red giant branch for metallicities in the interval  $[M/H] \in [-0.20, +0.20]$  dex.

## 2.4. Confidence limits on age

The confidence Bayesian interval is defined as

$$1 - \alpha = \int_{\tau_{\text{low}}}^{\tau_{\text{up}}} \psi(\tau | x) d\tau \quad (9)$$

where  $\tau_{\text{low}}$  and  $\tau_{\text{up}}$  are the boundaries of the confidence interval to be estimated. We determined  $\tau_{\text{low}}$  and  $\tau_{\text{up}}$  for a confidence level  $1 - \alpha = 68\%$  corresponding to a  $1\sigma$  error and to the mode of the  $\psi(\tau | x)$  function.

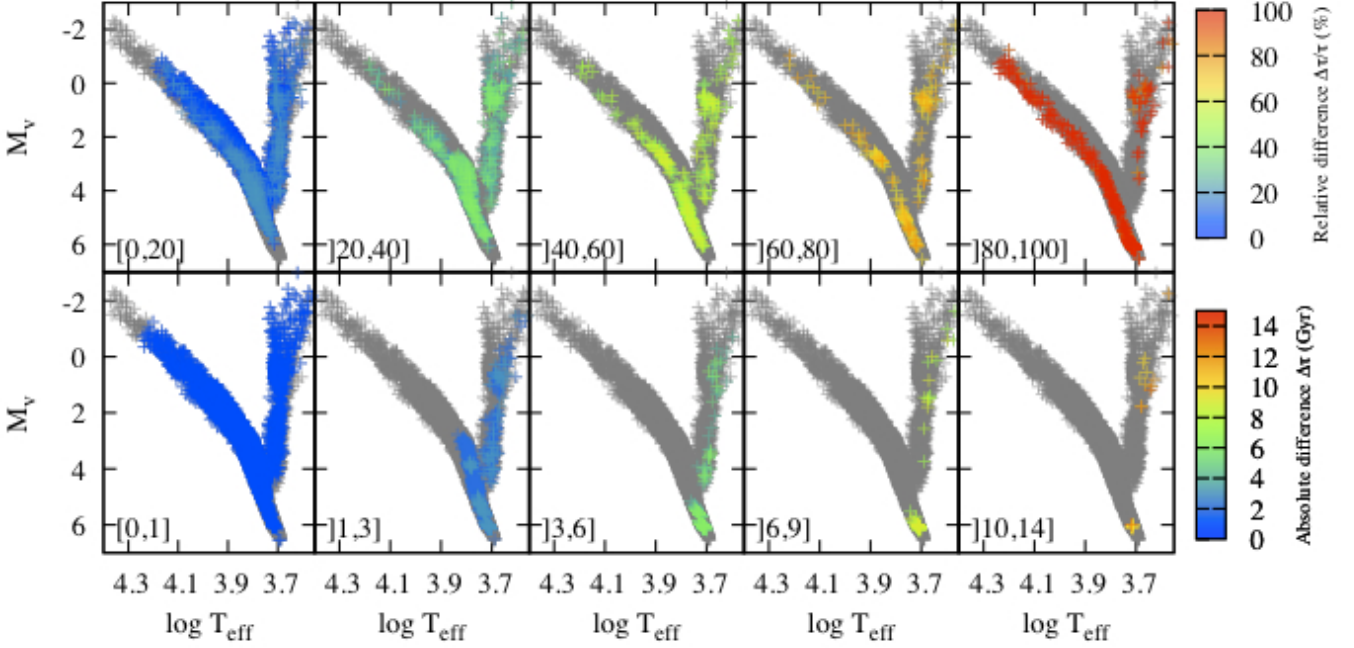
To verify that the confidence interval so obtained is correct we proceeded as follows. We generated -by a Monte Carlo simulation- a sample of stars in the vicinity of a reference star in the catalogue : we simulated 30 stars for a given metallicity and repeated the simulation for 30 values of the metallicity which leads to a sample of 900 stars. Then we determined the age of each star in the sample and verified that more than 68 per cents of the stars lie in the confidence interval of the original reference star. We paved the HRD of the simulated catalogue to select 75 reference stars to which we applied the Monte Carlo simulation. We found that for 62 per cents of these reference stars the confidence interval as well-defined. The remaining stars are divided in two categories, (i) the stars located in the RGB (13 per cents) and (ii) the stars which  $\psi(\tau | x)$  function shows a tiny peak located close to the ZAMS (see the morphology of  $\psi(\tau | x)$  functions in Appendix B). For the latter case the error bar is very small because the age resolution of the grid is too coarse.

## 3. Age-dating results and uncertainties

### 3.1. Ages inferred from HRD parameters and [M/H]

As a case study, we built a simulated catalogue of 10,000 stars, all located at 1 kpc, the error bars on the observational parameters of which being chosen to correspond to the Gaia specifications. The compiling of the catalogue is described in Appendix A. We then estimated the age of each star in the catalogue through Bayesian inversion and compared it to the simulated “true” age (Fig. 5). We obtained the following results. First we could not determine the age of 24 per cents of the 10,000 stars. There are two main reasons for that: (i) a small proportion (about 1 per cent) of the stars in the catalogue do not pass the  $\chi^2$  rejection test discussed in Sect. 2.1 because they are massive ( $\mathfrak{M} > 3.5 M_{\odot}$ ) and lie in the upper MS, (ii) about 23 per cent of the stars have an ill-determined  $\psi(\tau | x)$ -function which we will discuss later. Second, we could estimate the ages of 76 per cents of the 10,000 stars and only these 7,600 stars are discussed now. A total of 65 (respect. 44) per cents of them have a good recovered age with  $\Delta\tau/\tau < 20$  (respect. 10) per cents. Such stars with a well-determined age are preferentially located in the same regions of the HRD: the turn-off (TO), the sub-giant branch (SGB) and the intermediate mass stars domain ( $1.0 M_{\odot} < \mathfrak{M} < 2.0 M_{\odot}$ ) on the main sequence (MS). Finally 35 per cent of the stars have ill-determined ages: they are located in regions of the HRD where there is an age-degeneracy. As it is well-known, two regions of the HRD are mainly concerned: the region close to the zero-age MS (ZAMS) where the lower mass stars ( $\mathfrak{M} < 1.0 M_{\odot}$ ) are concentrated due to their slow evolution speed and the red giant branch (RGB) which stars ascend quite rapidly.

It is worth to be noticed that to each category of  $\psi(\tau | x)$  function described in Appendix B, a specific region of the HRD can be associated, as shown in Fig. 6. The percentages of stars in each category and the corresponding goodness of the age estimation are given in Table 2. Although rapidly evolving stars



**Fig. 5.** Comparison of the simulated “true” age and of the inferred one in the HRD of the simulated catalogue. *Top panel:* colors represent the relative age difference  $\Delta\tau/\tau$ . The stars in blue have a well-determined age while the stars in red have an ill-determined age. *Bottom panel:* colors represent the absolute age difference  $\Delta\tau$ . The stars in blue have a small difference between the simulated “true” age and the inferred one while the stars in red have a huge difference.

on the RGB have functions with an oscillatory behaviour we could infer the age of 57 per cents of them. Functions with multiple peaks are caused by complex morphologies of evolutionary tracks. Those associated to the RGB result from loops in the HRD in the vicinity of the clump while those associated to the MS of stars in the mass range  $\sim 0.9$  to  $\sim 1.4M_{\odot}$  result from the hook at the end of core H-burning occurring for stars that possessed a convective core on the MS (i.e. stars with  $\mathfrak{M} \gg 1.2M_{\odot}$  depending on the metallicity). For 75 per cents of those stars the highest peak corresponds to the “true” age. Only upper limit  $\psi(\tau | x)$  functions are found for 34 per cents of the stars and 4 per cents for the lower limit only. As explained in Appendix B upper (lower) limit only functions can be (i) well-defined with no plateau, (ii) ill-defined with a plateau and (iii) ill-defined, no limit functions. The upper limit only functions at young ages are located close to the ZAMS and those with a plateau have the lower masses. Similarly the lower limit only functions are located at the bottom of the MS.

On this basis, we have chosen to consider the ages of stars for a restricted range of  $\psi(\tau | x)$  functions: the well-defined functions with one or multiple peaks, functions showing oscillations and the upper and lower limit only functions with no plateau. For stars having other types of  $\psi(\tau | x)$  functions, the age-dating is too risky.

We point out that the results and trends presented herebefore are closely related to the present simulated catalogue, based on Gaia expected errors at a distance of 1 kpc. As detailed in the Appendix A, the observational errors on the parallax, on  $M_V$  and  $T_{\text{eff}}$  are functions of the apparent magnitude  $G$  of the Gaia system. Furthermore the ESA-Gaia Science Performance web page at [http://www.rssd.esa.int/index.php?page=Science\\_Performance&project=GAIA](http://www.rssd.esa.int/index.php?page=Science_Performance&project=GAIA) which provides the errors expected at the end of the mission is updated regularly. Since the accuracy on the observed parameters strongly impacts

**Table 2.** Percentage of stars for each category of  $\psi(\tau | x)$  function described in the Appendix B. The percentage of stars with  $\Delta\tau/\tau < 20\%$  is given for each case. We did not include the stars with an undetermined age (no fit) due to their location outside the regions covered by the evolutionary tracks (about 1 percent of the stars).

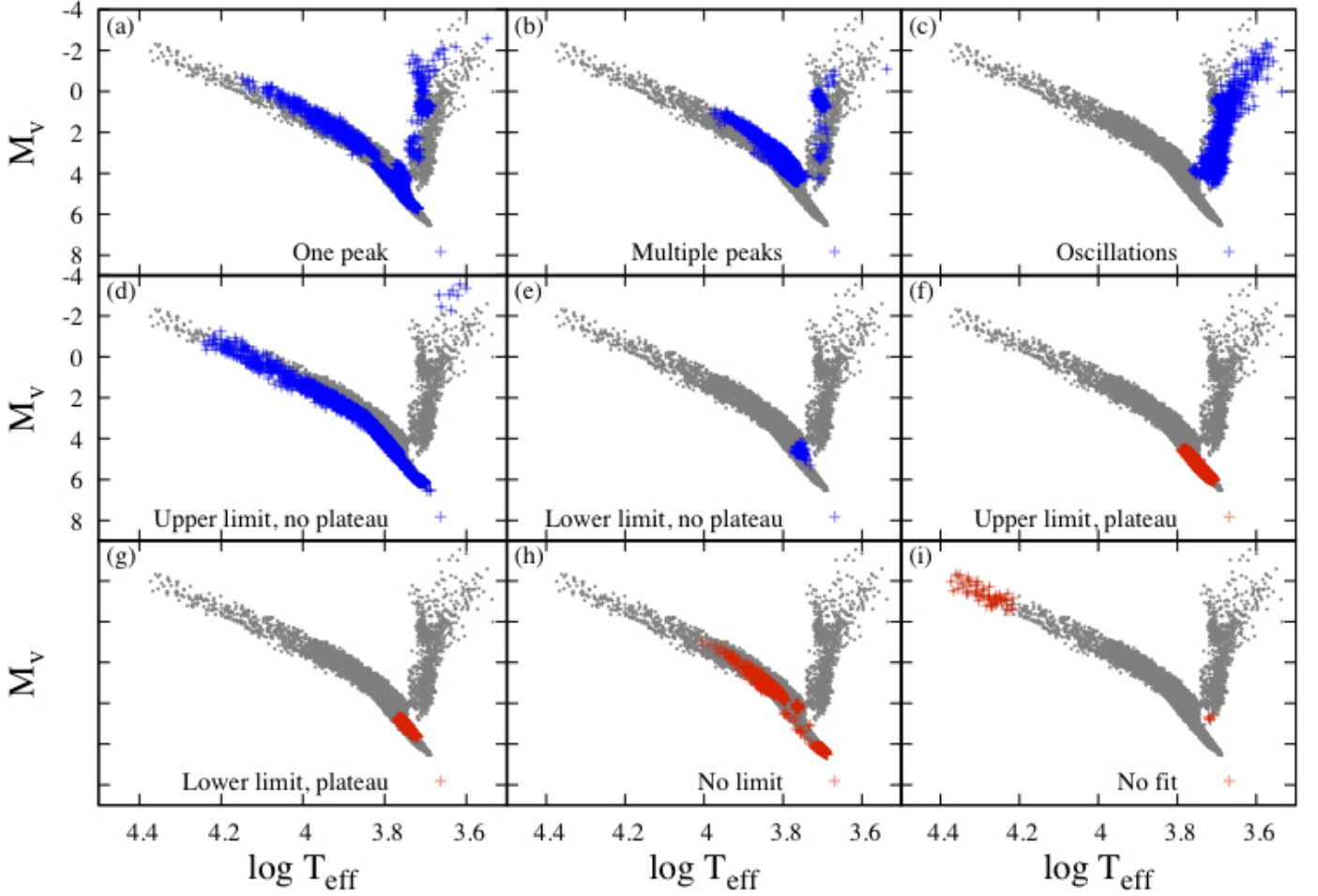
Type of functions	Stars	$\Delta\tau/\tau < 20\%$
One peak	16%	88%
Multiple peaks	18%	84%
Oscillations	8%	53%
Upper limit only, no plateau	26%	47%
Upper limit only, plateau	20%	19%
Lower limit only, no plateau	1%	100%
Lower limit only, plateau	4%	81%
No limit	6%	62%

the age-dating results, we will investigate age-dating at different distances and in different regions of the Galaxy in a forthcoming paper.

### 3.2. Effects of observational errors on the metallicity

To determine the age of a star, we considered all the evolutionary tracks that have a metallicity in the range  $[M/H] \pm 3\sigma_{[M/H]}$ . Therefore, depending on the observational error on the metallicity  $\sigma_{[M/H]}$ , a different number of tracks are taken into account in the age-dating process. To evaluate this effect, we determined and compared the ages of the stars in the simulated catalogue (Appendix A) for three values of  $\sigma_{[M/H]}$  (0.1, 0.2 and 0.3 dex). The comparison is presented in As expected, the smaller the error on the metallicity, the better the age restoration in particular in the regions of age-degeneracy in the HRD. A good accuracy





**Fig. 6.** HRD showing the location of each category of  $\psi(\tau | x)$  functions presented in Appendix B. The stars in the simulated catalogue are represented in grey. The stars in red have an ill-defined functions.

on the age requires to have precise measurements of the metallicity.

#### 4. Comparison of ages from different methods

We implemented a standard  $\chi^2$ -minimization approaches to compare the results with the bayesian estimation. The most probable age  $\hat{\tau}$  is the age of the point of the evolutionary track which gives the minimum of  $\chi^2$ . One advantage of the  $\chi^2$  minimization method is that once the point of the evolutionary track is identified, the mass of the star is inferred simultaneously. We used the evolutionary tracks of BaSTI (Sect. 2.3) and the  $\chi^2$  rejection test (Sect. 2.1).

We calculated the  $\chi^2$  only for points in the evolutionary tracks with  $M_{V,ijk}$  in the range  $\overline{M}_V \pm 3\sigma_{M_V}$ ,  $\log T_{\text{eff},ijk}$  in the range  $\log \overline{T}_{\text{eff}} \pm 3\sigma_{\log T_{\text{eff}}}$  and  $[M/H]_{ijk}$  in the range  $[\overline{M}/H] \pm 3\sigma_{[M/H]}$ :

$$\chi_{\text{ST}}^2 = \frac{(\log \overline{T}_{\text{eff}} - \log T_{\text{eff},ijk})^2}{\sigma_{\log T_{\text{eff}}}^2} + \frac{(\overline{M}_V - M_{V,ijk})^2}{\sigma_{M_V}^2} + \frac{([\overline{M}/H] - [M/H]_{ijk})^2}{\sigma_{[M/H]}^2}. \quad (10)$$

We determined the errors-bars with the propagation of uncertainty.

$$\begin{pmatrix} \sigma_{\log T_{\text{eff}}}^2 & \rho_{12}\sigma_{\log T_{\text{eff}}}\sigma_{\tau} \\ \rho_{21}\sigma_{\log T_{\text{eff}}}\sigma_{\tau} & \sigma_{\tau}^2 \end{pmatrix} = D\Sigma^x D^T, \quad (11)$$

$$\text{with } D = \begin{pmatrix} \frac{\partial \log T_{\text{eff}}}{\partial \tau} & \frac{\partial \log T_{\text{eff}}}{\partial M_V} & \frac{\partial \log T_{\text{eff}}}{\partial [M/H]} \\ \frac{\partial \tau}{\partial \log T_{\text{eff}}} & \frac{\partial \tau}{\partial M_V} & \frac{\partial \tau}{\partial [M/H]} \end{pmatrix}, \quad (12)$$

$$\text{and } \Sigma^x = \begin{pmatrix} \sigma_{\log T_{\text{eff}}}^2 & 0 & 0 \\ 0 & \sigma_{M_V}^2 & 0 \\ 0 & 0 & \sigma_{[M/H]}^2 \end{pmatrix}. \quad (13)$$

Where  $\Sigma^x$  is the covariance-variance matrix of the observational errors and  $D$  the design matrix. To estimate this matrix, we determined the age and mass of the considered star for six cases, according to the extreme observed values:  $\log T_{\text{eff}} \pm \sigma_{\log T_{\text{eff}}}$ ,  $M_V \pm \sigma_{M_V}$  and  $[M/H] \pm \sigma_{[M/H]}$ . We found that 61 per cents of the reference stars of the aforementioned MC simulation (Sec. 2.4) have a well-defined confidence level.

We implemented a  $\chi^2$ -minimizations approaches based on the work by NB (1998). It is only performed for a single value of the metallicity of the model tracks, the one closest to the observed metallicity. The minimization consists in searching the point in the evolutionary tracks which radius is the closest to the observed stellar radius, inferred from the Stefan-Boltzmann law:

$$\chi_{\text{NB(1998)}}^2 = \sqrt{(\overline{M}_V - M_{V,ijk})^2 + 16(\log \overline{T}_{\text{eff}} - \log T_{\text{eff},ijk})^2 / \sigma_{M_V}^2}.$$

NB (1998) neglected the observational error on the effective temperature on the basis that it is small compared to the observational error on the absolute magnitude. They used the same selection on  $M_{V,ijk}$  and  $\log T_{\text{eff},ijk}$  that in the NB (1998) method. The main difference between the NB (1998)  $\chi^2$  approach and the standard  $\chi^2$  method is in the treatment of the metallicity and of the observational error on the metallicity.

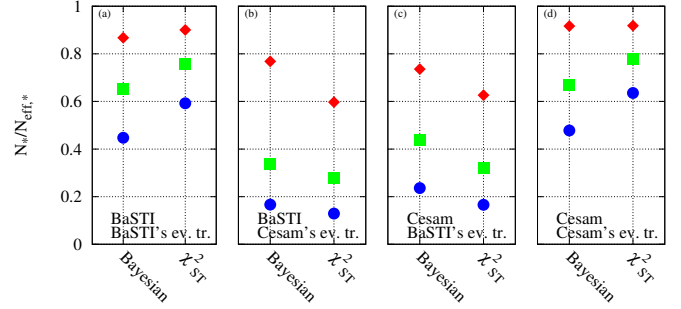
For purpose of comparison, we implemented also the Bayesian estimation methods of JL (2005), dSGP (2006) and CSA (2011) but for the priors,  $\chi^2$  rejection test and stellar evolution tracks chosen here (described in Sect. 2). The errors-bars on the age are evaluated with the credible age interval presented in Sect. 2.4. The main difference between these method and our method are listed below :

- The JL (2005) method use an implementation different of the priors than our method.
- dSGP (2006) used a simplified expression of the likelihood and the expression of the metallicity is different.
- CSA (2011) have an expression of the likelihood different. In their method, they included the parallax error because they want to take into account the gaussian nature of the error on the parallax. They neglected the error on the apparent magnitude because the error on absolute magnitude is dominated by the error on the parallax. We point out that CSA (2011) included a MDF in the age-dating to correct for the bias on the observed metallicity. As discussed in Sect. 2.2.2 we have chosen not to take into account such corrections.

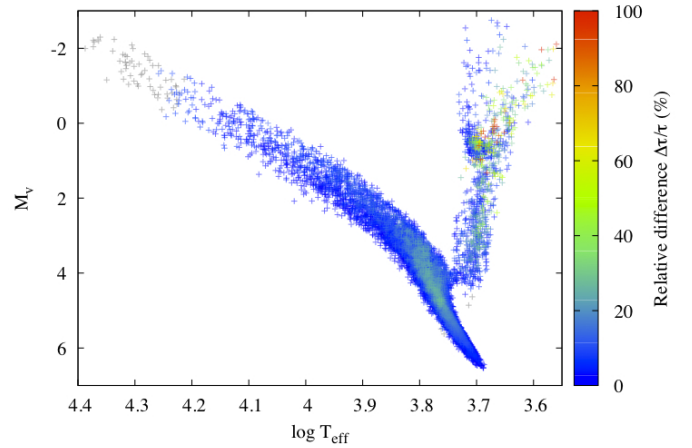
#### 4.1. Comparison of the methods

To compare the different Bayesian approaches we determined the ages of stars in the Gaia simulated catalogue for each method. The comparisons are presented in Fig. 7, Panel a. As expected, we found similar results for the present method and the JL (2005) and dSGP (2006) methods: about 65 per cents of the stars have  $\Delta\tau/\tau < 20$  per cents. On the other hand, the CSA (2011) method gives slightly different results because of a different treatment of the error on the observed parameters. We found that 42 per cents of the stars have  $\Delta\tau/\tau < 20$  per cents. We point out that here we considered a Gaussian error on absolute magnitude (calculated taking into account the errors on parallax, apparent magnitude and extinction). CSA (2011) proceeded differently and considered a Gaussian error on the parallax.

As for the comparison of the Bayesian method with the  $\chi^2$  approach, we only considered the standard  $\chi^2_{ST}$  since the  $\chi^2_{NB}$ -method does not take the error on metallicity into account. We compared the  $\chi^2$ -minimization and the adopted Bayesian estimation for two simulated stellar catalogues, all built as described in Appendix A: the first catalogue is built with the BaSTi evolutionary tracks and the second one with the Cesam evolutionary tracks. The age results are presented in Fig. 8. It is worth to point out that when the ages are determined using the same evolutionary tracks that the tracks used to build the simulated catalogue, the ages are better restored by the  $\chi^2$ -minimization. However if a different model grid is used to simulate the catalogue and to infer the ages, the Bayesian estimation leads to better results, because it is less sensitive to the stellar models. This is a strong point considering the fact that stellar models are still quite far to reproduce perfectly stellar evolution.



**Fig. 8.** Comparison of the ages determined with the presently adopted Bayesian method and with a standard  $\chi^2_{ST}$ -minimization. We considered and coupled two grids of stellar evolutionary tracks -BaSTi and Cesam- which are used either for the catalogue simulation or for the age restoration. The name of the grid used in the catalogue is given in the first line of the figure legend, the name of the grid used for age-dating is given in the second line. Colours are the same as in Fig. 3.



**Fig. 9.** Comparison of the simulated “true” mass and of the inferred one in the HRD of the simulated catalogue. The colour codes are the same as in Fig 5.

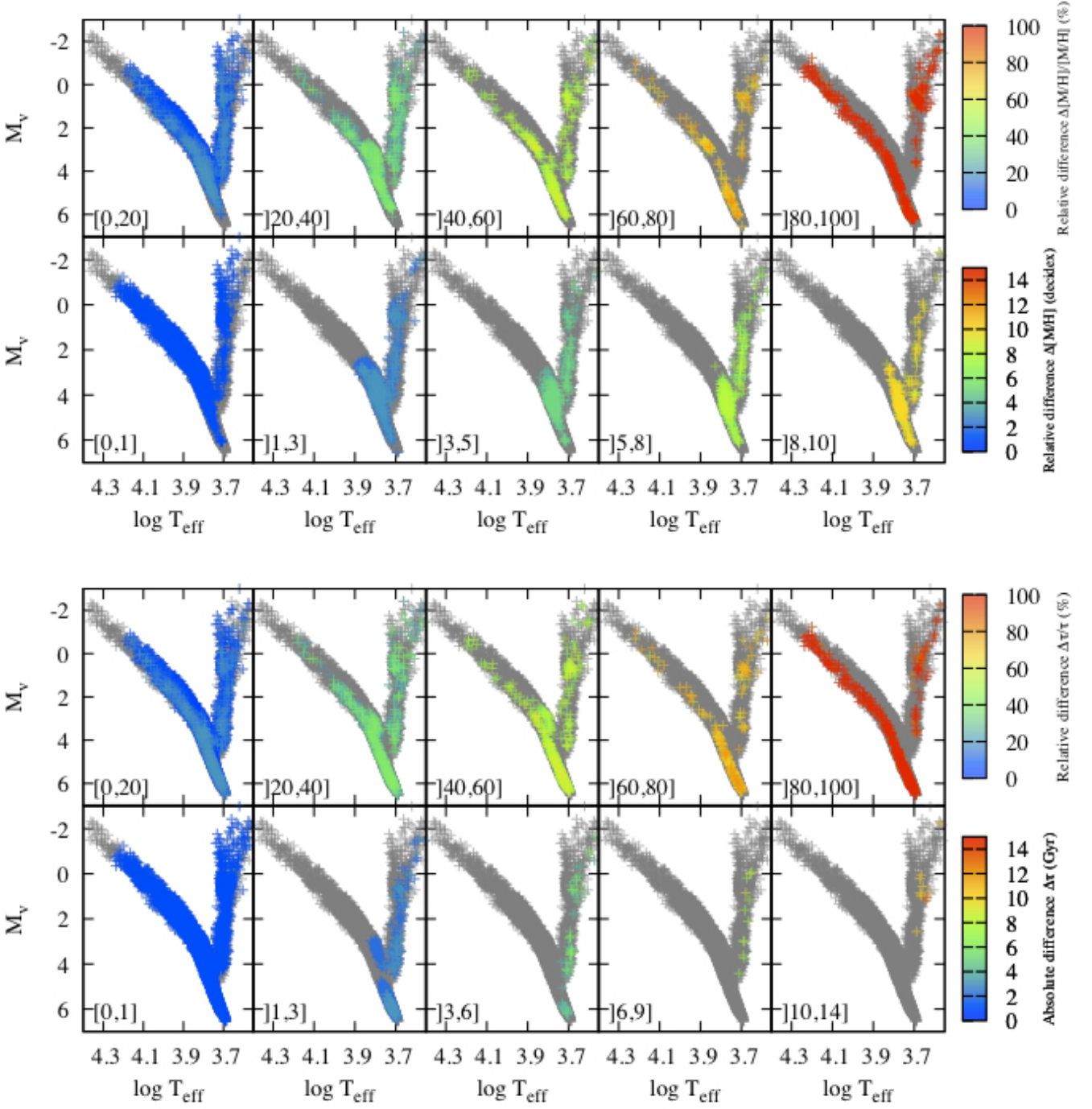
## 5. Application to the determination of the mass

The method described in Sect. 2 can be applied to stellar mass determination with a convenient choice of the priors. We estimated the masses using a flat IMF, not to introduce any bias. As shown in Fig. 9, 95 per cents of the stars have their mass restored to better than  $\Delta M/M < 20$  per cents. Ill-determined masses are mainly located in the RGB. The mass determination is easier than the age determination because masses and ages are not spread the same way in the HRD. In this respect, Prather (1976) have shown that along the HRD the relative variations of age and mass vary almost linearly, on average, as  $dt/t \approx -4dM/M$  (see Figs. 1 and 2 in Appendix b of Prather 1976). Indeed, we found that  $dt/t \approx -3dM/M$  for all metallicities of the Cesam evolutionary tracks.

## 6. Conclusions

In this paper, we used two inversion methods which are based on the isochrone placement: the Bayesian estimation and the  $\chi^2$ -minimization<sup>1</sup>. The ages are determined from three observables ( $T_{\text{eff}}$ ,  $M_V$  and  $[M/H]$ ) and the evolutionary tracks.

<sup>1</sup> <http://stellarafe.obspm.fr/index.html>



**Fig. 7.** Comparison of the simulated “true” age and of the inferred one in the HRD of the simulated catalogue for the JL (2005) method (top) and the  $\chi^2_{S7}$ -minimization (bottom). Colors are the same as in Fig 5.

We compared both methods using two catalogues simulated using the Gaia specifications, the first one based on the evolutionary tracks of BASTI, and the second one using Cesam’s evolutionary tracks.

We determine the ages of stars in the two catalogues with the Bayesian estimation and  $\chi^2$ -minimization for both sets of tracks (BASTI and CESAM). The determination of ages with the Bayesian estimation is two times longer than the determination with the  $\chi^2$ -minimization.

Firstly, the results show that the determinations are degenerated in two regions of the HRD: close to the ZAMS particularly for lower mass stars ( $M < 1.0 M_{\odot}$ ) and the RGB. This is due to the evolutionary speed, too slow close to the ZAMS and too fast on the RGB. The priors do not minimized the degeneracy.

Secondly, we found that the  $\chi^2$ -minimization gives better results than the Bayesian estimation when the ages are determined with the same evolutionary tracks than those used to generate the simulated catalogue. On the contrary, when the evolutionary



tracks are different, the Bayesian estimation is better. Thus, the  $\chi^2$ -minimization is more sensitive to the stellar models.

Reality is more likely to reproduce the former case, because our incomplete knowledge of stellar physics must produce systematic differences between stellar models and data, and the priors in the Bayesian estimation allows to compensate for these differences. Hence, our advice would be that, when possible, Bayesian estimation should be preferred.

Besides, the Bayesian estimation provides more information than  $\chi^2$ -minimization thanks to the  $\psi(\tau | x)$  function. Ill-behaved  $\psi(\tau | x)$  functions indicate an uncertainty on the age of the star. Finally, the treatment of the  $\psi(\tau | x)$  function of the Bayesian estimation allows to find the ill-defined  $\psi(\tau | x)$  functions that do not give a correct age. The  $\psi(\tau | x)$  functions also allow to determine the age of the different structures of the Galaxy or of the open clusters without losing information (see e.g. JL (2005)).

This work is important for the determination of the ages of the forthcoming Gaia mission. However, the derivation of ages assumes priors and evolutionary tracks that are dependent on our present knowledge of stellar and Galactic physics, which both will be improved by the Gaia data. It is important to note that Gaia will provide the individual parallaxes, bolometric corrections and extinction together with their error bars. Therefore the present programme will have to be adapted to the stellar samples to be studied. Moreover, other parameters such as the asteroseismic parameters, the exoplanetary transit parameter ( $\mathfrak{M}^{1/3}/R$ ) and the surface gravity, can be added to help lifting the degeneracies that affect parts of the HRD (Guédé et al. 2012). This will be the object of a forthcoming paper.

*Acknowledgements.* We are very grateful to Malek Mekkas and Sylvestre Taburet for designing the WEB page of the programme. We acknowledge Gaël Dréan for an earlier version of the programme. We are indebted to Frédéric Arenou, Santi Cassisi, Frédéric Thévenin, Bernard Pichon, Christophe Ordenovic for fruitful discussions and advice.

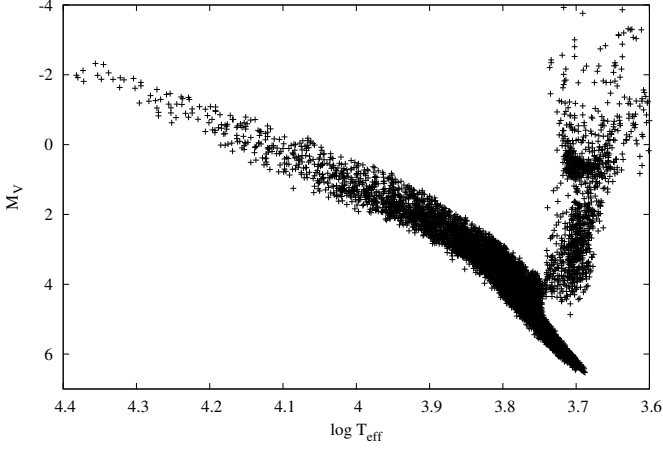
## References

- Baglin, A., Auvergne, M., Barge, P., et al. 2002, in ESA Special Publication, Vol. 485, *Stellar Structure and Habitable Planet Finding*, ed. B. Battrock, F. Favata, I. W. Roxburgh, & D. Galadi, 17–24
- Bailer-Jones, C. A. L. 2010, *MNRAS*, 403, 96
- Barnes, S. A. 2009, in IAU Symposium, Vol. 258, IAU Symposium, ed. E. E. Mamajek, D. R. Soderblom, & R. F. G. Wyse, 345–356
- Bessell, M. S., Castelli, F., & Plez, B. 1998, *A&A*, 333, 231
- Blaauw, A. 1964, *ARA&A*, 2, 213
- Casagrande, L., Schönrich, R., Asplund, M., et al. 2011, *A&A*, 530, A138
- Cassisi, S., Salaris, M., & Irwin, A. W. 2003, *ApJ*, 588, 862
- Castelli, F. 1998, *Mem. Soc. Astron. Italiana*, 69, 165
- Castelli, F. 1999, *A&A*, 346, 564
- Chabrier, G. 2003, *PASP*, 115, 763
- Cowan, J. J., Thielemann, F.-K., & Truran, J. W. 1991, *ARA&A*, 29, 447
- da Silva, L., Girardi, L., Pasquini, L., et al. 2006, *A&A*, 458, 609
- Edvardsson, B., Andersen, J., Gustafsson, B., et al. 1993, *A&A*, 275, 101
- Feltzing, S. & Bensby, T. 2009, in IAU Symposium, Vol. 258, IAU Symposium, ed. E. E. Mamajek, D. R. Soderblom, & R. F. G. Wyse, 23–30
- Freeman, K. C. 1993, in *Astronomical Society of the Pacific Conference Series*, Vol. 49, *Galaxy Evolution. The Milky Way Perspective*, ed. S. R. Majewski, 125
- Gilmore, G. 1999, *Baltic Astronomy*, 8, 203
- Grevesse, N. & Noels, A. 1993, in *Origin and Evolution of the Elements*, ed. N. Prantzos, E. Vangioni-Flam, & M. Casse, 15–25
- Guédé, C., Lebreton, Y., Babusiaux, C., & Haywood, M. 2012, in *SF2A-2012: Proceedings of the Annual meeting of the French Society of Astronomy and Astrophysics*, ed. S. Boissier, P. de Laverny, N. Nardetto, R. Samadi, D. Valls-Gabaud, & H. Wozniak, 195–198
- Havel, M., Guillot, T., Valencia, D., & Crida, A. 2011, *A&A*, 531, A3
- Hernandez, X., Gilmore, G., & Valls-Gabaud, D. 2000, *MNRAS*, 317, 831
- Hernandez, X. & Valls-Gabaud, D. 2008, *MNRAS*, 383, 1603
- Irwin, J. & Bouvier, J. 2009, in IAU Symposium, Vol. 258, IAU Symposium, ed. E. E. Mamajek, D. R. Soderblom, & R. F. G. Wyse, 363–374
- Jordi, C., Gebran, M., Carrasco, J. M., et al. 2010, *A&A*, 523, A48
- Jørgensen, B. R. & Lindegren, L. 2005, *A&A*, 436, 127
- Koch, D., Borucki, W., Jenkins, J., et al. 2010, in 38th COSPAR Scientific Assembly, Vol. 38, 2513
- Kroupa, P. 2001, *MNRAS*, 322, 231
- Kroupa, P. 2002, *Science*, 295, 82
- Kroupa, P., Tout, C. A., & Gilmore, G. 1993, *MNRAS*, 262, 545
- Lebreton, Y. & Montalbán, J. 2009, in IAU Symposium, Vol. 258, IAU Symposium, ed. E. E. Mamajek, D. R. Soderblom, & R. F. G. Wyse, 419–430
- Maeder, A. 1974, *A&A*, 32, 177
- Mamajek, E. E. 2009, in IAU Symposium, Vol. 258, IAU Symposium, ed. E. E. Mamajek, D. R. Soderblom, & R. F. G. Wyse, 375–382
- Mazumdar, A. 2005, *A&A*, 441, 1079
- Miller, G. E. & Scalo, J. M. 1979, *ApJS*, 41, 513
- Morel, P. & Lebreton, Y. 2008, *Ap&SS*, 316, 61
- Ng, Y. K. & Bertelli, G. 1998, *A&A*, 329, 943
- Nordström, B., Mayor, M., Andersen, J., et al. 2004, *A&A*, 418, 989
- Otí Floranes, H., Christensen-Dalsgaard, J., & Thompson, M. J. 2005, *MNRAS*, 356, 671
- Perryman, M. A. C., de Boer, K. S., Gilmore, G., et al. 2001, *A&A*, 369, 339
- Pietrinferni, A., Cassisi, S., Salaris, M., & Castelli, F. 2004, *ApJ*, 612, 168
- Pont, F. & Eyer, L. 2004, *MNRAS*, 351, 487
- Prather, M. J. 1976, PhD thesis, Yale University., New Haven, CT.
- Reimers, D. 1975, *Memoires of the Societe Royale des Sciences de Liege*, 8, 369
- Robert, C. 2007, *The Bayesian choice, From Decision-Theoretic Foundations to Computational Implementation*, 2nd edition (New York: Springer Verlag)
- Robin, A. C., Luri, X., Reylé, C., et al. 2012, *A&A*, 543, A100
- Salpeter, E. E. 1955, *ApJ*, 121, 161
- Scalo, J. 1998, in *Astronomical Society of the Pacific Conference Series*, Vol. 142, *The Stellar Initial Mass Function (38th Herstmonceux Conference)*, ed. G. Gilmore & D. Howell, 201
- Skumanich, A. 1972, *ApJ*, 171, 565
- Soderblom, D. R. 2010, *ARA&A*, 48, 581
- Soderblom, D. R., Stauffer, J. R., Hudon, J. D., & Jones, B. F. 1993, *ApJS*, 85, 315
- Takeda, G., Ford, E. B., Sills, A., et al. 2007, *ApJS*, 168, 297
- Watson, A. 1998, *Science*, 279, 981
- Wielen, R. 1977, *A&A*, 60, 263

## Appendix A: The Gaia simulated catalogue

To test the capability of age-dating methods to return the actual age of a star we simulated a stellar catalogue on the basis of the specifications of the Gaia mission (Perryman et al. 2001). The catalogue contains 10,000 stars all located at a distance of 1 kpc. To build the catalogue we proceeded as follows. Firstly we randomly drew the metallicity of the star according to the MDF in the solar neighbourhood (Eq. 7) and the stellar mass according to the K(2002) IMF. We identified the four BaSTI (Pietrinferni et al. 2004) evolutionary tracks corresponding to these random selections (i.e. for the two metallicities delimiting the selected one, there are two values of the mass delimiting the selected mass). Secondly, we randomly drew an age in these four stellar evolution tracks, with the adopted constraint on the SFR (Eq. 8). We realized several interpolations between the points (delimiting the selected age) of the four evolutionary tracks. This returned the global parameters of the simulated star ( $B$  and  $V$  indexes, effective temperature). We calculated the  $G$ -magnitude of the star using the transformation of Jordi et al. (2010)<sup>4</sup>. We only kept the simulated stars with an apparent  $G$ -magnitude greater than 6 since stars of lower magnitudes are too bright to be observable by Gaia. We estimated the accuracy on the parallax measurement from the law provided on the ESA website<sup>3</sup>, i.e.  $\sigma_{\pi} = (9.3 + 658.1z + 4.568z^2)^{1/2} \times [0.986 + (1 - 0.986)(V - J)]$  where  $z = \text{MAX}[10^{0.4(12-15)}, 10^{0.4(G-15)}]$  and we only retained the stars which accuracy on the parallax is better than 10 per cents.

The HRD of the stars in the simulated catalogue is plotted in Fig. A.1. The simulated stars have  $G$  magnitudes in the range



**Fig. A.1.** HRD of the 10,000 stars in the Gaia simulated catalogue at 1 kpc. The catalogue is obtained assuming a Gaussian MDF, the K(2002) IMF, a flat SFR and BaSTI stellar evolutionary tracks.

6 – 20 mag and their distribution as a function of the  $G$  magnitude is plotted in Fig. A.2. Each parameter is given an observational error corresponding to what is expected at the end of the Gaia mission (see Bailer-Jones 2010, and the ESA website)<sup>3</sup>. As shown in Fig. A.2, the relative accuracy on the effective temperature is 0.3 per cent for  $G \leq 15$  mag and increases linearly to 4 per cents for the dimmest stars at  $G = 20$  mag. The error on the absolute magnitude  $\sigma_{M_V}$  depends on the errors on the apparent magnitude  $\sigma_V$ , the parallax  $\sigma_\pi$  and the extinction  $\sigma_{A_V}$ . Here we assumed that  $\sigma_V = \sigma_G$  and we took  $\sigma_G = 10^{-3}(0.02076z^2 + 2.7224z + 0.004352)^{1/2}$  and  $\sigma_{A_V} = 0.1 \times A_V$  following Bailer-Jones (2010) and the ESA website<sup>3</sup>. It is worth to point out that, as shown in Fig. A.2, the error on  $M_G$  is larger for the brighter stars. We assumed that the metallicity accuracy will be of 0.2 dex.

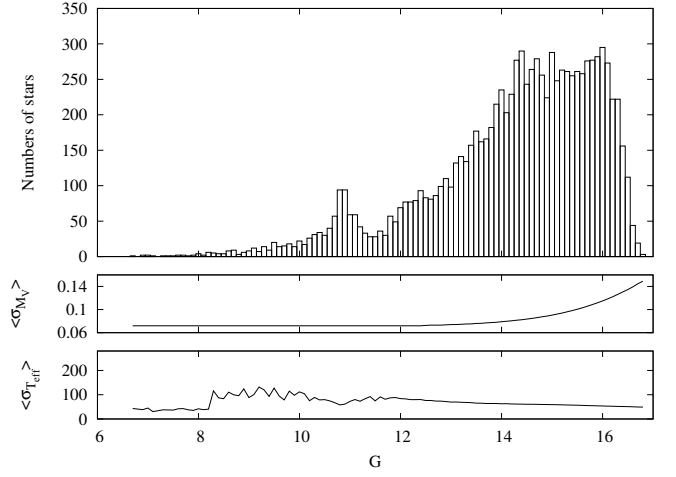
For comparison purposes, we also built a simulated catalogue using the Cesam evolutionary tracks (Morel & Lebreton 2008). The corresponding HRD is plotted on Fig. A.3.

## Appendix B: Treatment of the *a posteriori* probability density function $\psi(\tau | x)$

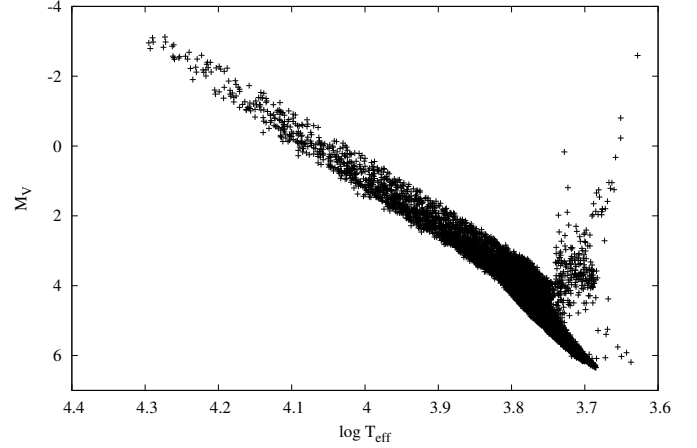
As already described by JL (2005) and by Takeda et al. (2007), the probability density functions  $\psi(\tau | x)$  can have different shapes. We found 9 main categories of  $\psi(\tau | x)$  functions that we plotted in Fig. B.1: (i) one peak, (ii) multiple peaks, (iii) oscillations, (iv) upper limit only with no plateau, (v) upper limit only with plateau, (vi) lower limit only with no plateau, (vii) lower limit only with plateau, (viii) no limit and (viii) no fit. The last case correspond to the star locate outside the regions covered by the evolutionary tracks, well-defined functions with several peaks, (iii) functions showing oscillations, (iv) continuously decreasing functions, (v) truncated functions with a plateau and (vi) truncated functions with no plateau. To properly attribute an age to a given star, we had to identify the morphology of its  $\psi(\tau | x)$  function and smooth it if necessary. We proceeded as follows.

<sup>3</sup> August 2012 version of [http://www.rssd.esa.int/index.php?project=GAIA&page=Science\\_Performance#table2](http://www.rssd.esa.int/index.php?project=GAIA&page=Science_Performance#table2)

<sup>4</sup>  $G = V - 0.0424 - 0.0851(B - V) - 0.3348(B - V)^2 + 0.0205(B - V)^3$



**Fig. A.2.** Top figure: distribution of the Gaia  $G$  magnitude of the 10,000 stars in the Gaia simulated catalogue at 1 kpc. Middle and bottom figures : variation of the error on  $T_{\text{eff}}$  (middle) and  $M_G$  (bottom) as a function of  $G$ .



**Fig. A.3.** HRD of the 10,000 stars in the Gaia simulated catalogue at 1 kpc. The catalogue is obtained assuming a Gaussian MDF, the K(2002) IMF, a flat SFR and Cesam evolutionary tracks.

1. We looked for spurious oscillations in the functions. In the case where a function exhibited more than 3 peaks we considered them as spurious. We smoothed the function a first time by making a convolution with a Gaussian kernel and a second smoothing was applied if more than 3 peaks remained after the first smoothing.
2. The most probable age of a star having  $\psi(\tau | x)$  functions (smoothed or not) with 1 to 3 peaks is given by the age of the highest peak.
3. To detect truncated functions, we measured the half full width at half maximum (FWHM<sub>left</sub> and FWHM<sub>right</sub>) on both sides of the most probable age  $\hat{\tau}$ . We considered the function to be truncated at low ages when  $\left| \frac{\tau_{\text{FWHM, right}} - \hat{\tau}}{\tau_{\text{FWHM, left}} - \hat{\tau}} \right| > 1.5$  and at high ages when  $\left| \frac{\tau_{\text{FWHM, left}} - \hat{\tau}}{\tau_{\text{FWHM, right}} - \hat{\tau}} \right| > 1.5$ .
4. This lead to three kinds of truncated functions:
  - Continuously increasing or decreasing functions. They generally have ill-defined ages (see text).
  - Some functions do not have a well-defined maximum but a large plateau. We decided arbitrarily that when the

# The Gaia astrophysical parameters inference system (Apsis)

## Pre-launch description

C.A.L. Bailer-Jones<sup>1\*</sup>, R. Andrae<sup>1</sup>, B. Arcay<sup>2</sup>, T. Astraatmadja<sup>1</sup>, I. Bellas-Velidis<sup>3</sup>, A. Berihuete<sup>4</sup>, A. Bijaoui<sup>5</sup>, C. Carrión<sup>6</sup>, C. Dafonte<sup>2</sup>, Y. Damerdjil<sup>7,8</sup>, A. Dapergolas<sup>3</sup>, P. de Laverny<sup>5</sup>, L. Delchambre<sup>7</sup>, P. Drazinos<sup>9</sup>, R. Drimmel<sup>10</sup>, Y. Frémat<sup>11</sup>, D. Fustes<sup>2</sup>, M. García-Torres<sup>12</sup>, C. Guédé<sup>13,14</sup>, U. Heiter<sup>15</sup>, A.-M. Janotto<sup>16</sup>, A. Karamelas<sup>9</sup>, D.-W. Kim<sup>1</sup>, J. Knude<sup>17</sup>, I. Kolka<sup>18</sup>, E. Kontizas<sup>3</sup>, M. Kontizas<sup>9</sup>, A.J. Korn<sup>15</sup>, A.C. Lanzafame<sup>19,20</sup>, Y. Lebreton<sup>13,14</sup>, H. Lindstrøm<sup>17,21</sup>, C. Liu<sup>1</sup>, E. Livanou<sup>9</sup>, A. Lobel<sup>11</sup>, M. Manteiga<sup>2</sup>, C. Martayan<sup>22</sup>, Ch. Ordenovic<sup>5</sup>, B. Pichon<sup>5</sup>, A. Recio-Blanco<sup>5</sup>, B. Rocca-Volmerange<sup>23,24</sup>, L.M. Sarro<sup>6</sup>, K. Smith<sup>1</sup>, R. Sordo<sup>25</sup>, C. Soubiran<sup>26</sup>, J. Surdej<sup>7</sup>, F. Thévenin<sup>5</sup>, P. Tsalmandza<sup>1</sup>, A. Vallenari<sup>25</sup>, and J. Zorec<sup>23</sup>

(Affiliations can be found after the references)

submitted 23 July 2013; revised 8 September 2013; accepted 9 September 2013

### ABSTRACT

The Gaia satellite will survey the entire celestial sphere down to 20th magnitude, obtaining astrometry, photometry, and low resolution spectrophotometry on one billion astronomical sources, plus radial velocities for over one hundred million stars. Its main objective is to take a census of the stellar content of our Galaxy, with the goal of revealing its formation and evolution. Gaia's unique feature is the measurement of parallaxes and proper motions with hitherto unparalleled accuracy for many objects. As a survey, the physical properties of most of these objects are unknown. Here we describe the data analysis system put together by the Gaia consortium to classify these objects and to infer their astrophysical properties using the satellite's data. This system covers single stars, (unresolved) binary stars, quasars, and galaxies, all covering a wide parameter space. Multiple methods are used for many types of stars, producing multiple results for the end user according to different models and assumptions. Prior to its application to real Gaia data the accuracy of these methods cannot be assessed definitively. But as an example of the current performance, we can attain internal accuracies (RMS residuals) on F,G,K,M dwarfs and giants at  $G = 15$  ( $V = 15-17$ ) for a wide range of metallicities and interstellar extinctions of around 100 K in effective temperature ( $T_{\text{eff}}$ ), 0.1 mag in extinction ( $A_0$ ), 0.2 dex in metallicity ( $[\text{Fe}/\text{H}]$ ), and 0.25 dex in surface gravity ( $\log g$ ). The accuracy is a strong function of the parameters themselves, varying by a factor of more than two up or down over this parameter range. After its launch in November 2013, Gaia will nominally observe for five years, during which the system we describe will continue to evolve in light of experience with the real data.

**Key words.** galaxies: fundamental parameters – methods: data analysis – methods: statistical – stars: fundamental parameters – surveys: Gaia

## 1. Introduction

The ESA Gaia satellite will provide the most extensive astrometric survey of our Galaxy to date. Its primary mission is to measure the positions, parallaxes, and proper motions for essentially all objects in the sky between visual (G-band) magnitudes 6 and 20, some  $10^9$  stars and several million galaxies and quasars. By revealing the three-dimensional distribution and space motions of a statistically significant sample of stars across the whole Galaxy, Gaia will enable a fundamentally new type of exploration of the structure, formation and evolution of our Galaxy. Furthermore, this exquisite astrometry – parallax uncertainties as low as  $10 \mu\text{as}$  – will promote major advances in our knowledge and understanding of stellar structure, open clusters, binary stars and exoplanets, lead to discoveries of near-earth asteroids and provide tests of general relativity (e.g.

Perryman et al. 2001, Turon et al. 2005, Lindegren et al. 2008, Casertano et al. 2008, Bailer-Jones 2009, Mignard & Klioner 2010, Tanga & Mignard 2012).

To achieve these goals, astrophysical information on the astrometrically measured sources is indispensable. For this reason Gaia is equipped with two low resolution prism spectrophotometers, which together provide the spectral energy distribution of all targets from 330 to 1050 nm. Data from these spectrophotometers (named BP and RP for “blue photometer” and “red photometer”) will be used to classify sources and to determine their astrophysical parameters (APs), such as stellar metallicities, line-of-sight extinctions, and the redshifts of quasars. The spectrophotometry is also required to correct the astrometry for colour-dependent shifts of the image centroids. Spectra from the higher resolution radial velocity spectrograph (RVS, 847–871 nm) on board will provide further information for esti-

\* Email: calj@mpia.de

rating APs as well as some individual abundances for the brighter stars.

Gaia scans the sky continuously, building up data on sources over the course of its five year mission. Its scanning strategy, plus the need for a sophisticated self-calibration of the astrometry, demands an elaborate data processing procedure. It involves numerous interdependent operations on the data, including photometric processing, epoch cross-matching, spectral reconstruction, CCD calibration, attitude modelling, astrometric parameter determination, flux calibration, astrophysical parameter estimation, and variability analysis, to name just a few. These tasks are the responsibility of a large academic consortium, the *Data Processing and Analysis Consortium* (DPAC), comprising over 400 members in 20 countries. The DPAC comprises nine co-ordination units (CUs), each dealing with a different aspect of the data processing.

One of these CUs, CU8 “Astrophysical Parameters”, is responsible for classifying and estimating the astrophysical parameters of the Gaia sources. In this article we describe the data processing system developed to achieve this goal. This system, called *Apsis*, comprises a number of modules, each of which will be described here.

The Gaia data processing is organized into a series of consecutive cycles centered around a versioned main data base (MDB). At the beginning of each operation cycle, the various processes read the data they need from the MDB. At the end of the cycle, the results are written to a new version of the MDB which, together with new data from the satellite, forms the MDB for the next processing cycle. In this way, all of the Gaia data will be sequentially processed until, at some version of the MDB several cycles after the end of observations, all data have received all necessary treatment and the final catalogue can be produced. Some suitably processed and calibrated data will be siphoned off during the processing into intermediate data releases, expected to start about two years after launch (Prusti 2012; DPAC 2012). For more details of the overall processing methodology see DPAC (2007), O’Mullane et al. (2007), and Mignard et al. (2008).

We continue our presentation of the Gaia astrophysical processing system in section 2 by looking more closely at the data Gaia will provide. Section 3 gives an overview of *Apsis*: the guiding concepts behind it, its component modules and how they interact, and how it will be used during the mission. Section 4 describes the data we have used for model training and testing. In section 5 we describe each of the modules and give some impression of the results which can be expected. More details on several of these can be found in published or soon-to-be published articles. In section 6 we outline how we plan to validate and calibrate the system once we get the Gaia data, and how we might improve the algorithms during the mission. We wrap up in section 7. More information on the Gaia mission, the data processing and planned data releases, as well as some of the DPAC technical notes cited, can be obtained from <http://www.rssd.esa.int/Gaia>.

## 2. Gaia observations and data

An overview of the Gaia instruments, their properties and expected performance can be found in de Bruijne (2012) and at <http://tinyurl.com/GaiaPerformance>. Here we

summarize some essential features relevant to our description of *Apsis*.

### 2.1. Overview and observation strategy

Gaia observes continuously, its two telescopes – which share a focal plane – scanning a great circle on the sky as the satellite rotates, once every six hours. The satellite simultaneously precesses with a period of 63 days. The combined result of these motions is that the entire sky is observed after 183 days. Each source is therefore observed a number of times over the course of the mission. These multiple observations, made at different points on Gaia’s orbit around the Sun, are the basis for the astrometric analysis (Lindgren et al. 2012).

As the satellite rotates, a source sweeps across a large focal plane mosaic of CCDs. These are read out synchronously with the source motion (“time-delayed integration”, TDI). Over the first  $0.7^\circ$  of the focal plane scan, the source is observed in unfiltered light – the *G*-band – for the purpose of the astrometry. Further along the light is dispersed by two prisms to produce the BP/RP spectrophotometry. At the trailing edge the light is dispersed by a spectrograph to deliver the RVS spectra.

Although Gaia observes the entire sky, not all CCD pixels are transmitted to the ground. Gaia selects, in real-time, windows around point sources brighter than  $G=20$ . The profile of the *G*-band, spanning 330–1050 nm, is defined by the mirror and CCD response (Jordi et al. 2010).<sup>1</sup> The source detection is near-diffraction limited to about  $0.1''$  (the primary mirrors have dimensions  $1.45\text{m}\times 0.5\text{m}$ ). While most of the  $10^9$  sources we expect Gaia to observe will be stars, a few million will be quasars and galaxies with point-like cores, and asteroids. Robin et al. (2012) give predictions of the number, distribution and types of sources which will be observed.

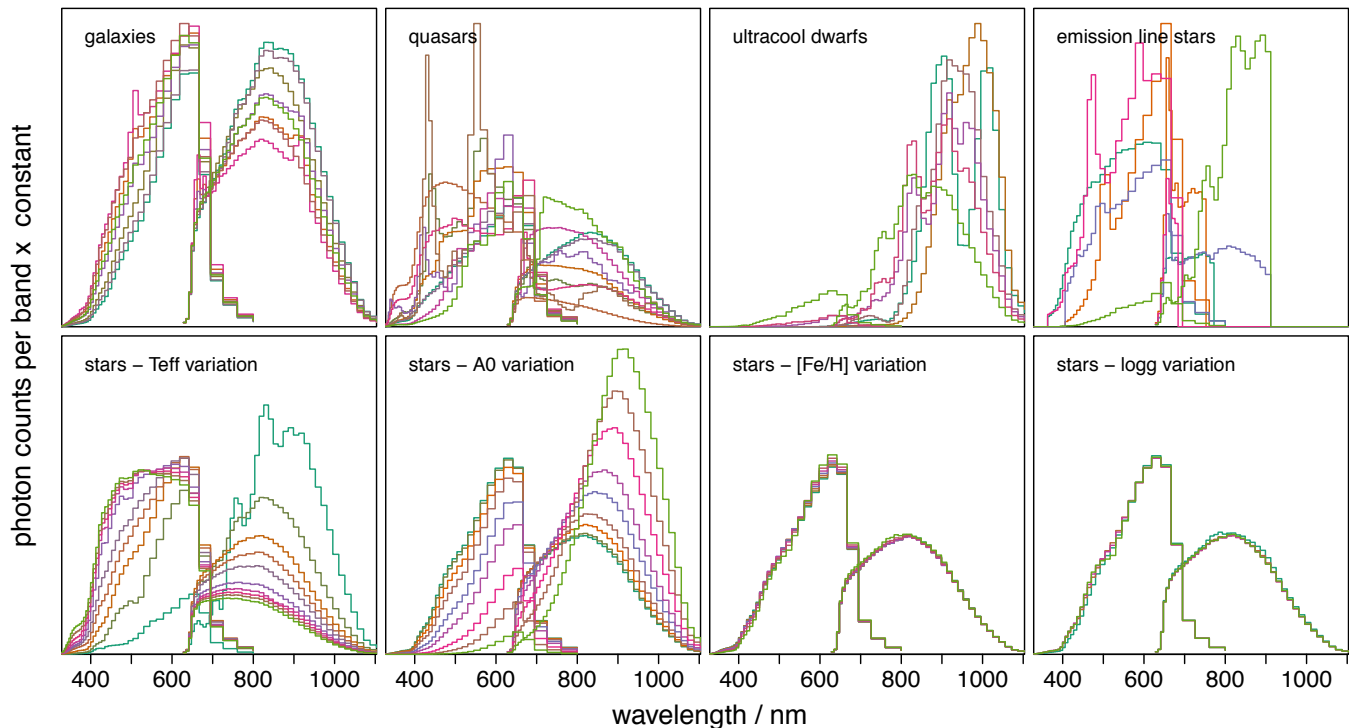
Each source will be observed between 40 and 250 times, depending primarily on its ecliptic latitude. BP/RP spectrophotometry are nominally obtained for all sources at every epoch. These are combined during the processing and calibration into a single BP/RP spectrum for each source. RVS spectra are obtained at fewer epochs due to the focal plane architecture, and these are also combined.

The accuracy of AP estimation depends strongly on the spectral signal-to-noise ratio (SNR) which, for a given number of epochs, is primarily a function of the source’s *G* magnitude. In the rest of this article we will consider a single BP/RP spectrum to be a combination of 70 observation epochs, which is the sky-averaged number of epochs per source (accounting also for various sources of epoch loss). For RVS it is 40 epochs. We refer to such combined spectra as “end-of-mission” spectra. All results in this article were obtained using (simulated) end-of-mission spectra.

### 2.2. Spectrophotometry (BP/RP) and spectroscopy (RVS)

BP and RP spectra are read out of the CCDs with 60 wavelength samples (or “bands”; they can be thought of as narrow overlapping filters). BP spans 330–680 nm with a reso-

<sup>1</sup> The  $G - V$  colours for B1V, G2V, and M6V stars are  $-0.01$ ,  $-0.18$  and  $-2.27$  mag respectively, so the Gaia limiting magnitude of  $G=20$  corresponds to  $V=20-22$  depending on the spectral type.



**Fig. 1.** Example BP/RP spectrophotometry. The spectra have been normalized to have the same number of photon counts over the spectral bands plotted. (This does not yield the same area under each spectrum as plotted – against wavelength – due to the nonlinear dispersion.) Except for the emission line stars, all spectra are noise-free synthetic spectra. Several examples of each type of object are shown in each panel (the line colours are arbitrary). The galaxies are for a range of types, all with zero redshift and zero Galactic extinction. The quasar spectra cover a range of emission line strengths, continuum slopes and redshifts. In the ultra cool dwarf panel six spectra are shown with  $T_{\text{eff}}$  ranging from 500–3000 K in steps of 500 K for  $\log g = 5$  dex. The top-right panel shows five emission line sources: Herbig Ae, PNe, T Tauri, WN4/WCE, dMe. The lack of red flux for these sources is a result of the input spectra used not spanning the full BP/RP wavelength range. The bottom row shows normal stars, in which just one parameter varies in each panel. From left to right these are:  $T_{\text{eff}} \in \{3000, 4000, 5000, 6000, 7000, 8500, 10000, 12000, 15000, 20000\}$  K;  $A_0 \in \{0.0, 0.1, 0.5, 1.0, 2.0, 3.0, 5.0, 7.0, 10.0\}$  mag;  $[\text{Fe}/\text{H}] \in \{-2.5, -1.5, -0.5, +0.5\}$  dex;  $\log g \in \{0, 2.5, 4, 5.5\}$  dex. The other parameters are held constant as appropriate at  $T_{\text{eff}} = 5000$  K,  $A_0 = 0$  mag,  $[\text{Fe}/\text{H}] = 0$  dex,  $\log g = 4.0$  dex. A common photon count scale is used in all the panels of the bottom row. The cooler/redder stars are those with increasingly more flux in the red part of the spectrum in the two lower left panels.

lution ( $=\lambda/\Delta\lambda$ ) varying from 85 to 13, and RP spans 640–1050 nm with a resolution of 26 to 17 (de Bruijne 2012). ( $\Delta\lambda$  is defined as the 76% energy width of the line spread function.) The resolution is considerably lower than what one would like for AP estimation, but limitations are set by numerous factors.<sup>2</sup> The upstream processing can in principle deliver spectra of higher resolution by a factor of a few for all sources, because the multiple epoch spectra are offset by fractions of a sample. Such “oversampled spectra” are not used in the present work. Examples of star, galaxy and quasar spectra are shown in Figure 1. Low SNR bands at the edges of both BP and RP have been omitted (approximately 8 bands from each end of both). Apsis will use BP/RP to classify all Gaia sources and to estimate APs down to the Gaia magnitude limit, although some “weaker” APs, such as  $\log g$ , will be poorly estimated at  $G = 20$ .

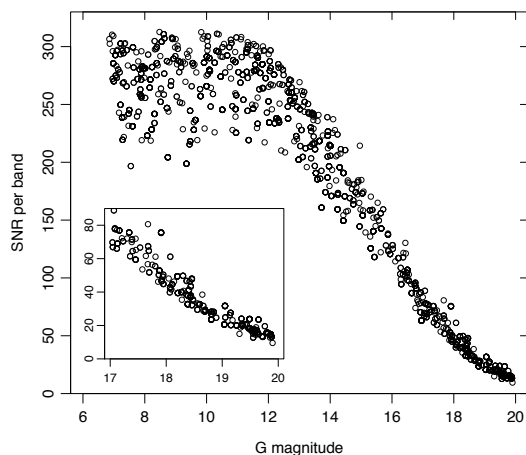
The expected variation of SNR of BP/RP with magnitude is shown in Figure 2. This plot includes a calibration error corresponding to 0.3% of the flux, an estimate based

both on past experience and our current understanding of the impact of systematic errors. This has not yet been included in the synthetic spectra used to train and test most Apsis modules, because it is difficult to estimate its magnitude in advance. Ignoring this calibration error increases the SNR at  $G = 15$  from about 175 to around 225, and for  $G > 17$  the difference in SNR is 10% or less, so most of our results are unaffected by this. Without systematic errors the SNR would extend to 1000–2000 for  $G < 12$ .

The radial velocity spectrograph (Katz et al. 2004; Cropper & Katz 2011) records spectra from 847 to 871 nm (the Ca II triplet region) at a resolution of 11 200; the figures given here reflect the manufactured instruments (T. Prusti September 2012, private communication). For SNR reasons, RVS does not extend to the  $G = 20$  limit of the other instruments, but will be limited to about  $G_{\text{RVS}} = 17$ , so is expected to deliver useable spectra for of order 200 million stars.<sup>3</sup> Spectra fainter than  $G_{\text{RVS}} = 10$  are binned on-chip by a factor of three in the dispersion direction in order to improve the SNR, at the cost of a lower spectral resolution.

<sup>2</sup> The Gaia consortium optimized a multi-band photometric system for Gaia, described by Jordi et al. (2006), but due to mission constraints this was not adopted.

<sup>3</sup>  $G_{\text{RVS}}$  is the photometric band formed by integrating the RVS spectrum, and in terms of magnitudes  $G_{\text{RVS}} \simeq I_c$ .

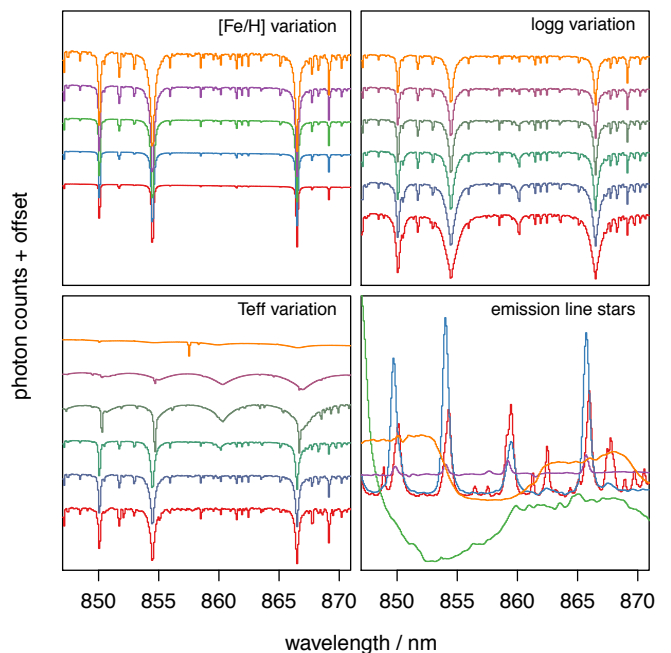


**Fig. 2.** Variation of the SNR per band in end-of-mission BP/RP for a set of 2000 stars covering the HR diagram. The inset is a zoom of the fainter magnitudes. Discontinuities occur at several brighter magnitudes (barely visible here) on account of the use of TDI gates to limit the integration time for bright stars in order to avoid saturation of the CCDs. The SNR for each spectrum is the mean over the bands plotted in Figure 1. In addition to the formal noise model errors, an additional error of 0.3% in the flux has been added in quadrature to accommodate calibration errors.

The main purpose of RVS is to measure radial velocities – the sixth component of the phase space. The radial velocity precision for most stars ranges from 1–15 km/s, depending strongly on both colour and magnitude (Katz et al. 2011, de Bruijne 2012). Apsis uses RVS data both for general stellar parameter estimation down to about  $G_{\text{RVS}} = 14.5$  (of order 35 million stars), and for characterizing specific types, such as emission line objects.

Examples of the RVS spectra are shown in Figure 3. The typical variation of the SNR with  $G_{\text{RVS}}$  is shown in Figure 4. This plot includes a 0.3% error assumed to arise from imperfect calibration and normalization. This was not included in the synthetic libraries used to train and test Apsis modules, although it decreases the SNR by no more than 15% for  $G_{\text{RVS}} > 10$ . It must be appreciated, however, that obtaining useable RVS spectra at the faint end depends critically on how well charge transfer inefficiency (CTI) effects in the CCDs can be modelled (Prod’homme et al. 2012).

The extraction, combination and calibration of both BP/RP and RVS spectra are complicated tasks which will not be discussed here. They are the responsibility of the coordination units CU5 (for BP/RP) and CU6 (for RVS) in DPAC, and are discussed in various technical notes (e.g. Jordi 2011; Katz et al. 2011; De Angeli et al. 2012). Apsis works with “internally calibrated” BP/RP and RVS spectra, by which we mean they are all on a common flux scale (and various CCD phenomena have been removed), but the instrumental profile and dispersion function have not been removed. The library spectra which form the basis of training our classification modules are projected into this data space using an instrument simulator (see section 4).



**Fig. 3.** Example RVS spectra. Each spectrum is noise free and at  $G_{\text{RVS}} = 12$  (i.e. with RVS in low resolution mode). The line colours are arbitrary. Three of the panels show the variation of one of the APs with the other two held constant, the constant values being  $T_{\text{eff}}=5500$  K,  $[\text{Fe}/\text{H}]=0$  dex,  $\log g=4.0$  dex.  $[\alpha/\text{Fe}]=0$  dex in all cases. In these cases the spectra in each panel have been offset vertically for clarity. The AP ranges (increasing from bottom to top in each panel) are:  $[\text{Fe}/\text{H}]=-2.5$  to 0.0 in steps of 0.5 dex;  $\log g=0$  to 5 in steps of 1 dex;  $T_{\text{eff}} \in \{4500, 5500, 6500, 8250, 14000, 40000\}$  K. The bottom right panel shows examples of five emission line stars (here the offset is zero). They are, from bottom to top around the feature at 859 nm: nova; WC star; O6f star; Be star; B[e] star.

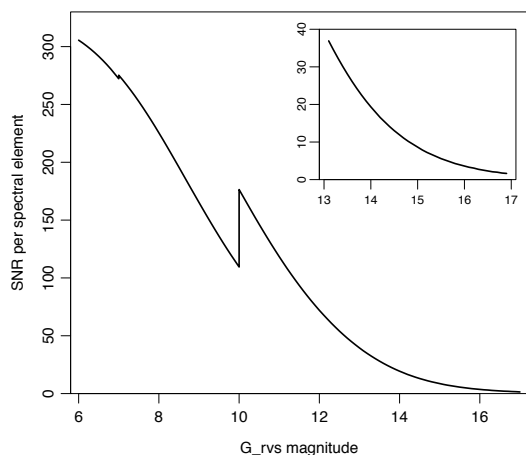
### 2.3. Photometry and astrometry

The sources’ G-band magnitudes are measured to a precision of 1–3mmag, limited by calibration errors even at  $G=20$ . The data processing will also produce integrated photometry for BP, RP and RVS, with magnitudes referred to as  $G_{\text{BP}}$ ,  $G_{\text{RP}}$ , and  $G_{\text{RVS}}$  respectively. For more details of these passbands including transformations between them and to non-Gaia passbands, see Jordi et al. (2010). These bands are used in Apsis primarily to assess (together with the number of observation epochs) the SNR of the spectra.

The astrometry is used in Apsis to help distinguish between Galactic and extragalactic objects, and parallaxes are also used in a few modules to aid stellar AP estimation. The astrometric accuracy is a function mostly of SNR and thus G magnitude. At  $G=15$ , 18.5, and 20 the sky-averaged parallax accuracy is 25–26  $\mu\text{as}$ , 137–145  $\mu\text{as}$ , and 328–347  $\mu\text{as}$  respectively, the ranges reflecting the colour dependence across early B to late M stellar spectral types (slightly better for earlier type stars).<sup>4</sup> For  $6 < G < 14$  the accuracy is 7–17  $\mu\text{as}$ , although the performance at the bright limit will depend on the actual TDI gate scheme

<sup>4</sup> On account of the large width of the G-band, the accuracies at constant V-band magnitude are quite different, e.g. ranging from 26  $\mu\text{as}$  for early B types to 9  $\mu\text{as}$  for late M types at  $V=15$ .





**Fig. 4.** Variation of the SNR per spectral element in end-of-mission RVS spectra with  $G_{RVS}$ . The inset is a zoom of the fainter magnitudes. The discontinuity at  $G_{RVS} = 10$  is due to the on-chip binning of the spectrum in the dispersion direction for fainter stars, and that at  $G_{RVS} = 7$  is due to on-chip binning perpendicular to the dispersion direction for fainter stars. In addition to the formal noise model errors, an additional error of 0.3% in the flux has been added in quadrature to accommodate calibration and normalization errors.

used to avoid saturating the bright stars.<sup>5</sup> The proper motion accuracies in  $\mu\text{as}/\text{yr}$  are about 0.5 times the size of the quoted parallax accuracies.

How the parallax accuracy converts to distance accuracy depends on the parallax itself. For example, an unreddened K1 giant at 5 kpc would have an apparent magnitude of  $G = 14.0$  and a distance accuracy of 9%. A G3 dwarf at 2 kpc has  $G = 16.5$  and a distance accuracy of 8%. When combining these accuracies with a model for the Galaxy, we expect the number of stars with distance determinations better than 0.1%, 1% and 10% to be of order  $10^5$ ,  $10^7$ , and  $10^8$  (respectively).

### 3. The astrophysical parameters inference system (Apsis)

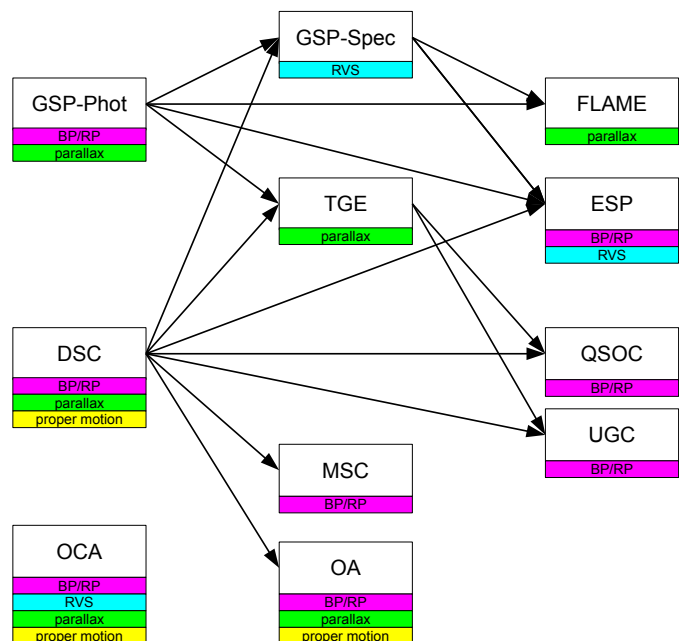
#### 3.1. Principles

The goal of Apsis is to classify and to estimate astrophysical parameters for the Gaia sources using the Gaia data. These APs will be part of the publicly released Gaia catalogue. They will also be used internally in the data processing, for example to help the template-based extraction of the RVS spectra and the identification of quasars used to fix the astrometric reference frame.

Our guiding principle for Apsis is to provide reasonably accurate estimates for a broad class of objects covering a large fraction of the catalogue, rather than to treat some specific types of objects exhaustively. To achieve this, Apsis consists of a number of modules with different functions.

The paradigm which underlies most of the Apsis modules is supervised learning. This means that the classes or

<sup>5</sup> The 5 000 or so stars brighter than  $G = 6$  will saturate in the focal plane, but may yield useful measurements if we can calibrate their diffraction spikes.



**Fig. 5.** Component modules in Apsis and their interdependency. The module names are defined in Table 1. The arrows indicate a dependency on the output of the preceding module. The coloured bars underneath each module indicate which data it uses. Most of the modules additionally use the photometry and some also the Galactic coordinates.

parameters of objects are determined according to the similarity of the data to a set of templates for which the parameters are already known, so-called “labelled” data. How this comparison is done – in particular, how we interpolate between the templates and how we use the data – is an important attribute distinguishing between the various machine learning (or pattern recognition) algorithms available. Our choices are based on their accuracy, utility and speed. The term “training” is used to describe the process by which the algorithm is fit to (learns from) the template data. For the most part we have, to date, used libraries of synthetic spectra as the basis for our training data, although we also use some semi-empirical libraries. These libraries and the construction of the training and testing data using a Gaia instrument simulator are described in section 4. Later, actual Gaia observations will be used to calibrate the synthetic spectral grids (see section 6).

#### 3.2. Architecture

Each of the modules in Apsis is described separately in section 5. Here we give an overview and describe their connectivity, which is summarized in Figure 5. The acronyms are defined in Table 1.

DSC performs a probabilistic classification into classes such as “(single) star”, “binary star”, “quasar”. This is used by many of the other modules to select sources for processing. GSP-Phot and GSP-Spec estimate stellar parameters using the BP/RP spectra (and parallaxes) and the RVS spectra respectively, whereby GSP-Phot also estimates the line-of-sight extinction to each star individually. Supporting these are a number of “extended stellar parametrization”

**Table 1.** Apsis modules

Acronym	name
DSC	Discrete Source Classifier
ESP	Extended Stellar Parametrizer:
-CS	ESP – Cool Stars
-ELS	ESP – Emission Line Stars
-HS	ESP – Hot Stars
-UCD	ESP – Ultra Cool Dwarfs
FLAME	Final Luminosity Age and Mass Estimator
GSP-Phot	Generalized Stellar Parametrizer – Photometry
GSP-Spec	Generalized Stellar Parametrizer – Spectroscopy
MSC	Multiple Star Classifier
OA	Outlier Analysis
OCA	Object Clustering Algorithm
QSOC	Quasar Classifier
TGE	Total Galactic Extinction
UGC	Unresolved Galaxy Classifier

modules, which operate on specific types of stars, their preliminary identification being taken from GSP-Phot and (if the stars are bright enough) GSP-Spec. These are ESP-ELS, ESP-HS, ESP-CS, and ESP-UCD. Although GSP-Phot is trained on a broad set of stars which includes all of these, these modules attempt to achieve more appropriate parameters estimates by making a more physically-motivated use of the data, and/or by using other stellar models. Using the outputs of GSP-Phot, FLAME uses isochrones to estimate stellar luminosities, masses and ages for certain types of stars. MSC attempts to estimate parameters of both components of systems suspected (by DSC) to be unresolved stellar binaries. QSOC and UGC estimate astrophysical parameters of quasars and (unresolved) galaxies, respectively. TGE will use the line-of-sight extinction estimates from GSP-Phot of the most distant stars to build a two-dimensional map of the total Galactic extinction over the whole sky. This may also be used as an input to QSOC.

The two remaining Apsis modules use the concept of unsupervised learning. OCA works independently of all other modules by using clustering techniques to detect “natural” patterns in the data, primarily for novelty detection. OA does something similar on the objects classified as “outliers” by DSC. Its purpose is to identify whether some of these outliers are known objects which were not, or were not correctly, modelled in the training data. Results from this can be used to improve the models in the next processing cycle.

### 3.3. Source selection

Which sources are processed by which modules depends on (1) the availability of the necessary data; (2) the SNR of the data; (3) the outputs from other modules.

DSC operates on all sources which have BP/RP data, which is nominally all Gaia sources. For each source, DSC assigns probabilities to a set of classes. This is the main output from DSC for the end-user. In addition, a single “best” class will be identified for each source. In principle this is just the class which receives the highest probability, but in practice this probability will also have to exceed some class-dependent threshold. Some sources may not attain this, in which case they will be classified as “unknown”.

GSP-Phot operates on all sources too. As more than 99% of sources are expected to be stars, there is little loss of efficiency if GSP-Phot is simply applied to everything, regardless of the DSC class. The GSP-Phot APs for what are later chosen to be non-stars based on the DSC class probabilities can then simply be ignored.

GSP-Spec will operate on all stars identified by DSC which have RVS spectra with sufficiently high SNR ( $G_{RVS} \lesssim 15$ ). FLAME operates on a subset of sources which have APs of sufficient precision from GSP-Phot and/or GSP-Spec. TGE selects a small fraction of distant stars assigned precise extinction estimates by GSP-Phot. The remaining modules, specifically ESP-HS, ESP-CS, ESP-ELS, ESP-UCD, MSC, QSOC, and UGC, will only be applied to objects of “their” class, as determined by the DSC class probabilities.

### 3.4. Multiple parameter assignments

A consequence of our system design is that any given source may be assigned multiple sets of APs. For example, a particular star could be assigned APs by GSP-Phot, ESP-CS, and GSP-Spec. This is an inevitable consequence of a diverse approach to inference: the conclusions we draw depend not only on the data we measure, but also on the stellar models we adopt (as embodied in the training data) and other assumptions made. We can never know the true APs with 100% confidence. All of these sets of APs will be reported in the MDB and the data releases, thus giving the end-user the freedom to choose among our models and assumptions. For those users who would rather forgo this choice, we will also provide the “best” set of APs for each source. We will establish the criteria for making this decision during the operations, based on experience with the data. GSP-Phot estimates APs for all stars, so there will always be a homogeneous set of stellar APs available.

The situation is actually more complex than this, because a few of the modules themselves comprise multiple algorithms, each providing separate estimates of the APs. One of the reasons for this is cross-checking: if two or more algorithms give similar results for the same source (and training data), our confidence in the results may be increased. A second reason is that different methods may make use of different data. For example, the Aeneas algorithm in GSP-Phot (section 5.2) can operate with or without the parallax. The former is potentially more accurate, yet makes more assumptions, so we may be interested in both results. A third reason for using multiple algorithms is that the best performing algorithm may be computationally too expensive to run on all sources.

### 3.5. Scope

One of the principles we adopt in DPAC is that the Gaia catalogue will be based only on Gaia data for the individual sources. (Non-Gaia observations are used for validation and calibration; see section 6.) “Better” AP estimates could be obtained for some sources by including external data in the analysis, such as higher resolution spectra or infrared photometry. The DPAC objective, however, is to produce a homogeneous Gaia catalogue by processing all sources in a consistent manner. We hope that the community at large



will extend our work by using the published data to make composite analyses where appropriate.

While Apsis tries to cover most types of objects, it does not include everything. Asteroids are excluded, for example. They will be detected by Gaia primarily via their very large proper motions, so they will be classified by the CU charged with their detection (CU4). Apsis presently ignores morphological information. Although Gaia only tracks point sources, two-dimensional images could be reconstructed using the multiple scans at different orientations over a source. This is planned by other CUs in DPAC, and could be useful for further galaxy characterization, for example. This may be introduced later into the data processing.

Apsis also does not take into account stellar variability. An entire CU in DPAC, CU7, is dedicated to classifying variable stars from, primarily, their G-band light curves. As Apsis works with combined epoch spectra, some types of variable source will receive spurious APs. During the course of the data processing we will investigate how and whether variability information can be introduced into our work.

### 3.6. Software, hardware and operations

The Apsis modules have been developed by various CU8 groups over the past years following a cyclic development process. They are written in the Java programming language according to DPAC-wide software engineering standards. The modules are integrated into a control system which deals with job allocation and data input/output.

As outlined in section 1, the Gaia data processing proceeds in cycles, centered around the MDB. When Apsis first runs (several months into the mission), essentially all Gaia sources will have been observed at least once. In succeeding cycles, Apsis will run again on the same set of sources, but the data are the combination of more observation epochs, so will have higher SNR and improved calibrations. The first significant, calibrated results from Apsis should appear about 2.5 years into the mission, and will be made available in the subsequent intermediate data release.

Apsis will run on multicore computers at CU8's data processing centre hosted by CNES in Toulouse. The time Apsis needs for processing is likely to vary considerably during the ongoing development, but as of late 2012, the supervised modules (i.e. excluding OA and OCA) together required of order 15 GFLOP (1 GFLOP =  $10^9$  floating point operations) for a single source. This is dominated by the Markov Chain Monte Carlo (MCMC) sampling performed by the Aeneas algorithm in GSP-Phot. A common CPU will today provide around 100 GFLOP per second, so processing all  $10^9$  Gaia sources in this way would take 1740 days. The Apsis processing is trivial to parallelize, so running it on 100 CPU cores reduces this to 17 days. However, the 15 GFLOPS figure neglects data input/output, which is likely to be a considerable fraction of the processing time. OCA and OA are also likely to add to this figure significantly. On the other hand, CU8 will have more like 400 CPU cores available full time for its processing. Given that we need to process all Gaia sources in one operation cycle (duration of 6–12 months), these figures are acceptable even if we assume some intra-cycle reprocessing.

## 4. Model training and testing

Supervised classification methods are based on the comparison of observed data with a set of templates. These are used to train the models in some way. For this purpose we may use either observed or synthetic templates, both of which have their advantages and disadvantages. Observed templates better represent the spectra one will actually encounter in the real data, but rarely cover the necessary parameter range with the required density, in particular not for a survey mission like Gaia. Synthetic templates allow us to characterize a wide parameter space, and also to model sources which are very rare or even which have not (yet) been observed. Intrinsically free of observational noise and interstellar extinction, they allow us to freely add these effects in a controlled manner. They are, however, simplifications of the complex physics and chemistry in real astrophysical sources, so they do not reproduce real spectra perfectly. This may be problematic for pattern recognition, so synthetic spectra will need calibration using the actual Gaia observations of known sources (see section 6).<sup>6</sup>

The training data for the Apsis modules are based on a mixture of observed (actually “semi-empirical”) and synthetic libraries for the main sources we expect to encounter. These are described below. Once the library spectra have been constructed, BP/RP and RVS spectra are artificially reddened, then simulated at the required G magnitude and with a SNR corresponding to end-of-mission spectra (see section 2) using the Gaia Object Generator (GOG, Luri et al. 2005).

### 4.1. Stellar spectral libraries

The Gaia community has calculated large libraries of synthetic spectra with improved physics for many types of stars. We are able to cover a broad AP space with some redundancy between libraries. Each library uses codes optimized for a given  $T_{\text{eff}}$  range, or for a specific object type, and includes as appropriate the following phenomena: departures from local thermodynamic equilibrium (LTE); dust; mass loss; circumstellar envelopes; magnetic fields; variations of single element abundances; chemical peculiarities. The libraries are listed in Table 2 with a summary of their properties and AP space. Not all of these libraries are used in the results reported in section 5. The synthetic stellar libraries are described in more detail in Sordo et al. (2010, 2011) together with details on their use in the Gaia context. The large synthetic grids for A, F, G, K, and M stars have been computed in LTE for both BP/RP and RVS. For OB stars, non-LTE (NLTE) line formation has been taken into account.

Synthetic spectra are of course not perfect. We cannot yet satisfactorily simulate some processes, such as emission line formation. To mitigate these drawbacks, observed spectra are included in the training dataset in the form of semi-empirical libraries. These are observed spectra to which APs have been assigned using synthetic spectra, and for which the wavelength coverage has been extended (as necessary) using the best fitting synthetic spectrum. Semi-empirical libraries have been constructed for “normal” stars using SDSS Tsalmanza & Bailer-Jones (2010b), and from

<sup>6</sup> Of course, to estimate *physical* parameters we must, at some point, use *physical* models, so dependence on synthetic spectra cannot be eliminated entirely.

**Table 2.** Stellar libraries used to simulate BP/RP and RVS spectra.  $N$  is the number of spectra in the library. Ap/Bp are peculiar stars; UCD are ultracool dwarfs; WR are Wolf Rayet stars; WD are white dwarfs.

Name	$N$	$T_{\text{eff}} / \text{K}$	$\log g / \text{dex}$	$[\text{Fe}/\text{H}] / \text{dex}$	Ref. <sup>†</sup>	Notes
OB stars	1296	15 000–55 000	1.75–4.75	0.0–0.6	1	TLUSTY code; NLTE, mass loss, $v_{\text{micro}}$
Ap/Bp stars	36	7000–16000	4.0	0.0	2	LLmodels code, chemical peculiarities
A stars	1450	6000–16 000	2.5–4.5	0.0	3	LLmodels code, $[\alpha/\text{Fe}] = 0.0, +0.4$
MARCS	1792	2800–8000	−0.5–5.5	−5.0–1.0	4	Galactic enrichment law for $[\alpha/\text{Fe}]$
Phoenix	4575	3000–10 000	−0.5–5.5	−2.5–0.5	5	$\Delta T_{\text{eff}} = 100 \text{ K}$
UCD	2560	400–4000	−0.5–5.5	−2.5–0.5	6	various dust models
C stars MARCS	428	4000–8000	0.0–5.0	−5.0–0.0	7	$[\text{C}/\text{Fe}]$ depends on $[\text{Fe}/\text{H}]$
Be	174	15 000–25 000	4.0	0.0	8	range of envelope to stellar radius ratios
WR	43	25 000–51 000	2.8–4.0	0.0	9	range of mass loss rates
WD	187	6000–90 000	7.0–9.0	0.0	10	WDA & WDB
MARCS NLTE	33	4000–6000	4.5–5.5	0.0	11	NLTE line profiles
MARCS RVS	146 394	2800–8000	−0.5–5.5	−5.0–1.0	12	variations in individual elements abundances
3D models	13	4500–6500	2.0–5.0	−2.0–0.0	13	StaggerCode models and Optim3D code
SDSS stars	50 000	3750–10 000	0.0–5.5	−2.5–0.5	14	semi-empirical library
Emission line stars	1620	–	–	–	15	semi-empirical library (see section 5.4)

<sup>†</sup>References: 1 Bouret et al. (2008); 2 Kochukhov & Shulyak (2008); 3 Shulyak et al. (2004); 4 Gustafsson et al. (2008); 5 Brott & Hauschildt (2005); 6 Allard et al. (2001); 7 Masseron, priv. comm.; 8, 9 Martayan et al. (2008); 10 Castanheira et al. (2006); 11 Korn et al. 2009, priv. comm.; 12 Recio-Blanco et al. 2011, priv. comm.; 13 Chiavassa et al. (2011); 14 Tsalmantza & Bailer-Jones (2010b); 15 Lobel et al. (2010)

other sources for emission line stars (Lobel et al. 2010; see section 5.4).

Starting from the available synthetic and semi-empirical libraries, two types of data set are produced. The first one mirrors the AP space of the spectral libraries, and is regularly spaced in some APs. The second one involves interpolation on some of the APs ( $T_{\text{eff}}$ ,  $\log g$ ,  $[\text{Fe}/\text{H}]$ ) but with no extrapolation (and we do not combine different libraries). See Sordo et al. (2011) for details on how the interpolation is done. Both datasets are intended for training the AP estimation modules, while the interpolated one serves also for testing. In both cases extinction is applied using Cardelli’s law (Cardelli et al. 1989), with a given set of extinction parameters. Extinction is represented using an extinction parameter,  $A_0$ , rather than the extinction in a particular band, as defined in section 2.2 of Bailer-Jones (2011). The parameter  $A_0$  corresponds to  $A_V$  in Cardelli’s formulation of the extinction law, but this new formulation is chosen to clarify that it is an extinction parameter, and not necessarily the extinction in the  $V$  band, because the extinction (for broad bands) depends also on the spectral energy distribution of the source.

Mass, radius, age, and absolute magnitudes of the stars are derived using the Padova evolutionary models (Bertelli et al. 2008), resulting in a full description of stellar sources. These models cover a wide range of masses up to  $100 M_{\odot}$  and metallicities for all evolutionary phases. Although this is not needed for most Apsis modules, it is required by the GSP-Phot module Aeneas when it is using parallax, in order to ensure consistency between parallax, apparent and absolute magnitude, and the stellar parameters in the training data set (see section 3.3 of Liu et al. 2012 for a discussion).

#### 4.2. Galaxy spectral libraries

For the classification of unresolved galaxies we have generated synthetic spectra of normal galaxies (Tsalmantza et al. 2007, 2009; Karampelas et al. 2012) using the galaxy evo-

lution model PÉGASE.2 (Fioc & Rocca-Volmerange 1997, 1999). The objective was not just to obtain a set of typical synthetic spectra, but to have a broad enough sample which can predict the full variety of galaxies we expect to observe with Gaia. Four galaxy spectral types have been adopted: early, spiral, irregular, quenched star formation. For each type, a restframe spectrum is characterized by four APs: the timescale of in-falling gas and three parameters which define the appropriate star formation law. The current library comprises 28 885 synthetic galaxy spectra at zero redshift. These are then simulated at a range of redshifts from 0 to 0.2, and a range of values of  $A_0$  from 0 to 10 mag in order to simulate extinction due to the interstellar medium of our Galaxy.

In addition, a semi-empirical library of 33 670 galaxy spectra has been produced fitting SDSS spectra to the synthetic galaxy library (Tsalmantza et al. 2012).

#### 4.3. Quasar spectral libraries

Two different libraries of synthetic spectra of quasars have been generated, one with regular AP space coverage (17 325 spectra) and one with random AP space coverage (20 000 spectra). Three QSO APs are sampled: redshift (from 0 to 5.5), slope of the continuum ( $\alpha$  from -4 to +3), and equivalent width of the emission lines (EW from  $10^1$  to  $10^5 \text{ nm}$ ). The libraries also sample a broad range of Galactic interstellar extinction,  $A_0 = 0\text{--}10 \text{ mag}$ .

A semi-empirical library of 70 556 DR7 SDSS spectra of quasars has also been generated. The majority of quasars in this library have redshift below 4,  $\alpha$  from -1 to +1 and EW from 0 to 400 nm (the distributions are very non-uniform). These are not artificially reddened, but some will have experienced a small amount of real extinction.

### 5. The Apsis modules

We now describe each of the modules in Apsis listed in Table 1 and summarized in Figure 5.

### 5.1. Discrete Source Classifier (DSC)

DSC performs the top-level classification of every Gaia source, assigning a probability to each of a number of classes. These are currently: star, white dwarf, binary star, galaxy, quasar. For this it uses three groups of input data: the BP/RP spectrophotometry; the proper motion and parallax; the position and the photometry in the  $G$ -band. Each group of input data is directed to a separate subclassifier (described below), each of which produces a vector of probabilities for the classes. The results from all the subclassifiers are combined into a single probability vector, and based on this a class label may be generated if the highest probability exceeds a certain threshold. An additional module using the  $G$ -band light curve (time series) – or rather metrics extracted from it – is under development. An earlier phase of the DSC development was presented in some detail by Bailer-Jones et al. (2008), who examine in particular the issue of trying to identify rare objects.

The photometric subclassifier works with the BP/RP spectra. The classification algorithm is the Support Vector Machine (SVM; Vapnik 1995; Cortes & Vapnik 1995; Burges 1998), which is widely used for analysing high-dimensional astrophysical data (e.g. Smith et al. 2010). We use the implementation libSVM (Chang & Lin 2011). (We have also tested other machine learning algorithms, such as random forests, and find similar overall performance.) A set of SVM models is trained, each at a different  $G$  magnitude range, and the observed BP/RP spectrum passed to the one appropriate to its measured magnitude. (This is done because SVMs work best when the training and test data have similar noise levels.) Each model contains two layers, the first trained on an astrophysically meaningful distribution of common objects, the second trained on a broad distribution of AP space and intended in particular to classify rare objects. For each layer, a front-end outlier detector identifies sources that do not resemble the training data sufficiently closely. Only objects rejected by the first layer are passed to the second layer for classification. Flags are set to indicate outliers detected by each layer. These will be studied by the OA module (see section 5.13).

The astrometric subclassifier uses the parallaxes and proper motions to help distinguish between Galactic and extragalactic objects. This uses a three-dimensional Gaussian mixture model trained on noise-free, simulated astrometry for Galactic and extragalactic objects. This model is convolved with the estimated uncertainties in the proper motion and parallax for each source, and a probability of the source being Galactic or extragalactic is calculated.

The position–magnitude subclassifier gives a probability of the source being Galactic or extragalactic based on the source’s position and brightness. This reflects our broad knowledge of the overall relative frequency of Galactic and extragalactic objects and how they vary as a function of magnitude and Galactic coordinates. For example, if we knew only that a source had  $G=14$  and were at low Galactic latitude, we would think it more likely to be Galactic than extragalactic. In the absence of more data, we should fall back on this prior information. This subclassifier quantifies this using a simple lookup table based on a simple universe model. For very informative spectra, this subclassifier would have little influence on the final probabilities.

DSC is trained on numerous data sets built from almost all of the spectral libraries described in section 4, in-

**Table 3.** Example of the DSC classification performance shown as a confusion matrix for sources with magnitudes in the range  $G=6.8–20$ . The rows indicate the true classes (the spectral libraries), the columns the DSC assigned class. Each cell gives the percentage of objects classified from each true class to each DSC class. The dashes indicate exactly zero.

Library	Output class				
	Star	WD	Binary	Quasar	Galaxy
Phoenix	91.9	–	7.1	–	1.0
Phoenix- $R_0$	89.9	3.0	7.1	–	–
A stars	79.9	–	20.0	–	0.1
OB stars	95.3	0.6	4.1	–	–
WD	17.4	79.1	3.5	–	–
UCDs	97.3	–	1.0	1.7	–
Binary stars	18.3	–	81.7	–	–
SDSS stars	94.1	–	5.9	–	–
SDSS quasars	5.9	3.0	0.1	78.3	12.7
SDSS galaxies	2.0	–	0.5	–	97.5

cluding blended spectra of different types of objects (e.g. optical stellar binaries). A selection of the results on independent test sets constructed from these libraries is shown in Table 3. Phoenix- $R_0$  is the Phoenix library but now also showing a large variation in the second extinction parameter  $R_0$ . (More detailed results from an earlier version of the software can be found in Smith 2011.) These results combine the outputs from all three subclassifiers, and is for Galactic objects with magnitudes ranging from  $G = 6.8–20$  and quasars and galaxies from  $G = 14–20$  (uniform distributions) in both the training and test sets. The synthetic spectra include the 0.3% calibration error mentioned in section 2.2. The SDSS stars, quasars and galaxies are the semi-empirical libraries. The performance for stars and galaxies is generally quite good. Some confusion between single stars and physical binaries is expected because the binary sample includes systems with very large brightness ratios (see section 5.8). The relatively poor performance on quasars arises mostly due to a confusion with galaxies. If only the photometric subclassifier is used then a similar performance is obtained, but the confusion is then mostly with white dwarfs. Note that these results already assume quasars to be rare (1 for every 100 stars). Note that for faint, distant stars the true parallax and proper motions can be comparable to the magnitude of the uncertainty in the Gaia measurements, in which case the astrometry does not allow a good discrimination between Galactic and extragalactic objects.

These results should not be over-interpreted, however, as the performance depends strongly on the number of classes included in the training, to the extent that excluding certain classes can lead to much better results. Performance also depends on the relative numbers of sources in each class as well as their parameter distributions in the training and test data sets. Optimizing these is an important part of the on-going work.

### 5.2. Stellar parameters from BP/RP (GSP-Phot)

The objective of GSP-Phot is to estimate  $T_{\text{eff}}$ ,  $[\text{Fe}/\text{H}]$ ,  $\log g$  and the line-of-sight extinction,  $A_0$ , for all single stars observed by Gaia. The extinction is effectively treated as a

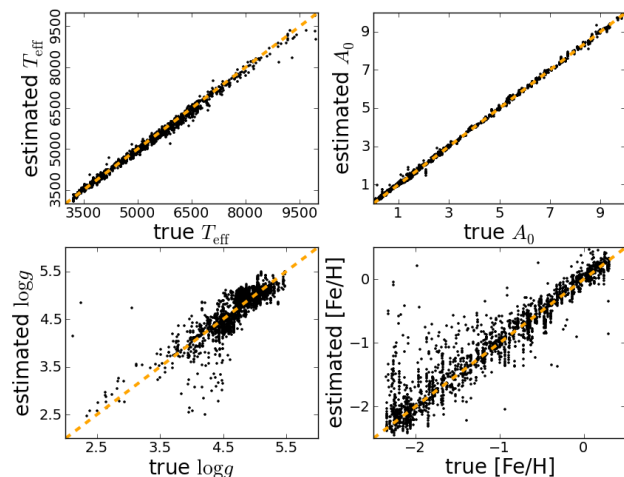
stellar parameter. (The total-to-selective extinction parameter,  $R_0$ , may be added at a later stage.) GSP-Phot uses the BP/RP spectrum and, in one algorithm, the parallax. In addition to being part of the Gaia catalogue, the AP estimates are used by several downstream algorithms in Apsis (Figure 5) and elsewhere in the Gaia data processing. The algorithms and their performance are described in more detail in Liu et al. (2012) and the references given below.

GSP-Phot is applied to all sources irrespective of class. While we could exclude those sources which DSC assigns a low star probability, the majority of Gaia sources are stars, so excluding them saves little computing time. A threshold on this probability can be applied by any user of the catalogue according to how pure or complete they wish their sample to be.

GSP-Phot contains four different algorithms. Each provides AP estimates for each target:

1. Priam (Kim 2013): Early in the mission, no calibrated BP/RP spectra are available. Priam uses only the integrated photometry ( $G$ ,  $G_{BP}$ ,  $G_{RP}$ ,  $G_{RVS}$ ) to estimate  $T_{\text{eff}}$  and  $A_0$  (see below). This algorithm uses SVM models trained on synthetic spectra.
2. SVM: An SVM is trained to estimate each of the four stellar APs using the BP/RP spectra. SVMs are computationally fast and relatively robust, but we find them not to be the most accurate method for GSP-Phot. Furthermore, a standard SVM does not provide uncertainty estimates (although techniques do exist for extracting these from SVMs). The SVM AP estimates will also be used to initialize the next two algorithms.
3. ILIUM (Bailer-Jones 2010): This uses a forward model, fit using labelled data, to predict APs given the observed BP/RP spectrum. An iterative Newton–Raphson minimization algorithm is used to find the best fitting forward-modelled spectrum, and thus the APs and their covariances. A two-component forward model is used to retain sensitivity to the “weak” APs  $\log g$  and  $[\text{Fe}/\text{H}]$  which only have a weak impact on the stellar spectrum compared to  $T_{\text{eff}}$  and  $A_0$ .
4. Aeneas (Bailer-Jones 2011): This is a Bayesian method employing a forward model and a Monte Carlo algorithm to sample the posterior probability density function over the APs, from which parameter estimates and associated uncertainties are extracted. Aeneas may be applied to the BP/RP spectrum alone, or together with the parallax. When using the parallax, the algorithm demands (in a probabilistic sense) that the inferred parameters be consistent not just with the spectrum, but also with the parallax and apparent magnitude. Consistency with the Hertzsprung-Russell diagram can also be imposed, thereby introducing constraints from stellar structure and evolution.

As single stars are the main Gaia target, we decided that multiple algorithms and multiple sets of AP estimates were desirable for the sake of consistency checking. Tests to date show that SVM, ILIUM, and Aeneas are each competitive in some part of AP space or SNR regime. Our plan is that individual results as well as a single set of “best” APs for each source will be published in the data releases (see section 3.4). Nonetheless, we may find during the data processing that some or all algorithms are unable to provide useful estimates of “weak” APs on fainter stars.



**Fig. 6.** Accuracy of AP estimation with the GSP-Phot algorithm Aeneas using BP/RP spectra of stars at  $G = 15$  covering the full AP space shown. There are a total of 2000 stars in this test sample. The vertical structure of points visible in the bottom panels is due to our procedure to generate test spectra from a limited supply of synthetic spectra. This causes spectra of identical AP values to appear multiple times in the test set, though these spectra differ in their noise realizations.

The APs provided by GSP-Phot are of course tied to the stellar libraries on which it was trained, and different libraries may produce different results. As no single library models the full AP space better than all others, we work with multiple libraries. We could attempt to merge all libraries into one, but this would hide the resulting inhomogeneities (or even introduce errors). We decided instead to train a GSP-Phot model on each library independently, and use each to estimate APs for a target source. The most appropriate set of results (i.e. library) can be decided post hoc based either on a model comparison approach, the estimated uncertainties, or perhaps a simple colour cut. This is still under investigation.

Since the publication of GSP-Phot results by Liu et al. (2012), SVM and in particular Aeneas have been improved. Figure 6 and Table 4 summarize the current internal accuracy of Aeneas, using parallaxes as well as BP/RP.

As noted above, the purpose of Priam is to characterize the stars in the early data releases only, before BP/RP is calibrated. As the  $G-G_{BP}$  and  $G-G_{RP}$  colours are almost perfectly correlated, these three bands yield essentially just one colour, making it impossible to estimate two APs ( $T_{\text{eff}}$  and  $A_0$ ) without using prior information. Assuming  $A_0 < 2$  mag (but with  $T_{\text{eff}} = 3000\text{--}10\,000$  K), we can estimate  $T_{\text{eff}}$  and  $A_0$  to an RMS accuracy of 1000 K and 0.4 mag respectively using three bands (the latter decreases to 0.3 mag if we introduce  $G_{RVS}$ ). If we can assume  $A_0 < 0.1$  mag, then  $T_{\text{eff}}$  can be estimated to an accuracy of 550 K using either three or four bands.

Aeneas has also been tested on real data. It was used by Bailer-Jones (2011) to estimate  $T_{\text{eff}}$  and  $A_0$  for 50 000 Hipparcos FGK stars cross-matched with 2MASS, using the parallax and five band photometry (two from Hipparcos, three from 2MASS). The forward model was fit to a subset of the observed photometry, with temperatures obtained from echelle spectroscopy, and extinction modelled by applying an extinction law to the photometry.  $T_{\text{eff}}$  and  $A_0$  could be estimated to precisions of 200 K and 0.2 mag re-

**Table 4.** Accuracy of AP estimation (internal RMS errors) with the GSP-Phot algorithm Aeneas using BP/RP spectra and parallaxes, for stars covering the full AP space shown in Figure 6.

	$G$ mag	$T_{\text{eff}}$ K	$A_0$ mag	$\log g$ dex	$[\text{Fe}/\text{H}]$ dex
A stars	9	340	0.08	0.43	0.86
	15	260	0.06	0.38	0.93
	19	400	0.15	0.51	0.74
F stars	9	150	0.06	0.36	0.36
	15	170	0.07	0.38	0.33
	19	630	0.35	0.37	0.60
G stars	9	140	0.07	0.31	0.14
	15	140	0.07	0.22	0.16
	19	450	0.33	0.45	0.65
K stars	9	100	0.09	0.26	0.19
	15	90	0.08	0.26	0.21
	19	230	0.23	0.36	0.48
M stars	9	60	0.13	0.15	0.21
	15	70	0.14	0.29	0.25
	19	90	0.13	0.17	0.29

spectively, from which a new HRD and 3D extinction map of the local neighbourhood could be constructed.

### 5.3. Stellar parameters from RVS (GSP-Spec)

GSP-Spec estimates  $T_{\text{eff}}$ ,  $\log g$ , global metallicity  $[\text{M}/\text{H}]$ , alpha element abundance  $[\alpha/\text{Fe}]$ , and some individual chemical abundances for single stars using continuum-normalized RVS spectra (i.e. each spectrum is divided by an estimate of its continuum). Source selection is based on the DSC single star probability, and GSP-Spec can optionally use the measured rotational velocities ( $v \sin i$ ) from CU6, as well as the stellar parameters from GSP-Phot.

Presently, three algorithms are integrated in the GSP-Spec module: MATISSE (Recio-Blanco et al. 2006), DEGAS (Kordopatis et al. 2011a; Bijaoui et al. 2012), and GAUGUIN (Bijaoui et al. 2012). MATISSE is a local multilinear regression method. The stellar parameters are determined through the projection of the input spectrum on a set of vectors, calculated during a training phase. The DEGAS method is based on an oblique k-d decision tree. GAUGUIN is a local optimization method implementing a Gauss-Newton algorithm, initialized by parameters determined by GSP-Phot or DEGAS. The algorithms perform differently in different parts of the AP and SNR space. Which results will be provided by which algorithm will be decided once we have experience with the real Gaia data. As the estimation of the atmospheric parameters and individual abundances from RVS is sensitive to the pseudo-continuum normalization, GSP-spec renormalizes the RVS spectra through an iterative procedure coupled with the stellar parameters as determined by the three algorithms (Kordopatis et al. 2011a).

**Table 5.** Accuracy of AP estimation (internal RMS errors) with GSP-Spec for RVS spectra for selected AP ranges. Thin disk dwarfs are defined as  $\log g > 3.9$  dex and  $-0.5 < [\text{M}/\text{H}] < -0.25$  dex, thick disk dwarfs as  $\log g > 3.9$  and  $-1.5 < [\text{M}/\text{H}] < -0.5$  dex, and halo giants as  $4000 < T_{\text{eff}} < 6000$  K,  $\log g < 3.5$  dex and  $-2.5 < [\text{M}/\text{H}] < -1.25$  dex.

	$G_{\text{RVS}}$ mag	$T_{\text{eff}}$ K	$\log g$ dex	$[\text{M}/\text{H}]$ dex
Thin disk dwarfs	10	60	0.08	0.09
	13	70	0.12	0.09
	15	270	0.39	0.30
Thick disk dwarfs	10	70	0.11	0.09
	13	110	0.17	0.12
	15	350	0.43	0.29
Halo giants	10	70	0.17	0.15
	13	90	0.28	0.17
	15	340	0.86	0.38

Performance estimates for GSP-Spec are shown in Table 5.<sup>7</sup> The individual abundances of several elements (Fe, Ca, Ti, Si) will be measured for brighter stars. Based on experience with the Gaia-ESO survey (Gilmore et al. 2012), we expect to achieve an internal precision of 0.1 dex for  $G_{\text{RVS}} < 13$ .

The parameterization algorithms in GSP-Spec have been applied to real data. MATISSE and DEGAS were used in a study of the thick disk outside the solar neighbourhood (700 stars) (Kordopatis et al. 2011b) and were used in the upcoming final data release (DR4) of the RAVE Galactic Survey (228 060 spectra). These two applications share almost the same wavelength range and resolution as RVS. MATISSE is used in the AMBRE project (de Laverny et al. 2012) to determine the parameters  $T_{\text{eff}}$ ,  $\log g$ ,  $[\text{M}/\text{H}]$ , and  $[\alpha/\text{Fe}]$  of high resolution stellar spectra in the ESO archive (see Worley et al. 2012 and other forthcoming publications). MATISSE has also been used to characterize fields observed by CoRoT (Gazzano et al. 2013, Gazzano et al. 2010) and is one of the algorithms being used to characterize FGK stars in the Gaia-ESO survey.

### 5.4. Special treatment for emission line stars (ESP-ELS)

The ELS module classifies emission-line stars, presently into seven discrete classes: PNe (planetary nebulae), WC (Wolf-Rayet carbon), WN (Wolf-Rayet nitrogen), dMe, Herbig AeBe, Be, and unclassified. Since some types of emission line star may deserve further treatment in the ESP-HS or ESP-CS modules, ESP-ELS is the first of the ESP modules to be applied to the data. ESP-ELS is triggered by receiving a classification label of “star” or “quasar” from DSC (the latter included in order to accommodate misclassification in DSC). The algorithm works on several characteristic features in the BP/RP and/or RVS spectra. These are centered on (wavelengths in nm):  $\text{H}\alpha$ ,  $\text{H}\beta$ , P14,  $\text{He II } \lambda 468.6$ ,  $\text{C III } \lambda 866$ ,  $\text{C IV } \lambda 580.8$  and  $886$ ,  $\text{N IV } \lambda 710$ ,  $\text{O III } \lambda 500.7$ ,

<sup>7</sup> These results are based on a slightly broader RVS pass band extending to 874 nm. A recent change in the RVS filter has cut this down to 871 nm. This excludes the Mg lines, which may affect these results and others using RVS quoted in this article.

**Table 6.** Example of ESP-ELS classification performance in terms of a confusion matrix. The rows indicate the true classes, the columns the ESP-ELS assigned class. Each cell gives the percentage of objects classified from each true class to each ESP-ELS class. The dashes indicate exactly zero. Uncl. = Unclassified; Star = Star without emission.

True class	Output class							
	PNe	DMe	AeBe	Be	WC	WN	Uncl.	Star
PNe	63	–	–	–	–	–	28	9
DMe	–	60	–	–	–	–	–	40
AeBe	–	–	48	9	–	–	–	43
Be	–	–	5	41	–	–	–	54
WC	–	–	–	–	74	1	21	4
WN	–	–	–	–	1	73	18	8

and the Ca II triplet in RVS. For each of these features a spectroscopic index is defined which minimizes sensitivity to interstellar reddening and instrumental response. We use the index definition described in Cenarro et al. (2001). If significant emission is detected in one or more of the indices, the source is classified using one or more methods, including a neural network, k-nearest neighbours, and an interactive graphical analysis of the distribution of various combinations of indices in two-dimensional diagrams. In this last case, a comparison with the distribution of the indices for template objects is then used to manually define optimal classification boundaries.

The set of template indices was constructed both from synthetic spectra, mainly of non-emission stars and quasars, and from observed spectra of various types of emission line stars collected from public telescope archives and online catalogues, and supplemented with dedicated ground-based observations. The resulting spectral library comprises 1620 spectra of stars belonging to 12 different ELS classes (Be, WN, WC, dMe, RS CVn, Symbiotic, T Tauri, Herbig AeBe, Pre-MS, Carbon Mira, Novae, PNe) and observed between 320 and 920 nm (Lobel et al. 2010). The spectra were processed with GOG from which the indices described above were derived.

Typical results of our classification with this template set are shown in Table 6. The initial selection thresholds on these indices were set to avoid processing non-emission line stars or quasars, with the risk that certain weak emission line stars will be excluded. This conservative approach, combined with the limited resolution and sensitivity drop in the blue wing of the RP H $\alpha$  line, leads to not detecting about half of the Be and Herbig AeBe stars. Most of the other misclassifications and false detections are due to overlapping spectroscopic index values. Using Gaia observations of a predefined list of known emission line stars, we hope to be able to improve this performance and to expand the number of emission line star classes during the mission.

### 5.5. Special treatment for hot stars (ESP-HS)

Emission lines in hot (OB) stars will confuse AP estimation methods which assume the entire spectrum has a temperature-based origin in the photosphere. As emission lines are difficult to model reliably, the ESP-HS package attempts to improve the classification of hot stars by omitting those regions of the BP/RP and RVS spectra dominated by

**Table 7.** Maximum fractional AP residuals, i.e.  $|\text{measured-true}|/\text{true}$ , for the ESP-HS algorithm as a function of the G magnitude.  $N$  is the number of cases for each magnitude range.

G	$\Delta T_{\text{eff}}$	$\Delta \log g$	$\Delta A_0$	$N$
0–10	0.10	0.15	0.08	504
10–15	0.17	0.31	0.12	1154
15–18	0.25	0.40	0.55	1102

emission lines. The comparison with the template spectra over the selected regions is achieved with a minimum distance method using simplex minimization, while error bars are derived in a second iteration by computing the local covariance matrix. This approach is similar to the one used by Frémat et al. 2006. The spectral regions to omit are selected based on the results of ESP-ELS. If that module detects emission, then those regions most affected by emission (for that ELS class) are omitted, otherwise the full spectrum is used.

ESP-HS is applied to all stars previously classified by GSP-Phot or GSP-Spec as early-B and O-type stars (specifically  $T_{\text{eff}} > 14\,000$  K). RVS spectra are used for sources which have  $G_{\text{RVS}} < 12$ , otherwise only BP/RP spectra are used. This may be extended to fainter magnitudes during the mission depending on the quality of the RVS spectra. ESP-HS always estimates  $T_{\text{eff}}$ ,  $\log g$ , and  $A_0$ . Assuming the BP spectrum is available  $[\text{Fe}/\text{H}]$  is also estimated. If an RVS spectrum is available and  $v \sin i$  has not been estimated already by CU6, ESP-HS will derive this too.

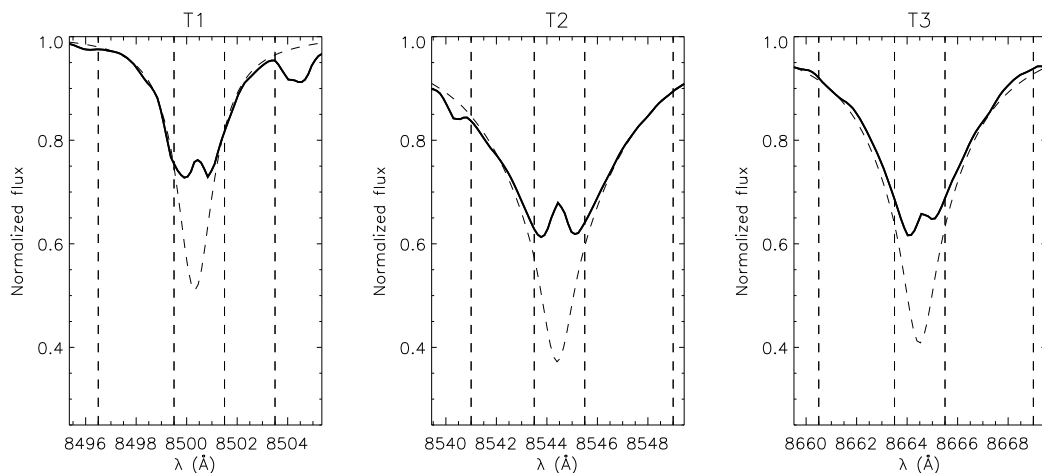
We applied the algorithm on spectra randomly spread over the parameter space ( $T_{\text{eff}} > 14\,000$  K). The maximum fractional residuals we found are given in Table 7. The current version of the algorithm is unable to correctly derive the APs from BP/RP for stars fainter than  $G = 18$ . During the mission, the results will be validated by comparing our derived APs to those obtained for a sample of reference O, B and A-type stars Lobel et al. (2013). Spectra for these are being collected and analysed as part of both the HERMES/MERCATOR project (Raskin et al. 2011) and the Gaia-ESO Survey.

### 5.6. Special treatment for cool stars (ESP-CS)

ESP-CS applies procedures for analysing peculiarities arising from magnetic activity and the presence of circumstellar material in stars with  $T_{\text{eff}}$  in the range 2500–7500 K.

Chromospheric activity can be detected via a fill-in of the Ca II infrared triplet lines in the RVS spectrum relative to the spectrum of an inactive star (Figure 7). Strong emission in the core of these lines for young stars generally indicates mass accretion. The degree of activity can be quantified by subtracting the spectrum of an inactive star from that of an active star with the same parameters, then measuring the difference in the central depth of a line,  $R_{\text{IRT}}$ , or in its integrated absorption,  $\Delta W$  (Busà et al. 2007). (These are the outputs from ESP-CS.) This is challenging because it requires knowledge of the star’s radial velocity, rotational velocity ( $v \sin i$ ),  $T_{\text{eff}}$ ,  $\log g$ , and  $[\text{Fe}/\text{H}]$ . NLTE effects must also be modelled (Andretta et al. 2005).

ESP-CS nominally adopts APs from GSP-Phot (and possibly GSP-Spec). However, these could be adversely affected by the star’s UV- or IR-excess from magnetic activity



**Fig. 7.** The three parts of the Ca II triplet of the active star HD82443 (solid line) observed at high resolution by TNG/SARG and processed by GOG to simulate an RVS spectrum at the noise level of a  $G = 8$  source. The dashed curve shows an NLTE synthetic spectrum for a pure photospheric contribution for comparison. Dashed vertical lines delimit the wings and the core regions.

and/or circumstellar material (and the GSP-Spec estimates could be distorted by high  $v \sin i$  in young active stars). For these reasons ESP-CS also estimates stellar APs itself, using  $\chi^2$  minimization against a set of templates of the Ca II wings. This takes into account rotational broadening and is unaffected by photometric excesses, so should provide a higher degree of internal consistency for the activity estimation. This approach is motivated by various studies in the literature showing that the wings of these lines are sensitive to all three stellar APs in some parts of the parameter space (Chmielewski 2000; Andretta et al. 2005). For metal rich dwarfs, the estimated activity level is actually not very sensitive to these stellar APs. Giants, in contrast, demand a higher accuracy of the APs, to better than 10% to get even a coarse estimate of chromospheric activity.

Based on existing  $R'_{\text{HK}}$  catalogues (e.g. Henry et al. 1996; Wright et al. 2004), the Besancon galaxy model (Robin et al. 2003), and the  $R_{\text{IRT}}$  and  $\Delta W$  correlations with  $R'_{\text{HK}}$  from Busà et al. (2007), we predict that Gaia will measure chromospheric activity to an accuracy of 10% in about 5000 main sequence field stars using  $\Delta W$ , and in about 10000 stars using  $R_{\text{IRT}}$ , which is some five times as many as existing activity measurements. We also expect to be able to measure activity in giants below the Linsky–Haish dividing line (Linsky & Haisch 1979), with the survey’s homogeneity being an added value for statistical studies. Finally, we expect to be able to identify very young low-mass stars in the field and in clusters down to  $G_{\text{RVS}} = 14$  via their accretion or chromospheric activity signatures in the Calcium triplet.

### 5.7. Special treatment for ultra-cool stars (ESP-UCD)

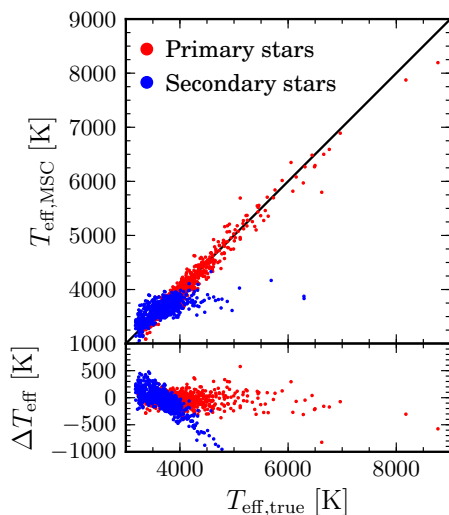
The ESP-UCD module provides physical parameters for the coolest stars observed by Gaia,  $T_{\text{eff}} < 2500$  K, hereafter referred to as ultra-cool dwarfs (UCDs) for brevity. The design of the module and its results are described in detail in Sarro et al. (2013), so we limit our description to the main features.

Based on empirical estimates of the local density of ultra-cool field dwarfs (Caballero et al. 2008), the BT-Settl family of synthetic models (Allard et al. 2012), and the Gaia instrument capabilities outlined in section 2, the expected number of Gaia detections per spectral type bin ranges from a few million at M5, down to a few thousand at L0, and several tens at L5. According to the Gaia pre-launch specifications, it should be possible to detect UCDs between L5 and late-T spectral types, although only 10–20 such sources are expected.

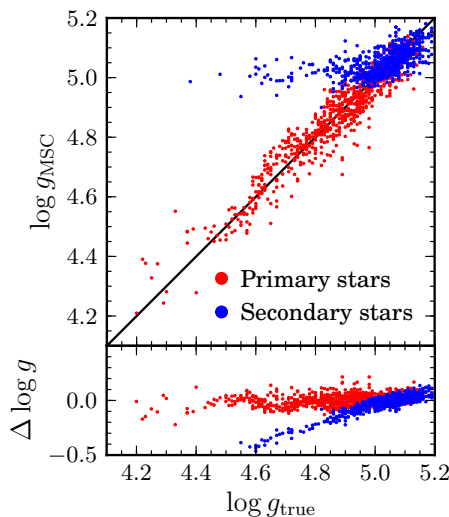
The ESP-UCD module comprises two stages: the select and process submodules. The select submodule identifies good UCD candidates for subsequent analysis by the process submodule. This selection is done according to predefined and non-conservative cuts in the proper motion, parallax,  $G$  magnitude, and  $G_{\text{BP}} - G_{\text{RP}}$  colour. The exact definition of the selection thresholds is based on the BT-Settl grid of synthetic models and is likely to change as a result of the internal validation of the module during the mission. In order to be complete at the hot boundary (2500 K) of the UCD domain, the module also selects stars which are hotter than this limit but fainter than the brightest UCD (since according to the models, a 2500 K low gravity star can be significantly brighter than hotter stars with higher gravities). Therefore, the ESP-UCD selection aims to be complete for sources brighter than  $G = 20$  and up to 2500 K, but will also contain sources between 2500 and approximately 2900 K. The module is actually trained on objects with  $T_{\text{eff}}$  up to 4000 K. During the data processing we may refine the selection and what we define as the hot boundary for the UCD definition.

The process submodule estimates  $T_{\text{eff}}$  and  $\log g$  from the RP spectrophotometry using three methods for each source:  $k$ -nearest neighbours (Cover & Hart 1967), Gaussian Processes (Rasmussen & Williams 2006), and Bayesian inference (Sivia & Skilling 2006). In all three cases the regression models are based on the relationship between physical parameters and observables defined by the BT-Settl library of models. ESP-UCD will provide all three estimates. The root-mean-square (RMS)  $T_{\text{eff}}$  error, estimated using spec-





**Fig. 8.** The performance of MSC at estimating  $T_{\text{eff}}$  for both components of an unresolved dwarf binary systems at  $G = 15$ . The upper panel shows the predicted vs. true  $T_{\text{eff}}$ , the lower the residual (predicted minus true) vs. true  $T_{\text{eff}}$ , for the primary component (red) and secondary (blue).



**Fig. 9.** As Figure 8, but for the surface gravity.

tra of real UCDs obtained with ground-based telescopes and simulated to look like RP spectra at  $G = 20$ , is 210 K. The lack of estimates of  $\log g$  for this set of ground-based spectra of UCDs prevents us from estimating the RMS error for this parameter, but the internal cross-validation experiments show an RMS error of 0.5 dex in  $\log g$ , which is probably an underestimate of the real uncertainty.

### 5.8. Multiple Star Classifier (MSC)

MSC uses the BP/RP spectrum to estimate the APs of sources identified as unresolved physical binaries by DSC. Currently it estimates  $A_0$ ,  $[\text{Fe}/\text{H}]$ , and the brightness ratio  $\text{BR} = \log_{10}(L_1/L_2)$ , of the system as a whole ( $L_1$  and  $L_2$  are the component luminosities), as well as the effective temper-

**Table 8.** Summary of MSC performance in terms of the RMS, median absolute residuals (MAR), and median residual (MR, as a measure of systematic errors). The subscripts 1 and 2 denote the primary and the secondary components respectively. The apparently good performance for the secondary component is mostly an artefact of the narrow distribution of the APs in our data set.

AP	RMS	MAR	MR
$T_{\text{eff},1}$	130	90	70
$T_{\text{eff},2}$	260	160	110
$\log g_1$	0.05	0.04	0.03
$\log g_2$	0.10	0.06	0.03
$[\text{Fe}/\text{H}]$	0.14	0.10	0.07
$A_0$	0.11	0.09	0.07
$BR$	0.36	0.26	0.19

ature and surface gravities of the two components individually. The APs are estimated using SVMs, using the same SVM implementation used for GSP-Phot (section 5.2). For more details see Tsalmantza & Bailer-Jones (2010a) and Tsalmantza & Bailer-Jones (2012).

The training and testing data for MSC are constructed by combining spectra for single stars into physically plausible binaries (Lanzafame 2007, private communication), and simulating them with GOG (Sordo & Vallenari 2013). A system age and metallicity is selected at random and masses are drawn from a Kroupa IMF and paired randomly. The corresponding atmospheric parameters are identified using Padova isochrone models. We then use the MARCS spectra (section 4) to represent the individual stars with the closest corresponding parameters.

The performance of MSC depends not only on the magnitude of the system, but also the brightness ratio of the two components. MSC has been trained on a data set of dwarf-dwarf binaries at  $G = 15$  with an exponentially decreasing distribution in BR from 0 to 5, such that the majority have  $\text{BR} < 2$ . Yet even at  $\text{BR} = 2$  the secondary is a hundred times fainter than the primary, so we should not expect good average performance on the secondary component.

Here we report results on a test data set limited to  $\text{BR} < 1.5$ . Figure 8 shows the  $T_{\text{eff}}$  residuals for both components as a function of  $T_{\text{eff}}$ . We see that we can predict  $T_{\text{eff}}$  for the primary star quite accurately, but not for the secondary star, despite the apparently good summary statistics given in Table 8. Figure 8 shows that for the secondary star, the estimated  $T_{\text{eff}}$  correlates poorly with the true  $T_{\text{eff}}$ . This is because the SVM is hardly able to learn the generally weak signature of  $T_{\text{eff},2}$  from the data, so assigns essentially random values from the training data distribution. The RMS is then relatively small simply because the spread in the  $T_{\text{eff},2}$  in the training data is also small.<sup>8</sup> This narrow  $T_{\text{eff},2}$  distribution is a consequence of the way the binary systems were constructed.

We see a similar problem with the determination of the surface gravity of the secondary in Figure 9. This is not at all surprising, because this is anyway a weak parameter. In contrast,  $\log g$  for the primary can be determined quite accurately (no significant systematics) despite the interfering

<sup>8</sup> This serves as a reminder that performance as measured by RMS residuals should be judged in comparison to the standard deviation in the data.



spectrum of the secondary. This is partly a consequence of a correlation between  $T_{\text{eff}}$  and  $\log g$  in our data set. On the other hand, when such correlations are real, they should be exploited.

If we extend the test data set to include systems with larger BR, then the accuracy of both components degrades. Conversely, limiting it to smaller BRs produces better average performance. Although the performance on the secondary components is relatively poor, MSC is nonetheless a useful algorithm because it gives a better performance on the primary components than does GSP-Phot (Tsalmantza & Bailer-Jones 2010a). That is, neglecting the existence of the secondary degrades the accuracy with which we can estimate the APs of the primary.

MSC will report statistical uncertainties on the predicted APs obtained from the residuals on a test data set: Given the predictions of the APs of an unknown object, we use a look-up table to find the typical errors obtained on a test set around those measured APs.

### 5.9. Stellar mass and age estimation (FLAME)

Using the atmospheric AP estimates from GSP-Phot, it is possible to infer, to a great or lesser accuracy, the fundamental stellar parameters of age and mass. This is the task of FLAME. First, the luminosity of each single star will be computed from the Gaia parallaxes and magnitudes, and suitable bolometric corrections calculated from the  $T_{\text{eff}}$ . The luminosity,  $T_{\text{eff}}$  and  $[\text{Fe}/\text{H}]$  estimates place the object in the Hertzsprung-Russell diagram. Comparison of this position with stellar evolutionary tracks for a range of masses and abundances then allows us to estimate the mass, age and metallicity fraction,  $Z$ , of the star through an inversion method. The initial helium abundance in mass fraction  $Y$  is assumed to follow the helium to metal enrichment law  $\frac{\Delta Y}{\Delta Z} = 1.45$ , based on the solar model calibration and a primordial helium abundance of  $Y = 0.245$ . Two different algorithms are used for the inversion (results will be provided from both): a classical  $\chi^2$  minimization algorithm (Ng & Bertelli 1998), and a forward modelling Bayesian method (Pont & Eyer 2004) using prior information on the initial mass function (only for age determination), stellar formation rate and metallicity distribution function. From the results the stellar radius can be calculated. This is important for correcting the zero point radial velocity of stars for the gravitational redshift.

When available, AP estimates (and additional abundance information) from GSP-Spec or ESP could be used by FLAME also.

For stars of A type and later, a 1% error in  $T_{\text{eff}}$  translates to a similar error in the mass determination. Age is more sensitive due in part to the degeneracy of evolutionary tracks in the HRD, so will have errors of at least 10%, or even 100% in the worst case. This all assumes fixed chemical composition, so the uncertainties will increase when the  $[\text{Fe}/\text{H}]$  uncertainty is taken into account.

### 5.10. Galaxy classification (UGC)

We expect to observe a few million unresolved galaxies with Gaia. UGC will use the BP/RP spectra to classify them into discrete classes and to estimate the redshift, the Galactic extinction, and parameters which determine the star for-

**Table 9.** Example of UGC AP estimation performance in terms of the RMS residual for sources at three different magnitudes. The results from the final step of a two-step approach are given.

AP	RMS at $G=$		
	15	18.5	20
$A_0$ / mag	0.04	0.15	0.35
redshift	0.002	0.011	0.028

**Table 10.** Example of UGC classification performance in terms of the true positive classification percentage. Results for sources of fixed magnitude as well as for a range of magnitudes are shown.

Galaxy type	True positive percentage at $G=$					
	15	13–16.5	18.5	16.5–19	19–20	20
early	93	91	78	79	50	44
spiral	98	95	90	92	73	64
irregular	89	86	51	54	26	28
quenched	99	98	94	94	83	83

mation law in the source galaxy. We use SVMs for both classification and parameter estimation, the former giving probabilities for each of the galaxy classes. The SVMs are trained on simulated BP/RP spectra generated from the Galaxy spectral libraries described in section 4.2, which also defines the four galaxy types.

UGC comprises two separate modules (Bellas-Velidis et al. 2012). The first, UGC-Learn, provides SVM tuning, training and testing functions for offline preparation of the models. A number of SVM models have been trained, arranged in a two-layer hierarchy. For each of three  $G$  magnitude ranges (13–16.5, 16.5–19, 19–20), an SVM in the first layer is trained to cover the total range of the extinction and redshift parameters. In the second layer, there is a set of SVM models, each dedicated to a narrower range of these two parameters (again for each magnitude range).

The second module, UGC-Apply, applies the fitted SVM models in a hierarchical manner, in two steps. It operates on sources identified by DSC as having a galaxy probability above a predefined threshold (section 5.1). In the first step, the “total range” SVM for the appropriate source magnitude is applied to provide an initial estimate of the redshift and extinction. In the second step, “specific range” SVM models corresponding to these initial parameter estimates are used to classify the spectrum and to estimate the parameters. The star formation parameters for all the galaxy types are estimated, independently of the best class predicted.

UGC shows good performance in predicting the extinction (range 0–6 mag) and redshift (range 0.0–0.2), as can be seen in Table 9. The residuals show no trend with the parameters. The classification performance is measured as the percentage of true-positive classifications, and is shown in Table 10. At all magnitudes the best performance is obtained on the spirals and quenched star formation galaxies. At  $G = 20$  the true-positive rate for the early and irregular classes is below 50%, and we are not able to estimate the star formation parameters. For the two brighter magni-

tude ranges the best performance is achieved for the early type parameters. For irregular and quenched star formation galaxy types the gas infall rate cannot be accurately predicted, even at  $G = 15$ . Applying the hierarchical approach to these parameters may help. Performance could be further increased by improving the libraries, representing more realistically the observed spectral types, eliminating overlaps among galaxy types, and providing larger numbers of spectra for the different training and testing data sets. We also plan to investigate using the total Galactic extinction estimates from TGE (section 5.12).

### 5.11. Quasar classification (QSOC)

QSOC processes the 500 000 or so quasars which we expect Gaia to observe. It has two goals. First, it classifies each quasar into the three classes type I, type II and BAL (broad absorption line) quasars. This is achieved using a standard SVM classifier. Second, QSOC estimates the redshift, total emission line equivalent width, and the slope of the power law continuum. These three APs are estimated using an ensemble of trees based on the Extremely Randomized Tree (ERT) algorithm (Geurts et al. 2006). The redshift is also estimated using an SVM in classification mode, in which each redshift bin of width 0.1 in redshift from  $z=0$  to 5.6 is represented as a separate class and assigned a probability. (It is done this way to search for possible multimodality in redshift estimation.) The SVM and ERT redshift estimates are then combined to give a single redshift estimate. The models are trained using the quasar spectral libraries described in section 4.

The module has been evaluated using K-fold cross validation. The results reported here are for training and testing with the semi-empirical library. The SVM classifier achieves an accuracy of 97.4%, 95.8%, and 91% for  $G = 15, 18.5,$  and  $20$  (respectively), for those quasars where the SVM redshift classifier gives a probability above some threshold in a single bin (i.e. high confidence). The AP predictions from each regression tree in the ERT ensemble could be combined in several different ways. While the mean minimizes the RMS, the median minimizes the mean absolute deviation and reduces the bias. We instead calculate a discretized mode: we form a histogram of the estimates (with an adapted bin width), and report the central value of the highest bin as our AP estimate. We find that this gives a higher accuracy than either the mean or the median. The redshift accuracy obtained in this way is 0.02, 0.03, and 0.04 for  $G = 15, 18.5,$  and  $20$  respectively. We observe some systematically wrong redshift estimates due to degeneracies in the identification of the rest wavelengths of pairs of strong emission lines, as well as some minor biases at small redshifts caused by the zero limit of the redshifts.

So far we do not try to deal with misclassifications from DSC, e.g. contamination from Be or WR stars (although by using parallaxes and proper motions, DSC attempts to minimize such misclassifications). We also do not take into account interstellar extinction (the training data, from SDSS DR9, are at high Galactic latitude and assumed to be extinction free), so QSOC may not produce reliable results at low latitudes. Part of the on-going development will be to use extinction estimates from TGE to overcome this limitation.

### 5.12. Total Galactic Extinction (TGE)

When estimating the APs of objects from their broad band spectra, it is important to take into account the impact of interstellar extinction. Some algorithms, such as GSP-Phot, estimate this for each star independently, effectively treating the extinction as an additional stellar parameter. But this approach works less well for some types of object, in particular quasars. The role of TGE is to estimate the total Galactic extinction (the extinction integrated to the edge of the Galaxy) towards an extragalactic source. It does this by combining the individual extinction estimates from GSP-Phot for distant stars (small parallaxes) in the source's direction.

As part of the DPAC processing, Gaia sources will be indexed according to their position on the sky using the Hierarchical Equal Area isoLatitude Pixelisation (HEALPix) scheme (Górski et al. 2005). This hierarchically partitions the celestial sphere into ever smaller levels, or HEALpixels. Apsis processes sources in blocks of data corresponding to level 6 HEALpixels, each of which covers 0.839 square degrees. Given all the data in one such HEALPix, TGE first selects stars to use as extinction tracers. It then uses the estimated extinctions and parallaxes of these tracers to derive the total Galactic extinction for that HEALPix. This is repeated for all HEALpixels over the whole sky. In regions where there are more tracers, a higher HEALPix level can be used in order to achieve a higher angular resolution map.

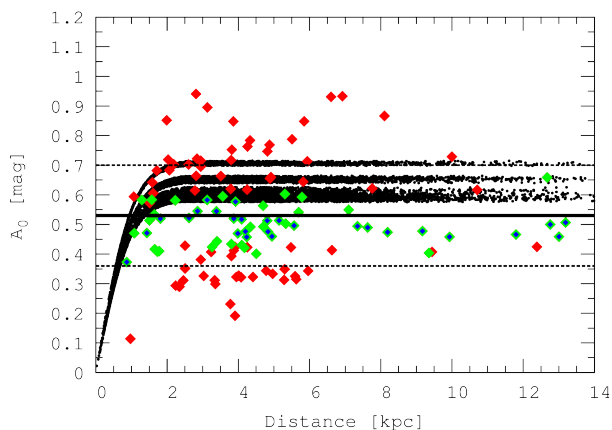
Candidate tracers are selected to be single stars (using the DSC probability) and non-variable (using indices from the photometric processing in CU5). Those with the most precise  $A_0$  estimates are selected based on the expected performance of GSP-Phot in different parts of the AP space and on the individual  $A_0$  uncertainty estimates provided by GSP-Phot. Finally, a parallax selection criterion is applied to the candidate tracers in order to select just those which are sufficiently far from the main gas and dust layer in the Galactic plane.

For the estimate of the total Galactic extinction for that HEALpixel,  $A_{TGE}$ , we report the mean  $A_0$  value of the selected tracers. The uncertainty in this we represent with the RMS of the  $A_0$  values. We are exploring the use of other estimators which enable more robust estimates of  $A_{TGE}$ , in particular at low Galactic latitudes. On account of the very high extinction in some fields at low latitudes, there will be insufficient distant tracers for TGE to make a reliable estimate of  $A_{TGE}$  (although we also then expect to detect fewer extragalactic objects on account of this same extinction).

As an example of how TGE works, Figure 10 shows the estimated  $A_0$  for stars and selected tracers in a particular HEALpixel (from Aeneas in GSP-Phot), using simulated Galactic data, and the value of  $A_{TGE}$  estimated from these. In addition to TGE providing extinction estimates for use in QSOC (and possibly UGC), it will also provide a two-dimensional map of the total extinction for most of the Galaxy, unique in that it will be derived from the individual extinction estimates of stars with measured parallax.

### 5.13. Outlier Analysis (OA)

Like any supervised classification algorithm, DSC can only reliably classify objects which are modelled accurately in its training set. By design, objects which do not match its

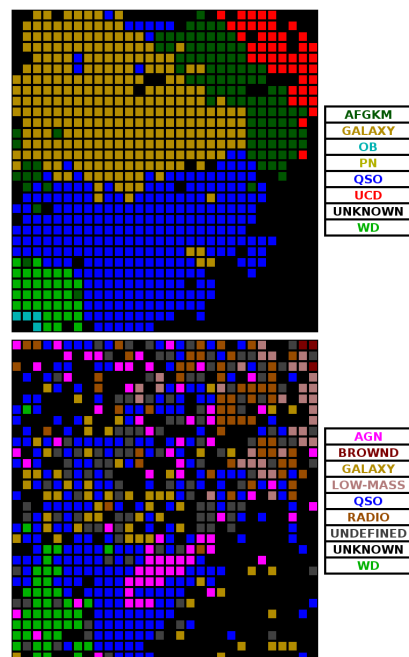


**Fig. 10.** Example of TGE extinction estimate in a simulated Galactic field at  $(l, b) = 60^\circ, 10^\circ$ . The black points show the true extinction,  $A_0$ , for all stars in the field, while the diamonds show the estimated extinction from GSP-Phot for the selected tracers, coloured according to the magnitude of the uncertainty of the  $A_0$  estimates: red for  $|\Delta A_0| > 0.1$ , green for  $|\Delta A_0| < 0.1$ , and blue for  $|\Delta A_0| < 0.05$  mag. The solid horizontal line shows the estimated  $A_{\text{TGE}}$  value and the dashed lines the  $\pm 1\sigma$  RMS.

training set achieve low probabilities for all classes and will therefore be labelled as unknowns, or outliers. These could be types of objects omitted from the DSC training data entirely, objects with poor spectral models, or instrumental artefacts not well modelled by GOG. They could also be previously unseen types of object. We estimate that 5% or more of the Gaia sources will be marked as outliers by DSC – more than 50 million objects – so some kind of automated analysis of these is mandatory. This is the task of OA. Its main purpose is to help improve the source and instrument modelling and thereby improve the training data sets during the mission.

OA uses a Self-Organizing Map (SOM, Kohonen et al. 2001). This projects the original data (BP/RP and astrometry) into a 2D grid of nodes in a way that attempts to preserve local topology, thereby clustering together similar objects which may be systematically rejected by DSC (Fustes et al. 2013b). Then follows an identification stage, where we try to discover whether any other known types of source are associated with any of these clusters. This stage could make use of data from other surveys and catalogues.

In order to study the behaviour of our algorithm with a realistic dataset, we compiled a semi-empirical BP/RP library from spectra that were classified as “unknown” by the SDSS spectroscopic classification pipeline. This dataset comprises 10 125 objects, which are mostly faint objects, incomplete spectra, or the result of a poor fibre alignment. We fit a SOM with  $30 \times 30$  nodes to these. As a first method of identification, we applied a k-nearest neighbours classifier on labelled objects from the simulated Gaia data (described in section 4). Observing where these tend to land in the SOM, we used this to label the SOM nodes. The results of this are shown in the upper panel of Figure 11. (Note that there is no physical meaning to the axes, size or shape of the SOM.) In a second method, we identified from the Simbad database the nearest object on the sky to each SDSS object and retrieved its Simbad class, if available (which is the case for about 3000 of the SDSS objects). If a large enough



**Fig. 11.** Identifications of SDSS outliers using the SOM in the OA module, obtained from Gaia simulations (top) and Simbad (bottom).

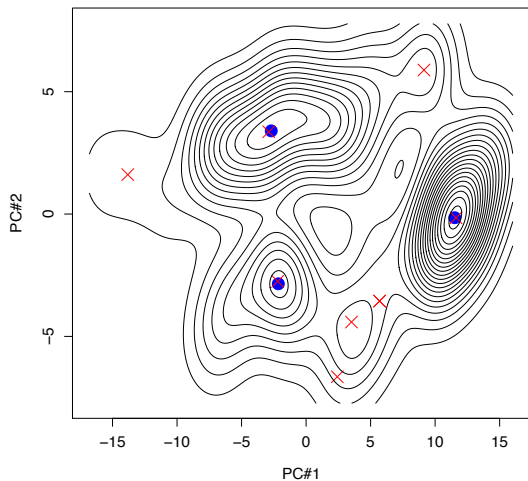
fraction of objects in a single SOM node share the same class, we label that node with that class, as shown in the lower panel of Figure 11. Otherwise the node is labelled “unknown”. Comparing the two maps, we see that quasars and white dwarfs are identified in similar regions in both cases, giving some confidence that these classifications are appropriate. Using this approach, we were able to identify 400 white dwarf candidates, 1000 quasar candidates, and 16 brown dwarf candidates from among the SDSS outliers. For further details, see Fustes et al. (2013a).

#### 5.14. Unsupervised clustering (OCA)

The supervised learning modules used in Apsis can only assign meaningful parameters to the types of objects they have been trained on. Inevitably, not all types of objects which Gaia will encounter are covered by the Apsis training sets. The aim of the Object Clustering Algorithm (OCA) is to identify some natural groups among the Gaia sources, independently of labelled training data sets. This can be used to help improve the training sets in a similar manner to OA, but may also identify potentially new types of object which can then be studied further.<sup>9</sup>

OCA implements a variant of the Hierarchical Mode Association Clustering algorithm (HMAC, Li et al. 2007), which analyses the density of sources in the multidimensional space formed by the data (here BP/RP, astrometry and, when available, RVS). In this framework, each individual source is associated to the closest mode (maximum) of the probability density landscape. Rather than explicitly computing the probability density, the Modal Expectation-Maximisation (MEM) algorithm is used to assign sources

<sup>9</sup> We use the standard term “natural groups” here, but in fact there is no such thing. The clustering found by any algorithm is determined by the similarity measure – or distance metric – we choose to adopt.



**Fig. 12.** Simplified example of the clustering performed by OCA to identify modes of density, shown in the space formed by the first two principal components. The black lines are the contours of the data density. The blue circles and red crosses correspond to the modes of the large (0.025) and small (0.015) kernel sizes respectively. At the lower resolution (blue circles), the three cluster modes obtained correspond to clusters dominated by stars, quasars and galaxies, although with significant contamination from other types.

to modes. This works by climbing to a local maximum in an iterative fashion. It is almost equivalent to assigning the source to the closest mode in a kernel density estimation (Wand & Jones 1995). The algorithm uses Gaussian kernels to find the modes, and it becomes hierarchical when we increase the size of the covariance matrix of these kernels, such that modes (and their associated sources) are merged into new modes/clusters at higher levels of the hierarchy.

The computational complexity of HMAc is quadratic in the number of sources, which is up to  $10^9$  in the case of Gaia. In order to fit within computational time and memory limitations, we use a divide-and-conquer strategy. The full sample of sources to be clustered is partitioned into disjoint subsamples corresponding to different HEALpixels on the celestial sphere. HMAc is applied to each of these subsamples in order to identify modes (cluster representatives). These modes are assigned a weight proportional to the number of sources that converged to it (the cluster size). An iterative process is then used to merge the modes across subsamples.

OCA has been tested on the semi-empirical SDSS star, galaxy and quasar libraries described in section 4 for sources with a range of magnitudes. The BP/RP spectra were normalized to unit area, and only the first 15 principal components (PCs) in each of RP and BP were used, accounting for 99% of the variance. We also included the first four moments of the BP and RP flux distributions, scaled to the range of values of the first PC. Figure 12 illustrates a simplified case of clustering for two kernel sizes and a reduced dataset.

## 6. Validation, calibration, and in-mission development

The data processing phase of the mission comprises three vital tasks beyond applying Apsis to the Gaia data. These are validation, calibration, and in-mission software development. (The subsequent task of catalogue production is not covered in this article.) These we now discuss in turn.

Apsis will produce an enormous set of AP estimates on many different types of objects. A critical assessment of these results is an important part of the data processing. We refer to this as validation, and it will take place in two ways. First, an internal validation examines (for example) the distributions of estimated parameters, their uncertainties, and correlations between them, and whether these agree with our expectations. For example, do we get an inordinately large number of low metallicity stars, or do we find unexpected correlations, such as  $T_{\text{eff}}$  increasing with  $A_0$ ?<sup>10</sup> Such analyses may allow us to identify problems and thus improve the training data and algorithms. The internal validation will also compare the AP estimates for common objects between the different modules in Apsis. Second, an external validation compares our AP estimates with external AP estimates, either for individual objects or for populations of objects. An example of the latter is to construct the HRD of known clusters, or compare metallicity or redshift distributions with published estimates. While we would not automatically take non-Gaia estimates as being true, systematic differences between our and non-Gaia estimates may be indicative of problems.

The Apsis algorithms have been developed over the past years using simulated Gaia data. The real Gaia data will of course differ from these. In particular, the response function and noise properties of the detectors may differ from expectations, and these will anyway evolve in unpredictable ways during the mission due to the progressive radiation damage of the CCDs. Upstream data processing tasks may also need to change the way they process the data, producing data with different properties. The result is that the shape and noise properties of the spectra are likely to deviate from our current simulations. Furthermore, the Apsis algorithms make extensive use of synthetic spectra for training. These differ from real spectra because of the approximations involved in modelling astrophysical sources.<sup>11</sup> These two issues – spectral simulation and instrument simulation – result in imperfectly modelled Gaia spectra, something we refer to as the “spectral mismatch problem”. As supervised algorithms depend on a match between their training data and the observed data, it is important that we accommodate these changes. It is the goal of the calibration of the Apsis algorithms to correct for this.

A calibration can be achieved by applying corrections either to the training data before it is used, or to the APs produced by the estimation algorithm. In the first approach,

<sup>10</sup> There is a known, strong degeneracy in the spectra between  $T_{\text{eff}}$  and  $A_0$  for individual stars (e.g. Bailer-Jones 2010; Liu et al. 2012) which we attempt to account for, but is not of physical origin.

<sup>11</sup> An example mismatch affecting RVS is incorrect broadening parameters and assumptions, which affects in particular the Paschen lines of stars with  $T_{\text{eff}}$  around 10 000 K (Fremat et al. 1996). In BP/RP, the largest effect is expected to arise for the cooler stars due to incomplete and poor molecular data, as shown for example by Plez (2011) and Lebzelter et al. (2012).

we use Gaia observations of labelled reference objects (i.e. with known APs) to modify the fluxes of the synthetic spectra, thereby producing a hybrid synthetic–real grid which is used for algorithm training. One specific idea is to use the denser synthetic spectral grids to model the small scale variations of fluxes with APs, and the sparser observed spectral grids to model the larger scale variations (Bailer-Jones 2010). This follows the assumption that synthetic spectra reproduce flux changes better than absolute fluxes. The feasibility of this approach is under investigation.<sup>12</sup> In the second approach, we instead model the AP deviations as a function of the main parameters. Although simpler, it is probably less accurate due to the loss of information from working with “faulty” spectra in the first place, so this is not being pursued. Both approaches require that we obtain accurate APs by independent methods for a set of reference objects which Gaia will observe. This is being done explicitly for Gaia using ground-based higher resolution spectra, as discussed in section 6.1.

Another aspect of the calibration work is to improve the representation of the synthetic stellar spectra in the first place. Two routes are currently being followed to improve the spectra of FGK stars. The first is to move from classical 1D stellar atmospheres to 3D radiation-hydrodynamics simulations in order to better represent the effects of convection (Chiavassa et al. 2011; Allende Prieto et al. 2013; Magic et al. 2013). Second, deviations from LTE will be taken into account by implementing the results of detailed statistical equilibrium calculations into the spectrum synthesis codes. This will be particularly important for modelling the calcium triplet lines dominating the RVS spectra (Mashonkina et al. 2007; Merle et al. 2011).

In addition to calibration and validation, we expect to have to adapt, during the data processing, how we use the Apsis algorithms. For example, we will inevitably have to modify our nominal strategy of which APs we attempt to estimate for which types of object at which magnitudes. We may even find that we need to modify or change algorithms, or introduce new algorithms to deal with additional classes of object. Indeed, we fully expect to have to modify and extend our spectral libraries to accommodate missing classes of objects or poorly modelled classes. We will also update the model of the Gaia instruments to match their in-flight properties as closely as possible, in order to produce more accurate training data sets. All of this will demand a continued software development and data simulation during the mission.

### 6.1. High resolution spectral observations for Apsis stellar calibration

To perform the calibrations described above, we need independent AP estimates of several thousand Gaia targets.

A two-level procedure for those algorithms which estimate  $T_{\text{eff}}$ ,  $\log g$ , and  $[\text{Fe}/\text{H}]$  is foreseen. At the first level, we define a set of benchmark stars made up of a small number of carefully selected, well-studied bright stars (around 40 FGKM and 20 OBA stars). Their Hipparcos parallaxes, angular diameters and bolometric fluxes are known, and

their masses have been determined in a homogeneous way, so their effective temperatures and surface gravities can be derived independently of spectroscopy. Reference metallicities for benchmark stars are determined from ground-based high-resolution spectra (see below) using several different methods. Details on the parameters and data for cool benchmark stars will be published in a series of forthcoming papers (and already in Lebzelter et al. 2012). At the second level, we define a much larger set of several hundred reference stars covering the AP space more densely than the benchmark stars. Homogeneous APs for these stars are being determined from high-resolution spectroscopy and calibrated to the benchmark stars.

The necessary high-resolution spectra are being obtained in various observing programs. OBA stars have been observed with the HERMES spectrograph on the Mercator Telescope in Spain. These will be supplemented with medium-resolution OBA (cluster) star spectra observed with VLT-Giraffe as part of the Gaia–ESO public spectroscopic survey. The FGKM stars have been observed with the NARVAL spectropolarimeter on the 2m Bernard Lyot Telescope at Pic du Midi in spectroscopic mode. High-resolution spectra of M dwarfs have been obtained in the infrared J-band with the CRIRES spectrograph (Önehag et al. 2012). High quality spectra are also retrieved from the various public archives. So far the library comprises 79 spectra of 35 cool benchmark stars from NARVAL, UVES and HARPS observations, with resolutions greater than 70 000 and SNRs greater than 200. The benchmark star spectra will be published online in the SPECTROWEB database<sup>13</sup> and the reference star spectra in the HHIGH-RESPECT database, currently under development.

## 7. Outlook

We have described the status of the Gaia classification system at the time of launch, prior to seeing any real data. During the course of the five year mission and the subsequent two or three years of processing before the final data release, this system will continue to evolve in light of the experience we gain with the data. Indeed, we anticipate substantial developments, which we will report in future publications.

Our classification approach involves a combination of supervised and unsupervised algorithms. The former are critically dependent on an accurate representation of the target sources, and our training data sets will need considerable optimization during the data processing phase. This will involve improvements to the simulations as well as the use of ground-based data to calibrate our training data.

The system developed so far makes some idealized assumptions about the data and the upstream spectral processing by the other CUs. For example, the exact impact of radiation damage on the CCDs and therefore on the combined BP/RP and RVS spectra is hard to model. Being a slitless spectrograph, some BP/RP spectra will overlap (and likewise for RVS). Although this is accommodated in the spectral extraction, imperfect removal of overlap will leave systematic residuals. Analysing the real data will be an important learning experience.

It should be appreciated that the main objective of Apsis is to provide reasonably accurate parameter estimates

<sup>12</sup> Ideally we would just use Gaia observations of labelled objects as the complete training data sets and dispense entirely with modifying synthetic grids, but a sufficient set of labelled objects does not exist for this purpose.

<sup>13</sup> <http://spectra.freeshell.org/spectroweb.html>



for a broad class of objects covering a large fraction of the catalogue. We do not aim to do everything possible. For example, while we try to identify white dwarfs, we do not (yet) attempt to estimate their parameters. We are likewise aware that almost any narrow class of objects could be given more targeted treatment, which may result in more accurate AP estimates through, for example, a more focused use of the data or by adopting different source models. A combination of Gaia data with non-Gaia data will be particularly beneficial for some classes of object, but this is beyond the remit of CU8. Some such work is planned by the DPAC within CU9 (responsible for the data releases), and the resulting hybrid catalogues would be published in the data releases. We note finally that we hope to publish not only the results from Apsis, but also the software, to enable the scientific community to obtain their own AP estimates with their own training data sets, for instance. There is no natural divide between “data processing” and “scientific analysis”, and we hope that in the course of exploiting the Gaia data the community will take up the challenge to extend and to improve our work.

*Acknowledgements.* We acknowledge the contributions of our former colleagues who were involved in earlier stages of this work. We remember in particular the contributions of our prematurely deceased colleague Joël Poels. We are grateful to the members of the Gaia DPAC CU2 for the development of the GOG simulator. Many of the GOG simulations used in this work were produced by CU2 using the supercomputer MareNostrum at the Barcelona Supercomputing Center – Centro Nacional de Supercomputación. Parts of this work have been funded by various funding agencies including: DLR (German space agency), grant 50 QG 1001; CNES (French space agency); ASI, under contract I/058/10/0 (Gaia Mission - The Italian Participation in DPAC); Spanish Ministry of Science and Innovation, through grants AyA2009-14648-C02-02 and AyA2012-39551-C02-02; Spanish Ministry of Economy, through grants AyA2008-02156 and AyA2011-24052; Swedish National Space Board; ESA PRODEX programme ExSciGaia (contract 4000106135); Belgian Federal Science Policy Office in connection with the ESA PRODEX programmes ‘Gaia-DPAC QSOs’ and ‘Binaries, extreme stars and solar system objects’ (contract C90290). SIGAMM computing centre of the Observatoire de la Côte d’Azur.

## References

- Allard, F., Hauschildt, P. H., Alexander, D. R., Tamanai, A., & Schweitzer, A. 2001, *ApJ*, 556, 357
- Allard, F., Homeier, D., & Freytag, B. 2012, *Royal Society of London Philosophical Transactions Series A*, 370, 2765
- Allende Prieto, C., Koesterke, L., Ludwig, H.-G., Freytag, B., & Caffau, E. 2013, *A&A*, 550, A103
- Andretta, V., Busà, I., Gomez, M. T., & Terranegra, L. 2005, *A&A*, 430, 669
- Bailer-Jones, C. A. L. 2009, in *IAU Symposium*, Vol. 254, *IAU Symposium*, ed. J. Andersen, B. Nordström, & J. Bland-Hawthorn, 475–482
- Bailer-Jones, C. A. L. 2010, *GAIA-C8-TN-MPIA-CBJ-044*, Tech. rep., Max Planck Institute for Astronomy, Heidelberg
- Bailer-Jones, C. A. L. 2010, *MNRAS*, 403, 96
- Bailer-Jones, C. A. L. 2011, *MNRAS*, 411, 435
- Bailer-Jones, C. A. L., Smith, K. W., Tiede, C., Sordo, R., & Vallenari, A. 2008, *MNRAS*, 391, 1838
- Bellas-Velidis, I., Kontizas, M., Dapergolas, A., et al. 2012, *Bulgarian Astronomical Journal*, 18, 020000
- Bertelli, G., Girardi, L., Marigo, P., & Nasi, E. 2008, *A&A*, 484, 815
- Bijaoui, A., Recio-Blanco, A., de Laverny, P., & Ordenovic, C. 2012, *Statistical Methodology*, 9, 55
- Bouret, J., Lanz, T., Frémat, Y., et al. 2008, in *Revista Mexicana de Astronomía y Astrofísica Conference Series*, Vol. 33, *Revista Mexicana de Astronomía y Astrofísica Conference Series*, 50–50
- Brott, I. & Hauschildt, P. H. 2005, in *ESA Special Publication*, Vol. 576, *The Three-Dimensional Universe with Gaia*, ed. C. Turon, K. S. O’Flaherty, & M. A. C. Perryman, 565–+
- Burges, C. J. C. 1998, *Data Min. Knowl. Discov.*, 2, 121
- Busà, I., Aznar Cuadrado, R., Terranegra, L., Andretta, V., & Gomez, M. T. 2007, *A&A*, 466, 1089
- Caballero, J. A., Burgasser, A. J., & Klement, R. 2008, *A&A*, 488, 181
- Cardelli, J. A., Clayton, G. C., & Mathis, J. S. 1989, *ApJ*, 345, 245
- Casertano, S., Lattanzi, M. G., Sozzetti, A., et al. 2008, *A&A*, 482, 699
- Castanheira, B. G., Kepler, S. O., Handler, G., & Koester, D. 2006, *A&A*, 450, 331
- Cenarro, A. J., Cardiel, N., Gorgas, J., et al. 2001, *MNRAS*, 326, 959
- Chang, C.-C. & Lin, C.-J. 2011, *ACM Transactions on Intelligent Systems and Technology*, 2, 27:1, software available at <http://www.csie.ntu.edu.tw/~cjlin/libsvm>
- Chiavassa, A., Bigot, L., Thévenin, F., et al. 2011, *Journal of Physics Conference Series*, 328, 012012
- Chmielewski, Y. 2000, *A&A*, 353, 666
- Cortes, C. & Vapnik, V. 1995, *Machine Learning*, 20, 273
- Cover, T. M. & Hart, P. E. 1967, *IEEE Transactions on Information Theory*, 13, 21
- Cropper, M. & Katz, D. 2011, in *EAS Publications Series*, Vol. 45, *EAS Publications Series*, 181–188
- De Angeli, F., Evans, D. W., & Brown, S. W. 2012, *GAIA-C5-TN-IOA-FDA-029*, Tech. rep., Data Processing and Analysis Consortium
- de Bruijne, J. H. J. 2012, *Ap&SS*, 341, 31
- de Laverny, P., Recio-Blanco, A., Worley, C. C., & Plez, B. 2012, *A&A*, 544, A126
- DPAC. 2007, *GAIA-CD-SP-DPAC-FM-030*, Tech. rep., Data Processing and Analysis Consortium
- DPAC. 2012, *GAIA-C9-PL-UB-XL-033*, Tech. rep., Data Processing and Analysis Consortium
- Fioc, M. & Rocca-Volmerange, B. 1997, *A&A*, 326, 950
- Fioc, M. & Rocca-Volmerange, B. 1999, arXiv:9912179
- Fremat, Y., Houziaux, L., & Andriolat, Y. 1996, *MNRAS*, 279, 25
- Frémat, Y., Neiner, C., Hubert, A.-M., et al. 2006, *A&A*, 451, 1053
- Fustes, D., Dafonte, C., Arcay, B., et al. 2013a, *A&A*, accepted
- Fustes, D., Dafonte, C., Arcay, B., et al. 2013b, *Expert Syst. Appl.*, 40, 1530
- Gazzano, J.-C., de Laverny, P., Deleuil, M., et al. 2010, *A&A*, 523, A91
- Gazzano, J.-C., Kordopatis, G., Deleuil, M., et al. 2013, *A&A*, 550, A125
- Geurts, P., Ernst, D., & Wehenkel, L. 2006, *Mach. Learn.*, 63, 3
- Gilmore, G., Randich, S., Asplund, M., et al. 2012, *The Messenger*, 147, 25
- Górski, K. M., Hivon, E., Banday, A. J., et al. 2005, *ApJ*, 622, 759
- Gustafsson, B., Edvardsson, B., Eriksson, K., et al. 2008, *A&A*, 486, 951
- Henry, T. J., Soderblom, D. R., Donahue, R. A., & Baliunas, S. L. 1996, *AJ*, 111, 439
- Jordi, C. 2011, in *EAS Publications Series*, Vol. 45, *EAS Publications Series*, 149–154
- Jordi, C., Gebran, M., Carrasco, J. M., et al. 2010, *A&A*, 523, A48
- Jordi, C., Høg, E., Brown, A. G. A., et al. 2006, *MNRAS*, 367, 290
- Karamelas, A., Kontizas, M., Rocca-Volmerange, B., et al. 2012, *A&A*, 538, A38
- Katz, D., Cropper, M., Meynadier, F., et al. 2011, in *EAS Publications Series*, Vol. 45, *EAS Publications Series*, 189–194
- Katz, D., Munari, U., Cropper, M., et al. 2004, *MNRAS*, 354, 1223
- Kim, D.-W. 2013, *GAIA-C8-SP-MPIA-DWK-001*, Tech. rep., Max Planck Institute for Astronomy, Heidelberg
- Kochukhov, O. & Shulyak, D. 2008, *Contributions of the Astronomical Observatory Skalnaté Pleso*, 38, 419
- Kohonen, T., Schroeder, M. R., & Huang, T. S., eds. 2001, *Self-Organizing Maps*, 3rd edn. (Secaucus, NJ, USA: Springer-Verlag New York, Inc.)
- Kordopatis, G., Recio-Blanco, A., de Laverny, P., et al. 2011a, *A&A*, 535, A106
- Kordopatis, G., Recio-Blanco, A., de Laverny, P., et al. 2011b, *A&A*, 535, A107
- Lebzelter, T., Heiter, U., Abia, C., et al. 2012, *A&A*, 547, A108
- Li, J., Ray, S., & Lindsay, B. G. 2007, *Journal of Machine Learning Research*, 8, 1687
- Lindgren, L., Babusiaux, C., Bailer-Jones, C. A. L., et al. 2008, in *IAU Symposium*, Vol. 248, *IAU Symposium*, ed. W. J. Jin, I. Platais, & M. A. C. Perryman, 217–223
- Lindgren, L., Lammers, U., Hobbbs, D., et al. 2012, *A&A*, 538, A78
- Linsky, J. L. & Haisch, B. M. 1979, *ApJ*, 229, L27

- Liu, C., Bailer-Jones, C. A. L., Sordo, R., et al. 2012, *MNRAS*, 426, 2463
- Lobel, A., Fremat, Y., & Blomme, R. 2010, GAIA-C8-TN-ROB-AJL-001, Tech. rep., Royal observatory of Belgium
- Lobel, A., Fremat, Y., Royer, P., & Blomme, R. 2013, GAIA-C8-TN-ROB-AJL-002, Tech. rep., Royal observatory of Belgium
- Luri, X., Babusiaux, C., & Masana, E. 2005, in *ESA Special Publication*, Vol. 576, *The Three-Dimensional Universe with Gaia*, ed. C. Turon, K. S. O’Flaherty, & M. A. C. Perryman, 357
- Magic, Z., Collet, R., Asplund, M., et al. 2013, arXiv:1302.2621
- Martayan, C., Frémat, Y., Blomme, R., et al. 2008, in *SF2A-2008*, ed. C. Charbonnel, F. Combes, & R. Samadi, 499–
- Mashonkina, L., Korn, A. J., & Przybilla, N. 2007, *A&A*, 461, 261
- Merle, T., Thévenin, F., Pichon, B., & Bigot, L. 2011, *MNRAS*, 418, 863
- Mignard, F., Bailer-Jones, C. A. L., Bastian, U., et al. 2008, in *IAU Symposium*, Vol. 248, *IAU Symposium*, ed. W. J. Jin, I. Platais, & M. A. C. Perryman, 224–230
- Mignard, F. & Klioner, S. A. 2010, in *IAU Symposium*, Vol. 261, *IAU Symposium*, ed. S. A. Klioner, P. K. Seidelmann, & M. H. Soffel, 306–314
- Ng, Y. K. & Bertelli, G. 1998, *A&A*, 329, 943
- O’Mullane, W., Lammers, U., Bailer-Jones, C. A. L., et al. 2007, in *Astronomical Society of the Pacific Conference Series*, Vol. 376, *Astronomical Data Analysis Software and Systems XVI*, ed. R. A. Shaw, F. Hill, & D. J. Bell, 99
- Önehag, A., Heiter, U., Gustafsson, B., et al. 2012, *A&A*, 542, A33
- Perryman, M. A. C., de Boer, K. S., Gilmore, G., et al. 2001, *A&A*, 369, 339
- Plez, B. 2011, *Journal of Physics Conference Series*, 328, 012005
- Pont, F. & Eyer, L. 2004, *MNRAS*, 351, 487
- Prod’homme, T., Holl, B., Lindegren, L., & Brown, A. G. A. 2012, *MNRAS*, 419, 2995
- Prusti, T. 2012, GAIA-CG-PL-ESA-TJP-011, Tech. rep., Data Processing and Analysis Consortium
- Raskin, G., van Winckel, H., Hensberge, H., et al. 2011, *A&A*, 526, A69
- Rasmussen, C. & Williams, C. 2006, *Gaussian processes for machine learning*, *Adaptive computation and machine learning* (MIT Press)
- Recio-Blanco, A., Bijaoui, A., & de Laverny, P. 2006, *MNRAS*, 370, 141
- Robin, A. C., Luri, X., Reylé, C., et al. 2012, *A&A*, 543, A100
- Robin, A. C., Reylé, C., Derrière, S., & Picaud, S. 2003, *A&A*, 409, 523
- Sarro, L. M., Berihuete, A., Carrión, C., et al. 2013, *A&A*, 550, A44
- Shulyak, D., Tsybmal, V., Ryabchikova, T., Stütz, C., & Weiss, W. W. 2004, *A&A*, 428, 993
- Sivia, D. & Skilling, J. 2006, *Data Analysis: A Bayesian Tutorial*, *Oxford Science Publications* (OUP Oxford)
- Smith, K. 2011, GAIA-C8-TN-MPIA-KS-019, Tech. rep., Max Planck Institute for Astronomy, Heidelberg
- Smith, K. W., Bailer-Jones, C. A. L., Klement, R. J., & Xue, X. X. 2010, *A&A*, 522, A88
- Sordo, R. & Vallenari, A. 2013, GAIA-C8-TN-MPIA-RS-010, Tech. rep., Data Processing and Analysis Consortium
- Sordo, R., Vallenari, A., Tantaló, R., et al. 2010, *Ap&SS*, 328, 331
- Sordo, R., Vallenari, A., Tantaló, R., et al. 2011, *Journal of Physics Conference Series*, 328, 012006
- Tanga, P. & Mignard, F. 2012, *Planet. Space Sci.*, 73, 5
- Tsalmantza, P. & Bailer-Jones, C. 2012, in *Springer Series in Astrostatistics*, Vol. 2, *Astrostatistics and Data Mining*, ed. L. M. Sarro, L. Eyer, W. O’Mullane, & J. De Ridder (Springer New York), 263–270
- Tsalmantza, P. & Bailer-Jones, C. A. L. 2010a, GAIA-C8-SP-MPIA-PAT-009, Tech. rep., Max Planck Institute for Astronomy, Heidelberg
- Tsalmantza, P. & Bailer-Jones, C. A. L. 2010b, GAIA-C8-TN-MPIA-PAT-008, Tech. rep., Max Planck Institute for Astronomy, Heidelberg
- Tsalmantza, P., Karamelas, A., Kontizas, M., et al. 2012, *A&A*, 537, A42
- Tsalmantza, P., Kontizas, M., Bailer-Jones, C. A. L., et al. 2007, *A&A*, 470, 761
- Tsalmantza, P., Kontizas, M., Rocca-Volmerange, B., et al. 2009, *A&A*, 504, 1071
- Turon, C., O’Flaherty, K. S., & Perryman, M. A. C., eds. 2005, *ESA Special Publication*, Vol. 576, *The Three-Dimensional Universe with Gaia*
- Vapnik, V. N. 1995, *The nature of statistical learning theory* (New York, NY, USA: Springer-Verlag New York, Inc.)
- Wand, M. P. & Jones, M. C. 1995, *Kernel smoothing* (Chapman and Hall)
- Worley, C. C., de Laverny, P., Recio-Blanco, A., et al. 2012, *A&A*, 542, A48
- Wright, J. T., Marcy, G. W., Butler, R. P., & Vogt, S. S. 2004, *ApJS*, 152, 261

- 1 Max Planck Institute for Astronomy, Königstuhl 17, 69117 Heidelberg, Germany
- 2 Fac. Informática, Universidade da Coruña, Campus de Elviña 15071, A Coruña, Spain
- 3 Institute for Astronomy, Astrophysics, Space Applications & Remote Sensing, National Observatory of Athens, PO Box 20048, 11810 Athens, Greece
- 4 Dpt. Statistics and Operations Research, University of Cádiz, Campus Universitario Río San Pedro s/n. 11510 Puerto Real, Cádiz, Spain
- 5 Laboratoire Lagrange (UMR7293), Université de Nice Sophia Antipolis, CNRS, Observatoire de la Côte d’Azur, BP 4229, 06304 Nice, France
- 6 Dpt. de Inteligencia Artificial, UNED, Juan del Rosal, 16, 28040 Madrid, Spain
- 7 Institut d’Astrophysique et de Géophysique, Université de Liège, Allée du 6 Août, 17, B5C, 4000 Sart Tilman, Belgium
- 8 Centre de Recherche en Astronomie, Astrophysique et Géophysique, Route de l’Observatoire Bp 63 Bouzareah, DZ-16340 Algiers, Algeria
- 9 Department of Astrophysics, Astronomy & Mechanics, Faculty of Physics, University of Athens, 15783 Athens, Greece
- 10 Osservatorio Astrofisico di Torino, Istituto Nazionale di Astrofisica (INAF), Pino Torinese, Italy
- 11 Royal Observatory of Belgium, 3 avenue circulaire, B1180 Brussels, Belgium
- 12 Area of Languages and Computer Systems, Pablo de Olavide University, Seville, Spain
- 13 Observatoire de Paris, GEPI, CNRS UMR 8111, 92195, Meudon, France
- 14 Institut de Physique de Rennes, Université de Rennes 1, CNRS UMR 6251, F-35042 Rennes, France
- 15 Division of Astronomy and Space Physics, Department of Physics and Astronomy, Uppsala University, Box 516, SE-751 20 Uppsala, Sweden
- 16 Centre National d’Etudes Spatiales, 18 av Edouard Belin, 31401 Toulouse, France
- 17 Niels Bohr Institute, Copenhagen University, Copenhagen, Denmark
- 18 Tartu Observatory, 61602 Toravere, Estonia
- 19 Dipartimento di Fisica e Astronomia, Sezione Astrofisica, Università di Catania, via S. Sofia 78, Catania, Italy
- 20 INAF - Osservatorio Astrofisico di Catania, via S. Sofia 78, Catania, Italy
- 21 CSC Danmark A/S, Retortvej 8, DK-2500 Valby, Denmark
- 22 European Southern Observatory, Alonso de Cordova 3107, Vitacura, Santiago de Chile, Chile
- 23 Institut d’Astrophysique de Paris, UMR 7095 CNRS – Université Pierre & Marie Curie, 98bis boulevard Arago, 75014 Paris, France
- 24 Université de Paris-Sud XI, IAS, 91405, Orsay Cedex, France
- 25 INAF - Osservatorio Astronomico di Padova, Padova, Italy
- 26 LAB UMR 5804, Univ. Bordeaux - CNRS, Floirac, France





ANNEXE F

## Autres publications

---

Determination of the ages of stars from their position in the HR diagram, C. Guédé, Y. Lebreton, and G. Dréan (Guédé et al. 2011).

SF2A-2011 : Proceedings of the Annual meeting of the French Society of Astronomy and Astrophysics, Paris, 20-24 juin 2011. Eds : G. Alecian and K. Belkacem and R. Samadi and D. Valls-Gabaud.

## DETERMINATION OF THE AGES OF STARS FROM THEIR POSITION IN THE HR DIAGRAM

C. Guédé<sup>1</sup>, Y. Lebreton<sup>1</sup> and G. Dréan<sup>2</sup>

**Abstract.** The determination of stellar ages is fundamental to understand the formation and evolution of the Galaxy. We determine the age of stars by combining their position in the HR diagram with stellar evolutionary tracks or isochrones. The goal of this study is to prepare the tools that will be used to age-date stars after the Gaia mission.

Keywords: Stars: fundamental parameters - Methods: statistical - (Stars:) Hertzsprung-Russell and C-M diagrams

### 1 Introduction

To understand the formation and evolution of the Galaxy it is necessary to determine the ages of its stars. There are several methods to determine the age of stars which are based on either the kinematics or expansion of stars, the lithium depletion, the gyrochronology, activity, asteroseismology or isochrones models. These methods are described by Soderblom 2010. Here we are interested in determining the ages of large samples of stars for which the method based on isochrones is applicable. The age of stars is determined by combining their position in the HR diagram and models isochrones (Ng & Bertelli 1998, Lachaume et al. 1999). We use a Bayesian estimation to determine the most probable age from stellar models ( Pont & Eyer 2004, Jørgensen & Lindegren 2005 and Casagrande et al. 2011). We adapt this method to use the stellar evolutionary tracks instead of the isochrones. We compare our results to Casagrande et al. (2011) work to validate our tools. This method will be used to determine the ages of stars that will be observed by the Gaia mission.

In Section 2 we describe the Bayesian estimation and the methods that we use. In section 3 we compare the ages obtained with evolutionary tracks and with isochrones. Section 4 describes the age-mass relation and age-metallicity relation and the comparison of these relations with Casagrande et al. (2011) work.

### 2 Determination of ages

We determine the age of the stars from their position in the HR diagram and either stellar evolutionary tracks or isochrones. In the region of isochrones where the stars evolve very quickly ( for example the turn-off) a star that we aim to date, can be adjusted by several isochrones. As an example in Fig 1 (Jørgensen & Lindegren 2005), for the star on the left there are three isochrones that adjust properly the star. Therefore, these stars have three possible ages. In this case, the question is how to choose the correct age? In order to answer this question we use a Bayesian approach: this method allows us to determine the most probable age with the *a priori* density function. The age of a star corresponds to the maximum of the *a posteriori* density function  $f(T, [Fe/H], m)$ , defined as

$$f(T, [Fe/H], m) \propto f_0(T, [Fe/H], m)L(T, [Fe/H], m) \quad (2.1)$$

where  $f_0(T, [Fe/H], m)$  is the *a priori* density function, which depends on the Initial Mass Function, Stellar Formation Rate and initial metallicity distribution. We choose to adopt a flat stellar formation rate and a

---

<sup>1</sup> GEPI UMR 8111, Observatoire de Paris-Meudon , France

<sup>2</sup> LTSI, INSERM U642 Université de Rennes 1, France

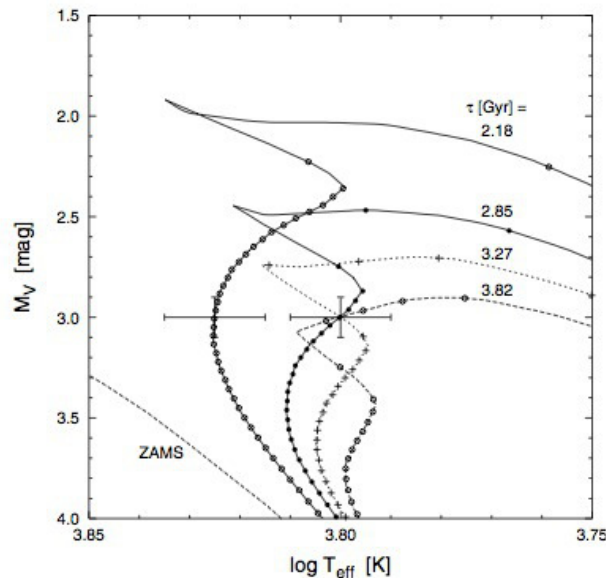
flat initial metallicity distribution. For the initial mass function we use the same than Jørgensen & Lindegren (2005), which is defined as  $\xi(m) = m^{-2.7}$  (it is based on the *IMF* of Kroupa et al. 1993).  $L(T, [Fe/H], m)$  is the likelihood defined as

$$L(T, [Fe/H], m) = \left( \prod_{i=1}^n \frac{1}{(2\pi)^{\frac{1}{2}} \sigma_i} \right) \exp\left(\frac{-\chi^2}{2}\right) \quad (2.2)$$

$$(2.3)$$

where the  $\chi^2$  parameter is calculated for the temperature  $T_{eff}$  (or color) of the stars, the magnitude  $M_v$  (or luminosity) and the metallicity  $[Fe/H]$ . The  $\sigma_i$  are the corresponding observational errors. For the numerical implementation, we sum the *a posteriori* density function for the evolutionary track that have a metallicity measuring range between  $[Fe/H]_{obs} \pm 3.5\sigma_{[Fe/H],obs}$  and for all masses. The age of the star corresponds to the maximum of the *a posteriori* density function.

The method of Jørgensen & Lindegren (2005) determines the age by using isochrones and we have adapted the program to determine the age directly from the evolutionary tracks. We compare both results in Section 3. This method is also well designed to calculate the mass of stars.



**Fig. 1.** Isochrones degeneracy in the HR diagram. After Jørgensen & Lindegren (2005)

### 3 Tracks vs. Isochrones

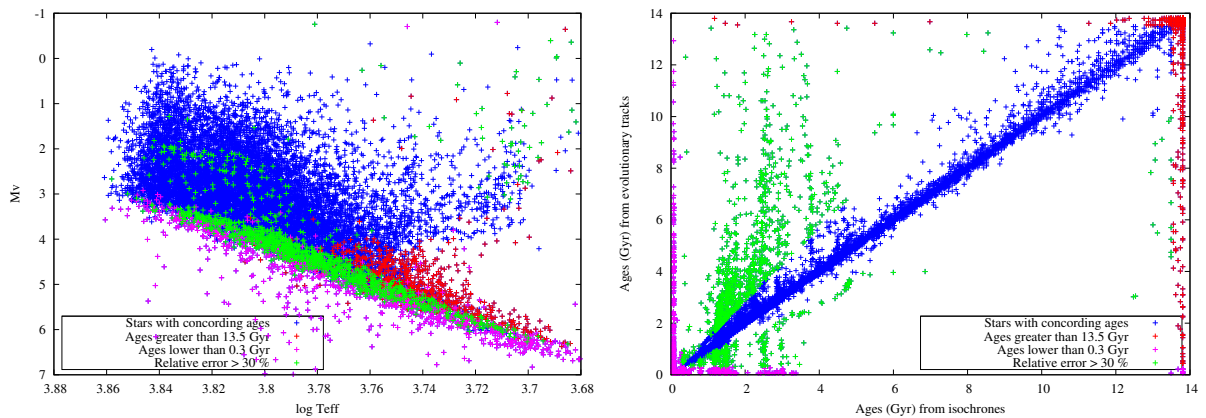
Traditionally, ages are derived from isochrones built by interpolation of stellar evolutionary tracks and provided by stellar modelers. To determine the ages we use the evolutionary tracks of Basti (Pietrinferni et al. 2004). We calculate the ages of 16 682 stars in the Geneva Copenhagen Survey of the solar neighbourhood (Casagrande et al. 2011 but see also Holmberg et al. 2009). These stars are represented in a diagram  $\log T_{eff} - M_v$  on the left Figure 2. We compare and classify the ages obtained with the isochrones and evolutionary tracks on the right Figure 2 and in the Table 1. We note that 72.2 % of the stars have similar ages : this shows that we have a good agreement between both methods. We note that the stars have different ages when they are located in the same region in the diagram  $\log T_{eff} - M_v$ , near the ZAMS. In the vicinity of the ZAMS, low mass stars evolve very slowly in the HR diagram so their age is poorly defined.

### 4 Age-mass relation and age-metallicity relation

We present the age-mass relation (on the left Figure 3) and the age-metallicity relation (on the right Figure 3) with ages and masses calculated by us with the evolutionary tracks method, for 6670 stars in the GCS

**Table 1.** Results of the comparison for the ages obtained with the isochrones and evolutionary tracks.

Similar ages	72.2 %
Relative difference exceeding 30%	7.5 %
Ages lower than 0.3 Gyr	15.2 %
<i>Stars with ages lower than 0.3 Gyr with both evolutionary tracks and isochrones</i>	47 %
Ages greater than 13.5 Gyr	5.1 %
<i>Stars with ages greater than 13.5 Gyr with both evolutionary tracks and isochrones</i>	68 %



**Fig. 2.** Left Figure : HR diagram with (i) in red, stars with age  $> 13.5$  Gyr, (ii) in pink stars with age  $< 0.3$  Gyr, (iii) in green stars with ages differing by more than 30 % and (iv) in blue the stars having similar ages. Right Figure : comparison between ages from isochrones and ages from evolutionary tracks.

catalogue. We obtain a relation that is similar to Casagrande et al. (2011). For the age-metallicity relation we see a concentration of stars at solar metallicity and small ages: when the age increases there is a metallicity dispersion due to the radial mixing of the stars. The relation allows to demonstrate that a subsample of stars belongs to the thin disk (Haywood 2008).

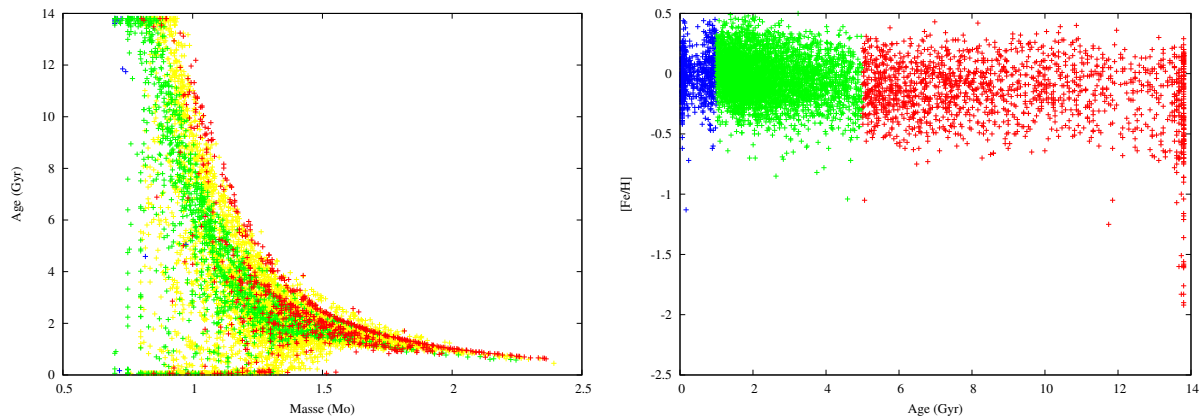
## 5 Conclusions

We adapted the method of Jørgensen & Lindegren (2005) to determine the age of stars from evolutionary tracks. The comparison of the isochrones ages with the evolutionary tracks ages shows that the results of both methods are similar except for the stars close to the ZAMS. In these regions, the stars have a low mass and they evolve very slowly, so their age is arduous to determine. The comparisons of our results with those of Casagrande et al. (2011) shows that we obtain the same trend for the age-mass and age-metallicity relations. These comparisons allow us to validate our program for age determination.

We warmly thank Misha Haywood for his advice and for discussions.

## References

- Casagrande, L., Schönrich, R., Asplund et al. S. 2011, A&A, 530, A138  
 Haywood, M. 2008, MNRAS, 388, 1175  
 Holmberg, J., Nordström, B., & Andersen, J. 2009, A&A, 501, 941  
 Jørgensen, B. R., & Lindegren, L. 2005, A&A, 436, 127  
 Kroupa, P., Tout, C. A., & Gilmore, G. 1993, MNRAS, 262, 545



**Fig. 3.** *Left Figure: Mass-Age relation for ages and masses from the evolutionary tracks. Colors indicate increasing metallicity  $[Fe/H]$  from metal-poor stars (in blue) to metal-rich stars (in red). Right Figure : Age- $[Fe/H]$  relation for ages from evolutionary tracks. In blue stars with age  $< 1$  Gyr, in green stars with  $1 \text{ Gyr} \leq \text{age} < 5$  Gyr, and in red stars with age  $\geq 5$  Gyr.*

Lachaume, R., Dominik, C., Lanz, T., & Habing, H. J. 1999, A&A, 348, 897

Ng, Y. K., & Bertelli, G. 1998, A&A, 329, 943

Pont, F., & Eyer, L. 2004, MNRAS, 351, 487

Pietrinferni, A., Cassisi, S., Salaris, M. et al. 2004, ApJ, 612, 168

Soderblom, D. R. 2010, ARA&A, 48, 581

Age dating large samples of stars : The Gaia context, C. Guédé, Y. Lebreton, C. Babusiaux and M. Haywood (Guédé et al. 2012a).

SF2A-2012 : Proceedings of the Annual meeting of the French Society of Astronomy and Astrophysics, Nice, 4-8 juin 2012. Eds : S. Boissier, P. de Laverny, N. Nardetto, R. Samadi, D. Valls-Gabaud and H. Wozniak.

## AGE DATING LARGE SAMPLES OF STARS: THE GAIA CONTEXT

C. Guédé<sup>1</sup>, Y. Lebreton<sup>1,2</sup>, C. Babusiaux<sup>1</sup> and M. Haywood<sup>1</sup>

**Abstract.** The Gaia ESA mission will measure stellar parameters for 1 billion of stars in the Galaxy. We aim at dating the stars and we use a method based on their observed position in the HR diagram and the evolutionary tracks. The method is based on a Bayesian estimation. We test the precision of the program with different simulated catalogs of Gaia stars at various distances and for a different observational error on the metallicity. The goal of this study is to prepare the tools that will be used to age-date stars after the Gaia mission.

Keywords: Stars: fundamental parameters - Methods: statistical - (Stars:) Hertzsprung-Russell and C-M diagrams

### 1 Introduction

The Gaia ESA mission will observe 1 billion of stars in photometry, spectroscopy and astrometry (Perryman et al. 2001). It will be launched in 2013. The age of these stars has to be estimated to obtain information about the Galaxy such as the stellar formation history and age metallicity relation which permit to understand its formation and evolution (Freeman 1993). Also to characterize the exoplanets it is necessary to date their host stars (Havel et al. 2011).

There are several methods to determine the age of stars described by Soderblom (2010). To determine the age of stars of the Gaia ESA mission it is necessary to use a method that applies to a large sample of stars. For this purpose we use a method based on isochrones placement (Edvardsson et al. 1993). A star is plotted in the Hertzsprung Russell Diagram (hereafter HRD) and the isochrone that adjusts its position is found. The most probable age corresponds to the isochrone that is closest to the star. In some regions of the HRD, the isochrones have a complex shape and stars can be adjusted by many isochrones. To select the most probable age it is useful to use a method based on the Bayesian estimation (Pont & Eyer 2004) to which we focus on the method of da Silva et al. (2006). We bring modifications on the choice of the *a priori* and on the stellar models. In the age dating we use the evolutionary tracks rather than the isochrones to reduce the number of interpolations and therefore the numerical errors.

We build a Gaia simulated catalog to evaluate the consistency between the simulated "true" age and the determined age. The catalog is constructed with the Basti evolutionary tracks and the Gaia specifications. We create several catalogs at different distances to test the precision of the determination of the ages as a function of the distance. The complementary spectroscopic observations will permit to obtain the metallicity with better observational errors. We study the effect on age determination of the observational errors on the metallicity for stars at different distances. We describe the Bayesian estimation method in Section 2 and the Gaia simulated catalogs in Section 3. We study the determination of ages as a function of the distance in Section 4 and the effect of observational error on the metallicity in the Section 5.

### 2 The Bayesian estimation

To date the stars we use three observables: the absolute magnitude  $M_V$ , the effective temperature  $T_{\text{eff}}$  and the metallicity  $[\text{Fe}/\text{H}]$ . The Bayesian estimation allows to determine the most probable age with the *a posteriori*

---

<sup>1</sup> GEPI UMR 8111, Observatoire de Paris-Meudon, France

<sup>2</sup> IPR, Université de Rennes 1, France

probability density function  $f(\tau, [\text{Fe}/\text{H}], m)$  which depends on the point of the evolutionary tracks with an age  $\tau$ , a metallicity  $[\text{Fe}/\text{H}]$  and a mass  $m$ . It is defined by  $f(\tau, [\text{Fe}/\text{H}], m) \propto f_0(\tau, [\text{Fe}/\text{H}], m)L(\tau, [\text{Fe}/\text{H}], m)$ , where  $f_0(\tau, [\text{Fe}/\text{H}], m)$  is the *a priori* density function that represents the stellar and galactic properties. The likelihood  $L(\tau, [\text{Fe}/\text{H}], m)$  is related to the distance between the models and the star. There are several adaptations of the Bayesian estimation for the determination of the ages. Here we have chosen to use the method based on da Silva et al. (2006, hereafter DG06) who define the *a posteriori* probability density function  $f(\tau_k)$  by:

$$f(\tau_k) = \sum_i p(\tau_k) \frac{1}{\sigma_{[\text{Fe}/\text{H}]} \sqrt{2\pi}} \exp \left[ -\frac{1}{2} \left( \frac{[\text{Fe}/\text{H}]_i - [\text{Fe}/\text{H}]^{\text{obs}}}{\sigma_{[\text{Fe}/\text{H}]}} \right)^2 \right], \quad (2.1)$$

$$\text{with } p(\tau_k) = \sum_j \int_{m_{ijk}^1}^{m_{ijk}^2} \xi(m_{ijk}) dm_{ijk} \psi(\tau_k) \exp \left[ -\frac{(M_{v,ijk} - M_v^{\text{obs}})^2}{\sigma_{M_v}^2} - \frac{(\log T_{\text{eff},ijk} - \log T_{\text{eff}}^{\text{obs}})^2}{\sigma_{\log T_{\text{eff}}}^2} \right]. \quad (2.2)$$

Where the index  $ijk$  correspond to the point of the evolutionary track with the metallicity  $i$ , the initial mass  $j$  and the age  $k$ .  $\xi(m_{ijk})$  is the initial mass function (IMF) and  $\psi(\tau_k)$  is the stellar formation rate (SFR).  $\sigma_{M_v}$ ,  $\sigma_{\log T_{\text{eff}}}$  and  $\sigma_{[\text{Fe}/\text{H}]}$  are the observational errors on the stellar parameters.

We bring several modifications to the method, in particular on the choice of the *a priori*. DG06 do not take into account any SFR. We adopt one because we have chosen not to take into account the stars that have an age greater than the Universe age. Indeed the evolutionary tracks have ages comprises between 0 and 22 Gyr for more massive stars. The SFR is defined by

$$\psi(\tau) = \begin{cases} 1 & \text{for } 0 \leq \tau \leq 14 \text{ Gyr} , \\ 0 & \text{elsewhere.} \end{cases} \quad (2.3)$$

We have chosen to use the stellar models of Basti (Pietrinferni et al. 2004). We use the evolutionary tracks rather than the isochrones to reduce the number of interpolations. Since DG06 do not specify the IMF, we have chosen to use the one of Kroupa (2002).

### 3 Gaia simulated catalogs

We create a simulated catalog of 10 000 stars to compare the simulated "true" age and the determined age. The simulated catalog is based on the Gaia mission specifications (Perryman et al. 2001). For each simulated star we select three parameters in the evolutionary tracks of Basti (Pietrinferni et al. 2004): metallicity, mass and age. The metallicity is selected randomly in the MDF. We defined the MDF with the 16 882 stars of the solar neighborhood of the Geneva Copenhagen Survey (GCS) of Casagrande et al. (2011). The MDF is defined by

$$\phi([\text{Fe}/\text{H}]) = \frac{1}{\sigma \sqrt{2\pi}} \exp \left( -\frac{([\text{Fe}/\text{H}] - \mu)^2}{2\sigma^2} \right), \quad (3.1)$$

with  $\mu = -0.043$  and  $\sigma = 0.19$ . We take randomly the mass in the IMF of Kroupa (2002) and the age is taken randomly in the SFR (Equation 2.3). The three parameters permit to select a point of the evolutionary tracks. To fulfill the Gaia mission specifications we keep the point if the magnitude  $G$  is greater than 6, beyond this limit the star is too bright to be observable by Gaia.

To test the effect on age of the distance  $d$  of stars we create several simulated catalogs for different distances. The selection of the distance is realized thanks to the selection on the precision of the parallax. Indeed the point is kept if it has  $\frac{\sigma_\pi}{\pi} < 10\%$ . The Gaia specification (ESA's website\*) gives  $\sigma_\pi = (9.3 + 658.1z + 4.568z^2)^{1/2} \times [0.986 + (1 - 0.986)(V - I)]$  where  $z = \text{MAX}[10^{0.4(12-15)}, 10^{0.4(G-15)}]$ .  $G$  is determined with the transformation of Jordi et al. (2010). We create several catalogs such as  $46 \text{ pc} \leq d \leq 12 \text{ kpc}$ . The Gaia simulated catalogs obtained for  $d = 46 \text{ pc}$ ,  $d = 1 \text{ kpc}$  and  $d = 10 \text{ kpc}$  are represented in Fig. 1.

We define an observation error for each stellar parameters based on the Gaia mission (Bailer-Jones (2010) and ESA's website\*). We assume an observational error on the effective temperature equal to 0.3% for  $G \leq 15$  and increasing linearly to 4% at  $G = 20$ . We assume that the observational error on the metallicity is constant and equal to 0.3 dex. The error on the absolute magnitude is based on the magnitude distance relation.  $\sigma_{M_v}$

\*[www.rssd.esa.int/index.php?project=GAI&page=Science\\_Performance#table2](http://www.rssd.esa.int/index.php?project=GAI&page=Science_Performance#table2)

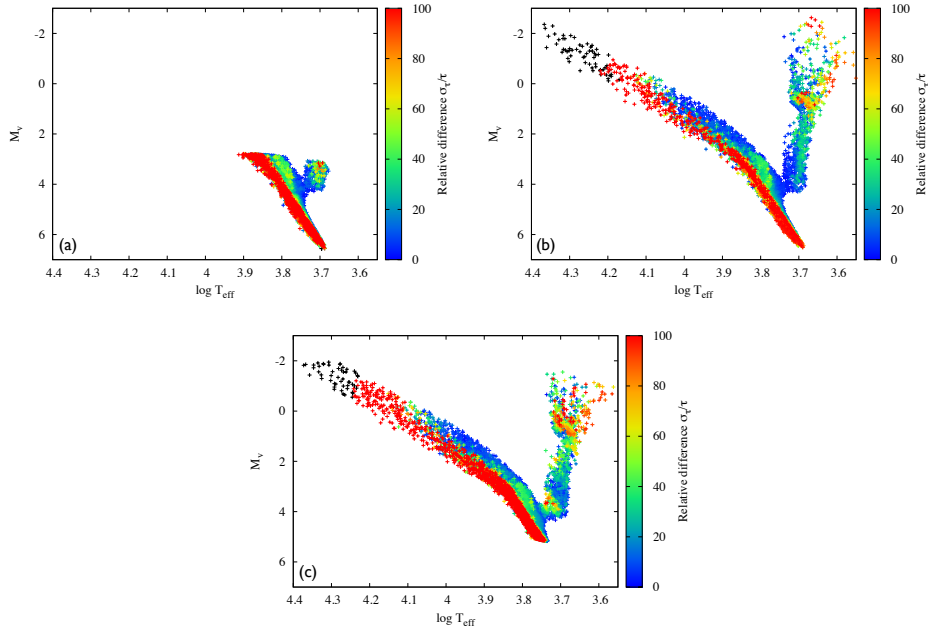


depends on the observational errors on the apparent magnitude  $\sigma_{m_V}$ , on the distance  $\sigma_d$  and on the extinction  $\sigma_{A_V}$ . We assume that  $\sigma_{m_V} = \sigma_{m_G}$  and the ESA's website\* gives  $\sigma_{m_G} = 10^{-3}(0.02076z^2 + 2.7224z + 0.004352)^{1/2}$ .  $\sigma_{A_V}$  is equal to 10% of the extinction  $A_V$  (Bailer-Jones 2010).

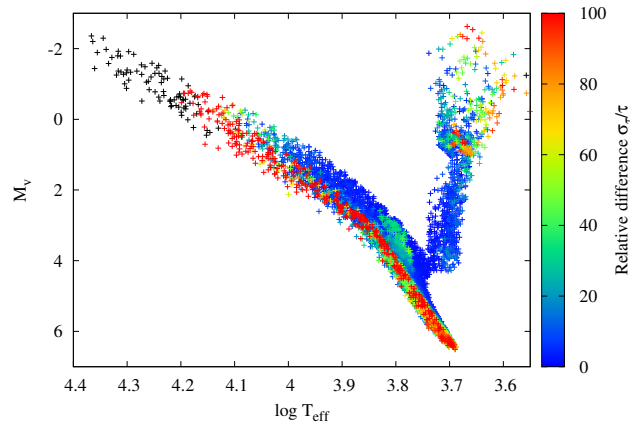
#### 4 Effect of the distance of the stars

We compare the simulated "true" age and the determined age for the Gaia simulated catalogs at different distances. For this purpose we define the relative difference between the two ages such as  $\sigma_\tau/\tau = (\tau_{\text{simulated}} - \tau_{\text{determined}})/\tau_{\text{simulated}}$ . We get the following results as a function of the distance.

1. We find 60 % of the stars with  $\sigma_\tau/\tau < 20\%$  for  $1 \leq d \leq 8$  kpc. These stars have a well-determined age. The results of the comparison are presented in Fig. 1, Panel b for  $d = 1$  kpc. The stars in red have an ill-determined age. They are located in the same problematic regions of the HRD: close to the zero age main sequence (ZAMS), regions of massive stars in the upper main sequence (MS) and red giant branch (RGB). The degeneracy in these regions is explained by the evolutionary speed of stars. In the vicinity of the ZAMS, low mass stars evolve slowly so their ages are poorly defined. On the contrary the stars on the RGB and in the upper MS evolve quickly.
2. There are 75 % of stars with  $\sigma_\tau/\tau < 20\%$  for  $d < 1$  kpc. The results for  $d = 46$  pc are represented in Fig. 1, Panel a. The limitation on the magnitude gives a catalog with faint stars at short distances. The stars are located in the bottom of the MS which is a degeneracy region. Thus we find more stars with a poorly defined age.
3. We find 50 % of stars with  $\sigma_\tau/\tau < 20\%$  for  $d > 8$  kpc. The HRD at 10 kpc is represented in Fig. 1, Panel c. When the distance is large, observational errors on the parameters become significant. It is more difficult to determine the ages.



**Fig. 1.** Comparison of the simulated "true" age and determined age in the  $\log T_{\text{eff}} - M_V$  diagram of the simulated catalogs for different distance and  $\sigma_{[Fe/H]} = 0.3$  dex. *Panel a:*  $d = 46$  pc. *Panel b:*  $d = 1$  kpc. *Panel c:*  $d = 10$  kpc. Colors represent the relative difference  $\sigma_\tau/\tau$ . The blue stars have a well-determined age ( $\sigma_\tau/\tau=0\%$ ) while the red stars have an ill-determined age ( $\sigma_\tau/\tau=100\%$ ). The age of black stars are undetermined.



**Fig. 2.** Comparison of true age and estimated age in the  $\log T_{\text{eff}} - M_V$  diagram of the Gaia simulated catalog with  $\sigma_{[Fe/H]} = 0.1$  dex. Colors are the same that in Fig. 1.

## 5 Effect of observational errors on the metallicity

The complementary spectroscopic observations of Gaia will permit to obtain the spectrum of the stars and will allow to determine the metallicity with a better observational error. We determine the ages for the same catalogs but now we take  $\sigma_{[Fe/H]} = 0.1$  dex. The results of the comparison at  $d = 1$  kpc are represented in Fig. 2. We compare the determined ages and the estimated "true" ages. We find a great improvement for the stars close to the ZAMS and in the RGB. Indeed there are more than 75 % of stars with  $\sigma_\tau/\tau < 20\%$  for the stars with a distance less than 10 kpc. For a distance greater than 10 kpc there are 55 % of stars with  $\sigma_\tau/\tau < 20\%$ . The improvement is worse for these stars because the observational errors are too important.

## 6 Conclusions

The Gaia ESA mission will observe 1 billion of stars. Different structures and objects will observe depending on the distances. At less than 200 pc the Jupiter mass planets will be detected. Age of their host stars will allow to characterize these exoplanets. The open clusters, the globular clusters and the disk will also be observed by Gaia. Their dating will allow to understand the formation and evolution of the Galaxy. At these distances, there are 60 % of stars with an age well determined with our method. To promote these results we need a good accuracy on the metallicity. The complementary spectroscopic observations will allow to obtain a better accuracy, allowing a better age determination for 15 % of the stars.

## References

- Bailer-Jones, C. A. L. 2010, MNRAS, 403, 96  
 Casagrande, L., Schönrich, R., Asplund, M., et al. 2011, A&A, 530, A138  
 da Silva, L., Girardi, L., Pasquini, L., et al. 2006, A&A, 458, 609  
 Edvardsson, B., Andersen, J., Gustafsson, B., et al. 1993, A&A, 275, 101  
 Freeman, K. C. 1993, in Astronomical Society of the Pacific Conference Series, Vol. 49, Galaxy Evolution. The Milky Way Perspective, ed. S. R. Majewski, 125  
 Havel, M., Guillot, T., Valencia, D., & Crida, A. 2011, A&A, 531, A3  
 Jordi, C., Gebran, M., Carrasco, J. M., et al. 2010, A&A, 523, A48  
 Kroupa, P. 2002, Science, 295, 82  
 Perryman, M. A. C., de Boer, K. S., Gilmore, G., et al. 2001, A&A, 369, 339  
 Pietrinferni, A., Cassisi, S., Salaris, M., & Castelli, F. 2004, ApJ, 612, 168  
 Pont, F. & Eyer, L. 2004, MNRAS, 351, 487  
 Soderblom, D. R. 2010, ARA&A, 48, 581

Age dating large samples of stars : Ways toward improved accuracy, C. Guédé, Y. Lebreton, C. Babusiaux and M. Haywood (Guédé et al. 2012b).

SF2A-2012 : Proceedings of the Annual meeting of the French Society of Astronomy and Astrophysics, Nice, 4-8 juin 2012. Eds : S. Boissier, P. de Laverny, N. Nardetto, R. Samadi, D. Valls-Gabaud and H. Wozniak.

## AGE DATING LARGE SAMPLES OF STARS: WAYS TOWARD IMPROVED ACCURACY

C. Guédé<sup>1</sup>, Y. Lebreton<sup>1,2</sup>, C. Babusiaux<sup>1</sup> and M. Haywood<sup>1</sup>

**Abstract.** The determination of stellar ages is essential in many fields of astrophysics, for instance to understand the formation and evolution of the Galaxy or to characterize exoplanets. We focus on age-dating combining the observed position of stars in the Hertzsprung–Russell diagram and evolutionary tracks. We use a Bayesian method that provides the most probable ages. We discuss the impact of including supplementary observational constraints (seismic parameters and exoplanetary transits) in the determination of the ages.

Keywords: Stars: fundamental parameters - Methods: statistical - (Stars:) Hertzsprung-Russell and C-M diagrams

### 1 Introduction

The Gaia ESA mission (Perryman et al. 2001) will be launched in 2013 and will observe 1 billion stars in the Galaxy. Gaia will provide global stellar parameters: effective temperature, absolute magnitude and metallicity. We aim at dating large samples of stars to be observed by Gaia. Among the various methods that can be used to determine the age of stars (see Soderblom 2010), the method based on stellar isochrones placement (Edvardsson et al. 1993) is well-suited for our purpose.

The method of the isochrones placement consists in adjusting the position of a star in the Hertzsprung–Russell diagram (hereafter HDR) by model isochrones. The age of the closest isochrone then corresponds to the most probable age of the star. In this work we have chosen to use the method of da Silva et al. (2006) which is based on Bayesian estimation. We have brought several modifications to the choice of the *a priori* (initial mass function, metallicity distribution function and stellar formation rate) and we have used evolutionary tracks rather than isochrones in order to reduce the number of interpolations and the numerical errors. The method used in the present paper is described in (Guédé et al. 2012, these proceedings).

We have built a Gaia simulated catalogue to evaluate the consistency between the "true" age of the star and the age determined with the Bayesian method. The catalogue is built from Basti evolutionary tracks (Pietrinferni et al. 2004) and Gaia specifications (Perryman et al. 2001). The Gaia simulated catalogue is described in (Guédé et al. 2012, these proceedings). For this work we took an observational error on the metallicity of 0.1 dex and we simulated stars at a distance of 1 kpc.

The CoRoT (Baglin et al. 2006) and Kepler (Christensen-Dalsgaard et al. 2007) missions allow to obtain the asteroseismic parameters of observed stars and/or the parameters of the exoplanetary transits. We have added this information as constraints for the age-dating in order to study the potential improvement of the age determination. In Section 2 we present results on age determination based on three observational constraints ( $M_v$ ,  $T_{\text{eff}}$ , [Fe/H]). In Section 3, we describe the improvement obtained when seismic and/or planetary transit constraints are added in the age determination process.

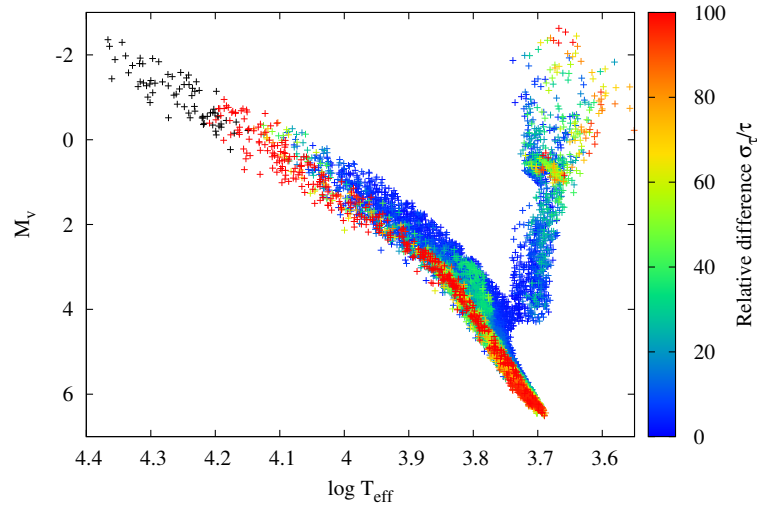
### 2 Ages from HRD constraints and metallicity

We determine the ages of stars in the Gaia simulated catalogue considering three observables: the absolute magnitude  $M_v$ , the effective temperature  $T_{\text{eff}}$  and the metallicity [Fe/H]. The comparison between the determined age and the simulated "true" age is presented in Fig. 1. We use the age relative difference defined by

<sup>1</sup> GEPI UMR 8111, Observatoire de Paris-Meudon, France

<sup>2</sup> IPR, Université de Rennes 1, France

$\sigma_\tau/\tau = (\tau_{\text{simulated}} - \tau_{\text{determined}})/\tau_{\text{simulated}}$ . We find 75 % of stars with  $\sigma_\tau/\tau < 20$  %. These stars have a rather well-determined age and are mainly located in the same regions of the HRD: the turn-off, the sub-giant branch and in the main-sequence (MS) of intermediate mass stars ( $1.0M_\odot < m < 2.0M_\odot$ ). On the other hand, the stars with an ill-determined age are located in three problematic regions of the HRD: close to the zero age main sequence (ZAMS) in particular close to the bottom of the MS ( $m < 1.0M_\odot$ ), in the massive stars MS region ( $m > 2.0M_\odot$ ) and at the top of the red giant branch (RGB). In these regions there is a well-known degeneracy in the age-dating that can be explained by the evolutionary speed. In the vicinity of the ZAMS, low mass stars evolve very slowly and their age cannot be accurately determined from their HR diagram location. On the other hand, stars in the RGB and massive stars on the MS evolve quickly making the age determination sensitive to the uncertainty on the observed parameters.



**Fig. 1.** Comparison of the simulated "true" age and determined age in the  $\log T_{\text{eff}} - M_v$  diagram of the Gaia simulated catalog. The ages are determined with three observables ( $M_v$ ,  $T_{\text{eff}}$ ,  $[\text{Fe}/\text{H}]$ ). Colors represent the relative difference  $\sigma_\tau/\tau$ . The blue stars have a well-determined age ( $\sigma_\tau/\tau \approx 0\%$ ) while the red stars have an ill-determined age ( $\sigma_\tau/\tau \approx 100\%$ ). The age of black stars cannot be determined.

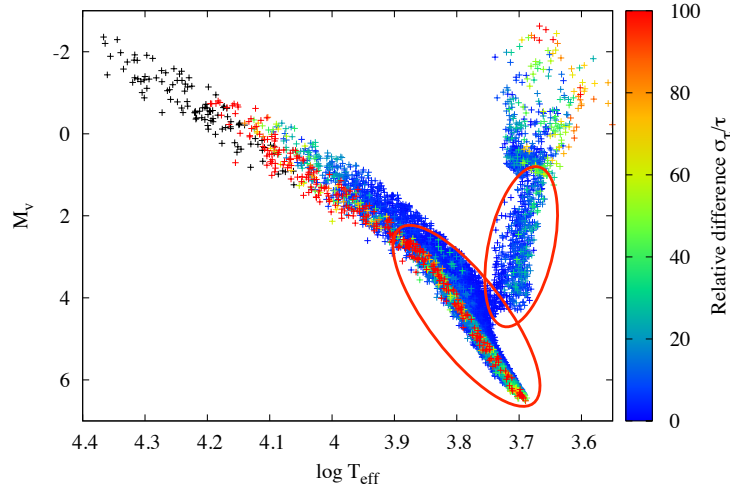
### 3 Way toward an age improvement

To reduce the age degeneracy we now add constraints provided by the high precision photometry missions observations CoRoT (Baglin et al. 2006) and Kepler (Christensen-Dalsgaard et al. 2007). These missions detect stellar oscillations which allow to probe the interior of stars in the whole H-R diagram in particular for low mass solar-type stars and red giants. CoRoT and Kepler also detect the transit of exoplanets in front of their host-star which provides information both on the planet and on the star (Southworth 2008 and Torres et al. 2008). We add these constraints in the age-dating to evaluate the improvement on the age determination.

#### 3.1 Adding seismic information

Seismic parameters are extracted from the oscillation power spectrum. We consider here the frequency at maximum power  $\nu_{\text{max}}$  and the large frequency separation  $\Delta\nu$  (Brown et al. 1991, Kjeldsen et al. 2008 and Mosser et al. 2010) which corresponds to the regular spacing in the power spectrum between low degree modes ( $\ell = 0, 1$ ) of same order. Kjeldsen & Bedding 1995 and Belkacem et al. 2011 have derived scaling relations that simply relate  $\Delta\nu$  and  $\nu_{\text{max}}$  to the global stellar parameters (mass, radius, effective temperature):

$$\frac{\Delta\nu}{\Delta\nu_\odot} = \left(\frac{M}{M_\odot}\right)^{\frac{1}{2}} \left(\frac{R}{R_\odot}\right)^{-\frac{3}{2}} \propto \left(\frac{\rho_{\text{mean}}}{\rho_{\text{mean},\odot}}\right)^{\frac{1}{2}}, \quad (3.1)$$



**Fig. 2.** Comparison of the estimated "true" age and of the determined age in the  $\log T_{\text{eff}} - M_V$  diagram of the Gaia simulated catalogue. The ages are determined from three observables ( $M_V$ ,  $T_{\text{eff}}$ ,  $[\text{Fe}/\text{H}]$ ) and two seismic constraints ( $\nu_{\text{max}}$ ,  $\Delta\nu$ ). Colors are the same as in Fig. 1. The red circles represent the region where solar-like oscillations are detected.

$$\frac{\nu_{\text{max}}}{\nu_{\text{max},\odot}} = \left(\frac{M}{M_{\odot}}\right) \left(\frac{R}{R_{\odot}}\right)^{-2} \left(\frac{T_{\text{eff}}}{T_{\text{eff},\odot}}\right)^{-\frac{1}{2}} \propto \frac{g}{g_{\odot}} \left(\frac{T_{\text{eff}}}{T_{\text{eff},\odot}}\right)^{-\frac{1}{2}}. \quad (3.2)$$

where  $\Delta\nu_{\odot} = 134.9 \mu\text{Hz}$  and  $\nu_{\text{max},\odot} = 3050.0 \mu\text{Hz}$  (Kjeldsen et al. 2008),  $g$  ( $g_{\odot}$ ) is the (solar) gravity and  $\rho_{\text{mean}}$  ( $\rho_{\text{mean},\odot}$ ) is the (solar) mean density.

The asteroseismic parameters can be calculated for stars in the Gaia simulated catalogue using the scaling relations. The observational errors on these constraints are also considered. For MS stars, we take  $\sigma_{\Delta\nu} = 0.018 \Delta\nu$  and  $\sigma_{\nu_{\text{max}}} = 0.038 \nu_{\text{max}}$  (Verner et al. 2011). For RGB stars, we adopt  $\sigma_{\Delta\nu} = 0.4 \mu\text{Hz}$ ,  $\sigma_{\nu_{\text{max}}} = 2.5 \mu\text{Hz}$  if  $\nu_{\text{max}} \leq 80 \mu\text{Hz}$  and  $\sigma_{\nu_{\text{max}}} = 1.0 \mu\text{Hz}$  if  $\nu_{\text{max}} > 80 \mu\text{Hz}$  (Hekker et al. 2011).

We determine the ages for all stars in the HRD but solar-like oscillations are observed in the low mass stars and on the RGB (red circles in Fig. 2). The comparison of the ages is presented in Fig. 2. We find 90 % of stars with  $\sigma_{\tau}/\tau < 20$  %. This is an improvement in the age-dating due to the fact that the values of  $\Delta\nu$  and  $\nu_{\text{max}}$  vary more than  $M_V$  and  $T_{\text{eff}}$  in the HRD.

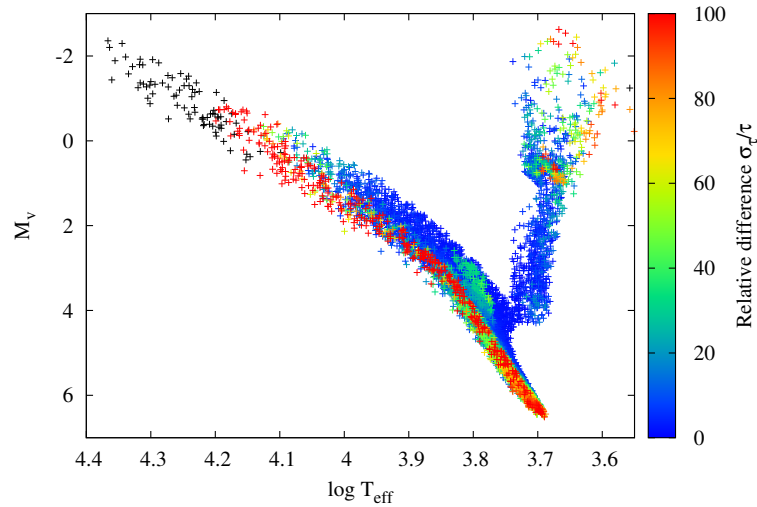
### 3.2 Adding information from planetary transits

The duration  $d$  of the exoplanetary transit is determined as a function of the period  $P$  of the orbital motion, of the projected distance  $b$  of the planet's center to the star's equator and of the mass  $M$  and radius  $R$  of the star (REF). It leads to the parameter:

$$\frac{M^{\frac{1}{3}}}{R} = \frac{1.8\sqrt{(1-b)^2 P^{\frac{1}{3}}}}{d} \propto \rho_{\text{mean}}^3 \quad (3.3)$$

We add  $M^{1/3}/R$  and observational error  $\sigma_{M^{1/3}/R}$  in the Gaia simulated catalogue. We take an observational error  $\sigma_{M^{1/3}/R} = 0.022 M^{1/3}/R$ , which is the mean of the observational errors found in the literature. The results of the comparison are presented in Fig. 3. We find 87 % of stars with  $\sigma_{\tau}/\tau < 20$  % which shows that the degeneracy is reduced for RGB stars and stars close to the ZAMS.

Both  $\Delta\nu$  and  $\sigma_{M^{1/3}/R}$  provide similar constraints to the age-dating because the mean density  $\rho_{\text{mean}}$  is proportional to  $M^{1/3}/R$  (Equation 3.3) and  $\Delta\nu$  (Equation 3.2).



**Fig. 3.** Comparison of the estimated "true" age and the determined age in the  $\log T_{\text{eff}} - M_v$  diagram of the simulated catalogue. The ages are determined on the basis of three observables ( $M_v$ ,  $T_{\text{eff}}$ ,  $[\text{Fe}/\text{H}]$ ) and exoplanetary transit constraint on the stellar mean density ( $M^{1/3}/R$ ). Colors are the same as in Fig. 1.

#### 4 Conclusions

We used a Bayesian estimation method to date the stars that will be observed by the Gaia ESA mission. This method allows to correctly date 60 % of the stars in a simulated catalogue on the basis of three stellar observables  $M_v$ ,  $T_{\text{eff}}$  and  $[\text{Fe}/\text{H}]$ . The other stars which are mainly located in three problematic regions of the HRD (close to the ZAMS, in the upper MS and in the RGB), have badly determined ages. We added new constraints in the age-dating to improve the ages. We took the seismic parameters and the mean stellar density which can be obtained with the CoRoT and Kepler missions. We showed that it improves the determination of the ages for more of 27 % of stars. Very interestingly there is a great improvement for stars close to the ZAMS which have traditionally poorly determined ages.

#### References

- Baglin, A., Auvergne, M., Barge, P., et al. 2006, in ESA Special Publication, Vol. 1306, ESA Special Publication, ed. M. Fridlund, A. Baglin, J. Lochard, & L. Conroy, 33
- Belkacem, K., Goupil, M. J., Dupret, M. A., et al. 2011, *A&A*, 530, A142
- Brown, T. M., Gilliland, R. L., Noyes, R. W., & Ramsey, L. W. 1991, *ApJ*, 368, 599
- Christensen-Dalsgaard, J., Arentoft, T., Brown, T. M., et al. 2007, *Communications in Asteroseismology*, 150, 350
- da Silva, L., Girardi, L., Pasquini, L., et al. 2006, *A&A*, 458, 609
- Edvardsson, B., Andersen, J., Gustafsson, B., et al. 1993, *A&A*, 275, 101
- Hekker, S., Elsworth, Y., De Ridder, J., et al. 2011, *A&A*, 525, A131
- Kjeldsen, H. & Bedding, T. R. 1995, *A&A*, 293, 87
- Kjeldsen, H., Bedding, T. R., & Christensen-Dalsgaard, J. 2008, in *American Institute of Physics Conference Series*, Vol. 1043, *American Institute of Physics Conference Series*, ed. V. Mioc, C. Dumitriche, & N. A. Popescu, 365–372
- Mosser, B., Belkacem, K., Goupil, M.-J., et al. 2010, *A&A*, 517, A22
- Perryman, M. A. C., de Boer, K. S., Gilmore, G., et al. 2001, *A&A*, 369, 339
- Pietrinferni, A., Cassisi, S., Salaris, M., & Castelli, F. 2004, *ApJ*, 612, 168
- Soderblom, D. R. 2010, *ARA&A*, 48, 581
- Southworth, J. 2008, *MNRAS*, 386, 1644
- Torres, G., Winn, J. N., & Holman, M. J. 2008, *ApJ*, 677, 1324
- Verner, G. A., Elsworth, Y., Chaplin, W. J., et al. 2011, *MNRAS*, 415, 3539

Stellar age-dating for different Gaia simulated catalogue, C. Guédé, Y. Lebreton, C. Babusiaux and M. Haywood.

Ewass 2012 : European Week of Astronomy and Space Science : Symposium 6 - "Stellar Populations 55 years after the Vatican Conference", Rome, 2-6 juillet 2012. Memorie della Società Astronomica Italiana.





# Stellar age-dating for different Gaia simulated catalogs

C. Guédé,<sup>1</sup> Y. Lebreton<sup>1,2</sup>, C. Babusiaux<sup>1</sup> and M. Haywood<sup>1</sup>

<sup>1</sup> GEPI UMR 8111, Observatoire de Paris-Meudon, France

<sup>2</sup> IPR, Université de Rennes 1, France e-mail: celine.guede@obspm.fr

**Abstract.** The determination of stellar ages is essential in many fields of astrophysics. For instance, it is used to understand the formation and evolution of the Galaxy. We focus on age-dating combining the observed position of stars in the HRD and stellar evolutionary models. Bayesian methods provide the most probable ages. We test the method on different simulated catalogs to test the precision of the method.

**Key words.** Stars: fundamental parameters - Methods: statistical - (Stars:) Hertzsprung-Russell and color-magnitude diagrams

## 1. Introduction

Stellar age determination is crucial for galactic evolution studies and many other fields of astrophysics. In particular, accurate ages are essential to establish the age metallicity relation and the stellar formation history. There are many methods to date the stars, we focus on the one based on stellar models which consists in adjusting the observed star on model isochrones or evolutionary tracks (Pont & Eyser 2004 and Jørgensen & Lindegren 2005).

## 2. Bayesian estimation

To date stars we use 3 observables: the effective temperature  $T_{\text{eff}}$ , the absolute magnitude  $M_v$  and the metallicity  $[\text{Fe}/\text{H}]$ . The position of the star is adjusted in the HRD<sup>1</sup> by isochrones. The nearest isochrone corresponds to the age of star. The selection is difficult in some regions of the HRD<sup>1</sup> where the isochrones have

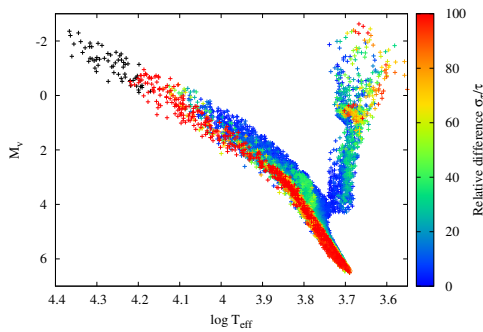
a complex shape. We select the more probable age with a Bayesian estimation. The PDF<sup>2</sup> *a posteriori* is defined by

$$f(\tau, [\text{Fe}/\text{H}], m) \propto f_0(\tau, [\text{Fe}/\text{H}], m) \times L(\tau, [\text{Fe}/\text{H}], m) \quad (1)$$

with  $f_0$  is the PDF<sup>2</sup> *a priori*.  $L$  is the likelihood, it determines the distance between the models and the star. We base our method on Da Silva et al. 2006's work to which we bring modifications. We have worked with evolutionary tracks (Basti, Pietrinferni et al. 2004) rather than isochrones, which reduces the number of interpolations and therefore the numerical errors. For the *a priori* we have used a flat stellar formation rate (SFR, from 0 to 14 Gyr and 0 elsewhere), the initial mass function (IMF) of Kroupa (2002) and no metallicity distribution function (MDF).

<sup>1</sup> Hertzsprung-Russell diagram

<sup>2</sup> Probability distribution function



**Fig. 1.** HRD of the simulated catalog to 1kpc. The color indicates the percentage of the relative age difference. The stars in blue have a well-determined age and the stars in red have an ill-determined age.

### 3. Gaia simulated catalogs

We create a simulated catalog to test the determination of ages. We simulate a catalog of 10 000 stars representative of Gaia observations at a distance of 1 kpc. For this purpose we randomly take three parameters in the Basti grid: the metallicity in the MDF, the mass in the Kroupa's IMF and the age in the SFR. We only keep stars with apparent magnitudes greater than 6 which is the Gaia lower threshold. The observational errors are those expected from the Gaia mission:  $\sigma_\pi$  depends on  $G$  and  $(V-I)$ ,  $\sigma_{T_{\text{eff}}}=0.3\%$  at  $G \leq 15$  mag and rises linearly to  $\sigma_{T_{\text{eff}}}=4\%$  at  $G=20$  mag,  $\sigma_G = \sigma_V = 10^{-3}$  mag,  $\sigma_{[\text{Fe}/\text{H}]}=0.3$  dex and  $\sigma_{A_V}=10\%$ . We have chosen to retain the stars with  $\sigma_\pi/\pi < 10\%$ . The catalog is represented in Fig. 1.

### 4. Age from HRD and $[\text{Fe}/\text{H}]$

We estimated the relative age difference ( $\sigma_\tau/\tau$ ) between the estimated age and the real simulated one. It is less than 10% for 38% of the stars in the catalog. The stars are represented in Fig. 1 with their relative difference. The stars close to the ZAMS (Zero age main sequence), the massive stars in the upper MS (Main sequence) and the stars on the RGB (Red giant branch) have very ill-determined ages.

### 5. Complementary spectroscopy

With the complementary spectroscopic observations planned in the Gaia context, the observational error on  $[\text{Fe}/\text{H}]$  will decrease to 0.1 dex. We determine the ages of the simulated catalog considering error. We find 62% of stars with  $\sigma_\tau/\tau < 10\%$ . The age are improved for the stars close to the ZAMS and in the RGB.

Spectroscopic observations will yield the surface gravity. Adding this constraint to the dating, we find 66% of stars with  $\sigma_\tau/\tau < 10\%$ . The degeneracy is minimized in the RGB where the gravity varies a lot.

### 6. Different distances

We created several catalogs for different distances with  $\sigma_{[\text{Fe}/\text{H}]}=0.1$  dex. For  $d < 100$  pc, 55% of the stars have  $\sigma_\tau/\tau < 10\%$ . The constraint  $G > 6$  mag favours the stars in the bottom of the HRD, where the age degeneracy is high. For  $100 \text{ pc} < d < 5 \text{ kpc}$ , 60% have  $\sigma_\tau/\tau < 10\%$ . For  $d > 5 \text{ kpc}$ , 40% have  $\sigma_\tau/\tau < 10\%$ . The ages are poorly defined because of the large observational errors.

### 7. Conclusions

The Gaia mission will observe 1 billion of stars, the age-dating of these stars using three observables allows to correctly determine 38% of the ages. There is a degeneracy in 3 problematic regions: close to the ZAMS, in the upper MS and in the RGB. With better observational errors there are 62% of the stars with good ages. This improvement will be crucial to characterize the exoplanetary systems and to study the galactic structure and evolution.

### References

- da Silva, L., et al. 2006, *A&A*, 458, 609
- Jørgensen, B. R., & Lindegren, L. 2005, *A&A*, 436, 127
- Kroupa, P. 2002, *Science*, 295, 82
- Pietrinferni, A., Cassisi, S., Salaris, M. et al. 2004, *ApJ*, 612, 168
- Pont, F., & Eyser, L. 2004, *MNRAS*, 351, 487

**Communications orales n'ayant pas donné lieu à des publications :**

Journées du GEPI, 13-14 janvier 2011. Détermination de l'âge des étoiles.

Séminaire au LESIA, 6 mars 2012. Détermination de l'âge des étoiles.

Journées du GEPI, 2-3 avril 2012. Détermination de l'âge des étoiles.

SF2A-2013. 4-7 juin 2013. Détermination de l'âge des étoiles dans le cadre de la mission spatiale Gaia.

Séminaire au GEPI, 24 juin 2013. Détermination de l'âge des étoiles dans le cadre de la mission spatiales Gaia.



ANNEXE G

# Nouveau chapitre de la thèse

---

*Valorisation des compétences, NCT<sup>®</sup>*

**Céline Guédé**

*Ecole doctorale : ED 127 Astronomie et Astrophysique d'île de France*

*Université ou organisme de rattachement : Observatoire de Paris*

*Nom du "mentor" : Philippe Bertrand*

**L'âge des étoiles**

**Sujet académique de la thèse : Détermination de l'âge des étoiles dans le cadre de la mission spatiale Gaia**

Nom de la directrice de thèse : Yveline Lebreton

Date probable de soutenance de la thèse : 11 Octobre 2013

## **1. Avant propos**

J'ai toujours été passionnée par l'astrophysique car cela permet de comprendre le monde qui nous entoure. Pour pouvoir travailler dans ce domaine je me suis orientée vers une licence de physique puis un master d'astronomie et d'astrophysique. J'ai choisi de faire de la recherche car c'est un métier enrichissant qui permet de toujours élargir ses connaissances. C'est pourquoi j'ai choisi de réaliser une thèse qui est indispensable pour entrer dans la recherche.

Le nombre de débouchés étant limité, j'ai choisi de me réorienter vers un autre secteur. De plus j'ai effectué ma thèse en télétravail car ma directrice de thèse est en mission permanente à Rennes. Maintenant je souhaite intégrer une équipe de travail afin de favoriser les échanges et les interactions dans l'équipe projet. J'écris cette présente note afin de valoriser mon travail de recherche et de présenter les compétences que j'ai acquises pendant ces trois années. Le sujet de ma thèse était « *la détermination de l'âge des étoiles dans le cadre de la mission Gaia* ».

## **1. Introduction**

### **2.1 Contexte**

La détermination de l'âge des étoiles est une donnée nécessaire dans plusieurs domaines de l'astrophysique. Par exemple, cette donnée est nécessaire à l'étude de la formation et de l'évolution de notre Galaxie. Il est également important de pouvoir dater précisément les étoiles les plus vieilles de la Galaxie afin d'avoir des indications supplémentaires sur l'âge de l'univers. De plus avec les récentes découvertes des planètes extrasolaires nous avons besoin de connaître l'âge de leur étoile hôte afin de pouvoir caractériser ces planètes. La future mission spatiale Gaia observera plus d'un milliard d'étoiles qu'il faudra dater.

### **2.2 Problématique**

L'âge est le paramètre stellaire le plus difficile à déterminer car il est impossible de savoir à quelle époque est née une étoile étant donnée sa durée de vie. Il faut donc développer des méthodes indirectes pour l'obtenir comme celles basées sur des relations empiriques ou des modèles d'évolution d'étoiles. L'utilisation de ces méthodes dépend des observations que l'on effectue c'est pourquoi certaines ne sont pas applicables à toutes les étoiles.

### **2.3 Objectifs**

Le but de mon projet de recherche est de développer un algorithme de détermination automatique de l'âge pour un grand échantillon d'étoiles. Ce projet est réalisé dans le cadre de la future mission Gaia. Cette mission spatiale de l'ESA, qui sera lancée en novembre 2013, observera plus d'un milliard d'étoiles dans les différentes structures de notre Galaxie. Cette mission permettra de caractériser notre Galaxie et d'améliorer nos connaissances sur la physique stellaire. Il est donc important d'avoir un programme automatique, optimisé et rapide pour dater les nombreuses étoiles observées par Gaia.

Pour déterminer les âges j'ai utilisé une méthode basée sur l'ajustement des observations par des modèles d'évolution d'étoiles. L'âge le plus probable est déterminé à partir d'une méthode statistique. Pour ce travail j'ai été amenée à comparer plusieurs méthodes (estimations bayésiennes et minimisations du chi-2) que j'ai implémentées. J'ai également analysé les effets des différents paramètres sur les âges. Enfin j'ai amélioré cette méthode en ajoutant des contraintes supplémentaires.

## **2.4 Enjeux**

Mon travail de recherche a des enjeux principalement scientifiques car les âges déterminés sont nécessaires dans de nombreux domaines de l'astrophysique. De plus mon travail a permis de faire une comparaison entre deux méthodes statistiques et ainsi de choisir celle qui est la plus optimale pour la mission. Mon travail a également un enjeu économique car il va être implémenté par le CNES pour la mission spatiale Gaia. Ce travail peut également être appliqué dans diverses études car j'ai étudié une méthode statistique de détermination de paramètres. Dans les cas où il est possible d'avoir un modèle théorique et des données alors la méthode statistique ainsi que l'algorithme que j'ai développé sont transposables à n'importe quel système.

## **1. Environnement de travail**

### **3.1 Activité**

J'ai réalisé ma thèse à l'Observatoire de Paris qui est un établissement de d'enseignement supérieur et de recherche. L'Observatoire de Paris, fondé en 1667, est le plus grand centre de recherche d'astrophysique en France et au monde. Il regroupe 30 % des astrophysiciens Français (650 permanents et 250 étudiants) sur trois sites : Paris, Meudon et Nançay. L'Observatoire de Paris est impliqué dans de nombreux projets de recherche Européens et internationaux. Pour cela il est associé à plusieurs laboratoires d'astrophysique Français. Il est également un des membres fondateurs de l'Initiative d'Excellence PSL, dont l'ambition est de créer en quatre ans une «université de recherche».

### **3.2 Organisation**

Depuis Janvier 2012 il est dirigé par Claude Catala et administré par un conseil administratif. Ils sont assistés par un conseil scientifique et un haut comité scientifique, composé de personnalités indépendantes, qui évalue régulièrement l'activité de l'établissement. L'Observatoire de Paris comporte cinq départements scientifiques:

- le LESIA : Laboratoire d'Etudes Spatiales et d'Instrumentation en Astrophysique
- le GEPI : Galaxies, Etoiles, Physique et Instrumentation
- le LERMA : Laboratoire d'Etudes du Rayonnement et de la Matière en Astrophysique
- le LUTH : Laboratoire Univers et THéories
- le SYRTE : SYstèmes de Référence Temps Espace

un institut :

- l'IMCCE : Institut de mécanique céleste et de calcul des éphémérides

deux services scientifiques :

- l'USN : Unité Scientifique de la Station de Nançay
- l'UFE : Unité de Formation et d'Enseignement

et un service commun qui regroupe l'ensemble des services de soutien aux missions de l'établissement ainsi que la bibliothèque.



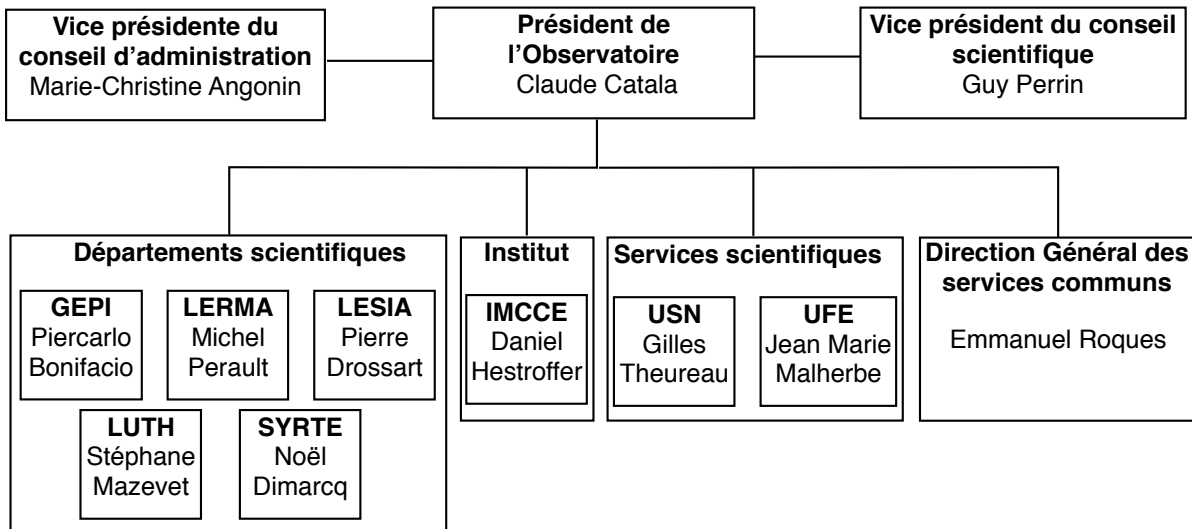


Fig. 1 : Organigramme de l'Observatoire de Paris

En 2010, le budget de l'Observatoire de Paris était de 24 millions d'euros, il était composé de 10,6 millions d'euros de crédits de fonctionnements, 9 millions d'euros de crédits d'investissements et 4,4 millions d'euros de crédits de personnels. Les dépenses s'élèvent à 14 millions d'euros avec 7,8 millions d'euros de dépenses de fonctionnement, 3 millions d'euros d'investissements et 2.9 millions d'euro de dépenses pour les charges sur le personnel.

### 3.3 Choix du laboratoire

J'ai effectué ma thèse au GEPI, le laboratoire Galaxies, Etoiles, Physique et Instrumentation dirigé par Piercarlo Bonifacio. C'est une unité mixte de recherche du Centre national de la recherche scientifique et il est associé à l'Université Paris-Diderot. Il regroupe 24 chercheurs, 49 ingénieurs, techniciens et administratifs, 7 post-docs et 6 doctorants. J'ai intégré ce laboratoire car il travaille sur la physique stellaire et Galactique et qu'il participe activement à la préparation de la mission Gaia, cela s'inscrit donc dans mon travail de thèse.

## 1. Activités conduites pendant la thèse

### 4.1 Recherche, stages, enseignements, formations et actions collectives

Durant la thèse, la plus grande partie de mon travail a été consacrée à la recherche, à la communication scientifique lors de colloques (présentations orales et posters) ainsi qu'à la publication de mes travaux dans des revues scientifiques.

Pendant 2 ans, j'ai également consacré un tiers de mon temps de travail à l'enseignement à l'Observatoire de Paris. J'ai été tutrice de plusieurs étudiants qui suivent une formation en ligne d'astrophysique. Cela consiste à répondre à leurs questions et corriger leurs exercices en ligne. J'ai également encadré des TPs d'informatique de M1, j'ai donné des cours et animé des TPs pour la formation des maîtres de l'Observatoire de Paris. Enfin j'ai participé à des événements pour le grand public à l'Observatoire ce qui m'a permis d'apprendre à vulgariser l'astrophysique et à communiquer.

Pendant ma thèse j'ai également suivi des formations «post-master» sur la vulgarisation scientifique, l'estimation bayésienne et la conduite de projets. Avec mon laboratoire j'ai participé à une formation du CNRS sur les risques au travail. De ma propre initiative j'ai suivi une formation sur le langage de programmation Python. Enfin dans le cadre de mon monitorat, j'ai suivi une formation pour apprendre à valoriser mes compétences ainsi qu'une formation aux premiers secours.

Lors de ma première année j'ai participé à une école d'une semaine sur les étoiles de petites masses. J'ai également participé à l'action collective dans l'Observatoire de Paris en étant représentante étudiante dans le comité hygiène, sécurité et conditions de travail pendant ma troisième année de thèse.

Ces activités ont été réalisées en parallèle du travail de recherche, ce qui m'a permis d'apprendre à gérer mon temps de travail de manière efficace.

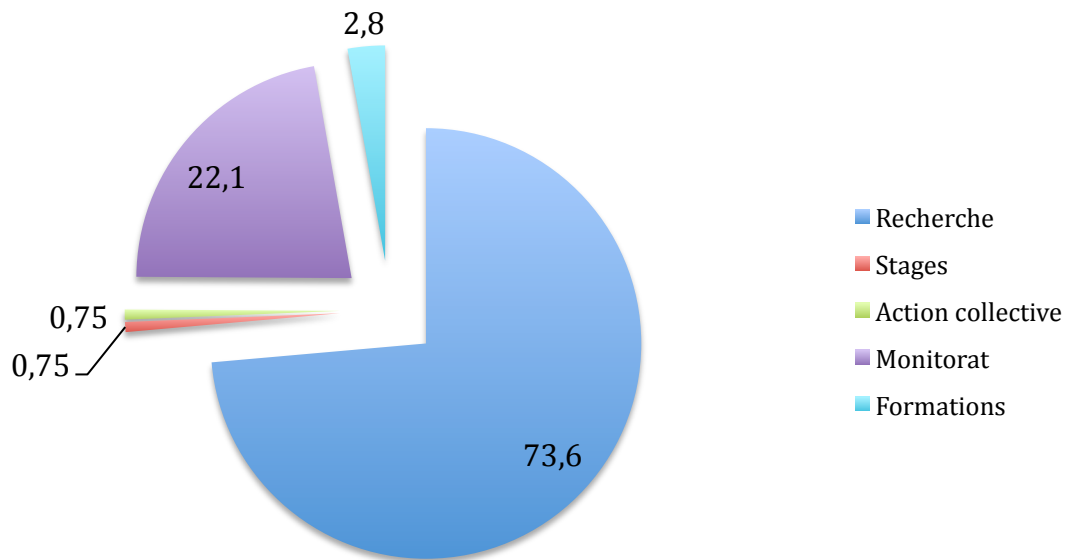


Fig. 2 : Activités conduites pendant la thèse

## 4.2 Conditions de thèse

J'ai bénéficié d'une bourse de trois ans de l'enseignement supérieur et de la recherche pour effectuer ma thèse. Elle a donc été encadrée par l'école doctorale d'astronomie et d'astrophysique d'Ile de France. Mes conditions de thèse étaient particulières car ma directrice de thèse est en mission permanente à Rennes alors que j'ai effectué ma thèse sur le site à Meudon. Chaque mois je suis allé en mission à Rennes afin de pouvoir travailler avec elle mais la grande partie de nos interactions s'est effectuée par télétravail. Cela m'a permis d'apprendre à travailler à distance et à devenir autonome dans mon travail.

### 1. Gestion du projet de recherche

#### 5.1 Equipe projet

L'équipe du projet est composée de 10 personnes mais j'ai principalement travaillé avec ma directrice de thèse Yveline Lebreton. J'ai travaillé avec trois chercheurs de l'observatoire (Misha Haywood, Carine Babusiaux et Didier Pelat) afin de collaborer sur certaines parties du projet. J'ai également interagi avec des chercheurs à Nice (Frédéric Thévenin, Bernard Pichon et Christophe Ordenovic) pour délivrer le programme à la mission spatiale Gaia du CNES. Enfin j'ai travaillé avec deux informaticiens (Malek Mekkas et Sylvestre Taburet) de mon laboratoire afin de créer une page web qui permet d'utiliser une version simplifiée de l'algorithme. Ce projet équivaut à employer une personne pendant presque 3 ans.

	Temps (mois)	Salaires et charges (k€)
Moi	23	69
Directrice de thèse	7	31,5
Collaborateurs (3)	2	9
Informaticiens (2)	2	6
Collaborateurs Gaia (3)	0,1	0,45
<b>Total</b>	<b>34,1</b>	<b>115,95</b>

*Tab. 1 : Temps de travail et salaire alloué au projet*

## 5.2 Etapes du projet

Lorsque j'ai commencé ma thèse la problématique et les objectifs étaient déjà définis par ma directrice de thèse. La première étape de la thèse a été de faire le point sur l'état des connaissances et des méthodes. Pour cela j'ai effectué une recherche bibliographique, ceci m'a permis de préciser les objectifs et de délimiter les champs de ma recherche. Ensuite j'ai défini les méthodes statistiques que j'ai utilisées durant mon travail. Puis j'ai commencé à développer l'algorithme de détermination des âges pour plusieurs méthodes. Puis j'ai réalisé plusieurs tests afin de choisir les paramètres les plus pertinents pour décrire l'étoile, son évolution et son environnement. Pour pouvoir réaliser ces tests, j'ai créé un catalogue d'étoiles simulées afin de vérifier les résultats. Par la suite, j'ai développé des outils afin de pouvoir comparer et exploiter les résultats. Dans cette étape j'ai comparé les différentes méthodes afin de sélectionner la plus performante. Ensuite, je suis passée à la rédaction d'articles pour présenter mon travail, cela permet également de mettre en perspective les prochains objectifs et les applications de mon travail. Enfin, afin de délivrer le programme à la mission Gaia, j'ai dû apprendre à travailler sous contrainte pour correspondre aux normes imposées par le CNES.

Le travail de thèse m'a permis d'acquérir des compétences en programmation, en informatique et en statistiques. J'ai également appris à communiquer mon travail à travers les colloques et les articles. Enfin, j'ai appris à travailler en équipe, à animer des réunions et à gérer un projet.

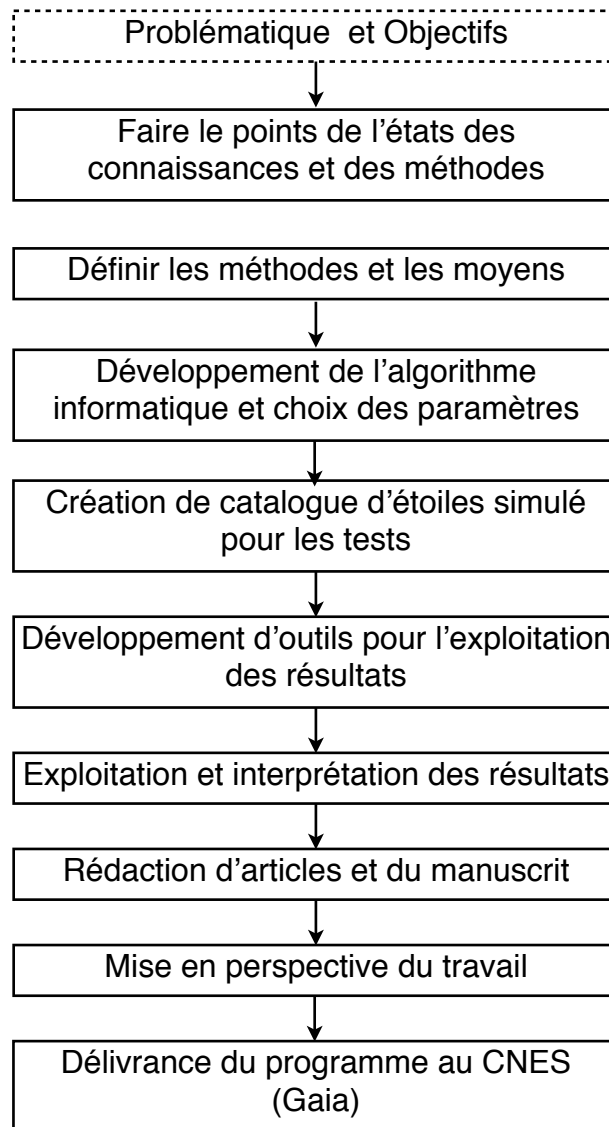


Fig. 3 : Etapes du projet de recherche

### 5.3 Budget du projet

Le projet de recherche a coûté 154 k€, dont la majorité a été allouée aux salaires et aux charges de l'équipe projet. J'ai bénéficié d'une bourse de l'enseignement supérieur de la recherche pour financer mes trois années de thèse. Les collaborateurs du projet dépendent du conseil national des astronomes et physiciens (CNAP) pour financer leurs recherches et les informaticiens dépendent du GEPI ou du CNRS. Durant ma thèse, j'ai effectué des calculs sur des serveurs qui sont financés par le GEPI. Il a également financé l'achat de l'ordinateur portable que j'ai utilisé pendant les trois ans ainsi qu'une partie de mes missions. Les autres missions ont été financées par le CNRS et la mission spatiale Gaia.

Dépenses (k€)		Financements (k€)	
Salaires et charges	115	Enseignement supérieur de la recherche	66
Coûts des infrastructures	33	CNAP	40
Investissement matériels	2	GEPI	39
Frais de fonctionnement		Observatoire de Paris	5
- missions	6	CNRS	3
- publications		AS Gaia	3
- consommables			
<b>Total</b>	<b>156</b>	<b>Total</b>	<b>156</b>

Tab. 2 : Dépenses et financements du projet de recherche

## 5.4 Organisation et suivi du projet

La majorité de mon travail de recherche est réalisé en autonomie avec un suivi régulier de ma directrice de thèse. Le télétravail m'a obligé à organiser des visioconférences ainsi que des réunions lors de mes déplacements à Rennes. Afin d'interagir avec les collaborateurs, nous avons effectué des réunions de travail qu'il m'est arrivé d'animer. Enfin j'ai également eu des réunions avec les informaticiens afin de créer une interface entre l'algorithme et la page web.

J'ai également eu plusieurs entrevues avec le directeur du GEPI qui souhaite rencontrer les étudiants afin de suivre l'évolution de la thèse et interagir sur les éventuelles difficultés rencontrées.

Ma thèse s'est effectuée au sein de l'école doctorale d'astronomie et d'astrophysique d'Ile de France ED 127. Afin de suivre l'avancement de mon travail de recherche, j'ai effectué une présentation mi-parcours devant les membres de l'école doctorale.

## 1. Compétences acquises

### 6.1 Expérience professionnelle

J'ai une expérience professionnelle de trois ans et demi dans la recherche. Cette expérience m'a permis d'apprendre à piloter et construire un projet de recherche et à travailler de façon autonome. J'ai également été assistante de recherche durant deux stages en Master. Le premier en première année sur *la conduction des neutrinos dans les supernovæ*. Durant le stage j'ai fait une recherche bibliographique sur l'état des connaissances et des méthodes sur ce domaine. Ce stage m'a permis d'apprendre à effectuer des recherches et à synthétiser les résultats. Le stage de la seconde année consistait à *comparer deux méthodes de détermination de la masse et du rayon des étoiles*. Ce stage m'a permis de me perfectionner en informatique et de faire mes premiers pas dans un projet de recherche. Enfin j'ai été enseignante durant deux ans pendant ma thèse (30% de mon temps de travail par an).

### 6.2 Compétences

Ce projet de recherche m'a permis de développer des connaissances scientifiques en

physique stellaire et en physique Galactique. J'ai également développé d'autres compétences :

### **1. Connaissance du monde de la recherche publique**

J'ai travaillé trois ans au sein de l'Observatoire de Paris ce qui m'a permis de connaître et de comprendre les acteurs de ce secteur, la gestion des crédits, le fonctionnement et la dynamique d'un institut de recherche publique international.

### **2. Gestion de projet de recherche et d'étude**

Ces trois années de thèse m'ont permis de gérer mon projet de recherche et d'étude en travaillant sur plusieurs aspects en parallèle. J'ai développé une capacité à résoudre les problèmes en utilisant différentes techniques. Grâce à mon projet de recherche j'ai développé une expertise en algorithmique statistique.

### **3. Connaissance de l'organisation d'une mission spatiale**

En travaillant sur la mission spatiale Gaia, j'ai appris à connaître le mode de fonctionnement et la gestion d'un projet spatial. Je comprends aussi l'interaction et la dynamique entre différents groupes de travail.

### **4. Maîtrise d'algorithme statistique**

La plus grande partie de mon travail a été de développer plusieurs algorithmes. J'ai également appris à les optimiser afin de pouvoir les délivrer au CNES pour la mission spatiale Gaia et pour l'interface avec la page web. Une grande partie de ces algorithmes sont statistiques, j'ai donc acquis des compétences en statistique en étudiant différentes méthodes que j'ai implémentées et comparées. Pour ce travail j'ai également dû comprendre et utiliser plusieurs méthodes mathématiques.

### **5. Maîtrise des outils numériques**

Je maîtrise plusieurs langages informatiques (Fortran, Python, C, Unix, IDL/GDL) et je suis actuellement capable d'apprendre un nouveau langage seule en quelques semaines. Comme j'ai utilisé couramment l'informatique je maîtrise également plusieurs systèmes d'exploitation (Mac OS, Linux, Windows) ainsi que des logiciels de présentation et de traitement de texte (LaTeX, Microsoft Word, Excel, Power Point).

Plus personnellement, je peux me décrire comme une personne motivée, enthousiaste et responsable mais je suis également timide et exigeante envers mon travail.

## **Projet professionnel**

Durant mes trois années de thèse j'ai développé un algorithme statistique afin de déterminer l'âge des étoiles. J'ai réussi à accomplir ce projet dans les délais qui m'étaient donnés. Ce programme va être implémenté par le CNES afin d'être utilisé dans l'exploitation des données de la mission spatiale Gaia. J'ai également présenté ce travail dans des revues scientifiques et lors de colloques internationaux.

Ce travail m'a permis de me conforter dans l'idée que le métier de chercheur est très enrichissant ce qui me conforte à continuer dans cette voie. Par contre la durée des projets en astrophysique est longue. En effet les algorithmes que j'ai développés pour la mission spatiale Gaia ne seront utilisés que dans quelques années. C'est pourquoi je souhaite travailler en tant qu'ingénieur d'études en Recherche et Développement dans l'industrie, afin de développer des algorithmes et des outils statistiques. La grande partie des interactions sur mon projet de recherche était par internet, maintenant je veux travailler dans un groupe de travail afin de favoriser les interactions et les synergies.

# Bibliographie

- Alonso, A., Arribas, S., & Martínez-Roger, C. 1996, *A&A*, 313, 873–108
- Alonso, A., Arribas, S., & Martínez-Roger, C. 1999, *A&AS*, 140, 261–114
- Baglin, A., Auvergne, M., Barge, P., et al. 2006, in *ESA Special Publication*, Vol. 1306, *ESA Special Publication*, ed. M. Fridlund, A. Baglin, J. Lochard, & L. Conroy, 33–85, 87
- Bailer-Jones, C. A. L. 2010, *MNRAS*, 403, 96–54
- Barnes, S. A. 2009, in *IAU Symposium*, Vol. 258, *IAU Symposium*, ed. E. E. Mamajek, D. R. Soderblom, & R. F. G. Wyse, 345–356–8
- Bedding, T. R. & Kjeldsen, H. 2003, *PASA*, 20, 203–86
- Bertelli, G., Bressan, A., Chiosi, C., Fagotto, F., & Nasi, E. 1994, *A&AS*, 106, 275–30, 141, 143, 144
- Böhm-Vitense, E. 1958, *ZAp*, 46, 108–29
- Borucki, W., Koch, D., Boss, A., et al. 2004, in *ESA Special Publication*, Vol. 538, *Stellar Structure and Habitable Planet Finding*, ed. F. Favata, S. Aigrain, & A. Wilson, 177–182–85, 87
- Burkhart, C. & Coupry, M. F. 1991, *A&A*, 249, 205–109
- Casagrande, L., Ramírez, I., Meléndez, J., Bessell, M., & Asplund, M. 2010, *A&A*, 512, A54–xv, 11, 112, 113, 114
- Casagrande, L., Schönrich, R., Asplund, M., et al. 2011, *A&A*, 530, A138–viii, xiii, 20, 21, 22, 47, 48, 53, 76, 111, 127
- Cassisi, S., Salaris, M., & Irwin, A. W. 2003, *ApJ*, 588, 862–29
- Chabrier, G. 2003, *PASP*, 115, 763–26, 116
- Christensen-Dalsgaard, J. 2012, *African Skies*, 16, 74–88
- Cowan, J. J., Thielemann, F.-K., & Truran, J. W. 1991, *ARA&A*, 29, 447–8
- Crawford, D. L. 1975, *AJ*, 80, 955–107, 108
- Creevey, O. L., Doğan, G., Frasca, A., et al. 2012, *A&A*, 537, A111–95

- Cuisinier, F., Buser, R., Acker, A., et al. 1994, *A&A*, 285, 943–93
- da Silva, L., Girardi, L., Pasquini, L., et al. 2006, *A&A*, 458, 609–viii, 11, 19, 47, 48, 114, 115
- de Bruijne, J. H. J., Hoogerwerf, R., & de Zeeuw, P. T. 2001, *A&A*, 367, 111–120, 122
- Deliyannis, C. P. & Steinhauer, A. 2001, in *Bulletin of the American Astronomical Society*, Vol. 33, American Astronomical Society Meeting Abstracts 198, 843–120, 122
- di Benedetto, G. P. 1998, *A&A*, 339, 858–110
- Drimmel, R., Cabrera-Lavers, A., & López-Corredoira, M. 2003, *A&A*, 409, 205–54, 99, 100
- Edvardsson, B., Andersen, J., Gustafsson, B., et al. 1993, *A&A*, 275, 101–9, 11, 109
- Feltzing, S. & Bensby, T. 2009, in *IAU Symposium*, Vol. 258, IAU Symposium, ed. E. E. Mamajek, D. R. Soderblom, & R. F. G. Wyse, 23–30–8
- Flynn, C. & Morell, O. 1997, *MNRAS*, 286, 617–109
- Gilmore, G., Randich, S., Asplund, M., et al. 2012, *The Messenger*, 147, 25–126
- Girardi, L., Bressan, A., Bertelli, G., & Chiosi, C. 2000, *A&AS*, 141, 371–30, 110, 115, 141
- Girardi, L., Bressan, A., Chiosi, C., Bertelli, G., & Nasi, E. 1996, *A&AS*, 117, 113–30, 141, 143, 144
- Glaspey, J. W., Pritchett, C. J., & Stetson, P. B. 1994, *AJ*, 108, 271–109
- Grevesse, N. & Noels, A. 1993, in *Origin and Evolution of the Elements*, ed. N. Prantzos, E. Vangioni-Flam, & M. Casse, 15–25–29
- Guédé, C., Lebreton, Y., Babusiaux, C., & Haywood, M. 2012a, in *SF2A-2012 : Proceedings of the Annual meeting of the French Society of Astronomy and Astrophysics*, ed. S. Boissier, P. de Laverny, N. Nardetto, R. Samadi, D. Valls-Gabaud, & H. Wozniak, 83–86–192



- Guédé, C., Lebreton, Y., Babusiaux, C., & Haywood, M. 2012b, in SF2A-2012 : Proceedings of the Annual meeting of the French Society of Astronomy and Astrophysics, ed. S. Boissier, P. de Laverny, N. Nardetto, R. Samadi, D. Valls-Gabaud, & H. Wozniak, 195–198 197
- Guédé, C., Lebreton, Y., & Dréan, G. 2011, in SF2A-2011 : Proceedings of the Annual meeting of the French Society of Astronomy and Astrophysics, ed. G. Alecian, K. Belkacem, R. Samadi, & D. Valls-Gabaud, 295–298 187
- Hasegawa, T., Malasan, H. L., Kawakita, H., et al. 2004, PASJ, 56, 295 120, 122
- Hekker, S., Elsworth, Y., De Ridder, J., et al. 2011, A&A, 525, A131 89
- Hernandez, X. & Valls-Gabaud, D. 2008, MNRAS, 383, 1603 11, 66
- Høg, E., Fabricius, C., Makarov, V. V., et al. 2000, A&A, 355, L27 112
- Holmberg, J., Nordström, B., & Andersen, J. 2007, A&A, 475, 519 110
- Holmberg, J., Nordström, B., & Andersen, J. 2009, A&A, 501, 941 111, 114
- Huber, D., Bedding, T. R., Stello, D., et al. 2011, ApJ, 743, 143 87, 89
- Irwin, J. & Bouvier, J. 2009, in IAU Symposium, Vol. 258, IAU Symposium, ed. E. E. Mamajek, D. R. Soderblom, & R. F. G. Wyse, 363–374 8
- Johnson, H. L. & Morgan, W. W. 1953, ApJ, 117, 313 18
- Jordi, C., Gebran, M., Carrasco, J. M., et al. 2010, A&A, 523, A48 5, 6, 53
- Jørgensen, B. R. & Lindegren, L. 2005, A&A, 436, 127 vii, viii, 11, 12, 17, 19, 20, 36, 39, 40, 42, 47, 110, 111
- Keenan, P. C. & Morgan, W. W. 1951, in 50th Anniversary of the Yerkes Observatory and Half a Century of Progress in Astrophysics, ed. J. A. Hynek, 12 10
- Kippenhahn, R., Weigert, A., & Weiss, A. 2013, Stellar Structure and Evolution 28
- Kiraga, M. & Stepien, K. 2007, Acta Astron., 57, 149 1
- Kjeldsen, H. & Bedding, T. R. 1995, A&A, 293, 87 86, 87

- Kjeldsen, H., Bedding, T. R., & Christensen-Dalsgaard, J. 2008, in American Institute of Physics Conference Series, Vol. 1043, American Institute of Physics Conference Series, ed. V. Mioc, C. Dumitrche, & N. A. Popescu, 365–372 89
- Kroupa, P. 2001, MNRAS, 322, 231–25
- Kroupa, P. 2002, Science, 295, 82–20, 25, 26, 71
- Kroupa, P., Tout, C. A., & Gilmore, G. 1993, MNRAS, 262, 545–24, 116
- Lakatos, S. L., Nandy, D., & Martens, P. 2005, in Bulletin of the American Astronomical Society, Vol. 37, American Astronomical Society Meeting Abstracts, 111.04–1
- Lebreton, Y., Fernandes, J., & Lejeune, T. 2001, A&A, 374, 540–29, 122
- Lebreton, Y. & Montalbán, J. 2009, in IAU Symposium, Vol. 258, IAU Symposium, ed. E. E. Mamajek, D. R. Soderblom, & R. F. G. Wyse, 419–430 9
- Lebreton, Y., Monteiro, M. J. P. F. G., Montalbán, J., et al. 2008, Ap&SS, 316, 1–89
- Liu, C., Bailer-Jones, C. A. L., Sordo, R., et al. 2012, MNRAS, 426, 2463–54
- Mamajek, E. E. 2009, in IAU Symposium, Vol. 258, IAU Symposium, ed. E. E. Mamajek, D. R. Soderblom, & R. F. G. Wyse, 375–382 8
- Mazumdar, A. 2005, A&A, 441, 1079–9
- Miller, G. E. & Scalo, J. M. 1979, ApJS, 41, 513–24
- Morel, P. & Lebreton, Y. 2008, Ap&SS, 316, 61–29, 141
- Moutou, C. & Pont, F. 2006, Ecole de Goutelas, 28, 55–92
- Mowlavi, N., Schaerer, D., Meynet, G., et al. 1998, A&AS, 128, 471–110
- Ng, Y. K. & Bertelli, G. 1998, A&A, 329, 943–viii, 11, 47, 49, 50
- Nordström, B., Mayor, M., Andersen, J., et al. 2004, A&A, 418, 989–107
- Olsen, E. H. 1983, A&AS, 54, 55–107
- Olsen, E. H. 1984, A&AS, 57, 443–108
- Olsen, E. H. 1988, A&A, 189, 173–107, 108

- Olsen, E. H. 1993, *A&AS*, 102, 89–107
- Olsen, E. H. 1994a, *A&AS*, 104, 429–107
- Olsen, E. H. 1994b, *A&AS*, 106, 257–107
- Otí Floranes, H., Christensen-Dalsgaard, J., & Thompson, M. J. 2005, *MNRAS*, 356, 671–9
- Peimbert, M., Luridiana, V., & Peimbert, A. 2007, *ApJ*, 666, 636–29
- Perryman, M. A. C. 1989, *Nature*, 340, 111–2
- Perryman, M. A. C., de Boer, K. S., Gilmore, G., et al. 2001, *A&A*, 369, 339–4, 53
- Pietrinferni, A., Cassisi, S., Salaris, M., & Castelli, F. 2004, *ApJ*, 612, 168–20, 29, 53, 141, 142
- Pont, F. & Eyer, L. 2004, *MNRAS*, 351, 487–11
- Prather, M. J. 1976, PhD thesis, Yale University., New Haven, CT. 70, 71, 141
- Reimers, D. 1975, *Memoires of the Societe Royale des Sciences de Liege*, 8, 369–29
- Robert, C. 2007, *The Bayesian choice, From Decision-Theoretic Foundations to Computational Implementation*, 2nd edition (New York : Springer Verlag) 18
- Robin, A. C., Luri, X., Reylé, C., et al. 2012, *A&A*, 543, A100–57, 100
- Ryter, C. E. 1996, *Ap&SS*, 236, 285–121
- Salaris, M., Chieffi, A., & Straniero, O. 1993, *ApJ*, 414, 580–110
- Salpeter, E. E. 1955, *ApJ*, 121, 161–23
- Santos, N. C., Israelian, G., & Mayor, M. 2001, *A&A*, 373, 1019–109
- Scalo, J. 1998, in *Astronomical Society of the Pacific Conference Series*, Vol. 142, *The Stellar Initial Mass Function (38th Herstmonceux Conference)*, ed. G. Gilmore & D. Howell, 201–25, 116
- Schuster, W. J. & Nissen, P. E. 1989, *A&A*, 221, 65–109

- Setiawan, J., Pasquini, L., da Silva, L., et al. 2004, *A&A*, 421, 241 ix, xiii, 107, 114, 115, 116
- Skrutskie, M. F., Cutri, R. M., Stiening, R., et al. 2006, *AJ*, 131, 1163 112
- Skumanich, A. 1972, *ApJ*, 171, 565 8
- Soderblom, D. R. 2010, *ARA&A*, 48, 581 8
- Soderblom, D. R., Stauffer, J. R., Hudon, J. D., & Jones, B. F. 1993, *ApJS*, 85, 315 8
- Southworth, J. 2008, *MNRAS*, 386, 1644 92
- Strömberg, B. 1965, in *Stellar Structure - Stars and Stellar Systems*, ed. L. H. Aller & D. B. McLaughlin, 269 107
- Takeda, G., Ford, E. B., Sills, A., et al. 2007, *ApJS*, 168, 297 11, 36, 96
- Taylor, B. J. 2003, *A&A*, 398, 731 109
- Thorén, P. & Feltzing, S. 2000, *A&A*, 363, 692 109
- Tomkin, J. & Lambert, D. L. 1999, *ApJ*, 523, 234 109
- Torres, G., Winn, J. N., & Holman, M. J. 2008, *ApJ*, 677, 1324 92
- Turon, C. & Arenou, F. 2009, *L'astronomie*, 13, 16 4
- van Leeuwen, F. 2007, *A&A*, 474, 653 111
- VandenBerg, D. A. & Clem, J. L. 2003, *AJ*, 126, 778 55
- VandenBerg, D. A., Swenson, F. J., Rogers, F. J., Iglesias, C. A., & Alexander, D. R. 2000, *ApJ*, 532, 430 110
- Verner, G. A., Elsworth, Y., Chaplin, W. J., et al. 2011, *MNRAS*, 415, 3539 89
- Watson, A. 1998, *Science*, 279, 981 2
- Wielen, R. 1977, *A&A*, 60, 263 8
- Yadav, R. K. S. & Sagar, R. 2004, *MNRAS*, 349, 1481 120, 121, 122

