

# THÈSE DE DOCTORAT

de l'Université de recherche Paris Sciences et Lettres  
PSL Research University

Préparée dans le cadre d'une cotutelle entre  
l'Observatoire de Paris  
et le CNES

Gaia: from the data validation to the Red Clump parameters

**Gaia : de la validation des données aux paramètres du Red Clump**

**Ecole doctorale n°127**

ASTRONOMIE ET ASTROPHYSIQUE D'ÎLE-DE-FRANCE

**Spécialité** ASTRONOMIE ET ASTROPHYSIQUE

**Soutenue par Laura RUIZ-DERN  
le 8 novembre 2016**

Dirigée par Carine BABUSIAUX  
et Frédéric ARENOU

## COMPOSITION DU JURY :

M. Martin GROENEWEGEN  
Observatoire Royal de Belgique  
Rapporteur

Mme Carme JORDI  
Universitat de Barcelona, ICCUB-IEEC  
Examinatrice

M. Réza SAMADI  
Observatoire de Paris  
Examineur

M. Alessandro SPAGNA  
INAF, Osservatorio Astronomico di Torino  
Rapporteur

l'Observatoire  
de Paris

 **cnés**

CENTRE NATIONAL D'ÉTUDES SPATIALES

**PSL**   
RESEARCH UNIVERSITY PARIS





Thèse préparée à

**l'Observatoire de Paris**

Galaxies, Étoiles, Physique et Instrumentation (GEPI, CNRS UMR 8111)

5 Place Jules Janssen

92190 Meudon CEDEX (France)



*To my family and friends*



# Aknowledgements

*"It always seems impossible until it's done"*

– Nelson Mandela

Après trois ans de recherche, dont six mois de rédaction de ce mémoire, vous conviendrez avec moi que cette page sera probablement la plus lue. Mais ne venez pas chercher votre nom ici. Le nombre de personnes qui m'ont aidé, soutenu, encouragé et contribué à la réalisation de cette thèse, d'une façon ou d'une autre, visible ou invisible, de près ou de loin, consciemment ou inconsciemment, est trop grand. Dans l'impossibilité de vous mentionner tous, et devant le risque d'en oublier: merci infiniment à tous.

Je tiens toutefois à exprimer toute ma reconnaissance et ma gratitude envers mes directeurs de thèse, Carine Babusiaux et Frédéric Arenou. À Carine pour tout le temps qu'elle m'a consacré, son aide, son exigence, sa vision critique de la recherche, et tout ce qu'elle m'a appris. À Frédéric, pour sa disponibilité constante, sa préoccupation, sa confiance en moi, sa patience, son implication et son aide précieuse, même au-delà de la thèse.

Je voudrais spécialement remercier Catherine Turon et Noël Robichon, ainsi que souligner combien j'ai apprécié la compagnie des collègues du bâtiment Hipparque. Merci pour votre accueil, votre aide et vos encouragements au cours de ces années.

Je désire en outre remercier Carme Jordi, Cesca Figueras et Xavi Luri, de l'*Universitat de Barcelona*, pour m'avoir soutenu, mis sur la voie et permis de faire cette thèse.

Et curieusement, je souhaite également remercier les enfants et les enseignants des parrainages qui, sans le savoir, m'ont souvent donné la force dont j'avais besoin pour arriver jusqu'à la fin.

*I finalment el meu més profund agraïment l'adreço als meus amics - "bslcrmae" - per haver-me acompanyat en la distància i la dificultat. I sobretot a la meva família, en especial als meus pares i ma germana, per haver-me ajudat, recolzat, animat i valorat en tot moment. En poques paraules: per tot allò que realment importa. Sóc conscient que sense vosaltres no hagués mai arribat fins aquí, i que és a vosaltres a qui ho dec tot. Gràcies és poc, però... gràcies. A vosaltres dedico aquesta tesis.*

Un grand merci à tous.





## Abstract

The Gaia mission of the European Space Agency (ESA) aims to map our galaxy with an unprecedented astrometric precision. It is therefore very important that the data that will be published be rigorously validated to ensure an optimal quality in the Catalogue. These validations are done by one of the teams of the coordination unit CU9 of the Gaia DPAC Consortium (Data Processing and Analysis Consortium) commissioned by ESA of the Gaia catalogue production. As part of this thesis, we implemented all the necessary infrastructure to validate the Gaia catalogue by comparison with external catalogues. This last manages all the interactions with the global environment of validations and with the Gaia database. Then we developed a set of statistical tests to validate the data from the first Gaia catalogue (DR1). These tests relate in particular to the homogeneity of data on the sky, the quality of the positions and of photometry of all the stars of DR1 (more than a billion stars,  $V < 20$ ) as well as that of the parallaxes and proper motions for *Tycho-Gaia* Astrometric Solution (TGAS) stars, around two million stars in common in Gaia and Tycho-2 catalogues ( $V < 12$ ).

These DR1 statistical tests are operational and were already applied very recently on preliminary data. This has improved the data (thus the quality of the catalogue) as well as allowed to characterize the statistical properties. This characterisation is essential for a correct scientific exploitation of the data. The first Gaia catalogue will be released in late summer 2016.

Among the objects that Gaia observes, there is a population of stars particularly interesting, the Red Clump (RC) stars, widely used for distance indicators. We developed and tested two methods to model the colour-colour (CC) and effective temperature - colour relations in all photometric bands, from the ultraviolet to the near infrared. They will allow us to characterize the RC in the Gaia G band upon publication of the catalogue: 1. using theoretical models, and 2. empirically, based on a Monte Carlo Markov Chain (MCMC) method. To do so, we have very carefully selected samples of stars with a good photometric quality, a good metallicity determined by spectroscopy, an homogeneous effective temperature and a low interstellar extinction.

From these CC and temperature-colour calibrations, we then developed a method that allows to derive intrinsic magnitudes, temperatures and extinctions of the RC stars. Estimates of colours and extinctions are tested on stars with spectroscopically measured effective temperatures and an extinction determined by the measurement of Diffuse Interstellar Bands (DIB). These intrinsic properties of RC stars will allow to characterize the Gaia RC and calibrate, within the Gaia G band, the absolute magnitude of this standard candle, first essential step of determining distances in the Universe.



## Résumé

La mission Gaia de l'Agence Spatiale Européenne (ESA) a pour objectif de cartographier notre galaxie avec une précision astrométrique jamais atteinte auparavant. Il est donc particulièrement important que les données qui seront publiées soient rigoureusement validées afin d'assurer une qualité optimum au Catalogue. Ces validations sont faites par l'une des équipes de l'unité de coordination CU9 du Consortium Gaia DPAC (Data Processing and Analysis Consortium) chargé par l'ESA de la production du Catalogue Gaia. Dans le cadre de cette thèse, nous avons mis en place toute l'infrastructure nécessaire à la validation du catalogue Gaia par comparaison avec des catalogues externes. Celle-ci gère toutes les interactions avec l'environnement global des validations et avec la base de données Gaia. Ensuite nous avons développé un ensemble de tests statistiques pour valider les données du premier catalogue Gaia (DR1). Ces tests concernent notamment l'homogénéité des données sur le ciel, la qualité des positions et de la photométrie de l'ensemble des étoiles de DR1 (plus d'un milliard d'étoiles,  $V < 20$ ) ainsi que celle des parallaxes et mouvements propres des étoiles de *Tycho-Gaia* Astrometric Solution (TGAS), environ deux millions d'étoiles communes aux catalogues Gaia et Tycho-2 ( $V < 12$ ).

Ces tests statistiques sur la DR1 sont opérationnels et ont déjà été appliqués très récemment sur des données préliminaires. Cela a déjà permis d'améliorer ces données (donc la qualité du catalogue), et d'en caractériser les propriétés statistiques. Cette caractérisation est essentielle à une exploitation scientifique correcte des données. Le premier catalogue Gaia sera publié à la fin de l'été 2016.

Parmi les objets observés par Gaia, il y a une population d'étoiles particulièrement intéressantes, les étoiles du Red Clump (RC), très utilisées comme étalons de distance. Nous avons développé et testé deux méthodes pour modéliser les relations couleur-couleur (CC) et température effective - couleur dans toutes les bandes photométriques, de l'ultraviolet au proche-infrarouge. Elles permettront de caractériser le RC dans la bande G de Gaia dès la publication du catalogue: 1. en utilisant des modèles théoriques, et 2. empiriquement, en se basant sur une méthode Monte Carlo Markov Chain (MCMC). Pour cela nous avons très rigoureusement sélectionné des échantillons d'étoiles avec une bonne qualité photométrique, une bonne métallicité, déterminée par spectroscopie, une température effective homogène et une faible extinction interstellaire.

À partir de ces calibrations CC et température-couleur, nous avons ensuite développé une méthode qui permet de déterminer les magnitudes intrinsèques, les températures et les extinctions des étoiles du RC. Les couleurs et extinctions ainsi obtenues ont été testées sur des étoiles avec des températures effectives mesurées spectroscopiquement et une extinction déterminée par la mesure des Bandes Diffuses Interstellaires (DIB). Ces propriétés intrinsèques des étoiles du RC vont permettre de caractériser le Red Clump Gaia et de calibrer, dans la bande Gaia, la magnitude absolue de cet étalon de distance, premier échelon essentiel de la détermination des distances dans l'Univers.



# Contents

<b>Contents</b>	<b>13</b>
<b>List of Figures</b>	<b>17</b>
<b>List of Tables</b>	<b>21</b>
<b>Abbreviations</b>	<b>23</b>
<b>Introduction</b>	<b>27</b>
<b>I Calibrating the Gaia Red Clump</b>	<b>29</b>
<b>1 Giants: state of knowledge</b>	<b>31</b>
1.1 Becoming a Giant star . . . . .	32
1.1.1 Formation . . . . .	32
1.1.2 Evolution to a giant . . . . .	32
1.1.3 Giant phases . . . . .	33
1.2 Giant features on a Colour-Magnitude Diagram . . . . .	36
1.2.1 Red Clump . . . . .	36
1.2.2 Vertical Red Clump . . . . .	42
1.2.3 Secondary Red Clump . . . . .	43
1.2.4 Red and Asymptotic Giant Branch Bumps . . . . .	45
1.2.5 Double RC-RGBB pairs . . . . .	50
1.2.6 An unexplained feature? . . . . .	55
1.2.7 Synopsis . . . . .	55
1.3 Red Clump characterization . . . . .	58
1.3.1 Modelling the RC . . . . .	58
1.3.2 The need of a calibration: this thesis . . . . .	60
<b>2 Red Clump calibration (I): atmosphere models</b>	<b>63</b>
2.1 Observations and models . . . . .	64
2.2 Data selection . . . . .	64
2.2.1 RC sample . . . . .	64
2.2.2 Isochrones . . . . .	68
2.3 Gap correction: MLE method . . . . .	71
2.4 Results . . . . .	76
2.5 Discussion . . . . .	77

2.5.1	Spectral libraries: observed and theoretical . . . . .	77
2.5.2	Clusters . . . . .	78
<b>3</b>	<b>Red Clump calibration (II): empirical (Final solution)</b>	<b>81</b>
3.1	Overview . . . . .	82
3.2	Colour-colour calibration . . . . .	82
3.2.1	Data selection . . . . .	82
3.2.2	MCMC method . . . . .	88
3.2.3	Results . . . . .	91
3.3	Effective Temperature - Colour calibration . . . . .	95
3.3.1	Data selection . . . . .	95
3.3.2	MCMC method . . . . .	96
3.3.3	Results . . . . .	97
3.4	Discussion . . . . .	101
3.4.1	Comparison with other $T_{\text{eff}}$ -Colour relations . . . . .	101
<b>4</b>	<b>Deriving RC physical parameters</b>	<b>105</b>
4.1	Introduction . . . . .	106
4.2	Interstellar extinction coefficients model . . . . .	106
4.3	$[A_0, T_{\text{eff}}, G_0]$ -method . . . . .	106
4.4	Applications . . . . .	110
4.4.1	Galactic Bulge . . . . .	110
4.4.2	APOGEE survey . . . . .	114
4.4.3	Applications with TGAS . . . . .	114
4.4.4	Discussion . . . . .	117
<b>II</b>	<b>Gaia: confrontation with external archives</b>	<b>119</b>
<b>1</b>	<b>The Gaia mission</b>	<b>121</b>
1.1	An astrometric space observatory . . . . .	122
1.1.1	Overview . . . . .	122
1.1.2	Science objectives . . . . .	123
1.1.3	The spacecraft . . . . .	124
1.1.4	Observation principle . . . . .	127
1.1.5	Detected problems and final performances . . . . .	129
1.1.6	Catalogue Releases . . . . .	131
1.2	News from Gaia . . . . .	132
1.3	The Gaia Data Processing and Analysis Consortium . . . . .	134
1.3.1	CU9: the last step . . . . .	135
<b>2</b>	<b>WP944: confrontation with external archives</b>	<b>137</b>
2.1	Underneath framework and purposes of WP944 . . . . .	138
2.1.1	Goals of WP944 . . . . .	139
2.1.2	Specificity: external archives . . . . .	139
2.1.3	Coverage: parameters and objects . . . . .	140
2.2	Infrastructure . . . . .	141
2.2.1	Start-up . . . . .	141
2.2.2	Crossmatch processing . . . . .	141
2.2.3	Global tests management . . . . .	143

2.3	General validation process . . . . .	144
2.3.1	Databases used in validation . . . . .	144
2.3.2	Validation process . . . . .	144
2.3.3	Pass/Fail criteria . . . . .	145
2.3.4	JUnit tests . . . . .	145
<b>3</b>	<b>Validation tests and contributions</b>	<b>147</b>
3.1	Overview . . . . .	148
3.2	External parallaxes and proper motions . . . . .	149
3.2.1	Preliminary external catalogues . . . . .	149
3.2.2	$\chi^2$ method . . . . .	149
3.3	Parallax Zero-Point and precision . . . . .	151
3.3.1	External Distances: RR Lyrae . . . . .	152
3.3.2	Very distant stars: LMC/SMC . . . . .	156
3.4	Specific contributions . . . . .	157
3.4.1	Sky Homogeneity: HEALPix maps . . . . .	157
3.4.2	Epoch Transformation . . . . .	159
<b>4</b>	<b>Improving the Gaia Data Release 1 solution</b>	<b>165</b>
4.1	CU9 validation releases . . . . .	166
4.2	Parallaxes . . . . .	166
4.2.1	Under-estimated uncertainties . . . . .	166
4.2.2	Local zero-point variations . . . . .	167
4.3	Proper Motions . . . . .	168
4.3.1	Under-estimated errors . . . . .	168
4.3.2	Missing High Proper Motion stars . . . . .	169
4.3.3	Gaia reference system rotation . . . . .	169
4.4	Sky Coverage . . . . .	170
4.4.1	Missing sources . . . . .	170
4.4.2	Duplicated sources . . . . .	171
4.5	Outliers . . . . .	172
4.5.1	Astrometric . . . . .	172
4.5.2	Photometric . . . . .	172
	<b>Conclusions and Perspectives</b>	<b>175</b>
4.6	Conclusions . . . . .	176
4.6.1	The Red Clump photometric calibration . . . . .	176
4.6.2	TGAS and the Gaia Data Release 1 . . . . .	178
4.7	Perspectives . . . . .	179
	<b>Bibliography</b>	<b>181</b>
	<b>Appendix A 3D Covariance Matrix</b>	<b>199</b>
A.1	Notions of Normal Density Distributions . . . . .	200
A.2	Gaia and External catalogues distributions . . . . .	201
A.2.1	Convolved distribution . . . . .	202
A.2.2	Global covariance matrix . . . . .	206
A.2.3	The $\chi^2$ variate . . . . .	207
	<b>Appendix B WP944 v18.1 Detailed Description and Validation Tests Report</b>	<b>209</b>





# List of Figures

## I Calibrating the Gaia Red Clump

1.1	Evolutionary Tracks in an HR diagram . . . . .	32
1.2	Star nuclei evolution from MS to AGB . . . . .	35
1.3	Identification of RC stars using the Gaussian model . . . . .	39
1.4	Median period variation ( $\Delta P$ ) vs frequency variation ( $\Delta \nu$ ) to identify giant phases . . . . .	39
1.5	Prediction for the conditional distribution of stars in surface gravity given their effective temperature from the PARSEC stellar isochrones and data from the APOKASC catalog, in two metallicity bins . . . . .	40
1.6	RC fine structure in synthetic $M_B - M_V$ and $M_V - M_I$ CMDs . . . . .	41
1.7	Vertical Red Clump in the <i>Hipparcos</i> CMD . . . . .	43
1.8	Distribution of RC stars in the $M_V$ versus $B - V$ diagram, from theoretical models and <i>Hipparcos</i> data . . . . .	43
1.9	<i>APOKASC</i> stars identified asteroseismically as secondary RC, RC, or first-ascent RGB . . . . .	44
1.10	Theoretical dependence of the red-giant luminosity function with the abundance, mass, and metallicity . . . . .	45
1.11	Colour-magnitude diagrams and magnitude distributions of globular clusters <i>47 Tuc</i> and <i>NGC 1261</i> . . . . .	46
1.12	The AGBB of <i>47 Tuc</i> . . . . .	47
1.13	Luminosity function for the RGB and AGB populations predicted by Padova stellar models. Observed and synthetic CMDs for M31 halo field . . . . .	48
1.14	Supraclump and VRC in the calibrated composite ( $V - R, V$ ) CMDs for four LMC fields . . . . .	48
1.15	Bump-HB magnitude difference dependence on metallicity . . . . .	49
1.16	Asteroseismic secondary RC from <i>Kepler</i> , wrong identification? . . . . .	51
1.17	Asteroseismic secondary RC using <i>APOKASC</i> . . . . .	51
1.18	Asteroseismic RC and RGBB in a $T_{\text{eff}}\text{-logg}$ diagram from combined <i>LAMOST</i> and <i>Kepler</i> surveys . . . . .	52
1.19	Double RC in the CMD of <i>GC Terzan 5</i> . . . . .	53
1.20	Double RC in the CMD of the Galactic Bulge . . . . .	54
1.21	Surface gravity versus effective temperature for 3865 <i>LAMOST-Kepler</i> stars . . . . .	55
1.22	Surface gravity versus effective temperature for <i>UVES</i> stars in the <i>Gaia-ESO Survey</i> . . . . .	56
1.23	Giant's CMD features schematic representation . . . . .	56
1.24	<i>Hipparcos</i> RC modelling . . . . .	59
1.25	<i>Gaia</i> RC modelling . . . . .	60
1.26	<i>Hipparcos Survey</i> Hertzsprung-Russell diagram . . . . .	61

1.27	$B - V$ vs $V - I$ and $J - K$ vs $V - K$ colour-colour diagrams of RC stars: observations versus isochrones . . . . .	62
2.1	RC sample on the <i>Hipparcos</i> HR diagram . . . . .	66
2.2	Distribution of the RC sample on an Aitoff map projection in galactic coordinates . . . . .	66
2.3	Distributions of colours and spectral types of the RC sample . . . . .	67
2.4	Metallicity distribution of the RC sample . . . . .	67
2.5	Extinction distribution of the RC sample . . . . .	68
2.6	<i>Padova</i> isochrones evolution with age and metallicity in a $B - V$ vs $V - I$ CC diagram . . . . .	69
2.7	<i>Dartmouth</i> isochrones evolution with age and metallicity in a $B - V$ vs $V - I$ CC diagram . . . . .	70
2.8	RC stars with metallicity data on the $B - V$ vs $V - I$ diagram . . . . .	71
2.9	Wavelength dependence of extinction in the IR/optical region . . . . .	74
2.10	Schema MLE method: calibration using atmosphere models . . . . .	75
2.11	Isochrone shift through the MLE <i>gap</i> correction method on four CC diagrams . . . . .	76
2.12	MLE results comparison with 9 observed Pickles giants' spectra . . . . .	78
2.13	MLE results comparison with 9 theoretical Atlas giants' spectra . . . . .	78
2.14	MLE results comparison with a Calspec giant spectra . . . . .	79
2.15	MLE results comparison with 379 NGSL giants' spectra . . . . .	79
2.16	$B - V$ vs $V - I$ CC diagrams of red giants and MS stars in NGC 2682 . . . . .	80
3.1	Giants selection by $\log g$ and parallax on the <i>Hipparcos</i> HR diagram . . . . .	83
3.2	Colours and metallicity distributions of the giants CC sample . . . . .	85
3.3	Distribution of the CC sample on an Aitoff map projection in galactic coordinates . . . . .	85
3.4	DIB in the Gaia high-resolution spectra of HD 190603 . . . . .	86
3.5	R. Lallement 3D extinction chart . . . . .	87
3.6	Schlegel 2D extinction maps . . . . .	87
3.7	Colour-Colour diagrams pairs . . . . .	88
3.8	Schema of the MCMC method implemented to calibrate the RC colour-colour relations . . . . .	91
3.9	Empirical colour- $H_p - K$ calibrations . . . . .	93
3.10	$T_{\text{eff}}$ vs $H_p - K$ diagram with stars from the five selected samples . . . . .	97
3.11	Empirical effective temperature - $H_p - K$ calibrations for the selected samples . . . . .	99
3.12	Empirical effective temperature - $H_p - K$ calibrations for the selected samples (cont.) . . . . .	100
3.13	Comparison between the five effective temperature - $H_p - K$ calibrations obtained in this work . . . . .	100
3.14	Comparison of the $T_{\text{eff}}$ -colour relation for giants of solar metallicity from the current work with the literature, for the five selected samples . . . . .	102
3.15	Comparison of the $T_{\text{eff}}$ -colour relation for giants of solar metallicity from the current work with the literature, for the five selected samples (cont.) . . . . .	103
4.1	Interstellar extinction coefficients model for $k_G$ as a function of $T_{\text{eff}}$ . . . . .	107
4.2	Density distribution of the $A_0$ from this work, the DIBs and an extinction map . . . . .	112
4.3	Residuals of the interstellar extinction measurements between this work and Hill et al. [2011] and Sumi [2004] . . . . .	112
4.4	Effective temperature measurements comparison between this work and Hill et al. [2011] and the <i>Gaia-ESO Survey</i> . . . . .	113
4.5	Residuals between the estimated interstellar extinction and the one given by the 3D maps, with respect to the metallicity and the $H_p - K$ colour for the <i>APOGEE</i> red giant stars . . . . .	114
4.6	Effective temperature comparison between the values obtained in this work and those from the <i>APOGEE</i> raw data, for the <i>APOGEE</i> red giant stars . . . . .	115
4.7	<i>Hipparcos</i> RC colour-magnitude diagrams . . . . .	116

4.8	$M_K$ vs $V - K$ HR diagram of solar metallicity <i>Hipparcos</i> red giant stars . . . . .	117
4.9	Residuals between the estimated interstellar extinction and the one given by the 3D maps, with respect to the metallicity and the $H_p - K$ colour for the <i>Hipparcos</i> RC stars . . . . .	117

## II Gaia: confrontation with external archives

1.1	Gaia trajectory to its final operational orbit around the L2 Lagrange points of the Earth-Sun system . . . . .	122
1.2	L2 Lagrange point position and orbit of the Sun-Earth system . . . . .	123
1.3	Schematic view of Gaia's Lissajous orbit around L2 . . . . .	123
1.4	Science capabilities of the Gaia mission . . . . .	124
1.5	Scientific predicted performances for Gaia . . . . .	125
1.6	Gaia payload module diagram . . . . .	126
1.7	Focal plane diagram . . . . .	127
1.8	Gaia observation principle . . . . .	128
1.9	Gaia Nominal Scanning Law . . . . .	128
1.10	Gaia $G$ , $G_{BP}$ , $G_{RP}$ and $G_{RVs}$ normalised passbands . . . . .	129
1.11	Gaia's first Hertzsprung-Russell diagram . . . . .	133
1.12	Schematic view of the data flow during the processing phase . . . . .	134
2.1	Work flow of the crossmatch process by source position . . . . .	142
2.2	Work flow of WP944 general validation process . . . . .	145
3.1	Schema of the components of proper motion . . . . .	148
3.2	Effect of parallax on stellar images in the preceding and following field of view of Gaia, when the Sun is in two different positions relative to the stars . . . . .	152
3.3	The RR Lyrae sample on an Aitoff map projection in galactic coordinates . . . . .	154
3.4	RR Lyrae distance modulus distribution . . . . .	155
3.5	Milky Way Map of William Herschel (1785) . . . . .	157
3.6	Orthographic view of the HEALPix tessellation of a sphere, at four different pixel resolutions . . . . .	158
3.7	HEALPix map for <i>Tycho-2</i> catalogue . . . . .	159
4.1	Normalized density distribution of the parallax differences between LMC/SMC and both TGAS 00.01 and GDR1-BF . . . . .	167
4.2	Parallax bias for <i>Hipparcos</i> stars in GDR1 as a function of the Gaia ecliptic coordinates . . . . .	168
4.3	Sky distribution of the mean normalized residuals of the TGAS proper motions versus the <i>Tycho-2</i> , for the TGAS stars not in <i>Hipparcos</i> , in equatorial coordinates . . . . .	169
4.4	Normalized density distribution of <i>Simbad</i> HPM proper motions for TGAS 00.01 and GDR1-AF . . . . .	170
4.5	Missing <i>Tycho-2</i> sources in TGAS . . . . .	171
4.6	Missing <i>Tycho-2</i> sources in GDR1 . . . . .	171
4.7	Duplicated <i>Tycho-2</i> sources in GDR1, before filtering . . . . .	172
4.8	Parallax distribution of LMC/SMC stars confirmed to be in <i>Hipparcos</i> , for TGAS 00.01 and GDR1 . . . . .	173
4.9	Parallax outliers of <i>Hipparcos</i> stars in TGAS 00.01 and GDR1 . . . . .	173
4.10	Colour-colour diagram $G - H_p$ vs $B - V$ of <i>Hipparcos</i> stars in GDR1 . . . . .	174
4.11	Photometric outliers in the $G - H_p$ vs $B - V$ diagram of <i>Hipparcos</i> stars in GDR1 . . . . .	174



# List of Tables

## I Calibrating the Gaia Red Clump

1.1	Absolute magnitude calibrations of the <i>Hipparcos</i> RC in the literature . . . . .	37
1.2	CMD giant features summary . . . . .	57
2.1	Shifts obtained for each band through the MLE <i>gap</i> correction method . . . . .	76
2.2	Spectral libraries information and data of available giant spectra . . . . .	77
3.1	Extinction source for the 144 stars of the sample (with $A_V < 0.03$ ) . . . . .	87
3.2	Coefficients and range of applicability of colour- $H_p - K$ relations . . . . .	92
3.3	Best model, outliers, and computational time of colour- $H_p - K$ relations . . . . .	92
3.4	Colours of Arcturus derived using our colour- $H_p - K$ relations, compared to the literature . . . . .	94
3.5	Extinction sources for each $T_{\text{eff}}-H_p - K$ sample . . . . .	96
3.6	Coefficients and range of applicability of the $[\theta_{\text{eff}}, H_p - K]$ relation . . . . .	98
3.7	Best model, outliers, and computational time of the $[\theta_{\text{eff}}, H_p - K]$ relation . . . . .	98
4.1	Coefficients and range of applicability of the intrinsic colour- $V - K$ and $T_{\text{eff}}-V - K$ calibrations for the relations used in the $A_0$ -method . . . . .	111
4.2	Coefficients of the EW (units of $\lambda$ ) and $E_{(B-V)}$ linear relation: $E_{B-V\lambda} = a + b EW_\lambda$ . . . . .	111

## II Gaia: confrontation with external archives

1.1	Gaia spacecraft main specifications. Credits from ESA . . . . .	125
1.2	Predicted end-of-mission parallax standard errors $\sigma_{\overline{\theta}}$ averaged over the sky for a uniform distribution, for unreddened B1V, G2V, and M6V stars . . . . .	131
1.3	Predicted end-of-mission photometric errors in units of milli-magnitude, for unreddened B1V, G2V, and M6V stars . . . . .	131
1.4	Predicted end-of-mission radial-velocity robust formal errors $\sigma_{\text{vrad}}$ , averaged over the sky for a uniform distribution, for unreddened B1V, G2V, and K1III-MP (MP = metal-poor) stars . . . . .	131
1.5	Responsibilities of each Gaia DPAC Coordination Unit . . . . .	134
2.1	Responsibilities of each Work Package in CU9 . . . . .	138
4.1	CU9 validation releases . . . . .	166
4.2	Summary of the comparison between the GDR1-AF (TGAS 04.00) parallaxes (zero-point and extra-dispersion) and the external catalogues . . . . .	169



# Abbreviations

$T_{\text{effC}}$	Effective Temperature - Colour	89
AGB	Asymptotic Giant Branch	29
AGBB	Asymptotic Giant Branch Bump	30
AMR	Age-Metallicity Relation	31
ASPOS	Astronomic SPectra Observation Simulation	68
CC	Colour-Colour	53
CHeB	Carbon-Helium Burning	30
CMD	Colour-Magnitude Diagram	27
CU	Coordination Unit	128
DB	Data Base	135
DIB	Diffuse Interstellar Band	78
DIC	Deviance Information Criterion	83
DR1	Data Release 1	73
EC	External Catalogue	135
EW	Equivalent Width	104
GC	Globular Cluster	39
GDR1	Gaia Data Release 1	160
GDR1-AF	Gaia Data Release 1 - After Filtering	160
GDR1-BF	Gaia Data Release 1 - Before Filtering	160
HB	Horizontal Branch	28
HPM	High Proper Motion	163
HR	Hertzsprung-Russell (diagram)	26
ISM	Interstellar Medium	78
MCMC	Monte Carlo Markov Chain	82

---

MLE	Maximum Likelihood Estimator	65
MS	Main Sequence	27
OC	Open Cluster	30
RC	Red Clump	30
RGB	Red Giant Branch	27
RGBB	Red Giant Branch Bump	30
SFR	Star Formation Rate	31
SRC	Secondary Red Clump	30
TGAS	Tycho-Gaia Astrometric Solution	104
VRC	Vertical Red Clump	30
VTR	Validation Test Report	142
VTs	Validation Test Specification	134
WP	Work Package	129
XM	Cross-Match	135







# Introduction

*“Who are we? We live on an insignificant planet of a humdrum star lost in a galaxy tucked away in some forgotten corner of a universe in which there are far more galaxies than people”*

– Carl Sagan

Democritus (460 - 370 BC) suggested for the first time that the Milky Way consisted of a vast number of distant and unresolved stars. Centuries later, in 1610, Galileo Galilei turned his telescope toward the sky and then it became clear that Democritus was right: the filmy glow was nothing but the light of billions and billions of stars too faint to be seen individually. He realized that if the stars were like the Sun, they had to be at vast distances because of the range of brightnesses he could observe. For several hundred years after Galileo, astronomers have tried to map out the stars in the Milky Way. William Herschel, in 1785, was the first to start counting stars in order to make a map and measure the distribution of stars in space. And, somehow, we are still counting today.

According to Edwin Hubble’s classification system, the Milky Way is a barred spiral galaxy. Moreover, it has an age of about 13.6 Gyr, it contains more than 200 billion stars, it is about 30.7 kpc in diameter, and we know the Sun is located within it at about 8 kpc from the center in the Galactic plane. Indeed, we are inside. A fact that clearly complicates the study of its structure and content. However, we know that if we could observe the Milky Way from the outside, we could distinguish four main components: the bulge, the thin and thick disk and the stellar halo.

The bulge, a boxy, slightly peanut-like, shaped structure of a few kpc, is formed by a dense concentration of mostly old (about 12 Gyr) and metal-poor stars distributed around a super-massive central black hole and located at the central part of the Galaxy. Apparently new stars may still be born there, near the galactic nucleus, due to gas and dust coming from outlying regions by gravitation attraction towards the galactic core, which continuously enrich the medium. Then there is the disk, composed of two kinematically and chemically distinct populations, the thin and the thick disks. The thin disk extends to at least 15 kpc from the galactic center, and has a scale height of about 300 pc. Inside it there are most of the youngest stars of the Galaxy (up to 10 Gyr old), forming a roughly circular structure which rotates about the center at about 220 km/s. Because it contains on the order of 80% of the baryonic material of the Galaxy and lots of dust and gas, it is an active site for ongoing star formation, especially in the spiral arms where the highest density regions of interstellar gas, dust and stars are located. The thick disk, instead, is broader, with a scale height of about 1 kpc, it is mainly made of old (more than 10 Gyr) and metal-poor stars, it contains about 10% of the stellar mass of the Milky Way, and it rotates much slower, by about 40 km/s. And finally, the halo, a spherical quasi-non-rotating structure composed of very old (more than 12 Gyr) and very metal-poor stars, globular clusters, gravitationally captured satellite galaxies and stellar remnants (like white dwarfs), which surrounds the entire Galaxy.

But if we are inside, how could we determine the shape, the size and the structure of the Milky Way? Mainly by measuring stellar distances. Accurate measurement of distances is actually one of the most basic goals of astronomers, however it is also one of the most difficult. The most direct method for estimating astronomical distances is the trigonometric parallax. However, parallaxes may be precise up to a certain distance. In order to go further with high precision we need to use *standard candles*. That is, sources abundant enough within reach of parallax measurements and a sufficiently bright well-known luminosity

to be seen for instance in dense regions or out to local group galaxies. *Red Clump* stars fit precisely this description. Their importance as distance indicators is one of the motivations of this thesis.

Although we already know *grosso modo* how our Galaxy is, it remains still enigmatic in many respects. How was it formed? How many spiral arms it has? What is its future? Gaia, one of the most ambitious mission of the European Space Agency will tackle many of these fundamental questions through nothing less than astrometric, photometric and spectrometric measurements. Combined, all this information will build up an unprecedented picture of the evolution of the Milky Way, and will be without any doubt an outstanding opportunity to unravel many enigmas of Galactic archaeology, Galactic structure and dynamics, history of star formation, chemical evolution, interstellar medium,... The astrometric measurements of the Gaia satellite will provide us precise positions and parallaxes for a large number of *standard candles*, allowing to improve significantly their calibration. This is particularly important for *Red Clump* stars which will be numerous even in the first releases of Gaia. This will represent a breakthrough in our knowledge on these populations of stars and a clear improvement of the accuracy in distance determinations.

This thesis is divided in two main parts. The first part is devoted to the absolute calibration of *Red Clump* stars. We will first introduce what we know about the evolution of a giant star followed by the different observational features we may find in a Colour-Magnitude diagram and their role as distance indicators. We will then describe the two methods developed, one using theoretical models and another empirical, to calibrate the intrinsic colours and the spectroscopic effective temperatures of *Red Clump* stars from the *Johnson*, *2MASS*, *Hipparcos* and *Tycho* photometric systems. The second method will be afterwards used to estimate the interstellar extinction, the effective temperature and the Gaia G intrinsic and absolute magnitudes of those stars present in the first Gaia release, allowing a new calibration of the *Red Clump* as distance indicator.

In the end, the *standard candle* calibrations depend on trigonometric parallaxes. Before Gaia parallaxes are published and allow us to calibrate our *Red Clump* stars, we need to ensure the quality of these parallaxes. The second part of this thesis is thus dedicated to the statistical validation of the Gaia data before its publication. We will detail our contributions within the Gaia Data Processing and Analysis Consortium (DPAC) and the statistical tests developed to check the astrometric solution of Gaia with other already well-known catalogues. Finally we will show the improvements achieved throughout the whole validation period, the faced problems, how have they been solved and the final performances and contents of the Gaia Data Release 1, which will be released the 14th September 2016.

## **Part I**

# **Calibrating the Gaia Red Clump**



# Giants: state of knowledge

## Contents

---

<b>1.1</b>	<b>Becoming a Giant star</b>	<b>32</b>
1.1.1	Formation	32
1.1.2	Evolution to a giant	32
1.1.3	Giant phases	33
<b>1.2</b>	<b>Giant features on a Colour-Magnitude Diagram</b>	<b>36</b>
1.2.1	Red Clump	36
1.2.2	Vertical Red Clump	42
1.2.3	Secondary Red Clump	43
1.2.4	Red and Asymptotic Giant Branch Bumps	45
1.2.5	Double RC-RGBB pairs	50
1.2.6	An unexplained feature?	55
1.2.7	Synopsis	55
<b>1.3</b>	<b>Red Clump characterization</b>	<b>58</b>
1.3.1	Modelling the RC	58
1.3.2	The need of a calibration: this thesis	60

---

## 1.1 Becoming a Giant star

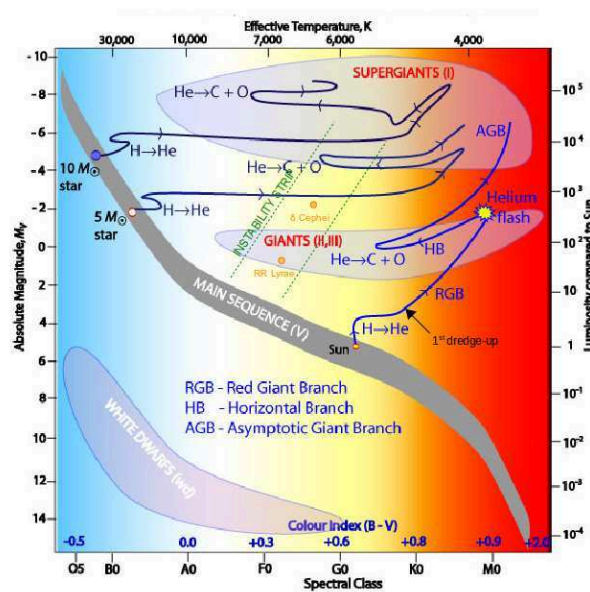
Stars were the first objects created in the Universe. Without stars, there would be no other elements aside from hydrogen and helium, there would be no galaxies, there would be no planets, there would be no life. They are the building blocks of the large scale structures of the Universe, and the Universe itself, thus the key to understand its origin and evolution, its history.

### 1.1.1 Formation

Stars are formed in vast clouds of cold molecular hydrogen ( $H_2$ ) and helium ( $He$ ), and other heavier elements. The clouds are in pressure equilibrium (hydrostatic equilibrium) with the surrounding interstellar medium. Eventually this stability may be disrupted by, for instance, a nearby supernova explosion, a collision with another cloud, the gravitational influence of other giant molecular clouds, or the pressure wave of a galaxy spiral arm passing through the region [Shu et al., 1987, Meyer et al., 2000, McKee and Ostriker, 2007]. This provokes the cloud to collapse and fragment into smaller clumps until the hydrostatic equilibrium is re-established in their cores, resulting in the birth of stars. Thousands of stars may be formed from a single molecular cloud.

### 1.1.2 Evolution to a giant

Ejnar Hertzsprung and Henry Noris Russell created in 1910 the so-called Hertzsprung-Russell (HR) diagram. The diagram represents the evolution of stars in terms of their luminosity and spectral type (or effective temperature), allowing to group stars in different populations depending on physical characteristics and properties (e.g. main sequence, giant branches, white dwarfs; see fig. 1.1). That is, stars go through different evolutionary stages as their energy source changes. This results in changes in the surface temperature (i.e. colour), the luminosity (the energy output per second), and their sizes. The HR diagram shows the trace of these changes for each star, from its birth to its death.



**Figure 1.1** – Evolutionary Tracks in an HR diagram. Adapted from a figure of the Australia’s National Science Organisation (CSIRO)

Once a star is formed, the type of life it will have, how long it will last on each stage, whether it dies as planetary nebula or a spectacular supernova, depends on its initial amount of mass. Higher the mass, higher



the energy, higher the capability of producing chemical reactions that will bring the star to different phases. That is, stars born with:

- $M \sim < 0.08M_{\odot}$  do not have enough mass for stellar ignition and directly become brown dwarfs
- $M \sim > 0.08M_{\odot}$  (but not higher than  $100M_{\odot}$  which would provoke collapse instabilities and the mass would be ejected) enter to the Main Sequence (MS) phase. Those with  $M > 0.5M_{\odot}$  will continue evolving to giant stages

When the star reaches the MS phase the abundance of  $He$  in the core starts to increase, due to the nuclear fusion of  $H$  into  $He$ . This is the longest and most stable phase in the lifetime of a star. Higher the mass faster the star evolves, so lower the time it remains in the MS. Changes do occur, but over long time scales, so that MS stars are indeed in thermal and hydrostatic equilibrium. A star remains on the MS as long as there is  $H$  in its core that it can fuse into  $He$ . As stated above, only those born with an initial mass of  $M > 0.5M_{\odot}$  will be able to leave the MS at a given moment, becoming giant stars.

### 1.1.3 Giant phases

Since the evolution of a star strongly depends on its mass, a clear way to introduce giant phases is to separate stars into three different groups according to their mass, their evolution through these post-MS phases and their final fate [Chiosi et al., 1992]:

- **Low-mass stars:** stars with  $M > 0.5M_{\odot}$  that develop an electron-degenerate  $He$  core immediately after leaving the MS. When their core mass ( $M_c$ ) attains a critical value of  $0.45M_{\odot}$ , the core  $He$ -burning starts through a *flash*, which continues until the degeneracy disappears (see details below in section 1.1.3). The maximum initial mass of the star (called  $M_{HeF}^1$ ) for this to occur is about 1.8 - 2.2  $M_{\odot}$  (depending on the initial chemical composition)
- **Intermediate-mass stars:** stars with  $M > M_{HeF}$  that develop a highly degenerate carbon-oxygen core after  $He$  exhaustion. They typically end up into a  $C-O$  core white dwarf (see section 1.1.3)
- **High-mass stars:** stars also with  $M > M_{HeF}$ , but massive enough to undergo degeneracy and allow a gradual carbon ignition. An iron core is created which will finally collapse and explode in a supernova (see section 1.1.3)

Using these definitions, below we describe the three main giant phases: *Red Giant Branch*, *Horizontal Branch*, *Asymptotic Giant Branch*, and processes involved in each one. These details are necessary to better understand the different observables that can be found in a Colour-Magnitude Diagram (CMD, next section 1.2). Figure 1.2 at the end of this section schematically shows the star's core evolution from the MS to these giant stages.

#### Red Giant Branch

When the  $H$  fuel is exhausted in the core of the MS star, nuclear reactions can no longer take place and fusion stops. To maintain hydrostatic equilibrium, the inner layers contract forming a  $He$  core to compensate the inward gravitational attraction (first dredge-up). The star enters to the *subgiant* stage.

The gravitational potential energy is therefore converted into thermal energy. Although there is no more fusion in the core, the increasing temperature heats up the shell of  $H$  surrounding it. Hydrogen fusion is then accelerated provoking the expansion of the outer layers of the star, increasing its radius.

By convection, the energy produced in the shell-burning region is spread over a larger surface area while the star expands, inducing an important decrease of the surface temperature (effective temperature). This effect produces a shift on the star luminosity towards the colour red, and that's why it is called a *red giant*. The star is now in the Red Giant Branch (RGB).

<sup>1</sup>The value of  $M_{HeF}$  depends on the extent of convective cores during the MS phase

## Helium Flash

Hydrogen fusion in the shell produces more  $He$ , which gets dumped onto the core causing an increase of its temperature, which in turn will induce  $He$  fusion. When the temperature of the core reaches around 100 million K, the  $He$  nuclei gets fused.

As explained, depending on the mass of the star, the  $He$  burning might be gradual or might begin with an explosive flash. In low-intermediate mass stars (typically  $0.5 < M < M_{HeF} \sim 2.2M_{\odot}$ ) the triple- $\alpha$  process<sup>2</sup> initiates in a matter of minutes or hours. The  $He$ -core is degenerate, hence the temperature can increase without the pressure adjusting along with it (as happens for ideal gases). Therefore, once the core reaches a mass of about  $M_c \sim 0.45M_{\odot}$ ,  $He$ -burning starts: the core rapidly gets hotter and the reaction quickly spreads throughout producing an enormous quantity of energy (just for a few seconds) which further accelerates the  $He$  burning rate. The overall effect is an explosive reaction: the *Helium Flash*. This flash occurs in the core deep inside the star, and all the released energy is absorbed by the layers of plasma above, so the effects are not seen from the exterior of the star. At that point the temperature has increased enough so that the thermal pressure becomes dominant again, eliminating the degeneracy. The core can then expand and cool down and a stable burning of  $He$  will continue.

Stars with masses slightly above  $M_{HeF}$  have a weakly degenerate core, hence  $He$ -ignition may start at a lower  $M_c$  (about  $0.33M_{\odot}$ ). High mass stars ( $M > M_{HeF}$ ) start to burn  $He$  via the triple- $\alpha$  process without its core becoming degenerate, and so does not exhibit this type of *Helium Flash*. And, worth to remember also that very low-mass stars ( $M < 0.5M_{\odot}$ ) do not reach a temperature high enough to ignite  $He$ , hence the degenerate  $He$  core keeps on contracting and the star ends up in a  $He$  white dwarf.

## Horizontal Branch

After the  $He$ -burning process begins (either explosively with a flash, or gradually for heavier stars) the star now changes to a new equilibrium state, and its evolutionary path switches from the RGB onto the Horizontal Branch (HB) of the HR diagram. This term derives from the fact that the luminosity of the star will stay relatively stable (and the luminosity distribution quite narrow) while the effective temperature increases, actually migrating horizontally across the HR diagram.

The effective temperature depends on the metallicity and the  $H$ -rich envelope mass. And the mass of the envelope depends on the initial mass of the star, but also on the mass loss during the RGB and the  $He$ -flash. Thus, lower metallicities and high mass envelopes cause bluer HB (BHB), while high metallicities and lower mass envelopes cause a redder HB (RHB), which corresponds to the *Red Clump* region (see later in section 1.2.1).

At this stage stars are powered by two sources of energy:  $He$  fusion in the core and  $H$  fusion in a shell around the core. Helium burns into carbon, and carbon combines with helium to make oxygen, making the nuclei rich in carbon and oxygen.

Worth to notice that, as they evolve, HB stars cross the *instability strip*, a narrow and almost vertical region in the HR diagram (see in fig. 1.1). For a short time, high mass stars will be Cepheid variables and low mass stars will be RR Lyrae stars (among other less frequent types of variable stars, like W Virginis or ZZ Ceti). Most stars more massive than the Sun enter the *instability strip* and become variable at least once after they have left the MS. It is within this region that they suffer instabilities that cause them to pulsate in size and vary in luminosity.

## Asymptotic Giant Branch

After a few million years,  $He$  in the core runs out and fusion stops: the star collapses once again causing the  $He$  in an outer shell to begin fusing.  $H$  may begin fusion in a shell just outside the burning  $He$  shell as

<sup>2</sup>Chemical process where three  $He$  nuclei (also known as  $\alpha$  particles) take part: two  $He$  nuclei are fused to form beryllium which immediately fuses in an almost instantaneous process with the third  $He$  nuclei to form carbon and  $\gamma$ -ray photons

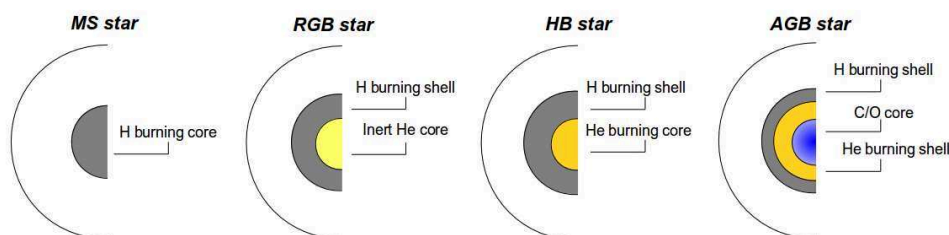
well. Thus, the structure of the stars is composed of a  $C-O$  core, a  $He$ -burning shell, and an  $H$ -rich envelope at the base of which an  $H$ -burning shell is active.

Due to the presence of two active burning shells, the core contracts, the  $He$ -rich layers above expand, and the outer envelope starts contracting. As the  $He$ -rich zone expands, the temperature in the  $H$ -shell decreases and the  $H$ -burning shell disappears. Now the entire envelope ( $He$ -rich layer plus  $H$ -outer envelope) starts expanding in response to core contraction. The star cools, expands, and grows in brightness, burning its nuclear fuel faster and faster, becoming thus an Asymptotic Giant Branch (AGB) star. During this phase it may also experience  $He$  shell flashes or thermal pulses. The  $He$ -burning shell gradually adds mass to the growing  $C-O$  core, which may become degenerate due to its increasing density.

Intermediate-mass stars develop degenerate  $C-O$  cores. In this case, for a star to be able to ignite carbon in highly degenerate conditions it needs to have a  $C-O$  core mass of  $1.4 M_{\odot}$  (the Chandrasekhar limit), which corresponds to an initial star mass  $M_w$  (considering the effect of mass loss by stellar wind during the RGB and AGB phases) of about  $4-6M_{\odot}$  [Iben and Renzini, 1983]. Knowing this, we may distinguish two different scenarios for AGB intermediate-mass stars:

- stars with  $M_{HeF} < M < M_w$ : they will never start fusion in their degenerate  $C-O$  core, since the maximum temperature is smaller than required to ignite carbon. Instead, at the end of the AGB phase, the outer layers of the star are blown off by the stellar wind, i.e. the envelope is ejected. What is left over is a cloud of gas and dust slowly expanding into space, with a very hot white dwarf in the center: a planetary nebula. The nebula disperses quickly, leaving an inert  $C-O$  white dwarf which slowly cools down
- stars with  $M_w < M < M_{up} \sim 7 - 9M_{\odot}$  (with  $M_{up}$  the minimum initial mass for which the core reach a sufficiently high temperature to undergo non-degenerate carbon ignition): if the core attains the Chandrasekhar limit, the stars may start  $C$ -ignition [Iben, 1985, 1991, Chiosi et al., 1992, Pagel, 2006], although it rarely happens and they usually end up into O-Mg or Mg-Ne white dwarfs

High-mass stars (larger than  $M_{up} \sim 9M_{\odot}$  and with a  $C-O$  core minimum mass of  $1.06M_{\odot}$ ) do not suffer core electron degeneracy, hence  $C$ -ignition can start in a gradual way. Carbon and oxygen fuse to form neon, then magnesium, then silicon. All forming into burning shells surrounding an iron ash core. When the core reaches a critical mass, all the material fall inward, collapsing into a supernova explosion. The scheme undergone will be different depending of a mass limit  $M_{mas} \sim 12M_{\odot}$ , being more complex for stars with  $M_{up} < M < M_{mas}$  and less complex for  $M > M_{mas}$ .



**Figure 1.2** – Star nuclei evolution. From left to right: Main Sequence star, Red Giant Branch star, Horizontal Branch star, Asymptotic Giant Branch

## 1.2 Giant features on a Colour-Magnitude Diagram

So far we have described the different phases of giant stars in terms of stellar evolution. From an observational point of view, however, we may also talk about other specific features. Starting from the subgiant phase, we can (sometimes) encounter the following sub-populations:

- On the RGB: *Red Giant Branch Bump* (RGBB)
- On the HB: *Red Clump* (RC) , *Vertical Red Clump* (VRC) , *Secondary Red Clump* (SRC)
- On the AGB: *Asymptotic Giant Branch Bump* (AGBB)

The identification, inner physics and differences between these observables have been largely discussed in the literature, and some questions remain still unanswered. The goal of this section is to review and clarify our actual knowledge about them, and give a general idea of all the processes involving giant stars. They will be introduced in order of importance (observationally speaking): first the RC and its counterparts SRC and VRC, and second the two *bumps* (RBBB and AGBB). Further, particular observable effects of some of these features might also be detected, such as the *Double Clump-Bump* (*double RC-RBBB*) pairs.

This overview allows to put in context the final aim of this first part of the thesis, i.e. the complete calibration of the most important feature, the *Red Clump*, and its application in determining stellar properties.

### 1.2.1 Red Clump

Because *He*-burning in a low-intermediate mass star degenerate core cannot start until the stellar core mass attains a critical value of about  $M_c \sim 0.45M_\odot$  (corresponding to an initial star mass  $M_{HeF}$  of 1.8-2.2 $M_\odot$ , see section 1.1.3), these stars will have similar core masses at the beginning of *He* burning, hence similar luminosities. This produces a *clump* of red HB stars: the *Red Clump* (RC)

That is, the RC is formed by core *He*-burning stars (CHeB) stars with masses between  $1M_\odot \leq M \leq M_{HeF}$  (or even lower for metal-rich systems) that evolved from RGB stars with degenerate cores (or around the transition to non-degeneracy), and are cooler than the *instability strip*. It appears as an overdensity in the CMD of populations with ages older than  $\sim 0.5 - 1\text{Gyr}$ , covering the range of spectral types *G8III - K2III* with  $4500\text{K} \leq T_{\text{eff}} \leq 5300\text{K}$  [Zhao et al., 2001, Soubiran et al., 2003]. Indeed, it represents the old and metal-rich counterpart of the HB (the RHB in stellar evolution, section 1.1.3). Fig. 1.11 (left) shows a clear example where the HB of the metal-rich GC 47 *Tuc* is a RC.

#### RC absolute magnitude

The luminosity (absolute magnitude) at which the *clump* is produced have been largely studied in the literature, and several calibrations have been carried out so far. Typically the range of magnitude and colour is:

$$0.6 < M_I < 1$$

$$0.8 < V - I < 1.3$$

The aim of all these works is also to determine the dependence or not of this parameter with the metallicity and colour. Table 1.1 summarizes some of these calibrations in the solar neighbourhood using *Hipparcos*<sup>3</sup> data (where 15% of the stars observed with  $\sigma_\pi/\pi < 5\%$  are RC stars,  $\sim 600$  stars).

All these calibrations imply that the absolute magnitude of the RC of the solar neighbourhood has a very weak or insignificant dependence on metallicity. However, this is not the case for galactic open clusters (OCs) for which a dependence has been found for  $M_V$  and  $M_I$  [Sarajedini, 1999, Twarog et al., 1999]. And

<sup>3</sup>All references in the table used data from the original *Hipparcos* catalogue except Groenewegen [2008] who made use of the new reduction of van Leeuwen [2007]

**Table 1.1** – Absolute magnitude calibrations of the *Hipparcos* RC in the literature. All authors used the original *Hipparcos* data except [Groenewegen \[2008\]](#) who used the new reduction of [van Leeuwen \[2007\]](#). Columns [Fe/H] and Colour indicate the degree of dependence of these parameters in their calibrations: none (n) or weak (w)

Author	Calibration	[Fe/H]	Colour	Sample
<a href="#">Paczynski and Stanek [1998]</a>	$M_I = -0.279 \pm 0.088$	n	n	228 RC stars with $\sigma_\pi/\pi < 10\%$ $0.8 \leq V - I \leq 1.25$ $-1.4 \leq M_I \leq 1.2$
<a href="#">Stanek and Garnavich [1998]</a>	$M_I = -0.23 \pm 0.03$	n	n	same as <a href="#">Paczynski and Stanek [1998]</a>
<a href="#">Udalski [2000]</a>	$M_I = (0.13 \pm 0.07)([\text{Fe}/\text{H}] + 0.25) - (0.26 \pm 0.22)$	w	n	same as <a href="#">Paczynski and Stanek [1998]</a>
<a href="#">Alves [2000]</a>	$M_K = (0.57 \pm 0.36)[\text{Fe}/\text{H}] - (1.64 \pm 0.07)$	w	n	238 <i>Hipparcos</i> RC stars with K-band in 2MASS catalogue
<a href="#">Groenewegen [2008]</a>	$M_I = (0.08 \pm 0.07)([\text{Fe}/\text{H}] + 0.15) - (0.22 \pm 0.03)$ $M_K = (-0.15 \pm 0.07)((V - K)_0 - 2.32) - (1.54 \pm 0.04)$	w n	n w	RC from the revised <i>Hipparcos</i> parallaxes of <a href="#">van Leeuwen [2007]</a>
<a href="#">Valentini and Munari [2010]</a>	$M_{V_T} = +0.54$	n	n	277 RC <i>Hipparcos</i> stars

actually, this result agrees with the theoretical models of stellar populations of [Girardi and Salaris \[2001\]](#), [Salaris and Girardi \[2002\]](#). Thus, according to these models and OC observations:

- $M_V$  and  $M_I$  become fainter (increase) with increasing metallicity at a given age
- $M_K$  behaves oppositely for ages above  $\sim 1.5\text{Gyr}$ , it becomes brighter (decreases) when increasing the metallicity. And, in general,  $M_K$  is more sensitive to age and metallicity than  $M_I$  for high ages and low metallicities
- For a given metallicity the range of RC luminosities is limited in all bands (0.2-0.3 mag). When the age decreases from 3 to 1 Gyr, the luminosity decreases, owing to the transition between degenerate and non-degenerate *He* cores at *He* ignition, followed by a sharp increase at an age of 0.5 Gyr, which signals the complete removal of electron degeneracy in the *He* core

These conclusions suggest that we must be careful when applying the RC  $M_I$  of the solar neighbourhood to RCs of other stellar populations with different ages or abundances compositions (like the Magellanic Clouds), since this can induce wrong results.

Other absolute magnitudes calibrations have also been done in the *ugriz* system [[Chen et al., 2009](#)], where it has been found that metallicity coefficients decrease, while age coefficients increase, from the blue *u* filter to the red *z* filter. They conclude that  $M_I$  is the less sensitive to metallicity and age, and that  $M_g$  and  $M_z$  show significant and opposite dependence on age. The use of the calibration in one band or another for further applications will depend on the knowledge and distribution of metallicity and age.

The reasons of these RC absolute magnitude differences and dependences with metallicities and ages may be explained by stellar models [[Girardi, 2002](#)]. The luminosity of a RC star is due to the luminosity of the *He*-burning core and the luminosity of the *H*-burning shell ( $L = L_H + L_{He}$ ).  $L_H$  depends on mass and metallicity: it increases monotonically with mass because the luminosity of the *H*-burning shell increases with the envelope mass; moreover, the efficiency of the *H*-burning shell decreases with [Fe/H].  $L_{He}$ , instead, is nearly constant for similar core masses, and decreases when  $M_c$  decreases with metallicity. Thus,  $L$  should decrease with metallicity when the  $M_c$  and *H*-burning shell efficiency decrease.

The variation of the luminosity with ages mainly occurs for CHeB intermediate-mass, due to their proportionality between the core mass stars at *He*-ignition and their initial mass (typical of stars that do not develop electron degenerate *He* cores).

From this theoretical view one may expect that age and metallicity distributions of RC stars in different stellar populations, such as galaxies, must be quite different. Indeed, it depends on the star formation rate (SFR) and the age-metallicity relation (AMR) of the population considered.

### RC identification in the CMD

This feature was identified and correctly interpreted by Cannon [1970] and Faulkner and Cannon [1973] in intermediate-age OCs. Since it is a remarkable and well populated observable of the CMD, it has been rapidly detected in other structures, such as GCs or more complex stellar populations, like galaxies. Moreover, thanks to the precise parallaxes determined with the *Hipparcos* mission, the RC has been also clearly identified in the solar neighbourhood [Girardi, 1999a], allowing to better understand RC properties.

Although it is relatively easy to visually identify the *clump* on the CMD because of their overdensity at a given colour range and luminosity of the RHB, the identification of individual RC candidates is not trivial because they do highly overlap with the RGB stars. For this reason, different methods have been developed in order to extract RC stars from RGB background. Among the most known and used methods we find:

- Gaussian model method (as in Paczynski and Stanek [1998]): it is based on the modelisation of the magnitude distribution of RGB stars with a linear polynomial (first two terms of equation 1.1) and with a Gaussian for RC stars (last term in eq. 1.1). The stars located in the region dominated by the Gaussian are quite probably RC stars. The authors applied this method to the *Hipparcos* catalogue in the direction of the Baade's Window<sup>4</sup> using the *I* magnitude and found 664 candidates. Figure 1.3 shows the colour magnitude diagram for this part of the Baade's Window.

$$n(I_0) = a + b(I_0 - I_{0,m}) + \frac{N_{RC}}{\sigma_{RC}\sqrt{2\pi}} \exp \left[ -\frac{1}{2} \left( \frac{I_0 - I_{0,m}}{\sigma_{RC}} \right)^2 \right] \quad (1.1)$$

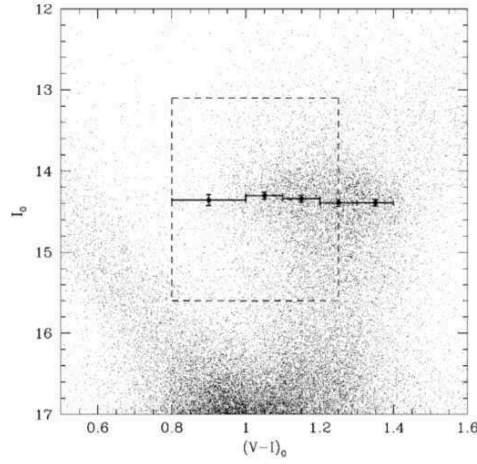
- Period spacing method (as in Bedding et al. [2011], Mosser et al. [2011], Stello et al. [2013]): it is based on solar-like oscillations. They extracted individual frequencies from asteroseismic data, and measured the period spacings of the dipole modes. The variation of periods with respect to the variation of frequencies for each star provides an identification of the different giant evolutionary phases (fig. 1.4).
- Photometric and spectrometric combination method (as in Bovy et al. [2014]): it is based on the selection of RC stars by their position in a colour - metallicity - surface gravity ( $\log g$ ) - effective temperature space (from *APOGEE* catalogue) using a method calibrated with stellar evolution models (*PARSEC* isochrones with a lognormal Chabrier [2001] distribution for the Initial Mass Function (IMF) and constant SFR), and high-quality asteroseismic data (mass, radius and  $\log g$  from *APOKASC* catalogue). Figure 1.5 compares the prediction from the stellar isochrones (density grayscale) for the distribution of stars in effective temperature and surface gravity with the data in the *APOKASC* catalogue, in two different metallicity bins. Coloured points correspond to *APOKASC* stars: red and yellow points correspond to Stello et al. [2013] classification of RGB and RC stars, respectively, and blue points to stars for which the evolutionary state is unknown.

### RC as a tool for distance determination

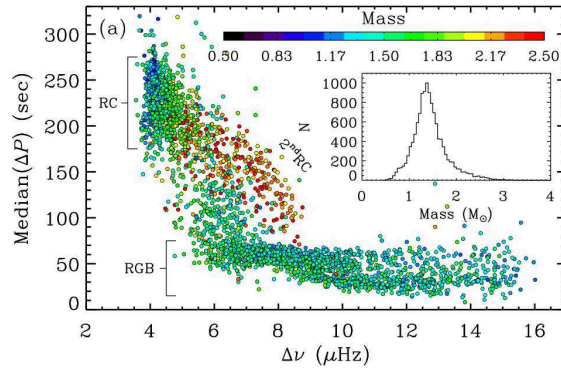
Thanks to the nearly-constant absolute magnitude and its weak dependence on metallicity, ages and colours, the RC is a powerful tool for astronomical distance determinations both within our galaxy and to nearby galaxies and clusters, as initially proposed by Hatzidimitriou and Hawkins [1989].

As stated by Paczynski and Stanek [1998], any method to obtain distances to large-scale structures suffers mainly from four problems: the accuracy of the absolute magnitude determination of the stars used, the interstellar extinction, the distribution of the inner properties of these stars (mass, age, chemical composition,...), and how large is the sample. Whether the use of the RC may be considered better than other *standard candles* such RR Lyrae or Cepheids, is precisely because they are very numerous. The

<sup>4</sup>Area of the sky on the Galactic Center region with relatively low amounts of interstellar dust along the line of sight from the Earth



**Figure 1.3** – CMD corrected for interstellar extinction for the stars in the part of Baade’s Window with a small extinction ( $A_V < 1.5$  mag). The dashed rectangle surrounds the RC region used for comparison with the local stars. The ridge of the highest density of the RC stars ( $I_{0,m}$ ) is shown for five color bins in the range  $0.8 < (V - I_0) < 1.4$  as calculated by eq. 1.1. Credits from Paczynski and Stanek [1998] (figure 1 of the article)



**Figure 1.4** – Median period variation ( $\Delta P$ ) for each star with respect to the frequency variation ( $\Delta \nu$ ). The different identified phases are indicated: RGB, RC and secondary RC (see later in section 1.2.3). Credits from Stello et al. [2013] (figure 4a of the article)

larger the sample used, the lower the statistical error in distance calculations. Moreover, with the *Hipparcos* mission, a high number of RC stars were observed in the solar neighbourhood with a parallax precision better than 10%. Thus the RC single-step method (equation 1.2) is one of the precise techniques most used so far for distance determination.

The distance modulus  $\mu$  is defined as:

$$\mu = (m - M) = \bar{X}_{Clump} - M_{X,Clump} - A_X \quad (1.2)$$

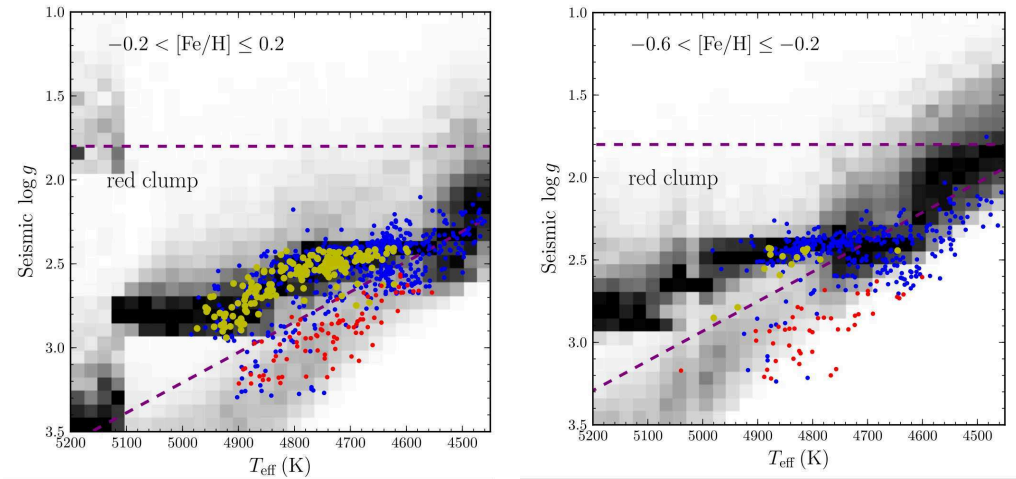
where  $\bar{X}_{Clump}$  is the mean magnitude of the RC of a given band  $X$ ,  $M_{X,Clump}$  is the absolute magnitude, and  $A_X$  is the reddening correction.

Its relation with the distance  $d$  (in parsecs) is given by:

$$\mu = 5 \log_{10}(d) - 5 \quad (1.3)$$

The RC as distance indicator has gained importance since the work of Paczynski and Stanek [1998], who determine the distance to the Galactic Center using these stars (in the I band) from the *Hipparcos* catalogue. The number of RC available is so large that the formal statistical error they obtained was only of about 1%.

These stars have also been used to highlight the structure of the Galactic Bar by using the Optical



**Figure 1.5** – Prediction for the conditional distribution of stars in surface gravity given their effective temperature from the PARSEC stellar isochrones (density grayscale) and data from the APOKASC catalog (colored points), in two metallicity bins. The prediction assumes a lognormal Chabrier [2001] model for the IMF, a metallicity distribution that matches that of the APOKASC data used in each bin, and a constant Star Formation History (SFH). Yellow and red points represent stars classified by Stello et al. [2013] as RC and RGB stars, respectively, from their measured period spacing gravity modes; blue points are stars for which the evolutionary state is unknown. The purple dashed lines represent our cuts in  $\log g$  and  $T_{\text{eff}}$  to separate RC stars from less luminous red giant branch stars (see Bovy et al. [2014] for more details). Credits from Bovy et al. [2014] (figure 1 of the article)

Gravitational Lensing Experiment (OGLE) data [Stanek et al., 1994, 1997]. Indeed the Galactic Bar was detected thanks to the observation of two groups of bulge RC stars, a faint and a bright one, belonging to the further and the nearest part of the Galactic Bar. This work was followed by, among others, Stanek et al. [1997], Babusiaux and Gilmore [2005], Nishiyama et al. [2005], Rattenbury et al. [2007b], Cabrera-Lavers et al. [2008], and extended to the study of the shape of the bulge [Nataf et al., 2010, McWilliam and Zoccali, 2010], the reddening in that region [Stanek et al., 1996, Stanek, 1996, Udalski, 2003, Sumi, 2004, Nishiyama et al., 2009], and its kinematics, dynamics and composition [Rattenbury et al., 2007a, Babusiaux et al., 2010].

Soubiran et al. [2003], Siebert et al. [2003] also characterised the thick disk using *clump* stars, while Bienaymé et al. [2003, 2005] determined the gravitational force perpendicular to the Galactic plane and the total surface mass density of the Galactic plane.

Further, RC stars may be used to estimate distances and infer shapes and orbits of other isolated components of the galaxy such as accreted stellar systems. This is the case of *Canis Major*, an accreted dwarf galaxy that appears as an elliptical shaped overdensity nearly coplanar with the Galactic disk [Newberg et al., 2002, Majewski et al., 2003, Bellazzini et al., 2006], or the *Sagittarius* stream, a stellar stream probably produced by tidal disruption of a dwarf galaxy within the Galactic potential [Ibata et al., 1998].

And regardless the inner galactic structure, these stars are also used to derive distances to other objects like: the Small Magellanic Cloud (SMC) [Udalski, 1998], the Large Magellanic Cloud (LMC) [Stanek and Garnavich, 1998, Salaris et al., 2003], *IC 1613* [Dolphin et al., 2001], *Carina* (M33) [Kim et al., 2002], and *Fornax* [Bersier, 2000, Rizzi et al., 2007].

However it is important to emphasize that, for these non-galactic objects, while some distances were coherent with previous values given in the literature (as for *Carina* and *Fornax*) this was not the case for the LMC, for which a 15% shorter distance was obtained. Although an uncertainty on extinction has to be considered, the main source of discrepancy was indeed, as predicted and discussed by Cole [1998], Girardi [1999a], Girardi and Salaris [2001], the dependence of the RC absolute magnitude to metallicity and age. As a matter of fact, the mean metallicity for the same RC colour interval ( $0.8 < V - I < 1.25$ ) is lower for the LMC compared to that of the solar neighbourhood, the bulge and *M31*.

Therefore, as already discussed above, absolute magnitude calibrations of table 1.1 cannot be used to

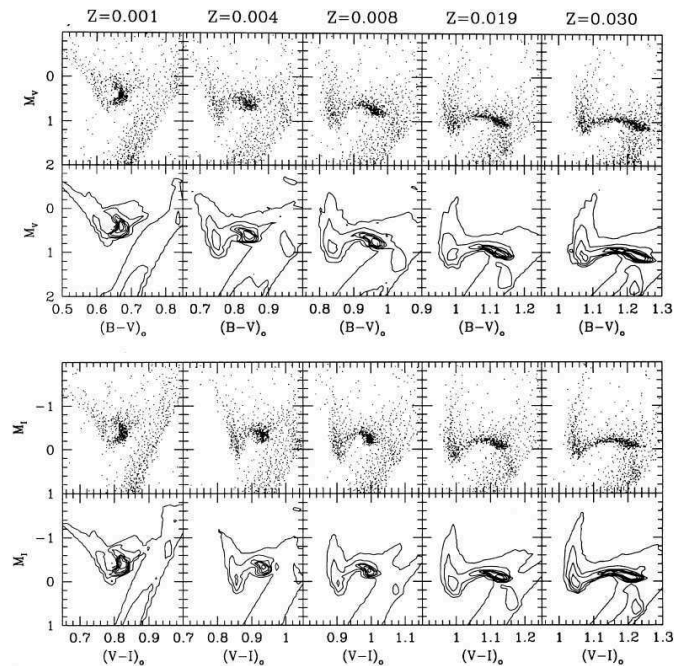


obtain distances to objects with clear differences on metallicity and stellar populations from the Milky Way. Like most stellar *standard candles*, the properties of the RC vary with the composition and age of the host galaxy. Hence, all RCs are not like the *Hipparcos* RC. Cole [1998] and Girardi and Salaris [2001] show that for those populations dominated by stars older than  $\sim 6$  Gyr, the *standard candle* approximation should be valid to a high degree for the mean absolute magnitude  $M_I$ , while for younger populations,  $M_I$  is probably brighter than the local  $M_I$ , and the correction may amount to as much as several tenths of a magnitude.

We can therefore conclude, that RC stars are a powerful tool for distance determinations if the history and metallicity of the object are known. For composite stellar populations, the age (mass) distribution depends in a complex way on the population's Star Formation History (SFH), the trend of *He*-burning timescale with stellar mass and metallicity, and the Initial Mass Function (IMF). We may say that, when the absolute magnitude difference between the external object and the local RC ( $\Delta M_{\lambda,RC}$ ) tends to zero, the local RC mean absolute magnitude can be used to determine distances [Salaris, 2013].

### RC fine structure

High-quality CMDs allow to identify with more detail the presence of sub-features of an apparently single observable. This means, for example, significant colour spreads or extensions to brighter and/or fainter luminosities. Such features are predicted by population synthesis models which consider stars in the complete relevant ranges of ages and metallicities, and are potentially useful for constraining the star formation histories of the parent galaxies over scales of gigayears. This detailed analysis is what is called the *fine structure*.



**Figure 1.6** – Synthetic CMDs showing the RC region for different metallicities, in the  $B - V$  (top panels) and  $V - I$  (bottom panels) planes. For each CMD series, the first row shows the synthetic CMD for 1000 stars, using models considering constant SFR from 0.1 to 10Gyr, constant mean metallicity, a small metallicity Gaussian dispersion of  $\sigma_z/Z = 0.1$ , and the effect of mass loss along the RGB. The second row presents six contour lines limiting regions with the same stellar density. As specified on the top of the figure, metallicities increase from left to right. The most densely populated region of the diagrams represents the main body of the RC. To its blue end, we have the structure that contains both the faint secondary clump and the plume of brighter clump stars. The RGB runs slightly to the red side of the main clump. It presents another concentration of stars (more evident in the right-hand panels), which corresponds to the RGBB. Credits from Girardi [1999b] (fig. 7 of the article)

The *fine structure* of the RC is not evident in the case of most ground-based observations of nearby galaxies, for which the clump spread in the CMD suffers the presence of large photometric errors. However,

since the *Hipparcos* mission and subsequent high precision surveys, we are nowadays able to unravel it.

Thus, the RC *fine structure* may be divided in three components: the *main red clump* (RC), the *vertical red clump* (VRC) and the *secondary red clump* (SRC). The following two sections introduce these last two RC components. To get a preliminary idea, figure 1.6 shows the synthetic CMDs for the  $M_B - M_V$  (top panels) and  $M_V - M_I$  (bottom panels) where the three observables and their evolution with increasing metallicity can be easily identified (see next sections for a better understanding).

## 1.2.2 Vertical Red Clump

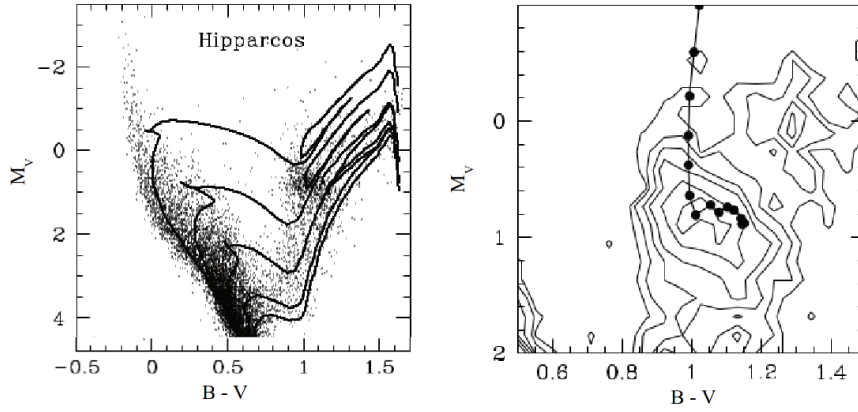
While studying the Large Magellanic Cloud (LMC), Zaritsky and Lin [1997] found another CMD feature close to the RC. This new population is also formed by CHeB stars that ignited *He* in non-degenerate conditions and has nearly the same color as the RC, but it departs from the bluest part of the RC and it is extended vertically towards higher luminosities (up to  $\approx 0.9$  mag brighter than the RC). The authors labelled this over-density as the *Vertical Red Clump* (VRC), and they concluded that it could be due to a foreground stellar population located between the LMC and us (at 33-35 kpc).

However, Beaulieu and Sackett [1998] and Ibata et al. [1998] argued that, if this was the reason of the VRC, that would have important consequences in the quantity of observed galactic microlensing events towards the LMC. These events generally correspond to compact objects distributed along the halo.

Instead, both Beaulieu and Sackett [1998] and Ibata et al. [1998] demonstrated that the origin of the VRC might be a fact of stellar evolution inside the LMC, caused by the youngest RC stars (between 300 Myr and 1 Gyr). Indeed, several reasons sustain this hypothesis: the authors show that the VRC is present in all LMC fields, that it has kinematic properties identical to that of the LMC itself, and that it is also in agreement with the physics of the *isochrones blue loops* of intermediate-age RC stars that have the same metallicity and age of the LMC. Its relative density is also consistent with the LMC star formation history and with stellar evolution in general. Moreover, the feature is also present in the solar neighbourhood as seen from the *Hipparcos* CMD. This clearly shows that it cannot be an intervening population.

As seen in section 1.1.2, stars in the HB burn *He* in their core while the *H* shell expands: the effective temperature increases and the core gains mass. When the core fusion stops and the *H* is exhausted, *He* begins to burn in the shell, and the star moves to the AGB. However, theoretical models predict that for an intermediate-mass star, this increase of temperature and mass when still in the HB, causes the star to spend more time near the bluest part of the HB ( $0.9 < B - V < 1.1$ ): it experiences a *blue loop* [Sweigart, 1987, Chiosi et al., 1992]. The colour remains roughly constant while the luminosity continues to increase. For RC stars, this translates to a vertical extension of the blue end of the RC in the HR diagram ( $-0.5 < M_V < 0.5$ ), the VRC. The agreement between the *isochrones blue loops* and the position of the VRC is observed in figure 1.7 with the *Hipparcos* sample, where this feature can be observed at  $B - V \sim 0.9$ , at the bluest part of the RC overdensity, where a theoretical *loop* is also predicted. Over-plotted isochrones correspond to a given metallicity  $Z = 0.02$  and different ages. Previous synthetic CMDs shown in figure 1.6 clearly manifest this *plume* of RC stars too, on the left side of each diagram.

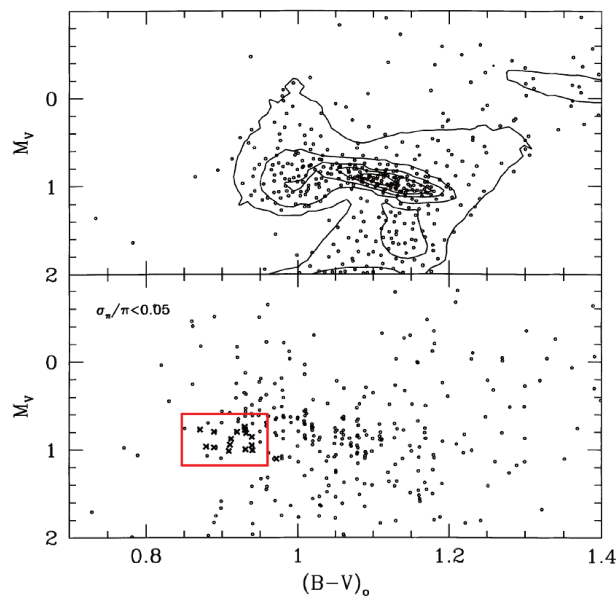
Once the *He* in the center is consumed the evolution towards the AGB continues as normal: the star moves to the red again in the HR diagram. The extension of the loop highly depends on the chemical abundances and their stratification throughout the star: the higher the *He* abundance and mass, the higher the luminosity, i.e. the loop extends further to the blue; the lower the *He* abundance and mass, the lower the luminosity, i.e. the loop is shortened [Sweigart, 1987]. A star can experience several *blue loops* during its post-main-sequence lifetime: for massive stars, this may frequently happen when new nuclear reactions start; for stars with  $2-3 M_{\odot}$  the effect is almost imperceptible and the star remains close to the RGB (although it depends on chemical abundances).



**Figure 1.7** – (Left) Hipparcos  $M_V$  vs  $B - V$  CMD containing over 16000 stars with accurate distances and photometry. The VRC can be slightly observed on the blueward side of the RC with nearly the same color as the peak clump density ( $B - V \sim 1$ ) and extending about a magnitude brighter. Isochrones with solar metallicity ( $Z = 0.02$ ) and  $\log(\text{age}) = 8.6, 9.0, 9.4, 9.8, 10$  are overplotted. Note that positions of the *blue loops* coincide with the vertical extension. (Right) Contour representation of the CMD density in the region of the RC for the solar neighbourhood as measured by *Hipparcos*, superposed on the mean locus of the core *He*-burning phase as a function of age for stars with metallicity and *He* abundance appropriate to the solar neighbourhood ( $Z = 0.02$  and  $Y = 0.28$ ). Separate scales for the abscissa of the top and bottom panels are indicated. The region of CMD shown has been divided into 20 bins along both the magnitude and color axes. For this binning, contour levels displayed are 35, 25, 15, 10, 5, and 2.5 stars. Adapted from [Beaulieu and Sackett \[1998\]](#) (fig. 8 of the article)

### 1.2.3 Secondary Red Clump

Still another sub-feature of the RC must be noticed in some CMDs. It corresponds to a small population also composed by CHeB stars, slightly more massive (hence younger) than those of the RC, and located at the bluest part of the main RC, like the VRC, but slightly below (about 0.4 magnitudes from the main *clump*). It can also be referred as the faint part of the VRC. This extension of the RC to lower luminosities was introduced by [Girardi et al. \[1998\]](#) based on the results of detailed populations synthesis models (fig. 1.8, and also in previous fig. 1.6, especially for high metallicities). The author called it the *Secondary Red Clump* (SRC).



**Figure 1.8** – Distribution of RC stars in the  $M_V$  versus  $B - V$  diagram, from (top) theoretical models and (bottom) *Hipparcos* data with parallax errors smaller than 5%. Stars in the red box belong to the secondary clump. In the top panel we may slightly see the VRC too. Adapted from [Girardi et al. \[1998\]](#) (fig. 4 of the article)

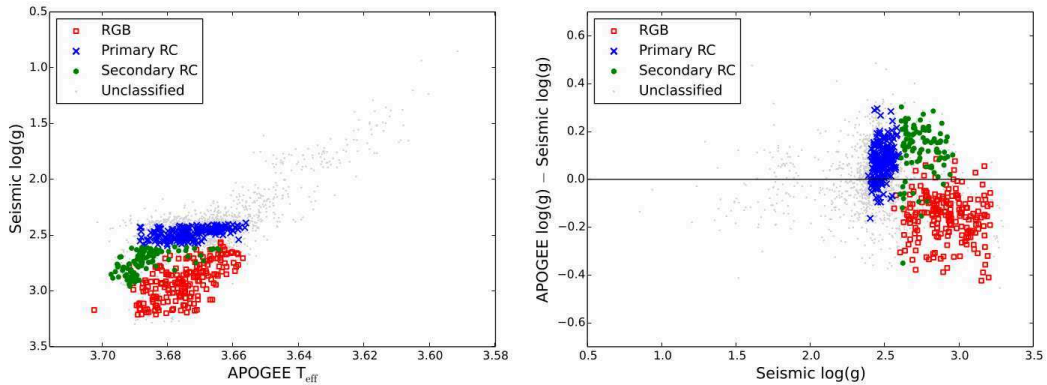
As explained in section 1.1.3, low-intermediate mass stars ( $0.5 < M < M_{HeF} \sim 2.2M_{\odot}$ ) need a core mass of  $0.45M_{\odot}$  to ignite *He*. However stars with masses slightly above  $M_{HeF}$  may start *He*-ignition with lower core masses (about  $0.33M_{\odot}$ ), because of their weak or non-degenerate cores. These last stars should correspond to this particular feature, and therefore belong to a narrow interval of masses ( $M_{HeF} \approx 2 M_{\odot}$ ) and ages ( $\approx 1$ Gyr). They are expected to be clearly separated from the main RC in the CMD of metal-rich populations ( $Z > 0.004$ , as already shown in fig. 1.6) and particularly observable in galaxies that increased their star formation rate in the past Gyr [Girardi, 1999a].

However, such a specific structure might be difficult to detect (i.e. indistinguishable from the main RC) due to:

- the low statistics: the observed field must contain an important amount of stars at 1Gyr and  $M \approx 2 M_{\odot}$
- the large dispersion in metallicity: there must be a significant number of metal-rich stars (more than  $Z = 0.004$ )
- large photometric errors, differential absorption (reddening) and distance dispersions: to be distinguishable, the combination of all these effects should not be larger than 0.2 mag [Girardi, 1999a]

Similar *secondary clumps* as the one predicted by the models and found in the *Hipparcos* solar neighbourhood [Girardi et al., 1998] have been identified in some fields of the Large Magellanic Cloud [Bica et al., 1998] or other galaxy fields, in several LMC star clusters like *NGC 419*, *NGC 1751*, *NGC 1783*, *NGC 1806*, *NGC 1846*, *NGC 1852* and *NGC 1917* [Girardi et al., 2000, Mackey et al., 2008, Milone et al., 2009, Girardi et al., 2009], as well as in some Milky Way OCs (like *NGC 752* and *NGC 7789*) [Girardi et al., 2000], in spectroscopic surveys as *LAMOST* [Wan et al., 2015], and in different asteroseismic surveys such as *Kepler* [Stello et al., 2013, Liu et al., 2015], *APOKASC* (*APOGEE* + *Kepler*) [Pinsonneault et al., 2014], among other works.

As seen in section 1.2.1, asteroseismic data is becoming a powerful tool for star characterisation and stellar population identification. Figure 1.4 (from Stello et al. [2013]) shows how a period spacing method based on solar-like oscillations allows to separate the RGB from the RC and SRC. Figure 1.9 presents these three populations identified asteroseismically using *APOKASC* asteroseismic and spectroscopic surface gravity, and spectroscopic effective temperature.



**Figure 1.9** – Stars from *APOKASC* identified asteroseismically as SRC (green), RC (blue) and first-ascent RGB (red). Gray dots represent stars without explicit classification. (Left) H-R diagram. (Right) difference between asteroseismic surface gravity and spectroscopic surface gravity as a function of asteroseismic  $\log g$ . Credits from Pinsonneault et al. [2014] (fig. 3 of the article)

The combination of main RC + secondary RC is sometimes referred in the literature as the *dual RC* [Girardi et al., 2000, 2009, 2010]

## 1.2.4 Red and Asymptotic Giant Branch Bumps

According to stellar evolution models, a similar phenomenon occurs during the RGB phase (after the first dredge-up) and at the beginning of the AGB: a *bump* in luminosity. In both cases the convective envelope of the star moves inwards. Depending on its deepest penetration, this convective envelope may enter the region that was previously occupied by  $H$ , producing a discontinuity in the  $H$  abundance at that particular point [Thomas, 1967, Iben, 1968, Sweigart, 1978, 1987]. The approach of the convective envelope to the  $H$  shell causes the first to retreat outwards, and as the  $H$ -burning shell expands, it could eventually reach and cross this discontinuity. Consequently, the interior structure changes: the fuel is increased while the core continues to gain mass and the matter in the shell expands and cools slightly. The evolution of the star is somehow slowed down, and the stellar surface luminosity drops temporarily.

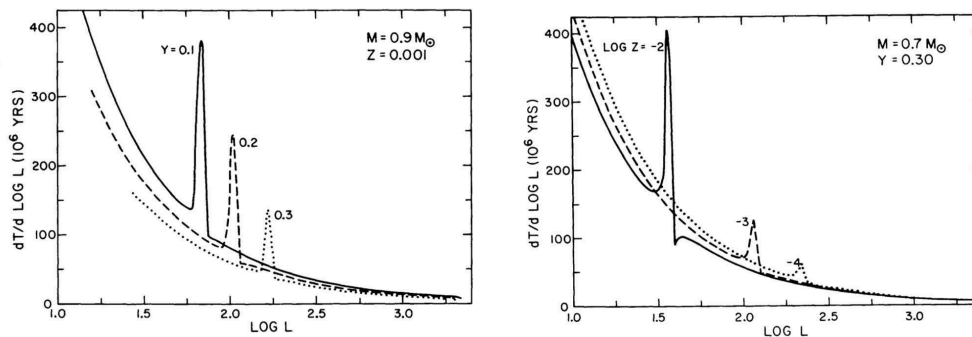
Once the extra fuel is consumed, the thermal equilibrium is restored, the shell reignites and the star begins brightening again, increasing again its luminosity monotonically.

As a consequence, the star has the same luminosity in three separate occasions since it crosses three times the same narrow luminosity interval. This leads to an over-density of stars at the given RGB or AGB luminosity range, and clearly resulting in a pronounced peak in the luminosity function of the given stellar population [Cassisi et al., 1997]. This excess in luminosity is known as the *bump*. Depending on the stellar evolution phase at which it occurs, RGB or AGB, it is called, respectively, the *Red Giant Branch Bump* (RGGB) or the *Asymptotic Giant Branch Bump* (AGBB; sometimes also called *AGB clump*).

### Red Giant Branch Bump

The RGGB occurs for low mass stars (below  $2.2M_{\odot}$ ). For massive stars  $He$  starts to ignite before the  $H$ -burning shell reaches the composition discontinuity [Christensen-Dalsgaard, 2015], hence, no *bump* is produced.

The properties of this peak as well as the position in magnitude of the *bump* in the CMD, mainly depends on the population age (the mass), the  $He$  abundance and the metallicity [Sweigart, 1987, Cassisi and Salaris, 1997, Bono et al., 2001, Riello et al., 2003, Bjork and Chaboyer, 2006, Di Cecco et al., 2010, Cassisi et al., 2011, Nataf et al., 2011a,b, 2013a]. For small values of abundance and large values of metallicity, the penetration of the convective envelope, hence the point where the hydrogen discontinuity is produced, will be deeper. Therefore, the peak will be more pronounced and the luminosity at which it occurs (the location of the *bump* in the CMD) will be fainter [Sweigart, 1978, Beaulieu and Sackett, 1998, Riello et al., 2003]. Indeed, the magnitude of the RGGB critically depends on the maximum depth of the convective envelope. Figure 1.10 from Sweigart [1978] show this dependence of the *bump* with the chemical composition.

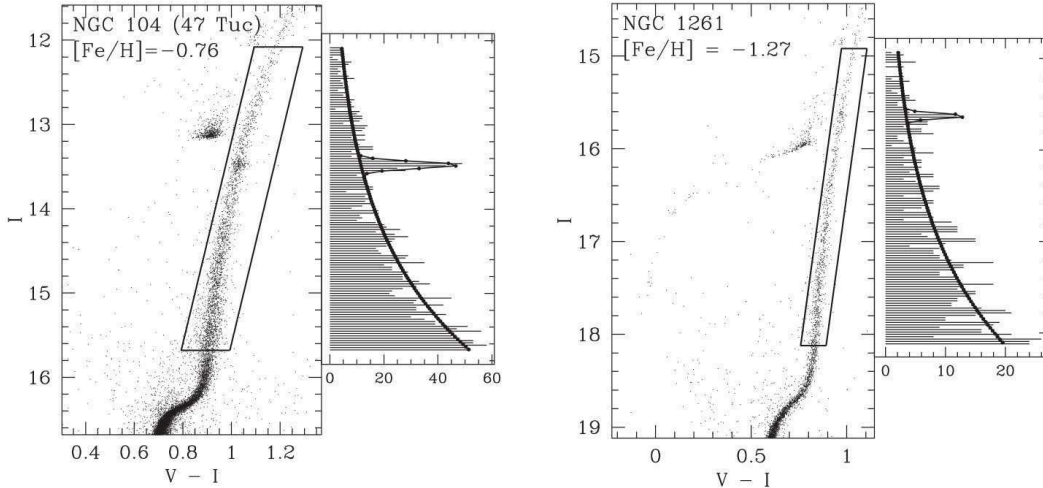


**Figure 1.10** – Theoretical dependence of the red-giant luminosity function with (left) the abundance  $Y$  for a given mass  $M$  and metallicity  $Z$ , and (right) with  $Z$  for a given  $M$  and  $Y$ . The ordinate corresponds to the time spent per unit interval in  $\log L$  in units of  $10^6$  yr. Credits from Sweigart [1978] (figs. 6 and 7 of the article)

This feature was theoretically predicted by Thomas [1967] and Iben [1968], and empirically confirmed for the first time by King et al. [1985] in the globular cluster *NGC 104 (47 Tuc)*. Since then, several

observational and theoretical investigations have been done on this subject.

As remarked by [Fusi Pecci et al. \[1990\]](#), large samples are needed to detect the *bump* especially in the most metal-poor populations. As indicated above, when increasing the metallicity (and age) the RGBB becomes bigger and appears at lower parts of the RGB (more populated zones), thus fainter than the HB. This fact can be observed in [Fig. 1.11](#) where the position and size of the *bump* is clearly different for two globular clusters (hereafter GCs) with different metallicities. This explains why there are more studies concerning metal-rich populations than metal-poor, as the detection is quite easier. So far, the RGBB has been detected in the following populations, regions and/or surveys:



**Figure 1.11** – Colour-magnitude diagrams and magnitude distributions of two globular clusters in *Hubble Space Telescope Advanced Camera for Surveys* (HST ACS) data: NGC 104 or 47 Tuc (left) and NGC 1261 (Right). For 47 Tuc there are 2415 RG+RGBB stars, 545 red HB stars and the RGBB is located at  $V - I = 1.03$  and  $I = 13.48 \pm 0.01$ . For NGC 1261 there are 808 RG+RGBB stars, and the RGBB is located at  $V - I = 0.972$  and  $I = 15.63 \pm 0.01$ . Credits from [Nataf et al. \[2013b\]](#) (Figs. 2 and 4 of the article)

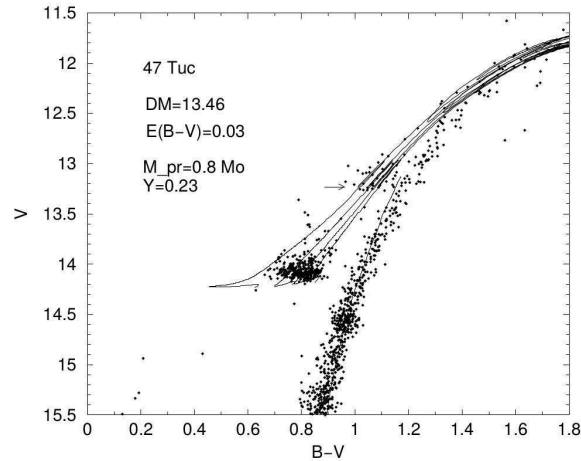
- **Globular Clusters (GC):** besides the so-known *47 Tuc* analysed by [King et al. \[1985\]](#), the RGBB have been also identified in several other 60 GCs (both metal-rich and metal-poor)
- **Local Group dwarf galaxies:** *Sculptor*, *Sextans*, *Sagittarius*, *Ursa Minor*, *LeoII*,...
- **Milky Way Bulge:** [Nataf et al. \[2011b\]](#) measured the RGBB of the Galactic bulge, using the OGLE-III photometric data. This *bump* is the most metal-rich ever detected, and even if the peak in luminosity corresponds to the expected magnitude, the star counts relative to the Galactic GCs is specially low, which means that the *He* abundance must be quite important and higher than in GCs. Related studies are linked with the X-shaped form of the bulge (see section 1.2.5).
- **Hipparcos survey:** [Girardi et al. \[1998\]](#) identified a *faint bump-like feature* at  $0.0 < M_I < 0.6$  magnitudes in the Hipparcos CMD diagram
- **Asteroseismic surveys:** asteroseismology is highly contributing to better determine stellar fundamental parameters. Therefore new and more precise surveys (i.e. *LAMOST*, *APOKASC*) have been generated allowing to analyse and review stellar populations in detail. Thanks to this accuracy, specific CMD features such as the RGBB can be clearly identified [[Kallinger et al., 2010](#), [Pinsonneault et al., 2014](#), [Liu et al., 2015](#), [Wan et al., 2015](#)].

The RGBB provides information about the inner evolution of giants at this stage, their chemical profile, and the properties (such as ages and metallicities) of mixed populations. Moreover, like the RC, it is also useful for distance determination: e.g. when the metallicity of a GC is known, it is possible to use the *bump* luminosity as *standard candle* [[Fusi Pecci et al., 1990](#), [Cassisi and Salaris, 1997](#), [Cho and Lee, 2002](#)].

### Asymptotic Giant Branch Bump

After the exhaustion of the central  $He$ , the  $He$  burning rapidly moves from the center toward the location attained by the convective core during the HB phase [Castellani et al., 1991]. This causes a rapid increase in luminosity, which observationally is translated into a *gap* between the HB and AGB phases in the CMD of the given stellar population, where a few stars should be found [Ferraro, 1992].

Once the shell  $He$ -burning stabilizes and the central  $He$  is exhausted, the *bump* occurs, just at the beginning of the AGB phase.



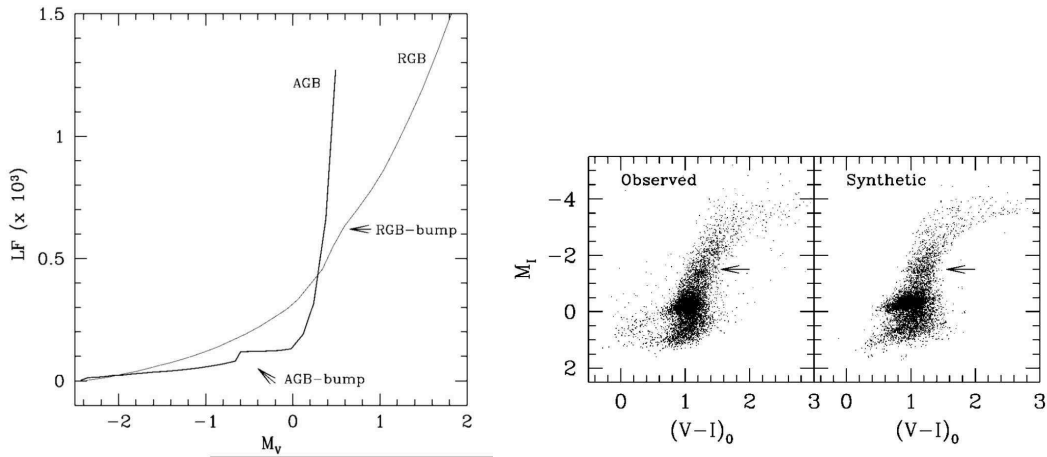
**Figure 1.12** – Observed CMD of 47 Tuc with superimposed selected evolutionary tracks of  $He$ -burning models with a  $0.80 M_{\odot}$  ( $Z = 0.006, Y = 0.23$ ) RGB progenitor. The horizontal arrow marks location of the AGBB, at  $V \sim 13.1$ . Notice that the RC and the RGBB can be clearly observed at  $V \sim 14$  and  $V \sim 14.5$ , respectively). Credits from Cassisi et al. [2001] (fig. 3 of the article)

Both observational effects, the *gap* and the *bump* had been highlighted in stellar evolutionary tracks by several authors [Castellani et al., 1991, Ferraro, 1992, Pulone, 1992, and references therein]. Ferraro [1992] also reported a preliminary identification of this feature in three GGCs ( $M5$ ,  $NGC 1261$ ,  $NGC 2808$ ), and Montegriffo et al. [1995] showed that it is clearly visible in 47 Tuc. Fig. 1.12 shows the CMD of 47 Tuc observed by Cassisi et al. [2001]. This is one of the very few diagrams we may find in the literature including three features at once: the AGBB, the RGBB and the RC.

Since then the AGBB has been identified in a large number of other GCs, nearby galaxies (like *Sextans* and *M31*). Gallart [1998] also noticed that the feature had already appeared before in other published data of GCs and galaxies. Figure 1.13 shows the prediction of the RGBB and AGBB in stellar models (left panel), as well as the observed and synthetic CMDs obtained for a *M31* halo field (right panel), where this feature is signalled with an arrow.

As a result of the short evolution time for the AGBB ( $\sim 10^7 yr$ ), the amount of data at this phase is quite low. Therefore, the identification of such a clump as a different population is difficult. This explains why it has passed unnoticed in previous works. Moreover, the AGBB involves a smaller fraction of stars than the RGBB, and generally it can only be clearly detected in a CMD with a relatively large number of stars.

The luminosity at which the AGBB occurs depends significantly on the maximum extension of the mixed core achieved during the  $He$  core burning phase. Hence, it indirectly depends on all parameters that influence the mass evolution, like the RGB mass-loss efficiency, the initial metallicity and the age of the stellar population [Ferraro, 1992, 2002, Salaris, 2013, Bossini et al., 2015]. Thus, higher masses tend to generate brighter and redder AGBBs [Castellani et al., 1991, Ferraro et al., 1999]. For metal-poor clusters (in general those with a blue HB) the AGBB is expected to become cooler, bluer and less clumpy with decreasing stellar mass, therefore more difficult to detect. Moreover, for a single-age population, the AGBB is bluer than the RGB stars at the same luminosity (as in fig. 1.12), while in composite stellar populations it appears to be entangled in the RGB-AGB area.



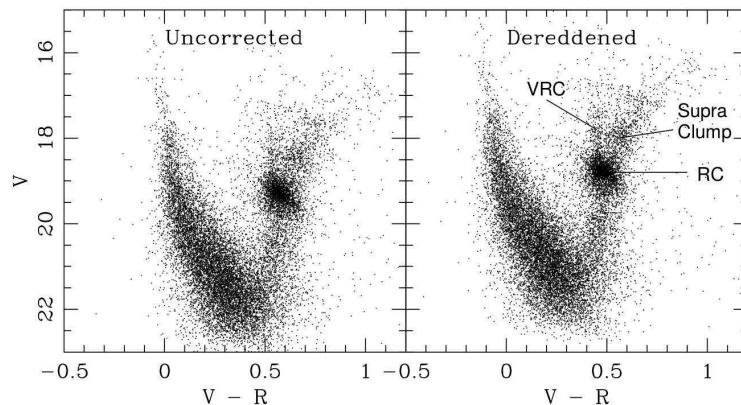
**Figure 1.13** – (Left) Luminosity function for the RGB and AGB populations predicted by Padova stellar models, with the positions of RGBB and AGBB indicated. (Right) Observed and synthetic CMDs of the M31 halo field. The arrow shows the position of the AGBB. Credits from [Gallart \[1998\]](#) (Figs. 2 and 5 of the article)

A further important property of the AGBB is the weak dependence of the difference in luminosity between the AGBB and the HB ( $\Delta V_{\text{HB}}^{\text{AGBB}}$ ) with metallicity [[Castellani et al., 1991](#)] and initial *He* abundance [[Bono et al., 1995](#)]. This makes the AGBB a potential *standard candle* for distance determination as well.

#### Metal-poor populations: RGBB or AGBB?

We have seen that the position of the RGBB is quite sensitive to stellar age and metallicity. Thus, it will be fainter than the RC for older and more metal-rich stars (as observed in the CMDs of the two GCs at different metallicities of fig. 1.11). If we read this the other way around, younger and metal-poor RGBBs will appear above the RC, like the AGBB. Therefore, in these cases there might be observational potential confusion between both features.

The first evidences of the RGBB-AGBB controversy appear in [Zaritsky and Lin \[1997\]](#) and [Beaulieu and Sackett \[1998\]](#). In these papers both authors studied the RC morphology in the LMC, with a special emphasis on the VRC. Nonetheless, figure 1 of [Zaritsky and Lin \[1997\]](#) includes the position of a RGBB above the RC (no details are given) and [Beaulieu and Sackett \[1998\]](#) find the same overdensity, calling it *supraclump* (fig. 1.14).



**Figure 1.14** – Calibrated composite ( $V-R, V$ ) CMDs for four LMC fields, uncorrected (left) and dereddened (right). Positions of the RC, the VRC, and the supraclump are indicated. Credits from [Beaulieu and Sackett \[1998\]](#) (fig. 1 of the article)

The last authors tried to identify this *supraclump* with both the RGBB and the AGBB, in comparison with different stellar models. While older models agreed with a RGBB feature, models from [Gallart \[1998\]](#)



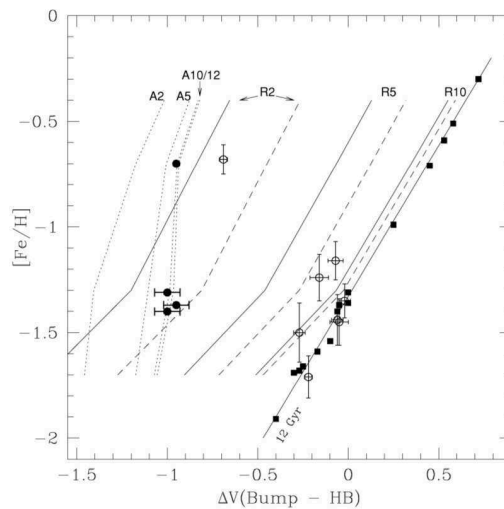
associated it with the AGBB. They finally consider the LMC overdensity as the AGBB, hence corresponding to stars belonging to a relatively long-lived phase of *He* core exhausted stars at the base of the AGB. That is, the RGBB of LMC giant stars is predicted to be located slightly below the RC, i.e. the upper overdensity actually belongs to the AGBB.

This was a case where the RGBB was just incorrectly located and misidentified. However, as indicated above, for metal-poor stellar populations the identification may be still more tricky because in this situation:

- both *bumps* appear above the RC (low metallicity or young populations implies bright bumps)
- the number of stars decreases with increasing luminosity along the RGB

The difficulty is compounded in young populations, since the duration of the pause in the star evolution due to the discontinuity in the chemical abundance profile near the convective envelope, is shorter relative to the pace of evolution up the RGB. Therefore, the contrast in number of stars between RGBB and RGB in the luminosity function is lower. Observationally, we would see less RGBB stars, hence it would be harder to distinguish the *bumps* on the RGB.

Alves and Sarajedini [1999] introduced the  $\Delta V_{\text{Bump}}^{\text{HB}} - [\text{Fe}/\text{H}]$  diagram (fig. 1.15) which shows the brightness difference between the AGBB or RGBB and the HB. Solid lines, labelled R2, R5 and R10, correspond to the locations of RGBB in stellar populations of ages 2, 5 and 10 Gyr, respectively, assuming the HB has no dependence on age, while the short-dashed lines show the same assuming certain dependence. The dotted lines, labelled A2, A5 and A10/12, show the locations of AGBBs, assuming age-dependent values (see the article for more details on these values). Filled squares and circles correspond respectively to the RGBB and AGBB observational data of 8 GCs from the SMC.



**Figure 1.15** –  $\Delta V_{\text{Bump}}^{\text{HB}} - [\text{Fe}/\text{H}]$  diagram which shows the brightness difference between the AGBB or RGBB and the HB. Filled circles and squares correspond to observational data for 8 GCs from the SMC. Solid lines show the position of the RGBB at different ages (R2, R5, R10, belonging to 2, 5, and 10 Gyr, respectively). Short-dashed lines indicate the same but assuming an HB dependence on age. Dotted lines show the location of the AGBB, also at 2, 5, and 10/12 Gyr. Credits from Alves and Sarajedini [1999] (fig. 6 of the article)

From this diagram, we may validate again that, indeed, the difference in magnitude (HB - *bump*) for the AGBB is weakly dependent on metallicity, whereas the RGBB is more dependent. It also predicts  $M_{V, \text{AGBB}}$  to be always brighter in the younger of two stellar populations with similar metallicity, which allows unambiguous identification of RGBBs in certain regions of this diagram. For example,  $\Delta V_{\text{Bump}}^{\text{HB}} > -0.8$  mag is likely to be an RGBB, regardless of the age or metallicity of the stellar population. But, if  $\Delta V_{\text{Bump}}^{\text{HB}} < -0.8$ , the situation is less clear, the *bump* may be a young RGBB or an AGBB.

The authors suggest that this diagram may be used as a consistency check for metallicity and age determinations of GCs measured via other techniques. Additionally, it can be used as a diagnostic tool for the

interpretation of CMDs of field populations in nearby galaxies, where the different ages and metallicities of the component stellar populations are not necessarily known with high accuracy. In this case, giant branch *bumps* may be identified as either RGBB or AGBB by their locations in the  $\Delta V_{\text{Bump}}^{\text{HB}} - [\text{Fe}/\text{H}]$  diagram.

In summary, the combination of star counts and luminosity differences between each *bump* and the HB, their luminosity functions, the stellar models predictions, their dependency level on metallicity and age, and the  $\Delta V_{\text{Bump}}^{\text{HB}} - [\text{Fe}/\text{H}]$  diagram, should aid to distinguish both features even in the more subtle and fuzzy cases.

### A Secondary Red Clump incorrectly identified?

As previously seen, there are some faint concentrations of stars that come out along the RGB and slightly below the main RC. This is the case of both the RGBB of metal-rich stellar populations and the SRC of some CHeB stars. One must then be careful not to confuse them.

In this sense, we would like to note the RC/SRC asteroseismic identification of Casagrande et al. [2014] in their figure 17 (here: fig. 1.16). In their top panel we may easily observe by eye the presence of the RC at an almost constant  $\log g = 2.5$ , and another elongated and downward tilted over-density just below it, which would correspond to the RGBB. In the low panels they show the distribution in mass of both the RC (blue) and the SRC (red) based on two different selections. In panel a) it is just based on the  $\log g$ ; as seen in the upper inset, both ranges are coherent with the regions where the RC and the SRC should be. In panel b), however, the selection was done based on a *certain clump classification* which is not described in the article. Considering the already mentioned constraints in mass ( $M > M_{\text{HeF}}$ ), metallicity ( $Z > 0.004$ ) and ages ( $> 1\text{Gyr}$ ) expected by stellar evolution for these features, the red points of the upper inset in panel b) would be covering more than the SRC, actually part of the RGBB.

In figure 1.17 we show the same kind of plot as their top panel, but using APOKASC data (seismic  $\log g$ ) with APOGEE  $T_{\text{eff}}$ . We have selected SRC stars from *Hipparcos* [van Leeuwen, 2007] by constraining masses ( $M > 2M_{\odot}$ ), surface gravities ( $2.6 < \log g < 3.0$ ) and metallicities ( $[\text{Fe}/\text{H}] > -0.5$ ). We also considered the cut on colour  $0.84 < V - I < 0.94$ , based on Girardi et al. [1998]. APOKASC does not contain photometric information, but all stars have a 2MASS identifier. Therefore, to do the same  $V - I$  cut, we first applied the  $V - I$  constraint on the X-matched *Hipparcos-2MASS*, thus obtaining the equivalent cut in  $J - K$ :  $-0.973 < J - K < 1.055$ . Then, we crossmatched APOKASC and 2MASS using the CDS X-Match Service<sup>5</sup>, and applied all the mentioned constraints.

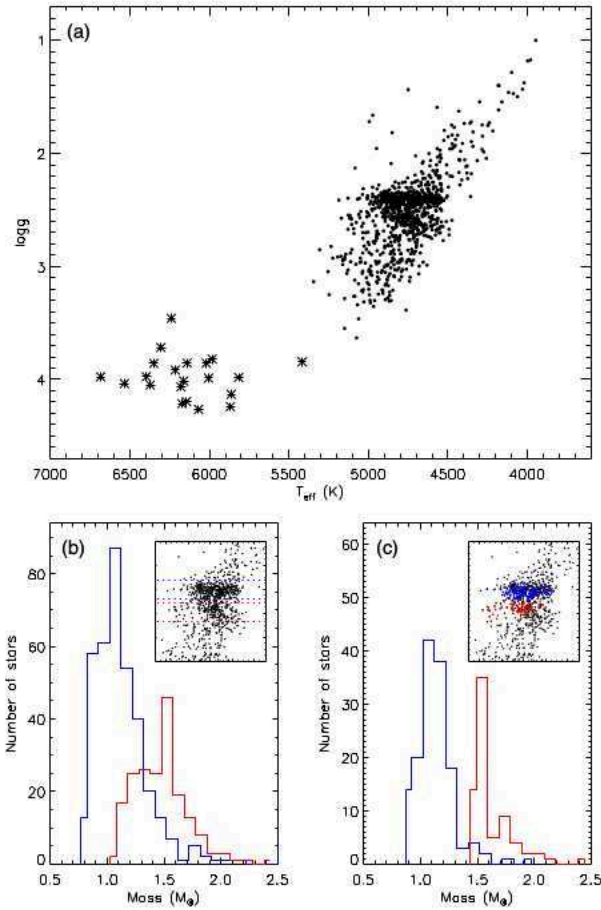
The result is a SRC similar to that of Girardi et al. [1998] shown in fig. 1.8. Additionally, fig. 1.18 from Liu et al. [2015] identifies the RC and RGBB also using spectroscopic and asteroseismic data (*LAMOST* + *Kepler*). We see that the shape of the observational data is similar as in Casagrande et al. [2014] and ours, and that they actually recognize the RGBB where it is expected. That leads to wonder whether the selection in Casagrande et al. [2014] belongs to a wrong identification. Nonetheless, none of the works that cites this paper gives a reference about this.

Finally, we also wonder why the VRC is not seen in none of the asteroseismic surveys published by all these authors.

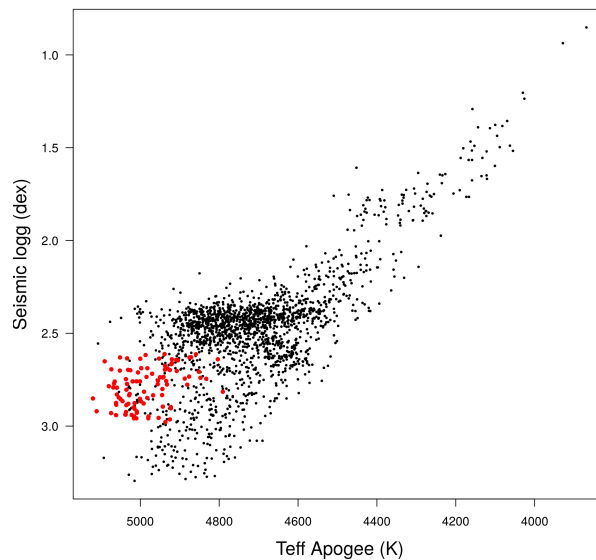
### 1.2.5 Double RC-RGBB pairs

Besides all previous features, we may also observe specific effects and combinations of them in some Colour Magnitude Diagrams. This is the case of the *double Red Clump* (double RC), the *double Red Giant Branch Bump* (double RGBB), or even both together (double RC-RGBB pair). Their occurrence may depend on the stellar composition, line of sight, geometric effects or photometric bands. Hence, the reasons for observing these "duplications" vary depending on the object we are looking to. We present here some cases where one or more of these features have been detected.

<sup>5</sup>CDS X-Match Service: <http://cdsxmatch.u-strasbg.fr/xmatch>



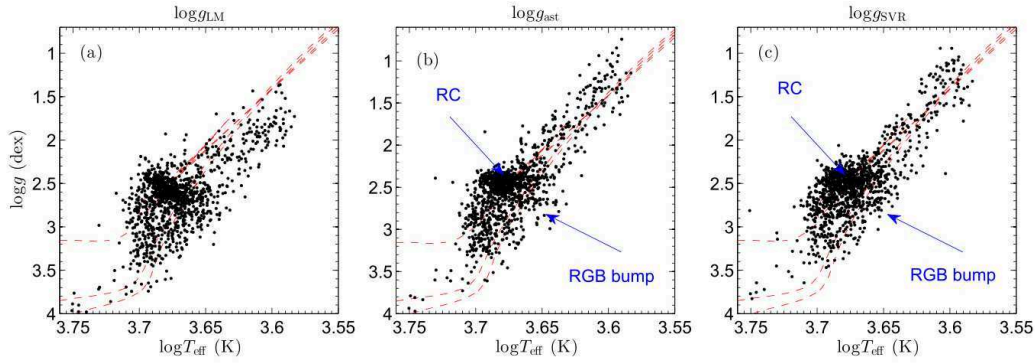
**Figure 1.16** – (Left, top)  $T_{\text{eff}}$  and  $\log g$  for *Kepler* stars. (Left, bottom) First (blue) and secondary (red) RC dissection and the corresponding histogram for their mass distribution; selection based on a)  $\log g$  (range indicated by dotted lines in the upper inset), b) a certain clump classification. Credits from Casagrande et al. [2014] (fig. 17 of the article)



**Figure 1.17** – Effective temperature versus  $\log g$  for the *APOKASC* data. Red points correspond to the selection of SRC stars based on mass ( $M > 2M_{\odot}$ ), surface gravity ( $2.6 < \log g < 3.0$ ), metallicity ( $[\text{Fe}/\text{H}] > -0.5$ ) and colour range ( $-0.973 < J - K < 1.055$ , based on Girardi et al. [1998]  $0.84 < V - I < 0.94$  cut)

### Double RGBB in dSph galaxies: AGBB contamination?

It has been found that in some dwarf spheroidal galaxies there exist more than one RGBB. Indeed, some of them present a *double RGBB*, as in *Sextans* [Bellazzini et al., 2001], *Sculptor* [Majewski et al., 1999] or



**Figure 1.18** –  $T_{\text{eff}}$  -  $\log g$  diagrams for the (a) *LAMOST* pipeline, (b) the *Kepler* asteroseismology, and (c) revised *LAMOST* + *Kepler*. The  $T_{\text{eff}}$  values in the three panels are all from the *LAMOST* pipeline. Red dashed lines correspond to isochrones with  $[\text{Fe}/\text{H}] = -0.2 \text{ dex}$  at the age of 1, 5, and 10 Gyr, from top to bottom. The red clump stars and the RGB bump stars in the second and third panels are indicated. Credits from Liu et al. [2015] (fig. 7 of the article)

*Leoll* [Bellazzini et al., 2005]. This feature has been in general interpreted as the presence of two stellar populations with different metallicities. However, it is argued whether the brighter *bump* correspond instead to the AGBB of the population that generated the RGBB (the fainter *bump*) [Lee et al., 2003]. Nevertheless, this hypothesis is not completely supported [Bellazzini et al., 2004] because if this was the case:

- the double *bump* should be detected in all galaxies where a RGBB is found. But this is not the case for *Ursa Minor* [Bellazzini et al., 2002] or *Sagittarius* [Monaco et al., 2002], where there is just one *bump*
- the ratio between the number of stars in the peak of each *bump* should be quite different from 1 since the number of stars for the RGBB is  $\approx 2$  times larger than for the AGBB. But a ratio  $\approx 1$  is found, i.e. the signal detected for the AGBB should be much lower than what is observed
- the magnitude of the second *bump* should be similar to the expected AGBB magnitude, but it results to be significantly brighter
- the luminosity difference between the brighter and the fainter *bumps* should not depend on metallicity if they were a RGBB and an AGBB (as predicted by Ferraro et al. [1999]), but it is found to decrease with increasing the metallicity

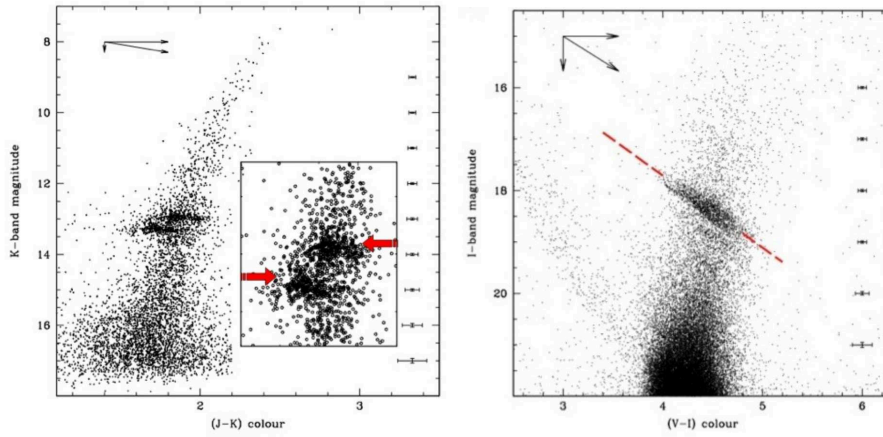
Therefore, it seems that the two *bumps* detected in some dSph's belong to a *double RGBB* and not to a combination of RGBB + AGBB. Assuming that the second *bump* is associated with the most metal poor populations observed in the considered dSph's, the *double RGBB* feature suggests that in these galaxies the SFH at early epochs was somehow structured in different main episodes [Bellazzini et al., 2001, 2004].

### Double RC in Globular Clusters

Another example is the presence of a *double RC* in the GC *Terzan 5* [Ferraro et al., 2009] (see figure 1.19), where both *clumps* are separated in magnitude and colour (0.15 mag in J-K), leading to interpret them as two populations with different ages and iron content, or similar ages with different iron and helium content [Nataf et al., 2011a]. Also in the *Sagittarius stream* Correnti et al. [2010] detect multiple *double RCs* often separated by 0.5 mag or greater, which can be geometrically explained due to the intersection between the RCs sightlines and the multiple wraps of stars.

### Double RC-RGBB pairs in the Galactic Bulge

Even more interesting was the discovery of a *double RC* toward the Milky Way Bulge region [McWilliam et al., 2010, Zoccali, 2010, McWilliam and Zoccali, 2010, Nataf et al., 2010]. As explained in section 1.2.1,



**Figure 1.19** – CMD of *Terzan 5* in the K-band (left) and I-band (right). In the magnified view of left panel the two RCs are marked with red arrows. In the right panel, still two different parallel structures are distinguishable above and below the dashed line. Notice that in both cases, the structures are nearly parallel and stretched along the reddening vector (shown in the upper left of each panel). Credits from [Ferraro et al. \[2009\]](#) (figs. 2 and supplement 1a of the article)

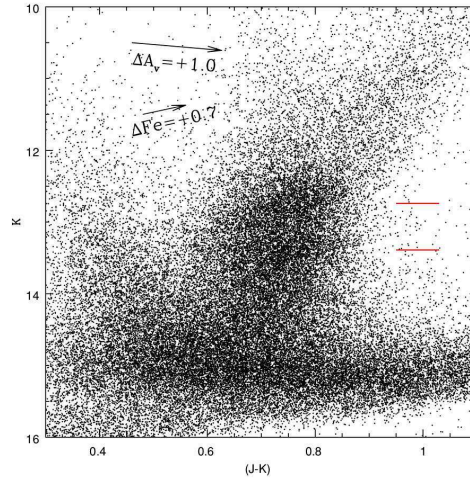
the RC is known to be a good *standard candle* for distance determination as well as for Milky Way structure indication. Therefore, the presence of two RCs in the direction of the Galactic Bulge provides a powerful tool to study its shape.

As seen in the *COBE/DIRBE* near and far-infrared light distribution [[Dwek et al., 1995](#), [Binney et al., 1997](#)] and the *2MASS* star counts [[López-Corredoira et al., 2005](#)], the Milky Way has a boxy/peanut-shaped bulge, because it contains a bar. Since the work of [Stanek et al. \[1994, 1997\]](#) we know that this bar is tilted by  $\approx 20^\circ - 25^\circ$  from the Sun-Galactic center direction, with the near side in the first Galactic quadrant. Similar results have been obtained using *OGLE-II* or near IR-photometry surveys along the plane, confirming this inclination [[Babusiaux and Gilmore, 2005](#), [Nishiyama et al., 2005, 2006](#), [Rattenbury et al., 2007b](#)].

Nevertheless, the distribution of stars associated with the boxy/peanut structure shows complexity at higher Galactic latitudes [[Saito et al., 2011, 2012](#)]. Indeed, near the minor axis, at  $b < -5^\circ$ , recent optical and near-IR stellar photometry analysis using *OGLE-II*, *OGLE-III* and *2MASS* surveys, show a split of about 0.5 mag of the RC along the line of sight (which corresponds to a distance of  $\approx 1.5\text{kpc}$ , one located at 6.5 kpc and the other at 8.8kpc). Figure 1.20 shows both bulge RCs for the field  $(l, b) = (-1, -8)$  in the *2MASS* catalogue. As suggested by [McWilliam and Zoccali \[2010\]](#), [Nataf et al. \[2010\]](#), this observation could be associated with a more complex X-shaped structure. That is, along the  $l = 0^\circ$  axis the RC appears to split into two components for  $|b| > 5^\circ$ , with the separation between the peaks increasing with the Galactic latitude. Hence, at the latitude of Baade's Window ( $b = -4^\circ$ ) both RCs are merged. This in agreement with the expectation and theory of an X-shaped structure. Actually, other boxy/peanut bulges with X-shaped structures have already been detected, like in *NGC 128*, *NGC 3625*, and *NGC 4469*. Their formation is predicted to be due to a particular strong buckling of a bar [[Patsis et al., 2002](#), [Athanasoula, 2005](#)].

The main characteristics of this observational feature can be summarized as follows:

- For  $|b| < 5^\circ$  only one RC is observed
- For  $|b| > 5^\circ$  the double RC is seen only close to the minor axis of the bulge ( $-4^\circ < l < +4^\circ$ ).
- Both RCs have equal  $V - I$  and  $J - K$  colours
- The second RC peak is  $\sim 0.5$  brighter than the first one; the magnitude separation between the two RCs increases with decreasing Galactic latitude from 0.4 mag at  $b = -5^\circ$  to 0.7 mag at  $b = -10^\circ$  [[Saito et al., 2011](#)]
- This separation as well as the variation of each RC magnitude, is independent of the observational band and longitude. This confirms that the *double RC* cannot be due to two extinction patches in this



**Figure 1.20** –  $K$  vs  $(J-K)$  CMD for the field  $(l, b) = (-1, -8)$  of the Galactic Bulge, with the *2MASS* catalogue data. The two RCs are visible showing a slight upward trend. Red lines indicate the observed  $K$ -band RC peaks at  $(l, b) = (+1, -8)$ . The upward pointing vector shows the change in RC  $K$  vs  $(J - K)$  from  $[Fe/H] = -0.70$  to  $[Fe/H] = 0.00$ , predicted from the Teramo stellar evolution models, and is consistent with the observed RC slopes. Also shown is a downward reddening vector for  $\Delta A_V = 1.0 \text{ mag}$ . Credits from [McWilliam and Zoccali \[2010\]](#) (fig. 2 of the article)

particular direction since the separation between the two peaks would be then smaller in the  $K$ -band, as compared to the  $I$ -band [[Gonzalez et al., 2015](#)]

- Both RCs are equally populated. However, at a fixed Galactic latitude, the fainter RC becomes gradually less populated towards lines of sight at increasing longitudes, until only the brighter RC can be detected. The opposite effect is observed towards negative Galactic longitudes [[McWilliam and Zoccali, 2010](#)]
- The fainter RC becomes narrower for sightlines further from the plane, while the brighter becomes broader
- The same characteristics are observed in both the northern and southern hemispheres
- The two peaks cannot be due to the RGBB, nor the AGBB, falling close to the RC because in that case the two would also occur in Baade’s Window. Moreover, number statistics also firmly rule out the possibility of having one RC and one RGBB or AGBB [[McWilliam and Zoccali, 2010](#)]

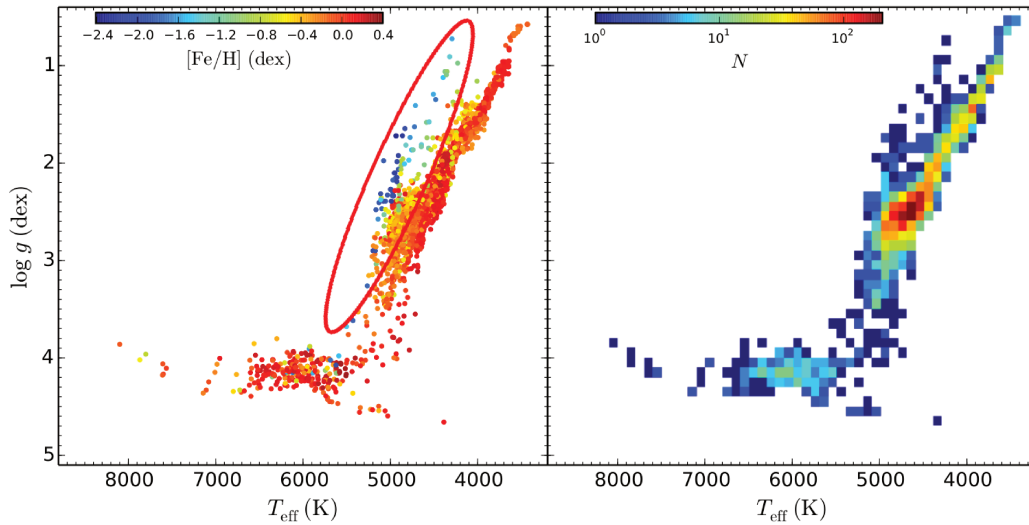
Apart from the bulge *double RC*, an additional much fainter component has also been detected at  $K_{s,0} \sim 14.3$  [[Gonzalez et al., 2015](#)] which could be linked to the RGBB [[Nataf et al., 2010](#)] or to the background stellar disc [[Gonzalez et al., 2011](#)]. [Nataf et al. \[2011b\]](#) suggest that sight lines with two RCs have two RGBBs with similar properties. By using residuals to a single-clump fit on small scales they show the clear signal of two clumps and two bumps: a *double RC-RGBB pair*. If both pairs are found to have the same separation in brightness, this will confirm the X-shaped Galactic Bulge. Even an AGBB could have been detected.

Finally, [Lee et al. \[2015\]](#) proposed another interpretation for this *double RC*. They consider a classical bulge with an additional population enriched in  $He$  co-existing with a bar. The Milky Way bar would not be sufficiently buckled to form an X-shaped structure, so the *double RC* would be generated by a massive classical bulge with an important fraction of stars enriched in  $He$ . However, as noted by [Gonzalez et al. \[2015\]](#), this theory would imply the existence of *double RCs* everywhere in the Galactic bulge, while it is observed just in a range of longitudes, and some of the above observational properties would not be explained or even possible. Variations in both iron and helium distributions would have to be considered. With the upcoming *Gaia* parallaxes, these discrepancies should be clarified.

### 1.2.6 An unexplained feature?

In a recent paper, Ren et al. [2016] derived surface gravities using spectroscopic and asteroseismic information for a set of stars selected from *LAMOST* and *Kepler* surveys. They also investigated the impact of  $T_{\text{eff}}$  and  $[\text{Fe}/\text{H}]$  estimates on asteroseismic determinations of  $\log g$ . They finally compared their results with previous works, with good agreement with all of them.

However, we would like to highlight the first figure of their article, where they plot their derived surface gravities as a function of the effective temperature. The sample used in this figure contains 3888 single stars for which they obtained their own stellar atmospheric parameters. For the objects that were observed more than once, they only used the atmospheric parameters of the *LAMOST* spectrum because they have higher signal-to-noise ratios (SNR): 3685 stars have SNRs higher than 10. Their plot is shown in fig. 1.21.



**Figure 1.21** – Surface gravity versus effective temperature for 3685 *LAMOST-Kepler* stars, varying in (left) metallicity and (right) in stellar number density (stars grouped in bins of 100 K by 0.1 dex in  $T_{\text{eff}}$  and  $\log g$ ). The unexplained feature is highlighted by a red ellipse in the left diagram. Credits from Ren et al. [2016] (fig. 1 of the article)

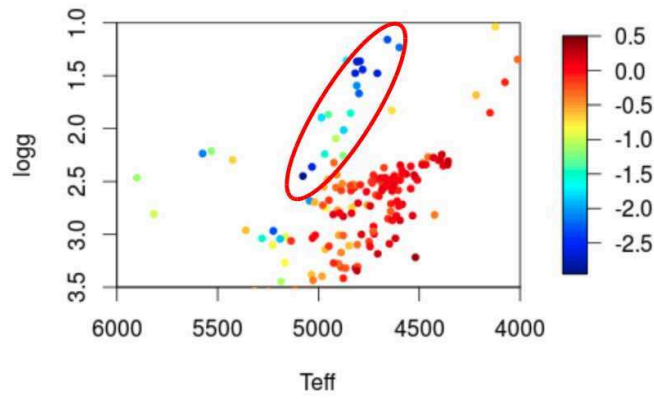
We note an unexpected feature in these diagrams (red ellipse): a plume of very low metallicity stars at the left side of the RGB (in the temperature range  $4500 < T_{\text{eff}} < 5500$ ) and having almost the same slope as the RGB, although much less populated. None of the features introduced in previous sections agree with these characteristics. We have further retrieved the *APOKASC* catalogue (also a combination of spectroscopic and asteroseismic information, *APOGEE* + *Kepler*) in order to check whether we could find the same structure. But as can be seen in previous figure 1.17 where we plot all *APOKASC* data, this feature does not appear. Either it is due to an *APOKASC* target selection bias, either there is a problem of parameters derivation in the Ren et al. [2016] work. Our final guess is that these stars belong to the galactic halo, i.e. too distant to be observed for instance in the *Hipparcos* HR diagram. This unexpected feature also appears in the *Gaia-ESO Survey*<sup>6</sup> (GES) data (fig. 1.22).

### 1.2.7 Synopsis

Table 1.2 summarizes all CMD observables described above, to facilitate comparison and clarify concepts. It briefly reviews the physics that causes each feature, the mass of the stars needed for them to occur, their observational effect on the CMD, in which evolutionary phase they are produced, their properties, in which objects they have been already detected, and their first references in the literature.

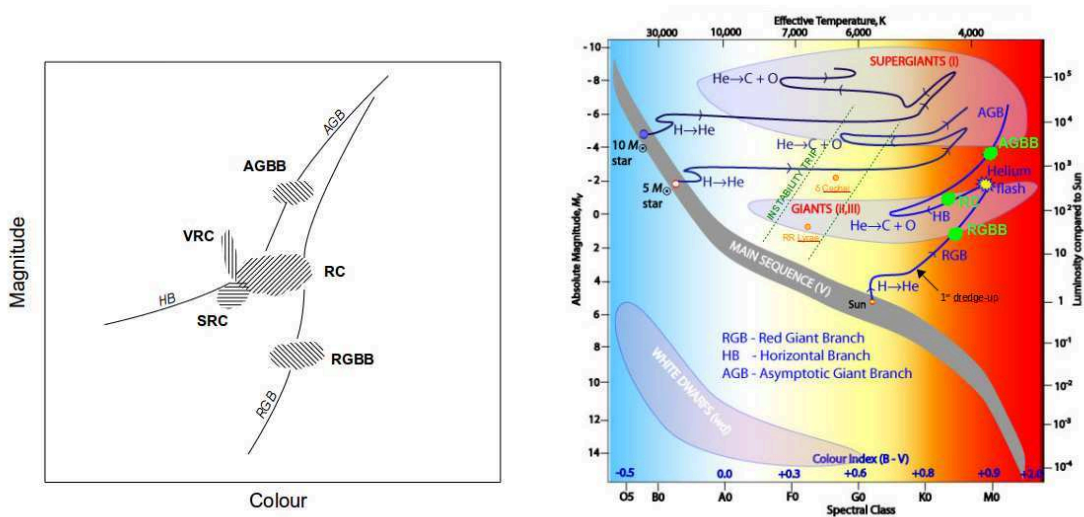
Similarly, in fig. 1.23 (left) all these observables are graphically represented in a same diagram for a composite stellar population. Their exact location will obviously depend on metallicity and age, especially

<sup>6</sup>*Gaia-ESO Survey*: public spectroscopic survey of approximately  $10^5$  stars of the Milky Way, from the halo to star-forming regions, observed by the VLT *FLAMES* spectrograph.



**Figure 1.22** – Surface gravity versus effective temperature for *UVES* (Ultraviolet and Visual Echelle Spectrograph) stars in the *Gaia-ESO Survey* (GES), varying in metallicity. Similarly as previous figure 1.21, the unexplained feature of metal-poor giant stars is highlighted with a red ellipse. Figure from C. Babusiaux (private communication)

(as already seen before) the RGBB would appear above the RC for metal-poor populations. The diagram on the right is the same as the shown at the very beginning of this chapter, fig. 1.1, but with the approximate positions of the RGBB, RC (main RC + SRC + VRC) and AGBB added in green.



**Figure 1.23** – (Left) Schematic representation of the different post-main sequence features produced by old-intermediate-age stars in a composite stellar population. (Right) Same HR diagram as fig. 1.1 with the observables indicated in green (here "RC" includes the main RC, the SRC and the VRC)



Table 1.2 – CMD giant features summary

Feature	Physics	Star masses	Observational effect	Evolutionary phase	Properties	Observed in...	First references		
RGBB	Convective envelope crosses $H$ discontinuity: the star passes three times through the same luminosity region	$M < 2.2M_{\odot}$	Overdensity	RGB	Location dependent on [Fe/H] High [Fe/H] $\rightarrow$ fainter bump	GCs, LG galaxies, Solar neighbourhood, Bulge	Thomas [1967], Iben [1968], King et al. [1985]		
RC Fine structure	Dual clump	Main RC	$He$ -burning needs $M_c \sim 0.45M_{\odot}$ (degenerate cores) to start: all stars begin with similar $M_c \rightarrow$ similar luminosities	$0.5M_{\odot} < M < 2M_{\odot}$	Overdensity	RHB	metal-rich $age > 1Gyr$ $G8III - K2III$ $4500 < T_{eff} < 5200$ $0.6 < M_I < 1.0$ almost constant $0.8 < V - I < 1.3$	GCs, OCs, LG galaxies, Solar neighbourhood, Bulge	Cannon [1970], Faulkner and Cannon [1973]
		SRC	same as for RC but for slightly massive stars (hence younger), thus a $M_c \sim 0.33M_{\odot}$ (weak degenerate cores) is needed for $He$ to ignite	$M \sim 2M_{\odot}$	Little extension of RC blue side to fainter magnitudes (up to 0.4 mag fainter)	RHB	metal-rich ( $Z > 0.004$ ) $age \sim 1Gyr$ or younger higher logg than the main RC	GCs, OCs, Solar neighbourhood, Bulge	Girardi et al. [1998], Girardi [1999a, 2000]
	VRC	$He$ ignited in non-degenerate conditions; the increase of $T_{eff}$ and mass causes the star to spend more time at the bluest part of the HB $\rightarrow$ the luminosity increases at constant colour: <i>blue loop</i> effect	$M \gg 2M_{\odot}$	Vertical plume departing from the bluest part of RC, up to 0.9 mag brighter	HB	youngest RC stars $0.9 < B - V < 1.1$	LMC, Solar neighbourhood	Zaritsky and Lin [1997], Beaulieu and Sackett [1998], Girardi et al. [1998]	
AGBB (AGB clump)	Convective envelope crosses discontinuity after $He$ exhaustion: the star passes three times through the same luminosity region	$M \gg 2M_{\odot}$	Overdensity	early AGB	always above RC	GCs, dwarf galaxies, LMC	Castellani et al. [1991], Ferraro [1992], Gallart [1998]		

## 1.3 Red Clump characterization

Among the different observables introduced in previous section 1.2, the main RC is the most powerful as distance indicator since its luminosity is almost constant, i.e. weakly dependent on metallicity, colour and age. This is not the case for the RGBB which varies significantly with metallicity, nor the VRC which is extended in luminosity along a constant range of colour, nor the SRC for which has a tilted shape. Moreover, we have seen that the RC is well populated in different objects/regions, especially in the solar neighbourhood but also in GCs and Local Group galaxies, contrary to the AGBB which is highly less populated, hence harder to detect. The RGBB and AGBB can likewise be used as *standard candles*, but the potential of the RC is clear.

With the upcoming *Gaia* data this fact becomes even more interesting, since we will have new and more accurate astrometric and photometric measurements for thousands of RC stars. However, the accuracy on distance determinations depends not only on good data, but also on the intrinsic knowledge of the RC.

A way to determine and find the physical characteristics of a population (that is: shape, structure, chemical composition, ages,...) is trying to reproduce it synthetically. Modelling implies guessing and assuming behaviours and a set of "ingredients" based on previous knowledge on individual stars or stellar populations formation and evolution. The goodness of fit between the model and the observations allows to either confirm the initial hypothesis (hence, to validate our comprehension of the physical processes behind that object or population) or to review them again if the results are not satisfactory.

Another way, is to use well known stellar models and check if observations correctly follow what models predict. If the match is not good enough, one may try to extract the main differences between both, or to calibrate empirically the observations and use the results to review the theory.

Both modelling and calibrating the RC are therefore required to guarantee a good understanding of its basic properties, and thereby their correct usage as distance indicators, especially because *Gaia* will soon provide high precision parallaxes for these stars.

### 1.3.1 Modelling the RC

To characterize the *Hipparcos* RC in terms of age and metallicity distribution, C. Babusiaux developed a model to simulate a synthetic stellar population, similar to TRILEGAL [Girardi et al., 2005], but using the age-velocity dispersion relation as a constraint to the scale height ( $H_z$ ) of the populations as done in Just and Jahreiß [2010] and assuming the galactic disk to be a set of isothermal mono-age ( $\tau$ ) populations. Given these assumptions, the Jeans equation becomes a *sech*<sup>2</sup> profile:

$$\rho(z) = \text{sech}^2\left(\frac{z}{2H_z}\right)$$

$$\text{with } \text{sech}(x) = \frac{2}{e^x + e^{-x}} \text{ and } H_z = \frac{\sigma_z}{\sqrt{8\pi G\rho_0}}$$

with  $\sigma_z$  the uncertainty on  $z$ ,  $G$  the gravitational constant and  $\rho_0$  the initial density.

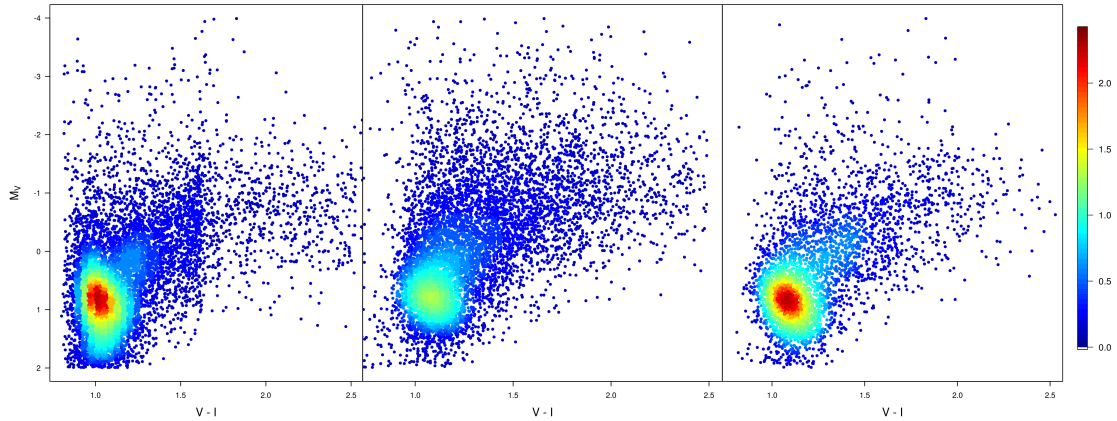
Given that for a given age  $\tau$ ,  $\rho_0$  depends on the Star Formation Rate (SFR)  $\psi(\tau)$  and assuming that all populations have the same scale length, the integration of the density law in the  $z$ -direction becomes:

$$\Sigma = \int \rho(z) dz = \left[ 2\rho_0 H_z \tanh\left(\frac{z}{2H_z}\right) \right]_0^\infty = 4\rho_0 H_z$$

Replacing  $H_z$  and combining:

$$H_z = \frac{\sigma_z^2}{2\pi G\Sigma} \propto \frac{\sigma_z^2}{\psi}$$

As initial values for the simulation, C. Babusiaux took the star physical parameters from the Padova isochrones [Girardi, 2002] assuming by default:



**Figure 1.24** – Hertzsprung-Russell diagrams of (left) observed *Hipparcos* RC stars with relative parallax precision  $< 10\%$ , and two simulated cases: (middle) Constant SFR and (right) decreasing SFR. All panels show the stars as individual points where possible and where the points overlap the relative source density is shown, with colours varying from blue to red, indicating increasing density on a logarithmic scale

- a normalisation constant, obtained considering that  $H_z(1\text{Gyr}) = 150\text{pc}$
- the [Chabrier \[2001\]](#) log-normal Initial Mass Function (IMF)
- an Age-Metalllicity Relation (AMR) from [Rocha-Pinto et al. \[2000\]](#)
- a constant Age-Velocity Relation (AVR) (from 10km/s at 1Gyr, to 40km/s at 13Gyr)
- the extinction model of [Arenou et al. \[1992\]](#)
- the *Hipparcos* astrometric and photometric errors

The star positions are simulated by direct Monte-Carlo integration.

The left panel in fig. 1.24 displays the observed RC from the *Hipparcos Survey*<sup>7</sup> with relative parallax precision  $< 10\%$ , while the middle and right panels show the simulations in two different cases:

1. Constant SFR (middle panel): the scale height is simplified as:

$$H_z = z_0(1 + \tau/t_0)^\alpha$$

retrieving the expression suggested in [Rana and Basu \[1992\]](#) and used in TRILEGAL with  $\alpha = 2$

2. Decreasing SFR (right panel)

These results illustrate the work (in progress) which is still needed to properly understand the stellar population content of the *Hipparcos* RC.

### **Gaia RC: expectations**

Using the same model, we also reproduce the expected RC for *Gaia*, using simulated G band and BP-RP colour. The results obtained for RC stars with parallax precision  $< 10\%$  are shown in fig. 1.25. Due to the huge amount of stars, we have plot the simulation in two parts. The left panel includes stars at high latitudes ( $|b| > 20^\circ$ ) while the red panel includes stars in the galactic plane ( $|b| < 10^\circ$ ).

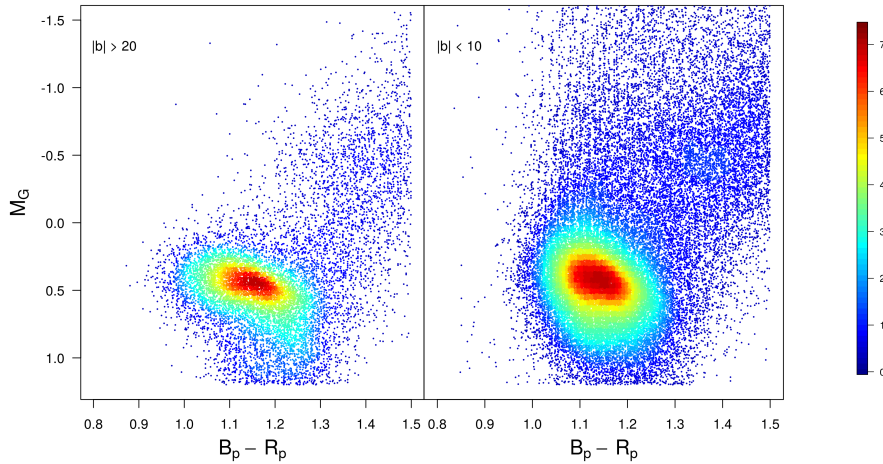
<sup>7</sup>*Hipparcos Survey*: basic list of bright stars added to and merged with the total list of proposed stars for the mission, to provide a stellar sample (almost) complete to well-defined limits. These stars are flagged with "S" in the *Hipparcos* catalogue's field H68. The limiting magnitude is a function of the star's spectral type and galactic latitude ( $b$ ):

$V \leq 7.9 + 1.1|\sin b|$  for spectral types earlier or equal to G5

$V \leq 7.3 + 1.1|\sin b|$  for spectral types later than G5

If no spectral type was available, the break was taken at  $B - V = 0.8$  mag.

52 045 *Hipparcos* catalogue are "survey" stars



**Figure 1.25** – Simulated Gaia Hertzsprung-Russell diagram of RC stars with relative parallax precision  $< 10\%$ . This illustration shows the differences of shape as a function of the stellar population. (Left) high latitudes,  $|b| > 20$ . (Right) low latitudes,  $|b| < 10$ . All panels show the stars as individual points where possible and where the points overlap the relative source density is shown, with colours varying from blue to red indicating increasing density on a logarithmic scale

This two samples show how the RC shape can vary in terms of different parameters such as metallicity, age, masses,... With the *Gaia* data we will therefore be able to subgroup the RC and better analyse these differences and inner properties. The *Gaia* RC will probably not be unique as in *Hipparcos* [Ruiz-Dern et al., 2014].

### 1.3.2 The need of a calibration: this thesis

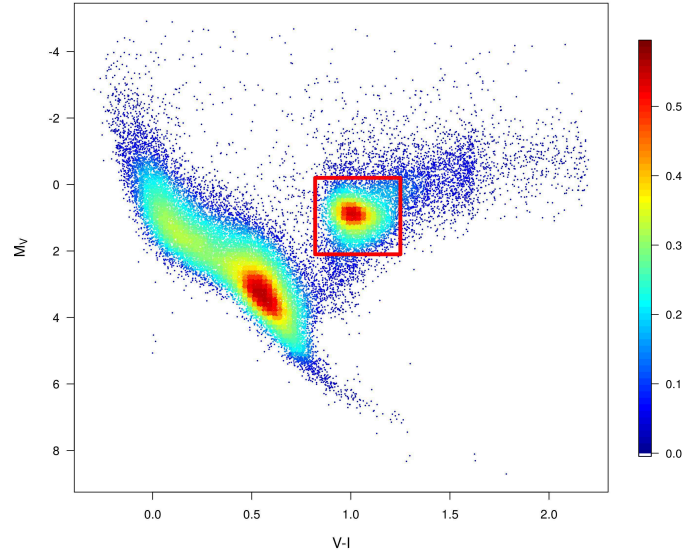
In order to study the RC we selected a sample from the *Hipparcos* catalogue, with relative parallax precision better than 20%, and a solution type  $S_n = 5$  (5-parameter solution, to reject potential double stars). For more details on this sample selection and the data used see later in chapter 2. These stars are located in the red box of fig. 1.26.

We then did a first inspection of the data by looking at different CMDs and colour-colour (hereafter CC) diagrams, in all photometric bands. Since atmosphere models are one of the most important ingredients when simulating the colours of a population of stars, we also checked the agreement between existing stellar models and our observed data. Interestingly we found a non-negligible discrepancy between both regardless of the model (ATLAS, PHOENIX, MARCS). This can be seen in the panels of figure 1.27, where we illustrate the  $B - V$  versus  $V - I$  and the  $J - K$  vs  $V - K$  CC diagrams for our RC stars with different isochrones overplotted. The diagrams on the left include Padova<sup>8</sup> 2.7 version isochrones (ATLAS atmosphere model) [Girardi, 2002, Tang et al., 2014, Chen et al., 2014], while those on the right include Dartmouth<sup>9</sup> isochrones (PHOENIX atmosphere model) [Dotter et al., 2008]. We use isochrones at 2 Gyr and at three different metallicities ( $[Fe/H] = +0.5$  [green],  $[Fe/H] = 0.0$  [red],  $[Fe/H] = -0.5$  [blue]), since according to both the literature and the models from C. Babusiaux, the solar neighbourhood RC is expected to have a mean age of  $1 - 2\text{Gyr}$ , and the predicted median metallicity is about  $[Fe/H] \sim [-0.3, 0.0]$  (for our sample  $[Fe/H] \sim -0.2$ ).

We aim to correct this difference and provide a complete colour-colour calibration for the RC covering all wavelengths from the ultraviolet to the near-infrared. There are several absolute magnitude and effective temperature - colour calibrations in the literature for RC and giants stars [Alonso et al., 1996, Bessell et al., 1998, Keenan and Barnbaum, 1999, Udalski, 1998, Alves, 2000, Ramírez and Meléndez, 2005, Masana

<sup>8</sup>Padova isochrones: <http://stev.oapd.inaf.it/cgi-bin/cmd>

<sup>9</sup>Dartmouth isochrones: <http://stellar.dartmouth.edu/models/>

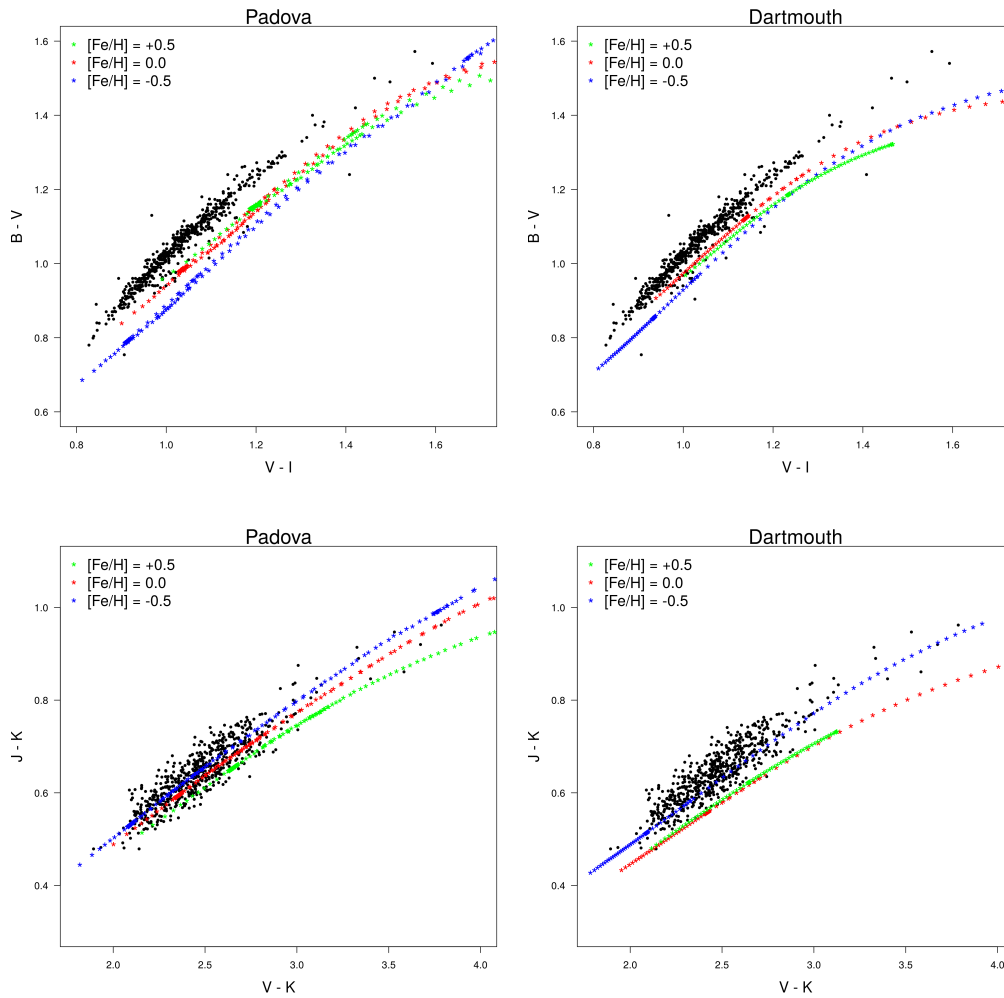


**Figure 1.26** – *Hipparcos* Survey Hertzsprung-Russell diagram, with relative parallax precision  $< 20\%$ . The red box indicated the location of RC stars. The stars are shown as individual points where possible and where the points overlap the relative source density is shown, with colours varying from blue to red indicating increasing density on a logarithmic scale

et al., 2006, González Hernández and Bonifacio, 2009, Casagrande et al., 2010, Bilir et al., 2013a,b, Karaali et al., 2013, Huang et al., 2015] but few for colour-colour [von Braun et al., 1998, Bessell et al., 1998]. Therefore, in this first part of the thesis, I developed methods to calibrate the RC colour-colour relations: in Chapter 2 we describe the first tests performed to obtain a robust calibration by comparing observations with isochrones and the issues encountered; in Chapter 3 we present the final solution adopted, a completely empirical calibration, where we also consider effective temperature - colour relations.

An additional and very important motivation to do this comes from the fact that the First Gaia Data Release will not provide the instrumental response. Therefore, a colour-colour calibration will be automatically needed.

Moreover, CC relations, if accurate, should allow us to infer reddening and other atmospheric parameters. Indeed, in Chapter 4 we take advantage of this and present a method to obtain extinction coefficients and effective temperatures from the previous calibrations. We finally test this tool in different sky regions and populations, especially in the Baade’s Window.



**Figure 1.27** –  $B - V$  vs  $V - I$  and  $J - K$  vs  $V - K$  colour-colour diagrams of 610 RC stars. The median metallicity of the sample is about  $-0.2$ . *Padova* (left) and *Dartmouth* (right) isochrones with a median age of 2Gyr are overlotted for three different metallicities: (green)  $[Fe/H] = +0.5$ , (red)  $[Fe/H] = 0.0$ , (blue)  $[Fe/H] = -0.5$ . Only the Red Giant Branch and the Early Asymptotic Giant Branch are shown

# Red Clump calibration (I): atmosphere models

## Contents

---

<b>2.1 Observations and models</b>	<b>64</b>
<b>2.2 Data selection</b>	<b>64</b>
2.2.1 RC sample	64
2.2.2 Isochrones	68
<b>2.3 Gap correction: MLE method</b>	<b>71</b>
<b>2.4 Results</b>	<b>76</b>
<b>2.5 Discussion</b>	<b>77</b>
2.5.1 Spectral libraries: observed and theoretical	77
2.5.2 Clusters	78

---

## 2.1 Observations and models

Atmosphere models are one of the most important ingredients when simulating the colours of a population of stars. That is, an atmosphere model aims to describe the outer layers of a star from which escapes the optical radiation. To do so, it needs to take into account all physical processes that occur (or may occur) at that level. The model will be more realistic as more detailed assumptions are considered: e.g. local thermodynamic equilibrium (LTE), a plane or spherical atmosphere shape, the expansion of the external layers, the mass loss and the hydrostatic equilibrium, the internal atoms/molecules motions (microturbulence), the convective motions (mixing length), the absorption of spectral lines (opacity),...

All these changes in the atmosphere of the star are logically induced by the changes in its internal structure. Indeed, stellar evolution models (isochrones) provide information about the effective temperature, mass, luminosity,... and, given these parameters, atmosphere models may simulate the associated colour indices of the stars.

Therefore, by comparing observations with isochrones we can directly constrain ages, metallicities, temperatures... However we showed in section 1.3.2 (fig. 1.27) that there is a *gap* between the observed data and stellar evolution models. Hence before using isochrones to derive RC parameters, we first need to model and correct this difference. The evidence of the *gap* is clearer in some colours than in others, but is found in all CC combinations.

In this chapter we describe the method developed to calibrate the RC by simultaneously considering the observed data and the models. In a few words, we aimed to determine a global photometric shift (for each band/colour) to be applied on isochrones in order to correct the *gap* and be able to reproduce the observations. Although, unfortunately, the results obtained were not as good as expected, we highlight all the tests performed to achieve this goal since they justify the adopted final empiric solution (chapter 3).

## 2.2 Data selection

### 2.2.1 RC sample

We selected solar neighbourhood red giant stars from the *Hipparcos* catalogue based on the following magnitude and colour conditions <sup>1</sup>:

$$M_K < 0$$

$$V - I > 0.8$$

set observationally from the *Hipparcos* HR diagram.

Additionally, to perform an accurate photometric calibration, we also considered extra selection criteria in order to ensure the data photometric quality, to reduce as much as possible the reddening and to avoid contamination from variables, binaries and multiple sources. The final sample used in this section was constructed according to the following criteria:

- *UBVIJHK photometry*: we selected those stars which have photometric data in the following surveys:

1. *UBV*: extracted from the UBV photometric system [Johnson and Morgan, 1953] of the General Catalogue of Photometric Data <sup>2</sup> (GCPD) database (by J.C. Mermilliod, B. Hauck, M. Mermilliod)
2. *I*: extracted from the *Hipparcos* V-I colour index

<sup>1</sup>We note that with this lower limit criteria AGB stars are not excluded. However the number of remaining AGB stars in the final sample is very low, and we checked that keeping these stars do not cause any significant impact on the results (contrary to dwarf stars which clearly contaminate the sample)

<sup>2</sup>General Catalogue of Photometric Data (GCPD): <http://obswww.unige.ch/gcpd/gcpd.html>



3. *JHK*: extracted from the *2MASS* catalogue

- *High infrared photometric quality*: the *2MASS* catalogue flags the quality of each photometric band. We selected only those with high quality in all three bands (i.e. flag  $q2M = \text{"AAA"}$ )
- *Parallax precision*: we applied a parallax constraint by selecting those stars with a precision better than 20% ( $|\sigma_{Plx}/Plx| < 0.2$ ) in the *Hipparcos* catalogue (New reduction, van Leeuwen [2007])
- *Variability, binarity and multiplicity*: we removed all stars flagged as variables, binaries (both photometric and spectroscopic) and belonging to multiple systems. To do so we took into account the specific flags in the *Hipparcos* catalogue as well as the last updated information from:
  - *9th Catalogue of Spectroscopic Binary Orbits*<sup>3</sup> (SB9) [Pourbaix et al., 2004, 2009]
  - *Catalogue of the Components of Doubles and Multiple Stars*<sup>4</sup> (CCDM) [Dommanget and Nys, 1994, 2002]
  - *The Washington Visual Double Star Catalog*<sup>5</sup> (WDS) [Mason et al., 2001]
- *Low latitudes*: to minimise reddening effects we removed stars from the Galactic plane ( $|b| < 10^\circ$ )
- *Extinction*: for the remaining stars we calculated their extinctions (see below) by using Arenou et al. [1992] tridimensional model and Lallement et al. [2014] 3D charts, and we removed those stars with  $A_V > 0.5$

Applying all these criteria, our sample contains 610 stars, 54 of them with metallicity information.

Metallicity data was extracted from the *Extended Hipparcos Compilation* (XHIP) [Anderson and Francis, 2012a,b] which mainly comes from the *PASTEL catalogue of stellar atmospheric parameters* [Soubiran et al., 2010a,b] and other spectroscopic sources (see Anderson and Francis [2012a] for more details about the compilation). We also added some metallicities found in Zieliński et al. [2012a,b]<sup>6</sup>.

Figures 2.1, 2.2, 2.3 show, respectively: the position of the sample on the *Hipparcos* HR diagram, their location on an Aitoff map projection in galactic coordinates, and the colour and spectral type distributions. Green points or regions belong to stars with known metallicity. Figure 2.4 displays the metallicity distribution of these stars.

### Interstellar extinction

To remove high extinction stars we used the Arenou et al. [1992] tridimensional model and the 3D chart from Lallement et al. [2014]. Both provide full sky extinction distributions within 1kpc from the Sun.

Arenou et al. [1992] modelled the whole sky extinction distribution as a function of the distance (up to 1kpc) and the galactic coordinates, by using spectral and photometric data.

$$A_V(r, b, l) = \alpha(l, b) r + \beta(l, b) r^2 \quad \text{if } r \leq r_0$$

$$A_V(r, b, l) = A_V(r_0, b, l) + \gamma(l, b) (r - r_0) \quad \text{if } r \geq r_0$$

with  $r_0$  representing the distance limit of the absorbing layers. If  $A_V < 0$ ,  $A_V = 0$  is adopted. The associated error was estimated as:

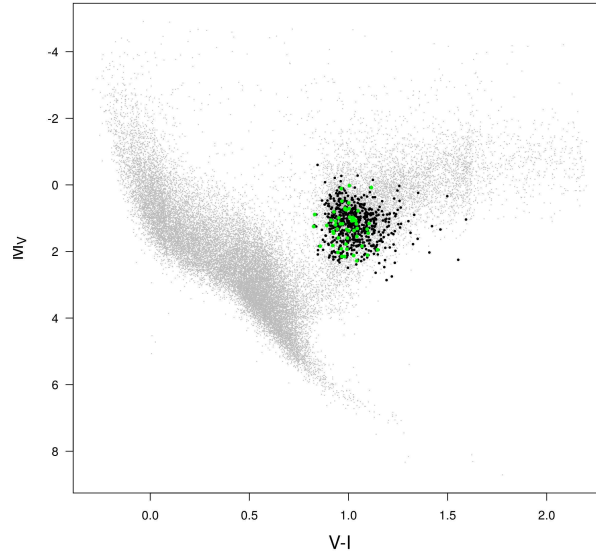
$$\sigma_{A_V} = \sqrt{0.15^2 + \left( \frac{\sigma_{A_V}}{A_V} A_V(r) \right)^2}$$

<sup>3</sup>SB9 VizieR reference: B/sb9 (<http://vizier.u-strasbg.fr/viz-bin/VizieR?-source=B/sb9>)

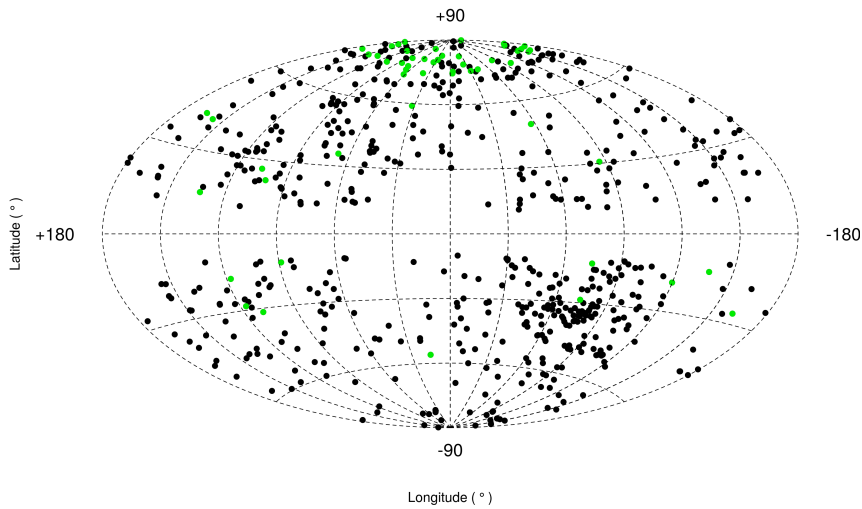
<sup>4</sup>CCDM VizieR reference: I/274/ccdm (<http://vizier.u-strasbg.fr/viz-bin/VizieR-3?-source=I/274/ccdm>)

<sup>5</sup>WDS VizieR reference: B/wds (<http://vizier.u-strasbg.fr/viz-bin/VizieR-3?-source=B/wds>)

<sup>6</sup>348 red giants VizieR reference: J/A+A/547/A91/stars (<http://vizier.u-strasbg.fr/viz-bin/VizieR-3?-source=J/A%2bA/547/A91>)



**Figure 2.1** – *Hipparcos* HR diagram with selected RC stars overlotted in black. Green points correspond to stars for which we have metallicity information



**Figure 2.2** – Distribution of the RC sample on an Aitoff map projection in galactic coordinates. Green points correspond to stars with known metallicity

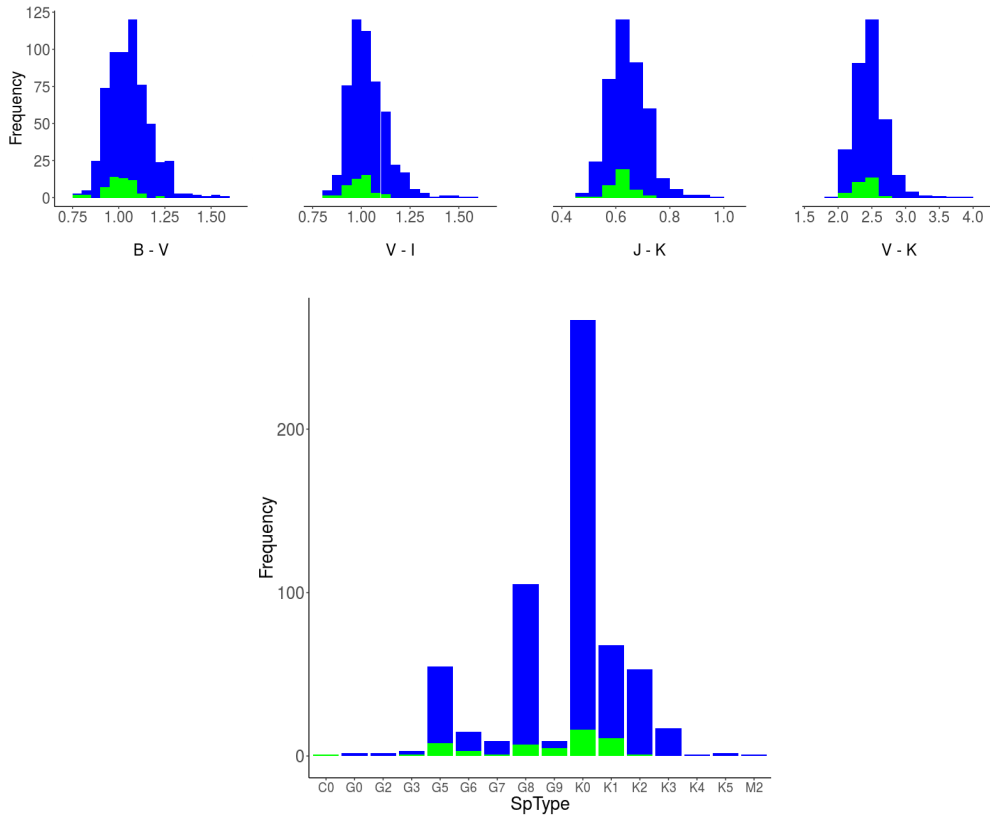
where 0.15 mag is the assumed uncertainty for  $A_V$  and the factor  $\frac{\sigma_{A_V}}{A_V}$  is given in the coefficient tables (available in their article) together with the values of  $\alpha$ ,  $\beta$ ,  $\gamma$  and  $r_0$ .

Lallement et al. [2014] developed a 3D local interstellar extinction map, within 1kpc from the Sun. The mapping is based on 23.000 optical colour excess measurements ( $E_{(B-V)} = A_B - A_V$ ) and *Hipparcos* distances, and it uses an inversion method [Tarantola and Valette, 1982] based on a regularized Bayesian approach.

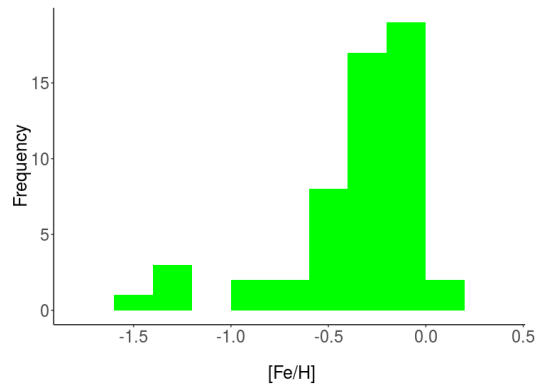
The mapped colour excesses for our stars were directly provided to us by the author. The extinction was immediately obtained through the relation:

$$A_V = 3.1 E_{(B-V)}$$

All stars with  $A_V > 0.5$  in at least one of the models were removed from the initial sample. Fig. 2.5



**Figure 2.3** – Distributions of colours (top) and spectral types for classes III and IV (bottom) of the RC sample. Overlapped in green is the distribution of stars with known metallicity

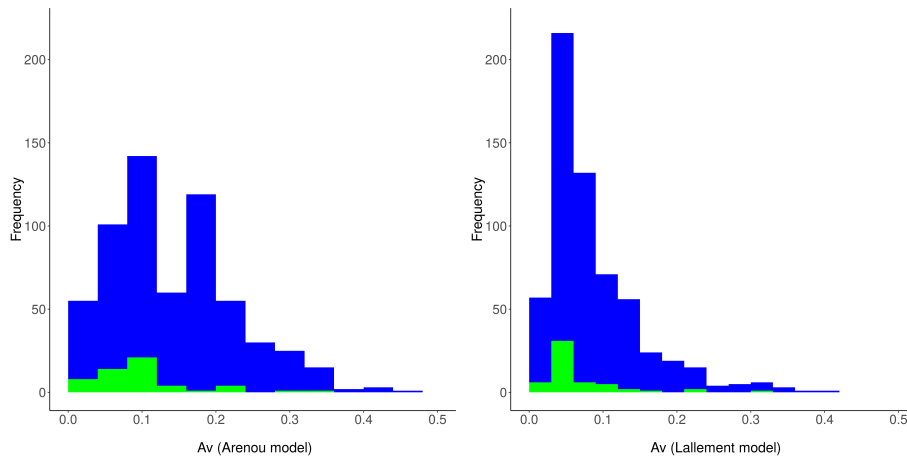


**Figure 2.4** – Metallicity distribution of the RC sample

shows the extinction distribution of the remained stars for each model.

In what follows we used the extinction from [Lallement et al. \[2014\]](#) (more recent model). The errors were calculated by using the left and right  $\sigma_{E_{(B-V)}}$  provided by the author and an arbitrary systematic of 0.05:

$$\sigma_{A_V} = \sqrt{\left[ \frac{3.1 \left( \sigma_{E_{(B-V)}}^+ + \sigma_{E_{(B-V)}}^- \right)}{2} \right]^2 + 0.05^2} \quad (2.1)$$



**Figure 2.5** – Extinction distribution of the RC sample for the [Arenou et al. \[1992\]](#) (left) and [Lallement et al. \[2014\]](#) (right) extinction maps. Overlapped in green is the distribution for stars with known metallicity

### 2.2.2 Isochrones

To highlight that the *gap* problem is a generalized issue regardless of the theoretical model used or the mean values of some physical properties of the stars (age, metallicity, mass), we checked it with two different isochrone models:

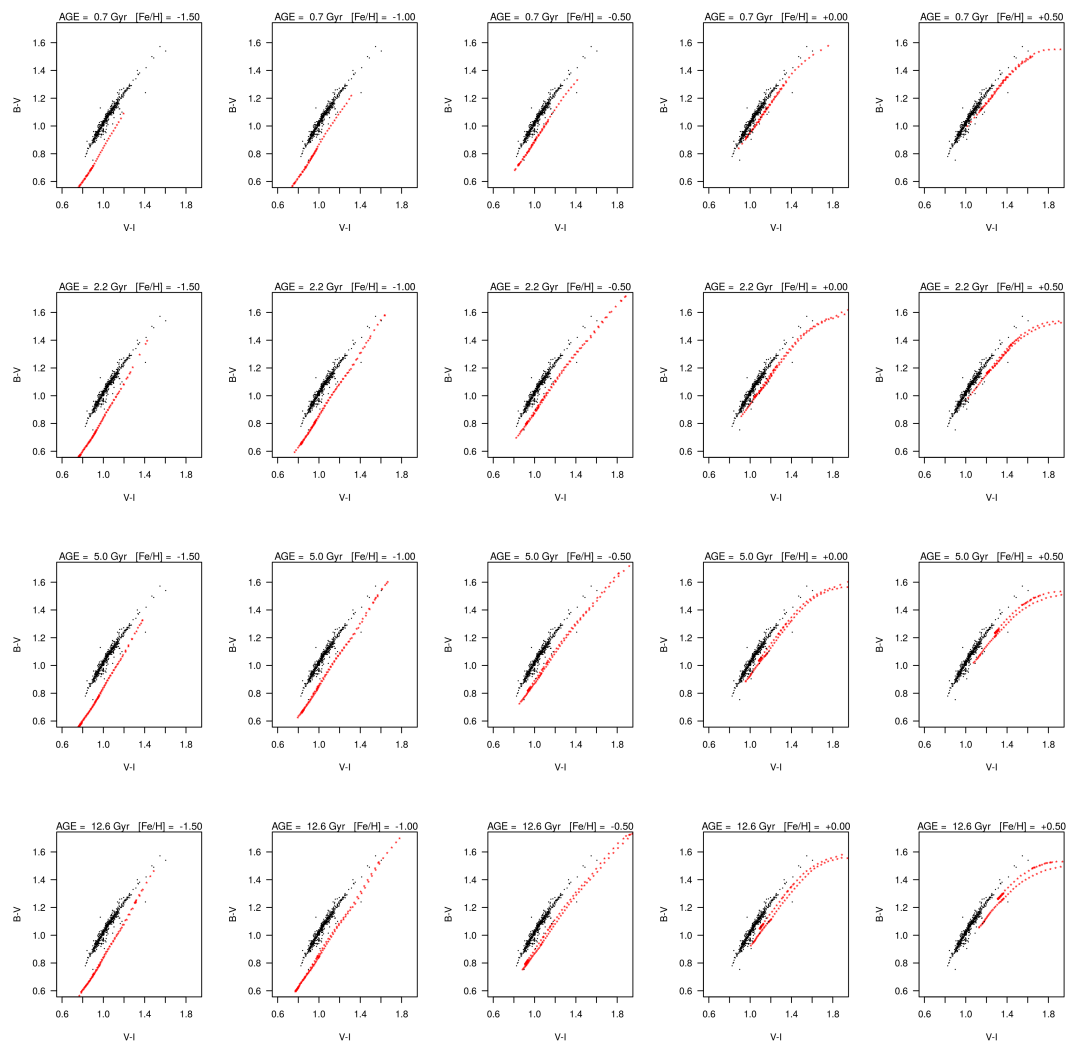
- *Padova* (2.7 version), which uses ATLAS<sup>7</sup> atmosphere model for  $T_{\text{eff}} > 4000\text{K}$
- *Dartmouth*, which uses PHOENIX<sup>8</sup> atmosphere model

Fig. 2.6 and 2.7 show the evolution of both *Padova* and *Dartmouth* isochrones, respectively, with age and metallicity. For *Padova* we used isochrones associated with the red and asymptotic giant branches (RGB and AGB), which correspond to the "stages" 3 to 7 of their tables. To do the same cut for *Dartmouth* isochrones, we took the minimum and maximum  $\log g$  of each *Padova* isochrone used (at a given metallicity and age), and we selected the same  $\log g$  range for *Dartmouth*. The figures present a set of  $B - V$  vs  $V - I$  CC diagrams. Each panel includes a different age-metallicity isochrone, with the ages increasing from top to bottom panels ( $[0.7, 12]$  Gyr), and metallicities from left to right panels ( $[\text{Fe}/\text{H}] = [-1.50, 0.5]$ ). The RC sample is overplotted to highlight the fact that the mentioned *gap* is neither caused by the chosen isochrone model, nor by a bias on the RC median metallicity or age.

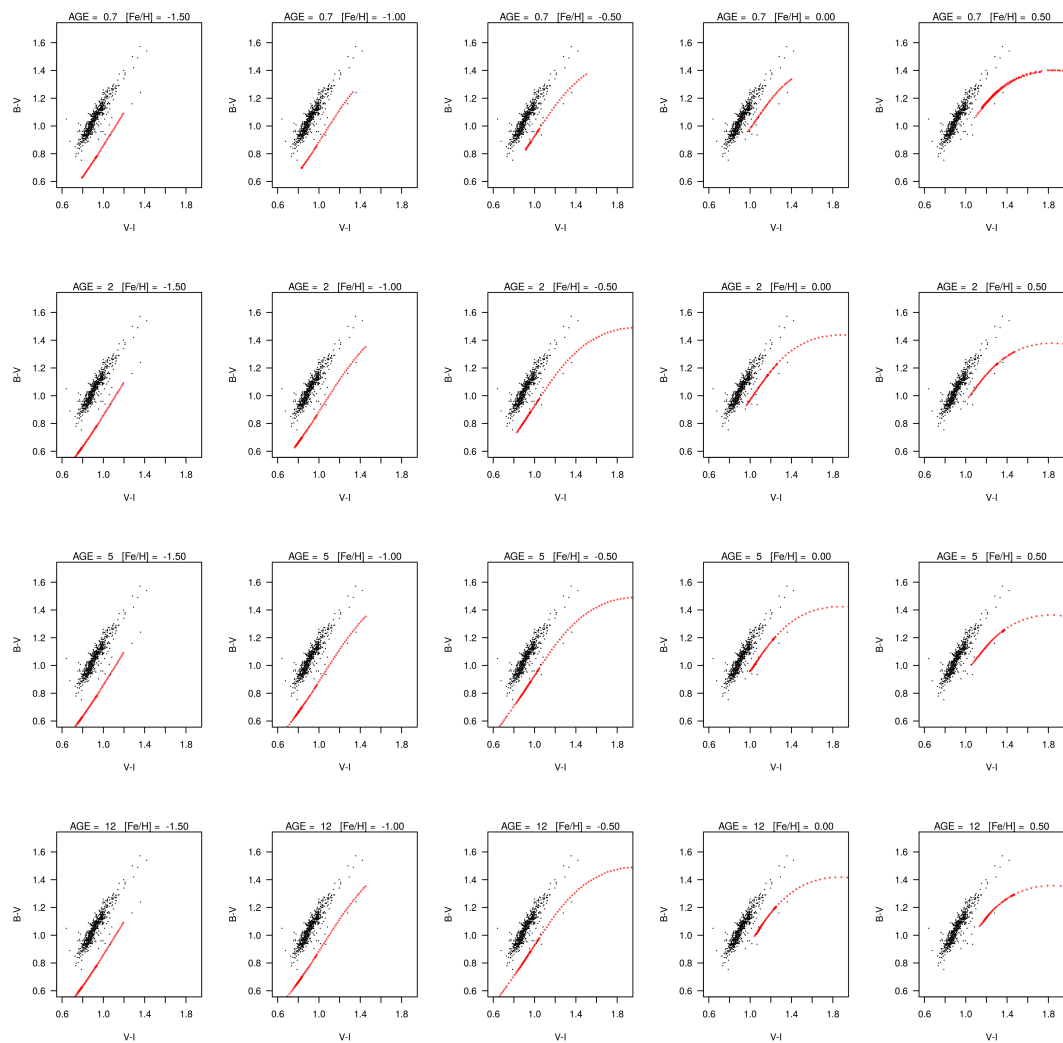
The first calibration tests were performed on both types of isochrones. However, given that the results were similar regardless of the theoretical model used, we only kept *Padova* for the complete implementation. The advantage of using these isochrones is that they include a large number of photometric filters available, the evolutionary stages are directly provided (no need to cut in  $\log g$ ), and they are periodically updated. We used 2.7 version, although at the time of writing this thesis version 2.8 is already available. We considered the default parameters, with an Initial Mass Function (IMF) from [Chabrier \[2001\]](#) lognormal, and only the stages 3 (Red Giant Branch), 4, 5, 6 (belonging to core *He* burning phases), and 7 (Asymptotic Giant Branch).

<sup>7</sup>ATLAS: [http://www.stsci.edu/hst/observatory/crds/castelli\\_kurucz\\_atlas.html](http://www.stsci.edu/hst/observatory/crds/castelli_kurucz_atlas.html)

<sup>8</sup>PHOENIX: [http://www.hs.uni-hamburg.de/index.php?option=com\\_content&view=article&id=14&Itemid=270&lang=de](http://www.hs.uni-hamburg.de/index.php?option=com_content&view=article&id=14&Itemid=270&lang=de)



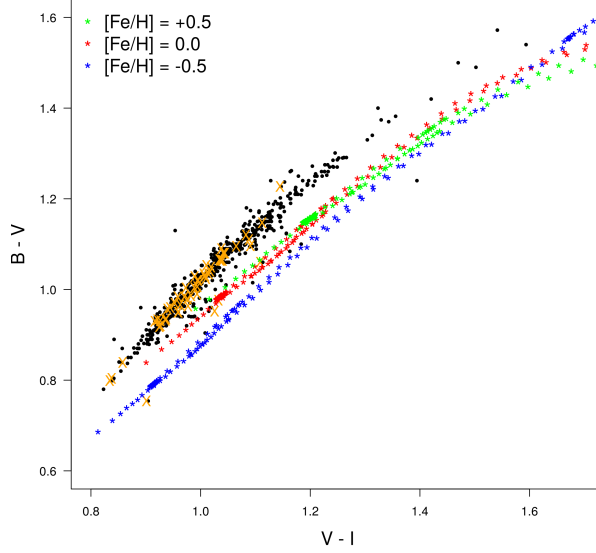
**Figure 2.6** – Evolution of *Padova* isochrones by age (panels from top to bottom) and metallicity (panels from left to right) in a  $B - V$  vs  $V - I$  CC diagram. Our RC sample is also plotted (black points). These graphs show that changing age and metallicity slightly varies the width of the *gap* but does not eliminate the *gap*



**Figure 2.7** – Evolution of *Dartmouth* isochrones by age (panels from top to bottom) and metallicity (panels from left to right) in a  $B - V$  vs  $V - I$  CC diagram. Our RC sample is also plotted (black points). These graphs show that changing age and metallicity slightly varies the width of the *gap* but does not eliminate the *gap*

## 2.3 Gap correction: MLE method

To determine the shifts to correct the *gap* between observations and models, we performed a Maximum Likelihood Estimator (MLE) method. Figure 2.8 displays again the *gap* between observations and three different metallicity isochrones. In this case, the stars used in the MLE method (those for which we had metallicity information) are highlighted in orange.



**Figure 2.8** –  $B - V$  vs  $V - I$  diagram of the 610 RC stars of our sample. Orange crosses correspond to the 54 stars with metallicity information, those used in the MLE method. Green, red and blue points belong to isochrones for  $[\text{Fe}/\text{H}] = 0.5$ ,  $[\text{Fe}/\text{H}] = 0.0$  and  $[\text{Fe}/\text{H}] = -0.5$ , respectively. The double curve one may see for each isochrone belongs to the Red Giant Branch and the Asymptotic Giant Branch

The method aims to derive unique shifts for each individual magnitude. Thus, it provides a correction for every photometric band ( $\text{shift}_B$ ,  $\text{shift}_V$ ,  $\text{shift}_K$ , ...) from which we can immediately calculate whichever colour pair (e.g.  $B - V_{\text{corrected}} = (B + \text{shift}_B) - (V + \text{shift}_V)$ ).

The MLE allows to estimate the parameters  $\Theta$  that maximizes  $P(\Theta|\tilde{x}_1, \dots, \tilde{x}_n)$ . Taking advantage of the Bayes' theorem and given that here the parameters to be estimated are the shifts for each photometric magnitude  $\{s\}$ , and that the observables (indicated with a  $\sim$ ) are the observed apparent magnitudes  $\{\tilde{x}\}$ , the parallax  $\tilde{\omega}$  and the extinction  $\tilde{A}_0$ , the probability function  $P(\{s\}|\{\tilde{x}\}, \tilde{\omega}, \tilde{A}_0)$  to maximize is:

$$P(\Theta|\tilde{x}_1, \dots, \tilde{x}_n) = \frac{f(\tilde{x}_1, \dots, \tilde{x}_n|\Theta)P(\Theta)}{P(\tilde{x}_1, \dots, \tilde{x}_n)} = \frac{P(\{\tilde{x}\}, \tilde{\omega}, \tilde{A}_0|\{s\})P(\{s\})}{P(\{\tilde{x}\}, \tilde{\omega}, \tilde{A}_0)} \quad (2.2)$$

$$\propto P(\{\tilde{x}\}, \tilde{\omega}, \tilde{A}_0|\{s\})P(\{s\}) \propto P(\{\tilde{x}\}, \tilde{\omega}, \tilde{A}_0|\{s\}) \quad (2.3)$$

where we consider  $P(\{s\})$  constant since we do not have any prior information of the shifts.

Given that the magnitudes are measured independently, the above expression can be written as the product of probabilities for each band:

$$P(\{s\}|\{\tilde{x}\}, \tilde{\omega}, \tilde{A}_0) \propto \prod_{\lambda} P(\tilde{x}_{\lambda}, \tilde{\omega}, \tilde{A}_0|s_{\lambda})$$

And since we are taking theoretical models into account, we must sum over all possible isochrones. To do so we apply:

$$P(A) = \int_B P(A, B) = \int_B P(A|B) P(B)$$

Thus,

$$P(\{s\}|\{\tilde{x}\}, \tilde{\omega}, \tilde{A}_0) \propto \prod_{\lambda} \int_i P(\tilde{x}_{\lambda}, \tilde{\omega}, \tilde{A}_0 | i, s_{\lambda}) P(i) \quad (2.4)$$

with  $i$  the isochrone point and  $P(i)$  the weight of each point of the *Padova* isochrones. Indeed, each isochrone point  $i$ , corresponding to a metallicity  $[M/H]_i$ , age  $\tau_i$  and mass  $\mathcal{M}_i$ , has a weight according to the IMF  $\xi(\mathcal{M})$  and SFR  $\psi(\tau)$ . We used here the [Chabrier \[2001\]](#) lognormal IMF (integrated over the mass interval between isochrone points) and a constant SFR (considering that we have a grid sampled in logAge, which means that the SFR associated weight is proportional to the age), and we did not introduce any age-metallicity correlation.

If we sum over all possible parallaxes:

$$P(\{s\}|\{\tilde{x}\}, \tilde{\omega}, \tilde{A}_0) \propto \prod_{\lambda} \int_i \int_{\varpi} P(\tilde{x}_{\lambda}, \tilde{\omega}, \tilde{A}_0 | i, s_{\lambda}, \varpi) P(i) P(\varpi) \quad (2.5)$$

where we consider  $P(\varpi)$  uniform.

Similarly, summing over all possible extinctions:

$$P(\{s\}|\{\tilde{x}\}, \tilde{\omega}, \tilde{A}_0) \propto \prod_{\lambda} \int_i \int_{\varpi} \int_{A_0} P(\tilde{x}_{\lambda}, \tilde{\omega}, \tilde{A}_0 | i, s_{\lambda}, \varpi, A_0) P(i) P(\varpi) P(A_0) \quad (2.6)$$

with  $P(A_0)$  uniform but null for negative values. The input interstellar extinctions provided by R. Lallement are already constrained to be always positive.

Moreover, the magnitudes, the parallaxes and the extinction are independent measurements, hence we may separate the above probability function  $P(\tilde{x}_{\lambda}, \tilde{\omega}, \tilde{A}_0 | i, s_{\lambda}, \varpi, A_0)$  in a product of three probability functions. The magnitude depends on all variables, while the observed parallax is only dependent on the true parallax and observed extinction only on the true extinction. Else, we know that the theoretic/expected value of the magnitude  $x_{\lambda}$  is a function of  $i$ ,  $\varpi$ ,  $A_0$  and the shift, thus:

$$P(\{s\}|\{\tilde{x}\}, \varpi, A_0) \propto \prod_{\lambda} \int_i \int_{\varpi} \int_{A_0} P(\tilde{x}_{\lambda} | x_{\lambda}) P(\tilde{\omega} | \varpi) P(\tilde{A}_0 | A_0) P(i) P(\varpi) P(A_0) \quad (2.7)$$

with

$$x_{\lambda} = M_i + k_{x_{\lambda}} A_0 + 5 \log_{10} \left( \frac{\varpi}{1000} \right) + 5 + s_{\lambda}$$

where  $M_i$  is the absolute magnitude of the isochrone point  $i$ , and  $k_{x_{\lambda}}$  are the extinction coefficients calculated using Kurucz atmosphere models. Indeed, here we are deriving  $A_0$  which is the absorption at a given wavelength (we consider  $\lambda = 550nm$ , which is the Gaia reference value). And  $A_0$  is related to the extinction on any other band ( $A_{\lambda}$ ) and the extinction coefficient ( $k_{\lambda}$ ) through the general extinction law:

$$k_{\lambda} = \frac{A_{\lambda}}{A_0} \quad (2.8)$$

Finally, given these estimates and the fact that the parameters are normally distributed, we can rewrite the probability function as follows (eq. 2.9):



$$P(\{s\}|\{\tilde{x}\}, \tilde{\omega}, \tilde{A}_0) \propto \prod_{\lambda} \int_i \int_{\tilde{\omega}} \int_{A_0} \mathcal{N}(\tilde{x}_{\lambda}, x_{\lambda}, \sigma_{\tilde{x}_{\lambda}}) \mathcal{N}(\tilde{\omega}, \omega, \sigma_{\tilde{\omega}}) \ln \mathcal{N}(A_0, \ln(\tilde{A}_0), \sigma_{\tilde{A}_0}) P(i) P(\tilde{\omega}) P(A_0) \quad (2.9)$$

We used logarithmic normal distribution for extinctions because, as explained at the end of section 2.2.1,  $\sigma_{\tilde{A}_0}$  was roughly estimated and the log-normal error model is consistent with the fact that the provided  $A_0$  has already been constrained to be positive.

## Implementation

Figure 2.10 shows a schema of the implemented process. Briefly, for each of the 54 stars with metallicity:

### 1. Select input:

- (a) complete photometric information ( $U, B, V, I, J, H, K$ ), parallax, metallicity and extinction
- (b) the closest 2 Gyr isochrone to each star, in terms of metallicity (2 Gyr being the typical age for a RC star)

2. **Determine photometric errors:** input errors for each band and star are taken into account. If no error is available for a given band (single observations in the General Catalogue of Photometric Data, Mermilliod), or whether it is lower than 0.01 (the estimated calibration error floor), the mean error of all stars with  $\sigma_{band} > 0.01$  is taken

3. **Calculate extinction coefficients:** the intrinsic colour of a star is the observed colour without the effects of the extinction. Thus, given two magnitudes  $m_1$  and  $m_2$ :

$$(m_1 - m_2)_{intrinsic} = (m_1 - m_2)_{observed} - E(m_1 - m_2)$$

where the colour excess  $E(m_1 - m_2) = A_{m_1} - A_{m_2}$ , with  $A_{m_1}$  and  $A_{m_2}$  the extinction on each band.

Thus, given the extinction law of eq. 2.8, the relation between the intrinsic and observed colours can be written as:

$$(m_1 - m_2)_{intrinsic} = (m_1 - m_2)_{observed} - (k_{m_1} - k_{m_2}) A_0 \quad (2.10)$$

with  $k_{m_1}$  and  $k_{m_2}$  the extinction coefficients for each band (different from  $k_{\lambda}$  which, as stated above, is the coefficient at a given wavelength).

As explained in Fitzpatrick [1999] and Fitzpatrick and Massa [2007] the shape of the extinction law varies from the UV to the IR regions (fig. 2.9). Indeed,

$$A_{\lambda} = m_{\lambda} - m_{\lambda}^0 = -2.5 \log_{10} \left( \frac{F_{\lambda}}{F_{\lambda}^0} \right)$$

So,

$$F_{\lambda} = F_{\lambda}^0 10^{-0.4 A_{\lambda}} = F_{\lambda}^0 10^{-0.4 k_{\lambda} A_0}$$

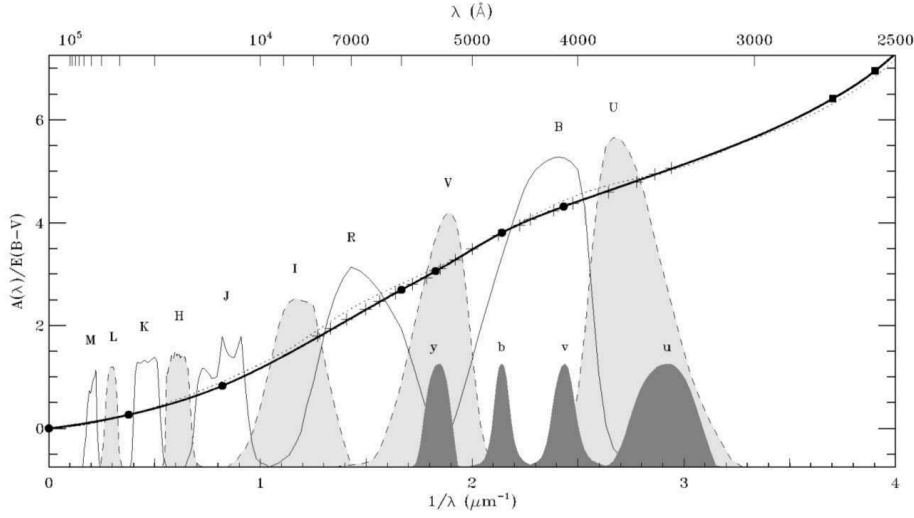
The magnitude in a photometric band  $m$  is:  $m = -2.5 \log_{10} (\int F_{\lambda} T_{\lambda} d\lambda) + m_0$

With  $T_{\lambda}$  the instrumental transmission (including CCD quantum efficiency, filter transmission, mirror coating,...)

$$m - m_0 = -2.5 \log_{10} \left( \frac{F}{F^0} \right) = -2.5 \log_{10} \left( \frac{\int F_\lambda T_\lambda E_\lambda^{A_0} d\lambda}{\int F_\lambda T_\lambda d\lambda} \right) \quad (2.11)$$

where [Fitzpatrick and Massa \[2007\]](#) used  $E_\lambda = 10^{-0.4 k_\lambda}$

By using the simulator ASPOS<sup>9</sup> of P. Sartoretti (private communication), which includes the [Fitzpatrick and Massa \[2007\]](#) extinction law, and the Kurucz spectra for a typical RC star ( $T_{\text{eff}} = 4750$ ,  $\log g = 2.5$ ,  $[\text{Fe}/\text{H}] = 0.0$ ), we obtained a list of  $A_\lambda$  values corresponding to an  $A_0$  that varies from 0.01 to 20.0, by increments of 0.01.



**Figure 2.9** – Wavelength dependence of extinction in the IR/optical region for  $R = 3.1$  (thick solid curve). The arbitrarily scaled profiles of the Johnson UBVRIJHKLM and Strömgren uvby filters are shown for comparison. The curve was constrained to reproduce the broadband and intermediate-band filter-based extinction measurements listed in [Fitzpatrick \[1999\]](#) table 2 and Baastiansen (1992) data. For  $\lambda > 2700\text{\AA}$  ( $1/\lambda < 3.7\mu\text{m}^{-1}$ ) the curve is constructed as a cubic spline curve interpolation between the points marked by the filled symbols. At wavelengths shortward of  $2700\text{\AA}$ , the curve is computed using the Fitzpatrick & Massa 1986, 1988 and 1990 fitting function. Credits from [Fitzpatrick \[1999\]](#) (figure 6 of the article)

To finally get the extinction coefficients  $k_\lambda$ , we interpolate between  $A_\lambda$  and  $A_0$  given by ASPOS. The interpolation is evaluated at each  $A_V$  obtained with [Lallement et al. \[2014\]](#) 3D maps. From the interpolated  $A_{\lambda,interp}$  and  $A_{0,interp}$ , and using eq. 2.8, we get  $k_\lambda$  for each photometric band of each star.

4. **Compute probability function and run MLE:** R m1e2 library is used to derive the MLE of the sum of probability functions. To speed up the computation we integrate over the parallax and the extinction only within their  $4\sigma$  confidence interval. As required by the method, we also set initial values for the shifts:

- **Parallaxes:**

$$\bar{\varpi}_{\text{interval}} \in [\text{Max}(0, \bar{\varpi} - 4 \sigma_{\bar{\varpi}}), \bar{\varpi} + 4 \sigma_{\bar{\varpi}}] \quad (2.12)$$

discretized by steps of  $\sigma_{\bar{\varpi}}/2$

- **Extinctions<sup>10</sup>:**

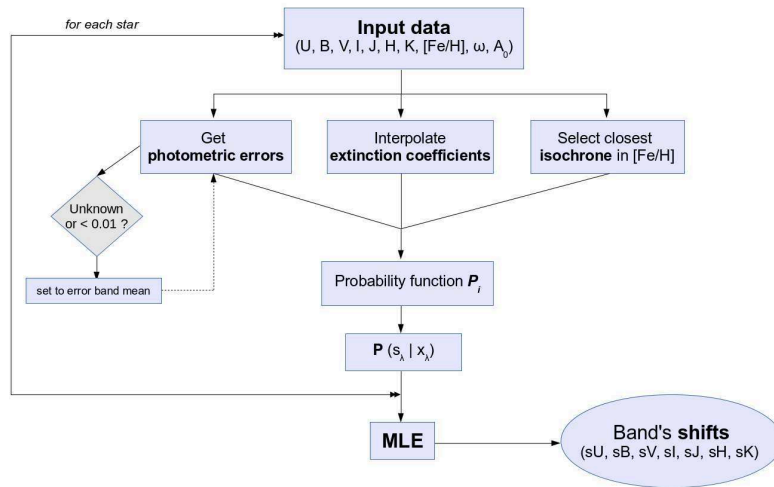
$$A_{0\text{interval}} \in [\text{Max}(0, \tilde{A}_0 - 0.4), \tilde{A}_0 + 0.4] \quad (2.13)$$

<sup>9</sup>Astronomic SPectra Observation Simulation

<sup>10</sup>The input  $\tilde{A}_V$  is assumed as  $\tilde{A}_0$

discretized by steps of 0.01

- **Initial values:** all shifts are initially set to 0



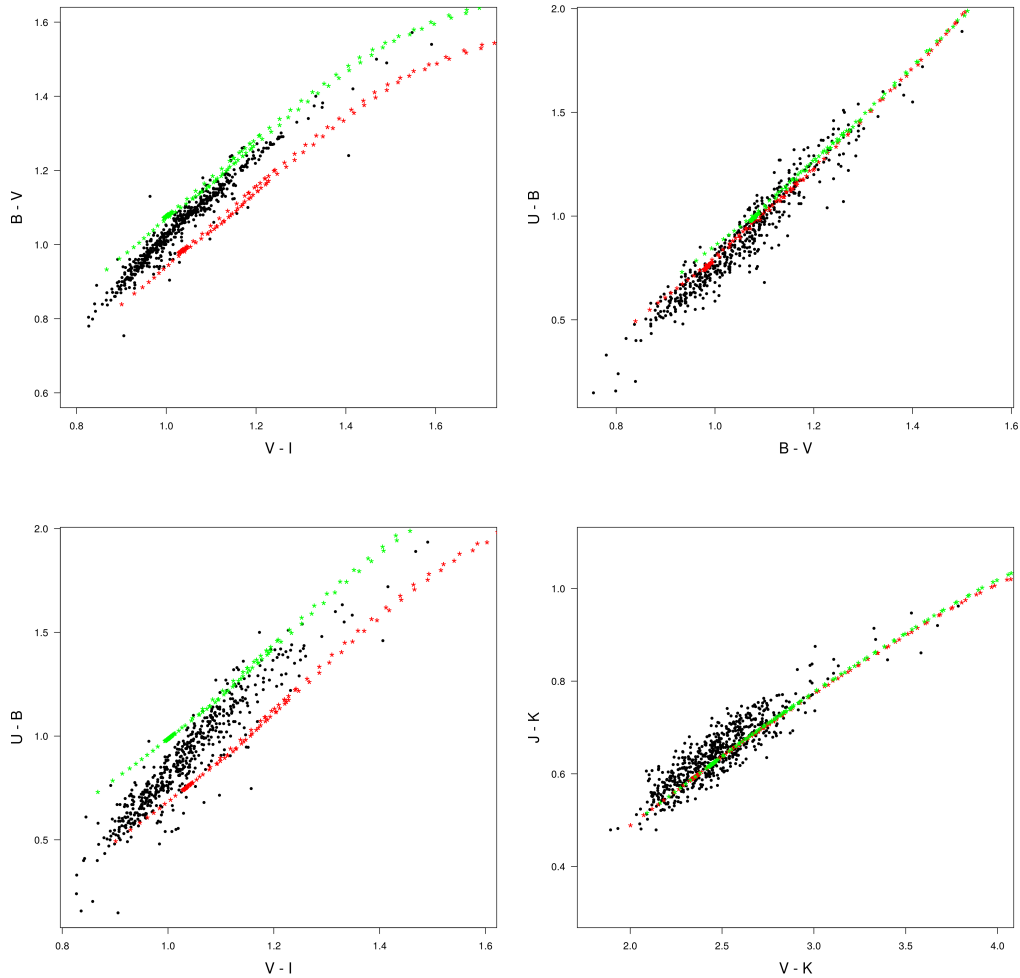
**Figure 2.10** – Schema of the MLE method performed to calibrate the RC taking atmosphere models into account

## 2.4 Results

The four panels in figure 2.11 show different CMDs displaying the 610 stars of our RC sample, the original isochrone (in red) and the shifted solution (green) obtained with the MLE method described above and using the 54 stars with metallicity data. Table 2.1 gives the shifts for each band.

**Table 2.1** – Shifts obtained for each band through the MLE *gap* correction method

Band	Shift
U	$0.354 \pm 0.004$
B	$0.117 \pm 0.004$
V	$0.023 \pm 0.003$
I	$0.057 \pm 0.003$
J	$-0.040 \pm 0.004$
H	$-0.029 \pm 0.004$
K	$-0.067 \pm 0.004$



**Figure 2.11** – CC diagrams showing the results of the MLE *gap* correction method. Black points correspond to the RC sample, and red and green dotted lines belong to the original and shifted isochrones ( $[Fe/H] = 0.0$ ,  $Age \sim 2Gyr$ ), respectively

## 2.5 Discussion

The results obtained clearly show that the method did not converge to the expected solution: the observations cannot be reproduced using atmosphere models. We also tried to consider ages as an extra constrain within the MLE method, to check if they caused some effect in the shift correction but they did not. We remind that although only results using ATLAS models have been presented, the tests were also performed with PHOENIX atmosphere models with similar results.

Two main reasons were therefore considered to justify why the observations are not reproduced when using a method that accounts for theoretical models:

- Filter modelling issues: according to Mann and von Braun [2014] using inaccurate filter profiles from the literature may create significant colour terms resulting in fluxes that deviate by  $\geq 10\%$  from actual values. At the time of getting our results, the mentioned authors proposed an empirical approach to improve the zero point accuracy to 0.5-1% or better: they iteratively adjust the literature filter profiles and zero points, convolve them with catalogue spectra, and compare them to the corresponding fluxes from the photometry (see the article for more details)
- Atmosphere models do not allow us to reproduce observations at the RC level

To check these two aspects and in order to try to understand which was the main problem, we performed two tests using spectral libraries (section 2.5.1) and open clusters (section 2.5.2).

### 2.5.1 Spectral libraries: observed and theoretical

In order to check whether the difference we found with RC stars between models and observations is generalized to giants, we selected different giant's spectra from observed and theoretical spectral libraries: Pickles<sup>11</sup>, Atlas9<sup>12</sup>, Calspec<sup>13</sup> and NGSL<sup>14</sup>. At the same time, P. Sartoretti updated ASPOS simulator to account for the revised filter profiles and zero points of Mann and von Braun [2014]. This simulator was then used to derive the magnitudes of all possible bands for the selected giant's spectra.

Table 2.2 summarizes the information used from each spectral library: observed/theoretical, units, number of selected giant stars' spectra, their spectral types and the wavelength coverage.

**Table 2.2** – Spectral libraries information and data of available giant spectra

Library	Type	Flux units	Giant stars	Spectral Types	Wavelength coverage
Pickles	Observed	$erg/cm^2/s/\text{\AA}$	9	G0III - K5III	IR/Optical
Atlas9	Theoretical (Kurucz model)	$erg/cm^2/s/\text{\AA}$	9	G0III - K5III	IR/Optical
Calspec	Observed	$erg/cm^2/s/\text{\AA}$	1	K1.5III	IR/Optical
NGSL	Observed	$erg/cm^2/s/\text{\AA}$	379	G2III - K5III	Optical

Figures 2.12, 2.13, 2.14 and 2.15 show the RC sample with the original and shifted isochrones as in fig. 2.11, but with the colours obtained for the considered spectra of each spectral library.

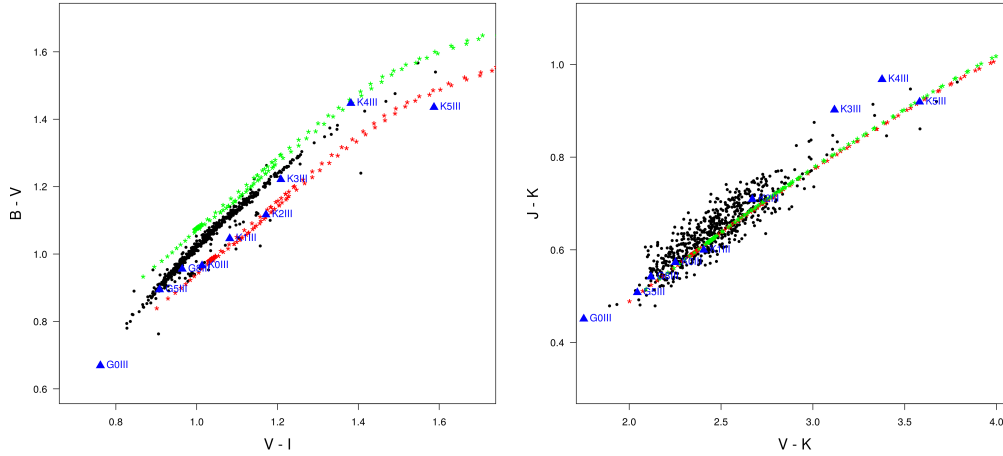
While doing these tests we verified that the colours derived from spectra were quite dependent on filter modelling. Using the revised filter profiles and zero points of Mann and von Braun [2014] the resulting colours from the observed spectra were improved. For Calspec star (fig. 2.14) the difference is due to high extinction (thus not explicitly used to conclude the problem between isochrones and observations).

<sup>11</sup>Pickles: <http://www.eso.org/sci/facilities/paranal/decommissioned/isaac/tools/lib.html>

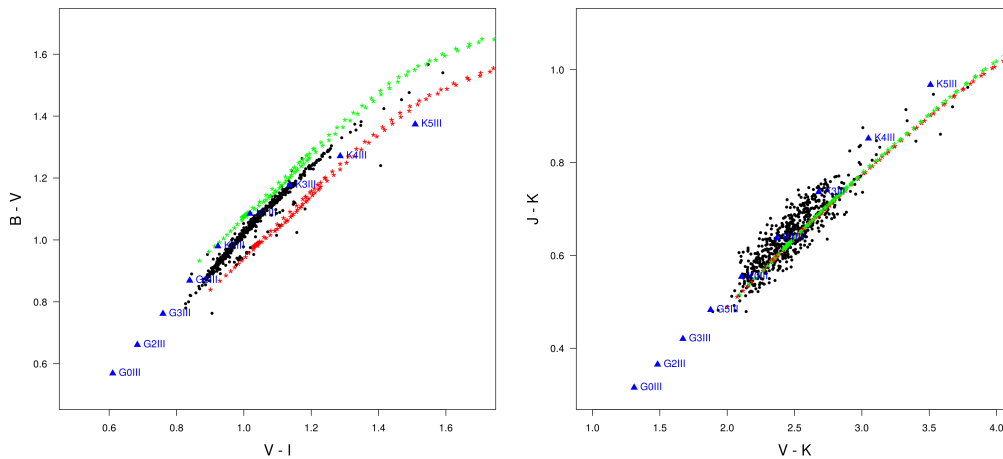
<sup>12</sup>Atlas9: [http://www.stsci.edu/hst/observatory/crds/castelli\\_kurucz\\_atlas.html](http://www.stsci.edu/hst/observatory/crds/castelli_kurucz_atlas.html)

<sup>13</sup>Calspec: [www.stsci.edu/hst/observatory/crds/calspec.html](http://www.stsci.edu/hst/observatory/crds/calspec.html)

<sup>14</sup>NGSL (Next Generation Spectral Library): <http://archive.stsci.edu/prepds/stisngsl/index.html>



**Figure 2.12** –  $B - V$  vs  $V - I$  and  $J - K$  vs  $V - K$  CC diagrams with the RC sample (black points), the original and shifted isochrones (red and green dotted lines, respectively) at  $[\text{Fe}/\text{H}] = 0.0$  and  $\text{Age} \sim 2\text{Gyr}$ , and the colours obtained with ASPOS simulator for the observed Pickles giant’s spectra (blue triangles). The spectral type of each Pickles’ star is given

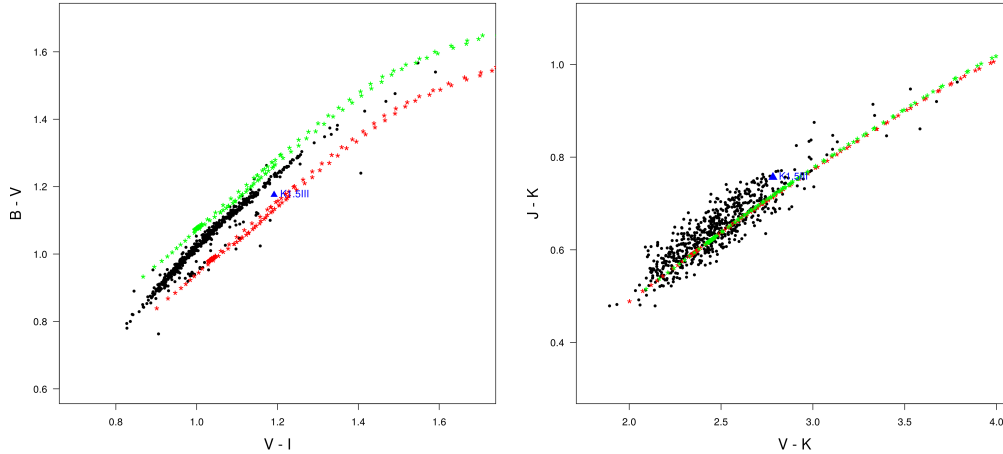


**Figure 2.13** – Same as in previous figure but with the theoretical Atlas giant’s spectra

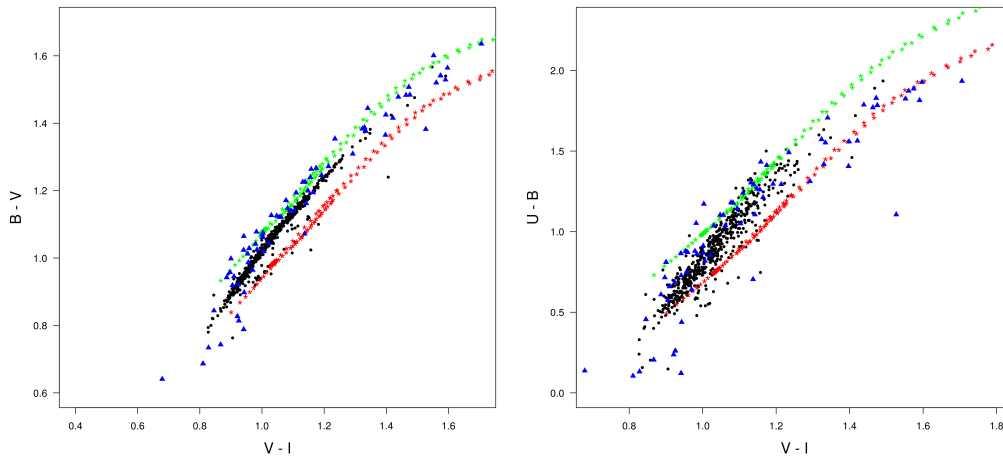
## 2.5.2 Clusters

We also considered red giants in clusters, where the metallicity is the same for all the stars. The criteria to choose the cluster were: young (around 2Gyr), with solar metallicity ( $[\text{Fe}/\text{H}] \sim 0.0$ ), maximum photometric information (UBVI and JHK if possible) and containing red giant stars. We selected NGC 2682: age = 2.56 Gyr,  $[\text{Fe}/\text{H}] = 0.0$ , BVI data from Montgomery 1993,  $E_{(B-V)} = 0.059$

Left panel in fig. 2.16 shows the  $B - V$  vs  $V - I$  CC diagram for the red giants in NGC 2682, corrected from extinction (assuming  $A_V$  as  $A_0$ , and using the Kurucz extinction model and the Fitzpatrick and Massa [2007] extinction law). The original and the shifted isochrones obtained with the MLE method are overplotted to show again that models for giant stars do not match the observations. On the right panel we show a set of main sequence stars from this cluster compared to the MS theoretical points of the same Padova isochrone (stage 1). For MS stars the agreement is good.



**Figure 2.14** – Same as in previous figure but with the observed CalSPEC giant spectra. Only one star with optical coverage was available

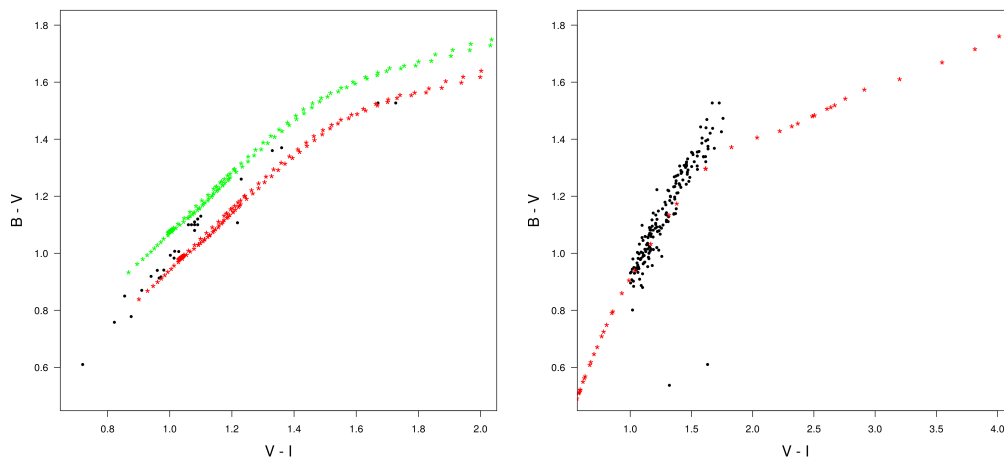


**Figure 2.15** – Same as in previous figure with the observed NGSL giant's spectra, but for  $B - V$  vs  $V - I$  and  $U - B$  vs  $V - I$  since NGSL spectra do not cover the infrared. The spectral types of these 379 stars extend from G2III to K5III

### Concluding remarks

The calibration method presented in this chapter intended to model RC colour-colour relations taking theoretical models into account. However, after considering different constraints (metallicities, parallaxes, extinctions, and even ages), analysed various possibilities and performed several tests, the method did not converge to the expected solution (i.e. agreement between observations and models). Not only the *gap* is not corrected, but the slope between observations and models is not good either (a clear example may be seen in the  $U - B$  vs  $V - I$  diagram in the bottom left panel of figure 2.11 of previous section 2.4). Thus only a shift is not enough. Moreover, the above tests (spectral libraries and clusters) showed us that filter modelling is correct and that it is most probably an issue of atmosphere models.

All these facts, combined with the unavailable Gaia calibrated filter model for the First Data Release (DR1), finally led us to discard the use of models and adopt a purely empirical calibration (see chapter 3).



**Figure 2.16** –  $B - V$  vs  $V - I$  CC diagrams of (left) red giants and (right) MS stars in NGC 2682. All data is corrected from extinction, assuming  $A_0 \sim A_V = E_{(B-V)} = 0.059$ , and using the Kurucz extinction model and the [Fitzpatrick and Massa \[2007\]](#) extinction law. Red points correspond to the 2Gyr Padova isochrone at 0 metallicity. Green points in the left diagram belong to the shifted isochrone obtained with MLE method



# Red Clump calibration (II): empirical

## Contents

---

<b>3.1 Overview</b> . . . . .	<b>82</b>
<b>3.2 Colour-colour calibration</b> . . . . .	<b>82</b>
3.2.1 Data selection . . . . .	82
3.2.2 MCMC method . . . . .	88
3.2.3 Results . . . . .	91
<b>3.3 Effective Temperature - Colour calibration</b> . . . . .	<b>95</b>
3.3.1 Data selection . . . . .	95
3.3.2 MCMC method . . . . .	96
3.3.3 Results . . . . .	97
<b>3.4 Discussion</b> . . . . .	<b>101</b>
3.4.1 Comparison with other $T_{\text{eff}}$ -Colour relations . . . . .	101

---

## 3.1 Overview

Besides their importance for use by themselves, accurate empirical calibrations have also the advantage of being used as a check of theoretical models, allowing to revisit the theory behind a particular population of stars. These calibrations are even more important in the current era of on-going large scale astrometric, photometric, spectroscopic and asteroseismic surveys, such as *Gaia*, *APOGEE*, *LAMOST*, *Kepler* or *CoRoT*. Indeed, all these data, especially the upcoming from *Gaia*, will clearly mark a milestone for stellar evolution and population synthesis, and will have an important impact on the understanding of the formation and evolution of the Milky Way.

In this chapter we aim to accurately calibrate the Gaia RC as standard candle through a completely empirical method. To do so we derived a complete and robust CC calibration covering wavelengths from the ultraviolet to the near-infrared, including the Gaia G band (section 3.2). Since at the time of writing this thesis the first Gaia data is still not available for scientific exploitation, the calibrations presented here use *Hipparcos* data and the  $H_p$  band instead. The first Gaia catalogue will be published in September 2016, at the same time of the end of this thesis. At that moment all this work will be fully updated and published. Moreover, using a similar method, in section 3.3 we extended the work to likewise derive RC effective temperature - colour relations (hereafter  $T_{\text{eff}}C$ ).

## 3.2 Colour-colour calibration

There exist very few empirical CC relations in the literature. von Braun et al. [1998] studied the dependence of  $U - B$ ,  $B - V$ ,  $V - R$ ,  $V - I$ ,  $J - K$  and  $H - K$  with metallicity and  $V - K$  colour for a set of red giant stars in the metal rich open cluster NGC 6791. They used Johnson-Cousins *UBVRI* photometry and *JHK* from the California Institute of Technology (CIT). They found a significant dependence for  $U - B$  and  $B - V$  whereas it is weak or negligible for the other colours. Bessell et al. [1998] compared CC relations and colour-bolometric correction relations for  $(J - K, V - K)$ ,  $(V - I, V - K)$ , and  $(B - V, V - I)$  indices for both dwarfs and giants, computed from synthetic spectra using the passbands of the Johnson-Cousins-Glass *UBVRIJHKL* photometry. They found a general good agreement with ATLAS9 and NMARCS<sup>1</sup> models except for a few colours, like  $B - V$  for giants and dwarfs and  $J - K$  for dwarfs.

In this section we provide CC metallicity-dependent calibrations for solar neighbourhood RC stars, using  $B, V, H_p, I, J, K, B_T, V_T$  magnitudes. All the relations are given versus the same colour index,  $H_p - K$  ( $G - K$  in the end).

### 3.2.1 Data selection

Giant stars may be easily chosen according to the surface gravity, since they have typical values of  $0.5 < \log g < 3.5$ . However, the selection was finally done using parallax and colour constraints. The reason is that even if we took care to consider only stars with spectroscopic  $\log g$  measurements from reliable sources (the same as those used for the spectroscopic metallicity, see later), we realised that the risk of incorrect measurements was high. To highlight this, we show in figure 3.1 the absolute magnitude-colour diagram for a set of supposed giant stars, selected with the mentioned  $\log g$  criteria (left) and with the final adopted parallax-colour criteria (right). We can see that the  $\log g$  sample is also including an important number of main sequence and dwarf stars, while with the parallax-colour cut we guarantee for sure that the stars will be only on the upper right part of the HR diagram, hence on the giant branch.

Therefore the main criteria considered were:

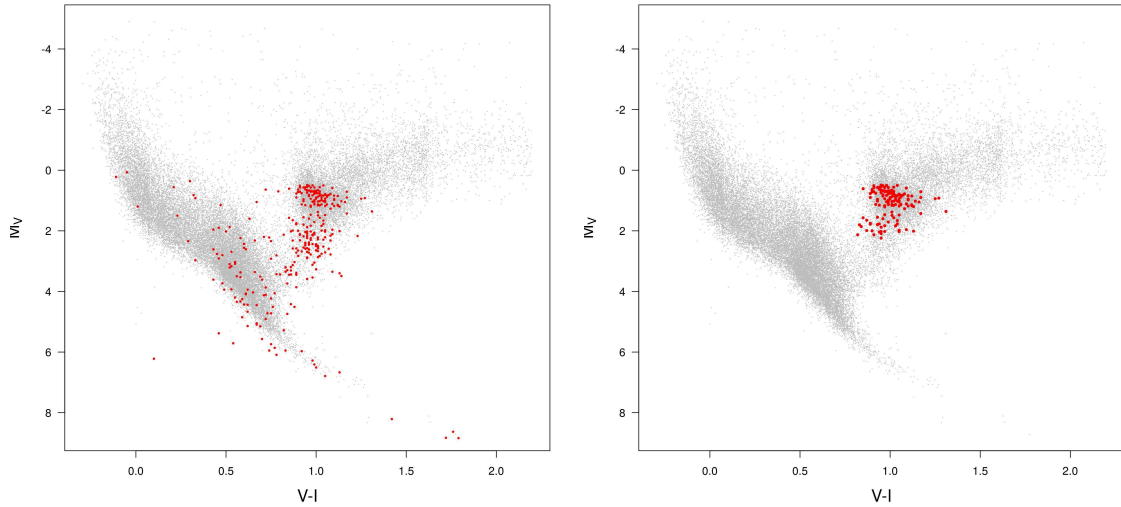
---

<sup>1</sup>Revised MARCS program

$$V - I > 0.8$$

$$m_V + 5 + 5 \log_{10} \left( \frac{\varpi + F^{-1}(0.99) \sigma_{\varpi}}{1000} \right) < 2.5$$

with the parallax  $\varpi$  within a confidence interval given by the quantile function  $F^{-1}(p)$  with  $p = 0.99$ , 0 mean and standard deviation 1.  $F^{-1}(0.99)$  is also known as the inverse cumulative distribution function (ICDF) at 99%, i.e. the boundary value at which the parallax probability density function covers an area of 99% (99th percentile of a normal distribution). This gives a value of  $\sim 2.32$ . Hence we selected all those stars with parallaxes within a range of  $\sim 2.32\sigma_{\varpi}$ .



**Figure 3.1** – Giants selection by (left)  $\log g$  [0.5 - 3.5] and (right) parallax on the *Hipparcos* HR diagram. The selection includes only stars with extinction  $A_V < 0.03$  according to *FEROS* and *SOPHIE* archives, [Lallement et al. \[2014\]](#) 3D and [Schlegel et al. \[1998\]](#) 2D extinction maps (see text for details)

To ensure the accuracy of the calibration, other conditions were also applied in order to have high photometric quality, only single stellar systems, and a set of stars with very low interstellar extinction:

- *BVH<sub>p</sub>IJK B<sub>T</sub>V<sub>T</sub> photometry*: we selected those stars which have photometric data in the following surveys:
  1. *BVH<sub>p</sub>I*: extracted from the *Hipparcos* catalogue
  2. *B<sub>T</sub>V<sub>T</sub>*: extracted from the *Tycho-2* catalogue
  3. *JK*: extracted from the *2MASS* catalogue and [Laney et al. \[2012\]](#)

We have not considered the U magnitude because of the few high quality data available for *Hipparcos* stars (a fact that would importantly reduce the sample), nor H magnitude because of its strong dispersion on the CC diagram. We expect to be able to include them when Gaia data is available.

- *High infrared photometric quality*: only stars with high quality in all infrared bands were considered: i.e. flag q2M = "AAA" for the *2MASS* catalogue and all the stars from [Laney et al. \[2012\]](#)
- *Binarity and multiplicity*: we removed all stars flagged as binaries and belonging to multiple systems. To do so we took into account, as in section 2.2, the specific flags in the *Hipparcos* catalogue as well as the last updated information from:

– *9th Catalogue of Spectroscopic Binary Orbits (SB9)* [[Pourbaix et al., 2004, 2009](#)]

- *Catalogue of the Components of Doubles and Multiple Stars* (CCDM) [Dommanget and Nys, 1994, 2002]
- *The Washington Visual Double Star Catalog* (WDS) [Mason et al., 2001]

In this chapter, we also considered only stars for which the proper motions from *Hipparcos* are consistent with those of *Tycho-2* (rejection p-value: 0.001). According to a specific test carried out in the framework of the *Gaia* data validation team (CU9 WP944) most of the stars for which the proper motions are not consistent between both catalogues, are expected to be long period binaries not detected in *Hipparcos*, and for which the longer time baseline of *Tycho-2* could have provided a more accurate value

- *Spectroscopic metallicity*: adopted from different validated sources, and only those data with error information: *RAVE*, *APOGEE*, *PASTEL*, *AMBRE*, *LAMOST* catalogues/surveys, Alves et al. [2015], Luck [2015], Jofré et al. [2015], Takeda et al. [2008], Valentini and Munari [2010], da Silva et al. [2015], Puzeras et al. [2010], Morel et al. [2014], Maldonado and Villaver [2016], Zieliński et al. [2012a], Bensby et al. [2014]

Note: these sources are the same as those used for the  $\log g$  selection in figure 3.1 (left). In particular, the red points appearing on the main sequence instead of on the giant branch (where they should be since their surface gravity falls within the  $0.5 < \log g < 3.5$  range) are mainly from *RAVE* and *PASTEL*, and few stars come also from *AMBRE*, *LAMOST* and Luck [2015]

- *Infrared Excesses of Hipparcos stars*: we checked whether there were stars with significant infrared excesses. McDonald et al. [2012] took advantage of several spectral energy distributions SEDs to identify and quantify infrared excesses ( $E_{IR}$ ) due to circumstellar matter on 107619 *Hipparcos* stars. This excess is defined as the average of all observed SED-modelled flux ratios in the 3-25  $\mu m$  region.

$$E_{IR} = \sum_{\lambda > 2.2\mu m} \frac{F_{\nu}^{obs} / F_{\nu}^{model}}{n_{obs}}$$

where  $F_{\nu}^{obs}$  and  $F_{\nu}^{model}$  are, respectively, the observed and modelled fluxes at frequency  $\nu$  (corresponding to wavelength  $\lambda$ ), and  $n_{obs}$  is the number of observations at wavelengths  $> 2.2\mu m$

For all our colour-colour combinations, we checked if there were outliers at more than  $3\sigma$ . If so, we looked for their associated  $E_{IR}$  with the goal of removing the stars with  $E_{IR} > 2.5$ . None of our stars had any significant IR excess

- *Interstellar extinction*: only stars with  $A_V < 0.03$  from a compilation of *FEROS* and *SOPHIE* archives, Lallement et al. [2014] 3D and Schlegel et al. [1998] 2D maps were kept (see details on the extinction treatment below)

The final sample contains 144 stars. Figures 3.2 and 3.3 show the parameter's distribution and the position on an Aitoff map of these stars.

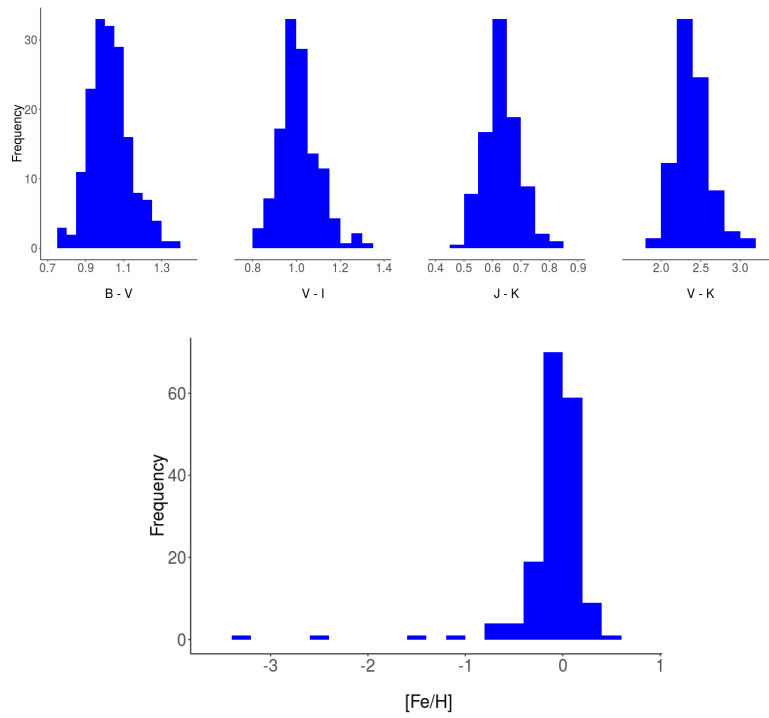
### Interstellar extinction

In order to minimize the effects of interstellar extinction on the photometric calibration, we kept only stars with very low (almost 0)  $A_V$ . Nevertheless, we are aware that this selection will inevitably introduce a bias in the calibrations, although minimally.

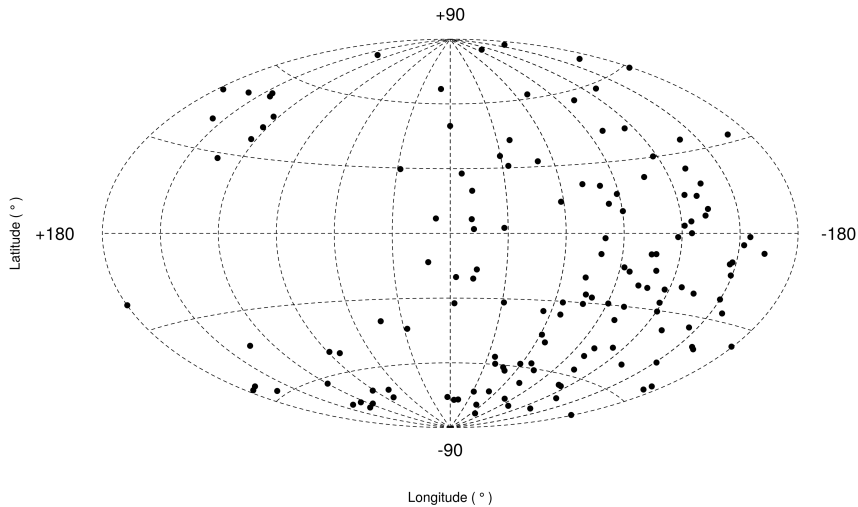
To do so, we considered the following two aspects:

- **Diffuse Interstellar Band: *FEROS* and *SOPHIE* archives**

The Diffuse Interstellar Bands (DIBs) are absorption lines observed in the spectra of stars and other astronomical objects, which are caused by the fraction of light absorbed by the interstellar medium



**Figure 3.2** – (Top) Colours and (bottom) metallicity distributions of the giants CC sample



**Figure 3.3** – Distribution of the CC sample on an Aitoff map projection in galactic coordinates

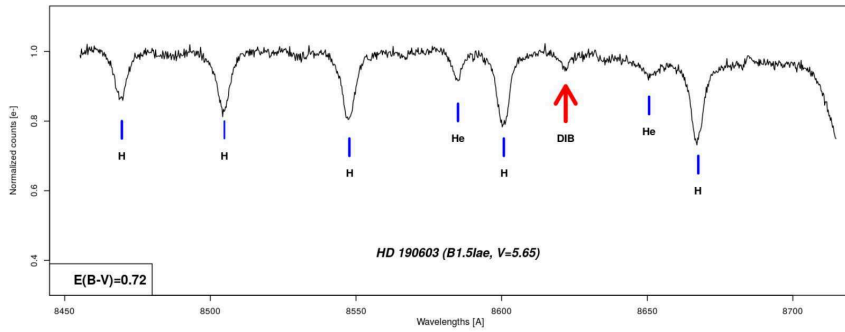
(ISM), composed of gas and dust. The strength of the DIB is a function of the amount of interstellar medium along the line of sight toward the star, hence between the star and the instrument. Therefore, it is a powerful tool to constrain the extinction of stars, and can be used to reconstruct the large-scale distribution of the ISM in the Milky Way

By using the spectra from two high-resolution echelle spectrographs: *FEROS* (Fibre-fed Optical Echelle Spectrograph, located in La Silla, Chile) and *SOPHIE* (*Spectrographe pour l'Observation des Phénomènes des Intérieurs stellaires et des Exoplanètes*, located at the *Observatoire de Haute-Provence*, France), R. Lallement is measuring the DIBs of lots of stars (many *Hipparcos* stars among them). The DIB equivalent width (EW) is linearly related with the reddening and can present a large dispersion [Raimond et al., 2012, Chen et al., 2013, Puspitarini et al., 2015]. R. Lallement provided

us the reddening obtained by averaging between 1 and 3 DIBs for each star.

The relation between DIB and extinction, although statistically correlated, is not ensured for individual stars as the DIBs are influenced by other material in the ISM than the dust. However, if no DIB is detected, the measurement becomes a strong observational proof of the absence or of a very low amount of extinction. Since in the current work we are interested in stars with negligible extinction, the use of DIBs is quite reliable.

It is worth to mention that the Radial Velocity Spectrograph (RVS) of Gaia will also provide spectra which will allow to determine the DIBs for a huge amount of stars, and thus to improve ISM measurements. This, together with the accuracy of Gaia magnitudes and parallaxes, will directly let us to improve our photometric calibration. Preliminary Gaia spectra have already clearly shown DIBs in several stars (see fig. 3.4 as an example).



**Figure 3.4** – DIB in the Gaia high-resolution spectra of HD 190603. The DIB position is marked with a red line. Other spectral lines such as those of Hydrogen and Helium have also been marked, but these lines are produced in the atmosphere of the star and not by the interstellar material. Credits from the ESA Gaia DPAC CU6 team at Observatoire de Paris-Meudon

### • 3D and 2D extinction maps

Since we still do not have DIB measurements for all of our selected stars, we also considered data from [Lallement et al. \[2014\]](#) 3D and [Schlegel et al. \[1998\]](#) 2D interstellar extinction maps.

As stated in section 2.2, the 3D extinction maps of [Lallement et al. \[2014\]](#) provide differential reddening caused by the interstellar dust within  $\sim 1$  kpc from the Sun (fig. 3.5). For stars outside the 3D map, we use as a maximum  $E_{(B-V)}$  value the 2D integrated map of [Schlegel et al. \[1998\]](#).

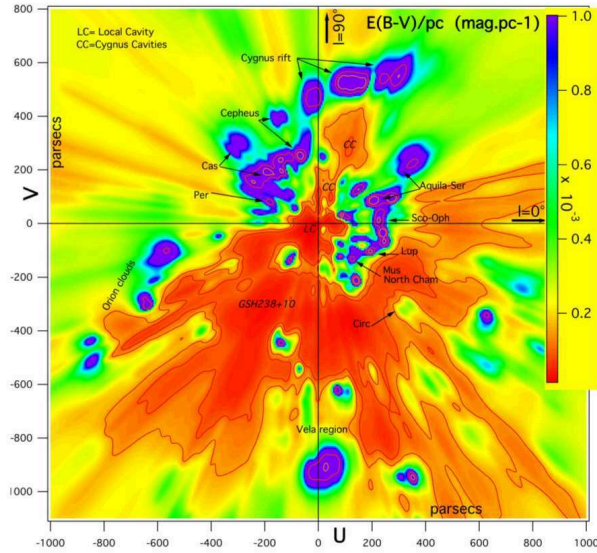
Again, maximum and minimum  $E_{(B-V)}$  values from 3D charts were directly provided to us by the author, corresponding to the values provided by the map within  $1\sigma$  parallax error. We used  $E_{(B-V)_{\max}}$  in order to select a subset of stars with negligible interstellar extinction. For the 2D maps, the reddening was obtained by using their conversion of galactic coordinates ( $l, b$ ) to pixel positions in a Lambert projection ( $x, y$ ). There are two maps (fig. 3.6), one for northern Galactic hemisphere, centered at  $b = +90$  deg, with latitude running clockwise, and another for the southern Galactic hemisphere, centered at  $b = -90$  deg, with latitude running counter-clockwise. The conversion is given by:

$$x = 2048 \sqrt{1 - n \sin(b)} \cos(l) + 2047.5$$

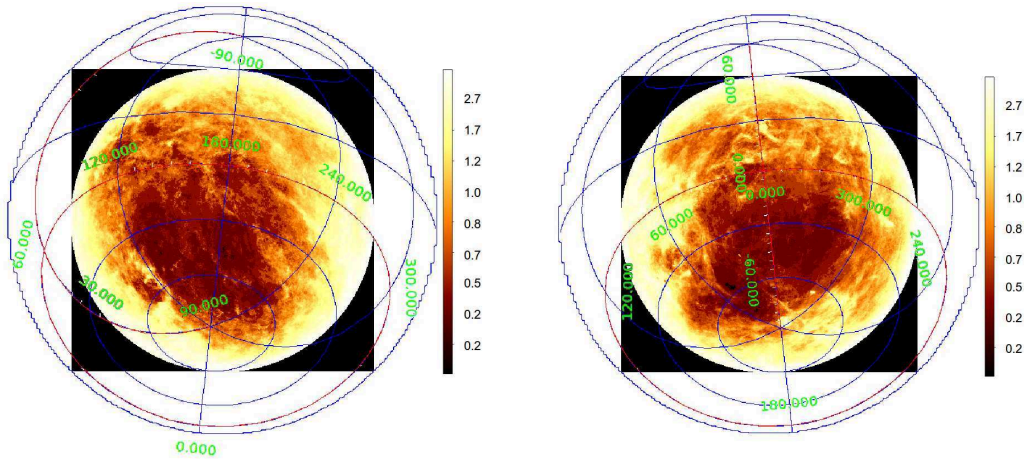
$$y = -2048 n \sqrt{1 - n \sin(b)} \sin(l) + 2047.5$$

where  $n = +1$  for the northern Galactic projection, and  $n = -1$  for the southern Galactic projection. The maps provide a value of  $E_{(B-V)}$  for each pixel position.

The values of  $A_V$  were assigned to our stars with the following priorities: 1. *FEROS*, 2. *SOPHIE*, 3. 3D maps, 4. 2D maps. That is, we associated the measurements from the DIBs first, and for those stars not in



**Figure 3.5** – Inverted differential opacity distribution in the Galactic plane. The Sun is at coordinates  $(x,y) = (0,0)$  and the Galactic center direction to the right. The opacity increases from red to violet (scale to the right). Credits from [Lallement et al. \[2014\]](#) (figure 1 of the article)



**Figure 3.6** – Schlegel 2D extinction maps ( $E_{(B-V)}$ ) for the two hemispheres: (left) North and (right) South. The grid of ecliptic coordinates is overlotted. The opacity increases from white to red on a logarithmic scale

the archives we looked for them in the 3D maps. If still not there, we gave them the  $A_{V_{\max}}$  provided by the 2D maps. Once the association done, all the stars with  $A_{V_{\max}} < 0.03$  were kept.

Given these priority criteria, the source of extinction for the 144 stars of the sample is distributed as follows in table 3.1:

**Table 3.1** – Extinction source for the 144 stars of the sample (with  $A_V < 0.03$ ). Extinction assignment is given with the following priorities: *FEROS*, *SOPHIE*, 3D maps, 2D maps (see text for details)

Extinction Source	# stars
<b>FEROS</b>	20
<b>SOPHIE</b>	0
<b>Lallement 3D map</b>	122
<b>Schlegel 2D map</b>	2

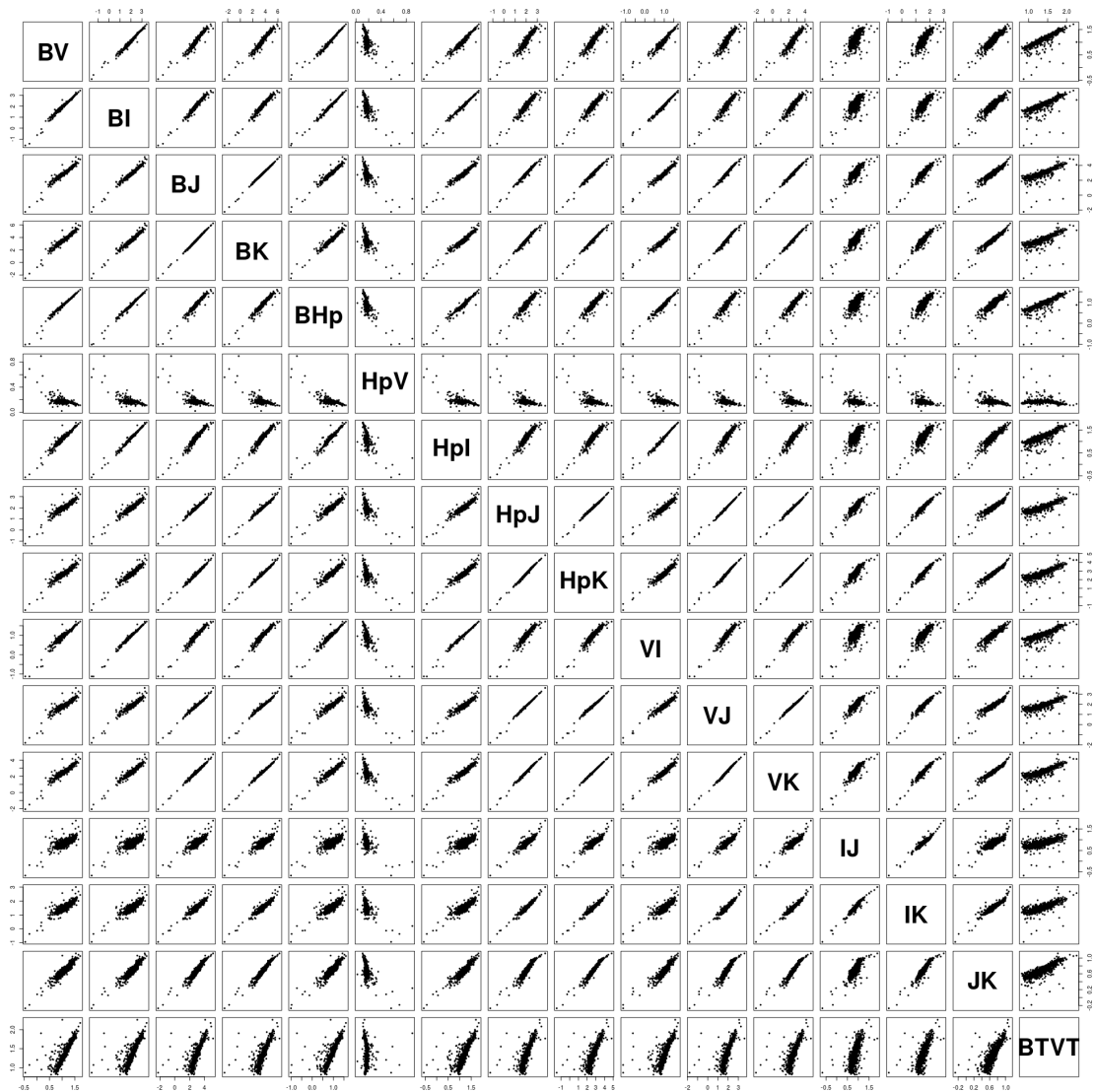
Notice that there is no stars with extinction measurements from the *SOPHIE* archive. However we decided to keep this information in the table because the archive is anyway taken into account in the selection process. The main reason is that with Gaia we will surely have more stars present in this archive. Similarly,

since all of our stars are *Hipparcos* stars, they are mostly included in the 3D maps (which were actually derived using *Hipparcos* stars). Hence following the priority criteria, the number of remaining stars to look for in the 2D charts of [Schlegel et al. \[1998\]](#) is obviously low.

### 3.2.2 MCMC method

To derive accurate CC relations, we implemented a Monte Carlo Markov Chain (MCMC) method. As can be seen in [fig. 3.7](#), all colour combinations (using the considered magnitudes mentioned above) can be well fitted by a polynomial. The choice of using a MCMC instead of others methods like multiple regression or the generalized linear model, lies in the fact that the MCMC allows to account and deal with the errors of both the predictor and response variables in a robust way.

Since the goal is to provide a complete calibration for the Gaia Red Clump, we considered the Gaia magnitude as reference together with the K band, which is the band less affected by extinction and there is high quality information available in the *2MASS* catalogue. Thus, from  $G - K$  (response variable) we may obtain the other colours (predictors). As explained, for the present implementation,  $H_p$  was used instead of  $G$ .



**Figure 3.7** – Colour-Colour diagrams pairs, for all colours considered in the empirical calibration

For each colour- $H_p - K$  relation, the process is as follows:



## 1. Input sample

- (a) Stars with full information: photometry and metallicity, with known errors
- (b) If the colour to be calibrated contains the  $I$  magnitude, the sample is limited to *Hipparcos* stars with selected sources of  $V - I$ . That is, in this catalogue twenty different methods were applied to estimate  $V - I$ . From these methods, ranked according to their hierarchy of application, we only considered those coming from: direct measurements in the Cousin's system (field  $H42 = A$ ), measurements in the Johnson's system then converted to Cousin's (field  $H42 = C$ ), measurements in the Kron-Eggen's system then converted to Cousin's (field  $H42 = E$ )
- (c) Due to the low number of metal-poor stars in the sample, we only considered solar metallicity stars ( $-0.5 \leq [\text{Fe}/\text{H}] \leq 0.5$ )

## 2. Polynomial models

- **Models:** the general fitting formula adopted is:

$$Y = a_0 + a_1 X + a_2 X^2 + a_3 [\text{Fe}/\text{H}] + a_4 [\text{Fe}/\text{H}]^2 + a_5 X [\text{Fe}/\text{H}] \quad (3.1)$$

a second order polynomial <sup>2</sup>, where  $X$  and  $Y$  are the colours to be calibrated and  $a_i$  the coefficients. In order to provide the most accurate fit for each CC pair, the process (see later) penalizes by the complex terms so that, in the end, seven different models may be tested for every relation (Model 7 being the more complex one):

Model 1 : Colour =  $a_0 + a_1 (H_p - K)$

Model 2 : Colour =  $a_0 + a_1 (H_p - K) + a_2 (H_p - K)^2$

Model 3 : Colour =  $a_0 + a_1 (H_p - K) + a_3 [\text{Fe}/\text{H}]$

Model 4 : Colour =  $a_0 + a_1 (H_p - K) + a_2 (H_p - K)^2 + a_3 [\text{Fe}/\text{H}]$

Model 5 : Colour =  $a_0 + a_1 (H_p - K) + a_3 [\text{Fe}/\text{H}] + a_4 [\text{Fe}/\text{H}]^2$

Model 6 : Colour =  $a_0 + a_1 (H_p - K) + a_2 (H_p - K)^2 + a_3 [\text{Fe}/\text{H}] + a_4 [\text{Fe}/\text{H}]^2$

Model 7 : Colour =  $a_0 + a_1 (H_p - K) + a_2 (H_p - K)^2 + a_3 [\text{Fe}/\text{H}] + a_4 [\text{Fe}/\text{H}]^2 + a_5 (H_p - K) [\text{Fe}/\text{H}]$

We note that for those colours including the  $I$  magnitude, we forced the MCMC to use only polynomials without dependence on metallicity (i.e. Models 1 and 2). Indeed, the strict selection in these particular cases caused the sample to be quite small, roughly half the size of the other colours' samples.

- **Errors:** input errors from photometry and metallicity are taken into account in the model

3. **Run MCMC:** a Monte Carlo Markov Chain was run for every model tested, with 10 chains and 10000 iterations for each. For this we used RJags <sup>3</sup> library from R programme language, which needs or allows to account for some other parameters:

- **Priors:** as part of the method and to ensure the convergence, RJags requires a prior for each  $a_i$  coefficient. A normal distribution with zero mean and a deviance of 0.01,  $\mathcal{N}(0, 0.01)$ , was set for each of these coefficients. We verified the distribution range by checking that the results fit perfectly into it
- **Initial values:** RJags allows to set an initial value for each coefficient which helps the convergence of the method. For this, we used the output coefficients obtained for every model through a multiple linear regression (simpler method which does not take uncertainties into account, but allows to obtain approximated values)

<sup>2</sup>Upper degrees were tested, but discarded by an Analysis of Variance test (ANOVA), meaning that simpler models were good enough to describe the data

<sup>3</sup>More information about RJags library in <https://cran.r-project.org/web/packages/rjags/rjags.pdf>

4. **Select best fitted model:** as stated above, the model selection was done through a process of penalization by the complex terms. To do so, we took advantage of the Deviance Information Criterion (*DIC*).

The *DIC* is, among other options (like the well known Likelihood ratio deviance test, the Wald Test, the Akaike Information Criterion or *AIC*, and the Bayesian Information Criterion or *BIC*), a way to evaluate models and choose the best one. The *DIC* can be roughly viewed as an *AIC* or *BIC* statistic for MCMC methods. A natural way to compare models is to use a criterion based on a trade-off between the fit of the data to the model and the corresponding complexity of the model. *DIC* does this in a Bayesian way:

$$\text{DIC} = \text{"goodness of fit"} + \text{"complexity"}$$

- (a) The *goodness of fit* is measured by the deviance  $D$ , obtained through the likelihood <sup>4</sup> function  $\mathcal{L}$ , with  $O$  the observables (the data) and  $\theta$  the parameters to determine

$$D(\theta) = -2\log \mathcal{L}(O|\theta)$$

- (b) The *complexity* is measured by an estimate of the "effective number of parameters" defined as:

$$p_D = E_{\theta|y}[D] - D(E_{\theta|y}[\theta]) = \bar{D} - D(\bar{\theta})$$

i.e.: posterior mean deviance minus the deviance evaluated at the posterior mean of the parameters

Hence, analogously to *AIC*, the *DIC* is defined as:

$$\text{DIC} = D(\bar{\theta}) + 2p_D = \bar{D} + p_D$$

i.e. the sum of the posterior mean deviance and the effective number of parameters.

As in *AIC*, models with smaller *DIC* agree more with the data, thus any decrease in *DIC* suggests a better model. However, as discussed by Plummer [2008], there is no absolute scale for comparison of two penalized deviance statistics. Thus, to ensure a good model selection and to do this in a more robust way, the difference must be calibrated with respect to the sample standard deviation of the individual contributions from each observed stochastic node (each term).

According to this, models were tested by pairs: a given complex model was compared to the next simpler model. However, in some cases two "simpler models" could be equally eligible. Below we summarise all the possible cases:

- Model 7 (full second degree polynomial) was always compared to Model 6 (model without the crossed term)
- Model 6 could be either compared to Model 5 (without  $(Hp - K)^2$ ) or Model 4 (without  $[\text{Fe}/\text{H}]^2$ )
- Model 5 was directly compared to Model 3 (one degree polynomial)
- Model 4 (without  $[\text{Fe}/\text{H}]^2$ ) could be either compared to Model 3 (without  $(Hp - K)^2$ ) or Model 2 (without  $[\text{Fe}/\text{H}]$ )
- Model 3 (without  $(Hp - K)^2$ ) was directly compared to Model 1 (only  $Hp - K$  contribution)

<sup>4</sup>Given a value of  $\theta$  the probability of observing  $O$  is  $P(O|\theta)$ . When modelling a real life stochastic process, we often do not know  $\theta$ : we simply observe  $O$  and the goal is to get an estimate for  $\theta$  that would be a plausible choice given the observed outcomes. A "natural" estimation process is to choose that value of  $\theta$  that would maximize the probability that we would actually observe  $O$ ,  $P(O|\theta)$ . We need to find the parameter values  $\theta$  that maximize  $\mathcal{L}(\theta|O) \propto P(O|\theta)$ , where  $\mathcal{L}(\theta|O)$  is the likelihood function

- Model 2 (without [Fe/H]) was also directly compared to Model 1 (only  $H_p - K$  contribution)

In those situations where two "simpler models" were possible, the least significant term was removed. Therefore, the method continuously looked for the next pair of models to be compared, run the MCMC for each and checked their DIC. The process kept on until:

- the DIC difference was significantly negative at  $1 \sigma$  (i.e.  $\Delta DIC + \sigma_{\Delta DIC} < 0$ ): the complex model of the given pair was kept
- only one model remained: i.e. Model 1 was kept

- Outliers:** once the best model determined, the method checked whether there were calibrated stars out of  $3\sigma$  from the model. If so, the furthest one was removed and the complete MCMC process was run again. Outliers were eliminated one by one to ensure that the further one was not causing a deviation in the model that brought to consider other stars as "false outliers"

The global process was parallelized to compute all colour- $H_p - K$  combinations at the same time. Figure 3.8 displays a schema of the implementation.

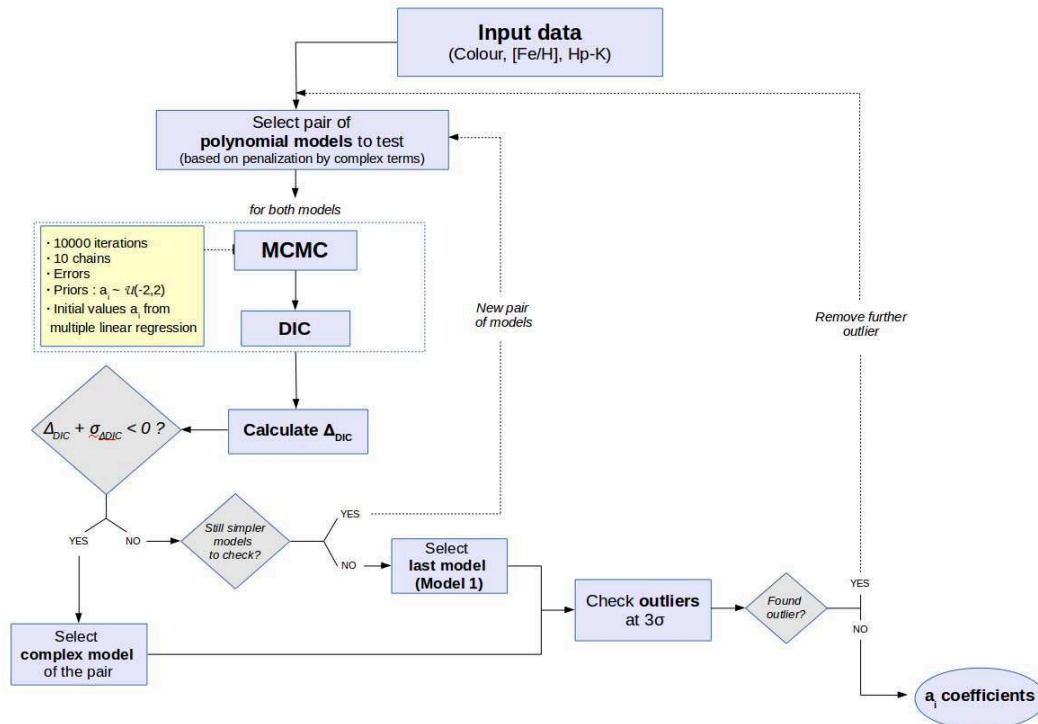


Figure 3.8 – Schema of the MCMC method implemented to calibrate the RC colour-colour relations

### 3.2.3 Results

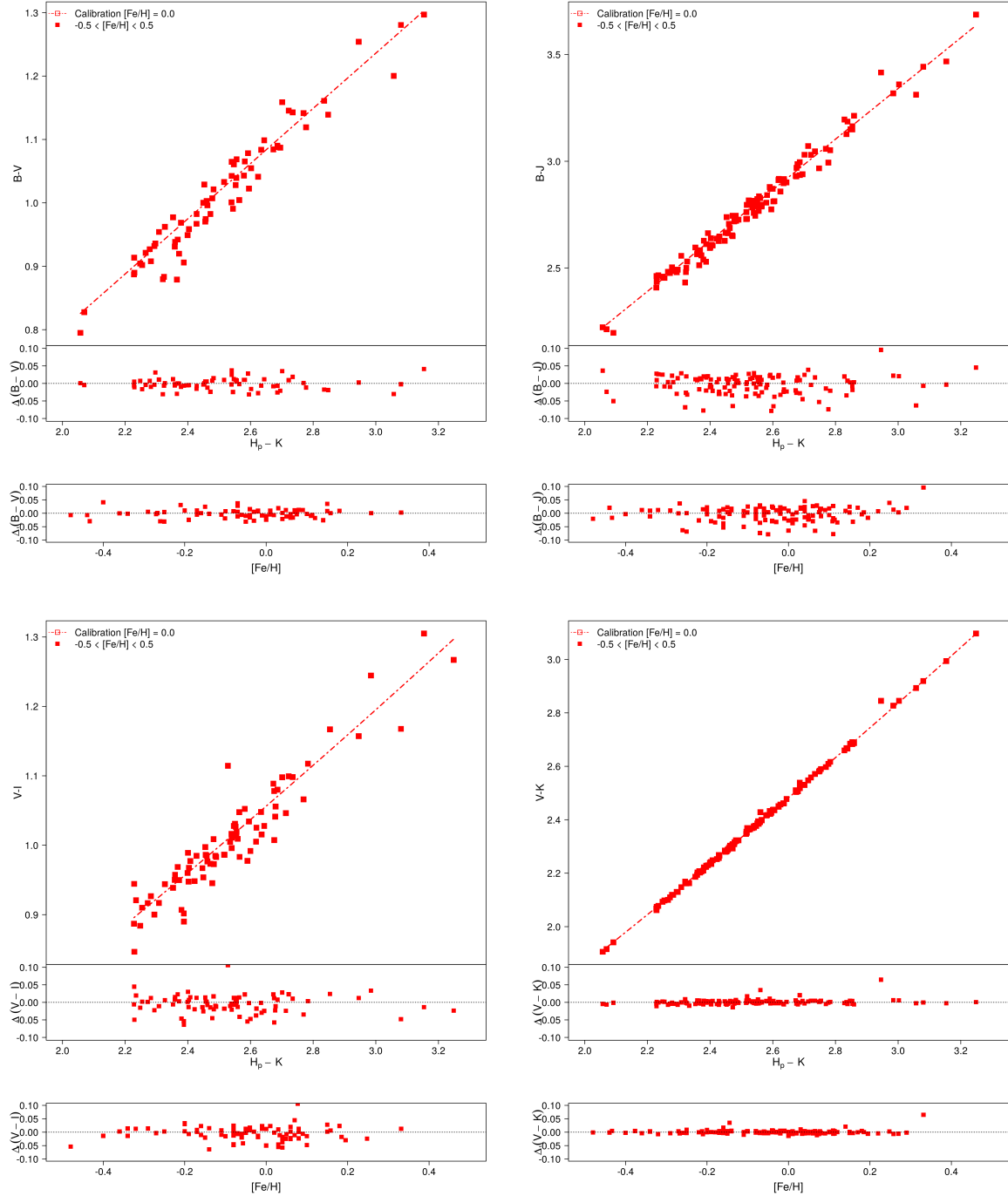
Table 3.2 gives the coefficients for each of the 15 colour- $H_p - K$  fit, together with the  $H_p - K$  and metallicity ranges of applicability (defined by the maximum and minimum values of each individual sample), as well as the number (N) of stars used. Fig. 3.9 shows the colour- $H_p - K$  relations obtained for four of the fifteen colour indices with the residuals of the fit as a function of the colour itself and of the metallicity. Note that the very small scatter in the residuals ( $\sim \pm 0.04$  globally) reflects the good quality and homogeneity of the input data. Table 3.3 specifies the selected model for each CC relation, the percentage of outliers removed and the computational time required for each fit (strongly dependent upon the number of outliers).

**Table 3.2** – Coefficients and range of applicability of colour- $H_p - K$  relations

Colour	$H_p - K$ range	[Fe/H] range	$a_0$	$a_1$	$a_2$	$a_3$	$a_4$	$a_5$	N
$B - V$	[2.1,3.2]	[-0.5,0.3]	$-0.070 \pm 0.024$	$0.435 \pm 0.010$	-	$0.118 \pm 0.012$	-	-	70
$B - I$	[2.2,3.2]	[-0.5,0.3]	$-0.111 \pm 0.043$	$0.851 \pm 0.017$	-	-	-	-	57
$B - J$	[2.1,3.2]	[-0.5,0.3]	$-0.224 \pm 0.029$	$1.188 \pm 0.012$	-	$0.132 \pm 0.015$	-	-	122
$B - K$	[2.1,3.2]	[-0.5,0.3]	$-0.263 \pm 0.027$	$1.450 \pm 0.011$	-	$0.113 \pm 0.014$	-	-	124
$B - H_p$	[2.1,3.2]	[-0.5,0.3]	$-0.275 \pm 0.031$	$0.454 \pm 0.013$	-	$0.113 \pm 0.011$	-	-	60
$H_p - V$	[2.1,3.2]	[-0.5,0.3]	$-0.104 \pm 0.031$	$0.198 \pm 0.024$	$-0.037 \pm 0.005$	$0.228 \pm 0.052$	$0.103 \pm 0.016$	$-0.089 \pm 0.021$	110
$H_p - I$	[2.2,3.2]	[-0.5,0.2]	$0.733 \pm 0.076$	$-0.067 \pm 0.059$	$0.096 \pm 0.011$	-	-	-	62
$H_p - J$	[2.1,3.2]	[-0.5,0.3]	$0.374 \pm 0.092$	$0.481 \pm 0.071$	$0.049 \pm 0.014$	$-0.268 \pm 0.128$	$-0.001 \pm 0.044$	$0.111 \pm 0.049$	134
$V - I$	[2.2,3.2]	[-0.5,0.3]	$0.227 \pm 0.204$	$0.222 \pm 0.154$	$0.035 \pm 0.029$	-	-	-	85
$V - J$	[2.1,3.2]	[-0.5,0.3]	$0.356 \pm 0.142$	$0.374 \pm 0.109$	$0.068 \pm 0.021$	-	-	-	127
$V - K$	[2.1,3.2]	[-0.5,0.3]	$0.208 \pm 0.111$	$0.721 \pm 0.085$	$0.052 \pm 0.016$	-	-	-	128
$I - J$	[2.2,3.2]	[-0.5,0.3]	$0.530 \pm 0.250$	$-0.176 \pm 0.191$	$0.101 \pm 0.037$	-	-	-	83
$I - K$	[2.2,3.2]	[-0.5,0.3]	$-0.150 \pm 0.032$	$0.595 \pm 0.013$	-	-	-	-	84
$J - K$	[2.1,3.2]	[-0.5,0.3]	$-0.042 \pm 0.021$	$0.262 \pm 0.008$	-	-	-	-	134
$B_T - V_T$	[2.1,3.2]	[-0.5,0.3]	$-0.266 \pm 0.027$	$0.585 \pm 0.011$	-	$0.186 \pm 0.012$	-	-	128

**Table 3.3** – Best model, outliers, and computational time of colour- $H_p - K$  relations

Colour	Model	Outliers (%)	Computational time (min)
$B - V$	3	43.5	12.5
$B - I$	1	32.9	103.3
$B - J$	3	4.7	131.5
$B - K$	3	3.1	96.9
$B - H_p$	3	53.1	15
$H_p - V$	10	14.1	329.3
$H_p - I$	2	27.1	91.3
$H_p - J$	10	0.7	32.6
$V - I$	2	0.0	10.2
$V - J$	2	0.8	47.7
$V - K$	2	0	26.3
$I - J$	2	2.4	21.1
$I - K$	1	1.2	17.1
$J - K$	1	0.7	46.6
$B_T - V_T$	3	5.2	151.8



**Figure 3.9** – Empirical colour- $H_p - K$  calibrations for  $B - V$ ,  $B - J$ ,  $V - I$  and  $V - K$  colour indices. The red squares belong to our sample of stars in the metallicity range of  $-0.5 \leq [\text{Fe}/\text{H}] \leq 0.5$ . The red dotted line correspond to our calibration at fixed  $[\text{Fe}/\text{H}] = 0.0$ . At the bottom of each calibration plot: residuals of the fit as a function of  $H_p - K$  and  $[\text{Fe}/\text{H}]$ . All plots are scaled to the same  $H_p - K$  colour, metallicity and colour residuals intervals

Table 3.4 provides the results obtained when testing our calibrations on the well-known K0 III red giant star Arcturus ( $\alpha$  Boötis, HIP 69673). We used  $H_p - K$  from the *Hipparcos* and *2MASS* catalogues, and metallicity information from Morel et al. [2014], and we compared the calibrated colours with the photometric information provided in those catalogues (corrected from extinction with the value given by the 3D maps of Lallement et al. [2014],  $E_{(B-V)} = 4.1710^{-4}$ ). We obtained globally a good agreement for all colours, mostly within  $1\sigma$  confidence interval,  $2\sigma$  for  $H_p - J$  and  $3\sigma$  for  $H_p - V$ ,  $I - J$  and  $J - K$ . The uncertainties have been calculated by taking advantage of the output traces of each MCMC (i.e. the variations of each polynomial coefficient during the MCMC iterations of all chains).

Note that this star was not included in our calibration sample because of some of the applied constraints:

- **Photometric quality:** its *2MASS* photometric quality flag for  $J$  and  $K$  bands is "C", whereas we only considered stars flagged as "A". This means that the photometric measurements are of good quality (i.e. we can use them as a comparison), but the signal to noise ratio of the scan was lower and the measurement uncertainties slightly less accurate and reliable ( $< 0.2714$ )
- **Binarity:** according to *Hipparcos*, Arcturus is a binary star with a companion about twenty times dimmer than the primary and orbiting too close from the primary to allow a clear detection. Its binary nature is still not conclusive but recent studies provide similar results

**Table 3.4** – Colours of Arcturus derived using our colour- $H_p - K$  relations, compared to the literature

Colour	Reference	
	Hipparcos/2MASS	This work
$B - V$	$1.239 \pm 0.006$	$1.164 \pm 0.076$
$B - I$	$2.459 \pm 0.015$	$2.529 \pm 0.208$
$B - J$	$3.441 \pm 0.157$	$3.276 \pm 0.204$
$B - K$	$4.100 \pm 0.170$	$4.041 \pm 0.247$
$B - H_p$	$1.078 \pm 0.005$	$1.018 \pm 0.079$
$H_p - V$	$0.161 \pm 0.005$	$0.239 \pm 0.025$
$H_p - I$	$1.381 \pm 0.014$	$1.453 \pm 0.106$
$H_p - J$	$2.363 \pm 0.157$	$2.231 \pm 0.123$
$V - I$	$1.220 \pm 0.020$	$1.297 \pm 0.117$
$V - J$	$2.202 \pm 0.157$	$2.112 \pm 0.134$
$V - K$	$2.861 \pm 0.170$	$2.860 \pm 0.175$
$I - J$	$0.982 \pm 0.158$	$0.783 \pm 0.068$
$I - K$	$1.640 \pm 0.171$	$1.568 \pm 0.099$
$J - K$	$0.659 \pm 0.231$	$0.751 \pm 0.045$
$B_T - V_T$	$1.343 \pm 0.013$	$1.376 \pm 0.102$

### 3.3 Effective Temperature - Colour calibration

The effective temperature is directly related to the observables of a star, like the photometric colours. Indeed, relations between both are widely used and crucial in the context of deriving reliable colours for theoretical stellar models, and thus to determine basic properties, as ages, distances and metallicities, for a given stellar population.

The determination of effective temperatures for giants stars is of utmost importance in stellar astrophysics. Accurate stellar effective temperatures are essential for reliable abundance estimates and analyses, and therefore for improving our understanding of Galactic chemical evolution.

However, estimations of this parameter may vary significantly from one method to another. For high resolution spectroscopy the typical techniques used are by fitting the Balmer lines profiles [Gray, 1992, Heiter et al., 2002, Nissen et al., 2007, Fuhrmann, 2008], by adjusting the ionisation and *excitation balance* of iron lines in a 1D LTE analysis [Takeda et al., 2002, Santos et al., 2004], or through the line-depth ratio technique Kovtyukh et al. [2003]. For low and intermediate resolution spectroscopy, instead, the techniques usually adopted are the fit between stellar continuum and model atmosphere fluxes [Smalley, 2005, Norris et al., 2013], and the strength of Balmer jump [Sokolov, 1995]. In almost all these techniques the effective temperature is determined from the measured spectrum simultaneously with the metallicity and the surface gravity, thus leading to somehow degenerated results [Huang et al., 2015].

We see that all these methods of  $T_{\text{eff}}$  estimation require reliable stellar atmosphere models and spectrum synthesis, and, consequently, lead to model-dependent results.

Another technique widely used nowadays and less model-dependent is the infrared flux method (IRFM) introduced by Blackwell and Shallis [1977], and based on the Kurucz model. Thanks to the fact that it is less model-dependent, the derived temperatures can be used to determine colour-temperature-metallicity relations spanning a wide range of parameters [Alonso et al., 1996, Ramírez and Meléndez, 2005, Casagrande et al., 2006, González Hernández and Bonifacio, 2009, Casagrande et al., 2010]. However, the reddening and the lack of absolute zero-point calibration of the IRFM may easily induce some systematic  $T_{\text{eff}}$  errors, up to  $\sim 100\text{K}$ , among the different studies (see discussion of Casagrande et al. [2010]). And changes of  $\sim 100\text{K}$  may result in important variations of, for instance, the oxygen-iron ratio and lithium initial abundances. Systematics can be thus found in most of the  $T_{\text{eff}}$  published values, regardless the technique used, leading to discrepancies among published  $T_{\text{eff}}$  values.

Moreover, we showed in chapter 2 the problem between models and observations for giant stars, especially with ATLAS model, which is the one used in the IFRM method. All these reasons finally led us to construct a robust and completely empirical metallicity-dependent  $T_{\text{eff}}$ -colour relation (hereafter  $T_{\text{eff}}\text{C}$ ).

Similarly as done for CC calibrations in previous section, we will describe the criteria and constraints applied to select the  $T_{\text{eff}}\text{C}$  sample, the implemented method and the results obtained using  $H_p - K$  as reference colour. We remind that in this work  $H_p$  is used instead of  $G$  since Gaia data will be published at the same time this thesis will finish.

#### 3.3.1 Data selection

The criteria to select the effective temperature sample were similar to those for CC calibrations. Again, selecting by  $\log g$  brings to an important number of non-giant stars. Therefore, we applied the same parallax-colour cut, as well as the multiplicity, binarity, spectrometric metallicity, IR excesses and extinction (cut at  $A_V = 0.03$  using *FEROS* and *SOPHIE* archives and 3D and 2D extinction maps) criteria. Since we aimed to obtain just a  $[T_{\text{eff}}, G - K]$  relation ( $H_p$  in this thesis), we also selected all the stars with 2MASS high photometric quality in the K magnitude.

The key for a reliable effective temperature calibration is homogeneity. Thus, one must be careful when combining parameters from different methods (photometry, spectrometry, interferometry, asteroseismology), studies or surveys, since there may be important variations or bias among them leading to inaccurate

results. However, it becomes a tricky aspect if we want to get a reasonable set of stars, which answers to all previous constraints and also has an homogeneous effective temperature.

Different samples were considered throughout this research in order to derive an accurate  $[T_{\text{eff}}, H_p - K]$  relation. The Gaia catalogue will provide precise photometric information for thousands of RC stars, which will allow us to obtain a larger, high quality and completely homogeneous sample, together with data from other asteroseismic and/or spectrometric full sky recent surveys. Meanwhile we used *Hipparcos* photometry with effective temperatures carefully selected from specific sources in order to develop the method and test it as accurately as possible. Five different samples were finally considered separately, all of them with spectroscopic  $T_{\text{eff}}$  (except for Luck [2015] who provides homogeneous photometric measurements, and PASTEL which combines photometric and spectroscopic temperatures), and only containing solar metallicity stars ( $-0.5 \leq [\text{Fe}/\text{H}] \leq 0.5$ ) because of the low number of metal-poor stars in each one:

- 45 stars from *APOGEE* (raw data, not the photometric corrections provided by default in the catalogue, see section 5.1 in Holtzman et al. [2015])
- 43 stars from Alves et al. [2015]
- 27 stars from Luck [2015] (homogeneous photometric  $T_{\text{eff}}$ )
- 16 stars from Kovtyukh et al. [2006]
- 116 stars from PASTEL (sample not homogeneous, since the catalogue combines spectroscopic and photometric  $T_{\text{eff}}$ )

Table 3.5 specifies the source of extinction information for each sample. As for the CC calibrations, none of our stars is in the *SOPHIE* archive. And since they are present in the 3D maps, given the priority criteria, no information from Schlegel et al. [1998] is used. But this will change with Gaia when we will have larger samples.

**Table 3.5** – Extinction sources for each  $T_{\text{eff}}-H_p - K$  sample. Extinction assignment is given by priority: *FEROS*, *SOPHIE*, 3D maps, 2D maps (see text for details)

Sample	# stars			
	FEROS	SOPHIE	Lallement 3D map	Schlegel 2D map
<i>APOGEE</i>	0	0	45	0
Alves et al. [2015]	4	0	39	0
Luck [2015]	1	0	26	0
Kovtyukh et al. [2006]	0	0	16	0
PASTEL	17	0	99	0

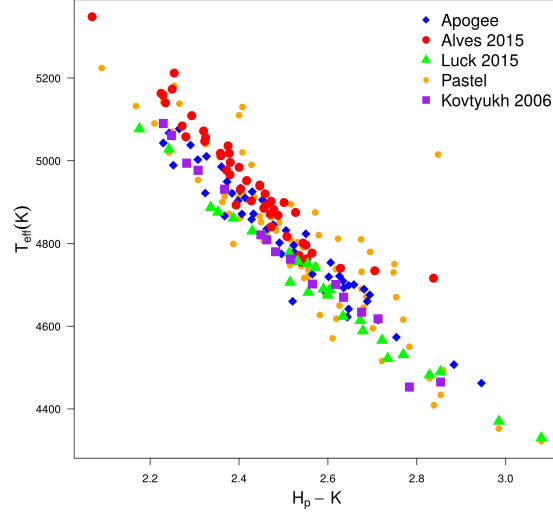
Fig. 3.10 shows the distribution of stars of these five samples on the  $[T_{\text{eff}}, H_p - K]$  diagram. The blue diamonds belong to *APOGEE*, the red filled circles to Alves et al. [2015], the green triangles to Luck [2015], the purple filled squares to Kovtyukh et al. [2006], and the orange points to PASTEL.

### 3.3.2 MCMC method

The implemented method follows the same schema as for CC relations (shown in Fig. 3.8). A second degree polynomial model accounting for both  $H_p - K$  and chemical composition was proposed, and we let the MCMC decide if simpler models fit better the data by penalizing by the complex terms. The MCMC process accounted for input errors, considered a uniform distribution  $\mathcal{U} \sim [-2, 2]$  as prior for each coefficient (RJags requirement), and set the output of a multiple linear regression as the initial values.

The fitting formula in 3.2 is the one generally adopted in most of the previous studies dealing with metallicity-dependent  $T_{\text{eff}}C$  calibrations (e.g. Alonso et al. [1996], Ramírez and Meléndez [2005], Masana et al. [2006], Casagrande et al. [2006], González Hernández and Bonifacio [2009], Casagrande et al. [2010],





**Figure 3.10** –  $[T_{\text{eff}}, H_p - K]$  diagram with stars from the five selected samples for the calibration: *APOGEE* (blue diamonds), *Alves et al. [2015]* (red filled circles), *Luck [2015]* (green triangles), *Kovtyukh et al. [2006]* (purple filled squares) and *PASTEL* (orange points)

*Huang et al. [2015]*). All  $T_{\text{eff}}C$  relations seem to adequately fit the observations by simpler polynomials, hence there is no need to get to higher orders to remove possible systematics.

$$\theta_{\text{eff}} = a_0 + a_1 (H_p - K) + a_2 (H_p - K)^2 + a_3 [\text{Fe}/\text{H}] + a_4 [\text{Fe}/\text{H}]^2 + a_5 (H_p - K) [\text{Fe}/\text{H}] \quad (3.2)$$

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

Although equation 3.2 is the conventional approach, we followed the same procedure as for CC calibrations. Thus, when penalizing by the complex terms seven possible models arise:

- Model 1 :  $\theta_{\text{eff}} = a_0 + a_1 (H_p - K)$
- Model 2 :  $\theta_{\text{eff}} = a_0 + a_1 (H_p - K) + a_2 (H_p - K)^2$
- Model 3 :  $\theta_{\text{eff}} = a_0 + a_1 (H_p - K) + a_3 [\text{Fe}/\text{H}]$
- Model 4 :  $\theta_{\text{eff}} = a_0 + a_1 (H_p - K) + a_2 (H_p - K)^2 + a_3 [\text{Fe}/\text{H}]$
- Model 5 :  $\theta_{\text{eff}} = a_0 + a_1 (H_p - K) + a_3 [\text{Fe}/\text{H}] + a_4 [\text{Fe}/\text{H}]^2$
- Model 6 :  $\theta_{\text{eff}} = a_0 + a_1 (H_p - K) + a_2 (H_p - K)^2 + a_3 [\text{Fe}/\text{H}] + a_4 [\text{Fe}/\text{H}]^2$
- Model 7 :  $\theta_{\text{eff}} = a_0 + a_1 (H_p - K) + a_2 (H_p - K)^2 + a_3 [\text{Fe}/\text{H}] + a_4 [\text{Fe}/\text{H}]^2 + a_5 (H_p - K) [\text{Fe}/\text{H}]$

with  $a_i$  the coefficients.

During the process outliers were rejected at more than  $3\sigma$  from the best fit. As in the CC method, they were treated one by one: every time outliers were found, the furthest one was removed and the MCMC fitting process was run again.

### 3.3.3 Results

Table 3.6 gives the coefficients of the fit for the five samples, together with the  $H_p - K$  and metallicity ranges of applicability, and the number (N) of stars used after the  $3\sigma$  clipping. Table 3.7 specifies the selected model in each case, the percentage of outliers removed and the computational time required. Figs. 3.11 and 3.12 show the  $[T_{\text{eff}}, H_p - K]$  relation obtained for each sample and their residuals as a function of  $H_p - K$  colour index and metallicity. For the case of the *PASTEL* sample (fig. 3.12) the discrepancies are larger, even more than  $100K$  for some stars, whilst for the other four samples the scatter is lower, up to  $\sim 60K$  maximum.

Finally, in figure 3.13 we compare the five calibrations obtained. Residuals were derived with respect to *Luck [2015]* and within its range of  $H_p - K$  colour, as long as it fell within the range of  $H_p - K$  of

**Table 3.6** – Coefficients and range of applicability of the  $[\theta_{\text{eff}}, H_p - K]$  relation

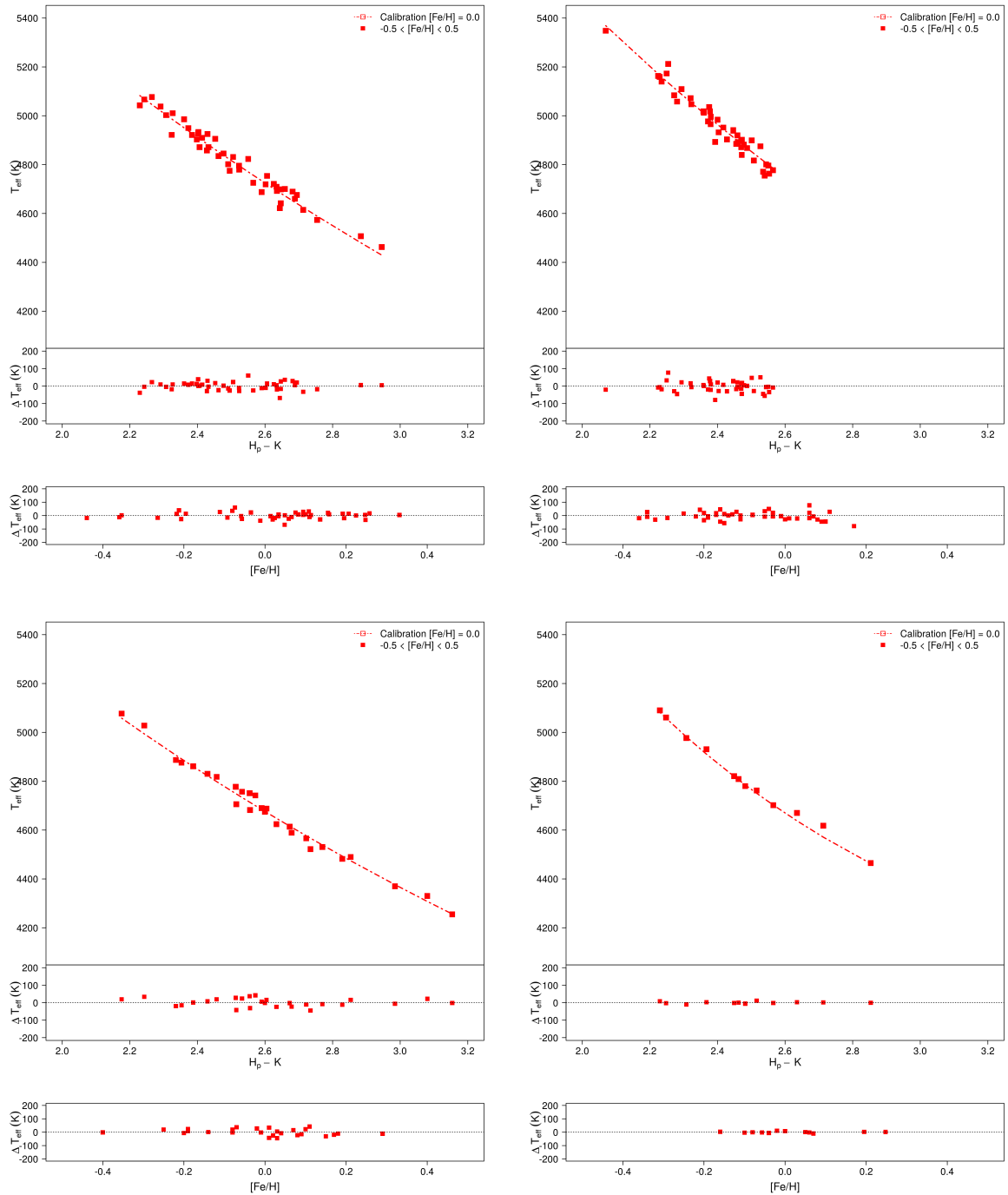
Sample	$H_p - K$ range	[Fe/H] range	$a_0$	$a_1$	$a_2$	$a_3$	$a_4$	$a_5$	N
<i>APOGEE</i>	[2.2,2.9]	[-0.4,0.3]	$0.567 \pm 0.254$	$0.180 \pm 0.201$	$0.0046 \pm 0.040$	$-0.021 \pm 0.018$	-	-	45
Alves et al. [2015]	[2.1,2.6]	[-0.4,0.2]	$0.460 \pm 0.023$	$0.231 \pm 0.009$	-	-	-	-	43
Luck [2015]	[2.2,3.2]	[-0.4,0.3]	$0.579 \pm 0.072$	$0.192 \pm 0.028$	-	-	-	-	27
Kovtyukh et al. [2006]	[2.2,2.9]	[-0.2,0.2]	$0.103 \pm 0.031$	$0.538 \pm 0.024$	$-0.062 \pm 0.005$	$-0.003 \pm 0.012$	$-0.163 \pm 0.071$	-	12
<i>PASTEL</i>	[2.1,3.1]	[-0.5,0.4]	$0.441 \pm 0.016$	$0.240 \pm 0.007$	-	-	-	-	96

**Table 3.7** – Best model, outliers, and computational time of the  $[\theta_{\text{eff}}, H_p - K]$  relation

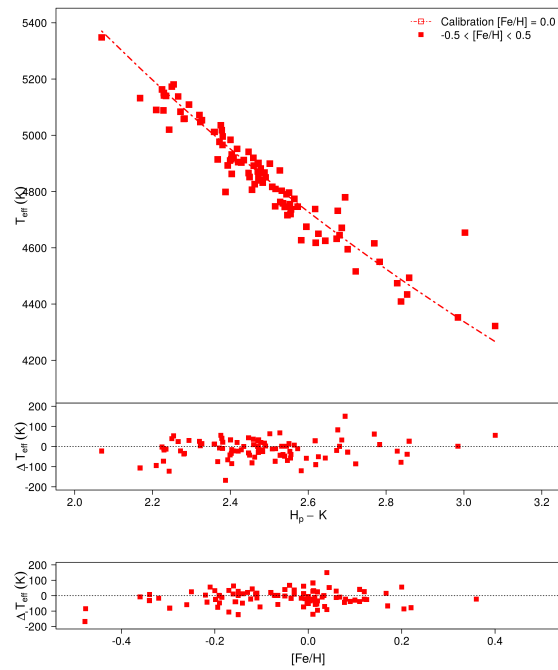
Sample	Model	Outliers (%)	Computational time (min)
<i>APOGEE</i>	4	0.0	12.3
Alves et al. [2015]	1	0.0	10.9
Luck [2015]	1	0.0	6.2
Kovtyukh et al. [2006]	6	25.0	11.0
<i>PASTEL</i>	1	4.0	116.3

each individual sample. We chose Luck [2015] as reference because, although it provides photometric temperatures rather than spectrometric, the sample is perfectly homogeneous, the dispersion is very small, it covers a larger range of  $H_p - K$  than the other homogeneous samples, and the calibration obtained is very good. All these properties were important in order to test the calibrations and develop the methods to characterise the RC (see next chapter 4). With Gaia data we will be able to use bigger samples with only spectrometric measurements and larger ranges of colour and metallicity.

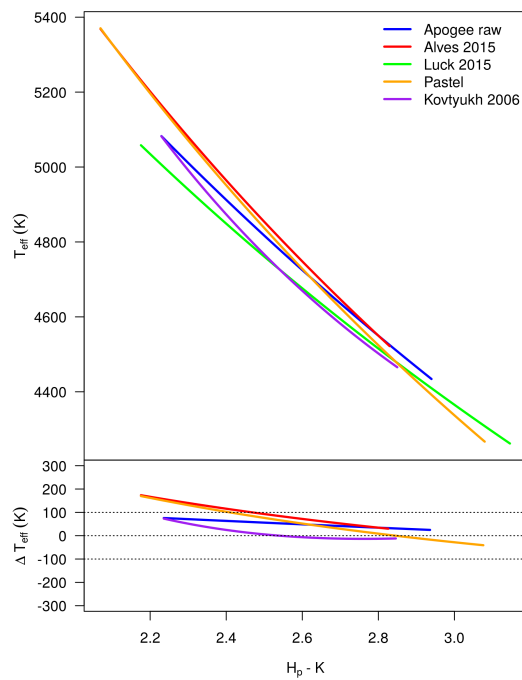
We find quite good agreement between Kovtyukh et al. [2006] and Luck [2015] with differences reaching the 80K only in the blue edge of the curve, and also with *APOGEE* with a mean difference of 50K. Alves et al. [2015] and *PASTEL* show higher discrepancies and even a different slope, although they agree among them. We remind that *PASTEL* is a compilation of effective temperatures derived through different techniques, thus the sample is less homogeneous which may induce to less accurate and reliable results. This proves the importance of having a completely homogeneous of high quality. All these results are in agreement with the ones discussed in next section 3.4, where we compare our relations with other from the literature.



**Figure 3.11** – Empirical effective temperature -  $H_p - K$  calibrations for the selected samples: (top left) *APOGEE*, (top right) *Alves et al. [2015]*, (bottom left) *Luck [2015]*, (bottom right) *Kovtyukh et al. [2006]*, (next figure 3.12) *PASTEL* (see text for details). The red squares belong to our sample of stars in the metallicity range of  $-0.5 \leq [\text{Fe}/\text{H}] \leq 0.5$ . The red dotted line correspond to our calibration at fixed  $[\text{Fe}/\text{H}] = 0.0$ . At the bottom of each calibration plot: residuals of the fit as a function of  $H_p - K$  and  $[\text{Fe}/\text{H}]$ . All plots are scaled to the same colour, metallicity and temperature intervals



**Figure 3.12** – Same as previous figure for *PASTEL*



**Figure 3.13** – Comparison between the five effective temperature -  $H_p - K$  calibrations obtained in this work: *APOGEE* (blue), *Alves et al. [2015]* (red), *Luck [2015]* (green), *PASTEL* (orange) and *Kovtyukh et al. [2006]* (purple). Residuals are derived with respect to *Luck [2015]* and within its range of  $H_p - K$  colour as long as it falls within the range of  $H_p - K$  of each individual sample calibration

## 3.4 Discussion

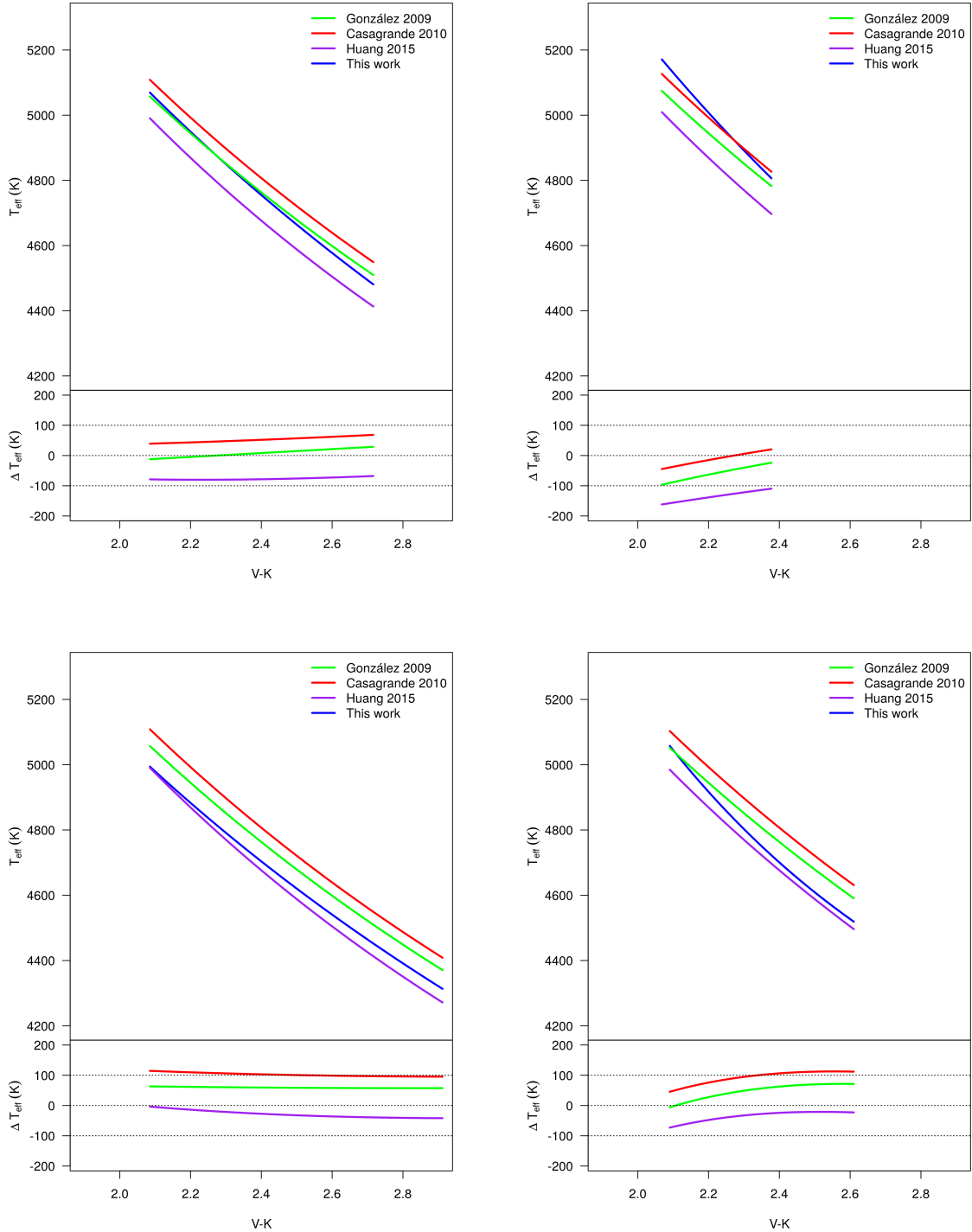
### 3.4.1 Comparison with other $T_{\text{eff}}$ -Colour relations

In order to test the metallicity-dependent  $T_{\text{eff}}$ C calibration derived in the current work, we took advantage of the already existing effective temperature relations provided by some of the previous studies mentioned at the beginning of section 3.3. For each of the five samples used in this section we determined the  $T_{\text{eff}}$  at solar metallicity ( $[\text{Fe}/\text{H}] = 0.0$ ) using our  $[T_{\text{eff}}, H_p - K]$  calibration. None of the  $T_{\text{eff}}$  relations in the literature uses  $H_p - K$  colour. However, the  $V$  photometric band is quite close to  $H_p$ . Therefore it is pretty feasible to compare our  $[T_{\text{eff}}, H_p - K]$  with other  $[T_{\text{eff}}, V - K]$  relations. To do so, we generated a set of  $H_p - K$  colour indices and metallicities within the range of applicability of each of our samples, and we obtained the  $V - K$  colour through the  $[V - K, H_p - K]$  calibration derived in section 3.2.3 (table 3.2). The effective temperatures were calculated through the  $[T_{\text{eff}}, V - K]$  from three different studies:

- **González Hernández and Bonifacio [2009]**: derived a new effective temperature scale for FGK stars through the infrared flux method (IRFM), by using the *2MASS* catalogue and theoretical fluxes computed from ATLAS models. Their  $T_{\text{eff}}$ -colour calibrations obtained with these temperatures are especially meant to be good for metal-poor stars. The calibrations were done using Johnson-Cousins  $BV(\text{RI})$ , the *2MASS*  $JHK_s$  photometric bands and the Strömgen  $b - y$  colour index
- **Casagrande et al. [2010]**: using also the IRFM, they derived effective temperatures and compared them to the **Ramírez and Meléndez [2005]** temperature scale (derived from IRFM as well). They also determined  $T_{\text{eff}}$ -colour calibrations from their own temperature scale. The  $T_{\text{eff}}$  obtained and used for the calibrations apply to solar twins dwarfs and subgiants, from super-solar metallicities to the most metal-poor stars. They agree with spectroscopic effective temperature scales at solar metallicities. The calibrations were also done using Johnson-Cousins  $BVR_{IC}$ , the *2MASS*  $JHK_s$  and the *Tycho*  $B_T V_T$  photometric bands
- **Huang et al. [2015]**: provided empirical metallicity-dependent effective temperature colour calibrations for dwarfs and giants based on a collection from the literature of about two hundred nearby stars (including 54 giants) with direct interferometry effective temperature measurements. Their giant's calibrations are valid for an effective temperature range of 3100 - 5700 K and spectral types K5 to G5. The calibrations were done using Johnson UBVRIJHK, the Cousins  $I_C R_C$ , the *2MASS*  $JHK_s$  and the *SDSS*  $g, r$  photometric bands

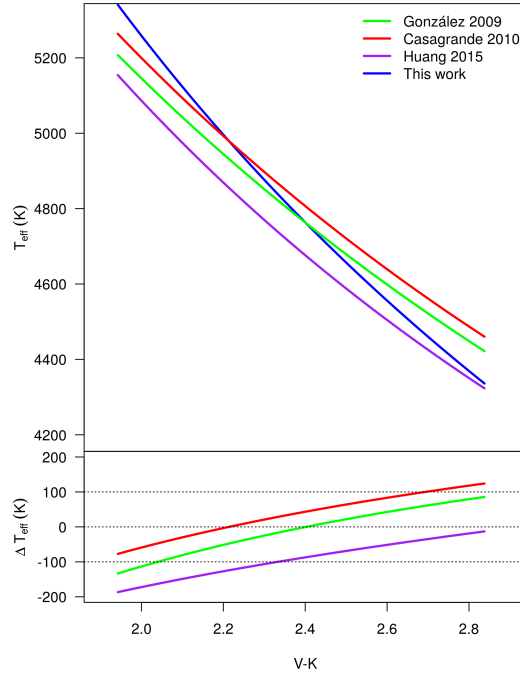
Fig. 3.14 and 3.15 compare these various  $[T_{\text{eff}}, V - K]$  literature relations with those derived in the current work. All diagrams are scaled to the same temperature and  $H_p - K$  intervals, and the calibrations tested in the range of applicability of our samples (shown in table 3.6).

The predicted  $T_{\text{eff}}$  are different depending on the input sample used for our calibration. For the **Alves et al. [2015]** and the *PASTEL* samples there are significant discrepancies respect to all authors, and they become larger for both blue (e.g.  $V - K < 2.15$  and 2.1, respectively) and red colours (e.g.  $V - K < 2.35$  and 2.6, respectively), the slope of the curve being different. For intermediate colours, the agreement is better with the **Casagrande et al. [2010]** relation and gets worse with **Huang et al. [2015]**, between 100 and 200K difference. For *APOGEE* the agreement is good with **González Hernández and Bonifacio [2009]**. However, **Holtzman et al. [2015]** compared the spectroscopically determined temperatures of *APOGEE* (those used in our sample, raw data) with the photometric calibration of **González Hernández and Bonifacio [2009]**, and found an offset of about  $\sim 50 - 90\text{K}$ . Hence the calibration from **González Hernández and Bonifacio [2009]** should differ more in this case. We suspect that this is due to a possible overestimation of the extinction in their reddening corrections (see section 4.3 of their article). The discrepancies of our calibration with respect to **Casagrande et al. [2010]** and **Huang et al. [2015]** are about 50K and 90K, respectively.



**Figure 3.14** – Comparison of the  $T_{\text{eff}}$ -colour relation for giants of solar metallicity from the current work (blue) with that from [González Hernández and Bonifacio \[2009\]](#) (green), [Casagrande et al. \[2010\]](#) (red) and [Huang et al. \[2015\]](#) (purple), for the five selected samples: (top left) *APOGEE*, (top right) [Alves et al. \[2015\]](#), (bottom left) [Luck \[2015\]](#), (bottom right) [Kovtyukh et al. \[2006\]](#) and (next figure 3.15) *PASTEL* (see text for details). The  $T_{\text{eff}}$  for this work was derived through the  $[T_{\text{eff}}, H_p - K]$  calibration, whereas those from [González Hernández and Bonifacio \[2009\]](#), [Casagrande et al. \[2010\]](#) and [Huang et al. \[2015\]](#) by using their  $[T_{\text{eff}}, V - K]$  relations. In these cases, the values of  $V - K$  were calculated using the  $V - K$  vs  $H_p - K$  calibration obtained in section 3.2.3 (table 3.2) for a set of  $H_p - K$  colours and metallicities within the range of applicability of each sample of the current work. Residuals respect to [Casagrande et al. \[2010\]](#) are shown at the bottom of each plot

Finally, for [Luck \[2015\]](#) and [Kovtyukh et al. \[2006\]](#) samples, the differences of temperature are quite homogeneous in the whole colour range, especially for [Luck \[2015\]](#), and the agreement with all authors is



**Figure 3.15** – Same as previous figure for *PASTEL*

rather good, being more significant for [Casagrande et al. \[2010\]](#) ( $\sim 50 - 100K$ ) and less for [Huang et al. \[2015\]](#) ( $\sim 20 - 50K$ ). These results are in agreement with the homogeneous temperature distribution of this set of stars, as seen in fig. 3.10.

Differences with [Casagrande et al. \[2010\]](#) can be globally explained by the fact that their calibrations were performed on dwarfs and subgiant stars. For [Huang et al. \[2015\]](#) the discrepancies are due to the different techniques used to determine the temperatures: spectroscopy for our stars and interferometry for their stars.

Considering all the above, in general we have good agreement with the literature and the accuracy achieved is very good, meaning that the sample selection process was really rigorous and that the implemented method is robust.

The several constraints required to have a high quality photometric and spectrometric set of stars, without contamination of binaries or multiple systems, nor effects of extinction, and mainly an homogeneous effective temperature, limit to a large degree the size of the sample. As seen, the results may vary considerably depending on the chosen input. Although having more data should somehow increase the accuracy of the fit, homogeneity is primordial. The first Gaia data will allow us to respect all these constraints for a much larger set of objects and, thus, to provide a better calibration within wider colour and metallicity ranges of applicability. Especially we will take advantage of the Gaia spectrophotometric standard stars survey [Pancino et al. \[2012\]](#) (a sample of  $\sim 200$  stars carefully selected for the absolute flux calibration of Gaia spectra and integrated magnitudes), as well as the Gaia FGK benchmark stars [[Heiter et al., 2015](#)].





# Deriving RC physical parameters

## Contents

---

<b>4.1</b>	<b>Introduction</b>	<b>106</b>
<b>4.2</b>	<b>Interstellar extinction coefficients model</b>	<b>106</b>
<b>4.3</b>	<b><math>[A_0, T_{\text{eff}}, G_0]</math>-method</b>	<b>106</b>
<b>4.4</b>	<b>Applications</b>	<b>110</b>
4.4.1	Galactic Bulge	110
4.4.2	APOGEE survey	114
4.4.3	Applications with TGAS	114
4.4.4	Discussion	117

---

## 4.1 Introduction

As already explained in chapter 1.2, RC stars are known to be good standard candles thanks to their almost constant luminosity, its global independence from metallicity and age, and the large number of RC stars in the solar neighbourhood. They are clearly a powerful tool for distance determinations to large scale structures.

However, to have accurate distances and avoid wrong or spurious results, one needs to take into account and handle the effects of interstellar extinction. Especially for sky regions with high levels of reddening, like the Galactic Bulge, these corrections are extremely important.

Throughout this thesis we have used different methods to determine the interstellar extinction for our samples, in order to keep only stars with  $\sim 0$  extinction: 3D and 2D maps, and estimates from the Diffuse Interstellar Band from *FEROS* and *SOPHIE* spectra. In this chapter we propose an empirical photometrically-based method to derive extinctions by combining the established CC and  $T_{\text{eff}}C$  empirical calibrations together with the well-known astrophysical relation between observed and intrinsic magnitudes (eq. 4.1):

$$m_{\text{observed}} = m_{\text{intrinsic}} + k_m A_0 \quad (4.1)$$

Thus, given  $G$  and  $K$  bands plus others available magnitudes, and the metallicity of a star, the method automatically provides the interstellar extinction  $A_0$ <sup>1</sup>, from which we may derive the  $T_{\text{eff}}$  and the intrinsic magnitude  $G_0$ .

## 4.2 Interstellar extinction coefficients model

To take into account the dependency of the extinction coefficients on the star Spectral Energy Distribution and on the extinction itself, we used an analytical model of those dependencies. C. Babusiaux computed the extinction coefficient using the Fitzpatrick and Massa [2007] extinction law for extinctions  $A_0$  from 0 to 5 and using Kurucz spectra for a  $\log g$  of 2.5 with  $T_{\text{eff}}$  from 4000 to 6500 K using ASPOS<sup>2</sup> (simulator developed by P. Sartoretti, private communication). She modelled the dependency of the extinction coefficient  $k$  (eq. 2.8) with  $A_0$  (absorption at a given wavelength, here the Gaia reference value of  $\lambda = 550\text{nm}$ ) and  $x$  ( $x$  being either  $T_{\text{eff}}$  or a colour) as:

$$k = a + b x + c x^2 + d x^3 + e A_0 + d A_0^2$$

The broader the photometric band, the higher those effects will be. Figure 4.1 shows the results for  $k_G$  as a function of  $T_{\text{eff}}$ : in black the coefficients derived from ASPOS and in green the values provided by this fit. The top line corresponds to  $A_0 = 1$ , the bottom line to  $A_0 = 5$ . The fit is not perfect but the formulae is simple to use and fit all  $T_{\text{eff}}$  and colours within the range of uncertainty coming from the template spectra and the extinction law itself.

## 4.3 $[A_0, T_{\text{eff}}, G_0]$ -method

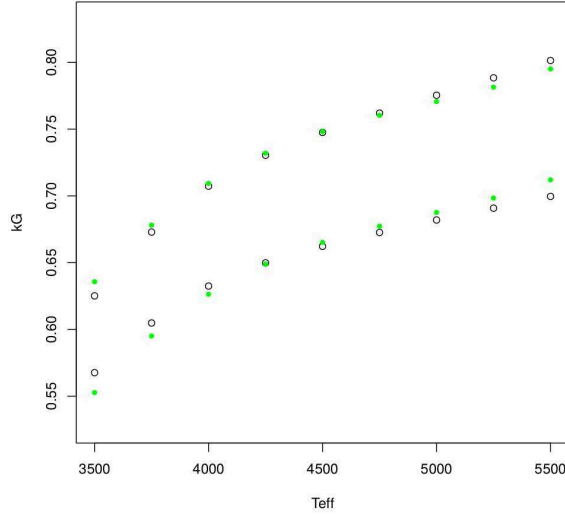
We can estimate the extinction  $A_0$  by computing the probability:

$$P(\Theta | \tilde{x}_1, \dots, \tilde{x}_n) = \frac{f(\tilde{x}_1, \dots, \tilde{x}_n | \Theta) P(\Theta)}{P(\tilde{x}_1, \dots, \tilde{x}_n)} = \frac{P(\{\tilde{x}\} | A_0) P(A_0)}{P(\{\tilde{x}\})} \propto P(\{\tilde{x}\} | A_0) P(A_0)$$

where  $\Theta$  are the parameters to estimate ( $A_0$  here) and  $\{\tilde{x}\}$  are the observed magnitudes.

<sup>1</sup>Absorption at  $\lambda = 550 \text{ nm}$  (Gaia reference value)

<sup>2</sup>Astronomic SPectra Observation Simulation



**Figure 4.1** – Interstellar extinction coefficients model for  $k_G$  as a function of  $T_{\text{eff}}$ . Black circles belong to the coefficients derived from ASPOS, green points to the values provided by the current fit. The top curve corresponds to a  $A_0 = 1$ , while the bottom curve to  $A_0 = 5$ . Figure provided by C. Babusiaux (private communication)

Given that each observable  $\tilde{x}_\lambda$  is measured independently we may write the last expression as the product of probabilities for each one:

$$P(A_0|\{\tilde{x}\}) \propto \prod_{\lambda} P(\tilde{x}_\lambda|A_0) P(A_0)$$

And since the magnitudes are normally distributed:

$$P(A_0|\{\tilde{x}\}) \propto \prod_{\lambda} \mathcal{N}(\tilde{x}_\lambda, x_\lambda, \sigma) P(A_0) \quad (4.2)$$

where  $x_\lambda$  is the true value for each magnitude.

Three parameters from this expression must be first determined or established in order to be able to calculate the probability  $P$ : the true value  $x_\lambda$ , the standard deviation or error  $\sigma$ , and the prior  $P(A_0)$ .

- **True magnitude  $x_\lambda$ :**

To derive this parameter we took advantage of three aspects:

1. the  $[T_{\text{eff}}, G_0 - K_0]$  calibration:

$$G_0 - K_0 = f(T_{\text{eff}}, [\text{Fe}/\text{H}])$$

2. the  $[x_{\lambda,0} - K_0, G_0 - K_0]$  CC relation:

$$x_{\lambda,0} - K_0 = f(G_0 - K_0, [\text{Fe}/\text{H}])$$

3. and the well-known relation of equation 4.1, between the observed and intrinsic magnitudes and the interstellar extinction:

$$x_\lambda = x_{\lambda,0} + k_{x_\lambda} A_0$$

with  $x_{\lambda,0}$  the intrinsic magnitude of each observable, and  $k_{x_\lambda}$  the extinction coefficient for the given band, calculated using the Babusiaux's model.

Combining the three relations:

$$x_\lambda = f(K_0, G_0 - K_0, [\text{Fe}/\text{H}]) + k_{x_\lambda} A_0 = f(T_{\text{eff}}, G_0, [\text{Fe}/\text{H}]) + k_{x_\lambda} A_0$$

Given that in Babusiaux's model  $k_{x_\lambda}$  depends on  $G_0 - K_0$ , we implemented a linear model of  $k_{x_\lambda}$  as a function of  $G - K$  in order to provide a first approximated value of  $G_0 - K_0$  from  $A_0$ .

- **Global magnitude standard deviation  $\sigma$**

It can be expressed as the quadratic sum of the input observed magnitude error ( $\sigma_{\tilde{x}_\lambda}^2$ ) and the error of the calibrated intrinsic magnitude ( $\sigma_{x_{\lambda,0}}^2$ ):

$$\sigma = \sqrt{\sigma_{\tilde{x}_\lambda}^2 + \sigma_{x_{\lambda,0}}^2}$$

The first term is known from the input, whereas we calculated the second by using the trace<sup>3</sup> of the MCMC of each  $[x_{\lambda,0} - K_0, G_0 - K_0]$  CC relation. To do so we generated a normally distributed sample for  $G_0 - K_0$  and  $[\text{Fe}/\text{H}]$ , respectively, from the input values, and we applied our CC calibration for each point with the coefficients provided by the trace of the MCMC. The value of  $\sigma_{x_{\lambda,0}}^2$  therefore belonged to the standard deviation of the final calibrated colour distribution.

- **Prior for the estimated parameter  $P(A_0)$**

This prior describes the constraints over the parameters to be estimated in order to ensure the convergence. We considered  $P(A_0)$  uniform with an always positive extinction, and within a range of feasible values specified in each application of the method (to speed up the computational time).

Thus, taking again equation 4.2 and replacing the above conditions and expressions, the probability of deriving the extinction finally becomes:

$$P(A_0|\{\tilde{x}\}) \propto \prod_{\lambda} \mathcal{N} \left( \tilde{x}_\lambda, f(T_{\text{eff}}, G_0, [\text{Fe}/\text{H}]) + k_{x_\lambda} A_0, \sqrt{\sigma_{\tilde{x}_\lambda}^2 + \sigma_{x_{\lambda,0}}^2} \right) \quad (4.3)$$

To compute the uncertainty of the estimated  $A_0$  we used the High Density Intervals (HDI). That is, we keep the maximum probability of the shortest interval that contained a given percentage of the posterior probability. We chose a highest density interval percentage of 68% (which corresponds to  $1\sigma$  in the Gaussian case). The final extinction value for the given star belong to the maximum probability (highest point) of the probability density. Its uncertainty is defined as the mean absolute difference between the chosen density interval and the extinction value derived.

Following a similar process, we also determined the probability of obtaining the effective temperature given the input observed magnitudes. That is, from one of the three parameters:  $A_0$ ,  $G_0 - K_0$  or  $T_{\text{eff}}$  we can deduce the others. However, by deriving the probability instead of doing a direct calculation we can ensure, as explained just above, that the value obtained is the correct one in case of degenerated or non-symmetric probability distributions. The probability of deriving the effective temperature is:

$$P(T_{\text{eff}}|\{\tilde{x}\}) \propto \prod_{\lambda} \mathcal{N} \left( \tilde{x}_\lambda, f(T_{\text{eff}}, G_0, [\text{Fe}/\text{H}]) + k_{x_\lambda} A_0, \sqrt{\sigma_{\tilde{x}_\lambda}^2 + \sigma_{x_{\lambda,0}}^2} \right) \quad (4.4)$$

with the corresponding prior  $P(T_{\text{eff}})$  considered uniform within a range of possible values specified in each application, and null elsewhere.

Further, the method ensures that only stars within the colour range of applicability of the calibrations are used. And it also computes the p-value in order to determine whether the results are within the normal

<sup>3</sup>Variation of the coefficients' values during the convergence process of the MCMC

range of expected values. This p-value is calculated through the  $\chi^2$  of each star, obtained while accounting for all the photometric bands:

$$\chi_{star}^2 = \sum_{\lambda} \left( \frac{\tilde{x}_{\lambda} - (x_{\lambda} + k_{x_{\lambda}} A_0)}{\sqrt{\sigma_{\tilde{x}_{\lambda}}^2 + \sigma_{x_{\lambda}.0}^2}} \right)^2$$

Outliers are checked at 1% level of rejection, i.e. at a p-value limit of 0.01. If the p-value obtained for the star is below this limit, the null hypothesis is rejected, hence the results obtained for that star are not considered as good enough (i.e. the star is removed).

## 4.4 Applications

In this last section we aim to test the  $A_0$ -method in different samples: the Galactic Bulge, *APOGEE* stars, the Globular Cluster *47 Tuc*, the Corot and Kepler fields and on the TGAS HR diagram. We derive the interstellar extinctions and the effective temperature of these stars and we compare the results with the literature.

However we will not present results of all the mentioned samples because of different limitations that currently impede us to provide reliable and accurate values. These limitations are specified below in each case. Nonetheless this will be solved with the upcoming TGAS release, and that is why we present here all what is already implemented. Therefore, we first show the analysis performed with the Galactic Bulge and the *APOGEE* stars and what could be done with the current calibrations, and we then finally describe the work prepared with a view to TGAS.

### 4.4.1 Galactic Bulge

One of our biggest motivations to develop a method to derive interstellar extinctions was to be able to apply it to Bulge stars. Especially in the Baade's Window, a region of the sky located at  $b \sim -4^\circ$  (galactic latitude) that along the line of sight from the Earth shows relatively low amounts of interstellar dust. The advantage of this sky area for bulge studies is hence evident, and the presence of lots of RC stars on that direction allows to draw the bulge structure thanks to their properties as distance indicators [Babusiaux and Gilmore, 2005, Nishiyama et al., 2006].

We used a sample of 219 stars from a compilation of bulge clump giants in the Baade's Window from Hill et al. [2011]<sup>4</sup> and Chen et al. [2013]<sup>5</sup>. The authors determined, respectively, the metallicity and the diffuse interstellar bands (DIBs) observed at high spectral resolution (*GIRAFFE* spectrograph at the ESO *Very Large Telescope* (VLT)). The sample contains atmospheric parameters ( $\log g$ ,  $T_{\text{eff}}$  and  $[\text{Fe}/\text{H}]$ ) derived spectroscopically, photometry for the *U B V I J H K* (*OGLE-2MASS*) bands, the equivalent widths (EW) of three DIBs ( $\lambda = 6196, 6204, 6284 \text{ \AA}$ ), and the reddening ( $E_{(B-V)}$ ) derived from the Sumi [2004] extinction map by interpolation. Since the authors did not consider any criterion on the quality of the *2MASS* photometry, we first crossmatched both catalogues and selected only stars with high quality measurements of J and K (remember that H was not taken into account in our calibrations). We also removed all metal-poor stars from the sample, thus keeping those within the interval of metallicity of our calibrations (i.e.  $-0.5 \leq [\text{Fe}/\text{H}] \leq 0.5$ ). With these criteria, 144 stars remained.

As explained, to use the  $A_0$ -method we need the CC and  $T_{\text{eff}}C$  relations derived in chapter 3. However, in that chapter we obtained the calibration with respect to the  $H_p - K$  colour, but there is no  $H_p$  magnitude available for the bulge stars. Therefore, by using the same calibration method described in section 3.2.2, we derived the empirical calibrations for  $V - K$ . Table 4.1 summarizes the coefficients obtained for the relationships used in the  $A_0$ -method. We note that, as done in that chapter, we considered the different colour-metallicity-dependent polynomials (up to a second degree, as we remind in eq. 4.5) and we tested them by penalizing by the complex terms. Since for the I magnitude we did not have too much homogeneous and good measurements, for the IK colour we forced the MCMC to use only non metallicity-dependent relationships.

$$Y = a_0 + a_1 X + a_2 X^2 + a_3 [\text{Fe}/\text{H}] + a_4 [\text{Fe}/\text{H}]^2 + a_5 X [\text{Fe}/\text{H}] \quad (4.5)$$

To obtain the  $T_{\text{eff}}-V - K$  calibration we used the sample constructed with effective temperature information from Luck [2015] (see section 3.3.1), because of their larger colour range of applicability.

After removing outliers, 111 stars remained. We compared the estimated extinction of these stars with respect to the values provided by Sumi [2004] and to those extracted from the DIBs. To derive the latter, we

<sup>4</sup>VizieR reference: J/A+A/534/A80/

<sup>5</sup>VizieR reference: J/A+A/550/A62/

**Table 4.1** – Coefficients and range of applicability of the intrinsic colour- $V - K$  and  $T_{\text{eff}}-V - K$  calibrations for the relations used in the  $A_0$ -method. We remind that  $\theta_{\text{eff}} = 5040/T_{\text{eff}}$

Colour	$V - K$ range	[Fe/H] range	$a_0$	$a_1$	$a_2$	$a_3$	$a_4$	$a_5$	N
$B - K$	[1.9,3.1]	[-0.5,0.3]	$-0.038 \pm 0.027$	$1.455 \pm 0.011$	-	$0.116 \pm 0.014$	-	-	122
$H_p - K$	[1.9,3.1]	[-0.5,0.3]	$0.159 \pm 0.018$	$1.002 \pm 0.007$	-	-	-	-	127
$I - K$	[2.1,3.1]	[-0.5,0.3]	$0.212 \pm 0.102$	$0.361 \pm 0.083$	$0.051 \pm 0.017$	-	-	-	66
$J - K$	[1.9,3.1]	[-0.5,0.3]	$-0.004 \pm 0.02$	$0.265 \pm 0.008$	-	-	-	-	127
$\theta_{\text{eff}}$	[2.0,3.0]	[-0.4,0.3]	$0.604 \pm 0.075$	$0.195 \pm 0.031$	-	-	-	-	27

used the linear relations between the equivalent width  $EW_\lambda$  and the reddening  $E_{(B-V)_\lambda}$  detailed in table 4.2. The coefficients were taken from the table 2 of [Chen et al. \[2013\]](#). We determined then the three  $E_{(B-V)_\lambda}$  and we computed the weighted mean  $E_{(B-V)_{\text{wm}}}$ . The interstellar extinction follows:  $A_V = 3.1E_{(B-V)_{\text{wm}}}$ .

**Table 4.2** – Coefficients of the EW (units of  $\lambda$ ) and  $E_{(B-V)}$  linear relation:  $E_{B-V_\lambda} = a + b EW_\lambda$

$DIB_\lambda$	Reference	a	b
6196.0	<a href="#">Puspitarini et al. [2013]</a>	0	$(1.48 \pm 0.09) 10^{-2}$
6204.5	<a href="#">Puspitarini et al. [2013]</a>	0	$(4.58 \pm 0.37) 10^{-3}$
6283.8	<a href="#">Raimond et al. [2012]</a>	$(-0.50 \pm 0.40) 10^{-2}$	$(6.37 \pm 0.23) 10^{-4}$

Figure 4.2 shows the extinction density distribution of the three methods for the 111 stars: (from left to right)  $A_0$ -method of this work, [Sumi \[2004\]](#) map and DIBs. The median in each case is (indicated as a vertical dotted red line in the graphs):

$$\tilde{A}_{0_{\text{This work}}} = 1.29 \pm 0.03$$

$$\tilde{A}_{V_{\text{Sumi}}} = 1.35 \pm 0.01$$

$$\tilde{A}_{V_{\text{DIB}}} = 1.67 \pm 0.03$$

In general, our method tends to give lower amounts of extinction<sup>6</sup> than those determined by [Sumi \[2004\]](#), which in turn are also lower than the ones provided by the DIBs. These differences mainly come from the systematics due to the various calibrations used in each method. The agreement with [Sumi \[2004\]](#) is, however, quite good. The dispersion is higher from left to right panels ( $\sigma_{\text{This work}} > \sigma_{\text{DIB}} > \sigma_{\text{Sumi}}$ ), meaning that the uncertainties obtained are too large to be able to detect correlations at the level of each individual star. Nonetheless, compared to the procedure using the DIBs, which is clearly an analysis star per star, our results show a statistically coherent behaviour.

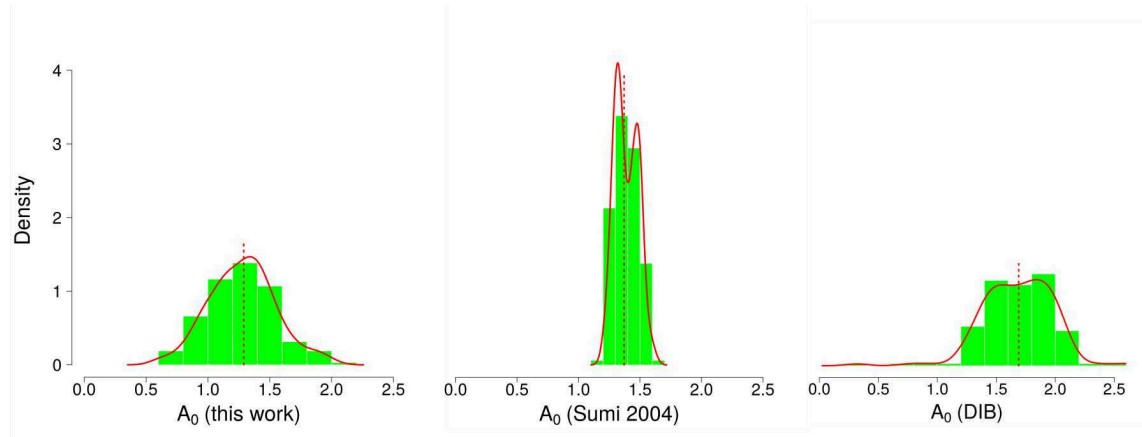
Indeed, we see in figure 4.3 the residuals of the interstellar extinction measurements between this work and [Chen et al. \[2013\]](#) (top panels) and between this work and [Sumi \[2004\]](#) (bottom panels), with respect to the metallicity (left panels) and the  $V - K$  colour (right panels). We may observe a slight correlation with metallicity in both cases. The agreement with the median of [Sumi \[2004\]](#) is again confirmed by the low residuals dispersion in this figure.

### Effective temperature

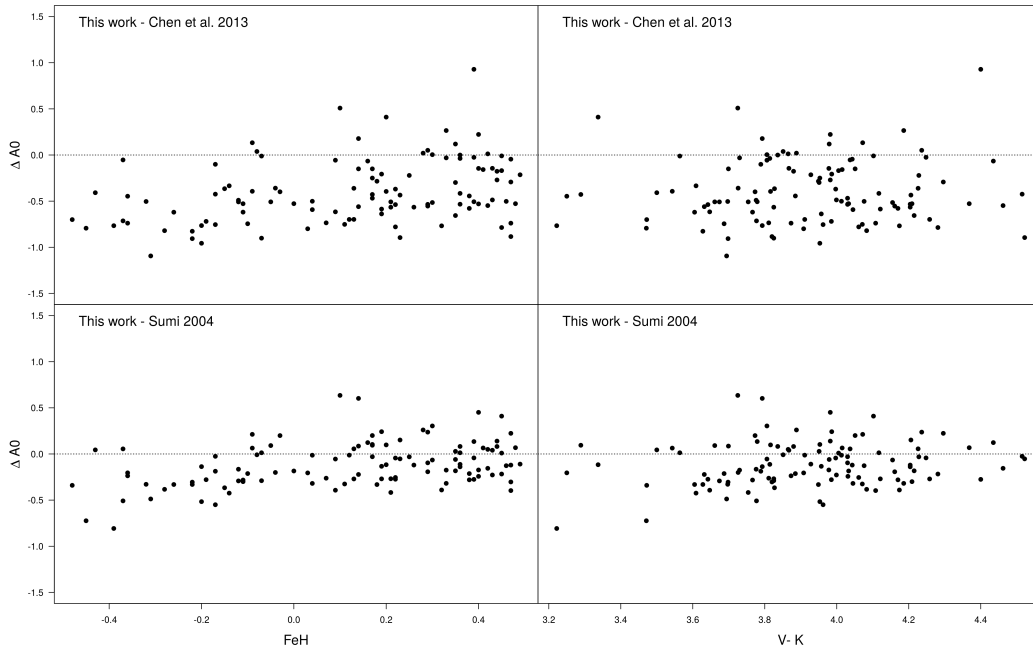
We also computed the photometric effective temperatures for these 111 stars. In figure 4.4 we compare the results obtained with respect to the spectroscopic values provided by [Hill et al. \[2011\]](#) and the ones available in the *Gaia-ESO Survey* (GES). We also show the relation between these two, i.e. [Hill et al. \[2011\]](#) and GES.

As we found with the extinctions, in general our estimated temperatures tend to be lower than those from the literature. The median absolute difference is about 310K with [Hill et al. \[2011\]](#) whereas with

<sup>6</sup>We note that although  $A_V$  is not  $A_0$  they are very close, hence we may compare both



**Figure 4.2** – Density distribution of the  $A_0$  derived in this work (left), that from the extinction map of [Sumi \[2004\]](#) (middle), and that from the DIBs (right). We remind that the distributions of [Sumi \[2004\]](#) and the DIBs are  $A_V$  distributions and not  $A_0$  distributions, but that they are very close (to facilitate the graphs reading we put  $A_0$  in the axis for all). The vertical dotted red line in each histogram indicates the location of the median. The three graphs are plotted with the same scale

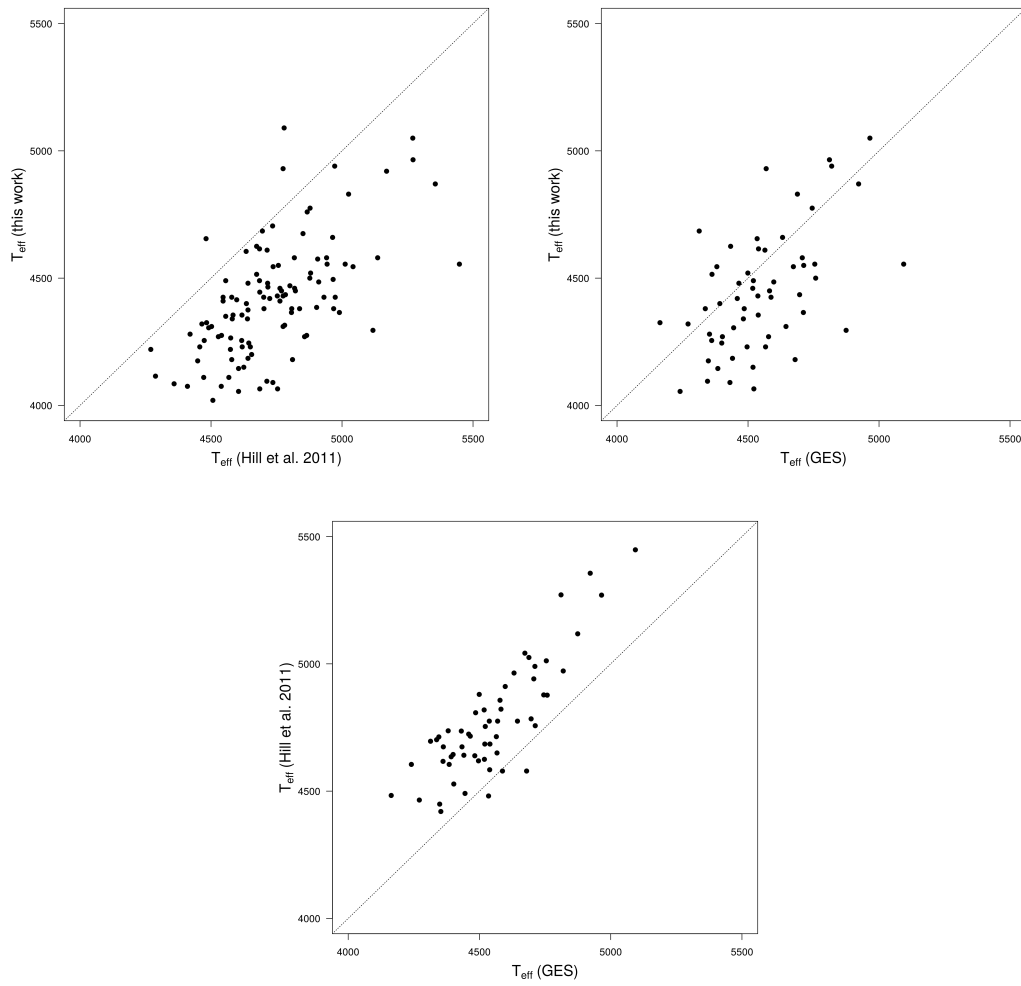


**Figure 4.3** – Residuals of the interstellar extinction measurements between (top panels) this work and [Chen et al. \[2013\]](#), and (bottom panels) this work and [Sumi \[2004\]](#), with respect to (left panels) the metallicity and (right panels) the  $V - K$  colour

GES is about 120K. We see in the third panel of figure 4.4 that there is also a median shift of  $\sim 240$ K between both [Hill et al. \[2011\]](#) and GES measurements. We remind as well that in the current work we derived photometric temperatures that we are comparing to spectroscopic, i.e. systematic differences may be found.

Finally we note that, before the work performed in this thesis, C. Babusiaux and V. Hill also tried to derive the effective temperatures of these stars but by taking advantage of the *Padova* isochrones. However, although they considered several estimations of the interstellar extinction, the p-value check always removed the ensemble of the stars, meaning that none of the values obtained for the temperature was compatible with the photometry. With the present method, only the 23% of the stars gave a p-value outside the range of expected values (p-value  $< 0.01$ ).



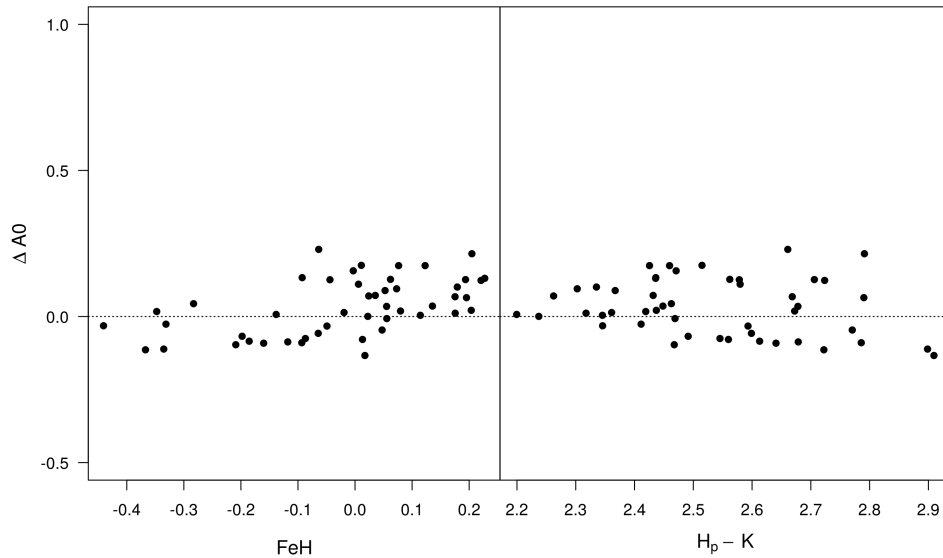


**Figure 4.4** – Effective temperature measurements comparison between (top left) this work and Hill et al. [2011], (top right) this work and the *Gaia-ESO Survey*, (bottom) Hill et al. [2011] and the *Gaia-ESO Survey*. The three graphs are plotted with the same scale

### 4.4.2 APOGEE survey

We selected 139 red giant stars from the *APOGEE* survey. These stars have *Hipparcos* and *2MASS* high quality photometry,  $H_p - K$  intrinsic colour within the range of the CC calibrations (see table 3.2 in section 3.2.3), spectroscopic effective temperatures and solar metallicity ( $-0.5 \leq [\text{Fe}/\text{H}] \leq 0.5$ , raw  $[M/H]$  data).

48 stars remained after the p-value clipping. Figure 4.5 shows the residuals between the estimated extinction for these stars using our  $A_0$ -method and the values provided by the 3D maps of [Lallement et al. \[2014\]](#), with respect to the metallicity and the observed  $H_p - K$  colour. The low dispersion highlights the goodness of the results, although we may also see a slight correlation with metallicity, as observed with the Bulge.



**Figure 4.5** – Residuals between the estimated interstellar extinction and the one given by the 3D maps, with respect to the metallicity and the  $H_p - K$  colour for the *APOGEE* red giant stars

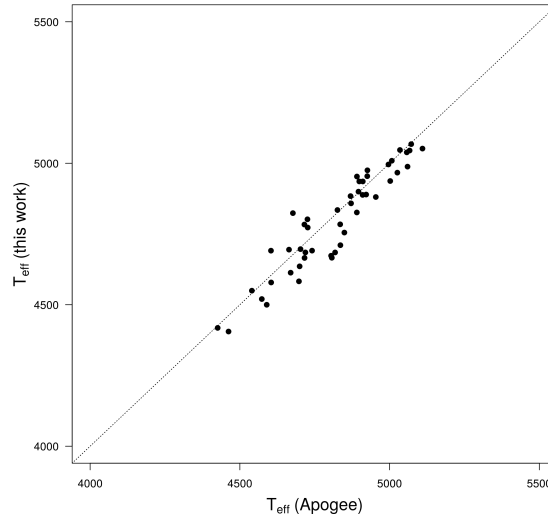
#### Effective temperature

Figure 4.6 shows the estimated photometric temperatures obtained with our method, compared to the raw spectroscopic measurements provided in *APOGEE*.

The absolute median discrepancies are about 27K, with the 85% of the stars within  $1\sigma$  confidence interval, and the 98% within  $2\sigma$ . The results are indeed good, although larger differences may be expected because of both the nature of the measurements (spectroscopy versus photometry) and the reddening treatment. However, for the sample used here, it is worth to notice that despite of the  $A_0$  dispersion observed in fig. 4.5, since the *APOGEE* stars are not too affected by extinction, the impact of this dispersion on the determination of the effective temperature is low.

### 4.4.3 Applications with TGAS

All the work done in this thesis is part of the Gaia mission. As previously stated, the methods to better characterise the RC were implemented with a view to its application with the Gaia data, while taking advantage of the large amount of data and precision that the satellite is providing. Given that the end of this thesis and the first Gaia publication finally overlap, we could not provide full results for the planned applications. These are described below, and we expect to use them on TGAS data soon.



**Figure 4.6** – Effective temperature comparison between the values obtained in this work and those from the *APOGEE* raw data, for the *APOGEE* red giant stars

#### 47 Tuc

47 Tuc (NGC 104) is one of the most massive globular clusters of the Galaxy, located at high galactic latitude ( $b = -44.9$ ) and at a relatively small distance from the Sun. These characteristics make it a particularly interesting object for several studies. Indeed it contains a larger number of different objects, from binary systems, cataclysmic variables, white dwarfs and blue stragglers, to X-ray binaries, neutron stars and many pulsars. And obviously it also contains a significant number of red giant stars.

We selected a sample of 50 red giant stars with derived spectroscopic abundances from [Cordero et al. \[2014\]](#)<sup>7</sup>. We then crossmatched this catalogue with the *Simbad* database to obtain the information for the  $B V J H K$  (*Johnsons-2MASS*) photometric bands. We kept only those with high photometric quality and dereddened  $V - K$  values within the range of applicability of our calibrations. To obtain these dereddened values (just as a first estimation in order to select the sample), we used the value of  $E_{(B-V)} = 0.04$  [[Salaris et al., 2007](#)] for all the stars.

The metallicity range of applicability of our calibrations is a condition that clearly constraints their use. 47 Tuc has a mean metallicity of  $\sim -0.7$ , and the selected stars are within the interval  $-0.95 < [\text{Fe}/\text{H}] < -0.65$ . This limitation impedes us, for the time being, to apply the  $A_0$ -method to derive the stellar parameters of this cluster and compare them with the already well-known values. With TGAS this will not be a problem any more and we will be able to fully test the method with these stars.

#### Kepler and CoRoT fields

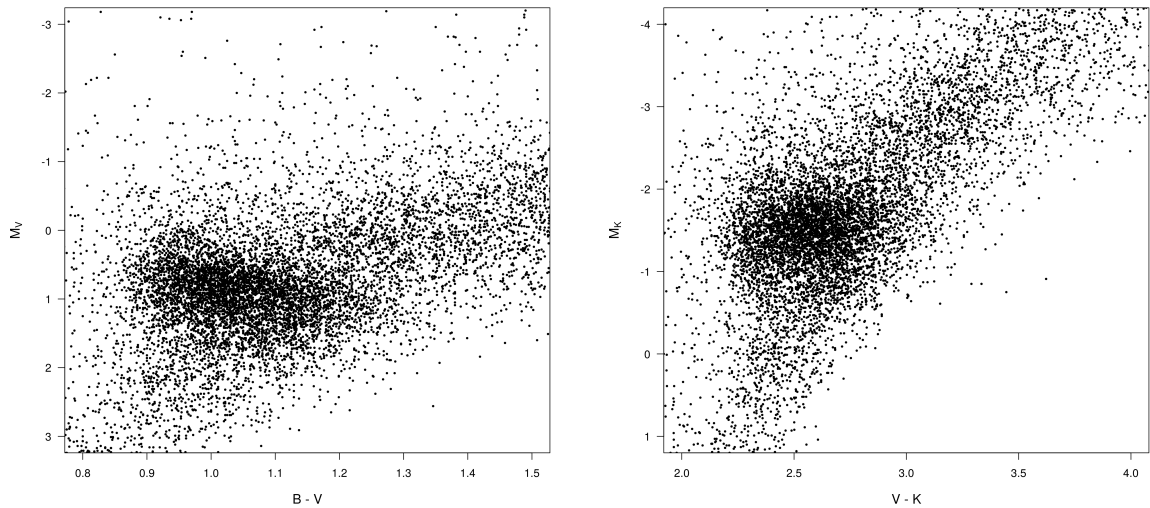
Kepler and CoRoT (*Convection, Rotation and planetary Transits*) are two space missions from NASA and CNES/European partners, respectively, devoted to exoplanetary research. Apart from planet-finding purposes, both missions also allowed to study the stellar oscillations with high precision (asteroseismology). That is, to determine the small and periodic variations in the brightness of the stars, which may be used to accurately analyse and constraint their inner structure: stellar masses, radii, surface gravity,... As we saw in section 1.3 (figures 1.4, 1.5 and 1.9), a good knowledge of these parameters allows to clearly disentangle the different red giant's populations, such as the red giant branch stars, the RC stars and even the secondary RC stars. We have therefore special interest in calibrating and photometrically characterising RC stars in those fields, where moreover the extinction is not negligible at all.

<sup>7</sup>VizieR reference: J/ApJ/780/94/

The sample selection in this case is thus based on asteroseismic constraints ( $0.5 < \log g < 3.5$ ), spectroscopic metallicity and photometric quality. The new photometric temperatures are compared to those derived spectroscopically from *APOGEE*, *LAMOST* and the *Gaia-ESO Survey*. However we are currently limited by the available photometry information for these stars, needed for the  $A_0$ -method and to obtain the effective temperature. We need hence to crossmatch these stars with TGAS data to get the Gaia G magnitude.

### TGAS versus Hipparcos RC

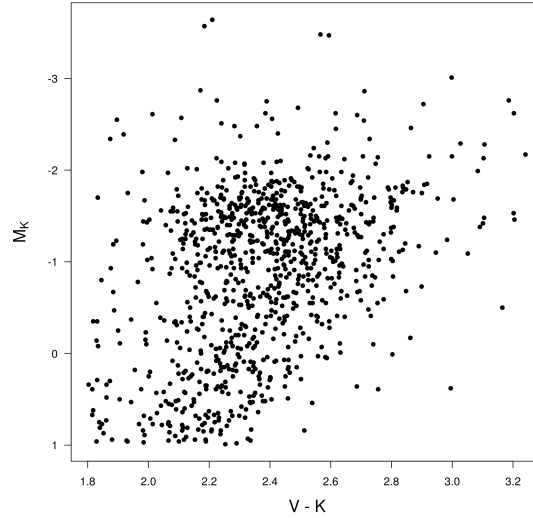
The high accuracy achieved with the *Hipparcos* satellite allowed, twenty years ago, to identify with an unprecedented precision and clearness the RC in a colour-magnitude diagram (figure 4.7). It also permitted to detect other RC components, such as the vertical and the secondary RCs (the known RC *fine structure* described in section 1.2.1). With a precision of one hundred times better, we expect with Gaia to unravel even more this stellar population. We showed in section 1.3.1 a simulated Gaia RC for stars with parallax precision  $< 10\%$ , and we saw how its shape could vary at different latitudes (fig. 1.25), depending on the metallicity, the age or the mass of the stars.



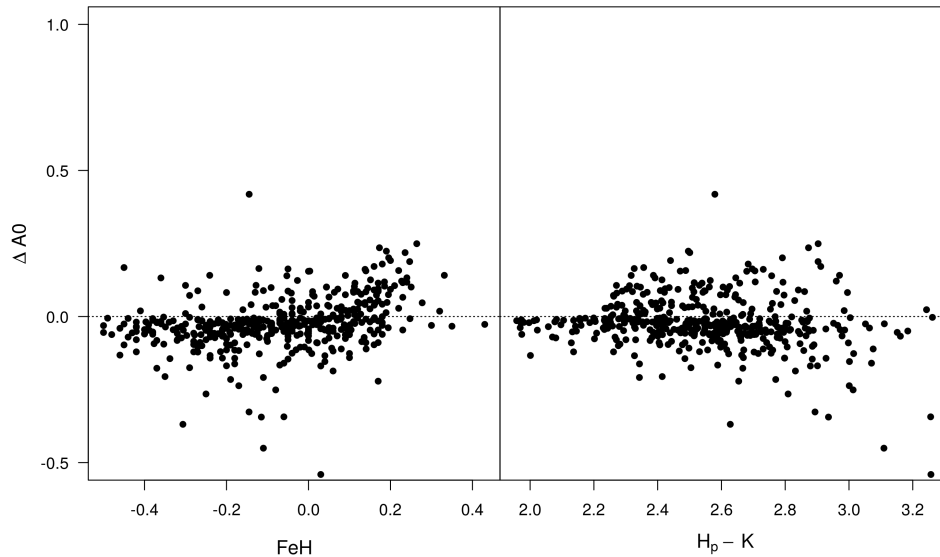
**Figure 4.7** – *Hipparcos* RC colour-magnitude diagrams: (left)  $M_V$  vs  $B - V$ , (right)  $M_K$  vs  $V - K$

Although with the first Gaia release we will still not have all this information, we will be able to compare the shape of the TGAS RC with the one of *Hipparcos*. With that purpose, we already selected a sample of 969 solar metallicity ( $-0.5 \leq [\text{Fe}/\text{H}] \leq 0.5$ ) *Hipparcos* red giant stars with high photometric quality. However, as we see in figure 4.8, these constraints result in a diagram where the RC is not well-defined, hence we may not distinguish with clarity its shape. With Gaia we will not only be able to provide calibrations with larger metallicity range of applicability (already with TGAS, which will obviously increase the sample), but we will also (in later releases) have metallicities for more *Hipparcos* stars. Both aspects will allow us to better define the *Hipparcos* RC shape and compare it to that of TGAS/Gaia.

Anyway, in order to test our method, we determined the extinction for these 969 stars, and we compare the results with the 3D maps of Lallement et al. [2014]. 491 remained after the p-value check. As expected by the maps, the extinction for these solar-neighbourhood stars is rather low. Figure 4.9 shows the residuals between the estimated extinction and the value provided by the 3D maps, with respect to the metallicity and the  $V - K$  colour. The dispersion is low, but we also observe the correlation with metallicity as with the Bulge and *APOGEE* stars.



**Figure 4.8** –  $M_K$  vs  $V - K$  HR diagram of solar metallicity *Hipparcos* red giant stars



**Figure 4.9** – Residuals between the estimated interstellar extinction and the one given by the 3D maps of Lallement et al. [2014], with respect to the metallicity and the  $H_p - K$  colour for the *Hipparcos* RC stars

#### 4.4.4 Discussion

Given that the range of applicability of our calibrations are currently limited we may still not provide quantitative outcomes, but qualitative. Statistically we got coherent results for the Bulge and the *APOGEE* stars, for both the extinction and the effective temperature. A correlation with metallicity has been identified in all the cases, and is being analysed. With the TGAS data we will be able to study it more in detail given that the amount of data will be larger.

Indeed, the upcoming Gaia data will definitely improve these results, since we will have more stars with metallicity and photometry information. This will allow us to enlarge the applicability of the calibrations, to have bigger samples for testing the  $A_0$ -method, ameliorate it, extend its applications, and to provide more quantitative results. Moreover the precision obtained with the  $A_0$ -method will be surely increased since another magnitude, G, will be added.

At this point, and given the accuracy that will be achieved, we will be able to determine the absolute

magnitude of the RC stars and determine distances, with a proper handling of the interstellar extinction, especially for the Galactic Bulge and the CoRoT and Kepler fields.

## **Part II**

# **Gaia: confrontation with external archives**





# The Gaia mission

## Contents

---

<b>1.1 An astrometric space observatory</b> . . . . .	<b>122</b>
1.1.1 Overview . . . . .	122
1.1.2 Science objectives . . . . .	123
1.1.3 The spacecraft . . . . .	124
1.1.4 Observation principle . . . . .	127
1.1.5 Detected problems and final performances . . . . .	129
1.1.6 Catalogue Releases . . . . .	131
<b>1.2 News from Gaia</b> . . . . .	<b>132</b>
<b>1.3 The Gaia Data Processing and Analysis Consortium</b> . . . . .	<b>134</b>
1.3.1 CU9: the last step . . . . .	135

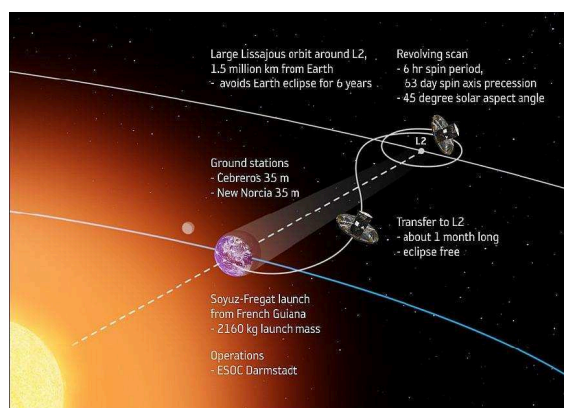
---

## 1.1 An astrometric space observatory

### 1.1.1 Overview

Gaia<sup>1</sup> is a satellite of the European Space Agency (ESA), whose main objective is to provide an unprecedented precise three-dimensional map of our Galaxy, the Milky Way, by performing a complete survey of the sky up to the magnitude  $V \sim 20$ . At the end of the mission Gaia will have observed around a billion stars with a precision of the order of a few micro-arc-seconds ( $\mu\text{as}$ ), equivalent to distinguish from Earth, to the naked eye, a human nail on the surface of the Moon, or to measure the width of a human hair located at 2000km. Ten million stars will be measured with a distance accuracy of better than 1% and a 150 million to better than 10% (at least up to 36000 ly). Compared to its predecessor, the *Hipparcos* mission, this translates to an improvement of the parallaxes and proper motions accuracy by almost 100 times and the number of stars observed increased ten thousand times.

Gaia was launched from the Europe's Spaceport in Kourou (French Guiana) by a Soyuz-Fregat vehicle on 19 December 2013 (09h12m19s UTC). The Fregat-Gaia composite was initially placed on a low-earth parking orbit, then the launcher was separated and the spacecraft injected into an hyperbolic transfer orbit (fig. 1.1) towards the Lagrange point L2 of the Sun-Earth system (a gravitationally stable virtual region in space), moment at which the in-orbit commissioning phase started. During this phase, the deployment of the solar array and the initialisation of the axis rotation were performed autonomously by the satellite, and all the science driving functions were tested and calibrated, including micropropulsion and payload.

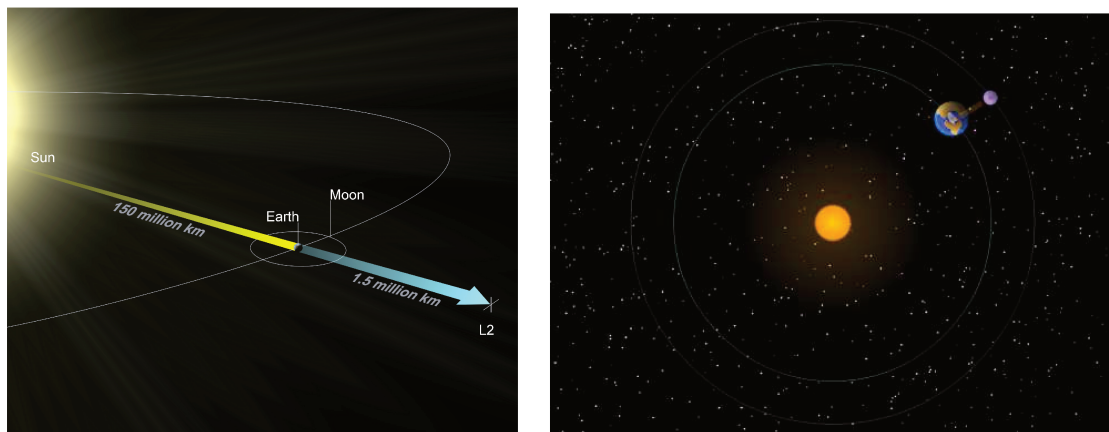


**Figure 1.1** – Gaia trajectory to its final operational orbit around the L2 Lagrange points of the Earth-Sun system. Credits from ESA

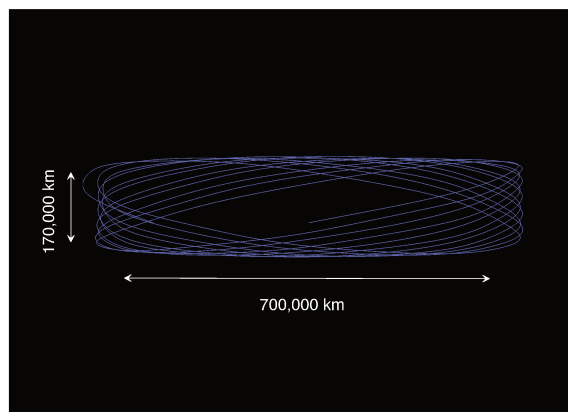
Between January 7 and January 13 2014, about one month later from the launch, the Lissajous-type orbit (orbit around L2) insertion manoeuvres took place successfully placing Gaia into its final operational orbit, at 1.5 million km from the Earth in the anti-Sun direction (fig. 1.2), free from Earth eclipses, and in a very stable thermal low radiation environment with high observing efficiency. The orbit period is about 180 days and the size of the orbit is typically 340000 x 90000 km. The name Lissajous refers to the shape of the path traced out by the orbit as seen from Earth, which moves above and below the ecliptic plane, sometimes leading and sometimes lagging the Earth (fig. 1.3). All these conditions allow an uninterrupted mapping of the sky during the whole mission.

Calibration activities over the ensemble of on-board instruments continued during weeks after the insertion, as well as the alignment and focusing of the Gaia telescopes. The satellite started routine scientific operations on 25 July 2014. The first 28 days it operated in a special scanning mode that sampled great

<sup>1</sup>The name "Gaia" was originally derived as an acronym for *Global Astrometric Interferometer for Astrophysics*, because it referred to the optical technique of interferometry planned at the very beginning. Although the acronym is no longer applicable, since the working method evolved and changed, the name was kept



**Figure 1.2** – L2 Lagrange point position and orbit of the Sun-Earth system. Credits from ESA



**Figure 1.3** – Schematic view of Gaia's Lissajous orbit around L2. Credits from ESA

circles on the sky, always including the ecliptic poles. That way it could observe the stars in those regions many times, providing an invaluable database for Gaia's initial calibration. The 5-year long nominal mission, thus its main survey operation, finally started on 21 August 2014 (consumables allocated for 6.5 years).

### 1.1.2 Science objectives

The scientific measurements of the all-sky Gaia mission cover three main areas: astrometry, photometry and spectrometry.

The immediate objectives for space astrometry are the same as those of the *Hipparcos* mission, i.e. determining the five astrometric parameters: positions at epoch, proper motions components and parallaxes. From these measurements, distances to each object as well as motions within double and multiple systems can be directly obtained. Additional parameters will be derived for other complex cases, such the orbital elements of close binary systems.

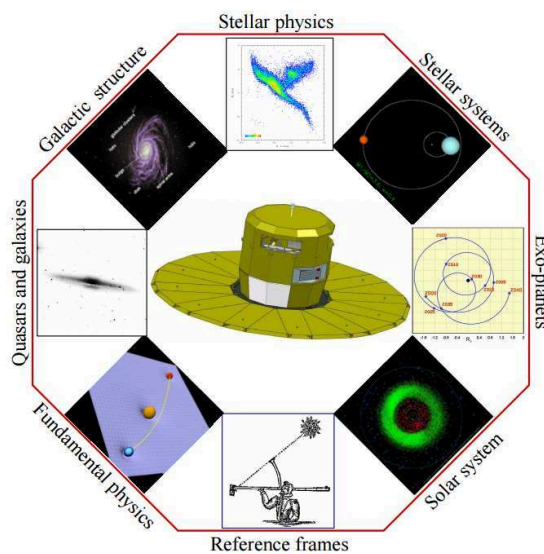
The current generation of CCD-based astrometric instruments may also be used as stable photometers to acquire photometric measurements. Thus, besides the positional information, Gaia also provides the spectral energy distribution of every object sampled, thanks to a spectrophotometric instrument which gives low-resolution spectra in the visible and infrared regions. This will allow to provide a mission-averaged G magnitude (plus a measure of its variability), colours from the red and blue bands, epoch spectrophotometry (for high variable stars), as well as other stellar parameters, such as effective temperatures, surface gravities, metallicities and interstellar reddening.

In addition, Gaia is measuring radial velocities for millions of stars with a precision starting from 1 km/s depending on the magnitude and the spectral type of the stars [Katz et al., 2004]. From these high

resolution spectra, stellar information may also be derived, as again atmospheric parameters, interstellar reddening, elemental abundances for different chemical species, and rotational velocities for stars brighter than  $V \sim 13$  mag.

Besides the  $10^9$  stars, Gaia is also observing other types of objects. At the end of the mission Gaia will have discovered around 30000 extrasolar planets (from both their astrometric wobble and from photometric transits) and their detailed orbits and masses, and observed some 500000 quasars, 20000 white dwarfs, 5000 brown dwarfs, 200000 asteroids, as well as satellites, comets, minor planets, near-Earth objects, inner Trojans and even new trans-Neptunian objects (including Plutinos), 10000 novae and supernovae explosions, 1 to 10 million of resolved galaxies, 200 microlensing events,...

Further, Gaia will follow the bending of star light by the Sun and major planets over the entire celestial sphere. This translates to a direct observation of space-time, and may reveal the long-sought scalar correction of General Relativistic light bending to its tensor form.



**Figure 1.4** – Science capabilities of the Gaia mission. Credits from ESA

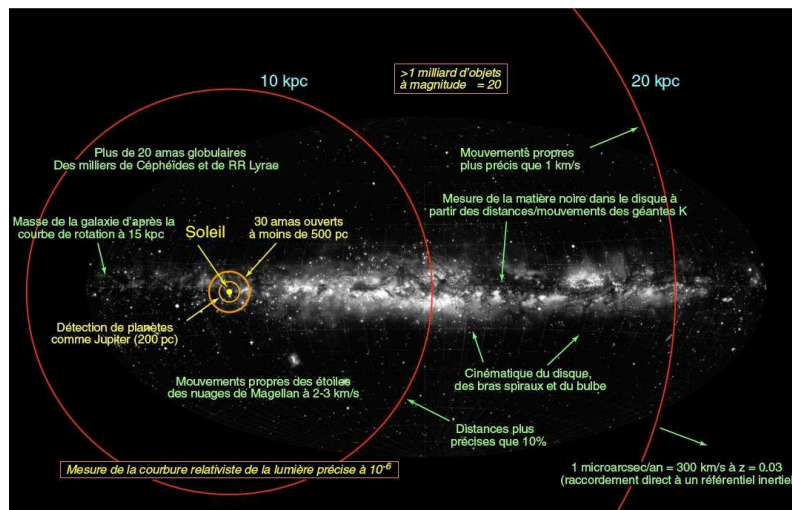
The scientific impact of the mission is therefore outstanding (fig. 1.4). With all these data, Gaia will be able to clarify the structure, kinematics, dynamics, content and evolution of our Galaxy, the star formation and internal stellar structure, as well as identify relics of tidally-disrupted accretion debris, establish a rigorous distance scale framework throughout the Galaxy and beyond, classify star formation and kinematical and dynamical behaviour within the Local Group of galaxies, probe the distribution of dark matter, observe effects predicted by General Relativity (as the deflection of the light when passing next to a very massive body), and perhaps even check if the gravity constant  $G$  of Newton's law is really constant with time (fig. 1.5).

Even if all Gaia observations will just represent a 1% of the Milky Way, Gaia will incredibly help us to tackle still unanswered questions about the origin and formation history of our home Galaxy and beyond.

### 1.1.3 The spacecraft

Technologically Gaia is one of the most precise and advanced space instrument ever built. To achieve the scientific objectives mentioned above, with high accuracy and covering the three main areas (astrometry, photometry and spectroscopy), the spacecraft carries out a payload of three main instruments:

- **Astrometric Instrument:** determines the positions of stars in the sky and provides data to track their proper motion and parallax



**Figure 1.5** – Scientific predicted performances for Gaia. Credits from GEPI Gaia team, Observatoire de Paris-Meudon

- Spectrophotometric Instrument: creates two low-resolution spectra in the red and blue ranges to provide temperature, mass and chemical composition of stars
- Radial Velocity Spectrometer (RVS): determines the velocity of a star along the line of sight of Gaia using the Doppler shift technique in a high-resolution spectrum ( $R = \lambda/\Delta\lambda \sim 11500$ ), in the narrow near-infrared band 847-874 nm.

Although these three instruments are destined to complete three different tasks, the entire payload is a single integrated and complex assembly as most equipment is shared by the three. Figure 1.6 shows the Gaia optical device. Table 1.1 details the spacecraft specifications. The main scientific shared components are the telescopes and the 106 CCD (*Charge-Coupled Device*) detectors of the focal plane.

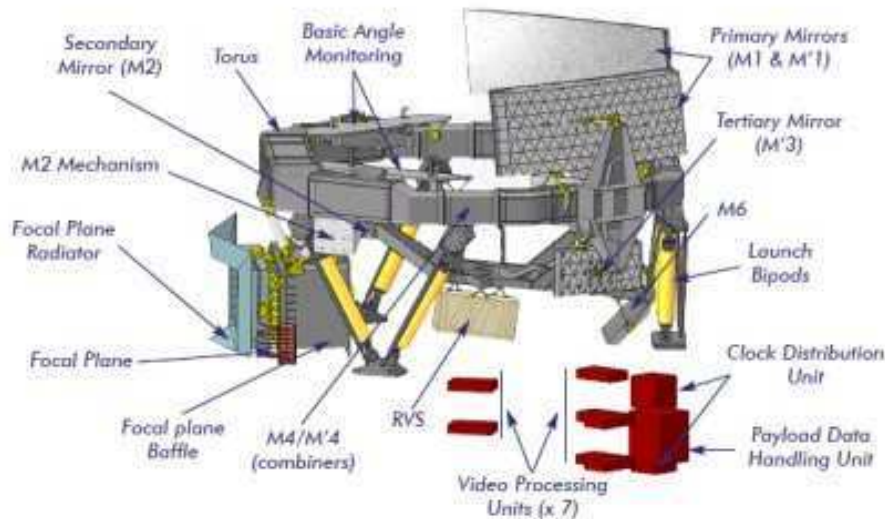
**Table 1.1** – Gaia spacecraft main specifications. Credits from ESA

Mass	2030 Kg
Sunshield diameter	11 meters
Focal length	35 meters
Primary mirrors	1.45 x 0.5 meters
Number of mirrors	10
Number of CCDs	106
Number of pixels per CCDs	4500 x 1966 pixels
Total number of pixels	938 Megapixels
Pixel size	10 x 30 micrometers
Pixel angular size	59 x 177 milliarcsec
Power consumption	1720 W
Lifetime	5 years
Cost	700 million euros

## Telescopes

To observe the stars and other celestial objects, the Gaia's science payload consists of two identical telescopes mounted on a resistant light silicon carbide toroidal structure. They are separated by 106.5 degrees in order to observe toward two different viewing directions in the sky. Indeed, global astrometry requires the simultaneous observation of two fields of view in which the star positions are measured and constantly correlated. Thus, this angular separation (known as *basic angle*) must be precisely known at every moment.

Each telescope has a primary mirror of 1.45m x 0.50m (M1 and M1'). After being reflected by the primary mirrors, the light is bounced back and forth between the M2/M2' and M3/M3' mirrors (0.65m x 0.275m) to create a total focal length of 35m. The M4/M4' mirrors act as combiners that reflect the light



**Figure 1.6** – Gaia payload module diagram. Credits from ESA

onto the focal plane assembly via the M5 and M6 folding mirrors (0.54m x 0.36m). A schematic picture of these mirrors can be found both in figure 1.6 and at the bottom of figure 1.7.

### Focal plane and CCD cameras

The focal plane towards which both viewing directions are projected consists of 106 detectors of 4500 x 1966 pixels (total amount of almost 1000 million pixels) working at  $\sim 170\text{K}$ . This focal plane is the largest focal plane ever built for a space mission (104 x 42 cm). The detectors are divided in groups depending on their functionality. From left to right in figure 1.7:

- 2 CCDs devoted to control the Basic Angle between the two telescopes, and measure possible fluctuations (Basic Angle Monitor, BAM)
- 2 CCDs devoted to control and measure the optical quality of each telescope (Wave Front Sensor, WFS)
- 2x7 CCDs distributed in two strips designed to detect real-time sources in the field of view (Sky Mapper, SM1 and SM2)
- 62 CCDs distributed in nine strips devoted to measure very accurately the position and brightness of the stars in the sky (Astrometric Field, AF). The Gaia spinning movement makes the image of the star transit through the different CCD strips, collecting more and more light
- 14 CCDs distributed in two strips of 7 CCDs, each strip associated to one spectrophotometer (Blue Photometer, BP, and Red Photometer, RP). They detect the light of the stars after passing through two different prisms, one producing a spectra in the blue range (360-680 nm) and the other in the red range (640-1050 nm). These low resolution spectra are used later to derive the astrophysical parameters of the observed objects, such as temperature, surface gravity and chemical composition
- 12 CCDs devoted to the high resolution spectrograph (Radial Velocity Spectrometer, RVS), which measures the radial velocity of objects through the radial movement of the spectral lines (Doppler effect). This allows to detect if the source is moving toward us/Gaia (blue-shift) or away from us/Gaia (red-shift)

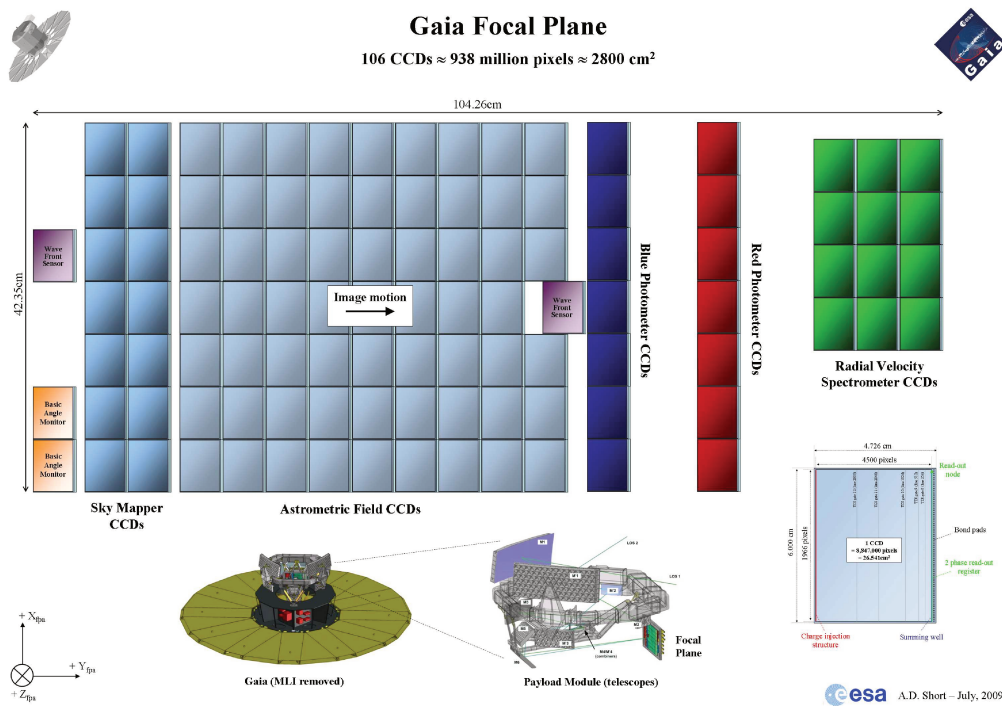


Figure 1.7 – Focal plane diagram. Credits from ESA

## Data handling

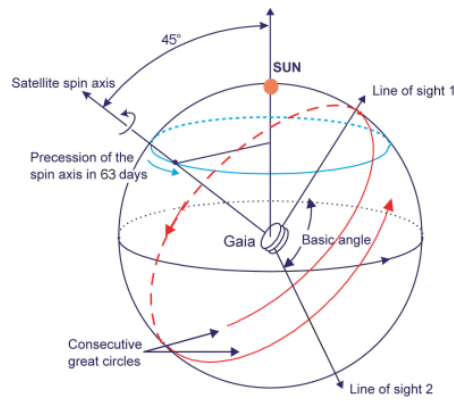
Besides these components, the spacecraft also includes on-board 7 Video Processing Units (VPU), each containing 3 PowerPC processors (for triple redundancy). They detect, process and compress the images of the sources seen through the focal plane CCDs, generating over 5Mbps on average. The on-board storage (about 100GB) stores this data, and once there is contact with the ground-station (which happens during 8 hours every day on average), it is sent to us. A special ground-based "phased array antenna" was required to receive the data, since the stability requirements on-board Gaia are so strict that a normal parabolic antenna would lead to unacceptable vibrations. These are two 35-meter antenna placed at ESA ground stations in Cebreros (Spain) and New Norcia (Australia). The whole communication system developed for Gaia constitutes one of the most powerful computing systems on-board a scientific satellite.

### 1.1.4 Observation principle

The spacecraft has a rotation axis that has an inclination angle of  $45^\circ$  with the Sun direction. This allows to constantly charge its solar panels and to have all the observational instruments always looking in the anti-Sun direction (payload shading).

Gaia rotates at a constant angular rate of  $1^\circ/\text{min}$ , which corresponds to one revolution every 6 hours. Since the spin axis is perpendicular to the field of view of the two telescope apertures, this allows them to scan across the sky, describing a great circle over the period of six hours, i.e. a single spin period. The exact spin rate is measured by on-board sensors and is important for the image read-out times on the focal plane because integration times have to be adjusted to microsecond accuracy based on actual spin rates. Thanks to the basic angle of  $106.5^\circ$  between both telescopes, an object seen by the first telescope will transit in front of the second one 106.5 minutes later (fig. 1.8). This *modus operandi* is the same astrometric measurement principle as the one used for the *Hipparcos* mission: each object position is systematically measured twice, one in each field of view. Comparing the position of the stars in different parts of the sky during its 5 years mission, a consistent and precise global map can be built.

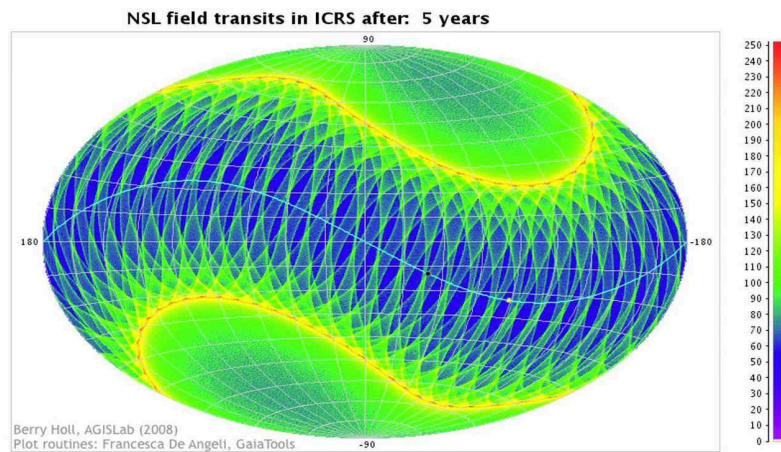
Further, to achieve an all-sky survey and not scanning always the same great circle, the rotation axis is in a slow precession motion around the Sun-to-Earth direction, producing a continuous drift of the great



**Figure 1.8** – Gaia observation principle. Credits from ESA

circle on the sky (indicated in fig. 1.8 as well). The average precession period is of 63.12 days.

The combination of both movements (axis rotation and precession) will allow Gaia to map each star from 40 to 220 times (70 times on average) during the whole mission (fig. 1.9). For all these observations, the basic angle between the two fields of view has to be constant.



**Figure 1.9** – Gaia Nominal Scanning Law. Credits from Berry Holl, AGISLab (2008)

### When Gaia detects light

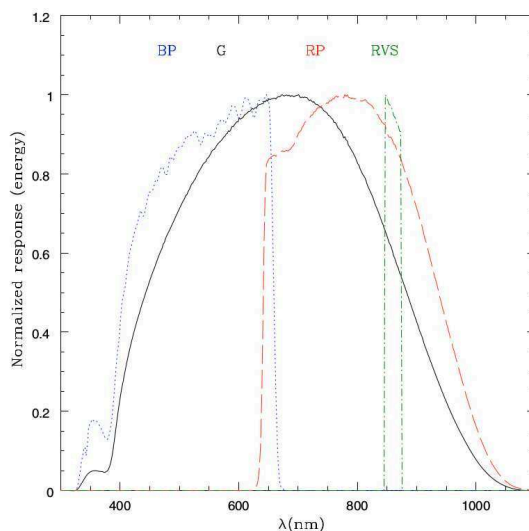
As light from a particular source brighter than  $G = 20$  gets into the instrument, it first enters the Sky Mapper portion of the focal plane, being registered by SM1 or SM2 (depending on which telescope detected it). As stated above, the spinning movement makes the image of each observed source transit through the different CCD strips. Hence objects that are registered by the Sky Mapper are then tracked as they cross the focal plane, allowing the individual instrument CCDs to extract the information of the detected sources.

After passing the Sky Mapper, the sources are detected by the AF (Astrometric Field) CCDs. The first CCD column is used to verify that the SM registered actual objects in order to filter out false readings (due for instance to cosmic rays). The image then progressively crosses the next eight AF CCD columns. The data obtained from this astrometric instrument allows to measure image centroids and thus the relative separation of the thousands of stars simultaneously present in the combined field of view. The five-year period of the Gaia scanning process will allow to provide, after a high complex and accurate ground-based data reduction (which includes geometrical calibration of the focal plane, calibrations of instrument attitude and basic angle, and corrections for chromatic shifts, aberration and relativistic light bending), the five main astrometric parameters: the position components ( $\alpha, \delta$ ), the two from the proper motion ( $\mu_\alpha, \mu_\delta$ ), and the parallax ( $\varpi$ ). The already mentioned additional parameters, for objects such as binaries, extra-solar planets



and solar system objects, can be determined by making multiple observations. And apart from stars, the astrometric instrument is also able to detect quasars, near-Earth objects, asteroids and supernovae, as long as they are brighter than the capabilities of the SM, i.e. about  $G = 20.7$ .

All the objects analysed by the astrometric instrument are also covered by the Photometric instrument. Before reaching the BP and RP CCD detectors, the light crosses two low-resolution fused silica prisms that act as dispersive elements with broadband filters to block unwanted light. Each prism operates within the corresponding wavelength range of each photometer (we remind from section 1.1.3 that BP: 360nm-680nm and RP: 640nm-1050nm). Combined, the two spectra cover the entire visible spectrum and the near-infrared range (fig. 1.10). From this data, the spectral energy distribution of all observed objects may be measured. This can be used to both correcting the centroid positions in the main astrometric field for systematic chromatic shifts, and to determine physical properties such as effective temperature, mass, age and chemical composition.



**Figure 1.10** – Gaia  $G$  (solid line),  $G_{BP}$  (dotted line, 360nm-680nm),  $G_{RP}$  (dashed line, 640nm-1050nm) and  $G_{RVS}$  (dot-dashed line, 847nm-874nm) normalised passbands. Credits from [Jordi et al. \[2010\]](#) (fig. 3 of the article)

Finally, the image reaches the RVS CCD detectors. RVS also uses the Sky Mapper function for object detection and confirmation, although the final object selection is made via data provided by the RP of the Photometric instrument (which considers only stars brighter than about  $16^m$ ). A passband filter restricts the throughput to the desired wavelength range of 847nm - 874nm (fig. 1.10). RVS analysis of each observed object will be performed on the average 40 times throughout the Gaia mission. From these spectra, radial velocities are then used to study the kinematic and dynamic evolution of the Milky Way. To derive radial velocities three isolated Calcium lines at 849.8nm, 854.2nm and 855.2nm are used. Other lines in the 847nm-874nm range can provide data on star composition, surface gravity, and metal abundance.

### 1.1.5 Detected problems and final performances

#### Problems

Although in general the assembly of on-board instruments as well as communication and ground-based data reception work fine since the beginning of the mission, some problems were detected early after launch. These issues initially affect the final accuracy of Gaia data and the precision expected for each scientific instrument. Enormous efforts were done to successfully mitigate and/or eliminate those problems, thus guaranteeing the quality of the data and the goals of the mission. The encountered issues and the way they were (and are) handled are briefly describe below:

- **Ice contamination**

Apparently some water vapour was likely trapped in the spacecraft before launch and emerged as water-ice deposits on the telescopes mirrors once it was in the vacuum (Gaia's payload has temperatures between  $-100$  and  $-150^{\circ}\text{C}$  in the dark behind the 10m diameter sunshield). Some ice on the mirrors was expected (that is why the mirrors are equipped with heaters), but the amount detected was higher than expected. And this caused a temporary and significant decrease in the overall transmission of the optics. Different "decontamination" cycles were performed by heating the affected optics to remove the ice, thus solving the problem. As the spacecraft continues to outgas for a while, future "decontamination" campaigns will be required during the mission. However, whilst the ice contamination has returned periodically, the rate at which it returns has dramatically reduced and the time between required decontamination operations has increased. Last decontamination (at the time of writing this thesis) was performed in June 2015 and since then there has been no perceptible return of the ice.

- **Straylight**

Once the satellite was placed in its final operational orbit, higher levels of straylight than those predicted before launch were reaching Gaia's focal plane, causing a significant excess of background light that varied with the spacecraft rotation phase. Since both the ice contamination and the straylight were simultaneously detected, it was initially thought that both were due to ice deposits: the first because of ice on the mirrors and the second because of an ice layer formed on the inner ceiling of the thermal tent structure surrounding the payload. However, while the mirrors have heaters incorporated, the roof does not.

Therefore two in-flight tests were performed: the satellite's nominal Sun Aspect Angle (SAA) of  $45^{\circ}$  was reprogrammed and changed first to  $42^{\circ}$  and then to  $0^{\circ}$ . The straylight was dramatically reduced at  $\text{SAA} = 0^{\circ}$  thus confirming that its origin came from the sunlight. Moreover the tests also rejected the hypothesis of ice contamination on the roof, a theory also disproved by ground laboratory tests which proved that ice layers on the sunshield could not be able to increase straylight in line with the in-flight measurements.

The final conclusion was that this extra straylight was caused by fibers protruding beyond the edge of the sunshield, which boosted diffracted light from the sun. On April-May 2015, an update of the on-board VPU software to a version that includes straylight mitigation measures was performed.

Nevertheless there is still some impact on the data, especially for the fainter stars where the noise associated with the stellar light itself is sometimes comparable to that from the background. Since the straylight is variable across Gaia's focal plane and variable with time, it has a different effect on each of Gaia's science instruments. Indeed, the impact is significant for the fainter stars seen by the high resolution spectrometer (RVS), and about 1.5 magnitudes of sensitivity could be lost. For astrometry this is negligible for objects brighter than 15 magnitudes, whereas it affects the positional accuracy for faint stars, reaching 50% for stars at the Gaia's nominal faint limit of magnitude 20. The final expected performances are detailed below.

- **Basic Angle Monitor variations**

As already mentioned, this instrument measures the angle between both telescopes to an accuracy of  $5 \mu\text{arcsec}$  every few minutes. This is necessary in order to ensure the accuracy of data reduction, by correcting for variations in the separation angle caused by thermal changes in the payload as Gaia spins. The system is working fine, but the periodic variations are larger than expected. These variations were carefully analysed during the commissioning phase and after, and they are now much better characterized, although definitive conclusions on their detailed origins are not yet known. Nevertheless, several tests were carried out with the first data to strength the confidence that these angle variations could be correctly dealt within the astrometric data processing.

## Final performances

At the time of the In-Orbit Review and start of the Nominal mission (July 2014), after the detailed calibration of all on-board instruments and analysis and treatment of the mentioned issues, the final Gaia performances resulted in the ones detailed in the following tables 1.2, 1.3 and 1.4 (credits from ESA <sup>2</sup>).

**Table 1.2** – Predicted end-of-mission parallax standard errors  $\sigma_{\varpi}$  averaged over the sky for a uniform distribution, for unreddened B1V, G2V, and M6V stars

	<b>B1V</b>	<b>G2V</b>	<b>M6V</b>
$V - I_C$ [mag]	-0.22	0.75	3.85
Bright stars	5-16 $\mu\text{as}$ (3mag < $V$ < 12mag)	5-16 $\mu\text{as}$ (3mag < $V$ < 12mag)	5-16 $\mu\text{as}$ (5mag < $V$ < 14mag)
$V = 15$ mag	26 $\mu\text{as}$	24 $\mu\text{as}$	9 $\mu\text{as}$
$V = 20$ mag	600 $\mu\text{as}$	540 $\mu\text{as}$	130 $\mu\text{as}$

**Table 1.3** – Predicted end-of-mission photometric errors in units of milli-magnitude, for unreddened B1V, G2V, and M6V stars

G [mag]	<b>B1V</b>			<b>G2V</b>			<b>M6V</b>		
	G	BP	RP	G	BP	RP	G	BP	RP
3 - 13	1.4	4	4	1.4	4	4	1.4	4	4
14	1.4	4	4	1.4	4	4	1.4	5	4
15	1.4	4	5	1.4	4	4	1.4	6	4
16	1.5	5	5	1.5	5	5	1.5	11	4
17	1.5	5	8	1.5	6	6	1.5	26	5
18	1.7	8	16	1.7	10	9	1.7	63	6
19	2.3	16	37	2.3	23	20	2.3	158	9
20	4.0	37	91	4.0	56	48	4.0	395	20

**Table 1.4** – Predicted end-of-mission radial-velocity robust formal errors  $\sigma_{v_{rad}}$ , averaged over the sky for a uniform distribution, for unreddened B1V, G2V, and K1III-MP (MP = metal-poor) stars

Spectral type	V [mag]	Radial-velocity error [km/s]
B1V	7.5	1
	11.3	15
G2V	12.3	1
	15.2	15
K1III-MP (metal-poor)	12.8	1
	15.7	15

### 1.1.6 Catalogue Releases

Due to the complexity of the data processing, the delivery of the Gaia catalogue was subdivided in five consecutive releases. That way, scientific exploitation can be performed from the very beginning, although the accuracy of the Gaia data will be obviously improved from one release to another, and thereby the scientific results and applications. Below we briefly describe the contents expected for each one (see Gaia website for future updates on this information <sup>3</sup>):

- **Data Release 1 (DR1)**

Positions ( $\alpha$ ,  $\delta$ ), Gaia magnitudes ( $G$ ) of most of the stars with acceptable precisions, and epoch photometric data of RR Lyrae and Cepheid variable stars. In addition, a sub-catalogue, TGAS (Tycho-Gaia Astrometric solution) [Arenou and Michalik, 2015, Michalik et al., 2015], will be published. It

<sup>2</sup>Gaia Performances: <http://www.cosmos.esa.int/web/gaia/science-performance>

<sup>3</sup>Data Release Scenario: <http://www.cosmos.esa.int/web/gaia/release>

will contain 5-parameter astrometric solutions for about two million stars in common with the Tycho Catalogue, using as first epoch *Hipparcos* or *Tycho* positions.

This release will include 14 months of observations, providing the most accurate sky-map at HST-like resolution, a large increase in parallax information, and much more accurate proper motions for *Hipparcos* and *Tycho-2* stars. In order to ensure the quality of the published data, sources with too few observations, or without astrometric or photometric information, or with parallax, positions or photometric errors larger than a given upper limit, will be omitted (A. Brown, EWASS, Athens, 4th July 2016).

All this filtering process is carried out by data validation done at AGIS (Gaia Astrometric Global Iterative Solution), at a specific level by PhotPipe (photometric pipeline) and VariPipe (variability pipeline), and at a global level by the Gaia DPAC Coordination Unit 9 in which this thesis is involved (details in section 1.3.1 and chapter 2).

Delivery date: 14 September 2016 (announced by A. Brown <sup>4</sup> the 4th July 2016)

- **Data Release 2 (DR2)**

Five-parameter astrometric solution for single objects, integrated BP/RP photometry and mean radial velocities for objects with constant radial velocity and verified astrophysical parameters

Delivery date: end 2017

- **Data Release 3 (DR3)**

In addition to previous updated data, DR3 will also include orbital solutions, system radial velocities and five-parameter astrometric solution for some binaries, as well as BP/RP/RVS spectra and astrophysical parameters for spectrophotometrically well-behaved objects, and mean radial velocities for non-variables stars and verified astrophysical parameters

Delivery date: summer 2018 (TBC)

- **Data Release 4 (DR4)**

In addition to previous updated data, DR4 will also include variable stars classification with epoch photometry, preliminary orbital solutions for Solar System objects and non-single objects catalogue

Delivery date: summer 2019 (TBC)

- **Final Release (DR5)**

This will be the final Gaia catalogue, consisting of a complete astrometric, photometric and radial-velocity all-sky survey for a billion stars, with variable stars and multiple systems solutions, astrophysical parameters (effective temperatures, surface gravities, metallicities...) derived from BP/RP colours, RVS and astrometry for stars, unresolved binaries, galaxies, and quasars, a list of exoplanets and epochs and transits for all sources. All ground-based observations made for data-processing purposes will also be included.

Delivery date: summer 2022, 3 years after the end of the mission (TBC)

## 1.2 News from Gaia

The first full Gaia sky scan was achieved six months after the scientific operations started. Just within one year, the satellite recorded 272 billion astrometric measurements, 54.4 billion brightness or photometric data points, and 5.4 billion spectra, and an average of 14 measurements of each star on the sky. This was more than initially expected since Gaia has been able to observe objects with fainter magnitudes. And

<sup>4</sup>Chair Gaia Data Processing and Analysis Consortium (DPAC) Executive (2012-)

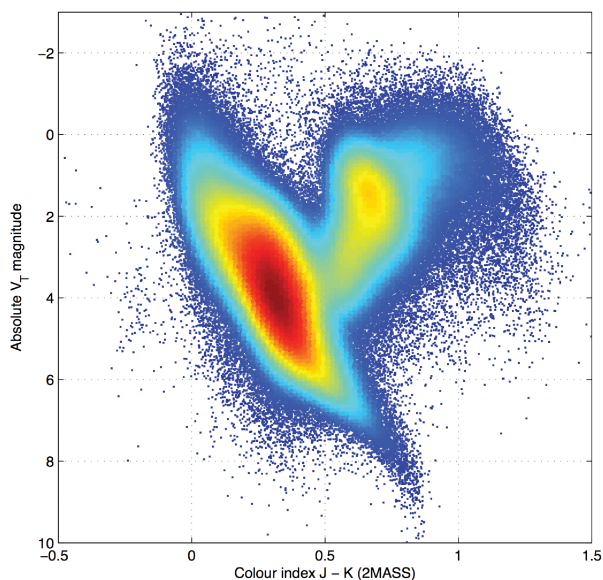
almost two years later, Gaia continues scanning the sky, mapping on average 50 million stars a day (with a peak of 250 million), that cross the focal plane triggering astrometric, photometric and spectroscopic measurements. The spacecraft is currently performing astrometry and photometry for stars brighter than  $G = 20.7$  mag and spectroscopy till  $G_{\text{RVS}} = 16.2$  mag. At the time of writing this thesis, Gaia has already obtained 450 billion astrometric measurements, 110 billion photometric sources and 9.3 billion spectra (last update by T. Prusti<sup>5</sup>, EWASS, Athens, 4th July 2016).

Around September 2015, the first version of TGAS was provided to the data validation team to check the first Gaia data, its quality, detect reduction problems, possible anomalies, and so on. The outputs from these validations were crucial to improve the next TGAS versions as well as the ensemble of Gaia data. On April 2016, the Gaia-DR1 was also transferred to the validation team, which during last months carried out a rigorous data check. All data accepted by the validation tests will be published the 14th September 2016.

### Some events

Since the beginning of the mission, some particular astrophysical events have occurred. We briefly highlight some of them here:

- **First Supernova:** on August 2014 Gaia discovered its first Type Ia supernova (called *Gaia14aaa*) in a far galaxy at some 500 million light-years. The event was detected because of a sudden rise of about two magnitudes in the galaxy's brightness between two Gaia observations separated by one month (31 July and 30 August 2014).
- **First Gaia HR diagram:** with one year of observations and combining Gaia data with positions extracted from *Tycho 2* catalogue, the parallaxes and proper motions for two million stars (*Tycho 2* stars) were determined. Thus using luminosity and colour measurements based on data from *Hipparcos* and ground-based observations, a preliminary HR diagram could be constructed. This can be seen in fig. 1.11.



**Figure 1.11** – Gaia's first Hertzsprung-Russell diagram. Credits from ESA/Gaia/DPAC/IDT/FL/DPCE/AGIS

- **Comet 67P/Churyumov-Gerasimenko:** at the end of 2015, Gaia could also detect the *Comet67P/Churyumov-Gerasimenko*, were ESA's *Rosetta* mission just put a robot, *Philae* on its surface. Indeed, thanks to its capabilities, Gaia is also able to observe and register faint Solar System bodies, like asteroids and

<sup>5</sup>Gaia Project Scientist, European Space Agency

comets. It has already identified tens of thousands of asteroids, for which accurate orbits, composition and surface properties will be determined.

### 1.3 The Gaia Data Processing and Analysis Consortium

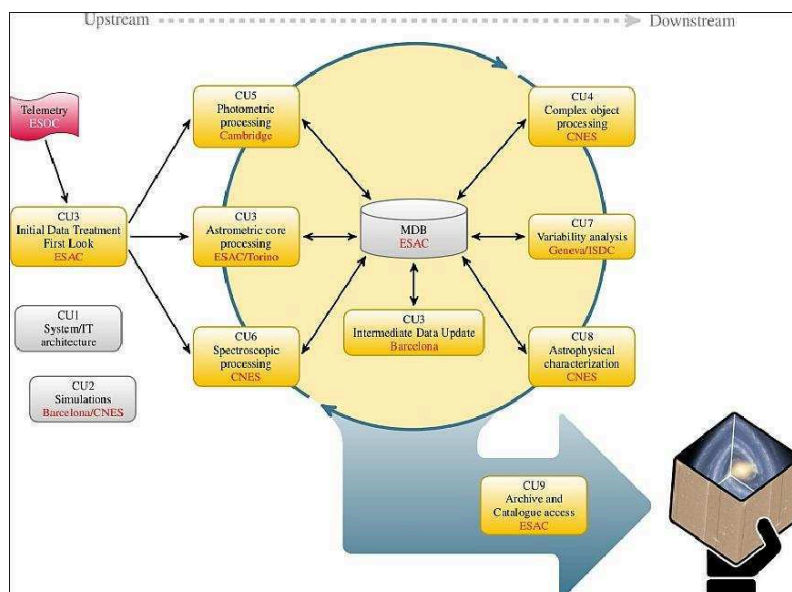
In order to process the huge amount of data and prepare the data analysis, a Gaia European consortium was created in 2006: the Data Processing Analysis Consortium (DPAC). It is a collaboration of more than 450 scientists and software developers from about 25 different countries, and six large Data Processing Centres (DPC-E:ESAC, Madrid; DPC-C: CNES, Toulouse; DPC-I: IoA, Cambridge; DPC-G: Observatoire de Genève/ISDC, Geneva; DPC-T: INAF-OATo, Torino; DPC-B: BSC and CESCA, Barcelona), created to handle and process Gaia data. The main role of DPAC is indeed to develop tools and algorithms to analyse the data, as well as establish the whole infrastructure to process this data, and prepare the successive Catalogue releases.

The DPAC is subdivided in nine Coordination Units (CU) , each one responsible of a particular aspect of the overall Gaia data processing system (astrometry, photometry, astrophysical parameters, variability, science alerts, simulations, final catalogue preparation, delivery to the scientific community,...). They are specified in table 1.5. Each DPC supports at least one CU.

**Table 1.5** – Responsibilities of each Gaia DPAC Coordination Unit

Coordination Unit	Responsibility
CU1	System architecture
CU2	Data simulations
CU3	Core processing
CU4	Object processing
CU5	Photometric processing
CU6	Spectroscopic reduction
CU7	Variability processing
CU8	Astrophysical parameters
CU9	Catalogue access

Figure 1.4 shows a schematic flow of the full processing phase, how the different CUs are involved and which are their associated DPCs.



**Figure 1.12** – Schematic view of the data flow during the processing phase. Credits from ESA

### 1.3.1 CU9: the last step

The Coordination Unit 9 is devoted to the preparation of the Gaia Catalogue, the delivery of the successive releases to the scientific community and the preparation of tools to facilitate its access and visualization. It actually represents the end of the full data processing phase and the last step before the publication of the data (as schematically shown in fig. 1.12). To be able to handle all these tasks, CU9 is in turn subdivided in eight Work Packages (WP) :

- WP910 - Management: coordinates and manages the ensemble of CU9 groups, and its link with the rest of DPAC. It has also created and maintains the overall plan towards the Gaia catalogue releases
- WP920 - Documentation: sets the DPAC guidelines and templates to document all the work performed during the mission
- WP930 - Archive architecture: designs and develops the core services and infrastructure to host the Gaia data, including databases and access services to be used by the rest of the work packages to develop their respective technical and scientific tools
- WP940 - Validation: develops statistical tools and tests in order to detect errors or anomalies in the Gaia data, such as contaminations, parameters' incoherences, outliers,... before each release of the catalogue
- WP950 - Operations: monitors, maintains and provides scientific, technical and user support for all the systems and data regarding the archive
- WP960 - Education and Outreach: responsible for providing informations, documentaries, media, platforms, brochures, movies, and all kinds of outreach tools to show to the public the goals, updates and whereabouts of the Gaia mission
- WP970 - Science enable applications: provides the user of the Archive the tools to perform advanced data mining, Big Data treatment and management, crossmatching with external catalogues, and all operations that involve large amounts of data and complex scenario situations
- WP980 - Visualisation: provides advanced visualisation tools to enable a global exploration of the Gaia archive, including interface services for the user, visualisation infrastructure needed for the rendering of data, data selection, etc

This thesis takes part of the WP940 validation team, more specifically the WP944 (see next chapter). Chapter 2 details the purposes of the specific validation actions performed within this research work, as well as the infrastructure developed to facilitate the interaction with the global environment of validations and the Gaia database. In chapter 3 the different tests and statistical contributions are described. And finally, in chapter 4 we show some of the results obtained during the validation process, and the improvements achieved on the data itself over the last months thanks to all these actions. The work done in WP940 will guarantee the quality of the Data Release 1, delivered on next 14 September 2016, as well as that from future releases.





# WP944: confrontation with external archives

## Contents

---

<b>2.1 Underneath framework and purposes of WP944</b>	<b>138</b>
2.1.1 Goals of WP944	139
2.1.2 Specificity: external archives	139
2.1.3 Coverage: parameters and objects	140
<b>2.2 Infrastructure</b>	<b>141</b>
2.2.1 Start-up	141
2.2.2 Crossmatch processing	141
2.2.3 Gobal tests management	143
<b>2.3 General validation process</b>	<b>144</b>
2.3.1 Databases used in validation	144
2.3.2 Validation process	144
2.3.3 Pass/Fail criteria	145
2.3.4 JUnit tests	145

---

## 2.1 Underneath framework and purposes of WP944

The Data Processing Center responsible for the CU9 operations is the European Space Astronomy Center (ESAC, Madrid). Once all the data has been analysed and processed by all the other CUs, it is integrated in the Main Data Base (MDB) at ESAC and transferred to the CU9-WP940 to carry out the last checks. Each individual CU possesses its own verification modules, i.e. the data arriving to CU9 has already been filtered and checked. However, since possible inconsistencies in cross-CU validations (ensuring that the outputs provided by the different CUs are consistent) and data modifications (data transformation, star filtering) can occur between the CUs and CU9, extra tests need to be applied to highly ensure the quality of the data and the reliability of the final catalogue.

The Work Package WP940 - Validation, is also subdivided in different groups depending on the kind of validation tasks, and the objects and parameters to be checked (table 2.1):

**Table 2.1** – Responsibilities of each Work Package in CU9. WP944, in which this thesis is involved, is highlighted in bold

Coordination Unit	Responsibility
WP941	Validation Management
WP942	Internal consistency and more complex scenarios
WP943	
<b>WP944</b>	<b>Confrontation with external archives</b>
WP945	Statistical tools
WP946	Time series and variability
WP947	Clusters as validation tools
WP948	Solar System Objects

All the test cases developed within each of these sub-working packages must use and interact with the following tools of the global validation environment, in order to facilitate the management of the ensemble of validation tasks, their output results and the final integration of the Gaia catalogue:

- **ValidationTools**

All validation tests developed in the framework of WP940 directly depend on the validation environment provided by the `ValidationTools` package [Shih, 2016a]. This package includes a collection of tools [Shih, 2016b] to help building and running Gaia archive validation tests within the constrained environment imposed by Gaia DPAC. It mainly manages and handles the access to data resources and the communication with the Gaia Archive at the ESAC Gaia Processing Center.

- **Gaia Archive Core Systems**

The Gaia Archive Core Systems <sup>1</sup> (GACS) are a set of subsystems developed for the Gaia project by the ESAC Science Data Centre that allow to search, download and explore the Gaia data. During CU9 Operation, all the validation software is ran on a virtual machine (VM) set up inside ESAC. Each test case is executed autonomously depending on the configuration. Once each test case completes its task, no matter what its final status is, it updates some of the results to a Validation Database and generates a report which is delivered to its developers. There are two ways to access the Gaia Archive: through the Table Access Protocol (TAP) database service, and through a series of Gaia files called (GBIN). The Gbin files are being used for the operation on the whole Catalogue. As WP944 either use small Catalogues to compare to Gaia stars, or needs cross-matching first, a direct access by identifier or position is needed (which is done through the TAP service).

<sup>1</sup>Gaia Archive: <http://gaia.esac.esa.int/archive/>

- **Table Access Protocol**

The Table Access Protocol <sup>2</sup> (TAP) is a web-service protocol for accessing collections of tabular data. Access is provided for both database and table metadata as well as for actual table data. TAP accepts queries written in different languages, although for the interaction with the Gaia Archive, only the Astronomical Data Query Language (ADQL) is used. In this case, the table metadata must be checked first (names of tables and columns) in order to formulate the query. The query is then posed against a table and TAP returns its response as another table (normally in VOTable format). The service allows to execute multi-table operations such as joins or crossmatches, and includes support for both synchronous and asynchronous queries.

### 2.1.1 Goals of WP944

The WP944 aims to validate and compare Gaia data with that from other catalogues and surveys. By doing this comparison it intends to check:

- the completeness of the Catalogue and its global sky distribution
- the accuracy and precision of parameters: positions, parallaxes, proper motions, magnitudes, colours and radial velocities of stars, quasars and galaxies. Lots of other parameters such as the astrophysical parameters, the extinction, the binarity and the variability, will be tested in future releases as well
- the correct distribution of these parameters, its coherence with previous studies, correlations between them, detection of possible bias
- the presence or absence of outliers and data processing problems

### 2.1.2 Specificity: external archives

Gaia has no other comparable all-sky survey with the same angular resolution and multiple star discovering power. Indeed, it is unique in many senses since it combines the spatial resolution of Hubble Space telescope with an all-sky survey in optical wavelength, down to  $G \sim 20$ , unprecedented astrometric accuracy and homogeneous photometric data. Nevertheless, there exist other smaller catalogues than can be used as well. This requires, however, taking especial care of the precision, coverage and inner performances of these external archives with which we are confronting Gaia to.

In WP944 we consider three groups of external catalogues:

1. **All-sky catalogues with lower performances:** like *2MASS*, *UCAC4*, *Tycho-2*, *OGLE...*, which allow to check, for instance, completeness and sky homogeneity
2. **Catalogues with added information:** e.g. those used to check distances and parallax zero-points, for which the distance modulus may be calculated from astrophysical parameters (spectroscopy or variability)
3. **Small but detailed catalogues:** this concerns both individual well-known stars, and thorough works on the determination of parameters for a specific star sample

All these external archives are selected, checked and tested individually and independently before being used in any Gaia validation test.

<sup>2</sup>TAP: <http://www.ivoa.net/documents/TAP/20100327/REC-TAP-1.0.pdf>

### 2.1.3 Coverage: parameters and objects

At the time of writing this thesis, the amount of Validation Test Specifications (VTS) defined for the WP944 have significantly increased, now covering a large field of different parameters and objects.

Currently there are seven blocks of VTS, each one including a set of specific validation tests <sup>3</sup> [Arenou et al., 2016c]:

- **VTS 010 - Sky coverage:** sky homogeneity, completeness, issues for specific stars (e.g. those detected in previous releases), problematic areas (such as Barnard's star area, or around Vega and Sirius), completeness as a function of magnitude and separation of visual binaries. For DR1 we also consider issues identified in some *Tycho-2* and *Hipparcos* stars

Catalogues: *2MASS*, *Tycho-2*, *U.S. Naval Observatory CCD Astropgraph Catalog (UCAC4)*, *Hubble Space Telescope (HST)*, *Washington Double Star Catalogue (WDS)*, *Optical Gravitational Lensing Experiment (OGLE)*, *Sloan Digital Sky Survey (SDSS)*

- **VTS 030 - External parallaxes and proper motions:** check these parameters with other catalogues, also specific for *Hipparcos* astrometry and high proper motion stars; parallax zero-point

Catalogues: *Very Long Baseline Interferometry (VLBI)*, *HST*, *Research Consortium On Nearby Stars (RECONS)*, *UCAC4*, *Tycho-2*, *Hipparcos*, *Simbad* database, *Large and Small Magellanic Clouds (LMC/SMC)*. Distance modulus estimation from variability: *RRLyrae*, *Cepheids*; from spectrophotometry: *RAial Velocity Experiment (RAVE)*, *PAramètres STELLaires (PASTEL)*, *APO Galaxy Evolution Experiment (APOGEE)*, *Large Sky Area Multi-Object Fibre Spectroscopic Telescope (LAMOST)*, *Gaia-ESO Survey (GES)*,...; and asteroseismology: *APOGEE + Kepler catalogue (APOKASC)*

- **VTS 040 - Biases of radial velocities:** global sky distribution of radial velocities differences of stars, their magnitude, colour and spectral type. This VTS will be activated in DR2

Catalogues: RVs velocities literature

- **VTS 050 - External astrophysical parameters and photometry:** colour-colour relations and effective temperature-colour relations from small compiled catalogues

Catalogues: *Hipparcos*, *HST*, *Tycho-2*, *SDSS*, standard stars (Landolt, Calspec...). Much more to be added when CU8 outputs will be available

- **VTS 060 - Quasars consistency:** cross-identification, sky distribution, completeness, parallaxes and proper motions, global consistency

Catalogues: *Large Quasar Astrometric Catalog (LQAC3)*, *Sloan Digital Sky Survey Quasar Catalog (SDSS DR10Q)*, *Gaia Initial Quasar Catalogue (GIQC)*, *UCAC4*, *International Celestial Reference Frame (ICRF2)*

- **VTS 070 - Known multiple stars:** consistency, equal parallaxes for wide binaries, proper motions, metallicities, ages, components' separation, spectroscopic binaries

Catalogues: *Tycho Double Star Catalogue (TDSC)*, *Hipparcos Double and Multiple Systems Annex (Hipparcos DMSA)*, *WDS*

- **VTS 080 - Galaxies consistency:** sky distribution, parallaxes and proper motions

Catalogues: *SDSS*

My work covers mainly VTS 030 and VTS 010 (details in section 3).

<sup>3</sup>Note: The reader will notice that there is no VTS 020. Indeed, due to updates of the ensemble of Validation Tests Specifications, VTS 020 was removed and its tasks distributed among the other VTS

## 2.2 Infrastructure

This section describes the work implemented in the framework of this thesis to carry out the start-up of WP944, which began in 2014. This concerns mainly the first steps taken on the development of the general infrastructure of WP944, the external dependencies and tools needed, the encountered issues, how they were handled, and the global management of the validation tests.

Since then the whole Work Package has hugely evolved. On one hand because its construction has been performed in parallel with the development of some of the main validation external dependencies (`ValidationTools`, *Gaia Archive*,...), thus we have had to continuously adapt the infrastructure to the new changes, improvements and updates; and on the other hand, because several processes have been generalized and optimised as more tests have been developed and added, and as clearer the needs have shown up.

### 2.2.1 Start-up

As already indicated in the previous section, all test cases must interact with the `ValidationTools` framework which provides global validation tools for all WP940 and it is, roughly speaking, the bridge between the VTS and the Gaia Archive. Since they were being developed in parallel with this work, the start-up of WP944 concerned several testing tasks, which can be summarized within the following points:

- test the capabilities, limitations and tools of the Gaia Archive Core Systems: data handle, queries format, jobs management,...
- test the best way to deal with external tables in order to perform crossmatches and data comparisons: input and output format tables, metadata, output information,...
- require the insertion of big full-sky surveys to avoid upload tasks for them (*UCAC4*, *Tycho-2*, *2MASS*, *Hipparcos*,...)
- require already crossmatched tables between big catalogues and "Gaia" (the *Initial Gaia Source List* and TGAS simulations at the time)
- test updates and new functionalities proposed in both `ValidationTools` package and the Gaia Archive at ESAC, in order to check their utility, limitations and possible drawbacks or issues for validation tasks (such as security access requirements, table sharing functionalities, input, output and temporary file and results locations, or some standard ADQL queries)
- suggest improvements, changes and global needs for validation that arose while developing the first tests, such as moving some frequently used methods to the `ValidationTools` level

WP944 was the first work package dealing with external catalogues to begin its tasks within the CU9 Validation team. Therefore, the enormous effort invested (briefly detailed above) to develop a global validation infrastructure and general tools within and outside WP944 has been of high utility for other validation WP, which started their tasks later.

### 2.2.2 Crossmatch processing

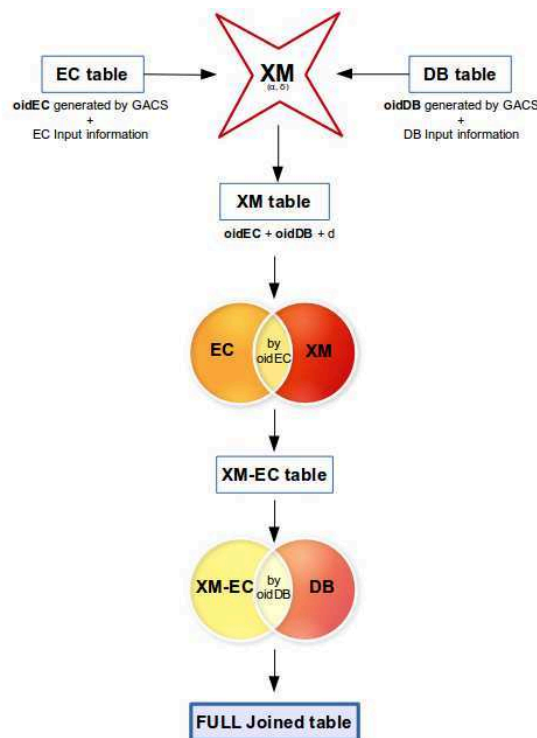
To crossmatch (hereafter XM) the external catalogues (hereafter EC) and the Gaia database (hereafter DB), two main ways have been implemented: by identifier and by position. This concerns only those small catalogues prepared specifically for WP944. For big catalogues (such as *2MASS*, *Tycho-2*, *UCAC4*,...) the XM is done in the WP974.

**By identifier** Especially for the TGAS DB, which includes all the stars from the Tycho-2 catalogue, the Tycho-2 identifier (Id) is given. Therefore, the easiest and fastest way to XM the external tables with TGAS is directly by the name of each source, its Id.

To do so, a list of identifiers is generated from the input table and an ADQL query is constructed to look for these identifiers in the DB. Since the ADQL query can only handle a limited list of values (around 350 sources), when the list size is higher than the limit, a temporary VOTable with the identifiers needs to be created and uploaded to the archive. Once there, the temporary table is *joined* with the DB by the identifiers, and the query response returns the full information of the join.

**By position** For the other objects, the XM must be done by coordinates. A positional XM ADQL query was defined within `ValidationTools` to make it available for everyone in WP940. Once the EC table is uploaded to the archive, for each source in the input EC the query does a cone search in the DB within a given radius (specified in each test case) that depends on the precision of the EC.

Every time a table is uploaded (by the user or by ESAC), the service gives an object Id to each source (independent of the original source Id). Once the XM is done, all the objects found for each source are added into a new XM table which includes the object Id of both initial tables (EC and DB) and the distance. To get the full information of each source, a *join* TAP query by object Id must be done with the EC and the DB separately. Figure 2.1 shows the work flow of a crossmatch by coordinates.



**Figure 2.1** – Work flow of the crossmatch process by source position. *oidEC* and *oidDB* are the default “object Id” given by GACS for the External Catalogue and the DataBase, respectively, and *d* the angular distance between the input EC objects and their crossmatches found in the DB

Before using the resulting table for validation test purposes, it is first analysed to check whether it was done correctly, that the information is coherent, and possible XM issues. This analysis allows to be sure of the reliability of the information of each source finally considered for the test case, and to provide a summary statistics of the XM itself. Three aspects are mainly checked:

- ***NaN*<sup>4</sup> or 0 values**

The input EC are always checked before doing the XM. And normally, after CUs processing there should be no *NaN* values for any of the parameters or zeros in, for instance, the parameter's errors. Nevertheless, since modifications can occur during the data transfer to CU9, this must be checked. At the time of developing all this infrastructure and the firsts tests, this was also important since there was no Gaia data yet. Instead of Gaia we used IGSL (the Initial Gaia Source List, a Gaia-size compilation containing data from several other catalogues) or simulations in the case of TGAS, which we knew they could include wrong information or estimations.

- **Not found sources**

Statistically it is important to know how many sources were not found during the XM. Especially when done by identifier, since we expect Gaia to observe "at least" all what has already been observed before. Not found sources may show possible integration issues or some kind of previous filtering for specific objects. This last option can occur for the first releases where there are still no good solutions for all known objects. These objects will be published in next releases. These cases are detailed in the documentation that will be delivered to the scientific community for each Data Release publication [Arenou et al., 2016d, in prep. for GDR1].

- **Multiple X-match**

Although it may also occur for XM by identifier, where for whatever reason we can find two sources with the same Id, this check is mainly for XM by position. Doing a cone search of a given radius in the DB, around the coordinates of a known source (given by an EC), may return more objects around that position since the precision of the DB is better (thus more objects observed within the same region). To validate which one corresponds to the EC's object, a check in magnitude is done: using the closest magnitude from G that is given (or can be derived) from the EC as reference, we verify whether the difference between both is lower than 2 mag (i.e. a reasonable limit). If so but still more than one source for the given input in the EC, these objects are kept apart to be inspected or validated later by hand.

### 2.2.3 Global tests management

ValidationTools takes in charge the ensemble of WP940 validation tests to facilitate their running in the virtual machine of the Gaia Archive. Nevertheless, each sub-WP also needs to manage its own tests as optimised and handy as possible.

In that sense, the overall infrastructure of WP944 constitutes of a series of general functions and tools, established to allow a global control of all VTS cases. That is:

- Initialisation of global parameters (such as XM radius, statistical confidence interval, p-value,...)
- Definition of input and output files formats
- Handling of XM processing
- Specification of available DB tables (see next section 2.3.1)
- Management of EC and DB field names. We use standard names for the common parameters, like *REFMAG* (reference magnitude), *ALPHA* (right ascension), *MU\_ALPHA* (proper motion, right ascension component),... This way, regardless the field names of each input catalogues, or their possible modifications in the Data Model, all VTS will run correctly
- Interaction with R program language used for validation tasks
- Set Test Flag Results (pass, error, exception, warning,...)

---

<sup>4</sup>NaN: Not A Number

## 2.3 General validation process

### 2.3.1 Databases used in validation

Since validation tasks started before Gaia data was available, we used other DB in order to develop the tests, ensure beforehand the correct functioning, and also to check the consistency of the results with real data when possible.

- **Initial Gaia Source List (IGSL)** [Smart and Nicastro, 2014]: we started using this compilation of the best optical astrometry and photometry information on celestial objects available before Gaia. It nominally includes all known objects brighter than 21 magnitude in the Gaia G band, i.e. it represents roughly the expected Gaia catalogue size. It has been used as the starting point for the data treatment and as a comparison catalogue to trigger early science alerts
- **Tycho-Gaia Astrometric Solution simulation (TGAS simulation)** [Arenou and Michalik, 2015, Michalik et al., 2015]: with the TGAS proposal, all the tests were adapted in order to give priority to TGAS validation. Therefore, instead of using IGSL we used simulated TGAS data (catalogue prepared by F. Arenou), in order to facilitate validation tasks for this particular catalogue
- **Hipparcos New Reduction** [van Leeuwen, 2007]: when applicable, we have also used the Hipparcos catalogue in place of Gaia, to check that the tests were working well on real data (specially those concerning astrometric parameters)
- **Gaia real data**: the first real TGAS data was delivered to the CU9 validation team on September 2015. We received several other preliminary versions of TGAS until the final Gaia Data Release 1 (GDR1) data received in April 2016 (see table 4.1). On the preliminary TGAS data we could both validate the provided data and test and improve our code

### 2.3.2 Validation process

WP944 aims to check Gaia data with other already known catalogues. Therefore all tests follow, globally, the same process schema (fig. 2.2).

Before using an EC for a specific validation test it is inspected and cleaned independently and outside the validation process. Once the verification done and the objects for testing selected, the final EC data table is included in the corresponding VTS environment.

For each test the first step is the XM between the EC and the DB chosen, through the TAP service. It is specified whether it is done by Identifier or by Coordinates. Then the X-matched table is checked and cleaned. And with the remained data, the Gaia statistical test is performed.

Some basic statistics are applied to almost all tests in WP944, e.g. the normalized differences of some parameters versus a normal distribution, the correlations between these parameters, their weighted mean and intrinsic dispersion, their distribution skewness, as well as possible outliers. For some tests specific statistical methods are also applied (see details later in section 3).

Finally the results of all the statistics are used to provide the output of each test. This output is constituted of three elements:

1. Summary of all the statistics performed
2. Graphs showing a visual and general idea of the parameter's behaviour
3. Pass/Fail test criteria (see 2.3.3)

All tests are run in a virtual machine at ESAC, and the above outputs returned to the developers for analysis and control of the quality of the Gaia data.



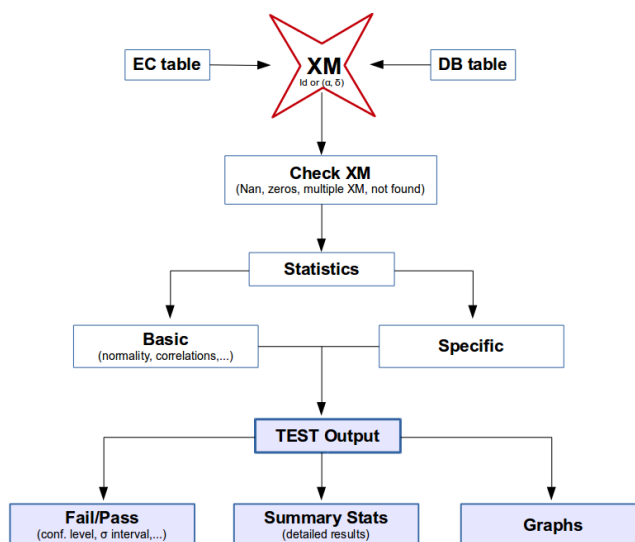


Figure 2.2 – Work flow of WP944 general validation process

### 2.3.3 Pass/Fail criteria

In general, a test case is considered Pass if all the steps within it have been correctly achieved. Otherwise it is considered as Fail. In practice, each validation test returns a global flag (fail/pass/warning/exception/...). Only fail/pass are described for each test individually. If the test throws a Java or R exception, it leads to an exception flag. If an exception is thrown due to the TAP service, it leads to an error flag. And if the test was not conclusive (no fail or pass explicit indication), a warning flag is received.

For the final review and acceptance of data for publication, the Pass/Fail criteria may slightly change. When an issue is encountered it is important to check with the other validation Work Packages and the other CUs whether they found similar results. For example, automatic correlation tests can indicate "Fail" because a correlation is found while such a correlation is actually expected. The comparison with the other WP940 Work Packages allows to see the results from different points of view, so we can confirm and quantify the possible problems. The interaction with the other CUs also allows to confirm or not the issue and, if so, to try to understand its origin.

All these checks may imply doing modifications in the database and reprocess everything (if there is still time before the publication of the given release, else it would be corrected for next release), or adding/correcting CU9 filters. The final encountered features are detailed in the release documentation.

### 2.3.4 JUnit tests

As part of the process and to ensure their proper functioning, each test case has its own *JUnit* test, which roughly speaking is a *test of the test*. It allows to constantly check that the VTS is doing what it has to do.

In a general way, they guarantee that all the TAP queries are correctly formulated and as expected, that there are no issues with the connection and response of the database at ESAC, and that the input catalogues are correctly loaded.

For each particular VTS other checks may be added. If a VTS uses a specific statistical method, the *JUnit* also validates whether the resulting value of the VTS for a given star is the expected one (from the literature). This is the case of the  $\chi^2$  method in the External Distances test (section 3.2) or the *calmos* program in the Parallax Zero-Point test (section 3.3). Another case is the Transformation Epoch method, where the results of the implemented rigorous treatment (see later in section 3.4.2) must agree with the simple treatment:

$$\alpha = \alpha_0 + (T - T_0) \mu_{\alpha_*} \sec \delta_0$$

$$\delta = \delta_0 + (T - T_0) \mu_{\delta_0}$$

This simple method is therefore included in the JUnit test, and it is used to check the results of the VTS for a given known star, the Barnard's star. That is, using as reference the astrometric information at  $T_0 = 1991.25$  ( $\alpha_0 = 269.45208$  deg,  $\delta_0 = 4.69339$  deg,  $\mu_{\alpha_*} = -798.58$  mas/yr,  $\mu_{\delta_0} = 10328.12$  mas/yr), it validates whether the rigorous transformation to  $T = 2000.0$  is in agreement with:

- the results from the above simplified treatment
- the values at epoch  $T$  given in Simbad

# Validation tests and contributions

## Contents

---

<b>3.1 Overview</b>	<b>148</b>
<b>3.2 External parallaxes and proper motions</b>	<b>149</b>
3.2.1 Preliminary external catalogues	149
3.2.2 $\chi^2$ method	149
<b>3.3 Parallax Zero-Point and precision</b>	<b>151</b>
3.3.1 External Distances: RR Lyrae	152
3.3.2 Very distant stars: LMC/SMC	156
<b>3.4 Specific contributions</b>	<b>157</b>
3.4.1 Sky Homogeneity: HEALPix maps	157
3.4.2 Epoch Transformation	159

---

### 3.1 Overview

Although the Gaia validation tasks in WP944 are not carried out by a single person but a team, this section aims to specifically describe the Gaia test cases and contributions developed in the framework of this thesis, beyond the work done in implementing and setting up the infrastructure of this Work Package (section 2.2 of previous chapter). Indeed, most of the tests are products of different collaborations involved in the external catalogues preparation, the interaction with the database, the development of specific methods or tools, the continuous updates and improvements, the JUnits tests, and the tests implementation itself.

Moreover, since the start of WP944, the whole system (including WP940 tasks and GACS interface in general) as well as the tests have largely evolved. Thus, new updates, modifications and improvements are being continuously applied, especially because of new data and new releases coming. The contributions done here must be considered as part of the start-up of WP944: they constitute the first actions performed in this Work Package, before the first Gaia data available. They mainly concern validation of parallaxes and proper motions.

These VTS were published in Ruiz-Dern et al. [2015], a Validation Test Report (VTR, Gaia internal publication) which also includes an ensemble of other WP944 tests developed in parallel by other WP944 collaborators. The complete publication can be found in Annex B. The VTR [Ruiz-Dern et al., 2017, in prep.] includes all the tests performed for the DR1.

The first section of this chapter, 3.2, is focused on the parallax and proper motion consistency tests. Then, section 3.3 concerns those aiming to check the parallax zero-point, using both distant and very distant stars. And, finally, section 3.4 includes the specific and punctual contributions, such as the star count technique to validate the sky homogeneity and the epoch transformation treatment of astrometric parameters.

We note beforehand that when we will talk about the proper motions components we will always refer to:

$$\vec{\mu} = (\mu_{\alpha_*}, \mu_{\delta}) = (\mu_{\alpha} \cos \delta, \mu_{\delta})$$

which takes into account the radius variation from the axis of a sphere to its surface, i.e. it diminishes towards the poles as  $\cos \delta$  (fig. 3.1). Physically this means that the component of velocity ( $v$ ) parallel to the equator corresponding to a given angular change in  $\alpha$  is smaller the further north the object's location.

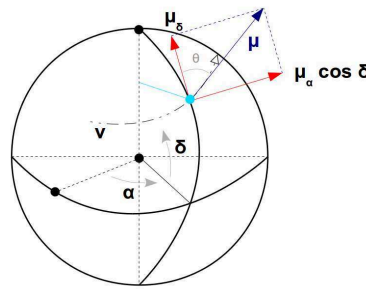


Figure 3.1 – Schema of the components of proper motion

The first tests developed, described in Ruiz-Dern et al. [2015], were checked at 5% level of rejection (i.e. a p-value limit of  $\alpha = 0.05$ , which corresponds to a gaussian  $\sigma \sim 2$ : less than 1 in 20 chance of being wrong). It means that statistic results giving p-values smaller than  $\alpha$  indicate a strong evidence against the null hypothesis, so the null hypothesis is rejected, the test fails. If the p-value is higher than  $\alpha$ , this evidence is weaker and so the null hypothesis cannot be rejected, the test passes. However, the comparison with *Hipparcos* prompted us to change the level of rejection to 1%, that is a p-value limit of  $\alpha = 0.01$ , which corresponds to a  $\sigma \sim 2.6$ . For the same reason, mean values, which in the VTR were checked at  $3\sigma$ ,

are also currently checked at  $2.6\sigma$  to be uniform. The limit for outliers is kept at  $5\sigma$ .

## 3.2 External parallaxes and proper motions

Since the main information expected from Gaia (already from TGAS and DR1 releases) are the astrometric data (parallaxes and proper motions), this VTS constitutes the first validation test carried out in WP944. Its development, therefore, was performed together with all the start-up tasks explained in section 2.2. As a matter of fact, it allowed us to test all the CU9 ValidationTools and GACS tools (i.e. ADQL queries, external tables handling and management, access to the database, control, format and required location of input files and output results,...), as well as to detect problems and to highlight new needs in WP944 (like global XM tools, access to other big catalogues, accepted formats for external tables, how to get back the information, or the way of dealing with the ADQL query outputs).

In this sense, we started by using three preliminary small catalogues (briefly described below in section 3.2.1) which allowed us to check all these aspects. Later, once we focused the tests on the TGAS catalogue, a compilation of VLBI (*Very-long-baseline interferometry*) stars was used instead. The HST (*Hubble Space Telescope*) data are currently being used too. The VLBI compilation contains only stars expected to be in TGAS. Since most of the stars of the three preliminary used catalogues are already included in this compilation, only the VLBI were kept for the VTS. For more details on this compilation see section 4.2 of annex B.

### 3.2.1 Preliminary external catalogues

The three catalogues initially used were:

- **HST Four Classical Novae** [Harrison et al., 2013]: contains precise astrometric parallaxes derived using data from the Fine Guidance Sensors on the Hubble Space Telescope, for four classical novae: *V603 Aql*, *DQ Her*, *GK Per*, and *RR Pic*.
- **Distances of 14 Cataclysmic Variables**<sup>1</sup> [Thorstensen, 2003]: provides trigonometric parallaxes with uncertainties below 1 mas (in the best cases), derived using data from the 2.4m Hiltner Telescope at MDM Observatory, for 14 cataclysmic variables (binary stars in which a white dwarf accretes matter from a close companion, which usually resembles a low main-sequence star).
- **Deep Astrometric Catalog in 20 sky fields**<sup>2</sup> [Lazorenko et al., 2014]: contains astrometric data with an accuracy of  $\sim 100$  micro-arcseconds derived using images from the FORS2/VLT camera. These are faint reference stars in 20 fields located close to the Galactic plane, observed in the framework of an astrometric planet search around ultracool dwarfs.

### 3.2.2 $\chi^2$ method

Ideally the difference between two measurements of the same parameter should be zero. That is, whether we compare, for instance, the parallax of a star from two surveys, Gaia (G) and an external (E):

$$\varpi_G - \varpi_E = 0$$

In practice, variations are found due to, among others, the capabilities of the instruments and the measurement techniques. Therefore, when comparing data coming from different sources, the precision of each as well as the correlations with other variables must be taken into account.

<sup>1</sup>Distances of Cataclysmic variables (VizieR reference): J/AJ/126/3017/table4

<sup>2</sup>Deep Astrometric Catalog (VizieR reference): J/A+A/565/A21/cat\_fors

We may consider that both parallaxes and proper motions are normally distributed. Thus, for each catalogue (Gaia and the external one), the density function is defined as:

$$\phi_G(\mu_{\alpha_G}, \mu_{\delta_G}, \varpi_G) = \frac{1}{\sqrt{(2\pi)^3 |\Sigma_G|}} \exp \left\{ -\frac{1}{2} (\vec{x}_G - \vec{x}_0)^T \Sigma_G^{-1} (\vec{x}_G - \vec{x}_0) \right\} \quad (3.1)$$

$$\phi_E(\mu_{\alpha_E}, \mu_{\delta_E}, \varpi_E) = \frac{1}{\sqrt{(2\pi)^3 |\Sigma_E|}} \exp \left\{ -\frac{1}{2} (\vec{x}_E - \vec{x}_0)^T \Sigma_E^{-1} (\vec{x}_E - \vec{x}_0) \right\} \quad (3.2)$$

$$\text{with } \vec{x}_G = (\mu_{\alpha_G}, \mu_{\delta_G}, \varpi_G), \vec{x}_E = (\mu_{\alpha_E}, \mu_{\delta_E}, \varpi_E), \vec{x}_0 = (\mu_{\alpha_0}, \mu_{\delta_0}, \varpi_0)$$

where  $\vec{x}_0$  is the expected value (which is the same for every star). And  $\Sigma_G$  and  $\Sigma_E$  the covariance matrices:

$$\Sigma_G = \begin{pmatrix} (\sigma_{\mu_\alpha}^2)_G & (\rho_{\mu_\alpha \mu_\delta} \sigma_{\mu_\alpha} \sigma_{\mu_\delta})_G & (\rho_{\mu_\alpha \varpi} \sigma_{\mu_\alpha} \sigma_\varpi)_G \\ (\rho_{\mu_\alpha \mu_\delta} \sigma_{\mu_\alpha} \sigma_{\mu_\delta})_G & (\sigma_{\mu_\delta}^2)_G & (\rho_{\mu_\delta \varpi} \sigma_{\mu_\delta} \sigma_\varpi)_G \\ (\rho_{\mu_\alpha \varpi} \sigma_{\mu_\alpha} \sigma_\varpi)_G & (\rho_{\mu_\delta \varpi} \sigma_{\mu_\delta} \sigma_\varpi)_G & (\sigma_\varpi^2)_G \end{pmatrix} \quad (3.3)$$

$$\Sigma_E = \begin{pmatrix} (\sigma_{\mu_\alpha}^2)_E & (\rho_{\mu_\alpha \mu_\delta} \sigma_{\mu_\alpha} \sigma_{\mu_\delta})_E & (\rho_{\mu_\alpha \varpi} \sigma_{\mu_\alpha} \sigma_\varpi)_E \\ (\rho_{\mu_\alpha \mu_\delta} \sigma_{\mu_\alpha} \sigma_{\mu_\delta})_E & (\sigma_{\mu_\delta}^2)_E & (\rho_{\mu_\delta \varpi} \sigma_{\mu_\delta} \sigma_\varpi)_E \\ (\rho_{\mu_\alpha \varpi} \sigma_{\mu_\alpha} \sigma_\varpi)_E & (\rho_{\mu_\delta \varpi} \sigma_{\mu_\delta} \sigma_\varpi)_E & (\sigma_\varpi^2)_E \end{pmatrix} \quad (3.4)$$

By convolving both distributions (see annex A.2.1 for mathematical details),  $\phi_G(\mu_{\alpha_G}, \mu_{\delta_G}, \varpi_G)$  and  $\phi_E(\mu_{\alpha_E}, \mu_{\delta_E}, \varpi_E)$  we obtain the distribution of the difference  $\phi_C = \phi_G - \phi_E$ :

$$\phi_C(\mu_\alpha, \mu_\delta, \varpi) = \frac{1}{\sqrt{(2\pi)^3 |\Sigma|}} \exp \left\{ -\frac{1}{2} (\vec{x}_G - \vec{x}_E)^T \Sigma^{-1} (\vec{x}_G - \vec{x}_E) \right\} \quad (3.5)$$

with  $\Sigma$  the global covariance matrix:

$$\Sigma = \Sigma_G + \Sigma_E = \left[ \Sigma_{\vec{x}_G - \vec{x}_E} \right]_{ij} = \Sigma_{G_{ij}} + \Sigma_{E_{ij}}$$

(by the properties of covariance matrices, see annex A.2.2)

And the term

$$(\vec{x}_G - \vec{x}_E)^T \Sigma^{-1} (\vec{x}_G - \vec{x}_E)$$

belongs to the quadratic form of the random variables  $x_i$ , i.e. a chi-square variate:

$$(\vec{x}_G - \vec{x}_E)^T \Sigma^{-1} (\vec{x}_G - \vec{x}_E) = \vec{X}^T \Sigma^{-1} \vec{X} = \chi^2 \quad (3.6)$$

To validate the proper motions and parallaxes consistency, we compare the p-value of the  $\chi^2$  for each star to the limit of 0.01 established for all tests. Given eq. 3.6 and the covariance matrices defined in equations 3.3 and 3.4, the  $\chi^2$  follows:

$$\chi^2 = \vec{X}^T \Sigma^{-1} \vec{X}$$

with

$$\vec{X} = \begin{pmatrix} \mu_{\alpha_G} \cos \delta_{\alpha_G} - \mu_{\alpha_E} \cos \delta_{\alpha_E} \\ \mu_{\delta_G} - \mu_{\delta_E} \\ \varpi_G - \varpi_E \end{pmatrix} \quad (3.7)$$

and

$$\Sigma = \begin{pmatrix} (\sigma_{\mu_\alpha}^2)_G + (\sigma_{\mu_\alpha}^2)_E & (\rho_{\mu_\alpha \mu_\delta} \sigma_{\mu_\alpha} \sigma_{\mu_\delta})_G + (\rho_{\mu_\alpha \mu_\delta} \sigma_{\mu_\alpha} \sigma_{\mu_\delta})_E & (\rho_{\mu_\alpha \varpi} \sigma_{\mu_\alpha} \sigma_\varpi)_G + (\rho_{\mu_\alpha \varpi} \sigma_{\mu_\alpha} \sigma_\varpi)_E \\ (\rho_{\mu_\alpha \mu_\delta} \sigma_{\mu_\alpha} \sigma_{\mu_\delta})_G + (\rho_{\mu_\alpha \mu_\delta} \sigma_{\mu_\alpha} \sigma_{\mu_\delta})_E & (\sigma_{\mu_\delta}^2)_G + (\sigma_{\mu_\delta}^2)_E & (\rho_{\mu_\delta \varpi} \sigma_{\mu_\delta} \sigma_\varpi)_G + (\rho_{\mu_\delta \varpi} \sigma_{\mu_\delta} \sigma_\varpi)_E \\ (\rho_{\mu_\alpha \varpi} \sigma_{\mu_\alpha} \sigma_\varpi)_G + (\rho_{\mu_\alpha \varpi} \sigma_{\mu_\alpha} \sigma_\varpi)_E & (\rho_{\mu_\delta \varpi} \sigma_{\mu_\delta} \sigma_\varpi)_G + (\rho_{\mu_\delta \varpi} \sigma_{\mu_\delta} \sigma_\varpi)_E & (\sigma_\varpi^2)_G + (\sigma_\varpi^2)_E \end{pmatrix} \quad (3.8)$$

Whether the observed p-value of the  $\chi^2$  is smaller than  $\alpha_{lim}$ , the null hypothesis is rejected, i.e. the astrometric solution is not good for the given star. Whether the observed p-value is greater, then we accept the measurements.

Finally, a global test is performed to validate if the whole catalogue has a good astrometric solution or not. That is, taking into account the calculated  $\chi^2$  for all the stars of the Gaia catalogue, we check if all together follow a  $\chi^2$  distribution. The test fails or passes with the same criteria mentioned above.

The results obtained using the compilation of VLBI stars when comparing to the *TGAS* simulation and the *Hipparcos* catalogue, may be found in section 4.5 of the VTR (annex B). In both cases the global  $\chi^2$  test failed since, at 1% level of rejection, the astrometric solution of 92.86% stars from the *TGAS* simulation and 43.75% from *Hipparcos*, was not good (proper motions and parallaxes do not match in the considered confidence interval). Such a bad result for the *Hipparcos* stars was studied in more details by C. Turon. She found:

1. an issue with VLBI proper motions observed before 2000 (old calibration maps) which has been removed from the compilation
2. a detailed inspection of the remaining outliers with Simbad showed that they are all peculiar targets: multiple systems, stars on nebulae, etc.

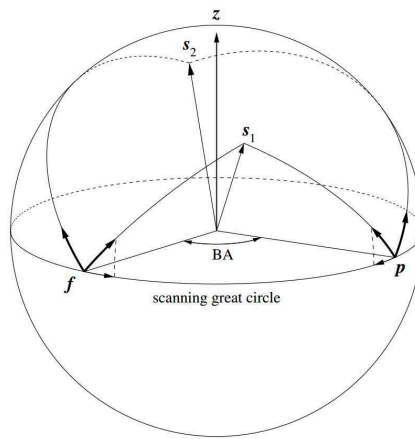
This will be described in [Ruiz-Dern et al., 2017, in prep.]. This test is therefore expected to fail on *TGAS* too.

The same statistical method has also been applied and is being used in other VTS aiming to check Gaia parallaxes and proper motions with respect to, for instance, *Tycho-2* or *Hipparcos* (section 5 in annex B), as well as for quasars (section 9 in annex B).

### 3.3 Parallax Zero-Point and precision

Thanks to the large difference in parallax factor between the two satellite's fields of view, Gaia allows, in principle, to measure absolute parallaxes (as with *Hipparcos*) [Lindgren and de Bruijne, 2005]. This is an advantage in contrast to ground-based observations which need to derive parallaxes relative to background sources. However, as occurred for *Hipparcos* too [Lindgren et al., 1992, Arenou et al., 1995], even small variations in the angle between both telescopes may lead to a global offset of the measured parallaxes [van Leeuwen, 2005, Mignard, 2011]. And these variations exist, for instance, because of the thermal changes in the payload as the satellite rotates (as stated in section 1.1.3). Indeed, along a great circle towards the Sun, stellar parallax causes an apparent shift. Thanks to its one-dimensional measurement principle, Gaia is only sensitive to the relative shift along each great circle, i.e.  $p$  (preceding) and  $f$  (following) points in fig. 3.2. Considering  $s_1$  and  $s_2$  the position of the Sun with respect to the star, at its closest and furthest location to the fields of view, these parallax shifts (indicated by the short arrows in  $p$  and  $f$ ) will be indistinguishable from a slight enlargement or reduction of the *basic angle*, respectively. Therefore, a temporal variation of the *basic angle* due to solar heating, would represent a global shift of the parallaxes [Windmark et al., 2011].

Of course, the satellite was designed to minimize and continuously control these effects by measuring the *basic angle* every few minutes, but even like that its resulting variations cannot be completely eliminated (although according to Gaia specifications the remaining parallax bias should theoretically be much less



**Figure 3.2** – Effect of parallax on stellar images in the preceding ( $p$ ) and following ( $f$ ) field of view of Gaia, when the Sun is in two different positions ( $s_1$  and  $s_2$ ) relative to the stars.  $z$  is the spin axis of Gaia, and the basic angle is marked between the two fields of view. The displacement due to parallax is always directed towards the Sun, as indicated by the thick arrows at  $p$  and  $f$ , and is measured by Gaia as projected along the scanning great circle. Credits from [Windmark et al. \[2011\]](#) (figure 1 of the article)

than  $0.1 \mu\text{as}$  after data processing). Therefore it is of great importance to verify the global zero-point and/or possible parallax bias. Moreover, we saw in section 1.1.5 that these angle variations are actually being larger than expected in the nominal performances. Although the reasons are not yet clearly known, the problem is being correctly handled and should not affect the final data. Nevertheless, this obviously reinforces the motives to carefully validate the Gaia parallax zero-point and their precision (whether the associated standard errors are correctly estimated).

To do so, we implemented two independent tests: the first one uses stars distant enough so that their *parallax uncertainties* (random and systematic) can be considered negligible compared to the Gaia astrometric precision; and the second, accounts for very distant stars, for which their *parallax* can be considered negligible compared to the Gaia astrometric precision. In both cases, we intend to ensure that what is found is only due to Gaia properties.

### 3.3.1 External Distances: RR Lyrae

Here we wanted to validate *TGAS* parallaxes, hence, given that the expected precision of *TGAS* parallaxes is about  $0.1 \text{ mas}$ , we need to select external parallaxes with a precision better than  $0.1 \text{ mas}$  ( $\sigma_{\varpi} < 0.1 \text{ mas}$ , including systematic errors).

#### Using distance modulus

As the star selection process introduces a bias<sup>3</sup>, we initially took advantage of the *calmos* program [[Arenou et al., 1995](#)], which mainly computes the maximum likelihood of the global parallax zero-point (bias) and the associated formal parallax errors (precision) of distant stars, explicitly taking into account all associated truncations on input data. The algorithm also allows to detect outliers (due to parallax, magnitude or classification errors), and to check the quality of the fit between the adopted model and the observations. With this method we may obtain an unbiased estimate of the global zero-point and of the external error. It was developed to validate *Hipparcos* data.

For each star, the conditional probability density function (pdf) to observe the Gaia parallax  $\varpi_G$  given its apparent distance modulus  $t$ <sup>4</sup>, its galactic latitude  $b$ , the unknown global zero-point  $z = \langle \varpi_G - \varpi \rangle$  (with

<sup>3</sup>Selecting photometric parallaxes with a precision smaller than some value implies selecting absolute magnitudes brighter than some value. However, the absolute magnitude is a statistical value, and we would also select preferentially the positive errors, thus biasing the distribution

<sup>4</sup>Note that we use  $t$  notation instead of  $\mu$  for distance modulus to avoid confusion with proper motions, usually referred as  $\mu$ . However, in [Ruiz-Dern et al. \[2015\]](#) (annex B) we did not need to do that difference and  $\mu$  was used for distance modulus as well



$\bar{\omega}$  the true parallax), and the unknown unit-weight error  $k = \left\langle \frac{\sigma_{ext}}{\sigma_G} \right\rangle$  (with  $\sigma_{ext}$  the external error), is:

$$P(\bar{\omega}_G|t, b, z, k) = \frac{P(\bar{\omega}_G, t, b|z, k)}{h(t, b|z, k)} = \frac{P(\bar{\omega}_G, t, b|z, k)}{\int_{-\infty}^{+\infty} P(\bar{\omega}_G, t, b|z, k) d\bar{\omega}_G} \quad (3.9)$$

with

$$P(\bar{\omega}_G, t, b|z, k) = \int_0^{+\infty} P(\bar{\omega}_G, t, b|\bar{\omega}, z, k) P(\bar{\omega}) d\bar{\omega}$$

where

$$P(\bar{\omega}_G, t, b|\bar{\omega}, z, k) = P(\bar{\omega}_G|\bar{\omega}, k, z) P(t|\bar{\omega}) P(b|\bar{\omega})$$

And the conditional probability density functions  $P(\bar{\omega}_G|\bar{\omega}, k, z)$ ,  $P(t|\bar{\omega})$ ,  $P(b|\bar{\omega})$  and  $P(\bar{\omega})$  defined as (see Arenou et al. [1995] for specific details on these calculations):

$$P(\bar{\omega}_G|\bar{\omega}, k, z) = \frac{\mathcal{G}_{\bar{\omega}_G}(\bar{\omega} + z, k\sigma_G)}{\int_{\bar{\omega}_G^-}^{\bar{\omega}_G^+} \mathcal{G}_{\bar{\omega}_G}(\bar{\omega} + z, k\sigma_G) d\bar{\omega}_G}$$

$$P(t|\bar{\omega}) = \frac{\mathcal{G}_t(-5 \ln \bar{\omega} - 5, \sigma_t)}{\int_{t^-}^{t^+} \mathcal{G}_t(-5 \ln \bar{\omega} - 5, \sigma_t) dt}$$

$$P(b|\bar{\omega}) P(\bar{\omega}) = \frac{\cos b}{2 h_Z \bar{\omega}^4} \exp\left(-\frac{|\sin b|}{\bar{\omega} h_Z}\right)$$

with  $h_Z$  the mean scale height, and  $Z$  the third cartesian coordinate ( $X, Y, Z$ ). To derive this  $P(b|\bar{\omega}) P(\bar{\omega})$ , they assumed that the space distribution is independent of the galactic longitude, and used a realistic model of the Galaxy, where the  $Z$ -space distribution was assumed to be an exponential law with  $h_Z$ :  $P(X, Y, Z) \propto \frac{1}{2h_Z} \exp\left(-\frac{|Z|}{h_Z}\right)$ .

In the best case  $z = 0$  and  $k = 1$ .

Together with F. Arenou, a method was developed to call `calmos`, originally written in C, and read it, compile it and execute it from Java. The implemented validation test, beyond the standard tasks of crossmatch, cleaning of NaN values or zero errors, removing duplicates, ..., automatically generated the information needed for `calmos` and used this method to derive the global zero-point and the parallax precision.

However, this method accounts for a particular galactic model ( $P(b|\bar{\omega})$ ) which is not suited to specific objects, such as the RR Lyrae (see later), neither globally to complex sample selections (e.g. stars showing asteroseismic variability), neither to data coming from spectrophotometry which already provide bayesian distance modulus (e.g. distance modulus of *RAVE*<sup>5</sup> stars already consider a more complex galactic model behind). Therefore, together with C. Babusiaux, we implemented a method in R programme language, more flexible but based on the same principles as those of `calmos`. We defined another maximum likelihood function to maximize, where the parameters to estimate are the bias  $z$  (following the notation of previous method) and the extra variance  $q$ :

$$P(z, q|t, \sigma_t, \bar{\omega}_G, \sigma_{\bar{\omega}_G}) \propto \int_{\bar{\omega}} \mathcal{N}[t, -5 \ln \bar{\omega} - 5, \sigma_t] \mathcal{N}\left[\bar{\omega}_G + z, \bar{\omega}, \sqrt{\sigma^2 \bar{\omega}_G + q}\right] d\bar{\omega} \quad (3.10)$$

where  $\bar{\omega}$  is the range of possible parallaxes within  $5\sigma$  confidence interval:

<sup>5</sup>RArial Velocity Experiment, multi-fiber spectroscopic astronomical survey of stars in the Milky Way using the 1.2-metre UK Schmidt Telescope of the Australian Astronomical Observatory (AAO)

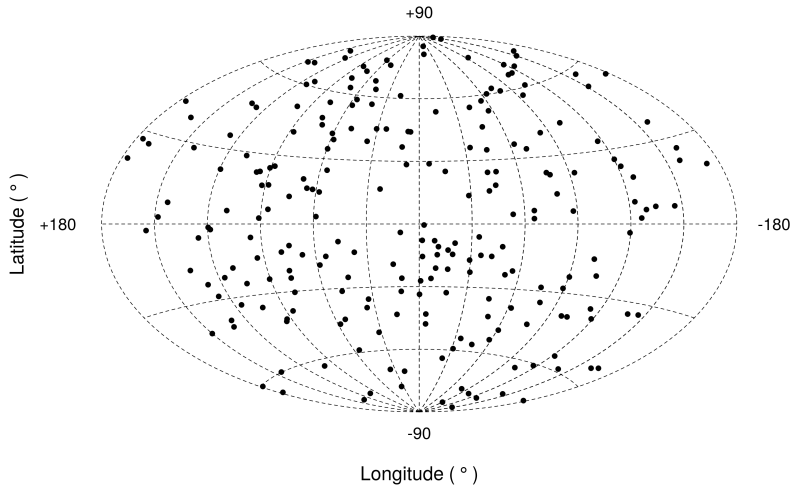
$$\bar{\varpi} \in [\text{Max}[\varpi - 5\sigma, 0], \varpi + 5\sigma]$$

The `mle2R` library is used to derive the corresponding MLE, with discretized steps of  $\min(\sigma_{\bar{\varpi}_G}/10, \sigma_{ext})$ . Basically we are doing the same thing as in `calmos` without accounting for a specific galactic model ( $P(b|\varpi)$ ). The main difference is that here we are testing whether an extra variance  $q$  is needed, while in `calmos` the author looked for a multiplicative factor  $k$ . Therefore, for our method, in the best case  $z = 0$  and  $q = 0$ .

### RR Lyrae sample

RR Lyrae stars are periodic radially pulsating variables. They are old, usually metal-poor, population II stars, connected with low-mass helium-burning stars on the horizontal branch of the CMD, and crossing the classical instability strip. They are abundant in globular clusters and in the halos of galaxies, and a perfect tool for studying the age, formation and structure of their parent stellar system. Moreover, they are widely used as *standard candles* to measure distances within the Milky Way and to Local Group galaxies because they have an almost constant absolute magnitude and their period-luminosity relations are well-known.

We selected RR Lyrae stars from [Maintz \[2005\]](#). Figure 3.3 shows the sky distribution of sample in an Aitoff map in galactic coordinates. This catalogue does not provide parallaxes nor distance modulus information, therefore we derived them through period-luminosity relations. The process and references used are detailed in section 7.2 of [Ruiz-Dern et al. \[2015\]](#) (annex B). Hence, in this section we will just briefly explain the process together with few implemented updates.



**Figure 3.3** – The RR Lyrae sample on an Aitoff map projection in galactic coordinates

Basically, we selected a subset of stars present in *Hipparcos* and *Tycho-2* catalogues, we added *2MASS* photometric information and we derived the Reddening-Independent Absolute Magnitude (equation 3.11):

$$M_{K_{J-K}} = M_K - \frac{A_K}{A_J - A_K} M_{J-K} \quad (3.11)$$

with the extinction calculated through [Schlegel et al. \[1998\]](#) 2D map, and using the extinction coefficients computed in the previous thesis part for Red Clump stars (using Kurucz extinction model and the [Fitzpatrick and Massa \[2007\]](#) extinction law). A sample of 236 stars was retrieved.

The distance modulus is then derived as follows:

$$t = m - M = m_{K_{J-K}} - M_{K_{J-K}}$$

To obtain the apparent magnitude  $m_{K_J-K}$  we used *2MASS* data, and for the absolute magnitude  $M_{K_J-K}$  we took advantage of the period-luminosity and period-abundance-luminosity relations of [Muraveva et al. \[2015\]](#) and [Catelan \[2004\]](#) (those with an \*), respectively, to derive  $M_K$  and  $M_{J-K}$ :

$$M_K = (-2.531 \pm 0.038) \log_{10} P - (0.95 \pm 0.01)$$

$$M_J^* = -0.1409 - 1.7734 \log_{10} P + 0.1899 \log_{10} Z$$

$$M_K^* = -0.5968 - 2.3529 \log_{10} P + 0.1746 \log_{10} Z$$

with  $P$  the period (in days), and  $Z$  the metallicity.

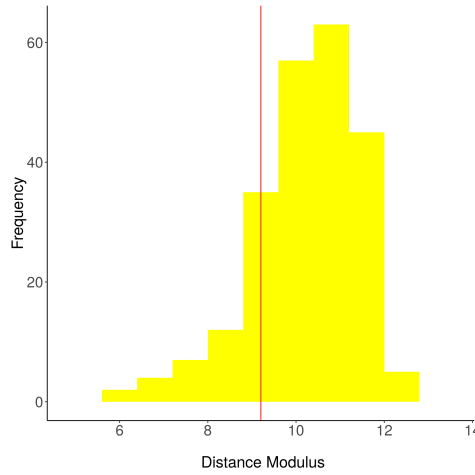
Since [Catelan \[2004\]](#) calibrations ( $M_J^*$ ,  $M_K^*$ ) are theoretical, the source of the final distance modulus uncertainties was only the [Muraveva et al. \[2015\]](#) relation for  $M_K$ , and the conversion between *Glass* and *2MASS* systems of [Catelan \[2004\]](#) relations for  $M_{J-K}$ <sup>6</sup>.

Given the distance modulus  $t$ , we can convert it to parallax in order to cut the sample at  $\sigma_{\varpi_{ext}} < 0.1$ :

$$\varpi = \exp \left\{ -\frac{t+5}{5} \ln(10) \right\} 1000 \quad (3.12)$$

$$\sigma_{\varpi} = \frac{\sigma_t}{5} \ln(10) \varpi \quad (3.13)$$

Figure 3.4 shows the distance modulus distribution of the RR Lyrae sample. The red line indicates the minimum distance value from which stars have  $\sigma_{\varpi_{ext}} < 0.1$  mas (limit established because of the *TGAS* astrometric precision). For the selected RR Lyrae sample, this value corresponds to  $t \sim 9.2$ . Only stars with larger distances, i.e. with  $\sigma_{\varpi}$  mas lower than 0.1mas in the external catalogue, are used for the test.



**Figure 3.4** – RR Lyrae distance modulus distribution. Red line marks the minimum distance from which  $\sigma_{\varpi_{ext}} < 0.1$  mas, which corresponds to  $t = 9.2$ . All stars with larger distances are used for the test

The results obtained when testing over *Hipparcos* and the *TGAS* simulation are shown in section 7.5.3 of chapter 7 of VTR [Ruiz-Dern et al. \[2015\]](#) (annex B). This published version of the test also includes Cepheids [[Ngeow, 2012](#)] and *RAVE* stars as external catalogues (samples prepared by C. Babusiaux). The results obtained can also be found in sections 7.5.2 and 7.5.4 of the VTR, respectively. More catalogues have been added for the DR1 validation: *APOGEE*, *APOKASC*, *LAMOST* and *PASTEL*.

<sup>6</sup>Conversion between the *2MASS* system and the *Glass* system for  $J-K$ :  $(J-K)_{2MASS} = (0.972 \pm 0.006)(J-K)_{Glass} + (-0.011 \pm 0.005)$

For both RR Lyrae and Cepheids the results on the parallax zero-point were good and as expected. However, for RR Lyrae the test indicated that *Hipparcos* parallax errors were underestimated, which translated to an underestimation of the input RRLyrae errors. For that reason, in an updated version of the VTR (Ruiz-Dern et al., 2016, in prep.), we used the period-metallicity-luminosity of Muraveva et al. [2015] instead (equation 17 of their article, eq. 3.14 here):

$$M_K = (-2.53 \pm 0.36) \log_{10} P + (0.07 \pm 0.04) [\text{Fe}/\text{H}] - (0.95 \pm 0.14) \quad (3.14)$$

in which we adopted a mean  $[\text{Fe}/\text{H}] = [-1.0 \pm 0.2]$  and we also assumed an error of  $\sigma_P = 0.1$  days for the period.

Thus:

$$\sigma_{M_K} = \sqrt{-0.36^2 |\log_{10} P| + \sigma_P^2 \frac{-2.53}{\ln(10) P} - 0.04^2 + 0.07 \sigma_{[\text{Fe}/\text{H}]}^2 + 0.14^2}$$

The final distance modulus error is estimated as:

$$\sigma_r = \sqrt{\sigma_{M_K}^2 + \sigma_{m_K}^2 + \frac{A_K}{A_J - A_K} (\sigma_{M_{JK}}^2 + \sigma_{m_J}^2 + \sigma_{m_K}^2)} \quad (3.15)$$

With this new approach the VTS returned "PASS" for *Hipparcos* for all the statistics.

### 3.3.2 Very distant stars: LMC/SMC

If we use even more distant stars we can safely assume that their Gaia parallaxes are zero, and thus use them to check the global zero-point and parallax precision. It is important to notice that, the size of the sample for this kind of test will be reduced with the increase of the Gaia astrometric accuracy at each release (the Gaia LMC/SMC parallaxes will not be able to be considered zero never more).

In this case, the method implemented is basically a check of the correct behaviour of the normalised parallax difference measurements and its errors distributions. To do so, we first derived the weighted mean of this normalised parallax difference:

$$\frac{\bar{w}_{ext} - \bar{w}_{Gaia}}{\sqrt{\sigma_{ext}^2 + \sigma_{Gaia}^2}}$$

and we validated whether this value fall within the reference p-value. If so, it means that external catalogue and Gaia parallaxes are in agreement, i.e. Gaia measurements are not significantly different from 0.

For the extra variance we took advantage of the two-sample Kolmogorov-Smirnov test, which compares the empirical cumulative distribution function (cdf) of the sample with that of a normal cdf. The resulting p-value must be higher than the reference p-value, so that the distribution of the normalized residuals is not significantly different from the normal distribution

Extra statistical checks were applied later in a general way to almost all the VTS (corresponding to the *basic statistics* mentioned in 2.3 and shown in the VTS flow of figure 2.2) in order to provide a complete view of the data behaviour and to complement the output results. All this can be found in section A.1 of VTR's Annex A.

### Large and Small Magellanic Clouds samples

The Large and Small Magellanic Clouds (LMC and SMC) are two irregular dwarf galaxies from the Local Group. They are orbiting the Milky Way every 1500 million years at about 160000 and 200000 ly away, respectively. At these far distances their parallaxes are indeed close to zero ( $\sim 0.016$  for the SMC and  $\sim 0.020$  for the LMC), thus they can be used to check the Gaia parallax zero-point. As indicated above, the number of LMC/SMC stars useful for this test will diminish with the improve of Gaia astrometric accuracy

at each data release, and will then fall in the previous test category (their error will be in principle small enough to be used in the above VTS).

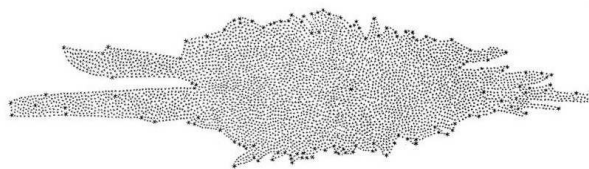
To test the *TGAS* catalogue, three samples were prepared by G. Costigan and C. Turon: 250 LMC Tycho stars, 79 SMC Tycho stars and 46 stars from both LMC and SMC but confirmed to be in the *Hipparcos* catalogues. More details on the construction of these samples may be found in section 8.2 of the VTR, as well as in its Annex B.

Once outliers of the parallax normalised difference were removed at  $5\sigma$  confidence interval, the test was implemented. Similarly as for the RR Lyrae test, the results obtained were already presented in the VTR, in section 8.5. They showed good agreement when tested on *Hipparcos*. For *TGAS*, local variation of the zero point was found. The test was therefore slightly modified to take into account the mean parallax expected for the SMC/LMC and test its influence on the result. Extra filter on the SMC/LMC input catalogue is now also tested using the mean parallax and proper motion of the clouds using a  $\chi^2$  test.

## 3.4 Specific contributions

### 3.4.1 Sky Homogeneity: HEALPix maps

In 1785 William Herschel attempt for the first time to map out the Galaxy, by counting every star he could see along the line of sight of 683 directions to the apparent brightness limit of his telescope ( $\sim 180$  cm diameter). He assumed that all stars had the same intrinsic luminosity so that he could estimate the distance to each star from the Earth. He also assumed that stars do not extend beyond the edges of the Milky Way, and that he could resolve all stars up to these edges. From this he deduced that the number of stars steadily increased toward one side of the sky, in the direction of the Milky Way core, and that we were at the center of the Galactic disk (fig. 3.5).



**Figure 3.5** – Milky Way Map of William Herschel (1785). Credits from *On the Construction of the Heavens* by William Herschel, published in *Philosophical Transactions of the Royal Society of London*, Vol. 75 (1785), pp. 213-266

Although we know today that some of his assumptions and conclusions were wrong, especially because he did not account for absorption by dust (an aspect still unknown at that epoch), this represented an important step in Galactic structure. Since then, several works have been focused on reconstructing the structure of the Galaxy by making star counts of the whole sky, i.e. on studying the star density distribution.

In order to check the Gaia star distribution on the sky, we can compare it to the distribution of other all-sky surveys, such as *Tycho-2*, *2MASS* or *UCAC4*.

To count stars and be able to do this comparison in a proper way, we may take advantage of the HEALPix (Hierarchical Equal Area isoLatitude Pixelation) scheme [Gorski et al., 1999, Górski et al., 2002, 2005]. It is a mathematical structure that allows spatial splitting of the sky in equal area regions, i.e. where each pixel covers the same surface area as every other pixel. It was originally developed to address the data processing and analysis needs of the present generation of cosmic microwave background experiments. Since it was mainly created to allow discretization of functions on a sphere at sufficiently high resolution, and to facilitate fast and accurate statistical and astrophysical analysis of massive full-sky data sets, it can also be used for other big surveys such as Gaia [O’Mullane et al., 2001].

HEALPix may be seen as a division of the sphere into 12 base pixels, which are in turn subdivided up to

the desired resolution (the lowest resolution would be the 12 base pixels). For each pixel subdivision, four new ones are created, while maintaining the original pixel boundaries. These boundaries are non-geodesic and take the simple form:

$$\cos \Theta = a + \frac{b}{\Phi^2} \quad \text{in the North cap}$$

$$\cos \Theta = a \pm b \Phi \quad \text{in the equatorial zone}$$

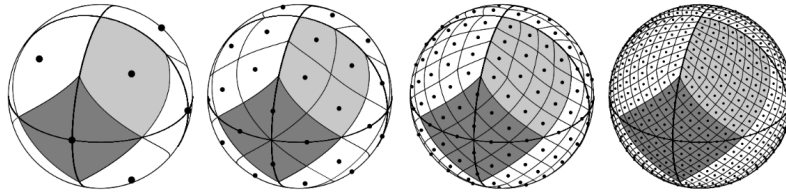
$$\cos \Theta = a + \frac{b}{(\pi/2 - \Phi)^2} \quad \text{in the South cap}$$

with  $\Theta$  and  $\Phi$  the spherical coordinates.

This allows to verify by analytical integration the exact area equality among pixels, and to compute efficiently more complex objects.

The amount of subdivision is given by the  $N_{side}$  number, where  $N_{side}$  is always a power of 2 (hence 1,2,4,8,16,32,...). Equation 3.16 gives the resolution (number of pixels) as a function of  $N_{side}$ . Thus, this resolution increases by three steps from the base level, as we show for instance in figure 3.6, where the sphere is partitioned, from left to right, into 12, 48, 192 and 768 pixels.

$$N_{pixel} = 12 N_{side}^2 \quad (3.16)$$



**Figure 3.6** – Orthographic view of the HEALPix tessellation of a sphere. The overplot of the equator and meridians illustrates the octahedral symmetry of the HEALPix construction. Light-grey shading shows one of the eight (four north, and four south) identical polar base resolution pixels. Dark-grey shading shows one of the four identical equatorial base-resolution pixels. From left to right, the grid is hierarchically subdivided with the grid resolution parameter equal to  $N_{side} = 1, 2, 4, 8$ , and the total number of pixels equal to  $N_{pixel} = 12 N_{side}^2 = 12, 48, 192, 768$ . All pixel centers are located on  $N_{ring} = 4 N_{side} - 1$  rings of constant latitude. Within each panel the areas of all pixels are identical. Figure adapted from Górski et al. [2002] (fig. 1 of the article)

Pixels are distributed on lines of constant latitude. This iso-latitude nature of HEALpix means that pixels align on rings. HEALPix lays a grid on the sphere so any pixel may be referred to by its ring number and offset, or  $i_{\Theta}$  and  $i_{\Phi}$  as it is sometimes referred to. All pixel centers are located on  $N_{ring} = 4 N_{side} - 1$  rings of constant latitude. This property is essential to increase the speed of computation of all harmonic analysis applications involving spherical harmonics [Górski et al., 2005].

Indeed, specific geometrical properties allow HEALPix to support two different numbering schemes for the pixels Gorski et al. [1999]:

- **RING**: pixels are counted from the north to the south pole along each is-latitude ring. As indicated above, this scheme allows to easy implement Fourier transforms with spherical harmonics
- **NESTED**: pixel indices are arranged in twelve tree structures, belonging to the mentioned base-resolution pixels. This structure allows to implement, for instance, all applications involving nearest-neighbour searches

In order to use both HEALPix schemes for sky homogeneity comparison with external catalogues, we took advantage of the *GaiaTools* package developed by the Gaia Coordination Unit 1 (CU1). As *TGAS*

had no vocation to be complete, in the framework of this thesis (mainly focused on *TGAS* validation) we set up the basis to access the specific HEALPix tools and create and HEALPix dataset from WP944, to be used later for DR1 and next releases. As a test for the implementation, we used the NESTED scheme with *Tycho-2* data.

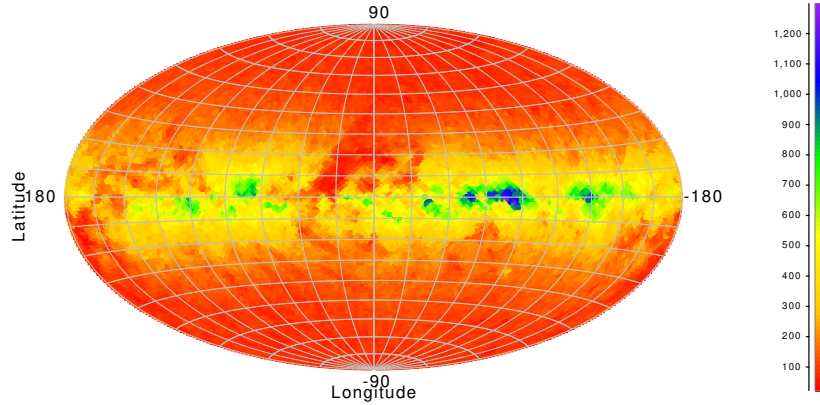
The equatorial coordinates (in radians) were converted to galactic (l,b), and then to spherical ( $\Theta$ ,  $\Phi$ ) through:

$$\Theta = (90 - b) (\pi/180) \quad \text{with } \Theta \in [0, \pi], \quad \Theta = 0: \text{ north pole}$$

$$\Phi = l \pi/180 \quad \text{with } \Phi \in [0, 2\pi]$$

Using the angular position of  $\Theta$  and  $\Phi$ , we determined the pixel number  $i_{pixel}$  containing the location, i.e. the point on the sphere at coordinates  $\Theta$  and  $\Phi$  given the map resolution parameter  $N_{side}$ . We used  $N_{side} = 16$  which corresponds (following equation 3.16) to 3072 pixels.

Given the  $i_{pixel}$  number for each pair of coordinates, we counted how many times each  $i_{pixel}$  index appeared. Then we generated an Aitoff Map Projection in galactic coordinates using the counts obtained in every HEALPix region. Figure 3.7 shows the HEALPix map obtained for *Tycho-2*.



**Figure 3.7** – HEALPix map generated for *Tycho-2* catalogue, with  $N_{side} = 16$ , belonging to a resolution of 3072 pixels

For DR1 those HEALPix counts have been used in particular to see the distribution of the duplicated and missing sources in the Gaia catalogue.

### 3.4.2 Epoch Transformation

When comparing with other catalogues one may take into account in which epoch the celestial coordinates were defined, i.e. in which moment in time the observation was made or the moment for which a prediction was calculated.

As the Earth rotates, its rotation axis gradually rotates as well although much slower than the daily rotation we see. Because of this precession, the positions of the stars change over time with a small motion every day. On small timescales this motion is not noticeable, however over decades it is: inexactitudes and other errors accumulate, creating more and greater errors of prediction, so ephemeris factors need to be recalculated from time to time, and that requires a new epoch to be defined.

Each survey gives its astrometric data in the epoch of the catalogue. For instance, the *Hipparcos* epoch was J1991.25, i.e. celestial coordinates as they were in 1991. And Gaia will provide coordinates in the new epoch J2015.0. To appropriately compare each EC to Gaia, we first need to transform the position of the stars as they are given in the EC epoch into the one expected for the Gaia epoch, by taking into account the proper motion of the stars. In WP944 we only considered so far catalogues in the ICRF reference frame.

This section describes the general procedure implemented for the transformation of the parameters from one initial epoch  $T_0$  to the arbitrary epoch  $T$ . It is based on the rigorous treatment employed for the *Hipparcos* and *Tycho* catalogues [Perryman et al., 1997, Vol. 1, Sect. 1.5.5].

### Astrometric Standard Model

In the standard astrometric model it is assumed that stars move through the space with a constant velocity vector, following a uniform rectilinear motion with respect to the solar-system baricenter. In reality this velocity may vary because of perturbations of other close sources, such as planets or companion stars, or even gravity fields from other stars. However, in general and for the majority of stars, these effects use to be negligible over time spans of century or less, even at Gaia precision level. For the purposes of testing Gaia positions, parallaxes and proper motions of single systems, and given that we always account for the external catalogues precision when confronting both archives, the standard model can be applied.

The barycentric position of a source at the epoch  $T$  is

$$\vec{b}(T) = \vec{b}(T_0) + (T - T_0) \vec{v} \quad (3.17)$$

where  $\vec{b}(T_0)$  is the barycentric position at the initial epoch  $T_0$  ( $b_0$ ) and  $\vec{v}$  the constant space velocity. Variables with subscript 0 belong quantities at  $T_0$  whilst the others correspond to epoch  $T$ . The difference of epochs is denoted as  $t$ :

$$t = T - T_0$$

Thus,

$$\vec{b} = \vec{b}_0 + t \vec{v}. \quad (3.18)$$

The expression for the space velocity in terms of the source parameters reads

$$\vec{v} = \frac{A}{\omega_0} (\vec{\mu}_0 + \vec{r}_0 \mu_{r_0}) = \frac{A}{\omega_0} (\vec{p}_0 \mu_{\alpha_0} + \vec{q}_0 \mu_{\delta_0} + \vec{r}_0 \mu_{r_0}), \quad (3.19)$$

where  $\vec{\mu}_0$  is the proper motion at the initial epoch,  $r_0$ ,  $p_0$  and  $q_0$  are the unit vectors constituting the orthogonal vector triad  $[\vec{p}_0, \vec{q}_0, \vec{r}_0]$ , known as *normal triad* (eq. 3.20), and  $A = 4.740470446$  equals the astronomical unit expressed in  $\text{km yr s}^{-1}$ . This relation implies that the parallax and proper motions are expressed in compatible units, for instance, mas and mas/yr, respectively.

$$\vec{r}_0 = \begin{pmatrix} \cos \delta_0 \cos \alpha_0 \\ \cos \delta_0 \sin \alpha_0 \\ \sin \delta_0 \end{pmatrix}, \quad \vec{p}_0 = \begin{pmatrix} -\sin \alpha_0 \\ \cos \alpha_0 \\ 0 \end{pmatrix}, \quad \vec{q}_0 = \begin{pmatrix} -\sin \delta_0 \cos \alpha_0 \\ -\sin \delta_0 \sin \alpha_0 \\ \cos \delta_0 \end{pmatrix} \quad (3.20)$$

### Propagation of the barycentric direction

Squaring both sides of the equation of propagation of the barycentric direction  $\vec{u} = \vec{b}/b$  (eq. 3.18), and using the relations  $\vec{b}'_0 \vec{v}_0 = b_0^2 \mu_{r_0}$  and  $v_0^2 = b_0^2 (\mu_{\alpha_0}^2 + \mu_{\delta_0}^2)$ , we can write:

$$b^2 = b_0^2 (1 + 2\mu_{r_0} t + (\mu_{\alpha_0}^2 + \mu_{\delta_0}^2) t^2),$$

where  $\mu_0^2 = \mu_{\alpha_0}^2 + \mu_{\delta_0}^2$ . Introducing the distance factor

$$f = b_0/b = [1 + 2\mu_{r_0} t + (\mu_{\alpha_0}^2 + \mu_{\delta_0}^2) t^2]^{-1/2}, \quad (3.21)$$

the propagation of the barycentric direction becomes:



$$\vec{u} = [\vec{r}_0 (1 + \mu_{r_0} t) + \vec{\mu}_0 t] f \quad (3.22)$$

and the propagation of the parallax:

$$\varpi = \varpi_0 f \quad (3.23)$$

The celestial coordinates  $(\alpha, \delta)$  at epoch  $T$  are obtained from  $\vec{u}$  in the usual manner and given expressions of the unit vectors of equation 3.20, the normal triad associated with the propagated direction is:

$$[\vec{p}, \vec{q}, \vec{r}] = \begin{bmatrix} -\sin \alpha & -\sin \delta \cos \alpha & \cos \delta \cos \alpha \\ \cos \alpha & -\sin \delta \sin \alpha & \cos \delta \sin \alpha \\ 0 & \cos \delta & \sin \delta \end{bmatrix}. \quad (3.24)$$

Differentiating 3.22 we obtain the propagated proper motion vector:

$$\vec{\mu} = \frac{d\vec{u}}{dt} = [\vec{\mu}_0 (1 + \mu_{r_0} t) - \vec{r}_0 \mu_0^2 t] f^3, \quad (3.25)$$

and the propagated radial proper motion is found to be

$$\mu_r = \frac{db}{dt} \frac{\varpi}{A} = [\mu_{r_0} + (\mu_0^2 + \mu_{r_0}^2) t] f^2. \quad (3.26)$$

To obtain the proper motion components  $(\mu_{\alpha_*}, \mu_{\delta})$  from vector  $\vec{\mu}$  it is necessary to resolve the latter along the tangential vectors  $\vec{p}$  and  $\vec{q}$  at the propagate direction, i.e. with the tangent point located at  $\vec{u}$  and with  $\vec{p}$  pointing towards the local north:

$$\mu_{\alpha_*} = \vec{p}' \vec{\mu}, \quad \mu_{\delta} = \vec{q}' \vec{\mu}.$$

The tangential vectors are defined in terms of the propagated  $\vec{u}$  or  $(\alpha, \delta)$  at the epoch  $T$  according to 3.24.

The above formulae describe the transformation of  $(\alpha_0, \delta_0, \varpi_0, \mu_{\alpha_0}, \mu_{\delta_0}, \mu_{r_0})$  at epoch  $T_0$  into  $(\alpha, \delta, \varpi, \mu_{\alpha_*}, \mu_{\delta}, \mu_r)$  at the arbitrary epoch  $T = T_0 + t$ . This is completely reversible, so that a second transformation from  $T$  to  $T_0$  recovers the original six parameters.

### Propagation of covariance

The elements of the variance-covariance matrix can be reconstructed from the standard errors and correlations by:

$$C_{ii} = \sigma_i^2 \quad C_{ij} = \rho_{ij} \sigma_i \sigma_j$$

The covariance matrix  $\vec{C}_0$  of the initial parameters and the matrix  $\vec{C}$  of the propagated parameters are related as

$$\vec{C} = \vec{J} \vec{C}_0 \vec{J}' \quad (3.27)$$

where  $\vec{J}$  is the Jacobian matrix of the source parameter transformation:

$$\vec{J} = \frac{\partial (\alpha, \delta, \varpi, \mu_{\alpha_*}, \mu_{\delta}, \mu_r)}{\partial (\alpha_0, \delta_0, \varpi_0, \mu_{\alpha_0}, \mu_{\delta_0}, \mu_{r_0})}$$

That is, the uncertainties in the source parameters  $\alpha, \delta, \varpi, \mu_{\alpha_*}, \mu_{\delta}, \mu_r$  and correlations between them are quantified by means of the  $6 \times 6$  covariance matrix  $\vec{C}$ . Therefore, the propagation of the covariances requires the calculation of all 36 partial derivatives constituting the Jacobian  $\vec{J}$  (see below).

In order to calculate the covariance matrix of the propagated astrometric parameters  $\vec{C}$ , the initial covariance matrix  $\vec{C}_0$  must be defined, using the parameters, their uncertainties and the correlation coefficient  $\rho$ . If not all parameters are known, the remaining ones must be treated as essentially unknown, which means that they can be set to some arbitrary values with very large uncertainties. Following Michalik et al. [2015], we set to zero all unknown parameters, with uncertainties  $\sigma_{\varpi} = 1000mas$ ,  $\sigma_{\mu_{\alpha_*}} = \sigma_{\mu_{\delta}} = 1000mas/yr$ , and  $\sigma_{\mu_r} = \sigma_{v_r} \sigma_{\varpi} / A$ , where  $\sigma_{v_r} = 100km/s$  is the prior radial velocity uncertainty.

Finally, the elements of the Jacobian matrix are:

$$\begin{aligned}
J_{11} &= \frac{\partial \alpha_*}{\partial \alpha_{*0}} = \vec{p}' \vec{p}_0 (1 + \mu_{r_0} t) f - \vec{p}' \vec{r}_0 \mu_{\alpha_*} t f \\
J_{12} &= \frac{\partial \alpha_*}{\partial \delta_0} = \vec{p}' \vec{q}_0 (1 + \mu_{r_0} t) f - \vec{p}' \vec{r}_0 \mu_{\delta_0} t f \\
J_{13} &= \frac{\partial \alpha_*}{\partial \varpi_0} = 0 \\
J_{14} &= \frac{\partial \alpha_*}{\partial \mu_{\alpha_*}} = \vec{p}' \vec{p}_0 t f \\
J_{15} &= \frac{\partial \alpha_*}{\partial \mu_{\delta_0}} = \vec{p}' \vec{q}_0 t f \\
J_{16} &= \frac{\partial \alpha_*}{\partial \mu_{r_0}} = -\mu_{\alpha_*} t^2 \\
J_{21} &= \frac{\partial \delta}{\partial \alpha_{*0}} = \vec{q}' \vec{p}_0 (1 + \mu_{r_0} t) f - \vec{q}' \vec{r}_0 \mu_{\alpha_*} t f \\
J_{22} &= \frac{\partial \delta}{\partial \delta_0} = \vec{q}' \vec{q}_0 (1 + \mu_{r_0} t) f - \vec{q}' \vec{r}_0 \mu_{\delta_0} t f \\
J_{23} &= \frac{\partial \delta}{\partial \varpi_0} = 0 \\
J_{24} &= \frac{\partial \delta}{\partial \mu_{\alpha_*}} = \vec{q}' \vec{p}_0 t f \\
J_{25} &= \frac{\partial \delta}{\partial \mu_{\delta_0}} = \vec{q}' \vec{q}_0 t f \\
J_{26} &= \frac{\partial \delta}{\partial \mu_{r_0}} = -\mu_{\delta} t^2 \\
J_{31} &= \frac{\partial \varpi}{\partial \alpha_{*0}} = 0 \\
J_{32} &= \frac{\partial \varpi}{\partial \delta_0} = 0 \\
J_{33} &= \frac{\partial \varpi}{\partial \varpi_0} = f \\
J_{34} &= \frac{\partial \varpi}{\partial \mu_{\alpha_*}} = -\varpi \mu_{\alpha_*} t^2 f^2 \\
J_{35} &= \frac{\partial \varpi}{\partial \mu_{\delta_0}} = -\varpi \mu_{\delta_0} t^2 f^2 \\
J_{36} &= \frac{\partial \varpi}{\partial \mu_{r_0}} = -\varpi (1 + \mu_{r_0} t) t f^2 \\
J_{41} &= \frac{\partial \mu_{\alpha_*}}{\partial \alpha_{*0}} = -\vec{p}' \vec{p}_0 \mu_0^2 t f^3 - \vec{p}' \vec{r}_0 \mu_{\alpha_*} (1 + \mu_{r_0} t) f^3 \\
J_{42} &= \frac{\partial \mu_{\alpha_*}}{\partial \delta_0} = -\vec{p}' \vec{q}_0 \mu_0^2 t f^3 - \vec{p}' \vec{r}_0 \mu_{\delta_0} (1 + \mu_{r_0} t) f^3 \\
J_{43} &= \frac{\partial \mu_{\alpha_*}}{\partial \varpi_0} = 0 \\
J_{44} &= \frac{\partial \mu_{\alpha_*}}{\partial \mu_{\alpha_*}} = \vec{p}' \vec{p}_0 (1 + \mu_{r_0} t) f^3 - 2 \vec{p}' \vec{r}_0 \mu_{\alpha_*} t f^3 - 3 \mu_{\alpha_*} \mu_{\alpha_*} t^2 f^2
\end{aligned}$$

$$\begin{aligned}
J_{45} &= \frac{\partial \mu_{\alpha_*}}{\partial \mu_{\delta_0}} = \vec{p}' \vec{q}_0 (1 + \mu_{r_0} t) f^3 - 2 \vec{p}' \vec{r}_0 \mu_{\delta_0} t f^3 - 3 \mu_{\alpha_*} \mu_{\delta_0} t^2 f^2 \\
J_{46} &= \frac{\partial \mu_{\alpha_*}}{\partial \mu_{r_0}} = \vec{p}' [\vec{\mu}_0 f - 3 \vec{\mu} (1 + \mu_{r_0} t)] t f^2 \\
J_{51} &= \frac{\partial \mu_{\delta}}{\partial \alpha_{*0}} = -\vec{q}' \vec{p}_0 \mu_0^2 t f^3 - \vec{q}' \vec{r}_0 \mu_{\alpha_{*0}} (1 + \mu_{r_0} t) f^3 \\
J_{52} &= \frac{\partial \mu_{\delta}}{\partial \delta_0} = -\vec{q}' \vec{q}_0 \mu_0^2 t f^3 - \vec{q}' \vec{r}_0 \mu_{\delta_0} (1 + \mu_{r_0} t) f^3 \\
J_{53} &= \frac{\partial \mu_{\delta}}{\partial \varpi_0} = 0 \\
J_{54} &= \frac{\partial \mu_{\delta}}{\partial \mu_{\alpha_{*0}}} = \vec{q}' \vec{p}_0 (1 + \mu_{r_0} t) f^3 - 2 \vec{q}' \vec{r}_0 \mu_{\alpha_{*0}} t f^3 - 3 \mu_{\delta} \mu_{\alpha_{*0}} t^2 f^2 \\
J_{55} &= \frac{\partial \mu_{\delta}}{\partial \mu_{\delta_0}} = \vec{q}' \vec{q}_0 (1 + \mu_{r_0} t) f^3 - 2 \vec{q}' \vec{r}_0 \mu_{\delta_0} t f^3 - 3 \mu_{\delta} \mu_{\delta_0} t^2 f^2 \\
J_{56} &= \frac{\partial \mu_{\delta}}{\partial \mu_{r_0}} = \vec{q}' [\vec{\mu}_0 f - 3 \vec{\mu} (1 + \mu_{r_0} t)] t f^2 \\
J_{61} &= \frac{\partial \mu_r}{\partial \alpha_{*0}} = 0 \\
J_{62} &= \frac{\partial \mu_r}{\partial \delta_0} = 0 \\
J_{63} &= \frac{\partial \mu_r}{\partial \varpi_0} = 0 \\
J_{64} &= \frac{\partial \mu_r}{\partial \mu_{\alpha_{*0}}} = 2 \mu_{\alpha_{*0}} (1 + \mu_{r_0} t) t f^4 \\
J_{65} &= \frac{\partial \mu_r}{\partial \mu_{\delta_0}} = 2 \mu_{\delta_0} (1 + \mu_{r_0} t) t f^4 \\
J_{66} &= \frac{\partial \mu_r}{\partial \mu_{r_0}} = \left[ (1 + \mu_{r_0} t)^2 - \mu_0^2 t^2 \right] f^4
\end{aligned}$$



# Improving the Gaia Data Release 1 solution

## Contents

---

<b>4.1</b>	<b>CU9 validation releases</b> . . . . .	<b>166</b>
<b>4.2</b>	<b>Parallaxes</b> . . . . .	<b>166</b>
4.2.1	Under-estimated uncertainties . . . . .	166
4.2.2	Local zero-point variations . . . . .	167
<b>4.3</b>	<b>Proper Motions</b> . . . . .	<b>168</b>
4.3.1	Under-estimated errors . . . . .	168
4.3.2	Missing High Proper Motion stars . . . . .	169
4.3.3	Gaia reference system rotation . . . . .	169
<b>4.4</b>	<b>Sky Coverage</b> . . . . .	<b>170</b>
4.4.1	Missing sources . . . . .	170
4.4.2	Duplicated sources . . . . .	171
<b>4.5</b>	<b>Outliers</b> . . . . .	<b>172</b>
4.5.1	Astrometric . . . . .	172
4.5.2	Photometric . . . . .	172

---

## 4.1 CU9 validation releases

Several *TGAS* preliminary releases have been delivered to the CU9 for validation purposes since mid-2015. That way we could check the parameters behaviour and detect possible data issues before the first Gaia official publication. The validation outputs of each preliminary release have allowed us to both improve the data reduction (thus the quality of the final catalogue) and the VTS by including more tests, homogenizing criteria for figures and outputs <sup>1</sup>, adding new features to be tested or using other external catalogues to ensure and reinforce the reliability of the results.

On the 9th of May 2016, the preliminary Data Release 1 was provided to CU9 for validation. This included the *TGAS* 04.00 solution. The Gaia mission is split artificially in consecutive time segments (of roughly six months): the observations of each segment are the input for the iterative data processing cycles. Gaia Data Release 1 (hereafter GDR1) includes the observations taken during segments 0 and 1, which correspond to the period 25th of July 2014 - 16th of September 2015.

In this chapter we aim to highlight the data quality improvements achieved from one preliminary release to another (up to the GDR1 solution), thanks to the continuous validations performed within the framework of the CU9 validation team. We will focus only on those aspects, tests and external catalogues referred to in chapter 3. The comparisons are mainly done between the *TGAS* 00.01 version and the GDR1 before and after CU9 filtering (hereafter GDR1-BF and GDR1-AF, respectively) . A number of issues found in GDR1-BF could be solved (or at least reduced) by simply removing some problematic sources. The CU9 filtering decision is discussed in detail in [Arenou et al. \[2016b\]](#). The CU9 filtering removed duplicated sources, sources with a too small number of observations for the G band and sources with obvious colour problems from BP/RP information. GDR1-AF is the one that will be published in September 2016. The *TGAS* 00.01 is the second preliminary release, covering all data from the 25th of July 2014 to the 29th of May 2015 ( $\sim 10$  months of data). It was provided to the CU9 on the 7th of August 2015. As stated above, GDR1 was delivered nine months later, covering  $\sim 14$  months of data. Table 4.1 summarizes the different releases delivered to CU9, their version in WP944, their associated Validation Test Report and the period of data they include.

**Table 4.1** – CU9 validation releases

DR name	WP944 version	VTR document	Period
TGAS 00.00	18.0	-	25/07/2014 - 29/05/2015 (10 months)
TGAS 00.01	18.1	<a href="#">Arenou et al. [2015]</a>	
TGAS 01.00	18.4	<a href="#">Arenou et al. [2016a]</a>	
GDR1 (TGAS 04.00)	18.5	GDR1-BF: <a href="#">Arenou et al. [2016b]</a> GDR1-AF: <a href="#">Arenou et al. [2016d]</a>	25/07/2014 - 16/09/2015 (14 months)

## 4.2 Parallaxes

### 4.2.1 Under-estimated uncertainties

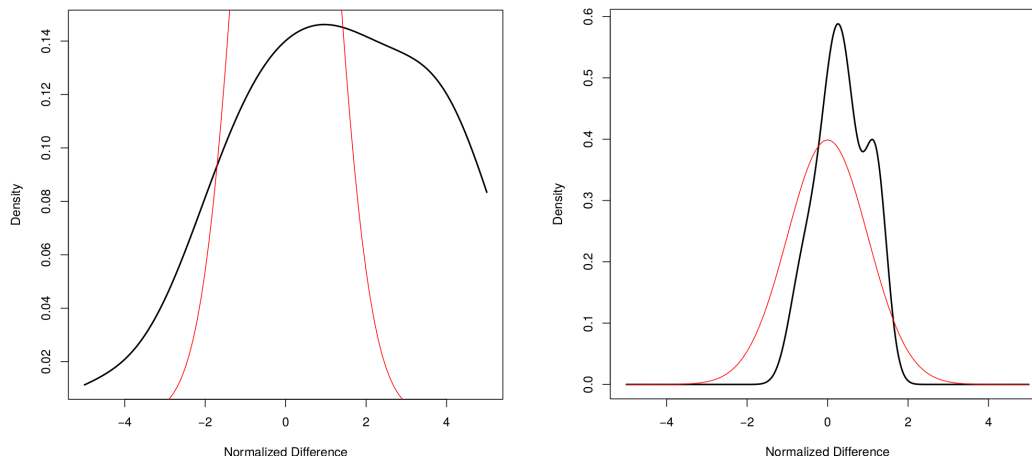
In the first *TGAS* releases for validation (*TGAS* 00.00, *TGAS* 00.01) a global under-estimation of parallax uncertainties was observed. Several VTS, among them the one checking the parallax consistency of VLBI stars and those using distant stars such as the RR Lyrae and stars in the LMC/SMC, failed because Gaia parallax uncertainties were smaller than expected. Therefore, all tests indicated that an extra variance was necessary.

For instance the test with RR Lyrae stars gave an extra dispersion of 0.09mas. The VLBI test, and two more comparing *TGAS* with both *Hipparcos* and quasars (both detailed in VTR [Ruiz-Dern et al. \[2015\]](#)

<sup>1</sup>For instance, the direction in which the bias and residuals are calculated. Currently: External catalogue - Gaia

attached in annex B) indicated a significant extra variance.

In TGAS 01.00 almost all these issues disappeared. The extra variance for VLBI stars was reduced to  $-0.02 \pm 0.03 \text{ mas}^2$ , that of RR Lyrae variable stars to  $-0.01 \pm 0.01 \text{ mas}^2$ , and those of LMC and SMC stars to  $-0.021 \pm 0.009 \text{ mas}^2$  (LMC),  $-0.0005 \pm 0.02 \text{ mas}^2$  (SMC) and  $-0.03 \pm 0.004 \text{ mas}^2$  stars from LMC/SMC confirmed to be in *Hipparcos*). We see that all these values are now negative, meaning that the uncertainties are just slightly overestimated. Fig. 4.1 shows the improvement in the width of the parallax distribution for the *Hipparcos* LMC/SMC sample obtained between TGAS 00.01 (left) and GDR1-BF (right). In the case of TGAS sources with VLBI counterparts, it is important to remark that the number of stars has been divided by two since TGAS 01.00, because most of them were peculiar and had large parallax uncertainties, i.e. they were filtered and removed.



**Figure 4.1** – Normalized density distribution of the parallax differences (difference divided by the formal error on this difference) between LMC/SMC (*Hipparcos*) and both TGAS 00.01 (left) and GDR1-BF (right). The red curve represents a normal distribution of unit variance. Note that the figures are not in the same scale

The *Hipparcos* (New reduction of van Leeuwen [2007]) sample used to confront with TGAS was defined by selecting only well-behaved stars: i.e. those with 5-parameter solution type with good astrometric solution (goodness of fit,  $|F2| < 5$ ), those not flagged as binaries according to the literature, and for those stars that were also in *Tycho-2* only those with consistent proper motions between both catalogues (at a p-value limit of 0.001) were kept. However, contrary to other VTS, this one is still pointing out a significant positive extra variance of  $0.34 \pm 0.005$ , hence to an underestimation of the parallax uncertainties. The results are better but there is still a non-zero parallax normalized residuals. Nonetheless, it doesn't mean that the fault is on the TGAS only as *Hipparcos* parallax uncertainties may also be slightly underestimated [by no more than 5%, Arenou et al., 1995, Arenou et al., 1997], i.e. increasing the uncertainty by about 0.3 mas ( $0.1 \text{ mas}^2$  extra variance). With the new reduction of van Leeuwen [2007] the uncertainty was improved although few tenths of mas cannot be excluded either [van Leeuwen, 2005].

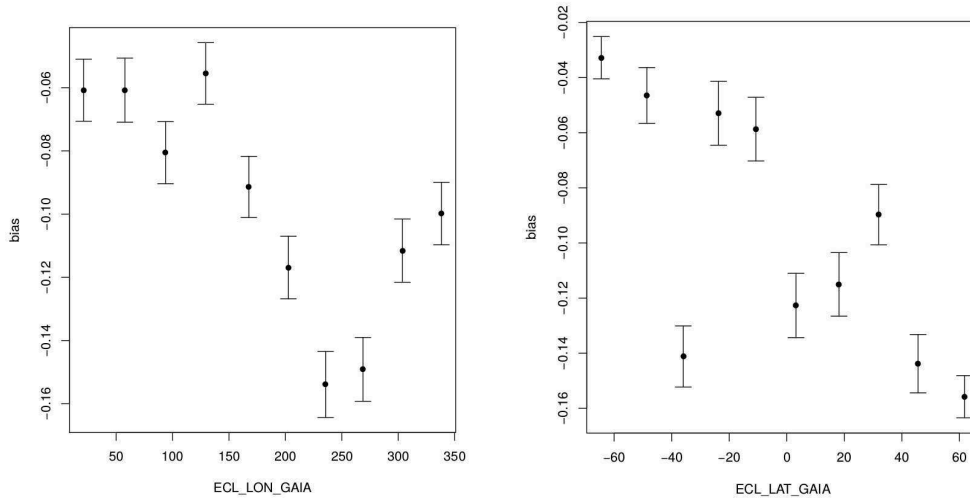
The last reprocessing (GDR1-BF and GDR1-AF), shows the same behaviour. Table 4.2 summarizes all these results for the GDR1 (TGAS 04.00) final solution.

## 4.2.2 Local zero-point variations

We have also checked the parallax accuracy by testing its zero-point with stars distant enough so that their estimated photometric parallax uncertainty (through the period-luminosity relation in the case of RR Lyrae stars) is smaller than 0.1 mas. The goal is to verify whether a bias should be added to the Gaia parallaxes to be consistent with the distance modulus estimates. Similarly, to test this feature, we have also used stars from the LMC/SMC which are even more distant. In this case, we assumed that their measured parallaxes cannot be distinguished from zero at the GDR1 expected precision level, i.e. the normalized parallax distribution of those stars should follow a normal distribution.

A significant parallax bias of  $-0.08$  mas was detected (at  $3\sigma$ ) for RR Lyrae in the first TGAS releases, due to the under-estimation of uncertainties explained above in section 4.2.1. It has been recently reduced down to  $-0.06 \pm 0.03$  mas in GDR1 release, which is no more sufficient to prove zero-point bias at  $3\sigma$ , as the number of RR Lyrae is not large enough. But when combining these stars with the Cepheids sample the zero-point bias becomes significant. For the LMC/SMC stars a bias of  $0.12 \pm 0.02$  mas (LMC),  $-0.12 \pm 0.050$  mas (SMC) and  $0.10 \pm 0.02$  mas (*Hipparcos* LMC/SMC stars) has been found.

Moreover, these results differ from the bias found through RR Lyrae stars, and also differ from the value found for the *Hipparcos* stars when comparing them to TGAS, which was  $-0.094 \pm 0.004$  mas (fig. 4.2). This shows a clear variation of the zero-point over the sky. Indeed, thanks to other tests too (e.g. quasars and clusters), we are observing a variation of at least a  $0.3$  mas amplitude of the parallax zero-point. In some cases, these biases can reach up to  $\pm 1$  mas in  $2^\circ$  radius regions [Arenou et al., 2016b].



**Figure 4.2** – Parallax bias for *Hipparcos* stars in GDR1-BF as a function of the Gaia ecliptic longitude (left, "ECL\_LON\_GAIA") and the ecliptic latitude (right, "ECL\_LAT\_GAIA"). The mean bias is  $-0.094 \pm 0.004$  mas

These variations at a local level mean that, for GDR1, we will have to pay attention on the samples used for scientific exploitation. Depending on the scientific purposes, we will not be able to use samples from the same sky region because the parameters will be strongly correlated. We will have to select all-sky samples instead. This is especially important, for instance, for our colour-colour Red Clump calibrations derived in the first part of this thesis, in order to provide accurate and absolute calibrations.

The origin of these local variations has been analysed and discussed with other WPs and CUs, and may be due to several problems not yet handled for the GDR1 solution: attitude modelling errors (e.g. micro-clanks and micro-meteoroid hits), colour-dependent PSF shifts (chromaticity) and basic-angle variations, which conspire to produce spatially-dependent systematics.

Whereas locally we are observing these variations, globally many independent estimation methods from several tests suggest a global negative parallax zero-point of  $\sim -0.04$  mas. Table 4.2 summarizes all these results for the GDR1 (TGAS 04.00) final solution.

## 4.3 Proper Motions

### 4.3.1 Under-estimated errors

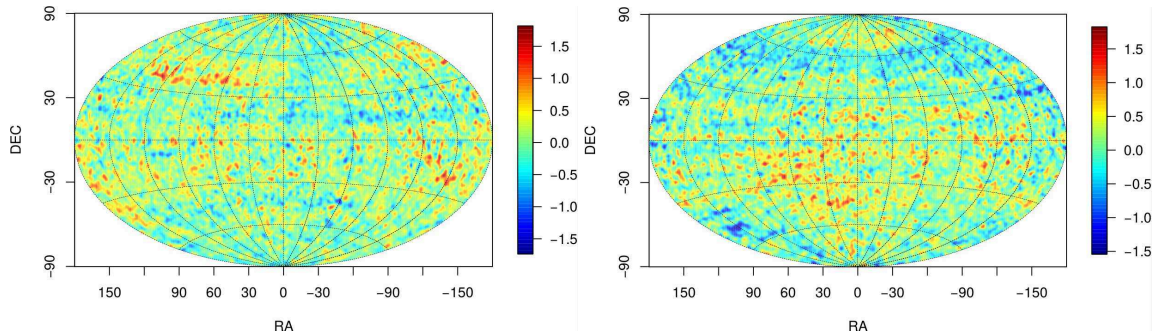
For the first preliminary TGAS 00.00 and TGAS 00.01 solutions, the proper motion was not consistent with the VLBI ones. For the same reason as was seen for the parallaxes, this was due to an under-estimation of the formal proper motion uncertainties. In TGAS 01.00 the uncertainties estimation was improved removing the significant bias on  $\mu_\alpha$ , and reducing that of  $\mu_\delta$ . In GDR1-AF/GDR1-BF, the discrepancy with VLBI



**Table 4.2** – Summary of the comparison between the GDR1-AF (TGAS 04.00) parallaxes (zero-point and extra-dispersion) and the external catalogues

Catalogue	GDR1 (TGAS 04.00)	
	$\varpi$ Zero-Point (mas)	$\varpi$ Extra-dispersion (mas)
RR Lyrae	$-0.07 \pm 0.02$	$0.16 \pm 0.02$
VLBI	$-0.083 \pm 0.12$	-
LMC	$0.11 \pm 0.02$	$-0.14 \pm 0.03$
SMC	$-0.12 \pm 0.05$	$-0.09 \pm 0.09$
QSO	$-0.046 \pm 0.01$	$-0.17 \pm 0.01$
Hipparcos	$-0.094 \pm 0.04$	$0.58 \pm 0.005$

stars looks as having been solved. For *Tycho-2* stars in common with *Hipparcos* no bias in the proper motion is seen neither. However a bias is seen for the *Tycho-2* sources not in *Hipparcos*. Those are most likely due to the *Tycho-2* catalogue as the residuals follow nicely the photographic plates used to derive those proper motions for that catalogue (fig. 4.3).

**Figure 4.3** – Sky distribution of the mean normalized residuals of the TGAS proper motions versus the *Tycho-2*, for the TGAS stars not in *Hipparcos*, in equatorial coordinates. Left:  $\mu_\alpha$ , right:  $\mu_\delta$ 

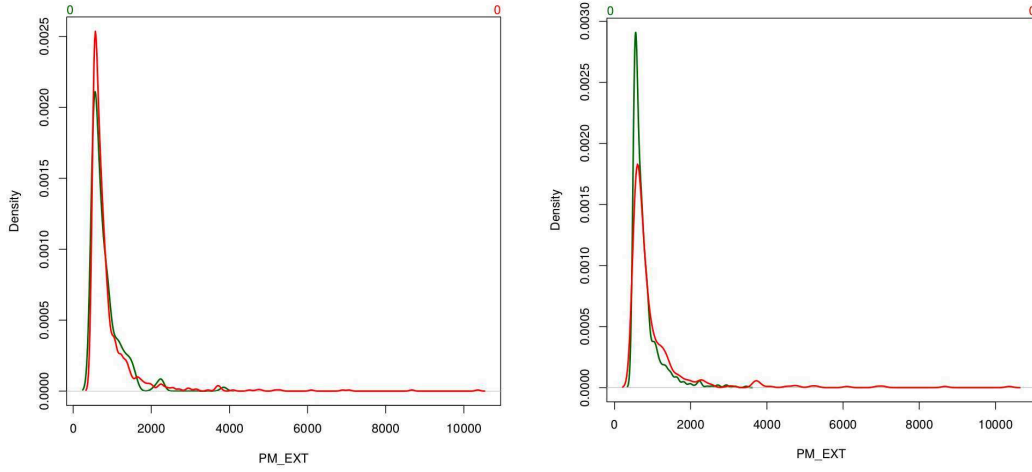
### 4.3.2 Missing High Proper Motion stars

Another observed issue is the huge number of High Proper Motion (HPM) stars missing in TGAS. Initially, in TGAS 00.00 and TGAS 00.01, the percentages of missing HPM stars were very significant, since 89% of *Tycho-2* HPM stars (total proper motion  $\mu > 1$  arcsec/yr) and 84% of *Simbad* HPM stars ( $\mu > 0.5$  arcsec/yr) did not appear in the TGAS solution. Moreover, six TGAS stars indicated as HPM with  $\mu > 1$  arcsec/yr are not HPM stars according to the *Simbad* database (which was logical since their associated *Tycho-2* identifier was 9999999, i.e. unknown).

After the reprocessing with TGAS 01.00 and later with the GDR1 solution, the number of missing stars has considerably decreased, although it is still important (fig. 4.4). Indeed, 40% of *Simbad* HPM stars are not found in TGAS. Two aspects could be identified: on one hand these stars seem to be fainter and redder than the found ones, and on the other hand, there are no HPM stars with  $\mu > 3.5$  arcsec/yr in the last TGAS solution. Although we could quantify and document this issue, it could not be solved for the GDR1-AF release, and may be due to a deeper problem in the initial source list and/or the cross-match. However, all detected TGAS HPM stars ( $\mu > 1$  mas/yr) are confirmed to be HPM in *Simbad*.

### 4.3.3 Gaia reference system rotation

During the first TGAS validation significant differences in proper motions (mainly in declination) were found for *Hipparcos* and *Tycho-2* stars. While looking for the origin of these discrepancies, we realised that we were not taking into account the fact that *Hipparcos* and Gaia have different reference frames. And



**Figure 4.4** – Normalized density distribution of *Simbad* HPM - TGAS 00.01 (left) and *Simbad* HPM - GDR1-AF (right) proper motions differences (divided by the formal error of the difference). "PM\_EXT" is the standard nomenclature adopted in WP944 to refer to the proper motion of the external catalogue (*Simbad* HPM in this case), in mas/yr. The green curve corresponds to the crossmatched stars *Simbad* versus Gaia database, and the red ones to the stars not found in the Gaia database

given their astrometric precision, the results may not be good if a solid rotation is not previously applied. Proper motions must be transformed from one system to another. In WP944 we transform the *Hipparcos* or *Tycho-2* values to the Gaia system.

The reference frame of GDR1 is aligned to the ICRF (*International Celestial Reference Frame*) by using the Gaia observations of the ICRF2 QSOs. Following [Lindegren et al. \[2016\]](#) the global solid rotation of *Hipparcos/Tycho-2* to the GDR1 system is:

$$w = (w_0, w_1, w_2) = (0.126, -0.185, -0.076) \text{mas/yr}$$

Given the coordinates  $(\alpha_G, \delta_G)$  of the Gaia catalogue, the proper motion components of the external catalogues in the Gaia system ( $\mu'_{\alpha_E}$  and  $\mu'_{\delta_E}$ ) are defined by:

$$\begin{aligned} \mu'_{\alpha_E} &= \mu_{\alpha_E} + q w \\ \mu'_{\delta_E} &= \mu_{\delta_E} + p w \end{aligned}$$

where  $q$  and  $p$  are two of the three unit vectors constituting the orthogonal vector triad  $[\vec{p}, \vec{q}, \vec{r}]$ :

$$\vec{r} = \begin{pmatrix} \cos \delta_G \cos \alpha_G \\ \cos \delta_G \sin \alpha_G \\ \sin \delta_G \end{pmatrix}, \quad \vec{p} = \begin{pmatrix} \sin \alpha_G \\ -\cos \alpha_G \\ 0 \end{pmatrix}, \quad \vec{q} = \begin{pmatrix} -\sin \delta_G \cos \alpha_G \\ -\sin \delta_G \sin \alpha_G \\ \cos \delta_G \end{pmatrix}$$

Therefore comparisons for both *Tycho-2* and *Hipparcos* proper motions with Gaia are done after subtracting this solid rotation from the proper motions to move them to the Gaia reference frame.

## 4.4 Sky Coverage

### 4.4.1 Missing sources

Analysing the sky homogeneity allows to check whether the distribution of stars on the sky is as expected, globally and locally per regions. Similarly we may test which fraction of stars we are recovering with each Gaia Data Release, as well as the type of objects. We have seen with HPM stars that in GDR1 we are loosing a large amount of them, especially those redder and fainter, and with proper motions larger than 3.5

arcsec/yr.

The completeness of the catalogue is reduced because of all those stars expected to be in TGAS (i.e. *Tycho-2* stars) and which are not. This may be due to problems in the astrometric solution (because of saturated sources, multiple objects, etc.) or to other particular problems such as small separation double stars missing or too small number of observations [Arenou et al., 2016b]. Figure 4.5 shows the distribution of missing *Tycho-2* stars in the last TGAS release (TGAS 04.00), which represents about 19% of the stars. Fig. 4.6 presents the same but for the whole catalogue, GDR1-BF, where the number is reduced to 3% of the stars. We see that the majority of missing stars are located near the galactic plane. The scanning law and crowding effects are clearly visible too [Arenou et al., 2016d].

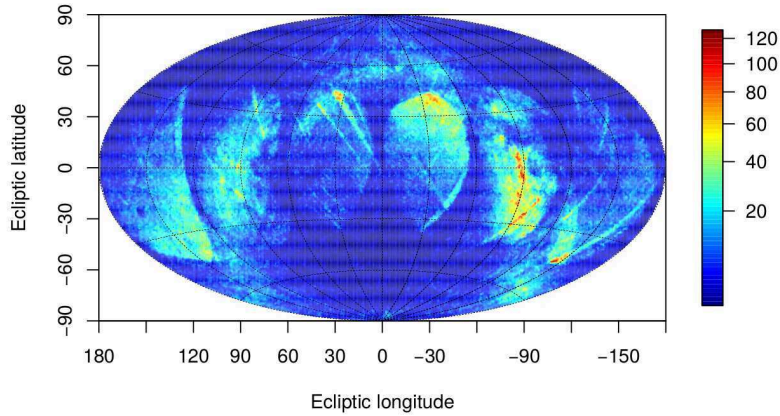


Figure 4.5 – Missing *Tycho-2* sources in TGAS, in ecliptic coordinates

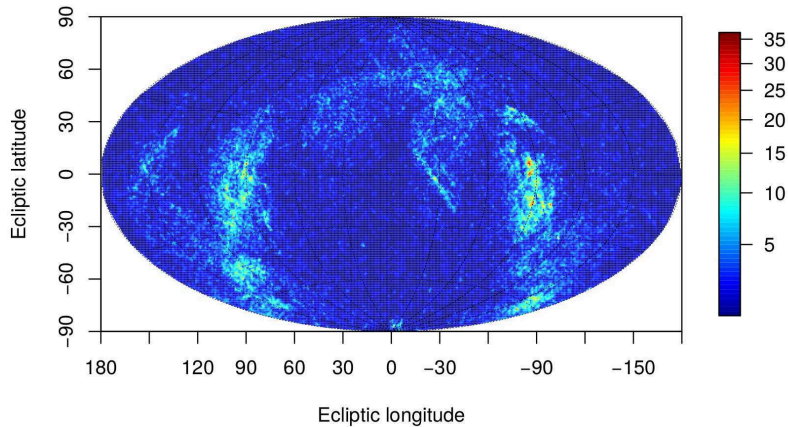
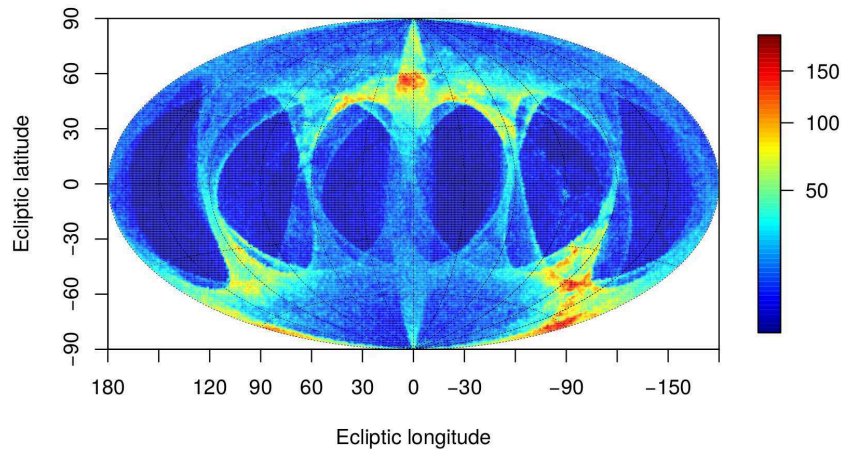


Figure 4.6 – Missing *Tycho-2* sources in GDR1, in ecliptic coordinates

#### 4.4.2 Duplicated sources

Wrong crossmatches not only lead to missing stars but also to duplicates. This has been seen from the first TGAS releases, and it is one of the main problems in GDR1 since the fraction of duplicated sources represents about the 3% of the whole Catalogue (for GRD1-BF), the majority of them located in the most scanned regions. The fraction increases to one third of the stars at  $G \sim 11$  [Arenou et al., 2016b]. These duplicates may come from spurious sources in the *IGSL* (Initial Gaia Source List), and/or from observational and processing problems when handling with close sources. Indeed, distinguishing between real and spurious pairs (or multiple) of sources is tricky when they fall in the same window: the parameters obtained may correspond to any or none of the objects.

For instance, in GDR1-BF, 32% of *Tycho-2* sources were duplicates, with a sky concentration clearly following the scanning law and a high over-density around ecliptic longitudes of  $180^\circ$  and  $0^\circ$  (fig. 4.7). For *2MASS* and *UCAC4* this represented 12% and 14%, respectively.



**Figure 4.7** – Duplicated *Tycho-2* sources in GDR1-BF

In order to clean and filter all candidate duplicate entries and to keep only one entry for catalogue publication, the following criterion has been adopted. For GDR1 it consists in keeping the source with the most precise parallax, then to the one with more observations, followed by the one with the best position or photometric uncertainties [Arenou et al., 2016d]. This process accounts for sources located within one pixel (59 mas) of each other or with positions consistent within  $5\sigma$ .

After this filtering process, more than 99% of duplicate sources were removed. However, for GDR1 the formal errors on positions may be underestimated, thus the adopted confidence interval of  $5\sigma$  could not be enough in some regions, especially in dense regions, and spurious sources still remain. Those duplicates are expected due to the crowding in the galactic plane.

It is important to remark that this issue will be solved in later data releases, and that for GDR1 all the stars with this problem will be flagged in order to be aware of using them carefully.

## 4.5 Outliers

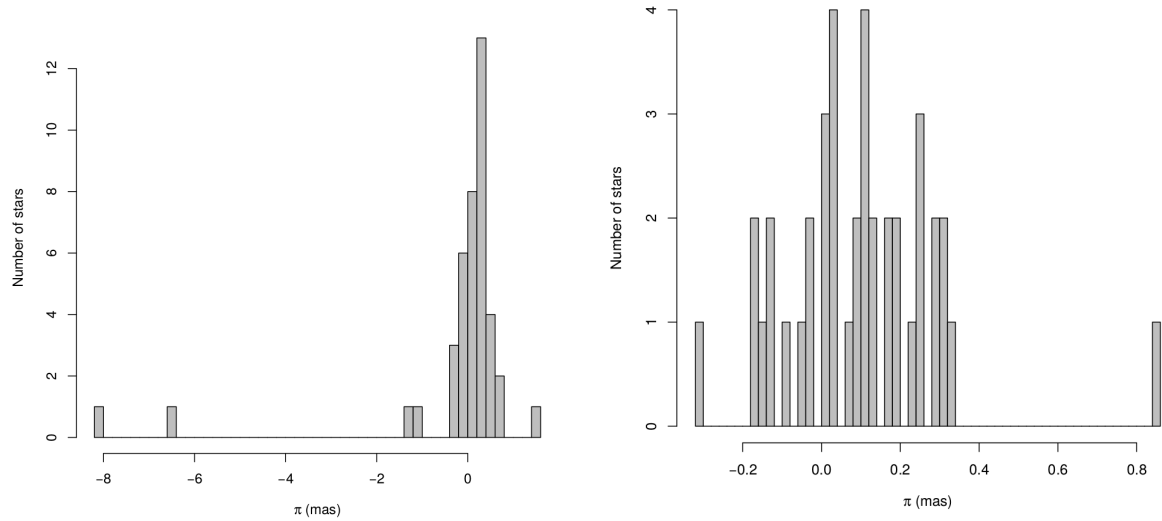
### 4.5.1 Astrometric

The number of astrometric outliers (defined in WP944 as stars with normalised difference larger than  $5\sigma$ , as detailed in chapter 3.1) has also been an important issue during the validation process. In TGAS 00.01, for instance, 23% of VLBI stars were considered as outliers. For RR Lyrae we obtained 6%. And for the LMC/SMC stars there were a few clearly identified outliers as may be seen in fig. 4.8 (left).

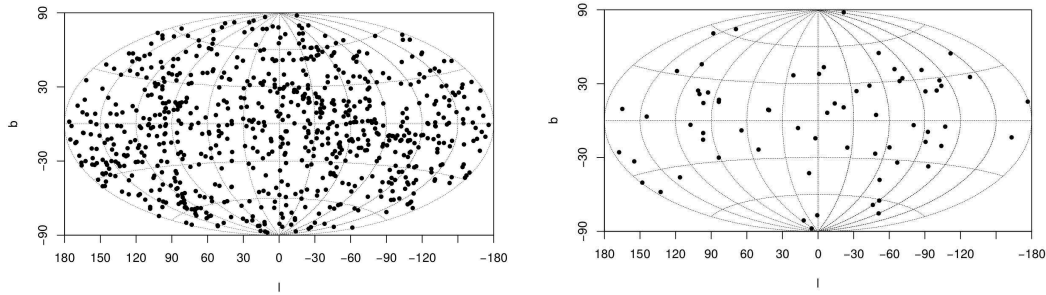
After new data reprocessing this amount strongly decreased, being almost 0% for all the samples. Figure 4.9 shows the outliers distribution in TGAS 00.01 and GDR1 for the *Hipparcos* parallaxes. The drop of the number of outliers has clearly improved the VTS results: most of them, such as the one with VLBI stars or *Hipparcos* stars, pass now correctly in GDR1 while it failed in TGAS 00.01.

### 4.5.2 Photometric

Photometric outliers have also been studied and also led to a specific filter applied by CU9. Indeed, a VTS has been also implemented in WP944 in order to check the Gaia photometry. One of the test cases aims to use the tight observed colour-colour relations (in the absence of extinction) to study departures from such a relation. Global splines are fitted on the observed colour-colour relations (i.e. G - EC reference magnitude



**Figure 4.8** – Parallax distribution of LMC/SMC stars confirmed to be in *Hipparcos*, for TGAS 00.01 (left) and GDR1 (right). The size of the sample is not the same because of the different filters applied between both Gaia versions. Note that the figures are not in the same scale



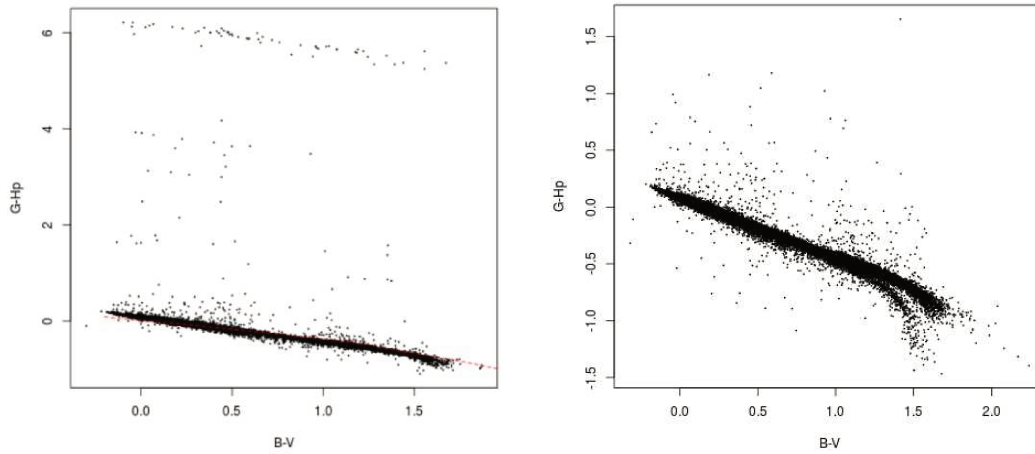
**Figure 4.9** – Parallax outliers of *Hipparcos* stars in TGAS 00.01 (left) and GDR1 (right)

versus EC reference colour) and residuals are studied. Moreover predictions of [Jordi et al. \[2010\]](#) and [Montaña et al. \[2016\]](#) are tested.

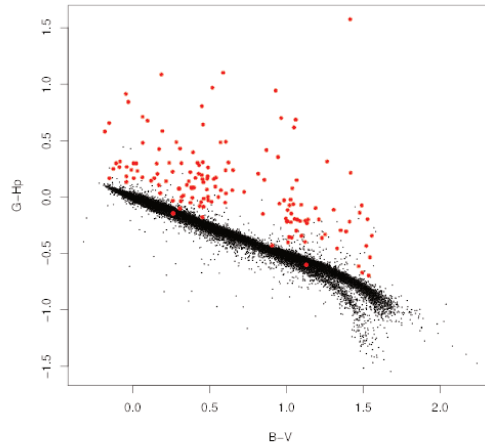
A large number of outliers were found in the  $G - H_p$  versus  $B - V$  diagram (fig. 4.10, left) of the *Hipparcos* stars (sample without binaries nor variable stars) in GDR1-BF. We may see on that figure that an important number of outliers are actually following the same colour-colour behaviour than most of the stars but shifted by 1 to 6 magnitudes. After deep analysis, the presence of these outliers seemed to agree with a gate information or to false detections because of saturated stars. This fact was also found and studied by the CU5 (photometric unit) members [[Evans, 2016](#), CU9 Validation Meeting]. All these outliers showed an issue on the number of observations: they had more RP (Red Photometer band) than G observations (less than 10), or directly no RP observations. Fig. 4.10 (right) displays the same CC plot after applying the CU9 filter, i.e. without the stars with less than 10 observations of G.

From this last diagram we see that there are still stars that do not fit the predicted CC relation. In fact, whereas the G magnitude uncertainty of the sample is about 0.0008 mag, for the remaining outliers (stars shifted around 0.5 magnitude from the main sample) it is 0.04 mag [[Arenou et al., 2016b](#)]. Fig. 4.11 shows the previous cleaned sample of fig. 4.10 (right) but highlighting in red those stars with  $\sigma_G > 0.02$  mag. We clearly see the impact on the colour-colour diagram of stars with less photometric accuracy. A filter on the photometric uncertainties is also necessary.

This test was not strictly implemented in the framework of this thesis but was developed with the purpose of using the empirical and robust CC calibrations provided in chapter 3. Indeed, the relationships predicted by [Jordi et al. \[2010\]](#) and [Montaña et al. \[2016\]](#) are based on the nominal Gaia passbands (calibrated before launch), but the *effective filters* actually slightly differ. This is due to the fact of averaging observations from



**Figure 4.10** – Colour-colour diagram  $G - H_p$  vs  $B - V$  of *Hipparcos* stars in GDR1 before filtering (left) and after filtering (right) stars with less than 10 observations of the  $G$  magnitude (see the text for details). Note that the figures are not in the same scale



**Figure 4.11** – Same as in figure 4.10 (right) but with those stars with  $\sigma_G > 0.02$  mag highlighted in red. The two curves belong to stars with  $\log g \leq 4.0$  (upper curve) and with  $\log g \geq 4.5$  (lower curve), both with  $A_V = 0$

both fields of view (FoVs), different CCDs, columns, gates and also due to contamination and aging effects. Therefore, we have special interest in using our CC empirical calibrations instead. Those calibrations will be updated soon to include the Gaia  $G$  magnitude, once the Gaia data will be available, and we will be able to use them within this VTS.

# Conclusions and Perspectives

## Contents

---

<b>4.6</b>	<b>Conclusions</b> . . . . .	<b>176</b>
4.6.1	The Red Clump photometric calibration . . . . .	176
4.6.2	TGAS and the Gaia Data Release 1 . . . . .	178
<b>4.7</b>	<b>Perspectives</b> . . . . .	<b>179</b>

---

## 4.6 Conclusions

This thesis was developed under an unprecedented and promising spatial mission to unravel the still many enigmas of our Galaxy: the Gaia satellite. Indeed, the high astrometric accuracy that Gaia will achieve will allow to derive precise distances. And distances are one of the major key to study the structure of the Milky Way. Especially, to go still further than the limit where Gaia will provide high accuracy trigonometric parallaxes, the use of *standard candles* as distance indicators is essential. The large number of *Red Clump* (RC) stars that Gaia will have observed in the end offers an important opportunity to better characterise them as *standard candles* and calibrate their absolute magnitude.

In this thesis I developed methods to calibrate the RC colour-colour (CC) relations which are key to be able to derive their astrophysical parameters. With the final goal of properly characterising the Gaia RC, I also took part in the statistical validation of the first Gaia data, mainly parallaxes. Ensuring the quality of these parameters is strongly necessary to guarantee the good calibration of RC stars and their use for scientific exploitation.

### 4.6.1 The Red Clump photometric calibration

RC stars are indeed well known to be good distance indicators, especially within the Galaxy because their number in the solar neighbourhood is higher than other *standard candles* such as the RR Lyrae or the Cepheid variable stars. However we have proved that they are still not well characterised since, on the one hand we encounter difficulties in reproducing or modelling these stars in a simulated HR diagram (work of C. Babusiaux and of the *Master 2* training period of M. Roelens) and, on the other hand, when we compare atmosphere models (isochrones) and high quality RC observations in a CC diagram they do not fit as expected regardless of the model used (work of this thesis). An accurate characterisation of these stars is therefore essential if we want to use them to determine distances with high precision. Especially with the upcoming Gaia data, the need of a new and complete photometric calibration, covering all wavelengths from the ultraviolet to the infrared, including the Gaia G band, is axiomatic. Since Gaia data will arrive just at the end of this thesis, I have meanwhile implemented all the methods with respect to the *Hipparcos* magnitude  $H_p$ . All the work presented here will be soon updated with new and larger samples selected from Gaia data, and including the Gaia G band.

To address the above goals we first considered accounting for theoretical models in order to provide a global photometric calibration of the RC linked to those models. Using isochrones is advantageous in the sense that they allow to directly derive physical properties of the stars, such as ages, metallicities or effective temperatures. Thus to easily describe a population.

In order to fit observations and models, and correct the *gap* we observe between them in the CC diagram, we first attempted to derive a shift to be added to each photometric magnitude. I developed a Maximum Likelihood Estimator method to estimate this shift while accounting for the parallax, the interstellar extinction, the metallicities and the photometric uncertainties. I made use of the isochrones of *Padova* as the reference theoretical model and of a sample of *Hipparcos* stars with high photometric quality, good parallax precision and low levels of interstellar extinction.

Many trials, improvements and tests were performed but, unfortunately, the results obtained always showed an incorrect convergence of the method to the expected solution. The problem was more complex than just providing a shift, since in some cases even the slope of the isochrone varies with respect to the observations. While we were trying to understand the possible reasons for these discrepancies, [Mann and von Braun \[2014\]](#) published an empirical approach to improve the accuracy of the literature photometric filter profiles and zero point. We then decided to test our calibration results by comparing them with photometric magnitudes from different observed and theoretical spectral libraries derived by using these new filters. With the same purpose, I also selected red giant and main sequence stars from an open cluster and checked separately their agreement with isochrones in the CC diagram. Both tests failed for red giant



stars, whereas they worked fine for main sequence stars. Hence, we assumed that we could not use the models to reproduce the observations, and to calibrate the RC.

These conclusions brought us to face the problem in a different way, and we decided to directly derive a purely empirical calibration. I also extended the work to effective temperature - colour ( $T_{\text{eff}}C$ ) relations. However, there are several challenges to be addressed that make difficult such a purpose if we look for accuracy and robustness. The first big challenge was getting a sample large enough that met the following strict criteria: high quality data in all photometric bands (*Johnsons* BVI, *Tycho*  $B_T V_T$ , *Hipparcos*  $H_p$  and *2MASS* JK), no binaries nor multiple sources, homogeneous atmospheric parameters, spectroscopic metallicity, and a low well-determined interstellar extinction.

Homogeneity of atmospheric parameters and well-determined extinction were indeed the most constraining conditions. Because each study uses its own methods and techniques, homogeneity is primordial in order to provide reliable results and to avoid systematic errors. Especially for the  $T_{\text{eff}}C$  calibration, having a good and large enough sample has been quite tricky. At the same time, we tried to make use of real and individual measurements of interstellar extinction derived from the diffuse interstellar band (DIB) of spectra. Although the correlation between the width of the DIBs and the amount of interstellar extinction of the star is still being studied, we may assume that if there is no apparent DIB the extinction is almost negligible. As we looked for stars with very low extinction ( $A_V < 0.03$ ), I took advantage of this assumption. For the remaining stars of our sample, for which DIB measurements were not available, I used the 3D and 2D maps of [Lallement et al. \[2014\]](#) and [Schlegel et al. \[1998\]](#), respectively. I ended up with 144 stars for the CC calibration. For the  $T_{\text{eff}}C$  relation I finally used four independent homogeneous samples between 16 and 45 stars, and one non-homogeneous of 116 stars.

A second challenge was to develop a method robust enough to provide a precise calibration. That meant a method able to account for the uncertainties of all variables and capable of selecting the best polynomial fit for each CC relations. I implemented a Monte Carlo Markov Chain (MCMC) method with a Deviance Information Criterion for model selection.

All calibrations were done with respect to  $H_p - K$  because the final goal is, once Gaia data be available, to provide the calibration with respect to the  $G - K$  colour index. The  $V - K$  colour is currently one of the most used temperature indicator as it is only weakly dependant on metallicity and the *2MASS* J, K photometry is available for all the sky. The *2MASS* photometry will also be available for a large portion of the bright Gaia stars, in particular those for which we will have good parallaxes in the DR1 and DR2. Other deep near infrared surveys (*UKIDSS*, *VISTA*) are currently underway. We therefore expect, due to its large availability, that G-K will soon be a largely used temperature indicator. Moreover, given the low number of metal-poor stars in all the samples, the results presented in this thesis only consider solar metallicity stars (i.e.  $-0.5 \leq [\text{Fe}/\text{H}] \leq 0.5$ ). We expect to increase this range of applicability with Gaia. In spite of all these challenges and limitations the results show that the calibrations obtained with this empirical method are of high accuracy, at a  $1\sigma$  level.

### Extinctions and temperatures of Red Clump stars

Additionally, once we had a robust colour and effective temperature calibration of the RC, we aimed at characterising this stellar population by determining their physical parameters. Since we cannot use the isochrones to derive them because of the discrepancies found between models and observations from the beginning of this work, I developed a new empirical photometrically-based Maximum Likelihood Estimator method which accounts for our CC and  $T_{\text{eff}}C$  relations. Given the observed magnitudes and the metallicity of a star, the method allows to provide a statistical estimation of the interstellar extinction and to automatically derive its effective temperature. The method was tested on a set of stars from the Galactic Bulge and *APOGEE* stars. We found a statistically good agreement with respect to their individually measured extinctions and temperatures. A correlation with the metallicity has been observed, which is being studied and will be hopefully corrected with the increased sample provided by TGAS.

We also applied this method on the *47 Tuc* globular cluster, on the *Kepler* and *CoRoT* fields, and on the *Hipparcos* RC. However, the current limitations in metallicity of our calibrations and/or the few available photometric quality data in some cases, significantly constraint the sample selection. The initial goal of this thesis was to implement everything directly on the Gaia data (including the empirical calibrations), but this has not been possible in the end. Indeed, the unavoidable overlap between the end of this thesis and the arrival of the first Gaia data, impeded us to provide results for these populations here. Nonetheless, everything is already implemented and will be applied soon, and the results for the Galactic Bulge and the *APOGEE* stars improved. Moreover, with Gaia, the magnitude G will be added in the method, thus improving the accuracy of the results.

#### 4.6.2 TGAS and the Gaia Data Release 1

Within the framework of the Coordination Unit 9 (CU9) of the Gaia Data Processing and Analysis Consortium (DPAC) I have actively participated in the statistical validation of the first Gaia catalogue release (DR1) and of a sub-catalogue (TGAS) containing the 5-parameter astrometric solutions for about two million stars in common with the *Tycho-2* Catalogue, using as first epoch *Hipparcos* or *Tycho-2* positions.

Within this thesis, I have started out and implemented all the necessary infrastructure of the WP944, the CU9 sub-unit in charge of the validation of the data with respect to external catalogues. At the time of starting this work, the platform to interact with the Gaia database (GACS) as well as other general statistical and validation tools were just prototypes, i.e. still in preliminary development. Therefore an important work was done to build up the whole structure of WP944, to test the interactions with these tools, services and databases, to implement the crossmatch with small external catalogues, and to set up the basis for the development of the different Validation Test Specifications (VTS) aiming to check all the parameters provided by Gaia.

Concerning these VTS I have also established the first tests to validate the TGAS parallaxes and proper motions. To do so I implemented a  $\chi^2$  method to check the astrometric solution consistency, and developed a method to study the zero-point of parallaxes by using distant stars (such as RR Lyrae stars) and very distant stars (such as stars in the Large and Small Magellanic Clouds) for which the parallax may be considered as zero. Additionally I have contributed to the implementation of specific tools needed in WP944, such as the basis to generate HEALPix maps and the method to transform the external catalogues' astrometry to the Gaia epoch. Moreover I have strongly collaborated with WP944 members responsible of the validation using quasars, and helped them to interact with the WP944 infrastructure we were constructing.

From September 2015 the CU9 has received several validation releases (of TGAS and recently of DR1). The statistical tests developed by us, together with all the VTS implemented by other CU9 members, has allowed us to continuously improve the data from one validation release to another up to the current final solution for DR1 (including TGAS). Also the tests have been updated and enhanced in order to provide more detailed outputs to better confirm possible issues or check the parameter's behaviours.

Our tests have allowed to highlight several aspects during this validation process. For instance, in the first releases we found a general underestimation of parallax and proper motion uncertainties that led to extra variances and significant bias. Moreover, whereas a global parallax zero-point of  $\sim -0.04$  has been found, we detected important local variations, probably due to the preliminary nature of the calibration, the attitude modelling errors, chromaticity or the basic-angle variations. This creates strong correlations between the parameters of the stars of the same sky region, thus the user will have to be careful and keep this in mind when selecting and using samples of stars (and depending on the scientific purposes). Missing and duplicated sources have also had an important impact on the filtering of the DR1 before publication; they are due to several problems, e.g. spurious IGSL sources, incorrect double detections, wrong crossmatches. And finally, a significant number of astrometric and photometric outliers have also been found. Most of them have been filtered out, especially the astrometric ones.

The interaction with the other WPs and CUs during the validation period has been of great importance

to understand the origin of all these issues, to solve them when possible and to define the filtering to be applied in CU9 for the final DR1, i.e. which sources are published and which are not. Since the goal is to provide the most accurate data and ensure the quality of the release, most of these problematic sources (that could not be solved for DR1 but they will for next releases) have been removed during the CU9 filtering. The details and the remaining features of the catalogue are described in the documentation delivered with the catalogue.

Besides these issues, the performances of the Gaia DR1 are very promising and will represent an important first step for the success of the mission.

## 4.7 Perspectives

The upcoming first Gaia data will allow us to apply my calibration method to the Gaia G band and, most important, to a larger and more homogeneous sample of RC stars. This will not only improve the accuracy of these calibrations but enlarge the range of applicability of colour and metallicity, thus to give them a wider use. For future analysis we might also consider extending the calibrations to other photometric systems, such as that of Strömgren or the *ugriz* used in the SDSS (*Sloan Digital Sky Survey*), as soon as we have enough high quality data.

Regarding the method to derive the interstellar extinction and effective temperatures, an immediate application once the TGAS data available, will be on the *47 Tuc* globular cluster and mainly on the *CoRoT* and *Kepler* fields with asteroseismic constraints (e.g. surface gravity). A collaboration with the *LESIA*<sup>2</sup> laboratory will also start in order to derive temperatures for all sources with seismic parameters. Those will be used to compute seismic luminosities that will be then compared to the Gaia absolute magnitudes. With our method we will also be able to provide new photometric temperatures and compare them to those derived spectroscopically from *APOGEE*, *LAMOST* and the *Gaia-ESO Survey*. Similarly the results provided for the Galactic Bulge and the *APOGEE* stars will be improved. And we will be able to provide the absolute magnitude for RC stars in the different sky regions, and derive new precise measurements of distances to large scale structures using the RC as *standard candle*. The shape of the derreddened *Hipparcos* RC and the one from TGAS will be also studied in next future.

Further, the calibration of the Gaia RC requires a prior estimate of the interstellar extinction. Indeed, any Gaia distance conversion into absolute luminosity needs to be corrected from extinction. Different 3D extinction maps are being constructed within our team through several methods, from the photometry to the diffuse interstellar bands (DIBs). The method developed in this thesis to derive interstellar extinctions may be extrapolated to other spectral types. We aim to use these photometric extinctions to allow the construction of these new 3D galactic extinction maps.

With the huge amount of data and the thousands of RC stars that are being observed by Gaia, we expect with no doubt to be able to disentangle the necessary ingredients to accurately position those stars in a theoretical HR diagram. This will help us to properly understand the RC stellar population content and, perhaps, tackle the *gap* found in the CC diagram between models (isochrones) and observations in a different way.

Finally, since the calibrations presented in this work allow to estimate distances of far red giant stars, we will be able to compare them to the parallaxes provided by Gaia (for which the relative precision diminishes with the distance). This will be therefore used to validate the data for the Gaia Data Release 2 (GDR2), and later releases.

The experience acquired before and during the Gaia DR1 validation process, as well as the robust tools developed in this thesis to calibrate and characterise the Gaia Red Clump, has allowed us to lay the groundwork for both the improvement of the statistical validation of the next releases and the scientific exploitation of the first data. The next 14th September 2016, a new door will open. The arrival of the DR1

<sup>2</sup>Laboratoire d'Études Spatiales et d'Instrumentation en Astrophysique, *Observatoire de Paris*

is just the beginning of a huge and promising study of that filmy glow we see every night. The same filmy glow that Democritus suggested it was made of plenty of stars, that Galileo confirmed and that Herschel started to count: our home Galaxy, the Milky Way.

# Bibliography

- A. Alonso, S. Arribas, and C. Martinez-Roger. The empirical scale of temperatures of the low main sequence (f0v-k5v). *A&A*, 313:873–890, 1996. ISSN 0004-6361. URL <http://adsabs.harvard.edu/abs/1996A%26A...313..873A>.
- D. R. Alves. K-Band Calibration of the Red Clump Luminosity. *ApJ*, 539:732–741, August 2000. doi: 10.1086/309278.
- David R. Alves and Ata Sarajedini. The age-dependent luminosities of the red giant branch bump, asymptotic giant branch bump, and horizontal branch red clump. *ApJ*, 511:225–234, 1999. ISSN 0004-637X. doi: 10.1086/306655. URL <http://adsabs.harvard.edu/abs/1999ApJ...511..225A>.
- S. Alves, L. Benamati, N. C. Santos, V. Zh. Adibekyan, S. G. Sousa, G. Israelian, J. R. De Medeiros, C. Lovis, and S. Udry. Determination of the spectroscopic stellar parameters for 257 field giant stars\*. *MNRAS*, 448:2749–2765, 2015. ISSN 0035-8711. doi: 10.1093/mnras/stv189. URL <http://adsabs.harvard.edu/abs/2015MNRAS.448.2749A>.
- E. Anderson and C. Francis. XHIP: An extended hipparcos compilation. *Astronomy Letters*, 38:331–346, May 2012a. doi: 10.1134/S1063773712050015.
- E. Anderson and C. Francis. VizieR Online Data Catalog: Extended Hipparcos Compilation (XHIP) (Anderson+, 2012). *VizieR Online Data Catalog*, 5137, March 2012b.
- F. Arenou and D. Michalik. Simulated TGAS data used for validation. Interface Control Document. *GAIA-C9-SP-OPM-FA-066*, May 2015.
- F. Arenou, M. Grenon, and A. Gomez. A tridimensional model of the galactic interstellar extinction. *A&A*, 258:104–111, 1992. ISSN 0004-6361. URL <http://adsabs.harvard.edu/abs/1992A%26A...258..104A>.
- F. Arenou, L. Lindegren, M. Froeschle, A. E. Gomez, C. Turon, M. A. C. Perryman, and R. Wielen. Zero-point and external errors of HIPPARCOS parallaxes. *A&A*, 304:52, 1995. ISSN 0004-6361. URL <http://cdsads.u-strasbg.fr/abs/1995A%26A...304...52A>.
- F. Arenou, F. Mignard, and J. Palasi. Verification of parallaxes. *ESA SP- 1200*, pages 433–446, December 1997.
- F. Arenou, C. Babusiaux, C. Fabricius, A. Helmi, A. Robin, A. Vallenari, E. Antiche, F. Figueras, M. Bredels, T. Cantat-Gaudin, K. Findeisen, G. Giuffrida, C. Reyl  , L. Ruiz-Dern, I.-C. Shih, and H. Ziaepour. CU9 Validation Tests Report: preliminary TGAS 00.01 data. *GAIA-C9-SP-OPM-FA-067*, November 2015.

- F. Arenou, C. Babusiaux, C. Fabricius, A. Helmi, A. Robin, A. Vallenari, E. Antiche, M. Breddels, T. Cantat-Gaudin, K. Findeisen, L. Ruiz-Dern, I.-C. Shih, and S. Soria. Validation Tests Report: preliminary TGAS 01.00 data. *GAIA-C9-SP-OPM-FA-070*, May 2016a.
- F. Arenou, C. Babusiaux, C. Fabricius, A. Helmi, A. Robin, A. Vallenari, C. Barache, M. Breddels, T. Cantat-Gaudin, S. Diakité, K. Findeisen, P. M. Marrese, C. Reylé, L. Ruiz-Dern, I.-C. Shih, S. Soria, R. Sordo, and C. Turon. Validation Tests Report: GDR1 (pre-filtering). *GAIA-C9-SP-OPM-FA-073*, June 2016b.
- F. Arenou, S. Babusiaux, C. Boudreault, P. Di Matteo, L. Eyer, C. Fabricius, K. Findeisen, E. Antiche, A. Helmi, F. Mignard, M. Manteiga-Outeiro, C. Reylé, A. Robin, P. Tanga, A. Vallenari, H. Ziaepour, and S. Soria. Validation Tests Specification (WP940). *GAIA-C9-SP-OPM-FA-061*, June 2016c.
- F. Arenou, X. Luri, A. Vallenari, T. Cantat-Gaudin, C. Fabricius, and others TBD. Gaia Data Release 1: Data validation: procedures, statistics and conclusions (in prep.). *A&A*, September 2016d.
- E. Athanassoula. On the nature of bulges in general and of box/peanut bulges in particular: input from N-body simulations. *MNRAS*, 358:1477–1488, April 2005. doi: 10.1111/j.1365-2966.2005.08872.x.
- C. Babusiaux and G. Gilmore. The structure of the Galactic bar. *MNRAS*, 358:1309–1319, April 2005. doi: 10.1111/j.1365-2966.2005.08828.x.
- C. Babusiaux, A. Gómez, V. Hill, F. Royer, M. Zoccali, F. Arenou, R. Fux, A. Lecureur, M. Schultheis, B. Barbuy, D. Minniti, and S. Ortolani. Insights on the Milky Way bulge formation from the correlations between kinematics and metallicity. *A&A*, 519:A77, September 2010. doi: 10.1051/0004-6361/201014353.
- J.-P. Beaulieu and P. D. Sackett. Red Clump Morphology as Evidence against a New Intervening Stellar Population as the Primary Source of Microlensing toward the Large Magellanic Cloud. *AJ*, 116:209–219, July 1998. doi: 10.1086/300400.
- T. R. Bedding, B. Mosser, D. Huber, J. Montalbán, P. Beck, J. Christensen-Dalsgaard, Y. P. Elsworth, R. A. García, A. Miglio, D. Stello, T. R. White, J. De Ridder, S. Hekker, C. Aerts, C. Barban, K. Belkacem, A.-M. Broomhall, T. M. Brown, D. L. Buzasi, F. Carrier, W. J. Chaplin, M. P. di Mauro, M.-A. Dupret, S. Frandsen, R. L. Gilliland, M.-J. Goupil, J. M. Jenkins, T. Kallinger, S. Kawaler, H. Kjeldsen, S. Mathur, A. Noels, V. Silva Aguirre, and P. Ventura. Gravity modes as a way to distinguish between hydrogen- and helium-burning red giant stars. *Nature*, 471:608–611, March 2011. doi: 10.1038/nature09935.
- M. Bellazzini, F. R. Ferraro, and E. Pancino. Multiple stellar populations in the Sextans dwarf spheroidal galaxy? *MNRAS*, 327:L15–L20, October 2001. doi: 10.1046/j.1365-8711.2001.04889.x.
- M. Bellazzini, F. R. Ferraro, L. Origlia, E. Pancino, L. Monaco, and E. Oliva. The Draco and Ursa Minor Dwarf Spheroidal Galaxies: A Comparative Study. *AJ*, 124:3222–3240, December 2002. doi: 10.1086/344794.
- M. Bellazzini, L. Origlia, L. Monaco, and E. Pancino. The RGB Bump in dwarf Spheroidal galaxies: discovery and perspectives. *Mem. Soc. Astron. Italiana*, 75:118, 2004.
- M. Bellazzini, N. Gennari, and F. R. Ferraro. The red giant branch tip and bump of the Leo II dwarf spheroidal galaxy. *MNRAS*, 360:185–193, June 2005. doi: 10.1111/j.1365-2966.2005.09027.x.
- M. Bellazzini, H. J. Newberg, M. Correnti, F. R. Ferraro, and L. Monaco. Detection of a population gradient in the Sagittarius stream. *A&A*, 457:L21–L24, October 2006. doi: 10.1051/0004-6361:20066002.

- T. Bensby, S. Feltzing, and M. S. Oey. Exploring the Milky Way stellar disk. A detailed elemental abundance study of 714 F and G dwarf stars in the solar neighbourhood. *A&A*, 562:A71, February 2014. doi: 10.1051/0004-6361/201322631.
- D. Bersier. The Distance to the Fornax Dwarf Galaxy Using Red Clump Stars and the Discrepancy between Red Clump and Tip of the Red Giant Branch Distances. *ApJ*, 543:L23–L26, November 2000. doi: 10.1086/318170.
- M. S. Bessell, F. Castelli, and B. Plez. Model atmospheres broad-band colors, bolometric corrections and temperature calibrations for O - M stars. *A&A*, 333:231–250, May 1998.
- E. Bica, D. Geisler, H. Dottori, J. J. Clariá, A. E. Piatti, and J. F. C. Santos, Jr. Ages and Metallicities of Star Clusters and Surrounding Fields in the Outer Disk of the Large Magellanic Cloud. *AJ*, 116:723–737, August 1998. doi: 10.1086/300448.
- O. Bienaymé, A. Siebert, and C. Soubiran. Observational constraints to the mass density in the Galactic plane. In U. Munari, editor, *GAIA Spectroscopy: Science and Technology*, volume 298 of *Astronomical Society of the Pacific Conference Series*, page 147, 2003.
- O. Bienaymé, C. Soubiran, T. Mishenina, V. Kovtyukh, and A. Siebert. Chemistry and Kinematics in the Solar Neighbourhood. In C. Turon, K. S. O’Flaherty, and M. A. C. Perryman, editors, *The Three-Dimensional Universe with Gaia*, volume 576 of *ESA Special Publication*, page 149, January 2005.
- S. Bilir, T. Ak, S. Ak, T. Yontan, and Z. F. Bostancı. A new absolute magnitude calibration for red clump stars. *New A*, 23:88–97, October 2013a. doi: 10.1016/j.newast.2013.03.006.
- S. Bilir, Ö. Önal, S. Karaali, A. Cabrera-Lavers, and H. Çakmak. Luminosity-colour relations for red clump stars. *Ap&SS*, 344:417–427, April 2013b. doi: 10.1007/s10509-012-1342-9.
- J. Binney, O. Gerhard, and D. Spergel. The photometric structure of the inner Galaxy. *MNRAS*, 288: 365–374, June 1997. doi: 10.1093/mnras/288.2.365.
- Stephan R. Bjork and Brian Chaboyer. Theoretical uncertainties in red giant branch evolution: The red giant branch bump. *AJ*, 641:1102–1112, 2006. ISSN 0004-637X. doi: 10.1086/500505. URL <http://adsabs.harvard.edu/abs/2006ApJ...641.1102B>.
- D. E. Blackwell and M. J. Shallis. Stellar angular diameters from infrared photometry - Application to Arcturus and other stars; with effective temperatures. *MNRAS*, 180:177–191, July 1977. doi: 10.1093/mnras/180.2.177.
- G. Bono, V. Castellani, S. degl’Innocenti, and L. Pulone. Advanced evolutionary phases of low-mass stars: The role of the original helium. *A&A*, 297:115–126, May 1995.
- G. Bono, S. Cassisi, M. Zoccali, and G. Piotto. Star counts across the red giant branch bump and below. *ApJ*, 546:L109–L113, 2001. ISSN 0004-637X. doi: 10.1086/318866. URL <http://adsabs.harvard.edu/abs/2001ApJ...546L.109B>.
- D. Bossini, A. Miglio, M. Salaris, A. Pietrinferni, J. Montalbán, A. Bressan, A. Noels, S. Cassisi, L. Girardi, and P. Marigo. Uncertainties on near-core mixing in red-clump stars: effects on the period spacing and on the luminosity of the AGB bump. *MNRAS*, 453:2290–2301, November 2015. doi: 10.1093/mnras/stv1738.
- Jo Bovy, David L. Nidever, Hans-Walter Rix, Léo Girardi, Gail Zasowski, S. Drew Chojnowski, Jon Holtzman, Courtney Epstein, Peter M. Frinchaboy, Michael R. Hayden, Thaíse S. Rodrigues, Steven R. Majewski, Jennifer A. Johnson, Marc H. Pinsonneault, Dennis Stello, Carlos Allende Prieto, Brett Andrews,

- Sarbani Basu, Timothy C. Beers, Dmitry Bizyaev, Adam Burton, William J. Chaplin, Katia Cunha, Yvonne Elsworth, Rafael A. García, Domingo A. García-Hernandez, Ana E. García Pérez, Fred R. Hearty, Saskia Hekker, Thomas Kallinger, Karen Kinemuchi, Lars Koesterke, Szabolcs Mészáros, Benoît Mosser, Robert W. O’Connell, Daniel Oravetz, Kaike Pan, Annie C. Robin, Ricardo P. Schiavon, Donald P. Schneider, Mathias Schultheis, Aldo Serenelli, Matthew Shetrone, Victor Silva Aguirre, Audrey Simmons, Michael Skrutskie, Verne V. Smith, Keivan Stassun, David H. Weinberg, John C. Wilson, and Olga Zamora. The APOGEE red-clump catalog: Precise distances, velocities, and high-resolution elemental abundances over a large area of the milky way’s disk. *AJ*, 790:127, 2014. ISSN 0004-637X. doi: 10.1088/0004-637X/790/2/127. URL <http://adsabs.harvard.edu/abs/2014ApJ...790..127B>.
- A. Cabrera-Lavers, C. González-Fernández, F. Garzón, P. L. Hammersley, and M. López-Corredoira. The long Galactic bar as seen by UKIDSS Galactic plane survey. *A&A*, 491:781–787, December 2008. doi: 10.1051/0004-6361:200810720.
- R. D. Cannon. Red giants in old open clusters. *MNRAS*, 150:111–135, 1970. doi: 10.1093/mnras/150.1.111.
- L. Casagrande, I. Ramírez, J. Meléndez, M. Bessell, and M. Asplund. An absolutely calibrated teff scale from the infrared flux method. dwarfs and subgiants. *A&A*, 512:A54, 2010. ISSN 0004-6361. doi: 10.1051/0004-6361/200913204. URL <http://adsabs.harvard.edu/abs/2010A%26A...512A..54C>.
- L. Casagrande, V. Silva Aguirre, D. Stello, D. Huber, A. M. Serenelli, S. Cassisi, A. Dotter, A. P. Milone, S. Hodgkin, A. F. Marino, M. N. Lund, A. Pietrinferni, M. Asplund, S. Feltzing, C. Flynn, F. Grundahl, P. E. Nissen, R. Schönrich, K. J. Schlesinger, and W. Wang. Strömgren survey for asteroseismology and galactic archaeology: Let the SAGA begin. *AJ*, 787:110, 2014. ISSN 0004-637X. doi: 10.1088/0004-637X/787/2/110. URL <http://adsabs.harvard.edu/abs/2014ApJ...787..110C>.
- Luca Casagrande, Laura Portinari, and Chris Flynn. Accurate fundamental parameters for lower main-sequence stars. *MNRAS*, 373:13–44, 2006. ISSN 0035-8711. doi: 10.1111/j.1365-2966.2006.10999.x. URL <http://adsabs.harvard.edu/abs/2006MNRAS.373...13C>.
- S. Cassisi, V. Castellani, S. Degl’Innocenti, G. Piotto, and M. Salaris. Asymptotic Giant Branch predictions: Theoretical uncertainties. *A&A*, 366:578–584, February 2001. doi: 10.1051/0004-6361:20000293.
- S. Cassisi, A. Marín-Franch, M. Salaris, A. Aparicio, M. Monelli, and A. Pietrinferni. The magnitude difference between the main sequence turn off and the red giant branch bump in galactic globular clusters. *A&A*, 527:A59, 2011. ISSN 0004-6361. doi: 10.1051/0004-6361/201016066. URL <http://adsabs.harvard.edu/abs/2011A%26A...527A..59C>.
- Santi Cassisi and Maurizio Salaris. A critical investigation on the discrepancy between the observational and the theoretical red giant luminosity function ‘bump’. *MNRAS*, 285:593–603, 1997. ISSN 0035-8711. URL <http://adsabs.harvard.edu/abs/1997MNRAS.285..593C>.
- Santi Cassisi, Scilla degl’Innocenti, and Maurizio Salaris. The effect of diffusion on the red giant luminosity function ‘bump’. *MNRAS*, 290:515–520, 1997. ISSN 0035-8711. URL <http://adsabs.harvard.edu/abs/1997MNRAS.290..515C>.
- V. Castellani, A. Chieffi, and L. Pulone. The evolution of He-burning stars - Horizontal and asymptotic branches in Galactic globulars. *ApJS*, 76:911–977, July 1991. doi: 10.1086/191584.
- M. Catelan. The RR Lyrae Period-Luminosity Relation. I. Theoretical Calibration. *ApJ*, 154:633–649, October 2004. doi: 10.1086/422916.
- G. Chabrier. The Galactic Disk Mass Budget. I. Stellar Mass Function and Density. *ApJ*, 554:1274–1281, June 2001. doi: 10.1086/321401.



- H.-C. Chen, R. Lallement, C. Babusiaux, L. Puspitarini, P. Bonifacio, and V. Hill. Extracting interstellar diffuse absorption bands from cool star spectra. application to bulge clump giants in baade's window. *A&A*, 550:A62, 2013. ISSN 0004-6361. doi: 10.1051/0004-6361/201220413. URL <http://adsabs.harvard.edu/abs/2013A%26A...550A..62C>.
- Y. Q. Chen, G. Zhao, and J. K. Zhao. The Absolute Magnitudes of Red Horizontal Branch Stars in the ugriz System. *ApJ*, 702:1336–1342, September 2009. doi: 10.1088/0004-637X/702/2/1336.
- Yang Chen, Léo Girardi, Alessandro Bressan, Paola Marigo, Mauro Barbieri, and Xu Kong. Improving PARSEC models for very low mass stars. *MNRAS*, 444:2525–2543, 2014. ISSN 0035-8711. doi: 10.1093/mnras/stu1605. URL <http://adsabs.harvard.edu/abs/2014MNRAS.444.2525C>.
- C. Chiosi, G. Bertelli, and A. Bressan. New developments in understanding the HR diagram. *ARA&A*, 30: 235–285, 1992. doi: 10.1146/annurev.aa.30.090192.001315.
- D.-H. Cho and S.-G. Lee. The Relation between Near-infrared Luminosity of Red Giant Branch Bumps and Metallicity of Galactic Globular Clusters. *AJ*, 124:977–988, August 2002. doi: 10.1086/341375.
- J. Christensen-Dalsgaard. On the red-giant luminosity bump. *MNRAS*, 453:666–670, October 2015. doi: 10.1093/mnras/stv1656.
- A. A. Cole. Age, Metallicity, and the Distance to the Magellanic Clouds From Red Clump Stars. *ApJ*, 500: L137–L140, June 1998. doi: 10.1086/311422.
- M. J. Cordero, C. A. Pilachowski, C. I. Johnson, I. McDonald, A. A. Zijlstra, and J. Simmerer. Detailed Abundances for a Large Sample of Giant Stars in the Globular Cluster 47 Tucanae (NGC 104). *ApJ*, 780: 94, January 2014. doi: 10.1088/0004-637X/780/1/94.
- M. Correnti, M. Bellazzini, R. A. Ibata, F. R. Ferraro, and A. Varghese. The Northern Wraps of the Sagittarius Stream as Traced by Red Clump Stars: Distances, Intrinsic Widths, and Stellar Densities. *ApJ*, 721:329–356, September 2010. doi: 10.1088/0004-637X/721/1/329.
- R. da Silva, A. d. C. Milone, and H. J. Rocha-Pinto. Homogeneous abundance analysis of FGK dwarf, subgiant, and giant stars with and without giant planets. *A&A*, 580:A24, August 2015. doi: 10.1051/0004-6361/201525770.
- A. Di Cecco, G. Bono, P. B. Stetson, A. Pietrinferni, R. Becucci, S. Cassisi, S. Degl'Innocenti, G. Iannicola, P. G. Prada Moroni, R. Buonanno, A. Calamida, F. Caputo, M. Castellani, C. E. Corsi, I. Ferraro, M. Dall'Ora, M. Monelli, M. Nonino, A. M. Piersimoni, L. Pulone, M. Romaniello, M. Salaris, A. R. Walker, and M. Zoccali. On the  $\delta v$  bump HB parameter in globular clusters. *ApJ*, 712:527–535, 2010. ISSN 0004-637X. doi: 10.1088/0004-637X/712/1/527. URL <http://adsabs.harvard.edu/abs/2010ApJ...712..527D>.
- A. E. Dolphin, A. Saha, E. D. Skillman, E. Tolstoy, A. A. Cole, R. C. Dohm-Palmer, J. S. Gallagher, M. Mateo, and J. G. Hoessel. Deep Hubble Space Telescope Imaging of IC 1613. I. Variable Stars and Distance. *ApJ*, 550:554–569, April 2001. doi: 10.1086/319809.
- J. Dommagnet and O. Nys. Catalogue of the Components of Double and Multiple stars (CCDM). First edition. *Communications de l'Observatoire Royal de Belgique*, 115, 1994.
- J. Dommagnet and O. Nys. VizieR Online Data Catalog: CCDM (Catalog of Components of Double Multiple stars) (Dommagnet+ 2002). *VizieR Online Data Catalog*, 1274, March 2002.
- Aaron Dotter, Brian Chaboyer, Darko Jevremovic, Veselin Kostov, E. Baron, and Jason W. Ferguson. The dartmouth stellar evolution database. *ApJS*, 178:89–101, 2008. ISSN 0067-0049. doi: 10.1086/589654. URL <http://adsabs.harvard.edu/abs/2008ApJS...178...89D>.

- E. Dwek, R. G. Arendt, M. G. Hauser, T. Kelsall, C. M. Lisse, S. H. Moseley, R. F. Silverberg, T. J. Sodroski, and J. L. Weiland. Morphology, near-infrared luminosity, and mass of the Galactic bulge from COBE DIRBE observations. *ApJ*, 445:716–730, June 1995. doi: 10.1086/175734.
- D. W. Evans. Validation in CU5. *CU9 Validation Meeting 2016*, May 2016.
- D. J. Faulkner and R. D. Cannon. Models of Population I Clump Giants. *ApJ*, 180:435–446, March 1973. doi: 10.1086/151974.
- F. R. Ferraro. The AGB and RGB clumps as standard candles. *Mem. Soc. Astron. Italiana*, 63:491–502, 1992.
- F. R. Ferraro. Photometric properties of Stellar Populations in GGCs: a multi-wavelength approach. In T. Lejeune and J. Fernandes, editors, *Observed HR Diagrams and Stellar Evolution*, volume 274 of *Astronomical Society of the Pacific Conference Series*, page 268, 2002.
- F. R. Ferraro, M. Messineo, F. Fusi Pecci, M. A. de Palo, O. Straniero, A. Chieffi, and M. Limongi. The giant, horizontal, and asymptotic branches of galactic globular clusters. i. the catalog, photometric observables, and features. *AJ*, 118:1738–1758, 1999. ISSN 0004-6256. doi: 10.1086/301029. URL <http://adsabs.harvard.edu/abs/1999AJ...118.1738F>.
- F. R. Ferraro, E. Dalessandro, A. Mucciarelli, G. Beccari, R. M. Rich, L. Origlia, B. Lanzoni, R. T. Rood, E. Valenti, M. Bellazzini, S. M. Ransom, and G. Coccozza. The cluster Terzan 5 as a remnant of a primordial building block of the Galactic bulge. *Nature*, 462:483–486, November 2009. doi: 10.1038/nature08581.
- E. L. Fitzpatrick. Correcting for the Effects of Interstellar Extinction. *PASP*, 111:63–75, January 1999. doi: 10.1086/316293.
- E. L. Fitzpatrick and D. Massa. An Analysis of the Shapes of Interstellar Extinction Curves. V. The IR-through-UV Curve Morphology. *ApJ*, 663:320–341, July 2007. doi: 10.1086/518158.
- K. Fuhrmann. Nearby stars of the Galactic disc and halo - IV. *MNRAS*, 384:173–224, February 2008. doi: 10.1111/j.1365-2966.2007.12671.x.
- F. Fusi Pecci, F. R. Ferraro, D. A. Crocker, R. T. Rood, and R. Buonanno. The variation of the red giant luminosity function 'bump' with metallicity and the age of the globular clusters. *A&A*, 238:95–110, 1990. ISSN 0004-6361. URL <http://adsabs.harvard.edu/abs/1990A%26A...238...95F>.
- C. Gallart. Observational Discovery of the Asymptotic Giant Branch Bump in Densely Populated Color-Magnitude Diagrams of Galaxies and Star Clusters. *ApJ*, 495:L43–L46, March 1998. doi: 10.1086/311218.
- L. Girardi, J.-C. Mermilliod, and G. Carraro. On the peculiar red clump morphology in the open clusters NGC 752 and NGC 7789. *A&A*, 354:892–898, February 2000.
- L. Girardi, M. A. T. Groenewegen, E. Hatziminaoglou, and L. da Costa. Star counts in the Galaxy. Simulating from very deep to very shallow photometric surveys with the TRILEGAL code. *A&A*, 436:895–915, June 2005. doi: 10.1051/0004-6361:20042352.
- L. Girardi, S. Rubele, and L. Kerber. Discovery of two distinct red clumps in NGC 419: a rare snapshot of a cluster at the onset of degeneracy. *MNRAS*, 394:L74–L78, March 2009. doi: 10.1111/j.1745-3933.2008.00614.x.
- L. Girardi, S. Rubele, and L. Kerber. Star clusters with dual red clumps. In R. de Grijs and J. R. D. Lépine, editors, *Star Clusters: Basic Galactic Building Blocks Throughout Time and Space*, volume 266 of *IAU Symposium*, pages 320–325, January 2010. doi: 10.1017/S1743921309991207.

- Leo Girardi. Clump stars in the solar neighbourhood. *ArXiv Astrophysics e-prints*, 1999a. URL <http://arxiv.org/abs/astro-ph/9912309>.
- Leo Girardi. Fine structure of the red clump in local group galaxies. *ArXiv Astrophysics e-prints*, pages 294–297, 2000. doi: 10.1007/10720961\_42. URL <http://arxiv.org/abs/astro-ph/9907086>.
- Leo Girardi, Martin A. T. Groenewegen, Achim Weiss, and Maurizio Salaris. Fine structure of the red giant clump from HIPPARCOS data, and distance determinations based on its mean magnitude. *MNRAS*, 301: 149–160, 1998. ISSN 0035-8711. doi: 10.1046/j.1365-8711.1998.02011.x. URL <http://adsabs.harvard.edu/abs/1998MNRAS.301..149G>.
- Léo Girardi. A secondary clump of red giant stars: why and where. *MNRAS*, 308:818–832, 1999b. ISSN 0035-8711. doi: 10.1046/j.1365-8711.1999.02746.x. URL <http://adsabs.harvard.edu/abs/1999MNRAS.308..818G>.
- Léo Girardi. Theoretical expectations for clump red giants as distance indicators. *Highlights of Astronomy*, 12:689–693, 2002. ISSN 1-58381-086-2. URL <http://adsabs.harvard.edu/abs/2002HiA...12..689G>.
- Léo Girardi and Maurizio Salaris. Population effects on the red giant clump absolute magnitude, and distance determinations to nearby galaxies. *MNRAS*, 323:109–129, 2001. ISSN 0035-8711. doi: 10.1046/j.1365-8711.2001.04084.x. URL <http://adsabs.harvard.edu/abs/2001MNRAS.323..109G>.
- O. A. Gonzalez, M. Rejkuba, D. Minniti, M. Zoccali, E. Valenti, and R. K. Saito. The inner Galactic bar traced by the VVV survey. *A&A*, 534:L14, October 2011. doi: 10.1051/0004-6361/201117959.
- O. A. Gonzalez, M. Zoccali, V. P. Debattista, J. Alonso-García, E. Valenti, and D. Minniti. Reinforcing the link between the double red clump and the X-shaped bulge of the Milky Way. *A&A*, 583:L5, November 2015. doi: 10.1051/0004-6361/201527434.
- J. I. González Hernández and P. Bonifacio. A new implementation of the infrared flux method using the 2mass catalogue. *A&A*, 497(2):497–509, 2009. ISSN 0004-6361, 1432-0746. doi: 10.1051/0004-6361/200810904. URL <http://www.aanda.org/articles/aa/pdf/2009/14/aa10904-08.pdf>.
- K. M. Gorski, B. D. Wandelt, F. K. Hansen, E. Hivon, and A. J. Banday. The HEALPix Primer. *ArXiv Astrophysics e-prints*, May 1999.
- K. M. Górski, A. J. Banday, E. Hivon, and B. D. Wandelt. HEALPix — a Framework for High Resolution, Fast Analysis on the Sphere. In D. A. Bohlender, D. Durand, and T. H. Handley, editors, *Astronomical Data Analysis Software and Systems XI*, volume 281 of *Astronomical Society of the Pacific Conference Series*, page 107, 2002.
- K. M. Górski, E. Hivon, A. J. Banday, B. D. Wandelt, F. K. Hansen, M. Reinecke, and M. Bartelmann. HEALPix: A Framework for High-Resolution Discretization and Fast Analysis of Data Distributed on the Sphere. *ApJ*, 622:759–771, April 2005. doi: 10.1086/427976.
- D. F. Gray. Books-Received - the Observation and Analysis of Stellar Photospheres - ED.2. *Journal of the British Astronomical Association*, 102:230, August 1992.
- M. A. T. Groenewegen. The red clump absolute magnitude based on revised hipparcos parallaxes. *A&A*, 488:935–941, 2008. ISSN 0004-6361. doi: 10.1051/0004-6361:200810201. URL <http://adsabs.harvard.edu/abs/2008A%26A...488..935G>.
- T. E. Harrison, J. Bornak, B. E. McArthur, and G. F. Benedict. Hubble Space Telescope Fine Guidance Sensor Parallaxes for Four Classical Novae. *ApJ*, 767:7, April 2013. doi: 10.1088/0004-637X/767/1/7.

- D. Hatzidimitriou and M. R. S. Hawkins. Stellar populations and large-scale structure of the SMC. II - Geometry of the north-eastern and south-western outlying regions. *MNRAS*, 241:667–690, December 1989. doi: 10.1093/mnras/241.4.667.
- U. Heiter, F. Kupka, C. van't Veer-Menneret, C. Barban, W. W. Weiss, M.-J. Goupil, W. Schmidt, D. Katz, and R. Garrido. New grids of ATLAS9 atmospheres I: Influence of convection treatments on model structure and on observable quantities. *A&A*, 392:619–636, September 2002. doi: 10.1051/0004-6361/20020788.
- U. Heiter, P. Jofré, B. Gustafsson, A. J. Korn, C. Soubiran, and F. Thévenin. Gaia FGK benchmark stars: Effective temperatures and surface gravities. *A&A*, 582:A49, October 2015. doi: 10.1051/0004-6361/201526319.
- V. Hill, A. Lecureur, A. Gómez, M. Zoccali, M. Schultheis, C. Babusiaux, F. Royer, B. Barbuy, F. Arenou, D. Minniti, and S. Ortolani. The metallicity distribution of bulge clump giants in Baade's window. *A&A*, 534:A80, October 2011. doi: 10.1051/0004-6361/200913757.
- J. A. Holtzman, M. Shetrone, J. A. Johnson, C. Allende Prieto, F. Anders, B. Andrews, T. C. Beers, D. Bizyaev, M. R. Blanton, J. Bovy, R. Carrera, S. D. Chojnowski, K. Cunha, D. J. Eisenstein, D. Feuillet, P. M. Frinchaboy, J. Galbraith-Frew, A. E. García Pérez, D. A. García-Hernández, S. Hasselquist, M. R. Hayden, F. R. Hearty, I. Ivans, S. R. Majewski, S. Martell, S. Meszaros, D. Muna, D. Nidever, D. C. Nguyen, R. W. O'Connell, K. Pan, M. Pinsonneault, A. C. Robin, R. P. Schiavon, N. Shane, J. Sobek, V. V. Smith, N. Troup, D. H. Weinberg, J. C. Wilson, W. M. Wood-Vasey, O. Zamora, and G. Zasowski. Abundances, Stellar Parameters, and Spectra from the SDSS-III/APOGEE Survey. *AJ*, 150:148, November 2015. doi: 10.1088/0004-6256/150/5/148.
- Y. Huang, X.-W. Liu, H.-B. Yuan, M.-S. Xiang, B.-Q. Chen, and H.-W. Zhang. Empirical metallicity-dependent calibrations of effective temperature against colours for dwarfs and giants based on interferometric data. *MNRAS*, 454:2863–2889, December 2015. doi: 10.1093/mnras/stv1991.
- R. A. Ibata, G. F. Lewis, and J.-P. Beaulieu. Reexamination of the Possible Tidal Stream in Front of the Large Magellanic Cloud. *ApJ*, 509:L29–L32, December 1998. doi: 10.1086/311749.
- I. Iben. Age and Initial Helium Abundance of Stars in the Globular Cluster M15. *Nature*, 220:143–146, October 1968. doi: 10.1038/220143a0.
- I. Iben, Jr. The life and times of an intermediate mass star - In isolation/in a close binary. *QJRAS*, 26:1–39, March 1985.
- I. Iben, Jr. Single and binary star evolution. *ApJS*, 76:55–114, May 1991. doi: 10.1086/191565.
- I. Iben, Jr. and A. Renzini. Asymptotic giant branch evolution and beyond. *ARA&A*, 21:271–342, 1983. doi: 10.1146/annurev.aa.21.090183.001415.
- E. Jofré, R. Petrucci, C. Saffe, L. Saker, E. A. de la Villarmois, C. Chavero, M. Gómez, and P. J. D. Mauas. Stellar parameters and chemical abundances of 223 evolved stars with and without planets. *A&A*, 574:A50, February 2015. doi: 10.1051/0004-6361/201424474.
- H. L. Johnson and W. W. Morgan. Fundamental stellar photometry for standards of spectral type on the revised system of the Yerkes spectral atlas. *ApJ*, 117:313, May 1953. doi: 10.1086/145697.
- C. Jordi, M. Gebran, J. M. Carrasco, J. de Bruijne, H. Voss, C. Fabricius, J. Knude, A. Vallenari, R. Kohley, and A. Mora. Gaia broad band photometry. *A&A*, 523:A48, 2010. ISSN 0004-6361. doi: 10.1051/0004-6361/201015441. URL <http://adsabs.harvard.edu/abs/2010A%26A...523A...48J>.

- A. Just and H. Jahreiß. Towards a fully consistent Milky Way disc model - I. The local model based on kinematic and photometric data. *MNRAS*, 402:461–478, February 2010. doi: 10.1111/j.1365-2966.2009.15893.x.
- T. Kallinger, B. Mosser, S. Hekker, D. Huber, D. Stello, S. Mathur, S. Basu, T. R. Bedding, W. J. Chaplin, J. De Ridder, Y. P. Elsworth, S. Frandsen, R. A. García, M. Gruberbauer, J. M. Matthews, W. J. Borucki, H. Bruntt, J. Christensen-Dalsgaard, R. L. Gilliland, H. Kjeldsen, and D. G. Koch. Asteroseismology of red giants from the first four months of kepler data: Fundamental stellar parameters. *A&A*, 522:A1, 2010. ISSN 0004-6361. doi: 10.1051/0004-6361/201015263. URL <http://adsabs.harvard.edu/abs/2010A%26A...522A...1K>.
- S. Karaali, S. Bilir, and E. Yaz Gökçe. Absolute magnitude calibration for red clump stars. *Ap&SS*, 346: 89–104, July 2013. doi: 10.1007/s10509-013-1449-7.
- D. Katz, U. Munari, M. Cropper, T. Zwitter, F. Thévenin, M. David, Y. Viala, F. Crifo, A. Gomboc, F. Royer, F. Arenou, P. Marrese, R. Sordo, M. Wilkinson, A. Vallenari, C. Turon, A. Helmi, G. Bono, M. Perryman, A. Gómez, L. Tomasella, F. Boschi, D. Morin, M. Haywood, C. Soubiran, F. Castelli, A. Bijaoui, G. Bertelli, A. Prsa, S. Mignot, A. Sellier, M.-O. Baylac, Y. Lebreton, U. Jauregi, A. Siviero, R. Bingham, F. Chemla, J. Coker, T. Dibbens, B. Hancock, A. Holland, D. Horville, J.-M. Huet, P. Laporte, T. Melse, F. Sayède, T.-J. Stevenson, P. Vola, D. Walton, and B. Winter. Spectroscopic survey of the Galaxy with Gaia- I. Design and performance of the Radial Velocity Spectrometer. *MNRAS*, 354:1223–1238, November 2004. doi: 10.1111/j.1365-2966.2004.08282.x.
- P. C. Keenan and C. Barnbaum. Revision and Calibration of MK Luminosity Classes for Cool Giants by HIPPARCOS Parallaxes. *ApJ*, 518:859–865, June 1999. doi: 10.1086/307311.
- M. Kim, E. Kim, M. G. Lee, A. Sarajedini, and D. Geisler. Determination of the Distance to M33 Based on the Tip of the Red Giant Branch and the Red Clump. *AJ*, 123:244–254, January 2002. doi: 10.1086/324639.
- C. R. King, G. S. Da Costa, and P. Demarque. The luminosity function on the subgiant branch of 47 tucanae a comparison of observation and theory. *ApJ*, 299:674–682, 1985. ISSN 0004-637X. doi: 10.1086/163733. URL <http://adsabs.harvard.edu/abs/1985ApJ...299..674K>.
- V. V. Kovtyukh, C. Soubiran, S. I. Belik, and N. I. Gorlova. High precision effective temperatures for 181 F-K dwarfs from line-depth ratios. *A&A*, 411:559–564, December 2003. doi: 10.1051/0004-6361:20031378.
- V. V. Kovtyukh, C. Soubiran, O. Bienaymé, T. V. Mishenina, and S. I. Belik. High-precision effective temperatures of 215 FGK giants from line-depth ratios. *MNRAS*, 371:879–884, September 2006. doi: 10.1111/j.1365-2966.2006.10719.x.
- R. Lallement, J.-L. Vergely, B. Valette, L. Puspitarini, L. Eyer, and L. Casagrande. 3D maps of the local ISM from inversion of individual color excess measurements. *A&A*, 561:A91, January 2014. doi: 10.1051/0004-6361/201322032.
- C. D. Laney, M. D. Joner, and G. Pietrzynski. VizieR Online Data Catalog: Nearby red clump stars JHK observations (Laney+, 2012). *VizieR Online Data Catalog*, 741, June 2012.
- P. F. Lazorenko, J. Sahlmann, D. Ségransan, E. L. Martín, M. Mayor, D. Queloz, and S. Udry. Astrometric planet search around southern ultracool dwarfs. II. Astrometric reduction methods and a deep astrometric catalogue. *A&A*, 565:A21, May 2014. doi: 10.1051/0004-6361/201323271.

- M. G. Lee, H. S. Park, J.-H. Park, Y.-J. Sohn, S. J. Oh, I.-S. Yuk, S.-C. Rey, S.-G. Lee, Y.-W. Lee, H.-I. Kim, W. Han, W.-K. Park, J. H. Lee, Y.-B. Jeon, and S. C. Kim. Deep Wide-Field BVI CCD Photometry of the Sextans Dwarf Spheroidal Galaxy. *AJ*, 126:2840–2866, December 2003. doi: 10.1086/379171.
- Y.-W. Lee, S.-J. Joo, and C. Chung. The Milky Way without X: an alternative interpretation of the double red clump in the Galactic bulge. *MNRAS*, 453:3906–3911, November 2015. doi: 10.1093/mnras/stv1980.
- L. Lindegren and J. H. J. de Bruijne. Performance of the Gaia mission. In P. K. Seidelmann and A. K. B. Monet, editors, *Astrometry in the Age of the Next Generation of Large Telescopes*, volume 338 of *Astronomical Society of the Pacific Conference Series*, page 25, October 2005.
- L. Lindegren, F. van Leeuwen, C. Petersen, M. A. C. Perryman, and S. Soderhjelm. Positions and parallaxes from the HIPPARCOS satellite - A first attempt at a global astrometric solution. *A&A*, 258:134–141, May 1992.
- L. Lindegren, U. Lammers, U. Bastian, J. Hernández, S. Klioner, D. Hobbs, A. Bombrun, D. Michalik, and et al. Gaia Data Release 1 Astrometry - one billion positions, two million proper motions and parallaxes. *LL-116-DRI-Astrometry*, 2016.
- C. Liu, M. Fang, Y. Wu, L.-C. Deng, L. Wang, W. Wang, J.-N. Fu, Y.-H. Hou, G.-W. Li, and Y. Zhang. Asteroseismic-based Estimation of the Surface Gravity for the LAMOST Giant Stars. *ApJ*, 807:4, July 2015. doi: 10.1088/0004-637X/807/1/4.
- M. López-Corredoira, A. Cabrera-Lavers, and O. E. Gerhard. A boxy bulge in the Milky Way. Inversion of the stellar statistics equation with 2MASS data. *A&A*, 439:107–110, August 2005. doi: 10.1051/0004-6361:20053075.
- R. E. Luck. Abundances in the Local Region. I. G and K Giants. *AJ*, 150:88, September 2015. doi: 10.1088/0004-6256/150/3/88.
- A. D. Mackey, P. Broby Nielsen, A. M. N. Ferguson, and J. C. Richardson. Multiple Stellar Populations in Three Rich Large Magellanic Cloud Star Clusters. *ApJ*, 681:L17, July 2008. doi: 10.1086/590343.
- G. Maintz. Proper identification of RR Lyrae stars brighter than 12.5 mag. *A&A*, 442:381–384, October 2005. doi: 10.1051/0004-6361:20053230.
- S. R. Majewski, M. H. Siegel, R. J. Patterson, and R. T. Rood. An Internal Second-Parameter Problem in the Sculptor Dwarf Spheroidal Galaxy. *ApJ*, 520:L33–L36, July 1999. doi: 10.1086/312133.
- S. R. Majewski, M. F. Skrutskie, M. D. Weinberg, and J. C. Ostheimer. A Two Micron All Sky Survey View of the Sagittarius Dwarf Galaxy. I. Morphology of the Sagittarius Core and Tidal Arms. *ApJ*, 599:1082–1115, December 2003. doi: 10.1086/379504.
- J. Maldonado and E. Villaver. Evolved stars and the origin of abundance trends in planet hosts. *A&A*, 588:A98, April 2016. doi: 10.1051/0004-6361/201527883.
- Andrew W. Mann and Kaspar von Braun. Revised filter profiles and zero points for broadband photometry. *ArXiv Astrophysics e-prints*, 2014. URL <http://arxiv.org/abs/1412.1474>.
- E. Masana, C. Jordi, and I. Ribas. Effective temperature scale and bolometric corrections from 2mass photometry. *A&A*, 450:735–746, 2006. ISSN 0004-6361. doi: 10.1051/0004-6361:20054021. URL <http://adsabs.harvard.edu/abs/2006A%26A...450..735M>.
- B. D. Mason, G. L. Wycoff, W. I. Hartkopf, G. G. Douglass, and C. E. Worley. The 2001 US Naval Observatory Double Star CD-ROM. I. The Washington Double Star Catalog. *AJ*, 122:3466–3471, December 2001. doi: 10.1086/323920.

- I. McDonald, A. A. Zijlstra, and M. L. Boyer. Fundamental parameters and infrared excesses of Hipparcos stars. *MNRAS*, 427:343–357, November 2012. doi: 10.1111/j.1365-2966.2012.21873.x.
- C. F. McKee and E. C. Ostriker. Theory of Star Formation. *ARA&A*, 45:565–687, September 2007. doi: 10.1146/annurev.astro.45.051806.110602.
- A. McWilliam, J. Fulbright, and R. M. Rich. Chemical Composition of the Galactic Bulge in Baade’s Window. In K. Cunha, M. Spite, and B. Barbuy, editors, *Chemical Abundances in the Universe: Connecting First Stars to Planets*, volume 265 of *IAU Symposium*, pages 279–284, March 2010. doi: 10.1017/S1743921310000748.
- Andrew McWilliam and Manuela Zoccali. Two red clumps and the x-shaped milky way bulge. *AJ*, 724: 1491–1502, 2010. ISSN 0004-637X. doi: 10.1088/0004-637X/724/2/1491. URL <http://adsabs.harvard.edu/abs/2010ApJ...724.1491M>.
- M. R. Meyer, F. C. Adams, L. A. Hillenbrand, J. M. Carpenter, and R. B. Larson. The Stellar Initial Mass Function: Constraints from Young Clusters, and Theoretical Perspectives. *Protostars and Planets IV*, page 121, May 2000.
- D. Michalik, L. Lindegren, and D. Hobbs. The Tycho-Gaia astrometric solution . How to get 2.5 million parallaxes with less than one year of Gaia data. *A&A*, 574:A115, February 2015. doi: 10.1051/0004-6361/201425310.
- F. Mignard. A few metrological aspects of the Gaia mission. *Advances in Space Research*, 47:356–364, January 2011. doi: 10.1016/j.asr.2010.05.020.
- A. P. Milone, L. R. Bedin, G. Piotto, and J. Anderson. Multiple stellar populations in Magellanic Cloud clusters. I. An ordinary feature for intermediate age globulars in the LMC? *A&A*, 497:755–771, April 2009. doi: 10.1051/0004-6361/200810870.
- L. Monaco, F. R. Ferraro, M. Bellazzini, and E. Pancino. First Detection of the Red Giant Branch Bump in the Sagittarius Dwarf Spheroidal Galaxy. *ApJ*, 578:L47–L50, October 2002. doi: 10.1086/344503.
- G. Montaña, J.M. Carrasco, H. Voss, and C. Jordi. Colour-colour relationships for photometric data validation. *GAIA-C5-TN-UB-GMF-001*, May 2016.
- P. Montegriffo, F. R. Ferraro, F. Fusi Pecci, and L. Origlia. IR-array photometry of Galactic globular clusters - II. JK photometry of 47 TUC. *MNRAS*, 276:739–752, October 1995. doi: 10.1093/mnras/276.3.739.
- T. Morel, A. Miglio, N. Lagarde, J. Montalbán, M. Rainer, E. Poretti, P. Eggenberger, S. Hekker, T. Kallinger, B. Mosser, M. Valentini, F. Carrier, M. Hareter, and L. Mantegazza. Atmospheric parameters and chemical properties of red giants in the CoRoT asteroseismology fields. *A&A*, 564:A119, 2014. ISSN 0004-6361. doi: 10.1051/0004-6361/201322810. URL <http://cdsads.u-strasbg.fr/abs/2014A%26A...564A.119M>.
- B. Mosser, C. Barban, J. Montalbán, P. G. Beck, A. Miglio, K. Belkacem, M. J. Goupil, S. Hekker, J. De Ridder, M. A. Dupret, Y. Elsworth, A. Noels, F. Baudin, E. Michel, R. Samadi, M. Auvergne, A. Baglin, and C. Catala. Mixed modes in red-giant stars observed with CoRoT. *A&A*, 532:A86, August 2011. doi: 10.1051/0004-6361/201116825.
- T. Muraveva, M. Palmer, G. Clementini, X. Luri, M.-R. L. Cioni, M. I. Moretti, M. Marconi, V. Ripepi, and S. Rubele. New Near-infrared Period-Luminosity-Metallicity Relations for RR Lyrae Stars and the Outlook for Gaia. *ApJ*, 807:127, July 2015. doi: 10.1088/0004-637X/807/2/127.
- D. M. Nataf, A. Udalski, A. Gould, P. Fouqué, and K. Z. Stanek. The Split Red Clump of the Galactic Bulge from OGLE-III. *ApJ*, 721:L28–L32, September 2010. doi: 10.1088/2041-8205/721/1/L28.

- D. M. Nataf, A. Gould, M. H. Pinsonneault, and P. B. Stetson. The gradients in the 47 Tuc red giant branch bump and horizontal branch are consistent with a centrally concentrated, helium-enriched second stellar generation. *AJ*, 736:94, 2011a. ISSN 0004-637X. doi: 10.1088/0004-637X/736/2/94. URL <http://adsabs.harvard.edu/abs/2011ApJ...736...94N>.
- D. M. Nataf, A. Udalski, A. Gould, and M. H. Pinsonneault. OGLE-III detection of the anomalous galactic bulge red giant branch bump: Evidence of enhanced helium enrichment. *AJ*, 730:118, 2011b. ISSN 0004-637X. doi: 10.1088/0004-637X/730/2/118. URL <http://adsabs.harvard.edu/abs/2011ApJ...730...118N>.
- David M. Nataf, Andrew Gould, Pascal Fouqué, Oscar A. Gonzalez, Jennifer A. Johnson, Jan Skowron, Andrzej Udalski, Michal K. Szymanski, Marcin Kubiak, Grzegorz Pietrzynski, Igor Soszynski, Krzysztof Ulaczyk, Lukasz Wyrzykowski, and Radoslaw Poleski. Reddening and extinction toward the galactic bulge from OGLE-III: The inner Milky Way's  $R_V \sim 2.5$  extinction curve. *AJ*, 769:88, 2013a. ISSN 0004-637X. doi: 10.1088/0004-637X/769/2/88. URL <http://adsabs.harvard.edu/abs/2013ApJ...769...88N>.
- David M. Nataf, Andrew P. Gould, Marc H. Pinsonneault, and Andrzej Udalski. Red giant branch bump brightness and number counts in 72 galactic globular clusters observed with the Hubble Space Telescope. *AJ*, 766:77, 2013b. ISSN 0004-637X. doi: 10.1088/0004-637X/766/2/77. URL <http://adsabs.harvard.edu/abs/2013ApJ...766...77N>.
- H. J. Newberg, B. Yanny, C. Rockosi, E. K. Grebel, H.-W. Rix, J. Brinkmann, I. Csabai, G. Hennessy, R. B. Hindsley, R. Ibata, Z. Ivezić, D. Lamb, E. T. Nash, M. Odenkirchen, H. A. Rave, D. P. Schneider, J. A. Smith, A. Stolte, and D. G. York. The Ghost of Sagittarius and Lumps in the Halo of the Milky Way. *ApJ*, 569:245–274, April 2002. doi: 10.1086/338983.
- C.-C. Ngeow. On the Application of Wesenheit Function in Deriving Distance to Galactic Cepheids. *ApJ*, 747:50, March 2012. doi: 10.1088/0004-637X/747/1/50.
- S. Nishiyama, T. Nagata, D. Baba, Y. Haba, R. Kadowaki, D. Kato, M. Kurita, C. Nagashima, T. Nagayama, Y. Murai, Y. Nakajima, M. Tamura, H. Nakaya, K. Sugitani, T. Naoi, N. Matsunaga, T. Tanabé, N. Kusakabe, and S. Sato. A Distinct Structure inside the Galactic Bar. *ApJ*, 621:L105–L108, March 2005. doi: 10.1086/429291.
- S. Nishiyama, T. Nagata, and IRSF/SIRIUS Team. Structure of the Galactic Bulge: Is the Milky Way a Double-barred Galaxy? *Journal of Physics Conference Series*, 54:62–66, December 2006. doi: 10.1088/1742-6596/54/1/011.
- Shogo Nishiyama, Motohide Tamura, Hirofumi Hatano, Daisuke Kato, Toshihiko Tanabé, Koji Sugitani, and Tetsuya Nagata. Interstellar extinction law toward the galactic center III: J, h, KS bands in the 2MASS and the MKO systems, and 3.6, 4.5, 5.8, 8.0  $\mu\text{m}$  in the Spitzer/IRAC system. *AJ*, 696:1407–1417, 2009. ISSN 0004-637X. doi: 10.1088/0004-637X/696/2/1407. URL <http://adsabs.harvard.edu/abs/2009ApJ...696.1407N>.
- P. E. Nissen, C. Akerman, M. Asplund, D. Fabbian, F. Kerber, H. U. Kaufl, and M. Pettini. Sulphur and zinc abundances in Galactic halo stars revisited. *A&A*, 469:319–330, July 2007. doi: 10.1051/0004-6361:20077344.
- J. E. Norris, M. S. Bessell, D. Yong, N. Christlieb, P. S. Barklem, M. Asplund, S. J. Murphy, T. C. Beers, A. Frebel, and S. G. Ryan. The Most Metal-poor Stars. I. Discovery, Data, and Atmospheric Parameters. *ApJ*, 762:25, January 2013. doi: 10.1088/0004-637X/762/1/25.



- W. O'Mullane, A. J. Banday, K. M. Górski, P. Kunszt, and A. S. Szalay. Splitting the Sky - HTM and HEALPix. In A. J. Banday, S. Zaroubi, and M. Bartelmann, editors, *Mining the Sky*, page 638, 2001. doi: 10.1007/10849171\_84.
- Bohdan Paczynski and Krzysztof Z. Stanek. Galactocentric distance with the optical gravitational lensing experiment and HIPPARCOS red clump stars. *ApJ*, 494:L219–L222, 1998. ISSN 0004-637X. doi: 10.1086/311181. URL <http://adsabs.harvard.edu/abs/1998ApJ...494L.219P>.
- B. Pagel. Book Review: EVOLUTION OF STARS AND STELLAR POPULATIONS / John Wiley & Sons, 2005. *The Observatory*, 126:306, August 2006.
- E. Pancino, G. Altavilla, S. Marinoni, G. Cocozza, J. M. Carrasco, M. Bellazzini, A. Bragaglia, L. Federici, E. Rossetti, C. Cacciari, L. Balaguer Nunez, A. Castro, F. Figueras, F. Fusi Pecci, S. Galleti, M. Gebran, C. Jordi, C. Lardo, E. Masana, M. Monguio, P. Montegriffo, S. Ragaini, W. Schuster, S. Trager, F. Villardell, and H. Voss. The gaia spectrophotometric standard stars survey. i. preliminary results. *ArXiv Astrophysics e-prints*, 2012. URL <http://arxiv.org/abs/1207.6042>.
- P. A. Patsis, C. Skokos, and E. Athanassoula. Orbital dynamics of three-dimensional bars - III. Boxy/peanut edge-on profiles. *MNRAS*, 337:578–596, December 2002. doi: 10.1046/j.1365-8711.2002.05943.x.
- M.A.C. Perryman, European Space Agency, and FAST Consortium. *The Hipparcos and Tycho Catalogues: The Hipparcos catalogue*. ESA SP. ESA Publications Division, 1997. URL <https://books.google.fr/books?id=MnzvAAAAMAAJ>.
- M. H. Pinsonneault, Y. Elsworth, C. Epstein, S. Hekker, S. Mészáros, W. J. Chaplin, J. A. Johnson, R. A. García, J. Holtzman, S. Mathur, A. García Pérez, V. Silva Aguirre, L. Girardi, S. Basu, M. Shetrone, D. Stello, C. Allende Prieto, D. An, P. Beck, T. C. Beers, D. Bizyaev, S. Bloemen, J. Bovy, K. Cunha, J. De Ridder, P. M. Frinchaboy, D. A. García-Hernández, R. Gilliland, P. Harding, F. R. Hearty, D. Huber, I. Ivans, T. Kallinger, S. R. Majewski, T. S. Metcalfe, A. Miglio, B. Mosser, D. Muna, D. L. Nidever, D. P. Schneider, A. Serenelli, V. V. Smith, J. Tayar, O. Zamora, and G. Zasowski. The APOKASC Catalog: An Asteroseismic and Spectroscopic Joint Survey of Targets in the Kepler Fields. *ApJS*, 215: 19, December 2014. doi: 10.1088/0067-0049/215/2/19.
- Martyn Plummer. Penalized loss functions for bayesian model comparison. *Biostatistics*, 9(3):523, 2008. doi: 10.1093/biostatistics/kxm049. URL [+http://dx.doi.org/10.1093/biostatistics/kxm049](http://dx.doi.org/10.1093/biostatistics/kxm049).
- D. Pourbaix, A. A. Tokovinin, A. H. Batten, F. C. Fekel, W. I. Hartkopf, H. Levato, N. I. Morrell, G. Torres, and S. Udry.  $S_{B<SUP>9</SUP>}$ : The ninth catalogue of spectroscopic binary orbits. *A&A*, 424:727–732, September 2004. doi: 10.1051/0004-6361:20041213.
- D. Pourbaix, A. A. Tokovinin, A. H. Batten, F. C. Fekel, W. I. Hartkopf, H. Levato, N. I. Morell, G. Torres, and S. Udry. VizieR Online Data Catalog: SB9: 9th Catalogue of Spectroscopic Binary Orbits (Pourbaix+ 2004-2014). *VizieR Online Data Catalog*, 1, August 2009.
- L. Pulone. AGB clump as standard candle. *Mem. Soc. Astron. Italiana*, 63:485–490, 1992.
- L. Puspitarini, R. Lallement, and H.-C. Chen. Automated measurements of diffuse interstellar bands in early-type star spectra. Correlations with the color excess. *A&A*, 555:A25, July 2013. doi: 10.1051/0004-6361/201321173.
- L. Puspitarini, R. Lallement, C. Babusiaux, H.-C. Chen, P. Bonifacio, L. Sbordone, E. Caffau, S. Duffau, V. Hill, A. Monreal-Ibero, F. Royer, F. Arenou, R. Peralta, J. E. Drew, R. Bonito, J. Lopez-Santiago, E. J. Alfaro, T. Bensby, A. Bragaglia, E. Flaccomio, A. C. Lanzafame, E. Pancino, A. Recio-Blanco, R. Smiljanic, M. T. Costado, C. Lardo, P. de Laverny, and T. Zwitter. The gaia-ESO survey: Extracting diffuse interstellar bands from cool star spectra. DIB-based interstellar medium line-of-sight structures

- at the kpc scale. *A&A*, 573:A35, 2015. ISSN 0004-6361. doi: 10.1051/0004-6361/201424391. URL <http://adsabs.harvard.edu/abs/2015A%26A...573A...35P>.
- E. Puzeras, G. Tautvaisiene, J. G. Cohen, D. F. Gray, S. J. Adelman, I. Ilyin, and Y. Chorniy. High-resolution spectroscopic study of red clump stars in the galaxy: iron-group elements. *MNRAS*, 408:1225–1232, 2010. ISSN 0035-8711. doi: 10.1111/j.1365-2966.2010.17195.x. URL <http://adsabs.harvard.edu/abs/2010MNRAS.408.1225P>.
- S. Raimond, R. Lallement, J. L. Vergely, C. Babusiaux, and L. Eyer. A southern hemisphere survey of the 5780 and 6284 Å diffuse interstellar bands: correlation with the extinction. *A&A*, 544:A136, August 2012. doi: 10.1051/0004-6361/201219191.
- I. Ramírez and J. Meléndez. The Effective Temperature Scale of FGK Stars. I. Determination of Temperatures and Angular Diameters with the Infrared Flux Method. *ApJ*, 626:446–464, June 2005. doi: 10.1086/430101.
- N. C. Rana and S. Basu. Derivation of the age-metallicity relation and initial mass function of stars in the solar neighbourhood. *A&A*, 265:499–503, November 1992.
- N. J. Rattenbury, S. Mao, V. P. Debattista, T. Sumi, O. Gerhard, and F. de Lorenzi. Proper motion dispersions of red clump giants in the galactic bulge: observations and model comparisons. *MNRAS*, 378:1165–1176, July 2007a. doi: 10.1111/j.1365-2966.2007.11851.x.
- N. J. Rattenbury, S. Mao, T. Sumi, and M. C. Smith. Modelling the Galactic bar using OGLE-II red clump giant stars. *MNRAS*, 378:1064–1078, July 2007b. doi: 10.1111/j.1365-2966.2007.11843.x.
- J.-J. Ren, X.-W. Liu, M.-S. Xiang, Y. Huang, S. Hekker, C. Wang, H.-B. Yuan, A. Rebassa-Mansergas, B.-Q. Chen, N.-C. Sun, H.-W. Zhang, Z.-Y. Huo, W. Zhang, Y. Zhang, Y.-H. Hou, and Y.-F. Wang. On the LSP3 estimates of surface gravity for LAMOST-Kepler stars with asteroseismic measurements. *Research in Astronomy and Astrophysics*, 16(3):009, March 2016. doi: 10.1088/1674-4527/16/3/045.
- M. Riello, S. Cassisi, G. Piotto, A. Recio-Blanco, F. De Angeli, M. Salaris, A. Pietrinferni, G. Bono, and M. Zoccali. The red giant branch luminosity function bump. *A&A*, 410:553–563, 2003. ISSN 0004-6361. doi: 10.1051/0004-6361:20031272. URL <http://adsabs.harvard.edu/abs/2003A%26A...410...553R>.
- L. Rizzi, E. V. Held, I. Saviane, R. B. Tully, and M. Gullieuszik. The distance to the Fornax dwarf spheroidal galaxy. *MNRAS*, 380:1255–1260, September 2007. doi: 10.1111/j.1365-2966.2007.12196.x.
- H. J. Rocha-Pinto, W. J. Maciel, J. Scalo, and C. Flynn. Chemical enrichment and star formation in the Milky Way disk. I. Sample description and chromospheric age-metallicity relation. *A&A*, 358:850–868, June 2000.
- L. Ruiz-Dern, C. Babusiaux, F. Arenou, and C. Turon. Characterisation of the Gaia Red Clump. In *EAS Publications Series*, volume 67 of *EAS Publications Series*, pages 395–396, July 2014. doi: 10.1051/eas/1567084.
- L. Ruiz-Dern, C. Babusiaux, C. Turon, G. Costigan, and F. Arenou. WP944 v18.1 Validation with external catalogues: Detailed Description and Validation Tests Report. *GAIA-C9-SP-OPM-LRD-001*, November 2015.
- L. Ruiz-Dern, C. Babusiaux, C. Barache, C. Turon, F. Arenou, G. Costigan, L. Galluccio, A. Krone-Martins, S. Lambert, and J. Souchay. WP944 v18.5 Validation with external catalogues: Detailed Description and Validation Tests Report (in prep.). *GAIA-C9-SP-OPM-LRD-002*, 2017.

- R. K. Saito, M. Zoccali, A. McWilliam, D. Minniti, O. A. Gonzalez, and V. Hill. Mapping the X-shaped Milky Way Bulge. *AJ*, 142:76, September 2011. doi: 10.1088/0004-6256/142/3/76.
- R. K. Saito, D. Minniti, B. Dias, M. Hempel, M. Rejkuba, J. Alonso-García, B. Barbuy, M. Catelan, J. P. Emerson, O. A. Gonzalez, P. W. Lucas, and M. Zoccali. Milky Way demographics with the VVV survey. I. The 84-million star colour-magnitude diagram of the Galactic bulge. *A&A*, 544:A147, August 2012. doi: 10.1051/0004-6361/201219448.
- M. Salaris. Evolutionary properties of stellar standard candles: Red clump, AGB clump and white dwarfs. In R. de Grijs, editor, *Advancing the Physics of Cosmic Distances*, volume 289 of *IAU Symposium*, pages 145–152, February 2013. doi: 10.1017/S1743921312021278.
- M. Salaris and L. Girardi. Population effects on the red giant clump absolute magnitude: the K band. *MNRAS*, 337:332–340, November 2002. doi: 10.1046/j.1365-8711.2002.05917.x.
- M. Salaris, S. Percival, E. Brocato, G. Raimondo, and A. R. Walker. The Distance to the Large Magellanic Cloud Cluster NGC 1866 and the Surrounding Field. *ApJ*, 588:801–804, May 2003. doi: 10.1086/374260.
- M. Salaris, E. V. Held, S. Ortolani, M. Gullieuszik, and Y. Momany. Deep near-infrared photometry of the globular cluster 47 Tucanae. Reconciling theory and observations. *A&A*, 476:243–253, December 2007. doi: 10.1051/0004-6361:20078445.
- N. C. Santos, G. Israelian, and M. Mayor. Spectroscopic [Fe/H] for 98 extra-solar planet-host stars. Exploring the probability of planet formation. *A&A*, 415:1153–1166, March 2004. doi: 10.1051/0004-6361:20034469.
- A. Sarajedini. WIYN Open Cluster Study. III. The Observed Variation of the Red Clump Luminosity and Color with Metallicity and Age. *AJ*, 118:2321–2326, November 1999. doi: 10.1086/301112.
- D. J. Schlegel, D. P. Finkbeiner, and M. Davis. Maps of Dust Infrared Emission for Use in Estimation of Reddening and Cosmic Microwave Background Radiation Foregrounds. *ApJ*, 500:525–553, June 1998. doi: 10.1086/305772.
- I.-C. Shih. Software User Manual for CU9 ValidationTools. *GAIA-C9-SP-OPM-IS-001-5*, March 2016a.
- I.-C. Shih. ValidationTools: Validation Procedure Specification. *GAIA-C9-SP-OPM-IS-004-2*, May 2016b.
- F. H. Shu, F. C. Adams, and S. Lizano. Star formation in molecular clouds - Observation and theory. *ARA&A*, 25:23–81, 1987. doi: 10.1146/annurev.aa.25.090187.000323.
- A. Siebert, O. Bienaymé, and C. Soubiran. Vertical distribution of Galactic disk stars. II. The surface mass density in the Galactic plane. *A&A*, 399:531–541, February 2003. doi: 10.1051/0004-6361:20021812.
- B. Smalley.  $T_{eff}$  and log g determinations. *Memorie della Societa Astronomica Italiana Supplementi*, 8: 130, 2005.
- R. L. Smart and L. Nicastro. The initial Gaia source list. *A&A*, 570:A87, October 2014. doi: 10.1051/0004-6361/201424241.
- N. A. Sokolov. The determination of  $T_{eff}$  of B, A and F main sequence stars from the continuum between 3200 Å and 3600 Å. *A&AS*, 110:553, May 1995.
- C. Soubiran, O. Bienaymé, and A. Siebert. Vertical distribution of Galactic disk stars. I. Kinematics and metallicity. *A&A*, 398:141–151, January 2003. doi: 10.1051/0004-6361:20021615.

- C. Soubiran, J.-F. Le Campion, G. Cayrel de Strobel, and A. Caillo. The PASTEL catalogue of stellar parameters. *A&A*, 515:A111, June 2010a. doi: 10.1051/0004-6361/201014247.
- C. Soubiran, J.-F. Le Campion, G. Cayrel de Strobel, and A. Caillo. VizieR Online Data Catalog: The PASTEL catalogue (Soubiran+, 2010-). *VizieR Online Data Catalog*, 1, January 2010b.
- K. Z. Stanek. Extinction Map of Baade's Window. *ApJ*, 460:L37, March 1996. doi: 10.1086/309976.
- K. Z. Stanek and P. M. Garnavich. Distance to m31 with the hubble space telescope and HIPPARCOS red clump stars. *ApJ*, 503:L131–L134, 1998. ISSN 0004-637X. doi: 10.1086/311539. URL <http://adsabs.harvard.edu/abs/1998ApJ...503L.131S>.
- K. Z. Stanek, M. Mateo, A. Udalski, M. Szymanski, J. Kaluzny, and M. Kubiak. Color-magnitude diagram distribution of the bulge red clump stars: Evidence for the galactic bar. *ApJ*, 429:L73–L76, July 1994. doi: 10.1086/187416.
- K. Z. Stanek, M. Mateo, A. Udalski, M. Szymański, J. Kaluzny, M. Kubiak, and W. Krzemiński. Evidence for the Galactic Bar from the Two Color Photometry of the Bulge Red Clump Stars. In L. Blitz and P. J. Teuben, editors, *Unsolved Problems of the Milky Way*, volume 169 of *IAU Symposium*, page 103, 1996.
- K. Z. Stanek, A. Udalski, M. Szymański, J. KaŁuŻny, Z. M. Kubiak, M. Mateo, and W. Krzemiński. Modeling the Galactic Bar Using Red Clump Giants. *ApJ*, 477:163–175, March 1997.
- D. Stello, D. Huber, T. R. Bedding, O. Benomar, L. Bildsten, Y. P. Elsworth, R. L. Gilliland, B. Mosser, B. Paxton, and T. R. White. Asteroseismic Classification of Stellar Populations among 13,000 Red Giants Observed by Kepler. *ApJ*, 765:L41, March 2013. doi: 10.1088/2041-8205/765/2/L41.
- T. Sumi. Extinction map of the Galactic centre: OGLE-II Galactic bulge fields. *MNRAS*, 349:193–204, March 2004. doi: 10.1111/j.1365-2966.2004.07482.x.
- A. V. Sweigart. Survey of some recent stellar-evolution calculations. In A. G. D. Philip and D. S. Hayes, editors, *The HR Diagram - The 100th Anniversary of Henry Norris Russell*, volume 80 of *IAU Symposium*, pages 333–343, 1978.
- A. V. Sweigart. Evolutionary sequences for horizontal branch stars. *ApJS*, 65:95–135, September 1987. doi: 10.1086/191219.
- Y. Takeda, B. Sato, E. Kambe, K. Sadakane, and M. Ohkubo. Spectroscopic Determination of Stellar Atmospheric Parameters: Application to Mid-F through Early-K Dwarfs and Subgiants. *PASJ*, 54:1041–1056, December 2002. doi: 10.1093/pasj/54.6.1041.
- Y. Takeda, B. Sato, and D. Murata. Stellar Parameters and Elemental Abundances of Late-G Giants. *PASJ*, 60:781–802, August 2008. doi: 10.1093/pasj/60.4.781.
- Jing Tang, Alessandro Bressan, Philip Rosenfield, Alessandra Slemmer, Paola Marigo, Léo Girardi, and Luciana Bianchi. New PARSEC evolutionary tracks of massive stars at low metallicity: testing canonical stellar evolution in nearby star-forming dwarf galaxies. *MNRAS*, 445:4287–4305, 2014. ISSN 0035-8711. doi: 10.1093/mnras/stu2029. URL <http://adsabs.harvard.edu/abs/2014MNRAS.445.4287T>.
- A. Tarantola and B. Valette. Generalized Nonlinear Inverse Problems Solved Using the Least Squares Criterion (Paper 1R1855). *Reviews of Geophysics and Space Physics*, 20:219, May 1982. doi: 10.1029/RG020i002p00219.
- H.-C. Thomas. Sternentwicklung VIII. der helium-flash bei einem stern von 1. 3 sonnenmassen. *ZAp*, 67:420, 1967. ISSN 0372-8331. URL <http://adsabs.harvard.edu/abs/1967ZA....67..420T>.

- J. R. Thorstensen. Parallaxes and Distance Estimates for 14 Cataclysmic Variable Stars. *AJ*, 126:3017–3029, December 2003. doi: 10.1086/379308.
- B. A. Twarog, B. J. Anthony-Twarog, and A. R. Bricker. Zeroing the Stellar Isochrone Scale: The Red Giant Clump Luminosity at Intermediate Metallicity. *AJ*, 117:1816–1826, April 1999. doi: 10.1086/300810.
- A. Udalski. The optical gravitational lensing experiment. the distance scale: Galactic bulge – LMC – SMC. *Acta Astron.*, 48:113–145, 1998. ISSN 0001-5237. URL <http://adsabs.harvard.edu/abs/1998AcA...48..113U>.
- A. Udalski. The optical gravitational lensing experiment: Red clump stars as a distance indicator. *ApJ*, 531:L25–L28, 2000. ISSN 0004-637X. doi: 10.1086/312513. URL <http://adsabs.harvard.edu/abs/2000ApJ...531L..25U>.
- A. Udalski. The Optical Gravitational Lensing Experiment: Is Interstellar Extinction toward the Galactic Center Anomalous? *ApJ*, 590:284–290, June 2003. doi: 10.1086/374861.
- M. Valentini and U. Munari. A spectroscopic survey of faint, high-galactic-latitude red clump stars. i. the high resolution sample. *A&A*, 522:A79, 2010. ISSN 0004-6361. doi: 10.1051/0004-6361/201014870. URL <http://adsabs.harvard.edu/abs/2010A%26A...522A..79V>.
- F. van Leeuwen. Rights and wrongs of the Hipparcos data. A critical quality assessment of the Hipparcos catalogue. *A&A*, 439:805–822, August 2005. doi: 10.1051/0004-6361:20053192.
- F. van Leeuwen, editor. *Hipparcos, the New Reduction of the Raw Data*, volume 350 of *Astrophysics and Space Science Library*, 2007. doi: 10.1007/978-1-4020-6342-8.
- K. von Braun, K. Chiboucas, J. K. Minske, J. F. Salgado, and G. Worthey. Color-Color Relations for Red Giants in Star Clusters. *PASP*, 110:810–820, July 1998. doi: 10.1086/316201.
- Junchen Wan, Chao Liu, Licai Deng, Wenyuan Cui, Yong Zhang, Yonghui Hou, Ming Yang, and Yue Wu. Red clump stars from the LAMOST data i: identification and distance. *ArXiv Astrophysics e-prints*, 2015. URL <http://arxiv.org/abs/1505.04878>.
- F. Windmark, L. Lindegren, and D. Hobbs. Using galactic cepheids to verify gaia parallaxes. *A&A*, 530:A76, 2011. ISSN 0004-6361. doi: 10.1051/0004-6361/201116929. URL <http://adsabs.harvard.edu/abs/2011A%26A...530A..76W>.
- D. Zaritsky and D. N. C. Lin. Evidence for an Intervening Stellar Population Toward the Large Magellanic Cloud. *AJ*, 114:2545, December 1997. doi: 10.1086/118666.
- G. Zhao, H. M. Qiu, and Shude Mao. High-resolution spectroscopic observations of hipparcos red clump giants: Metallicity and mass determinations. *ApJ*, 551(1):L85, 2001. ISSN 1538-4357. doi: 10.1086/319832. URL <http://iopscience.iop.org/1538-4357/551/1/L85>.
- P. Zieliński, A. Niedzielski, A. Wolszczan, M. Adamów, and G. Nowak. The Penn State-Toruń Centre for Astronomy Planet Search stars. I. Spectroscopic analysis of 348 red giants. *A&A*, 547:A91, November 2012a. doi: 10.1051/0004-6361/201117775.
- P. Zieliński, A. Niedzielski, A. Wolszczan, M. Adamów, and G. Nowak. VizieR online data catalog: Spectroscopic analysis of 348 red giants (zieliński+, 2012). *VizieR Online Data Catalog*, 354:79091, 2012b. URL <http://adsabs.harvard.edu/abs/2012yCat...35479091Z>.
- M. Zoccali. The Stellar Population of the Galactic Bulge. In K. Cunha, M. Spite, and B. Barbuy, editors, *Chemical Abundances in the Universe: Connecting First Stars to Planets*, volume 265 of *IAU Symposium*, pages 271–278, March 2010. doi: 10.1017/S1743921310000736.



## 3D Covariance Matrix

### Contents

---

<b>A.1</b>	<b>Notions of Normal Density Distributions</b>	<b>200</b>
<b>A.2</b>	<b>Gaia and External catalogues distributions</b>	<b>201</b>
A.2.1	Convolved distribution	202
A.2.2	Global covariance matrix	206
A.2.3	The $\chi^2$ variate	207

---

## A.1 Notions of Normal Density Distributions

### • Univariate Normal Distribution

In the case of a single random variable  $x$ , the probability distribution density or density function of that variable is represented by equation (A.1), called the Univariate Normal Distribution:

$$\phi(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{1}{2\sigma^2}(x-x_0)^2\right\} \quad -\infty < x < \infty \quad (\text{A.1})$$

where

$x_0$  : mean or expected value of the random variable  $x$

$\sigma^2$  : variance of the random variable  $x$

### • Bivariate Normal Distribution

The Bivariate Normal Distribution represents the joint distribution of two random variables  $x_1$  and  $x_2$  normally distributed with mean and variance:  $X_1 \approx \mathcal{N}(x_{10}, \sigma_1)$  and  $X_2 \approx \mathcal{N}(x_{20}, \sigma_2)$ , respectively. Both variables are related to each other in the sense that they are not independent of each other. This dependence is reflected by the correlation parameter  $\rho$ :

$$\rho = \text{corr}(x_1, x_2) = \frac{\text{COV}(x_1, x_2)}{\sigma_1 \sigma_2}$$

From the Univariate Distribution (equation A.1) we may notice that the term  $(x-x_0)^2/\sigma^2$  can be written as:

$$\left(\frac{x-x_0}{\sigma}\right)^2 = (x-x_0)(\sigma^2)^{-1}(x-x_0)$$

which generalized to more than one variable (a  $p \times 1$  vector  $\vec{x}$  of observations on several variables) becomes:

$$(\vec{x} - \vec{x}_0)^T \Sigma^{-1} (\vec{x} - \vec{x}_0) = \vec{X}^T \Sigma^{-1} \vec{X} \quad (\text{A.2})$$

where the  $p \times 1$  vector  $\vec{x}_0$  represents the expected value of the random vector  $\vec{X}$ , and the  $p \times p$  matrix  $\Sigma$  is the variance-covariance matrix (intuitively, the generalization of variance to multiple dimensions) of  $\vec{X}$ .

The covariance matrix is then the square matrix that contains the variances of the variables (diagonal elements) and the covariances between all possible pairs of variables (how much both change together; off-diagonal elements):

For two variables, it becomes:

$$\Sigma = \begin{pmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{pmatrix} = \begin{pmatrix} \sigma_1^2 & \rho \sigma_1 \sigma_2 \\ \rho \sigma_1 \sigma_2 & \sigma_2^2 \end{pmatrix}, \quad \sigma_{12} = \text{COV}(X_1, X_2)$$

And the inverse of  $\Sigma$ :

$$\Sigma^{-1} = \frac{1}{|\Sigma|} \Sigma^T = \frac{1}{\sigma_1^2 \sigma_2^2 (1-\rho^2)} \begin{pmatrix} \sigma_2^2 & -\rho \sigma_1 \sigma_2 \\ -\rho \sigma_1 \sigma_2 & \sigma_1^2 \end{pmatrix} = \frac{1}{1-\rho^2} \begin{pmatrix} 1/\sigma_1^2 & -\rho/\sigma_1 \sigma_2 \\ -\rho/\sigma_1 \sigma_2 & 1/\sigma_2^2 \end{pmatrix}$$



Thus, given two random variables and their covariance matrix  $\Sigma$ , we can now calculate the term in [A.2](#) for the bivariate case:

$$\begin{aligned}
(\vec{x} - \vec{x}_0)^T \Sigma^{-1} (\vec{x} - \vec{x}_0) &= \\
&= [x_1 - x_{10}, x_2 - x_{20}] \frac{1}{1 - \rho^2} \begin{pmatrix} 1/\sigma_1^2 & -\rho/\sigma_1\sigma_2 \\ -\rho/\sigma_1\sigma_2 & 1/\sigma_2^2 \end{pmatrix} \begin{pmatrix} x_1 - x_{10} \\ x_2 - x_{20} \end{pmatrix} \\
&= \frac{1}{1 - \rho^2} \left[ \frac{x_1 - x_{10}}{\sigma_1^2} - \frac{\rho(x_2 - x_{20})}{\sigma_1\sigma_2}, -\frac{\rho(x_2 - x_{20})}{\sigma_1\sigma_2} + \frac{x_2 - x_{20}}{\sigma_2^2} \right] \begin{pmatrix} x_1 - x_{10} \\ x_2 - x_{20} \end{pmatrix} \\
&= \frac{1}{1 - \rho^2} \left[ \frac{(x_1 - x_{10})^2}{\sigma_1^2} + \frac{(x_2 - x_{20})^2}{\sigma_2^2} - \frac{2\rho(x_1 - x_{10})(x_2 - x_{20})}{2\sigma_1\sigma_2} \right]
\end{aligned}$$

Therefore, the density function for the two variables jointly is (eq. [A.3](#)):

$$\phi(x_1, x_2) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \exp \left\{ -\frac{1}{2(1-\rho^2)} \left[ \frac{(x_1 - x_{10})^2}{\sigma_1^2} + \frac{(x_2 - x_{20})^2}{\sigma_2^2} - \frac{2\rho(x_1 - x_{10})(x_2 - x_{20})}{2\sigma_1\sigma_2} \right] \right\} \quad (\text{A.3})$$

Which written in matrix algebra becomes:

$$\phi(x_1, x_2) = \frac{1}{\sqrt{(2\pi)^2|\Sigma|}} \exp \left\{ -\frac{1}{2} (\vec{x} - \vec{x}_0)^T \Sigma^{-1} (\vec{x} - \vec{x}_0) \right\} \quad -\infty < x < \infty \quad (\text{A.4})$$

#### • Multivariate Normal Distribution

The generalization of equation [A.4](#) to multiple variables is called the Multivariate Normal Distribution. Just by inspection of the univariate and bivariate forms (equations [A.1](#) and [A.4](#)), we may easily deduce the multivariate distribution expression:

$$\phi(x_1, \dots, x_n) = \frac{1}{\sqrt{(2\pi)^n|\Sigma|}} \exp \left\{ -\frac{1}{2} (\vec{x} - \vec{x}_0)^T \Sigma^{-1} (\vec{x} - \vec{x}_0) \right\} \quad -\infty < x_i < \infty, i = 1, 2, \dots, n \quad (\text{A.5})$$

where  $n$  is the dimension of the distribution (i.e.  $n = 2$  for two variables)

The most important remark here is that expression [A.2](#) is actually the quadratic form of the random variables  $x_i$ , and is, therefore, a chi-square variate.

$$\vec{X}^T \Sigma^{-1} \vec{X} = \chi^2 \quad (\text{A.6})$$

## A.2 Gaia and External catalogues distributions

Each object considered has three correlated variables (without taking into account the position  $[\alpha, \delta]$ ): the two components of the proper motion ( $\mu_\alpha, \mu_\delta$ ) and the parallax ( $\varpi$ ), with their respective variances. We are indeed in the case of a 3-dimensional Multivariate Density Distribution.

However, what we want to validate is not whether each of these variables is coherent compared to an expected value:

$$\begin{aligned}
\mu_\alpha - \mu_{\alpha_0} &= 0 \\
\mu_\delta - \mu_{\delta_0} &= 0 \\
\varpi - \varpi_0 &= 0
\end{aligned}$$

, but whether they are coherent compared to another catalogue:

$$\begin{aligned}\mu_{\alpha_G} - \mu_{\alpha_E} &= 0 \\ \mu_{\delta_G} - \mu_{\delta_E} &= 0 \\ \overline{\omega}_G - \overline{\omega}_E &= 0\end{aligned}$$

with G: Gaia catalogue and E: External catalogue.

That means that each catalogue will have its own density distribution:  $\phi_G(\mu_{\alpha_G}, \mu_{\delta_G}, \overline{\omega}_G)$  for the Gaia catalogue, and  $\phi_E(\mu_{\alpha_E}, \mu_{\delta_E}, \overline{\omega}_E)$  for the External catalogue, but the expected value ( $x_0$ ) for each of the variables will be the same in both:

$$\begin{aligned}\phi_G(\mu_{\alpha_G}, \mu_{\delta_G}, \overline{\omega}_G) &= \frac{1}{\sqrt{(2\pi)^3 |\Sigma_G|}} \exp \left\{ -\frac{1}{2} (\vec{x}_G - \vec{x}_0)^T \Sigma_G^{-1} (\vec{x}_G - \vec{x}_0) \right\} \\ \vec{x}_G &= (\mu_{\alpha_G}, \mu_{\delta_G}, \overline{\omega}_G), \vec{x}_0 = (\mu_{\alpha_0}, \mu_{\delta_0}, \overline{\omega}_0) \\ \phi_E(\mu_{\alpha_E}, \mu_{\delta_E}, \overline{\omega}_E) &= \frac{1}{\sqrt{(2\pi)^3 |\Sigma_E|}} \exp \left\{ -\frac{1}{2} (\vec{x}_E - \vec{x}_0)^T \Sigma_E^{-1} (\vec{x}_E - \vec{x}_0) \right\} \\ \vec{x}_E &= (\mu_{\alpha_E}, \mu_{\delta_E}, \overline{\omega}_E), \vec{x}_0 = (\mu_{\alpha_0}, \mu_{\delta_0}, \overline{\omega}_0)\end{aligned}$$

To compute the difference between the two normal density distributions written above, we need to convolute them.

### A.2.1 Convolved distribution

Let  $X_C$  (C meaning "convolved") be a continuous random variable equal to the difference of the continuous random variables  $X_G$  and  $X_E$ :

$$X_C = X_G - X_E$$

To compute the density distribution of  $X_C$ ,  $\phi_C(x_C)$ , we need to integrate  $\phi_{GE}$  along the line where  $x_G - x_E = x_C$  or  $x_E = x_G - x_C$ :

$$\phi_C(x_C) = \int_{-\infty}^{\infty} \phi_{GE}(x_G, x_G - x_C) dx_G$$

Since  $X_G$  and  $X_E$  are independent random variables (parameter's measurements have been obtained in a completely different way and different instruments for every catalogue, so they are not related at all), it follows that:

$$\phi_{GE} = \phi_G(x_G) \phi_E(x_E)$$

therefore,

$$\phi_C(x_C) = \int_{-\infty}^{\infty} \phi_G(x_G) \phi_E(x_G - x_C) dx_G \quad (\text{A.7})$$

For simplicity, we do the integration in one dimension since the complete result will follow by induction on  $n$ . So, taking the equation of a univariate distribution, the normal densities for  $X_G$  and  $X_E$  are:

$$\phi_G(x_G) = \frac{1}{\sqrt{2\pi\sigma_G^2}} \exp \left\{ -\frac{1}{2\sigma_G^2} (x_G - x_{G_0})^2 \right\}$$

$$\phi_E(x_E) = \frac{1}{\sqrt{2\pi\sigma_E^2}} \exp\left\{-\frac{1}{2\sigma_E^2}(x_E - x_{E_0})^2\right\}$$

Notice that, to be mathematically generic, we consider here different expected values for each catalogue:  $x_{G_0}$  and  $x_{E_0}$ , instead of a unique  $x_0$ . This condition will be applied at the end.

Therefore, inserting both distributions into the integral:

$$\phi_C(x_C) = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma_G^2}} \exp\left\{-\frac{1}{2\sigma_G^2}(x_G - x_{G_0})^2\right\} \frac{1}{\sqrt{2\pi\sigma_E^2}} \exp\left\{-\frac{1}{2\sigma_E^2}(x_G - x_C - x_{E_0})^2\right\} dx_G$$

Multiplying the constants and the exponentials:

$$\phi_C(x_C) = \tag{A.8}$$

$$= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi(\sigma_G^2 + \sigma_E^2)}} \exp\left\{-\frac{1}{2}\left[\frac{(x_G - x_{G_0})^2}{\sigma_G^2} + \frac{(x_G - x_C - x_{E_0})^2}{\sigma_E^2}\right]\right\} dx_G \tag{A.9}$$

$$= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi(\sigma_G^2 + \sigma_E^2)}} \exp\left\{-\frac{1}{2}\left[\frac{\sigma_E^2(x_G - x_{G_0})^2 + \sigma_G^2(x_G - x_C - x_{E_0})^2}{\sigma_G^2\sigma_E^2}\right]\right\} dx_G \tag{A.10}$$

$$\tag{A.11}$$

Let's now take just the numerator ( $N$ ) of the term in the exponential and develop it individually:

$$N = \sigma_E^2(x_G - x_{G_0})^2 + \sigma_G^2(x_G - x_C - x_{E_0})^2$$

$N =$

$$\begin{aligned} &= \sigma_E^2 x_G^2 - 2\sigma_E^2 x_G x_{G_0} + \sigma_E^2 x_{G_0}^2 + \sigma_G^2 (x_G^2 - 2x_C x_G - 2x_G x_{E_0} - 2x_C x_{E_0} + x_C^2 + x_C x_{E_0} - x_G x_{E_0} + x_C x_{E_0} + x_{E_0}^2) \\ &= \sigma_E^2 x_G^2 - 2\sigma_E^2 x_G x_{G_0} + \sigma_E^2 x_{G_0}^2 + \sigma_G^2 + \sigma_G^2 x_G^2 - 2x_C x_G \sigma_G^2 - 2x_G x_{E_0} \sigma_G^2 + 2x_C x_{E_0} \sigma_G^2 + x_C^2 \sigma_G^2 + x_{E_0}^2 \sigma_G^2 \\ &= (\sigma_G^2 + \sigma_E^2) x_G^2 - 2\sigma_E^2 x_G x_{G_0} + \sigma_E^2 x_{G_0}^2 - 2x_C x_G \sigma_G^2 - 2x_G x_{E_0} \sigma_G^2 + 2x_C x_{E_0} \sigma_G^2 + x_C^2 \sigma_G^2 + x_{E_0}^2 \sigma_G^2 \end{aligned}$$

[Putting together all terms with  $x_G$ , with a common constant of  $(\sigma_G^2 + \sigma_E^2)$ ]

$$\begin{aligned} &= (\sigma_G^2 + \sigma_E^2) \left[ x_G^2 - \left( \frac{2\sigma_E^2 x_{G_0} + 2x_C \sigma_G^2 + 2x_{E_0} \sigma_G^2}{\sigma_G^2 + \sigma_E^2} \right) x_G \right] + \sigma_E^2 x_{G_0}^2 + 2x_C x_{E_0} \sigma_G^2 + x_C^2 \sigma_G^2 + x_{E_0}^2 \sigma_G^2 \\ &= (\sigma_G^2 + \sigma_E^2) \left[ x_G^2 - \left( \frac{2\sigma_E^2 x_{G_0} + 2(x_C + x_{E_0}) \sigma_G^2}{\sigma_G^2 + \sigma_E^2} \right) x_G \right] + \sigma_E^2 x_{G_0}^2 + 2x_C x_{E_0} \sigma_G^2 + x_C^2 \sigma_G^2 + x_{E_0}^2 \sigma_G^2 \end{aligned}$$

Taking a look to the first term (the ones dependent on  $x_G$ ) we can notice that it corresponds to the quadratic form of  $[x_G - \sigma_E^2 x_{G_0} - (x_C + x_{E_0}) \sigma_G^2]^2$  without the independent variable. Indeed:

$q =$

$$\begin{aligned} &= [x_G - \sigma_E^2 x_{G_0} - (x_C + x_{E_0}) \sigma_G^2]^2 \\ &= [x_G - \sigma_E^2 x_{G_0} - (x_C + x_{E_0}) \sigma_G^2] [(x_G - \sigma_E^2 x_{G_0} - (x_C + x_{E_0}) \sigma_G^2)] \\ &= x_G^2 - x_G \sigma_E^2 x_{G_0} - x_G (x_C + x_{E_0}) \sigma_G^2 - x_G \sigma_E^2 x_{G_0} + \sigma_E^4 x_{G_0}^2 + \sigma_E^2 x_{G_0} (x_C + x_{E_0}) \sigma_G^2 \\ &\quad - x_G (x_C + x_{E_0}) \sigma_G^2 + \sigma_E^2 x_{G_0} (x_C + x_{E_0}) \sigma_G^2 + (x_C + x_{E_0})^2 \sigma_G^4 \\ &= x_G^2 - 2x_G \sigma_E^2 x_{G_0} - 2x_G (x_C + x_{E_0}) \sigma_G^2 + 2\sigma_E^2 x_{G_0} (x_C + x_{E_0}) \sigma_G^2 + \sigma_E^4 x_{G_0}^2 + (x_C + x_{E_0})^2 \sigma_G^4 \end{aligned}$$

$$= x_G^2 - 2x_G\sigma_E^2x_{G_0} - 2x_G(x_C + x_{E_0})\sigma_G^2 + [\sigma_E^2x_{G_0} + (x_C + x_{E_0})\sigma_G^2]^2$$

We may see that the first three terms of this expression correspond exactly to what we have in the first term of  $N$ . The other terms belong to the independent variable. Thus, to rewrite the dependent term of  $x_G$  in  $N$  in a quadratic form we need to sum and deduct this independent term:

$$\begin{aligned} N &= \\ &= (\sigma_G^2 + \sigma_E^2) \left[ x_G^2 - \left( \frac{2\sigma_E^2x_{G_0} + 2(x_C + x_{E_0})\sigma_G^2}{\sigma_G^2 + \sigma_E^2} \right) x_G + \left( \frac{\sigma_E^2x_{G_0} + (x_C + x_{E_0})\sigma_G^2}{\sigma_G^2 + \sigma_E^2} \right)^2 \right] \\ &\quad + \sigma_E^2x_{G_0}^2 + 2x_Cx_{E_0}\sigma_G^2 + x_C^2\sigma_G^2 + x_{E_0}^2\sigma_G^2 - \frac{[\sigma_E^2x_{G_0} + (x_C + x_{E_0})\sigma_G^2]^2}{\sigma_G^2 + \sigma_E^2} \\ &= (\sigma_G^2 + \sigma_E^2) \left[ x_G - \frac{\sigma_E^2x_{G_0} + (x_C + x_{E_0})\sigma_G^2}{\sigma_G^2 + \sigma_E^2} \right]^2 \\ &\quad + \sigma_E^2x_{G_0}^2 + 2x_Cx_{E_0}\sigma_G^2 + x_C^2\sigma_G^2 + x_{E_0}^2\sigma_G^2 - \frac{[\sigma_E^2x_{G_0} + (x_C + x_{E_0})\sigma_G^2]^2}{\sigma_G^2 + \sigma_E^2} \\ &= (\sigma_G^2 + \sigma_E^2) \left[ x_G - \frac{\sigma_E^2x_{G_0} + (x_C + x_{E_0})\sigma_G^2}{\sigma_G^2 + \sigma_E^2} \right]^2 + I \end{aligned}$$

where  $I$  regroups all the independent terms.

If we develop  $I$ :

$$\begin{aligned} I &= \sigma_E^2x_{G_0}^2 + 2x_Cx_{E_0}\sigma_G^2 + x_C^2\sigma_G^2 + x_{E_0}^2\sigma_G^2 - \frac{[\sigma_E^2x_{G_0} + (x_C + x_{E_0})\sigma_G^2]^2}{\sigma_G^2 + \sigma_E^2} \\ I &= \frac{(\sigma_G^2 + \sigma_E^2) [\sigma_E^2x_{G_0}^2 + 2x_Cx_{E_0}\sigma_G^2 + x_C^2\sigma_G^2 + x_{E_0}^2\sigma_G^2] - [\sigma_E^2x_{G_0} + (x_C + x_{E_0})\sigma_G^2]^2}{\sigma_G^2 + \sigma_E^2} \end{aligned}$$

Where the numerator ( $I_n$ ):

$$\begin{aligned} I_n &= \\ &= x_{G_0}^2\sigma_E^4 + x_{G_0}^2\sigma_G^2\sigma_E^2 + x_C^2\sigma_G^2\sigma_E^2 + 2x_Cx_{E_0}\sigma_G^2\sigma_E^2 + x_{E_0}^2\sigma_G^2\sigma_E^2 + x_C^2\sigma_G^4 + 2x_{E_0}x_C\sigma_G^4 \\ &\quad + x_{E_0}^2\sigma_G^4 - x_{G_0}^2\sigma_E^4 - 2x_{G_0}(x_C + x_{E_0})\sigma_G^2\sigma_E^2 - (x_C + x_{E_0})\sigma_G^4 \\ &= x_{G_0}^2\sigma_E^2 + x_{G_0}^2\sigma_G^2\sigma_E^2 + x_C^2\sigma_G^2\sigma_E^2 + 2x_Cx_{E_0}\sigma_G^2\sigma_E^2 + x_{E_0}^2\sigma_G^2\sigma_E^2 + x_C^2\sigma_G^4 + 2x_{E_0}x_C\sigma_G^4 \\ &\quad + x_{E_0}^2\sigma_G^4 - x_{G_0}^2\sigma_E^4 - 2x_{G_0}x_C\sigma_G^2\sigma_E^2 - 2x_{G_0}x_{E_0}\sigma_G^2\sigma_E^2 - x_C^2\sigma_G^4 - 2x_{E_0}x_C\sigma_G^4 - x_{E_0}^2\sigma_G^4 \end{aligned}$$

Simplifying the opposite terms and writing those with  $x_C$  dependence first:

$$\begin{aligned} I_n &= \\ &= x_C^2\sigma_G^2\sigma_E^2 + 2x_{E_0}x_C\sigma_G^2\sigma_E^2 - 2x_{G_0}x_C\sigma_G^2\sigma_E^2 + x_{G_0}^2\sigma_G^2\sigma_E^2 + x_{E_0}^2\sigma_G^2\sigma_E^2 - 2x_{G_0}x_{E_0}\sigma_G^2\sigma_E^2 \\ &= x_C^2\sigma_G^2\sigma_E^2 + 2x_C(x_{E_0} - x_{G_0})\sigma_G^2\sigma_E^2 + [x_{G_0}^2 + x_{E_0}^2 - 2(x_{G_0}x_{E_0})] \sigma_G^2\sigma_E^2 \\ &\text{[Last term can be written in quadratic form:]} \\ &= x_C^2\sigma_G^2\sigma_E^2 + 2x_C(x_{E_0} - x_{G_0})\sigma_G^2\sigma_E^2 + (x_{G_0} - x_{E_0})^2\sigma_G^2\sigma_E^2 \\ &\text{[And the rest also corresponds to a quadratic expression:]} \\ &= \sigma_G^2\sigma_E^2 [x_C - (x_{E_0} - x_{G_0})]^2 \end{aligned}$$

Therefore the global independent term  $I$  is reduced to:

$$I = \frac{\sigma_G^2 \sigma_E^2 [x_C - (x_{E_0} - x_{G_0})]^2}{\sigma_G^2 + \sigma_E^2}$$

And so, the numerator  $N$  becomes:

$$N = (\sigma_G^2 + \sigma_E^2) \left[ x_G - \frac{\sigma_E^2 x_{G_0} + (x_C + x_{E_0}) \sigma_G^2}{\sigma_G^2 + \sigma_E^2} \right]^2 + \frac{\sigma_G^2 \sigma_E^2 [x_C - (x_{E_0} - x_{G_0})]^2}{\sigma_G^2 + \sigma_E^2}$$

Introducing this into the integral of  $\phi_C(x_C)$  (eq. A.11):

$$\begin{aligned} \phi_C(x_C) &= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi(\sigma_G^2 + \sigma_E^2)}} \exp \left\{ -\frac{1}{2} \left[ \frac{\sigma_E^2 (x_G - x_{G_0})^2 + \sigma_G^2 (x_G - x_C - x_{E_0})^2}{\sigma_G^2 \sigma_E^2} \right] \right\} dx_G \\ \phi_C(x_C) &= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi(\sigma_G^2 + \sigma_E^2)}} \exp \left\{ -\frac{1}{2} \left[ \frac{(\sigma_G^2 + \sigma_E^2) \left[ x_G - \frac{\sigma_E^2 x_{G_0} + (x_C + x_{E_0}) \sigma_G^2}{\sigma_G^2 + \sigma_E^2} \right]^2 + \frac{\sigma_G^2 \sigma_E^2 [x_C - (x_{E_0} - x_{G_0})]^2}{\sigma_G^2 + \sigma_E^2}}{\sigma_G^2 \sigma_E^2} \right] \right\} dx_G \end{aligned}$$

Separating again the exponential into two exponentials, one dependent on  $x_G$  and the other independent:

$$\phi_C(x_C) = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi(\sigma_G^2 + \sigma_E^2)}} \exp \left\{ -\frac{\sigma_G^2 + \sigma_E^2}{2\sigma_G^2 \sigma_E^2} \left[ x_G - \frac{\sigma_E^2 x_{G_0} + (x_C + x_{E_0}) \sigma_G^2}{\sigma_G^2 + \sigma_E^2} \right]^2 \right\} \exp \left\{ -\frac{1}{2} \frac{[x_C - (x_{E_0} - x_{G_0})]^2}{\sigma_G^2 + \sigma_E^2} \right\} dx_G$$

Since the second exponential is not dependent on  $x_G$  we can put it out of the integral. And reorganizing the constant term of the other exponential, we get:

$$\phi_C(x_C) = \frac{1}{\sqrt{2\pi(\sigma_G^2 + \sigma_E^2)}} \exp \left\{ -\frac{1}{2} \frac{[x_C - (x_{E_0} - x_{G_0})]^2}{\sigma_G^2 + \sigma_E^2} \right\} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi} \frac{\sigma_G \sigma_E}{\sqrt{\sigma_G^2 + \sigma_E^2}}} \exp \left\{ -\frac{\left( x_G - \frac{\sigma_E^2 x_{G_0} + (x_C + x_{E_0}) \sigma_G^2}{\sigma_G^2 + \sigma_E^2} \right)^2}{2 \left( \frac{\sigma_G \sigma_E}{\sqrt{\sigma_G^2 + \sigma_E^2}} \right)^2} \right\} dx_G$$

Written like this we may notice that the term in the integral corresponds to a normal density distribution on  $x_G$ , which means that the whole integral is equal to 1.

Therefore, the convolved distribution function finally becomes:

$$\phi_C(x_C) = \frac{1}{\sqrt{2\pi(\sigma_G^2 + \sigma_E^2)}} \exp \left\{ -\frac{[x_C - (x_{G_0} - x_{E_0})]^2}{2(\sigma_G^2 + \sigma_E^2)} \right\} \quad \text{with } x_C = x_G - x_E$$

which is also a normal distribution  $\phi_C(x_C) = \mathcal{N}(x_C, \sigma_C^2) = \mathcal{N}(x_{G_0} - x_{E_0}, \sigma_G^2 + \sigma_E^2)$

If the expected value for both variables is the same (as it is for the Gaia catalogue and the external one), then  $x_{G_0} - x_{E_0} = 0$ , and:

$$\phi_C(x_C) = \frac{1}{\sqrt{2\pi(\sigma_G^2 + \sigma_E^2)}} \exp \left\{ -\frac{(x_G - x_E)^2}{2(\sigma_G^2 + \sigma_E^2)} \right\}$$

Which generalized to a 3-variate normal distribution (following eq. A.5) and applied to the three astrophysical parameters ( $\mu_\alpha$ ,  $\mu_\delta$ , and  $\varpi$ ), it becomes:

$$\phi_C(\mu_\alpha, \mu_\delta, \varpi) = \frac{1}{\sqrt{(2\pi)^3 |\Sigma|}} \exp \left\{ -\frac{1}{2} (\vec{x}_G - \vec{x}_E)^T \Sigma^{-1} (\vec{x}_G - \vec{x}_E) \right\} \quad (\text{A.12})$$

with

$$\vec{x}_G = (\mu_{\alpha_G}, \mu_{\delta_G}, \varpi_G)$$

$$\vec{x}_E = (\mu_{\alpha_E}, \mu_{\delta_E}, \varpi_E)$$

$$\Sigma = \Sigma_G + \Sigma_E$$

Where this global covariance matrix  $\Sigma$  is the one that needs to be calculated to obtain the  $\chi^2$  variate of eq. A.6.

### A.2.2 Global covariance matrix

By definition of covariance matrix,  $\Sigma_G$  and  $\Sigma_E$  are defined as:

$$\Sigma_G = \begin{pmatrix} (\sigma_{\mu_\alpha}^2)_G & (\rho_{\mu_\alpha \mu_\delta} \sigma_{\mu_\alpha} \sigma_{\mu_\delta})_G & (\rho_{\mu_\alpha \varpi} \sigma_{\mu_\alpha} \sigma_\varpi)_G \\ (\rho_{\mu_\alpha \mu_\delta} \sigma_{\mu_\alpha} \sigma_{\mu_\delta})_G & (\sigma_{\mu_\delta}^2)_G & (\rho_{\mu_\delta \varpi} \sigma_{\mu_\delta} \sigma_\varpi)_G \\ (\rho_{\mu_\alpha \varpi} \sigma_{\mu_\alpha} \sigma_\varpi)_G & (\rho_{\mu_\delta \varpi} \sigma_{\mu_\delta} \sigma_\varpi)_G & (\sigma_\varpi^2)_G \end{pmatrix} \quad (\text{A.13})$$

$$\Sigma_E = \begin{pmatrix} (\sigma_{\mu_\alpha}^2)_E & (\rho_{\mu_\alpha \mu_\delta} \sigma_{\mu_\alpha} \sigma_{\mu_\delta})_E & (\rho_{\mu_\alpha \varpi} \sigma_{\mu_\alpha} \sigma_\varpi)_E \\ (\rho_{\mu_\alpha \mu_\delta} \sigma_{\mu_\alpha} \sigma_{\mu_\delta})_E & (\sigma_{\mu_\delta}^2)_E & (\rho_{\mu_\delta \varpi} \sigma_{\mu_\delta} \sigma_\varpi)_E \\ (\rho_{\mu_\alpha \varpi} \sigma_{\mu_\alpha} \sigma_\varpi)_E & (\rho_{\mu_\delta \varpi} \sigma_{\mu_\delta} \sigma_\varpi)_E & (\sigma_\varpi^2)_E \end{pmatrix} \quad (\text{A.14})$$

To calculate  $\Sigma$  we can make use of some properties of covariances in order to prove whether the sum of two covariance matrices is simply the sum of the two matrices. Consider X and Y be random variables defined on the same probability space, both with expectations and variances. Then the covariance is defined by:

$$COV(X, Y) = E[(X - x_0)(Y - y_0)]$$

which brings us to the follow two properties:

$$1. COV(X, Y) = E[XY] - E[X]E[Y]$$

$$2. COV(X, Y) = 0 \text{ if X and Y are independent, since in that case } E[XY] = E[X]E[Y]$$

Using the first property  $\Sigma$  can be written as:

$$\begin{aligned} \left[ \Sigma_{\vec{x}_G - \vec{x}_E} \right]_{ij} &= \\ &= E[(\vec{X}_{G_i} - \vec{X}_{E_i}, \vec{X}_{G_j} - \vec{X}_{E_j})] - E[(\vec{X}_{G_i} - \vec{X}_{E_i})]E[(\vec{X}_{G_j} - \vec{X}_{E_j})] \\ &= E[(\vec{X}_{G_i} \vec{X}_{G_j} - \vec{X}_{G_i} \vec{X}_{E_j} - \vec{X}_{E_i} \vec{X}_{G_j} + \vec{X}_{E_i} \vec{X}_{E_j})] - (E[\vec{X}_{G_i}] - E[\vec{X}_{E_i}])(E[\vec{X}_{G_j}] - E[\vec{X}_{E_j}]) \\ &= E[\vec{X}_{G_i} \vec{X}_{G_j}] - E[\vec{X}_{G_i} \vec{X}_{E_j}] - E[\vec{X}_{E_i} \vec{X}_{G_j}] + E[\vec{X}_{E_i} \vec{X}_{E_j}] \\ &\quad - E[\vec{X}_{G_i}]E[\vec{X}_{G_j}] - E[\vec{X}_{G_i}]E[\vec{X}_{E_j}] - E[\vec{X}_{E_i}]E[\vec{X}_{G_j}] + E[\vec{X}_{E_i}]E[\vec{X}_{E_j}] \\ &= COV(\vec{X}_{G_i}, \vec{X}_{G_j}) - COV(\vec{X}_{G_i}, \vec{X}_{E_j}) - COV(\vec{X}_{E_i}, \vec{X}_{G_j}) + COV(\vec{X}_{E_i}, \vec{X}_{E_j}) \end{aligned}$$

Since  $\vec{X}_G$  and  $\vec{X}_E$  are completely independent, we can apply the second property. Thus:

$$\Sigma = \left[ \Sigma_{\vec{x}_G - \vec{x}_E} \right]_{ij} = COV(\vec{X}_{G_i}, \vec{X}_{G_j}) + COV(\vec{X}_{E_i}, \vec{X}_{E_j}) = \Sigma_{G_{ij}} + \Sigma_{E_{ij}} \quad (\text{A.15})$$

### A.2.3 The $\chi^2$ variate

Finally, by combining:

1. The relation between the quadratic form of the random variables and the chi-square variate (eq. A.6)
2. The convolved distribution for three correlated variables between the Gaia catalogue distribution and the External catalogue distribution (eq. A.12)
3. The global covariance matrix  $\Sigma$ , which has been proved (eq. A.15) to be just the sum of the two matrices defined in equations A.13 and A.14

the  $\chi^2$  to be calculated for each star is:

$$\chi^2 = \vec{X}^T \Sigma^{-1} \vec{X}$$

with

$$\vec{X} = \begin{pmatrix} \mu_{\alpha_G} \cos \delta_{\alpha_G} - \mu_{\alpha_E} \cos \delta_{\alpha_E} \\ \mu_{\delta_G} - \mu_{\delta_E} \\ \varpi_G - \varpi_E \end{pmatrix} \quad (\text{A.16})$$

and

$$\Sigma = \begin{pmatrix} (\sigma_{\mu_\alpha}^2)_G + (\sigma_{\mu_\alpha}^2)_E & (\rho_{\mu_\alpha \mu_\delta} \sigma_{\mu_\alpha} \sigma_{\mu_\delta})_G + (\rho_{\mu_\alpha \mu_\delta} \sigma_{\mu_\alpha} \sigma_{\mu_\delta})_E & (\rho_{\mu_\alpha \varpi} \sigma_{\mu_\alpha} \sigma_\varpi)_G + (\rho_{\mu_\alpha \varpi} \sigma_{\mu_\alpha} \sigma_\varpi)_E \\ (\rho_{\mu_\alpha \mu_\delta} \sigma_{\mu_\alpha} \sigma_{\mu_\delta})_G + (\rho_{\mu_\alpha \mu_\delta} \sigma_{\mu_\alpha} \sigma_{\mu_\delta})_E & (\sigma_{\mu_\delta}^2)_G + (\sigma_{\mu_\delta}^2)_E & (\rho_{\mu_\delta \varpi} \sigma_{\mu_\delta} \sigma_\varpi)_G + (\rho_{\mu_\delta \varpi} \sigma_{\mu_\delta} \sigma_\varpi)_E \\ (\rho_{\mu_\alpha \varpi} \sigma_{\mu_\alpha} \sigma_\varpi)_G + (\rho_{\mu_\alpha \varpi} \sigma_{\mu_\alpha} \sigma_\varpi)_E & (\rho_{\mu_\delta \varpi} \sigma_{\mu_\delta} \sigma_\varpi)_G + (\rho_{\mu_\delta \varpi} \sigma_{\mu_\delta} \sigma_\varpi)_E & (\sigma_\varpi^2)_G + (\sigma_\varpi^2)_E \end{pmatrix} \quad (\text{A.17})$$





## WP944 v18.1 Detailed Description and Validation Tests Report

Gaia DPAC Validation Test Report published in Gaia Livelink [Ruiz-Dern et al., 2015]. The co-authors contributions were the development of some Validation Test Specifications as well as preparation of external catalogues. I implemented, together with C. Babusiaux, the WP944 VTS framework integrating different tools to handle the ensemble of tests, to facilitate the interaction with the database, and to provide global statistical methods or algorithms need for several VTS. I developed the statistical tests aiming to check Gaia parallaxes and proper motions consistency (VTS 030001), in which I implemented the  $\chi^2$  method using the covariance matrices, and the parallax zero-point (VTS 030005 and VTS 030006), for which I also prepared the RR Lyrae external catalogue for validation purposes. I also implemented the JUnit tests of the mentioned VTS to ensure their proper functioning, although not presented on the report.



WP944 v18.1  
Validation with external catalogues:  
  
Detailed Description  
and  
Validation Tests Report

---

prepared by: Laura Ruiz-Dern, Carine Babusiaux, Catherine  
Turon, Gráinne Costigan, Frédéric Arenou  
approved by: CU9  
reference: GAIA-C9-SP-OPM-LRD-001  
issue: A  
revision: 1  
date: 2015-11-25  
status: Issued

## Abstract

This is a Validation Test Report of WP944 and deliverable D5.6 of GENIUS (*Comparison with external catalogues*) which describes the details of the currently operational tests for TGAS catalogue (WP944 version 18.1). This document presents the results obtained when testing them on TGAS simulated data and on Hipparcos data. They will be applied later to real TGAS and other Gaia data releases.

## Document History

Issue	Revision	Date	Author	Comment
D	4	2015-11-01	LRD	
D	3	2015-09-15	LRD	
D	2	2015-09-04	CB	
D	1	2015-08-25	LRD	First draft

## Contents

<b>1</b>	<b>Introduction</b>	<b>8</b>
1.1	Objectives . . . . .	8
1.2	System Overview . . . . .	8
1.3	References . . . . .	9
1.4	Definitions, acronyms, and abbreviations . . . . .	12
1.5	Document Overview . . . . .	13
<b>2</b>	<b>VTS_010_001: Sky homogeneity</b>	<b>15</b>
2.1	Goals . . . . .	15
2.2	Input: data selection . . . . .	15
2.3	Process . . . . .	15
2.4	Statistical Techniques . . . . .	15
2.5	Test Results . . . . .	15
<b>3</b>	<b>VTS_010_003: Known issues for specific stars</b>	<b>17</b>
3.1	Goals . . . . .	17
3.2	Input: data selection . . . . .	17
3.3	Process . . . . .	17
3.4	Statistical Techniques . . . . .	17
3.5	Test Results . . . . .	18
<b>4</b>	<b>VTS_030_001: External parallaxes and proper motions</b>	<b>19</b>

4.1	Goals . . . . .	19
4.2	Input: data selection . . . . .	19
4.3	Process . . . . .	19
4.4	Statistical Techniques . . . . .	20
4.5	Test Results . . . . .	20
<b>5</b>	<b>VTS_030_003: Comparison to Hipparcos Astrometry</b>	<b>25</b>
5.1	Goals . . . . .	25
5.2	Input: data selection . . . . .	25
5.3	Process . . . . .	25
5.4	Statistical Techniques . . . . .	26
5.5	Test Results . . . . .	27
<b>6</b>	<b>VTS_030_004: Known High Proper Motion Stars</b>	<b>36</b>
6.1	Goals . . . . .	36
6.2	Input: data selection . . . . .	36
6.3	Process . . . . .	36
6.4	Statistical Techniques . . . . .	37
6.5	Test Results . . . . .	38
<b>7</b>	<b>VTS_030_005: Parallax zero-point and precision using external distances</b>	<b>49</b>
7.1	Goals . . . . .	49
7.2	Input: data selection . . . . .	49

7.3	Process . . . . .	51
7.4	Statistical Techniques . . . . .	52
7.5	Test Results . . . . .	53
<b>8</b>	<b>VTS_030_006: Parallax zero-point and precision using very distant stars</b>	<b>60</b>
8.1	Goals . . . . .	60
8.2	Input: data selection . . . . .	60
8.3	Process . . . . .	61
8.4	Statistical Techniques . . . . .	62
8.5	Test Results . . . . .	62
<b>9</b>	<b>VTS_060_003: Check QSO parallaxes and proper motions</b>	<b>70</b>
9.1	Goals . . . . .	70
9.2	Input: data selection . . . . .	70
9.3	Process . . . . .	70
9.4	Statistical Techniques . . . . .	70
9.5	Test Results . . . . .	70
<b>10</b>	<b>Summary of the Tests Results</b>	<b>71</b>
10.1	Summary Table . . . . .	71
10.2	Overall Assessment . . . . .	71
10.3	Impact of Test Environment . . . . .	71
10.4	Recommended Improvements . . . . .	71

<b>A</b>	<b>ANNEX A: Global Statistical Techniques</b>	<b>72</b>
A.1	Main Statistical Analysis . . . . .	72
A.2	Parallax and proper motion consistency: $\chi^2$ test . . . . .	74
<b>B</b>	<b>ANNEX B: LMC/SMC catalogue generation</b>	<b>77</b>
B.1	LMC Catalogue . . . . .	77
B.2	SMC Catalogue . . . . .	82



# 1 Introduction

## 1.1 Objectives

This document constitutes a Validation Test Report of the CU9 WP944 / GENIUS WP540. It covers the detailed description of the operational tests for TGAS (*Tycho-Gaia* Astrometric Solution) and their scientific validation report. The overall tests are specified in the Validation Test Specification document FA-061.

The aim of these tests is to check that Gaia data is coherent compared to other known catalogues, so to accept or not them for publication. They concern several types of objects (simple stars, variables, binaries, multiples systems, quasars,...) as well as different kind of parameters (parallaxes, proper motions, radial velocities,...). The priorities have been selected depending on the information received on each Release.

To ensure that the tests are correctly working and return correct results, we tested them on simulated TGAS data (hereafter TGAS\_SIMU, catalogue prepared by Frédéric Arenou, FA-066) as well as on the Hipparcos catalogue when applicable. These are the results presented in this document.

These external validation tools that we have developed in the frame of CU9 WP944 / GENIUS WP540 have already been applied on preliminary data.

## 1.2 System Overview

The WP944 makes use of the CU9 ValidationTools software IS-001, which provides global methods and tools for the whole Validation team. It also has its own structure for internal tests management, organization and statistics.

Each Validation Test Specification (VTS) returns a global flag (fail/pass/warning/exception/...) as well as a log, a detailed summary of the results and eventually some graphs. Only fail/pass are described for each VTS individually. If the test throws a java or R exception, it leads to an exception flag. If an exception is thrown due to the TAP service, it leads to an error flag. If the test was not conclusive (no fail or pass explicit indication), a warning flag is received.

All graphs are produced automatically for all tests. Therefore by default the axis will always refer to "Gaia" and to the "External" catalogues; see the captions for specific information and other details. Especially notice that in this VTR what it is called by default "Gaia" may refer to TGAS simulation (TGAS\_SIMU) or to Hipparcos (as explained in 1.1).

In this VTR version all the statistics are checked at 5% level of rejection (i.e. a p-value limit

of  $\alpha = 0.05$ : less than 1 in 20 chance of being wrong). It means that statistic results giving p-values smaller than  $\alpha$  will indicate a strong evidence against the null hypothesis, so the null hypothesis is rejected, the test fails. If the p-value is higher than  $\alpha$ , this evidence is weaker and so the null hypothesis cannot be rejected, the test passes. Mean values are checked at  $3\sigma$ , while outliers are checked at  $5\sigma$ .

JUnit tests are developed to ensure the proper functioning of the VTS.

### 1.3 References

Alcock, C., Alves, D.R., Axelrod, T.S., et al., 2004, AJ, 127, 334, ADS Link

[FA-066], Arenou, F., Michalik, D., 2015, *Simulated TGAS data used for validation Interface Control Document*,

GAIA-C9-SP-OPM-FA-066,

URL <http://www.rssd.esa.int/cs/livelihood/open/3332814>

[FA-061], Arenou, F., Babusiaux, C., Boudreault, S., et al., 2015, *Validation Tests Specification (WP940)*,

GAIA-C9-SP-OPM-FA-061,

URL <http://www.rssd.esa.int/cs/livelihood/open/3311561>

Balbinot, E., Santiago, B.X., Girardi, L., et al., 2015, MNRAS, 449, 1129, ADS Link

Bonanos, A.Z., Massa, D.L., Sewilo, M., et al., 2009, AJ, 138, 1003, ADS Link

Bonanos, A.Z., Lennon, D.J., Köhlinger, F., et al., 2010, AJ, 140, 416, ADS Link

Brunet, J.P., Imbert, M., Martin, N., et al., 1975, A&AS, 21, 109, ADS Link

Carpenter, J.M., 2001, AJ, 121, 2851, ADS Link

Carrera, R., Gallart, C., Hardy, E., Aparicio, A., Zinn, R., 2008, AJ, 135, 836, ADS Link

Carrera, R., Gallart, C., Aparicio, A., Hardy, E., 2011, AJ, 142, 61, ADS Link

Catelan, M., 2004, ApJ, 154, 633, ADS Link

Cioni, M.R., Loup, C., Habing, H.J., et al., 2000, A&AS, 144, 235, ADS Link

Cole, A.A., Tolstoy, E., Gallagher, J.S., III, Smecker-Hane, T.A., 2005, AJ, 129, 1465, ADS Link

Dias, B., Kerber, L.O., Barbuy, B., et al., 2014, A&A, 561, A106, ADS Link

Evans, C.J., Howarth, I.D., Irwin, M.J., Burnley, A.W., Harries, T.J., 2004, MNRAS, 353, 601, ADS Link

Gallart, C., Stetson, P.B., Meschin, I.P., Pont, F., Hardy, E., 2008, ApJL, 682, L89, ADS Link

Glatt, K., Grebel, E.K., Koch, A., 2010, A&A, 517, A50, ADS Link

Graczyk, D., Soszyński, I., Poleski, R., et al., 2011, ActaAstronomica, 61, 103, ADS Link

Graczyk, D., Pietrzyński, G., Thompson, I.B., et al., 2014, ApJ, 780, 59, ADS Link

Graham, J.A., 1975, PASP, 87, 641, ADS Link

Gruendl, R.A., Chu, Y.H., 2009, ApJs, 184, 172, ADS Link

Hainich, R., Rühling, U., Todt, H., et al., 2014, A&A, 565, A27, ADS Link

Harris, J., Zaritsky, D., 2006, AJ, 131, 2514, ADS Link

Haschke, R., Grebel, E.K., Frebel, A., et al., 2012, AJ, 144, 88, ADS Link

Henize, K.G., 1956, ApJs, 2, 315, ADS Link

Hilditch, R.W., Howarth, I.D., Harries, T.J., 2005, MNRAS, 357, 304, ADS Link

Hill, V., Andrievsky, S., Spite, M., 1995, AAP, 293, 347, ADS Link

Howarth, I.D., 2013, A&A, 555, A141, ADS Link

Kordopatis, G., Gilmore, G., Steinmetz, M., et al., 2013, AJ, 146, 134, ADS Link

Kourniotis, M., Bonanos, A.Z., Soszyński, I., et al., 2014, A&A, 562, A125, ADS Link

Kunkel, W.E., Irwin, M.J., Demers, S., 1997, A&AS, 122, 463, ADS Link

Maintz, G., 2005, A&A, 442, 381, ADS Link

Martayan, C., Frémat, Y., Hubert, A.M., et al., 2007, A&A, 462, 683, ADS Link

Martayan, C., Baade, D., Fabregat, J., 2010, A&A, 509, A11, ADS Link

Massey, P., 2002, ApJs, 141, 81, ADS Link

Massey, P., Duffy, A.S., 2001, ApJ, 550, 713, ADS Link

Mennickent, R.E., Pietrzyński, G., Gieren, W., Szewczyk, O., 2002, A&A, 393, 887, ADS Link

Muraveva, T., Palmer, M., Clementini, G., et al., 2015, ApJ, 807, 127, ADS Link

Murphy, M.T., Bessell, M.S., 2000, MNRAS, 311, 741, ADS Link

- Neugent, K.F., Massey, P., Skiff, B., et al., 2010, ApJ, 719, 1784, ADS Link
- Neugent, K.F., Massey, P., Skiff, B., Meynet, G., 2012, ApJ, 749, 177, ADS Link
- Ngeow, C.C., 2012, ApJ, 747, 50, ADS Link
- Parisi, M.C., Grocholski, A.J., Geisler, D., Sarajedini, A., Clariá, J.J., 2009, AJ, 138, 517, ADS Link
- Prevot, L., Martin, N., Rebeiro, E., Maurice, E., Rousseau, J., 1983, A&AS, 53, 255, ADS Link
- Saha, A., Olszewski, E.W., Brondel, B., et al., 2010, AJ, 140, 1719, ADS Link
- Sanduleak, N., 1968, AJ, 73, 246, ADS Link
- Sanduleak, N., 1970, Contributions from the Cerro Tololo Inter-American Observatory, 89, ADS Link
- Schlegel, D.J., Finkbeiner, D.P., Davis, M., 1998, ApJ, 500, 525, ADS Link
- Schmidtke, P.C., Chobanian, J.B., Cowley, A.P., 2008, AJ, 135, 1350, ADS Link
- Scowcroft, V., Freedman, W.L., Madore, B.F., et al., 2015, ArXiv e-prints, ADS Link
- Seale, J.P., Looney, L.W., Chu, Y.H., et al., 2009, ApJ, 699, 150, ADS Link
- [IS-001]**, Shih, I., 2015, *Software User Manual for CU9 ValidationTools*,  
GAIA-C9-UG-OPM-IS-001,  
URL <http://www.rssd.esa.int/cs/livelihood/open/3367978>
- Simons, R., Thilker, D., Bianchi, L., Wyder, T., 2014, Advances in Space Research, 53, 939, ADS Link
- Soszynski, I., Zebrun, K., Udalski, A., et al., 2002, ACTAA, 52, 143, ADS Link
- Soszynski, I., Poleski, R., Udalski, A., et al., 2008, ActaAstronomica, 58, 163, ADS Link
- Soszyński, I., Udalski, A., Szymański, M.K., et al., 2009a, ActaAstronomica, 59, 1, ADS Link
- Soszyński, I., Udalski, A., Szymański, M.K., et al., 2009b, ActaAstronomica, 59, 239, ADS Link
- Soszyński, I., Udalski, A., Szymański, M.K., et al., 2010a, ACTAA, 60, 91, ADS Link
- Soszyński, I., Udalski, A., Szymański, M.K., et al., 2010b, ACTAA, 60, 165, ADS Link
- Soszyński, I., Udalski, A., Szymański, M.K., et al., 2011, ACTAA, 61, 217, ADS Link

Spezzi, L., de Marchi, G., Panagia, N., Sicilia-Aguilar, A., Ercolano, B., 2012, MNRAS, 421, 78, ADS Link

Székely, P., Kiss, L.L., Jackson, R., et al., 2007, A&A, 463, 589, ADS Link

Udalski, A., Szymanski, M., Kubiak, M., et al., 2000, ActaAstronomica, 50, 307, ADS Link

Vallenari, A., Ortolani, S., Chiosi, C., 1994, A&AS, 108, 571, ADS Link

Yokogawa, J., Imanishi, K., Tsujimoto, M., Koyama, K., Nishiuchi, M., 2003, PAsJ, 55, 161, ADS Link

Zaritsky, D., Harris, J., Thompson, I.B., Grebel, E.K., 2004, AJ, 128, 1606, ADS Link

Zhao, H., Ibata, R.A., Lewis, G.F., Irwin, M.J., 2003, MNRAS, 339, 701, ADS Link

GAIA-C9-PL-ESAC-GGA-038-D: CU9 roadmap for publication of TGAS results

Running Integration Test In Local Environment, GAIA-C9-TN-OPM-IS-004

Specification for CU9: Validation database operation for TGAS release, GAIA-C9-SP-OPM-IS-003-D

Simulated TGAS data used for validation, GAIA-C9-SP-OPM-FA-066

## 1.4 Definitions, acronyms, and abbreviations

The following table includes a complete list of the acronyms used in this report. It has been generated from the on-line Gaia acronym list:

The following table has been generated from the on-line Gaia acronym list:

Acronym	Description
2MASS	Two-Micron All Sky Survey
AGB	Asymptotic Giant Branch (star)
AGN	Active Galactic Nucleus
CCD	Charge-Coupled Device
DE	Developmental Ephemerides (from JPL/NASA)
DEC	Declination
EB	Eclipsing Binary
ESAC	European Space Astronomy Centre (VilSpa)
FLAMES	Fiber Large Array Multi-Element Spectrograph (VLT)

GALEX	GALaxy Evolution eXplorer
GCVS	General Catalogue of Variable Stars
GENIUS	Gaia European Network for Improved User Services
ID	Identifier (Identification)
LMC	Large Magellanic Cloud (special, high-density area on the sky)
LPV	Long-Period Variable (star)
MC	Master Clock
MLE	Maximum-Likelihood Estimate
MOST	Micro-variability and Oscillations of STars
NGC	North Galactic Cap
NOAO	National Optical Astronomy Observatories (USA)
OB	On-Board
OGLE	Optical Gravitational Lensing Experiment
PI	Principal Investigator
PL	Period-Luminosity (star)
PM	Polarisation Maintaining
QSO	Quasi-Stellar Object
RA	Right Ascension
RAVE	RAAdial Velocity Experiment
RGB	Red Giant Branch (star)
SED	Spectral Energy Distribution
SMC	Small Magellanic Cloud (special, high-density area on the sky)
SRV	Semi-Regular Variable (star)
TAP	Table Access Protocol
TGAS	Tycho-Gaia Astrometric Solution
VIM	Variability-Induced Mover (DMSA, Hipparcos Catalogue)
VLBI	Very Long Baseline Interferometry
VTR	Validation Test Report (CU9)
VTs	Validation Test Specification (CU9)
WR	Window in RVS (covering a star spectrum; obsolete)
XM	Cross-Matching

## 1.5 Document Overview

Each section of this document corresponds to a VTS: goals, processes, specific statistical techniques and results obtained. The following table shows the list of tests described here (and already running at ESAC), with their associated priority.

TEST CASE ID	PURPOSE	PRIORITY
--------------	---------	----------

VTS_010_001	Sky homogeneity	TGAS 1
VTS_010_003	Known issues for specific stars	TGAS 3
VTS_030_001	External parallaxes and proper motions	TGAS 3
VTS_030_003	Comparison to Hipparcos astrometry	TGAS 1
VTS_030_004	Known High Proper Motion stars	TGAS 2
VTS_030_005	Parallax Zero-Point using external distances	TGAS 3
VTS_030_006	Parallax Zero-Point using very distant stars	TGAS 1
VTS_060_003	Check QSO parallaxes and proper motions	TGAS 4

A global summary of the results of all VTS is presented in the last section. The Annex A includes all those global statistical techniques that are common in several tests.

## 2 VTS\_010\_001: Sky homogeneity

### 2.1 Goals

This test aims to check the stellar full sky distribution homogeneity by comparison with other homogeneous full sky catalogues. By changing over different magnitudes ranges, it checks the number counts on different regions of the sky.

This full sky comparison will be done once the first Gaia releases come. As TGAS has no vocation to be complete, this test for TGAS just checks that there are no duplicated Tycho2 identifiers.

### 2.2 Input: data selection

The available test case only uses TGAS. External full sky catalogues will be used for the complete VTS once Gaia data arrive.

### 2.3 Process

This test case looks for Tycho2 identifiers in TGAS with more than one entries.

- FAIL: the same identifier appears more than once
- PASS: there are no duplicates

### 2.4 Statistical Techniques

There are no specific statistical techniques for this test case.

### 2.5 Test Results

#### 2.5.1 Summary

TEST CASE	TGAS_SIMU	SUMMARY
VTS_010_001_Tycho2Duplicates	PASSED	no duplicates



## 2.5.2 Detailed results

VTS\_010\_001\_Tycho2Duplicates : PASSED : no duplicates

## 3 VTS\_010\_003: Known issues for specific stars

### 3.1 Goals

This test aims to check whether known issues on specific stars found in previous catalogues and/or releases appear again or are solved with Gaia.

### 3.2 Input: data selection

From Hipparcos catalogue, we have checked their association with Tycho2 objects. Only one star with more than one identifier in Tycho2 was found: **HIP 67626**: This star is associated with two sources in Tycho2 (TYC 7291-1949-1 and TYC 7291-1949-2, see in Simbad) with the same brightness, while inspection of the Tycho2 raw data by Claus Fabricius reveals no case for such 2 Tycho2 stars.

### 3.3 Process

A XM by Tycho2 identifier is done between the input list and TGAS\_SIMU. The resulting list is used to determine if the test Fails or Passes, under the hypothesis that Gaia must associate only one identifier for this object (as it is in Hipparcos).

- FAIL:
  - No corresponding Tycho2 identifier found in Gaia DR1+
  - All Tycho2 identifiers associated to this object have also been found in Gaia
  - Gaia gives more identifiers for this object than in the input list
- PASS:
  - No corresponding Tycho2 identifier found in TGAS (as TGAS is not designed to be complete)
  - Gaia offers only one solution for this object

### 3.4 Statistical Techniques

There are no specific statistical techniques for this test.

## 3.5 Test Results

### 3.5.1 Summary

<b>TEST CASE</b>	<b>TGAS_SIMU</b>	<b>SUMMARY</b>
VTS_010_003_hip67626	PASSED	only one source for hip67626, all ok!

### 3.5.2 Detailed results

VTS\_010\_003\_hip67626 : PASSED : only one source for hip67626, all ok!

## 4 VTS\_030\_001: External parallaxes and proper motions

### 4.1 Goals

This test aims to check the Gaia parallaxes and proper motions versus literature values

### 4.2 Input: data selection

Currently a single catalogue is used:

- VLBI parallaxes: A compilation 28 VLBI parallaxes (including 18 tycho2) from : Asaki et 2010 ApJ 721 267, Imai et al 2012 PASJ 64 14, Kamezaki et al 2012 PASJ 64 7, Kuramaya et al 2005 ApJ 627 L49, Kusuno et al 2013 ApJ 774 107, Loinard et al 2007 ApJ 671 546, Melis et al 2014 Science 345 1029, Menten et al 2007 AA 474 515, Min et al 2014 PASJ 66 38, Nakagawa et al 2008 PASJ 1013 21, Nakagawa et al 2014 PASJ 66 101, Nyu et al 2011 PASJ 63 63, Torres et al 2007 ApJ 671 1813, Torres et al 2009 ApJ 698 242, Torres et al 2012 ApJ 747 18, Vlemmings et al 2003 AA 407 213, Vlemmings vanLangevelde 2007 AA 472 547, Zhang et al 2012 AA 544 A42, Zhang et al 2012 ApJ 744 23.

### 4.3 Process

- A XM by Tycho2 ID is done between the input and the Gaia catalogues
- Test case: Check parallax and proper motions following the  $\chi^2$  test described in Annex A.2
- Test case: Global test over the parallax
  - Global statistical analysis for  $\varpi_{EXT} - \varpi_{Gaia}$ , as detailed in Annex A.1. Correlations are checked against the  $V$  magnitude and the  $B - V$  colour.
- Test case: Global test over the proper motion
  - Global statistical analysis for both proper motion components:  $\mu_{\alpha_{Hip}} - \mu_{\alpha_{Gaia}}$  and  $\mu_{\delta_{Hip}} - \mu_{\delta_{Gaia}}$ , as detailed in Annex A.1. Correlations are checked against the magnitude and colour.
- FAIL:

- For the  $\chi^2$  test case: the p-value of the calculated  $\chi^2$  is lower than the given significance level ( $\alpha = 0.01$ )
- For the other test cases: at least one of the properties does not agree with the expected result
- PASS:
  - For the  $\chi^2$  test case: the p-value of the calculated  $\chi^2$  is higher than the given significance level ( $\alpha = 0.01$ )
  - For the other test cases: all properties are coherent with what is expected

## 4.4 Statistical Techniques

All statistical techniques applied here are defined in Annex A.

## 4.5 Test Results

### 4.5.1 Summary

TEST CASE	TGAS_SIMU	HIP
VTS_030_001_VLBI_tgas_CHI2	FAILED	FAILED
VTS_030_001_VLBI_tgas_PLX	PASSED	PASSED
VTS_030_001_VLBI_tgas_PM	FAILED	FAILED

### 4.5.2 Detailed results versus TGAS\_SIMU

**VLBI vs TGAS\_SIMU**

XM 14/18, 22.22 % not found

**Test Case:  $\chi^2$  on astrometric solution**

13 outliers (92.86%): proper motions of these objects do not match at 1% level of rejection. FAIL

**Test Case: Global statistical test over parallaxes**

Property	Obtained Value	Confidence Interval	Expected Value	PASS/FAIL	Related test	
					Test	p-value
Outliers	57.14% (8)	$5\sigma$	5.73e-05	-	-	-
Mean	$0.249 \pm 0.158$	$3\sigma$	-	PASS	t test	0.769
Standard Dev.	2.565	-	-	-	-	-
Median Abs. Dev.	2.918	-	-	-	-	-
Distr. Skewness	-0.085	-	0	-	agostino	-
Normality (KS test)	p-v: 0.430	p-v: 0.05	-	PASS	-	-
Corr with Ext Mag	p-v: 0.109	p-v: 0.05	-	PASS	-	-
Corr with Ext Colour	p-v: 0.131	p-v: 0.05	-	PASS	-	-

**Test Case: Global statistical test over proper motions**

Test on RA proper motion

Property	Obtained Value	Confidence Interval	Expected Value	PASS/FAIL	Related test	
					Test	p-value
Outliers	35.71% (5)	$5\sigma$	5.73e-05	-	-	-
Mean	$0.458 \pm 0.103$	$3\sigma$	-	FAIL	t test	0.016
Standard Dev.	1.611	-	-	-	-	-
Median Abs. Dev.	0.767	-	-	-	-	-
Distr. Skewness	-0.063	-	0	-	agostino	0.912
Normality (KS test)	p-v: 0.0001	p-v: 0.05	-	FAIL	-	-
Corr with Ext Mag	p-v: 0.841	p-v: 0.05	-	PASS	-	-
Corr with Ext Colour	p-v: 0.478	p-v: 0.05	-	PASS	-	-

Test on DEC proper motion

Property	Obtained Value	Confidence Interval	Expected Value	PASS/FAIL	Related test	
					Test	p-value
Outliers	64.29% (9)	$5\sigma$	5.73e-05	-	-	-
Mean	$0.372 \pm 0.477$	$3\sigma$	-	PASS	t test	0.537
Standard Dev.	1.149	-	-	-	-	-
Median Abs. Dev.	1.079	-	-	-	-	-
Distr. Skewness	0.537	-	0	-	agostino	-
Normality (KS test)	p-v: 0.946	p-v: 0.05	-	PASS	-	-
Corr with Ext Mag	p-v: 0.628	p-v: 0.05	-	PASS	-	-
Corr with Ext Colour	p-v: 0.994	p-v: 0.05	-	PASS	-	-

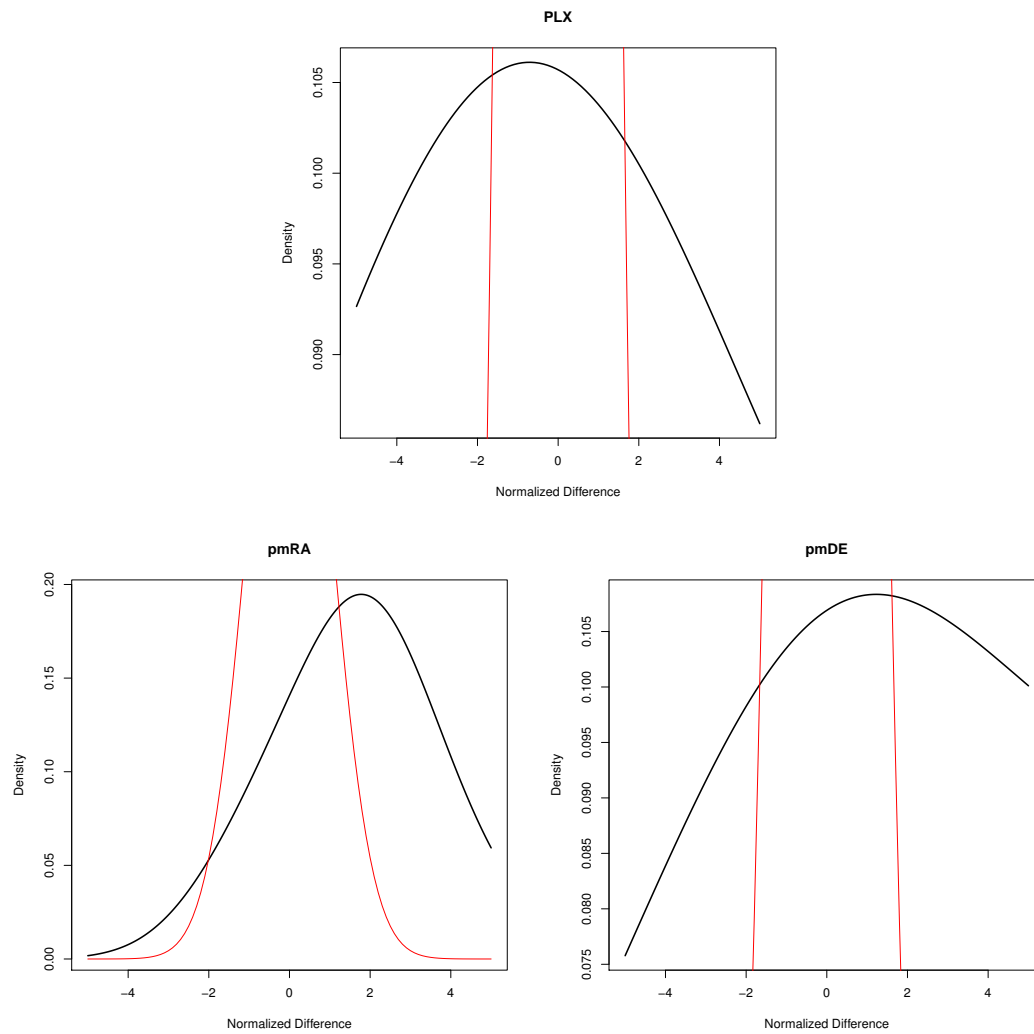


FIGURE 1: Normalized density distribution of VLBI-TGAS\_SIMU parallax difference (Top) and of the proper motion components difference (RA on the bottom-left, DEC on the bottom-right). The red curve corresponds to the expected distribution

### 4.5.3 Detailed results versus HIP

#### VLBI vs HIPPARCOS

XM 16/18, 11.11 % not found

---

**Test Case:  $\chi^2$  on astrometric solution**

 7 outliers (43.75%): proper motions of these objects do not match at 1% level of rejection. FAIL
 

---

**Test Case: Global statistical test over parallaxes**

Property	Obtained Value	Confidence Interval	Expected Value	PASS/FAIL	Related test	
					Test	p-value
Outliers	0% (0)	$5\sigma$	5.73e-05	-	-	-
Mean	$-0.713 \pm 0.338$	$3\sigma$	-	PASS	t test	0.098
Standard Dev.	1.035	-	-	-	-	-
Median Abs. Dev.	1.137	-	-	-	-	-
Distr. Skewness	-0.303	-	0	-	agostino	0.533
Normality (KS test)	p-v: 0.335	p-v: 0.05	-	PASS	-	-
Corr with Ext Mag	p-v: 0.762	p-v: 0.05	-	PASS	-	-
Corr with Ext Colour	p-v: 0.538	p-v: 0.05	-	PASS	-	-

**Test Case: Global statistical test over proper motions**

Test on RA proper motion

Property	Obtained Value	Confidence Interval	Expected Value	PASS/FAIL	Related test	
					Test	p-value
Outliers	6.25% (1)	$5\sigma$	5.73e-05	-	-	-
Mean	$0.685 \pm 0.308$	$3\sigma$	-	PASS	t test	0.312
Standard Dev.	2.372	-	-	-	-	-
Median Abs. Dev.	2.214	-	-	-	-	-
Distr. Skewness	-0.356	-	0	-	agostino	0.473
Normality (KS test)	p-v: 0.013	p-v: 0.05	-	FAIL	-	-
Corr with Ext Mag	p-v: 0.029	p-v: 0.05	-	FAIL	-	-
Corr with Ext Colour	p-v: 0.899	p-v: 0.05	-	PASS	-	-

Test on DEC proper motion

Property	Obtained Value	Confidence Interval	Expected Value	PASS/FAIL	Related test	
					Test	p-value
Outliers	6.25% (1)	$5\sigma$	5.73e-05	-	-	-
Mean	$0.251 \pm 0.333$	$3\sigma$	-	PASS	t test	0.519
Standard Dev.	1.319	-	-	-	-	-
Median Abs. Dev.	1.158	-	-	-	-	-
Distr. Skewness	-0.036	-	0	-	agostino	0.942
Normality (KS test)	p-v: 0.589	p-v: 0.05	-	PASS	-	-
Corr with Ext Mag	p-v: 0.915	p-v: 0.05	-	PASS	-	-
Corr with Ext Colour	p-v: 0.773	p-v: 0.05	-	PASS	-	-



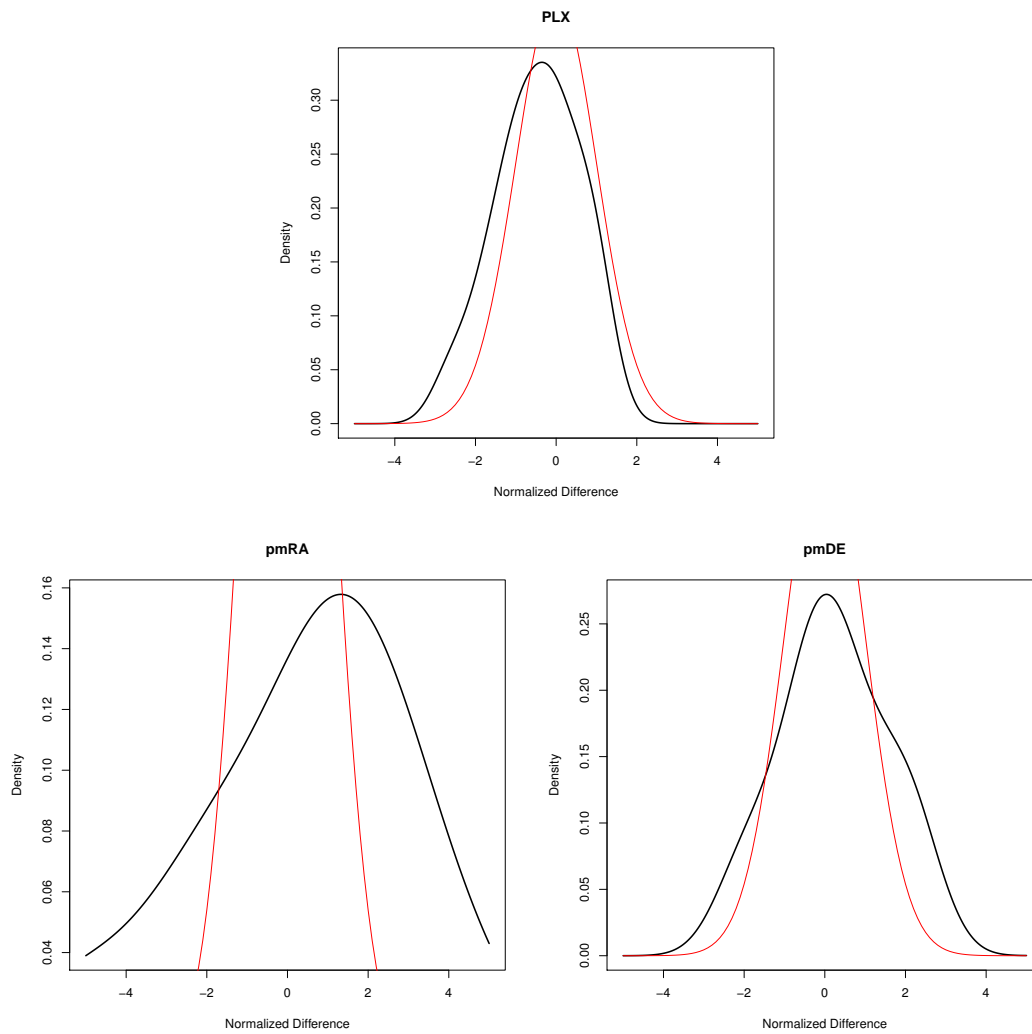


FIGURE 2: Normalized density distribution of the VLBI-Hipparcos parallax difference (Top) and of the proper motion components difference (RA on the bottom-left, DEC on the bottom-right). The red curve corresponds to the expected distribution

## 5 VTS\_030\_003: Comparison to Hipparcos Astrometry

### 5.1 Goals

This VTS aims to validate the Gaia astrometric solution versus Hipparcos. To do so, two subtests are implemented: the first one who directly checks Gaia parallaxes and proper motions versus Hipparcos; and the second one that verifies the Gaia solution for those stars with different proper motion (according to a  $\chi^2$  test) between Hipparcos and Tycho2 catalogues.

### 5.2 Input: data selection

Two input catalogues have been constructed from the HIP2 catalogue for each goal mentioned in previous section 5.1:

- VTS\_030\_003\_goodHip: "good" Hipparcos stars : Hipparcos stars having the solution type Sn=5 (5-parameter solution) without binarity flag in the literature (nor astrometric nor spectroscopic, mainly from the XHIP compilation) and with proper motion consistent with the Tycho2 ones (see next point). It leads to 93802 Hipparcos stars. All those stars have an original Hipparcos solution either So=0 (5-param), 1 (7 or 9 param), 2 (stochastic) or 5 (VIM), with only 2% with So>0.
- VTS\_030\_003\_badchi2HipTycho2: Bad  $\chi^2$  Hipparcos-Tycho2 : proper motion of Hipparcos stars with a 5-parameter solution and without binarity flag (see previous point) are compared to the Tycho2 ones.

$$\chi^2 = \frac{(\mu_{\alpha Tyc} - \mu_{\alpha Hip})^2}{\sigma_{\alpha Tyc}^2 + \sigma_{\alpha Hip}^2} + \frac{(\mu_{\delta Tyc} - \mu_{\delta Hip})^2}{\sigma_{\delta Tyc}^2 + \sigma_{\delta Hip}^2} > \chi_{0.999}^2 \quad (1)$$

$\chi_{0.999}^2$  corresponds to a p-value of 0.001, that is the 99.9th percentile value of the chi-square distribution of 2 degrees of freedom in our case:  $\chi_{0.999}^2 = 13.8$ . It leads to 1574 stars.

### 5.3 Process

For each subtest, the corresponding process applied is as follows:

#### 1. Testing Hipparcos versus Gaia parallaxes and proper motions

- A XM by Tycho2 ID is done between the input and the Gaia catalogues. Number of HIP stars without Gaia counterpart indicated (not a failing point for TGAS).

- Test case: Test over the parallax
  - Global statistical analysis for  $\varpi_{EXT} - \varpi_{Gaia}$ , as detailed in Annex A.1. Correlations are checked against the  $H_p$  magnitude and the  $V - I$  colour.
  - Aitoff skymap using calculated Gaia galactic coordinates to analyse parallax residuals and outliers.
- Test case: Test over the proper motion
  - Global statistical analysis for both proper motion components:  $\mu_{\alpha_{Hip}} - \mu_{\alpha_{Gaia}}$  and  $\mu_{\delta_{Hip}} - \mu_{\delta_{Gaia}}$ , as detailed in Annex A.1. Correlations are checked against the magnitude and colour.
  - Aitoff skymap using calculated Gaia galactic coordinates to analyse proper motion residuals and outliers.
- FAIL: at least one of the properties does not agree with the expected result
- PASS: all properties are coherent with what is expected

## 2. Testing Tycho2-Hipparcos proper motion inconsistency versus Gaia

- A XM by Tycho2 ID is done between the input and the Gaia catalogues. Number of HIP stars without Gaia counterpart indicated (not a failing point for TGAS).
- Test over the proper motion:
  - Global statistical analysis for both proper motion components between Tycho2 and Gaia, and between Hipparcos and Gaia:  $\mu_{\alpha_{Tycho2}} - \mu_{\alpha_{Gaia}}$ ,  $\mu_{\delta_{Tycho2}} - \mu_{\delta_{Gaia}}$ ,  $\mu_{\alpha_{Hip}} - \mu_{\alpha_{Gaia}}$  and  $\mu_{\delta_{Hip}} - \mu_{\delta_{Gaia}}$ , as detailed in Annex A.1.

An automatic fail/pass will be given by default by this VTS, but the final fail/pass summary provided by WP944 will actually be given by hand if it can be as there is no clear answer to whether Hipparcos or Tycho2 should give the right answer, although one would expect that unresolved long period binaries should have a better proper motion using a longer time base as is Tycho2.

## 5.4 Statistical Techniques

All statistical techniques applied here are defined in Annex A.1.

## 5.5 Test Results

### 5.5.1 Summary

TEST CASE	TGAS_SIMU
VTS_030_003_goodHip_PLX	FAILED
VTS_030_003_goodHip_PM	FAILED
VTS_030_003_badchi2HipTycho2	FAILED

### 5.5.2 Detailed results

The following tables and graphs show the detailed results for each test case run against TGAS\_SIMU. Note that the TGAS\_SIMU uses colours and magnitudes to compute simulated parallaxes (explaining the strong correlation in the first test case: if parallax errors are well estimated, then the normalised residuals should be independent of magnitudes and colours) and the Tycho2 proper motions (explaining the too good agreement of the second test case with the Tycho2 solution!).

#### 1. Testing Hipparcos versus Gaia parallaxes and proper motions

##### Parallaxes

##### Parallaxes Hipparcos vs TGAS\_SIMU

XM 90884/93802, 3.11 % not found

Property	Obtained Value	Confidence Interval	Expected Value	PASS/FAIL	Related test	
					Test	p-value
Outliers	1.02% (923)	$5\sigma$	5.73e-05	-	-	-
Mean	$0.059 \pm 0.003$	$3\sigma$	-	FAIL	t test	1.278
Standard Dev.	1.158	-	-	-	-	-
Median Abs. Dev.	0.909	-	-	-	-	-
Distr. Skewness	-0.674	-	0	-	agostino	-
Normality (KS test)	p-v: 0	p-v: 0.05	-	FAIL	-	-
Corr with Ext Mag	p-v: 0	p-v: 0.05	-	FAIL	-	-
Corr with Ext Colour	p-v: 0	p-v: 0.05	-	FAIL	-	-

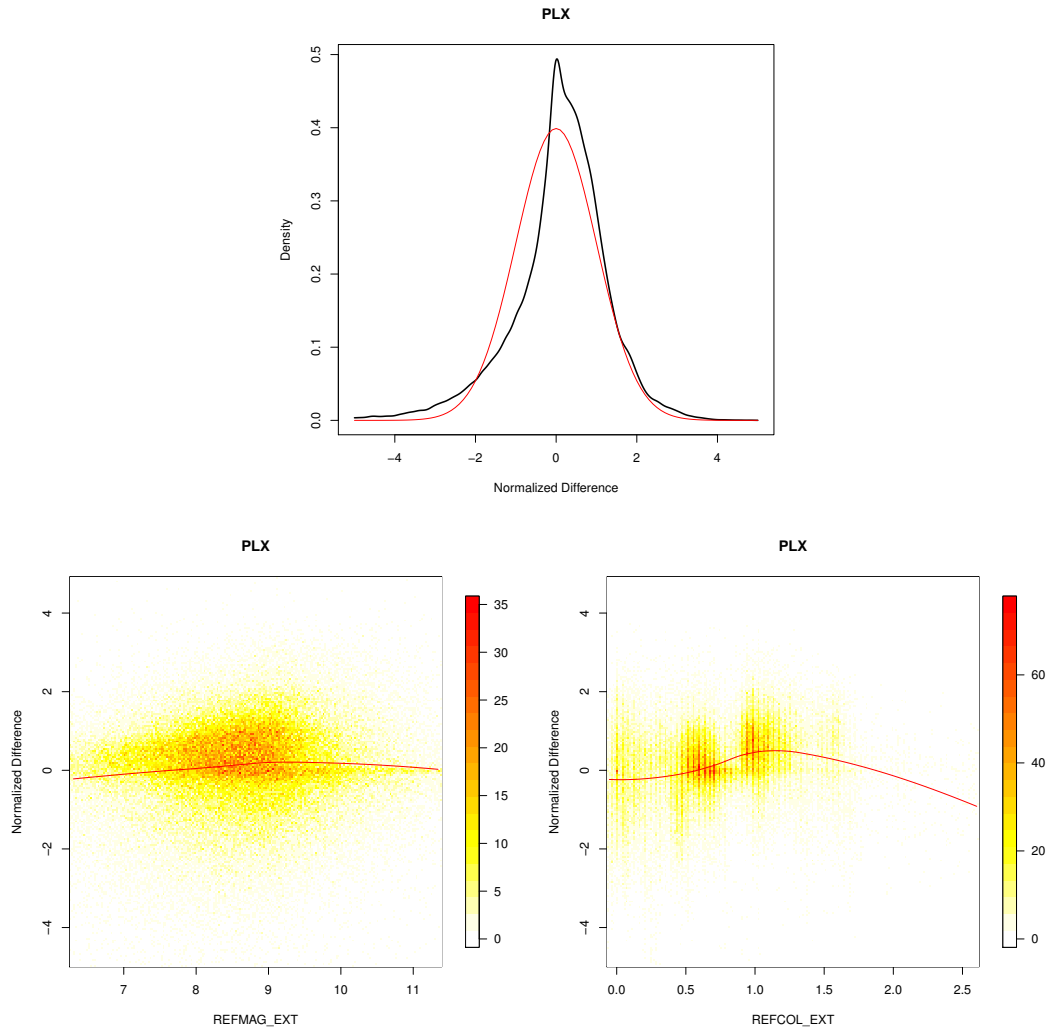


FIGURE 3: Normalized difference density distribution of the Hipparcos-TGAS SIMU parallax (Top), of the external magnitude  $H_p$  (bottom-left) and the external colour  $V - I$  (bottom-right). The red curve corresponds to the expected distribution

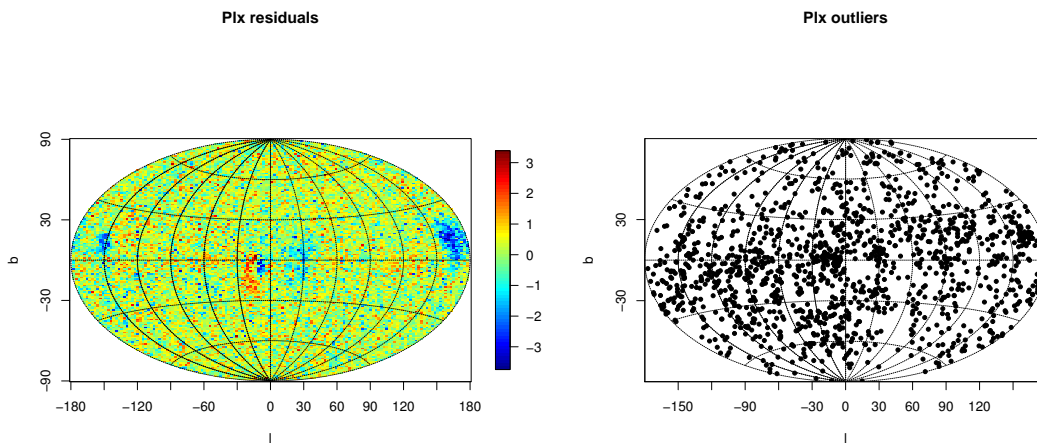


FIGURE 4: Aitoff skymaps (using calculated Gaia galactic coordinates) of the normalized difference of the TGAS\_SIMU parallax of the outliers (Left) and of the residuals (Right). Bad results around the galactic plane are probably due to extinction estimation problems in TGAS simulation

## Proper Motions

### Proper Motions Hipparcos vs TGAS\_SIMU

XM 90884/93802, 3.11 % not found, 336 external cat stars without PM

Test on RA proper motion

Property	Obtained Value	Confidence Interval	Expected Value	PASS/FAIL	Related test	
					Test	p-value
Outliers	0.039% (36)	$5\sigma$	5.73e-05	-	-	-
Mean	$0.003 \pm 0.004$	$3\sigma$	-	PASS	t test	0.014
Standard Dev.	0.237	-	-	-	-	-
Median Abs. Dev.	0.174	-	-	-	-	-
Distr. Skewness	-0.034	-	0	-	agostino	-
Normality (KS test)	p-v: 0	p-v: 0.05	-	FAIL	-	-
Corr with Ext Mag	p-v: 0.494	p-v: 0.05	-	PASS	-	-
Corr with Ext Colour	p-v: 1.55e-11	p-v: 0.05	-	FAIL	-	-

Test on DEC proper motion

Property	Obtained Value	Confidence Interval	Expected Value	PASS/FAIL	Related test	
					Test	p-value
Outliers	0.031% (28)	$5\sigma$	5.73e-05	-	-	-
Mean	$-0.007 \pm 0.004$	$3\sigma$	-	PASS	t test	6.75e-17
Standard Dev.	0.211	-	-	-	-	-
Median Abs. Dev.	0.154	-	-	-	-	-
Distr. Skewness	0.397	-	0	-	agostino	-
Normality (KS test)	p-v: 0	p-v: 0.05	-	FAIL	-	-
Corr with Ext Mag	p-v: 7.71e-08	p-v: 0.05	-	FAIL	-	-
Corr with Ext Colour	p-v: 9.79e-07	p-v: 0.05	-	FAIL	-	-

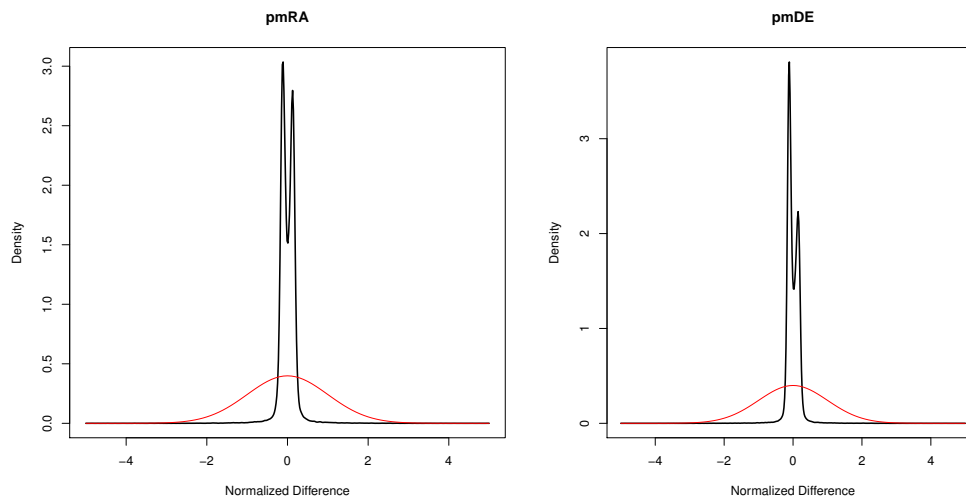


FIGURE 5: Normalized density distribution of the Hipparcos-TGAS.SIMU proper motion difference for each component (RA on the left, DEC on the right). The red curve corresponds to the expected distribution



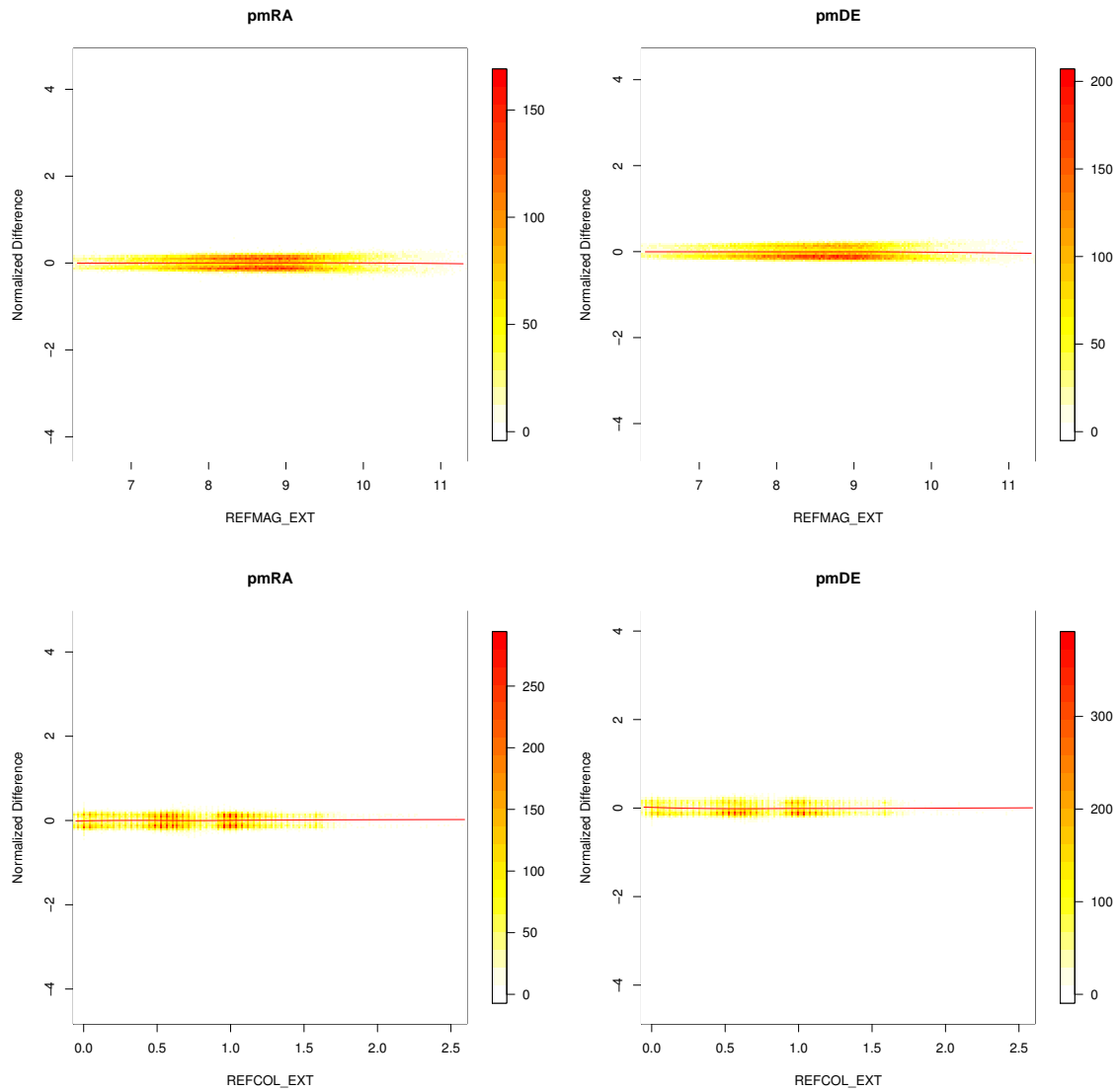


FIGURE 6: Normalized density distribution of the Hipparcos-TGAS\_SIMU reference magnitude  $H_p$  (top) and reference colour  $V - I$  (bottom), for RA proper motion component (left) and DEC component (right)

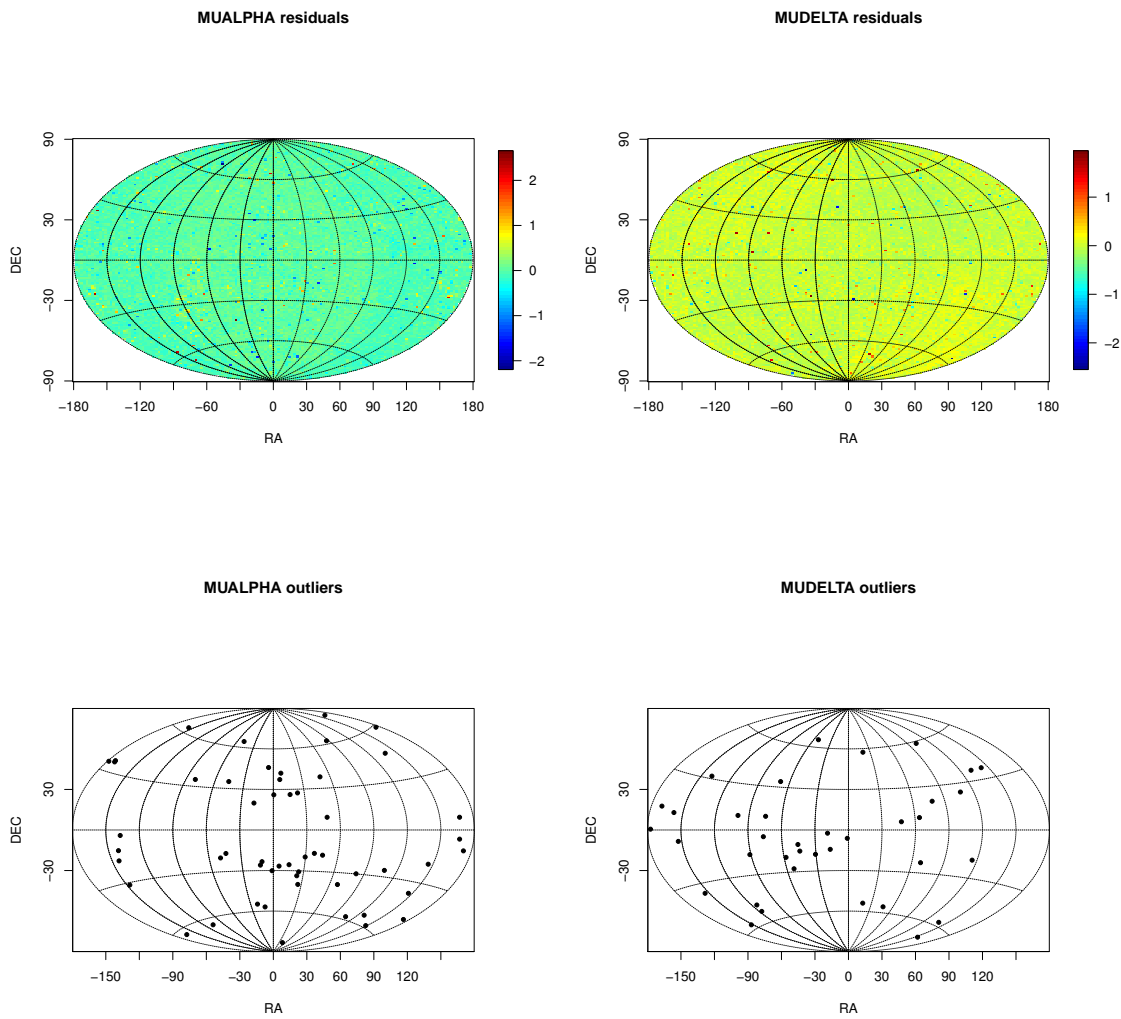


FIGURE 7: Aitoff skymaps (in equatorial coordinates) of the normalized difference of the TGAS.SIMU proper motions of the outliers (Top) and of the residuals (Bottom)

## 2. Testing Tycho2-Hipparcos proper motion inconsistency versus Gaia

**Unconsistent Hipparcos-Tycho2 PM vs TGAS\_SIMU**

XM 1461/1574, 3.11 % not found, 336 external cat stars without PM

## Test on Tycho2 RA proper motion

Property	Obtained Value	Confidence Interval	Expected Value	PASS/FAIL	Related test	
					Test	p-value
Outliers	0.14% (2)	$5\sigma$	5.73e-05	-	-	-
Mean	$0.0006 \pm 0.0029$	$3\sigma$	-	PASS	t test	0.691
Standard Dev.	0.326	-	-	-	-	-
Median Abs. Dev.	0.162	-	-	-	-	-
Distr. Skewness	1.879	-	0	-	agostino	1.92e-90
Normality (KS test)	p-v: 0	p-v: 0.05	-	FAIL	-	-

## Test on Hipparcos RA proper motion

Property	Obtained Value	Confidence Interval	Expected Value	PASS/FAIL	Related test	
					Test	p-value
Outliers	54.35% (794)	$5\sigma$	5.73e-05	-	-	-
Mean	$-0.041 \pm 0.031$	$3\sigma$	-	PASS	t test	0.690
Standard Dev.	2.983	-	-	-	-	-
Median Abs. Dev.	3.987	-	-	-	-	-
Distr. Skewness	0.042	-	0	-	agostino	0.658
Normality (KS test)	p-v: 0	p-v: 0.05	-	FAIL	-	-

## Test on Tycho2 DEC proper motion

Property	Obtained Value	Confidence Interval	Expected Value	PASS/FAIL	Related test	
					Test	p-value
Outliers	0.068% (1)	$5\sigma$	5.73e-05	-	-	-
Mean	$-0.008 \pm 0.029$	$3\sigma$	-	PASS	t test	0.544
Standard Dev.	0.308	-	-	-	-	-
Median Abs. Dev.	0.141	-	-	-	-	-
Distr. Skewness	3.624	-	0	-	agostino	2.41e-175
Normality (KS test)	p-v: 0	p-v: 0.05	-	FAIL	-	-

## Test on Hipparcos DEC proper motion

Property	Obtained Value	Confidence Interval	Expected Value	PASS/FAIL	Related test	
					Test	p-value
Outliers	64.48% (942)	$5\sigma$	5.73e-05	-	-	-
Mean	$-0.068 \pm 0.028$	$3\sigma$	-	PASS	t test	0.264
Standard Dev.	3.074	-	-	-	-	-
Median Abs. Dev.	4.152	-	-	-	-	-
Distr. Skewness	0.039	-	0	-	agostino	0.713
Normality (KS test)	p-v: 0	p-v: 0.05	-	FAIL	-	-

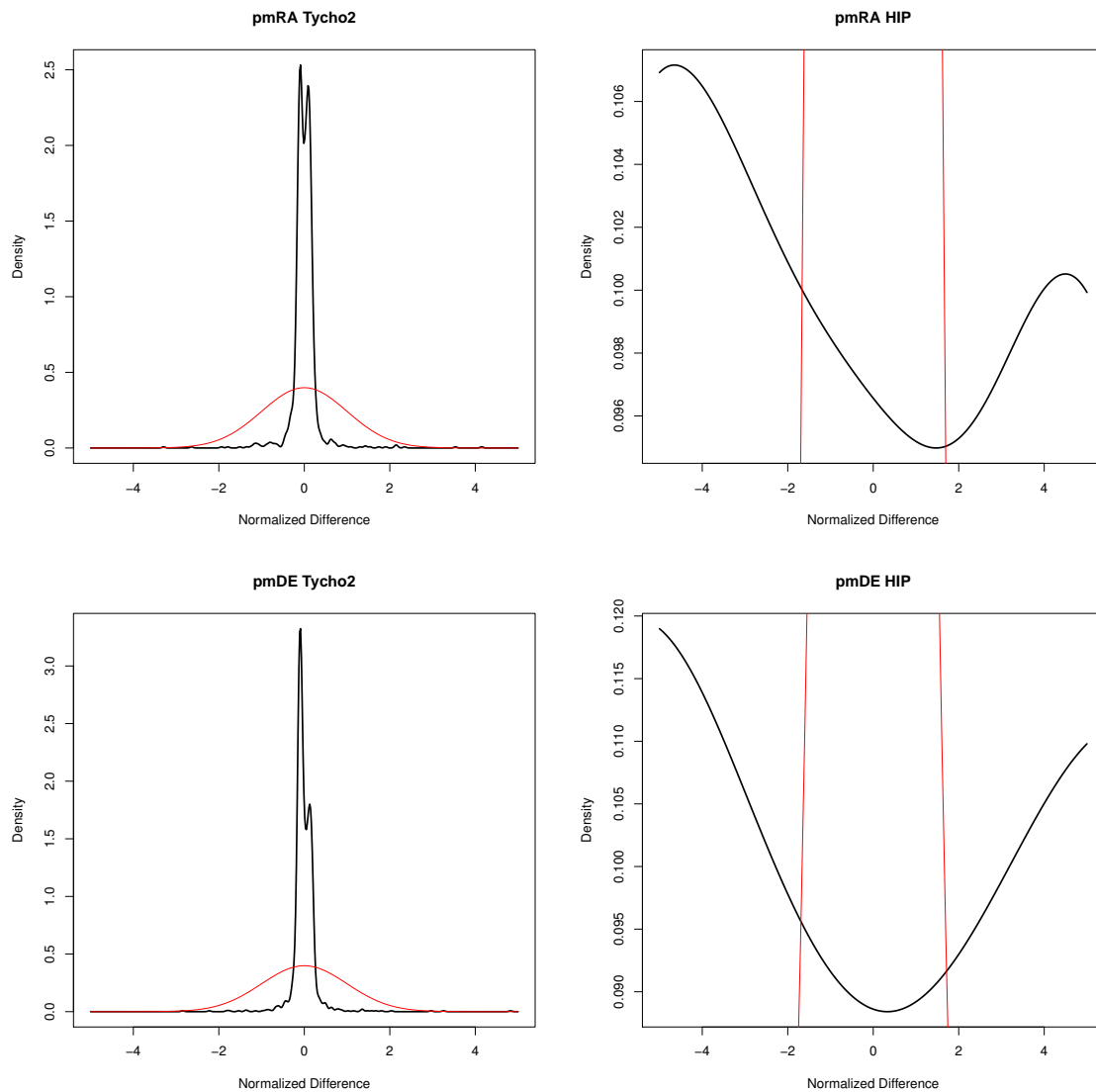


FIGURE 8: Normalized difference density distribution of the Tycho2-TGAS\_SIMU (Left) and Hipparcos-TGAS\_SIMU (Right) RA proper motion component (top) and DEC component (bottom). The red curve corresponds to the expected normal distribution

## 6 VTS\_030\_004: Known High Proper Motion Stars

### 6.1 Goals

This test aims to check the proper motions of stars with high velocities ( $\mu > 1$  arcsec), looking at their density distributions and their dependencies with other parameters.

### 6.2 Input: data selection

Three input catalogues were prepared for this test:

- VTS\_030\_004\_tycho2\_HV: Tycho2 High Proper Motion stars: Tycho2 stars with  $\mu > 1$  arcsec/yr
- VTS\_030\_004\_simbadHV\_tgas: Simbad-TGAS High Proper Motion stars: Simbad stars with Tycho2 identifiers and  $\mu > 0.5$  arcsec/yr
- VTS\_030\_004\_TGASHPM: same as previous catalogue including duplicated IDs due to stellar binarity

### 6.3 Process

The test is subdivided in two parts. The first two input catalogues (VTS\_030\_004\_tycho2\_HV and VTS\_030\_004\_simbadHV\_tgas) are used in the first part, and the last one (VTS\_030\_004\_TGASHPM) is used in the second part.

#### 1. Testing High Proper Motion (HPM) catalogues versus Gaia

*Note:* the reference magnitude and colour for each input catalogue are specified in the *Detailed results* section 6.5.2.

- A XM by Tycho2 identifier is done between the input catalogue and TGAS\_SIMU. The number of stars without TGAS\_SIMU counterpart is noted and manually checked against number of stars without cross-match in other VTS
- Check density distribution of the reference magnitude, reference colour and proper motion of the external catalogue (for both found and not found stars)
- If there is reference magnitude  $G$  for Gaia (which is not the case for TGAS), check colour comparison between the colour calculated using the reference magnitudes of each ( $G - magEXT$ ) and the reference colour of the external catalogue

- If the external catalogue contains proper motion components ( $\mu_\alpha$  and  $\mu_\delta$ ), they are analysed as detailed in Annex A.1, for  $\mu_{\alpha_{EXT}} - \mu_{\alpha_{Gaia}}$  and  $\mu_{\delta_{EXT}} - \mu_{\delta_{Gaia}}$ . Correlations are checked against the external magnitude and colour.
- If there are no proper motion components with errors (Simbad case in this version), then it checks the residuals of the proper motion, and the difference between Gaia and the External catalogue proper motions versus the reference external magnitude and versus the reference external colour

These results are used to determine if the test fails or passes:

- FAIL: at least one property of the residuals in one of the proper motion component deviates from the normalized Gaussian
- PASS: all properties are coherent with what is expected
- WARNING: no proper motion error provided in the input catalogue, therefore no statistical test done.

## 2. Testing High Proper Motion data of Gaia itself

The goal is to check if stars indicated in TGAS\_SIMU as HPM ( $\mu > 1$  arcsec) have really high proper motions (i.e. are considered as HPM in Simbad). To do so, a XM is done between the TGAS\_SIMU stars selected to have  $\mu > 1$  arcsec and the Simbad list of stars catalogued as having  $\mu > 0.5$  arcsec.

This test will fail or pass when:

- FAIL: there are TGAS stars not considered as HPM in Simbad
- PASS: all HPM TGAS stars are known to be HPM

## 6.4 Statistical Techniques

All statistical techniques applied here are defined in Annex A.1.

## 6.5 Test Results

### 6.5.1 Summary

TEST CASE	TGAS_SIMU	SUMMARY	HIP	SUMMARY
VTS_030_004_tycho2_HV VTS_030_004_simbadHV_tgas	FAILED WARNING		FAILED WARNING	
VTS_030_004_TGASHPM	FAILED	4 TGAS_SIMU HPM not HPM in Simbad	FAILED	1 HIP HPM not HPM in Simbad

### 6.5.2 Detailed results

The following tables and graphs show the results obtained for each test case.

## 1. Testing High Proper Motion (HPM) catalogues versus Gaia and Hipparcos

**Catalogue: VTS\_030\_004\_tycho2\_HV (Tycho2 HPM)**

 Reference magnitude:  $V_T$ , Reference colour:  $B - V, \mu > 1$  arcsec/yr

**HPM Tycho2 results on TGAS\_SIMU**

XM 118/178, 33.71 % not found

Test on RA proper motion						
Property	Obtained Value	Confidence Interval	Expected Value	PASS/FAIL	Related test	
					Test	p-value
Outliers	1.69% (2)	$5\sigma$	5.73e-05	-	-	-
Mean	$-0.146 \pm 0.135$	$3\sigma$	-	PASS	t test	0.120
Standard Dev.	0.885	-	-	-	-	-
Median Abs. Dev.	0.215	-	-	-	-	-
Distr. Skewness	-0.454	-	0	-	agostino	0.043
Normality (KS test)	p-v: 5.297e-08	p-v: 0.05	-	FAIL	-	-
Corr with Ext Mag	p-v: 0.874	p-v: 0.05	-	PASS	-	-
Corr with Ext Colour	p-v: 0.271	p-v: 0.05	-	PASS	-	-

Test on DEC proper motion						
Property	Obtained Value	Confidence Interval	Expected Value	PASS/FAIL	Related test	
					Test	p-value
Outliers	3.39% (4)	$5\sigma$	5.73e-05	-	-	-
Mean	$0.206 \pm 0.138$	$3\sigma$	-	PASS	t test	0.045
Standard Dev.	0.931	-	-	-	-	-
Median Abs. Dev.	0.179	-	-	-	-	-
Distr. Skewness	0.947	-	0	-	agostino	0.0001
Normality (KS test)	p-v: 9.541e-10	p-v: 0.05	-	FAIL	-	-
Corr with Ext Mag	p-v: 0.689	p-v: 0.05	-	PASS	-	-
Corr with Ext Colour	p-v: 0.136	p-v: 0.05	-	PASS	-	-



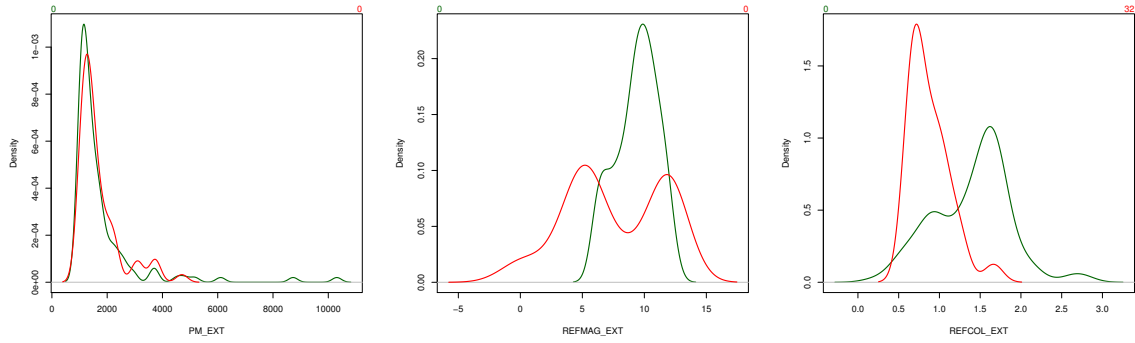


FIGURE 9: From left to Right: Normalized density distribution HPM Tycho2 proper motion, reference magnitude  $V_T$  and reference colour  $B - V$  of the crossmatched stars Tycho2 vs TGAS.SIMU (in green) and of the not found stars in TGAS.SIMU (in red)

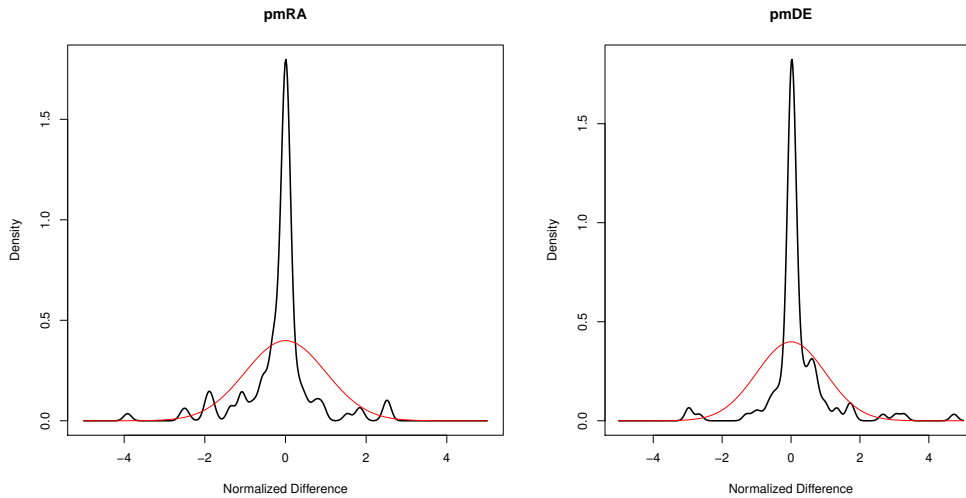


FIGURE 10: Normalized density distribution of the HPM Tycho2-TGAS\_SIMU proper motion components differences (RA on the left, DEC on the right). Red curve is the theoretical distribution

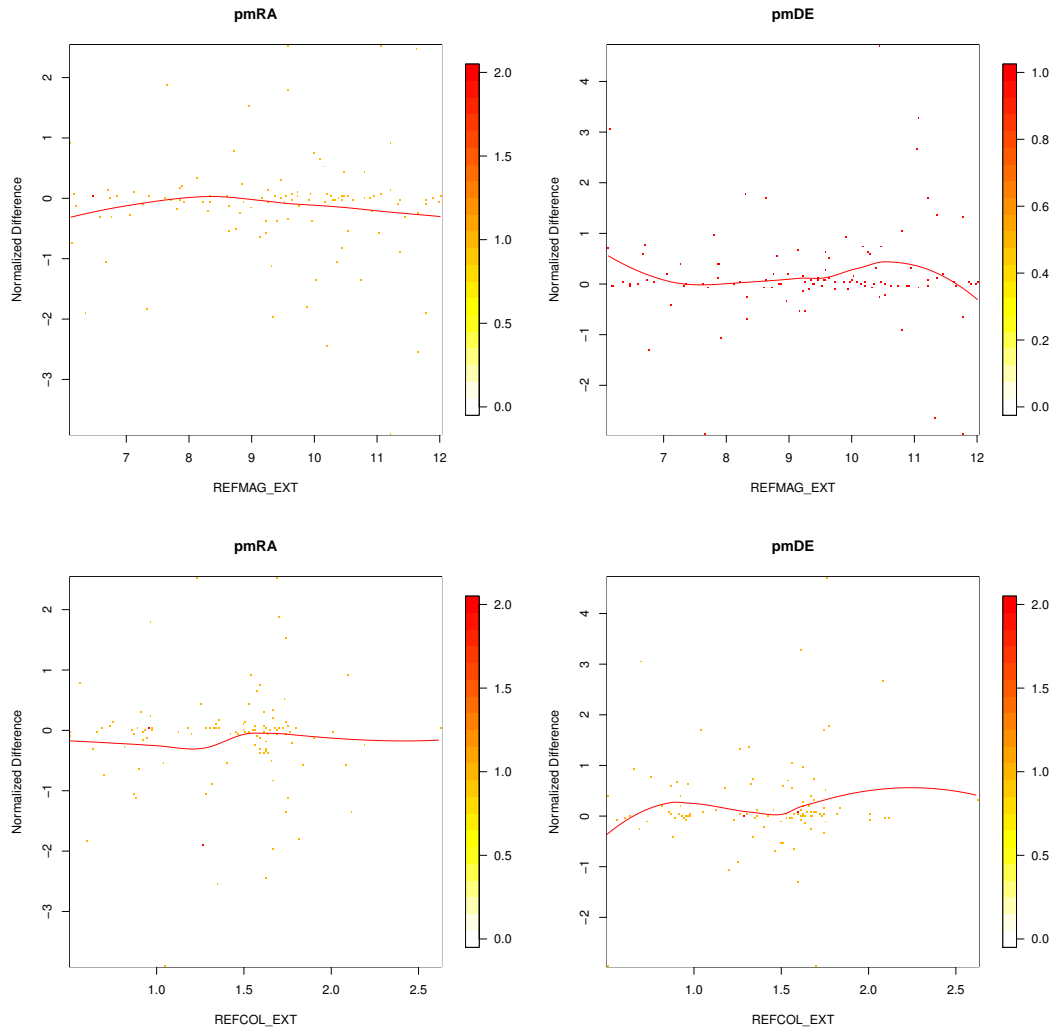


FIGURE 11: Normalized difference density distribution of the HPM Tycho2-TGAS\_SIMU proper motions respect to the HPM Tycho2 reference magnitude  $V_T$  (top) and respect to the HPM Tycho2 reference colour  $B - V$  (bottom), for RA proper motion component (left) and DEC component (right). The red curves correspond to automatic fits

**HPM Tycho2 results on HIPPARCOS**

XM 168/178, 5.62 % not found

Test on RA proper motion						
Property	Obtained Value	Confidence Interval	Expected Value	PASS/FAIL	Related test	
					Test	p-value
Outliers	7.14% (12)	$5\sigma$	5.73e-05	-	-	-
Mean	$0.472 \pm 0.132$	$3\sigma$	-	FAIL	t test	0.079
Standard Dev.	1.276	-	-	-	-	-
Median Abs. Dev.	1.009	-	-	-	-	-
Distr. Skewness	-0.047	-	0	-	agostino	0.800
Normality (KS test)	p-v: 0.121	p-v: 0.05	-	PASS	-	-
Corr with Ext Mag	p-v: 0.089	p-v: 0.05	-	PASS	-	-
Corr with Ext Colour	p-v: 0.608	p-v: 0.05	-	PASS	-	-

Test on DEC proper motion						
Property	Obtained Value	Confidence Interval	Expected Value	PASS/FAIL	Related test	
					Test	p-value
Outliers	8.33% (14)	$5\sigma$	5.73e-05	-	-	-
Mean	$0.254 \pm 0.124$	$3\sigma$	-	PASS	t test	0.244
Standard Dev.	1.390	-	-	-	-	-
Median Abs. Dev.	1.417	-	-	-	-	-
Distr. Skewness	-0.023	-	0	-	agostino	0.902
Normality (KS test)	p-v: 0.050	p-v: 0.05	-	PASS	-	-
Corr with Ext Mag	p-v: 0.024	p-v: 0.05	-	FAIL	-	-
Corr with Ext Colour	p-v: 0.335	p-v: 0.05	-	PASS	-	-

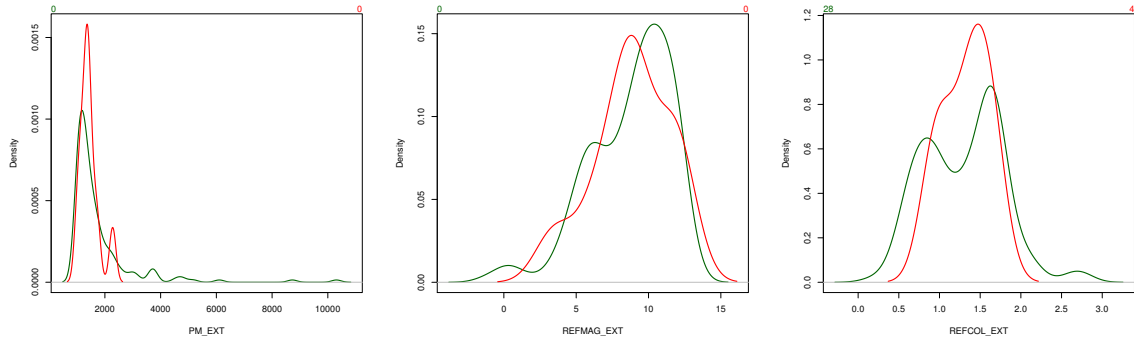


FIGURE 12: From left to Right: Normalized density distribution HPM Tycho2 proper motion, reference magnitude and reference colour of the crossmatched stars Tycho2 vs Hipparcos (in green) and of the not found stars in Hipparcos (in red)

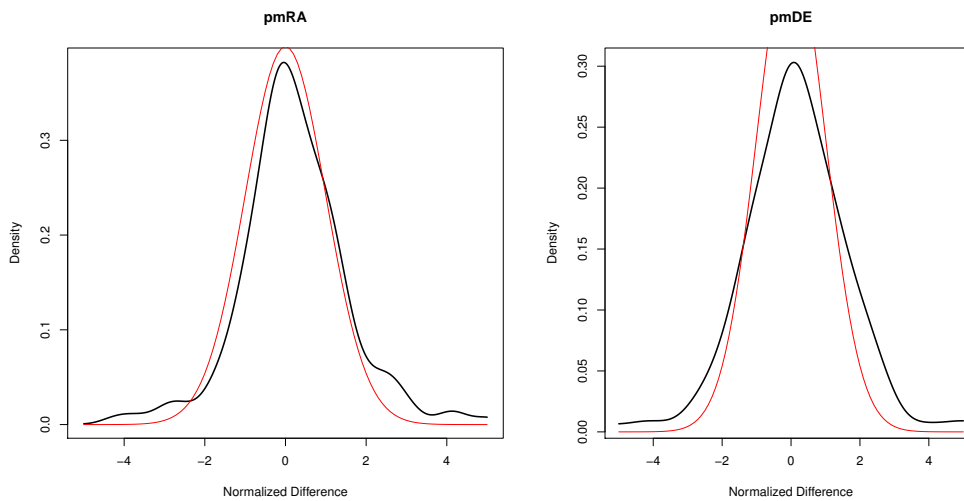


FIGURE 13: Normalized density distribution of the HPM Tycho2-Hipparcos proper motion components differences (RA on the left, DEC on the right). Red curve is the theoretical distribution. Notice that Hipparcos underestimates the errors for HPM stars since the measurements are more instantaneous than in Tycho2

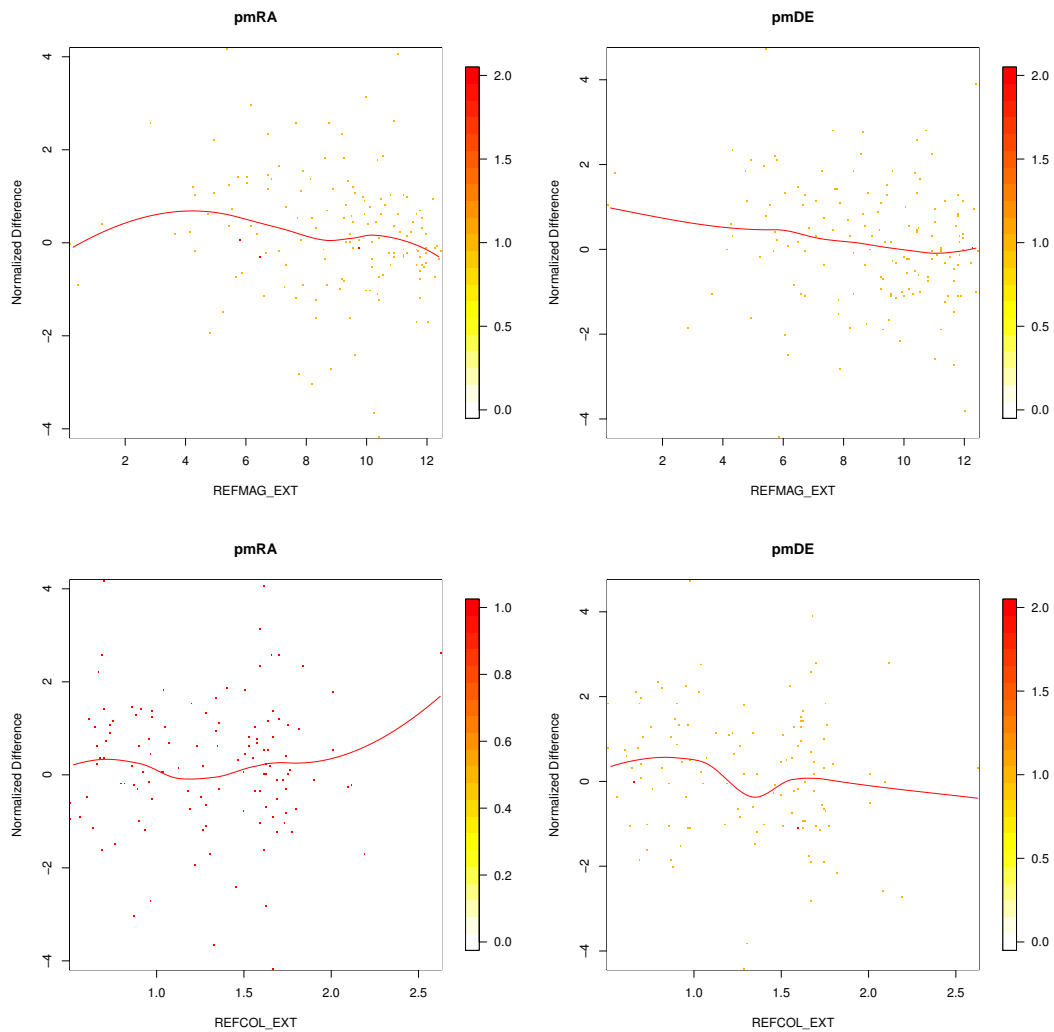


FIGURE 14: Normalized difference density distribution of the HPM Tycho2-Hipparcos proper motions respect to the HPM Tycho2 reference magnitude  $V_T$  (top) and respect to the HPM Tycho2 reference colour  $B - V$  (bottom), for RA proper motion component (left) and DEC component (right). The red curves correspond to automatic fits

**Catalogue: VTS\_030\_004\_simbadHV\_tgas (Simbad HPM)**

Reference magnitude:  $V$ , Reference colour:  $B - V$ ,  $\mu > 0.5$  arsec/yr

**HPM Simbad results on TGAS.SIMU**

XM 835/1094, 23.67 % not found

**No proper motion error components in Simbad (at the time of this version!)**

→ detailed analysis cannot be done

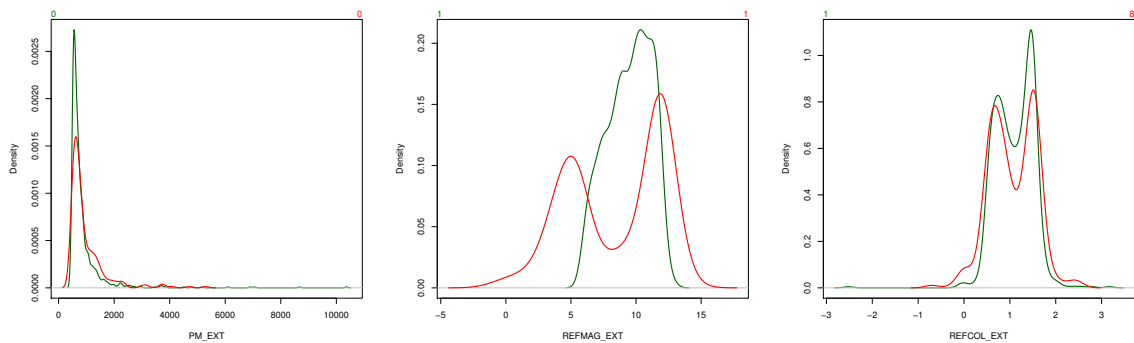


FIGURE 15: From left to right: Normalized density distribution HPM Simbad proper motion, reference magnitude  $V$  and reference colour  $B - V$  of the crossmatched stars Simbad vs TGAS.SIMU (in green) and of the not found stars in TGAS.SIMU (in red)

Since there are no proper motion components, only the residuals and the relation between the HPM Simbad-TGAS.SIMU pm difference and other parameters (magnitude and colour) can be reproduced (Fig. 16).

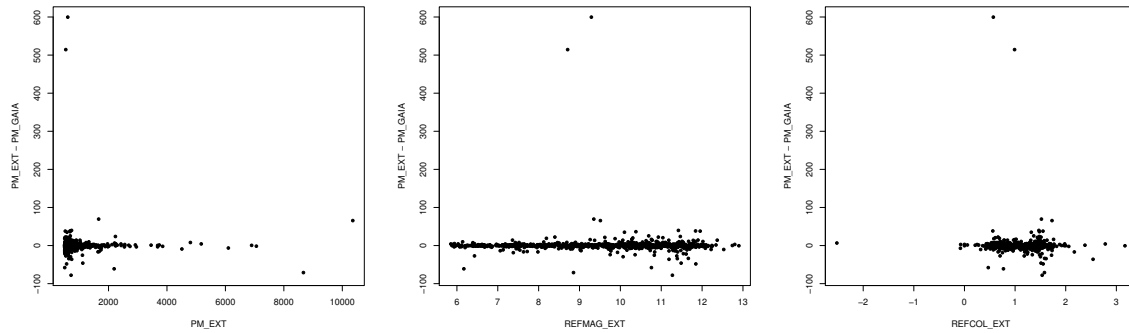


FIGURE 16: Left: residuals of proper motion. Middle: proper motion difference versus HPM Simbad reference magnitude  $V$ . Right: proper motion difference versus HPM Simbad reference colour  $B - V$

### HPM Simbad results on HIPPARCOS

XM 1011/1094, 7.59 % not found

**No proper motion error components in Simbad (at the time of this version!)**

→ detailed analysis cannot be done

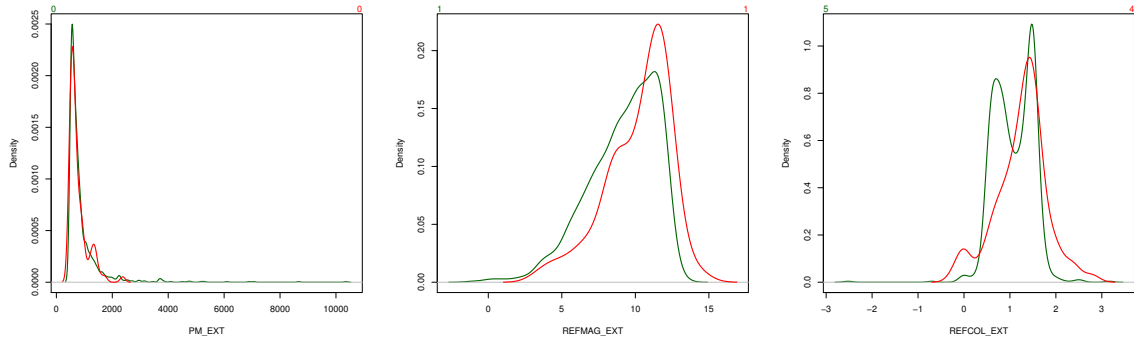


FIGURE 17: From left to right: Normalized density distribution External Catalogue proper motion, reference magnitude  $V$  and reference colour  $B - V$  of the crossmatched stars Simbad vs Hipparcos (in green) and of the not found stars in Hipparcos (in red)

Since there are no proper motion components only the residuals and the relation between the External-Hipparcos pm difference and other parameters (magnitude and colour) can be reproduced (Fig. 18).

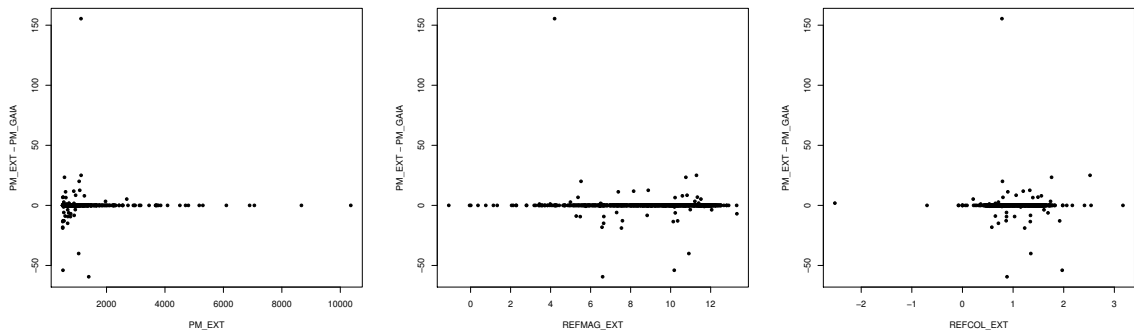


FIGURE 18: Left: residuals of proper motion. Middle: proper motion difference versus HPM Simbad  $V$ . Right: proper motion difference versus HPM Simbad  $B - V$

## 2. Testing High Proper Motion data of Gaia and Hipparcos versus Simbad

**TGAS\_SIMU HPM check:** 161 TGAS\_SIMU stars with  $PM > 1$  arcsec

Stars not found in Simbad: 4/161

TYC 1661-219-1

TYC 6588-404-1

TYC 8754-544-1

TYC 8367-1787-1



The test fails since these 4 stars are considered as HPM in TGAS\_SIMU while they are not in Simbad.

**Hipparcos HPM check:** 217 Hipparcos stars with  $PM > 1$  arcsec  

---

---

Stars not found in Simbad: 1/217  

---

---

TYC 1999-1085-1

The test fails since there is a star considered as HPM in Hipparcos while it is not in Simbad.

## 7 VTS\_030\_005: Parallax zero-point and precision using external distances

### 7.1 Goals

This test intends to check the accuracy and precision of the parallaxes. To do so it globally proves the parallax zero-point error as well as possible bad estimations of the formal errors on this parameter, i.e, it indicates if there is a need of adding a global bias or extra variance on the Gaia parallaxes and errors.

In future updates of this validation code, since the objects of some input catalogues are selected using classification and photometry, it will also be able to test possible correlations, classifications and/or photometric errors.

### 7.2 Input: data selection

To check the zero-point the test needs objects further enough so that their parallax errors (random and systematic) can be considered negligible compared to the Gaia astrometric precision. That way, the results obtained will be due to Gaia properties, not to those of the Catalogues. Hence, the stars selected for this test depend on the astrometric precision of the Gaia catalogue tested.

Since the results presented on this report correspond to TGAS, the selection is done at  $\sigma_{\omega} < 0.1$  mas (including systematic errors).

So far, three input catalogues have been prepared and are being used to check TGAS parallaxes:

- **VTS\_030\_005\_Ngeow\_tgas: Cepheids (Ngeow, 2012)**

The catalogue used is: *Distance to Cepheids using the Wesenheit function (Ngeow, 2012)*<sup>1</sup>

The distance modulus is directly provided in the catalogue. The distance modulus error is computed adding quadratically the dispersion around the Wesenheit function (0.069), the uncertainty on the distance modulus of the LMC used to calibrate this relation (0.043), the I-magnitude error (0.02) and the overall dispersion seen by (Ngeow, 2012) when comparing their distance modulus to other methods (0.2). Note that this last dispersion was not indicated by (Ngeow, 2012) but was needed to get distance modulus consistent with Hipparcos parallaxes.

---

<sup>1</sup>Cepheids catalogue in VizieR: <http://vizier.u-strasbg.fr/viz-bin/VizieR-3?-source=J/ApJ/747/50/table1>

- **VTS\_030\_005\_RRLyr: RR Lyrae (Maintz, 2005)**

The catalogue used is: *Precise positions of RR Lyrae Stars with  $V_{max} > 12.5$  (Maintz, 2005)<sup>2</sup>.*

Since there is no parallax neither distance modulus information in this catalogue, we need to calculate the parallax precision by using the period and photometry. These are the steps applied to select the sample of RR Lyrae for this VTS:

1. Using the extinction map of Schlegel et al. (1998), the reddening  $E(B-V)$  is obtained for each star
2. Taking into account the non-linearity of extinction coefficients and the Spectral Energy Distribution (SED) of a Red Clump star, we determine, by interpolation, the extinction coefficient of the K band ( $A_K$ ), using Kurucz<sup>3</sup> tables. We use the K band since it uses to be the less affected by extinction.
3. All Tycho and Hipparcos stars (named "T" and "H" in the input catalogue) are selected and crossmatched with Tycho2 catalogue (including the Supplement 1)
4. The result is crossmatched with 2MASS catalogue in order to recover infra-red information (JHK bands)
5. In order minimize the extinction effects, the best way to calculate distance modulus is to use the Reddening-Independent Absolute Magnitude, derived using any two colours for a given extinction law:

$$M_{K_{J-K}} = M_K - \frac{A_K}{A_K - A_J} M_{J-K}$$

Using this, the distance modulus will be:

$$\mu = m - M = m_{K_{J-K}} - M_{K_{J-K}}$$

The apparent magnitudes J and K are taken from 2MASS, therefore  $m_{K_{J-K}}$  can be directly obtained. The absolute magnitudes must be taken from period-luminosity relations:

- $M_K$  is calculated using the period-luminosity relation of Muraveva et al. (2015):

$$M_K = (-2.531 \pm 0.038) \log_{10} P - (0.95 \pm 0.01)$$

- Since there are no  $M_J$  period relations only luminosity dependent in the literature, we make use of the theoretical period-abundance-luminosity relations of Catelan (2004) (equations 4 and 6 of his article):

$$M_J^* = -0.1409 - 1.7734 \log_{10} P + 0.1899 \log_{10} Z$$

<sup>2</sup>RR Lyrae catalogue in VizierR: <http://vizier.u-strasbg.fr/viz-bin/VizieR-3?-source=J/A%2bA/442/381/table1>

<sup>3</sup>Kurucz Extinction tables:<ftp://ftp.stsci.edu/cdbs/grid/k93models/>

$$M_K^* = -0.5968 - 2.3529 \log_{10} P + 0.1746 \log_{10} Z$$

Taking an abundance of  $Z = 0.001$  for all the stars we can see that the factor depending on  $Z$  will be almost negligible when doing the difference of  $M_J^* - M_K^*$ .

\* These relations are in GLASS system. Transformations given by Carpenter (2001) (Appendix A of his article) were used to get the results in 2MASS system:

$$(J - K_s)_{2MASS} = (0.972 \pm 0.006)(J - K)_{GLASS} + (-0.011 \pm 0.005)$$

–  $M_{K_{J-K}}$  is finally obtained using the above calculated  $M_K$  and  $(M_{J-K_s})_{2MASS}$

- **VTS\_030\_005\_rave\_tgas: RAVE (Kordopatis et al., 2013)**

The catalogue used is: *RAVE 4th data release (Kordopatis et al., 2013)*<sup>4</sup>

This catalogue has parallaxes and distance modulus available. Hence only a cross-match with Tycho2 and a check of possible duplicates or wrong XM, was needed. The catalogue contains duplicated values per RAVE ID and duplicated RAVE ID per Tycho2 ID (due to rounding issues of the coordinates in RAVE ID construction), in those cases a weighted mean was applied. Distance modulus estimate was used instead of the parallax one as the RAVE distances are derived from magnitudes.

Other catalogues will be used in future releases:

- Spectroscopic surveys: PASTEL, APOGEE, GES
- Asterosismology: COROT KEPLER
- Eclipsing Binaries
- dSph stars

### 7.3 Process

For each of the catalogues:

- XM by identifier with TGAS

<sup>4</sup>RAVE catalogue in VizierR: <http://vizier.u-strasbg.fr/viz-bin/VizieR-3?-source=III/272/ravedr4>

- Calculates parallaxes from distance modulus of the external catalogue:

$$\varpi_{EXT} = \exp \left\{ -\frac{\mu_{EXT} + 5}{5} \ln(10) \right\} 1000$$

$$\sigma_{\varpi_{EXT}} = \frac{\sigma_{\mu_{EXT}}}{5} \ln(10) \varpi_{EXT}$$

- Plot external versus Gaia parallaxes
- Selects stars with external parallax error  $\sigma_{\varpi_{EXT}}$  lower than the parallax precision needed (0.1 for TGAS)
- VTS over this data selection:
  - Statistical analysis as detailed in Annex A.1 for  $\varpi_{EXT} - \varpi_{Gaia}$ . No correlations are checked.
  - Remove outliers out of  $5 \sigma$
  - Perform a Maximum Likelihood Estimation (MLE) to look for bias and extra variance (see section 7.4).
  - The confidence interval for each of the fitted parameters (bias and extra variance) is computed at a level of 99%

The final flag for the VTS (for each catalogue) will depend on the MLE analysis:

- PASS: 0 falls within the confidence intervals of the MLE estimate of the bias and the extra variance.
- FAIL: 0 is not within the confidence interval.

## 7.4 Statistical Techniques

### MLE estimate of Bias and extravarience

The Maximum Likelihood Estimation (MLE) allows to estimate the parameters  $\Theta$  that maximizes:

$$P(\Theta|x_1, \dots, x_n) = \frac{f(x_1, \dots, x_n|\Theta)P(\Theta)}{P(x_1, \dots, x_n)}$$

Here we want to derive the bias and extra variance parameters and will use a flat prior on those.

Therefore, the function to maximize becomes:

$$P(bias, extravar | \mu, \sigma_\mu, \varpi, \sigma_\varpi) \propto \int_{\bar{\varpi}} \mathcal{N} \left[ \mu, -5 \log_{10} \left( \frac{\bar{\varpi}}{1000} \right) - 5, \sigma_\mu \right] \mathcal{N} \left[ \varpi + bias, \bar{\varpi}, \sqrt{\sigma_\varpi^2 + extravar} \right] d\bar{\varpi}$$

where  $\bar{\varpi}$  is the range of possible parallaxes within the  $5\sigma$  confidence interval:

$$\bar{\varpi} \in [Max[\varpi - 5\sigma, 0], \varpi + 5\sigma]$$

The computation is done using discretized steps of  $\min(\sigma_\varpi/10, \sigma_{\varpi_{EXT}})$ .

R mle2 is used to derive the corresponding MLE.

## 7.5 Test Results

### 7.5.1 Summary

TEST CASE	TGAS_SIMU	HIP
VTS_030_005_Ngeow_tgas	PASSED	PASSED
VTS_030_005_RRLyrae	FAILED	PASSED
VTS_030_005_rave_tgas	FAILED	PASSED

### 7.5.2 Detailed results for Cepheids

The following tables and plots show the results obtained with TGAS\_SIMU and with Hipparcos, for the Cepheids.

**Cepheids ZP parallax results on TGAS\_SIMU**

XM 248/292, 15 % not found, 0 % outliers

<b>MLE test: performed on 67/248 stars</b>					
<b>Property</b>	<b>Estimate</b>	<b>p-value Lim.</b>	<b>z value</b>	<b>PASS/FAIL</b>	<b>Pr(z)</b>
Bias	$-0.128 \pm 0.052$	0.01	-2.496	PASS	0.0125
Extra Variance	$0.089 \pm 0.026$	0.01	3.422	-	0.001

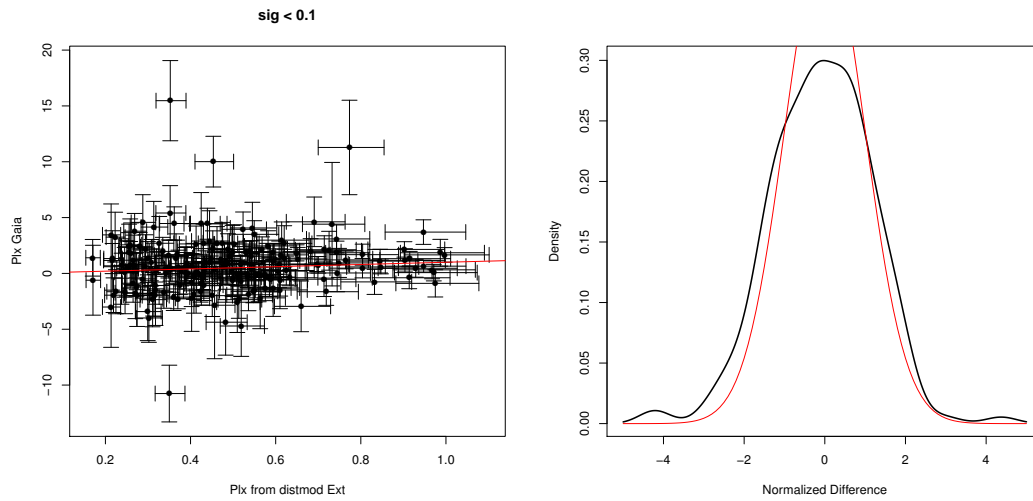


FIGURE 19: Left: Comparison between TGAS\_SIMU (y axis) and Cepheids (x axis) parallaxes (in mas); outliers out of  $5\sigma$  are not considered. Right: Normalized parallax density distribution (in red the expected normal distribution)

**Cepheids ZP parallax results on HIPPARCOS**

XM 248/292, 15 % not found, 0 % outliers

<b>MLE test: performed on 190/248 stars</b>					
<b>Property</b>	<b>Estimate</b>	<b>p-value Lim.</b>	<b>z value</b>	<b>PASS/FAIL</b>	<b>Pr(z)</b>
Bias	$-0.086 \pm 0.108$	0.01	-0.801	PASS	0.423
Extra Variance	$0.438 \pm 0.225$	0.01	1.949	-	0.051

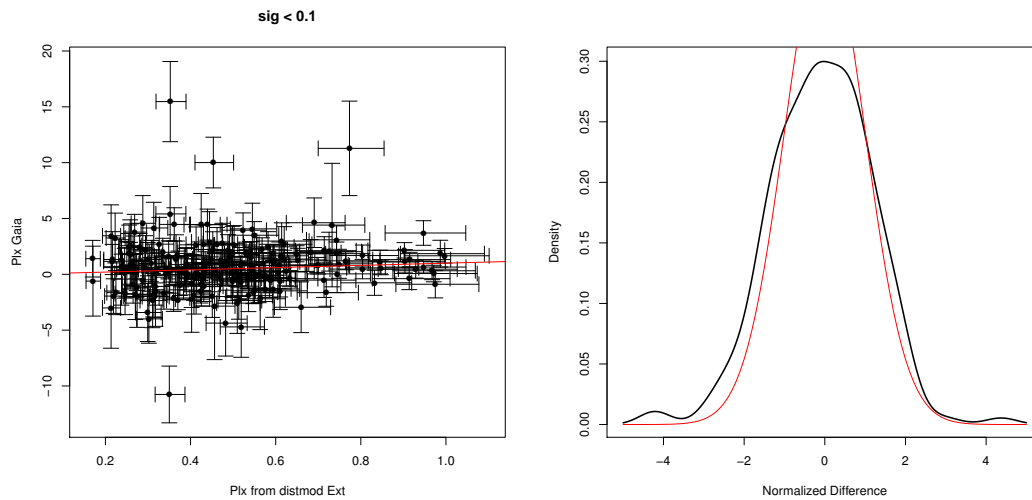


FIGURE 20: Left: Comparison between Hipparcos (y axis) and Cepheids (x axis) parallaxes (in mas); outliers out of  $5\sigma$  are not considered. Right: Normalized parallax density distribution (in red the expected normal distribution)



### 7.5.3 Detailed results for RR Lyrae

The following tables and plots show the results obtained with TGAS\_SIMU and with Hipparcos, for the RR Lyrae.

#### RR Lyrae ZP parallax results on TGAS\_SIMU XM 136/236, 42.4 % not found; 0% outliers

MLE test: performed on 123/136 stars					
Property	Estimate	p-value Lim.	z value	PASS/FAIL	Pr(z)
Bias	$-0.337 \pm 0.046$	0.01	-7.244	FAIL	4.34e-13
Extra Variance	$0.092 \pm 0.036$	0.01	2.593	-	9.51e-03

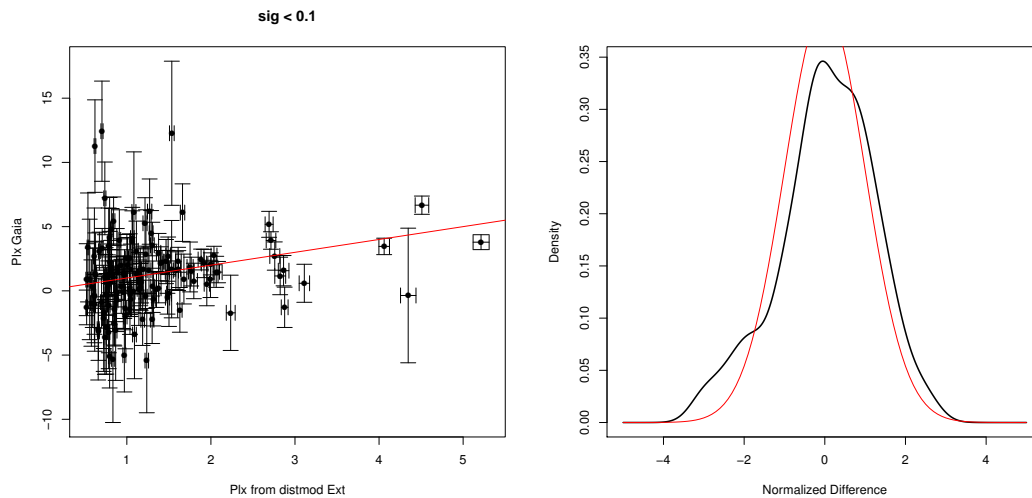


FIGURE 21: Left: Comparison between TGAS\_SIMU (y axis) and RR Lyrae (x axis) parallaxes (in mas); outliers out of  $5\sigma$  are not considered. Right: Normalized parallax density distribution (in red the expected normal distribution)

**RR Lyrae ZP parallax results on HIPPARCOS**

XM 136/236, 42.4 % not found, 0% outliers

<b>MLE test: performed on 134/136 stars</b>					
<b>Property</b>	<b>Estimate</b>	<b>p-value Lim.</b>	<b>z value</b>	<b>PASS/FAIL</b>	<b>Pr(z)</b>
Bias	$0.082 \pm 0.169$	0.01	0.486	PASS	0.627
Extra Variance	$0.718 \pm 0.380$	0.01	1.887	-	0.059

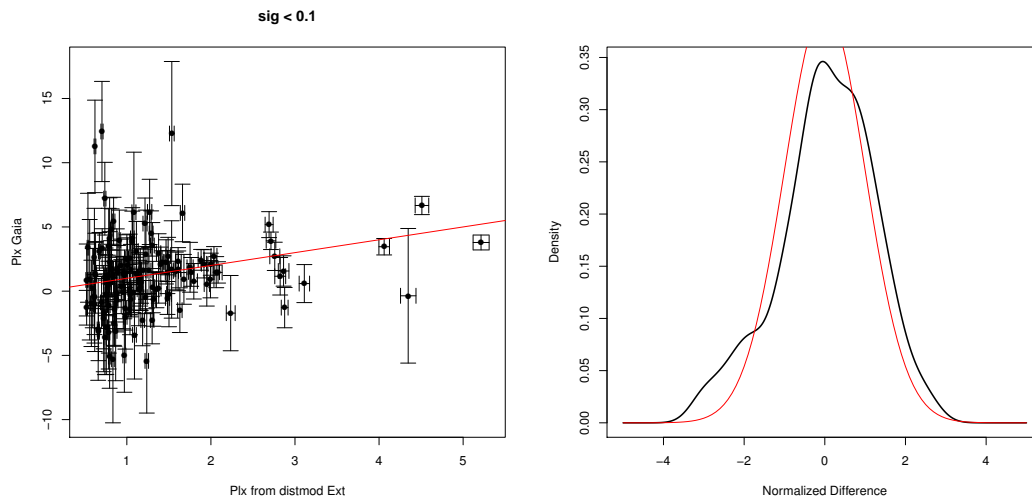


FIGURE 22: Left: Comparison between Hipparcos (y axis) and RR Lyrae (x axis) parallaxes (in mas); outliers out of  $5\sigma$  are not considered. Right: Normalized parallax density distribution (in red the expected normal distribution)

### 7.5.4 Detailed results for RAVE

For RAVE, a large number of outliers is found. Close inspection on Hipparcos stars seems to indicate mis-classification of dwarfs as giants... Those are removed by the outlier criteria on the Hipparcos test. The RAVE catalogue results will therefore have to be treated with caution.

The following tables and plots show the results obtained with TGAS\_SIMU and with Hipparcos, for RAVE.

**RAVE ZP parallax results on TGAS\_SIMU**  
XM 6245/6872, 9.04 % not found, 3.27% outliers

MLE test: performed on 6030/6872 stars					
Property	Estimate	p-value Lim.	z value	PASS/FAIL	Pr(z)
Bias	$-0.219 \pm 0.006$	0.01	-38.807	FAIL	0.0
Extra Variance	$0.049 \pm 0.003$	0.01	16.449	-	8.55e-61

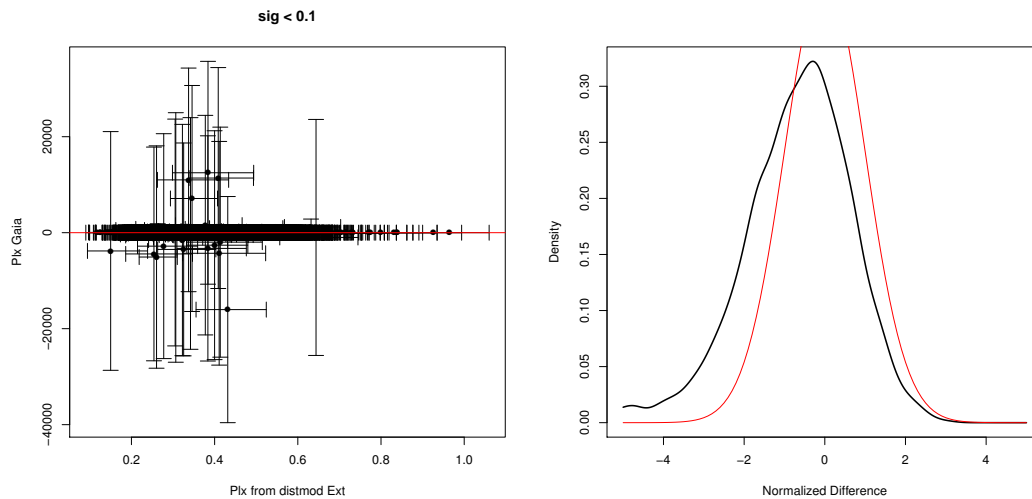


FIGURE 23: Left: Comparison between TGAS\_SIMU (y axis) and RAVE (x axis) parallaxes (in mas); outliers out of  $5\sigma$  are not considered. Right: Normalized parallax density distribution (in red the expected normal distribution)

**RAVE ZP parallax results on HIPPARCOS**

XM 25/6872, 99.6 % not found, 24% outliers

**MLE test: performed on 19/25 stars**

Property	Estimate	p-value Lim.	z value	PASS/FAIL	Pr(z)
Bias	$-0.660 \pm 0.375$	0.01	-1.761	PASS	0.078
Extra Variance	$0.148 \pm 0.564$	0.01	0.263	-	0.792

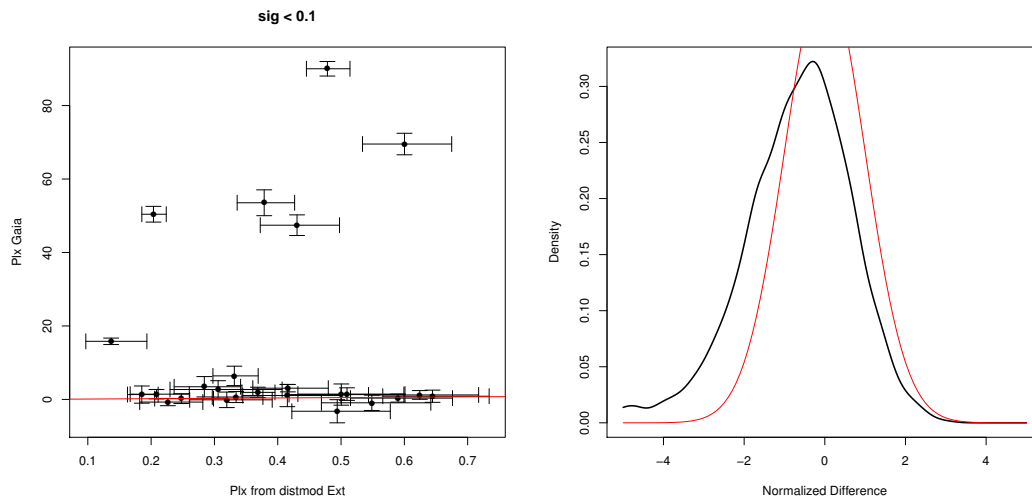


FIGURE 24: Left: Comparison between Hipparcos (y axis) and RAVE (x axis) parallaxes (in mas); outliers out of  $5\sigma$  are not considered. Right: Normalized parallax density distribution (in red the expected normal distribution)

## 8 VTS\_030\_006: Parallax zero-point and precision using very distant stars

### 8.1 Goals

This VTS also aims to check the accuracy (zero-point) and precision (whether the associated standard errors are correctly estimated) of the parallaxes, but using stars known to be at very large distances, which allows to assume that their Gaia parallaxes should be zero.

In future updates of this VTS, it will also test the classification and photometry of those selected stars being well known.

Notice that the number of stars useful for this test will diminish with the increase of astrometric accuracy at each Gaia release.

### 8.2 Input: data selection

To check the zero-point the test needs objects further enough so that their parallax errors (random and systematic) can be considered negligible compared to the Gaia astrometric precision. That way, the results obtained will be due to Gaia properties, not to those of the Catalogues. Hence, the stars selected for this test depend on the astrometric precision of the Gaia catalogue tested.

Since the results presented on this report correspond to TGAS, the selection is done at  $\varpi < 0.1$  mas (including systematic errors).

So far only stars from the Large and Small Magellanic Clouds (LMC and SMC) are being used to check TGAS parallaxes in this test. Three input catalogues have been prepared using these stars: LMC, SMC, all LMC and SMC stars we are sure they are present in the Hipparcos catalogue:

- **VTS\_030\_006\_LMC\_tgas: LMC**

See detailed description of the construction of the LMC catalogue in Annex B. Cross-match with Tycho2 is done with a search radius of 1". Duplicated sources are removed. The catalogue contains 250 Tycho stars.

- **VTS\_030\_006\_SMC\_tgas: SMC**

See detailed description of the construction of the SMC catalogue in Annex B. 2 confirmed Hipparcos SMC stars (see next item) are added. Cross-match with

Tycho2 is done with a search radius of 1". Duplicated sources are removed. The catalogue contains 79 Tycho stars.

- **VTS\_030\_006\_LMCSMC\_hipconfirmed: LMC + SMC in Hipparcos**

Hipparcos stars selected as LMC/SMC members in the Hipparcos input catalogue have been manually checked to be LMC/SMC members on the Hipparcos astrometry, radial velocity and spectral type indication from Simbad. The catalogue contains 46 stars.

Other catalogues will be used in future releases:

- Spectroscopic surveys: APOGEE, GES
- Cepheids
- RR Lyrae
- Eclipsing Binaries
- dSph stars

### 8.3 Process

For each catalogue:

- A XM is done with the database (TGAS) by Tycho2 identifier
- The Gaia parallax distribution is checked
- The statistical analysis defined in Annex A.1 is applied for  $\varpi_{Gaia}$ , to verify that Gaia parallaxes are not significantly different from 0, and that the variance of the distribution of the normalized residuals is not significantly different from 1 (at 95% level of confidence).

The test will fail or pass according to:

- FAIL: one or both conditions (parallax not significantly different from 0, and variance not significantly different from 1) is not accomplished
- PASS: both conditions result with what is expected

## 8.4 Statistical Techniques

All statistical techniques applied here are defined in Annex A.1.

## 8.5 Test Results

### 8.5.1 Summary

TEST CASE	TGAS_SIMU	HIP
VTS_030_006_LMC_tgas	FAILED	PASSED
VTS_030_006_SMC_tgas	FAILED	PASSED
VTS_030_006_LMCSMC_hipconfirmed	FAILED	PASSED

### 8.5.2 Detailed results for LMC

The following tables and plots show the results obtained with TGAS\_SIMU and with Hipparcos, for LMC stars.

#### LMC ZP parallax results on TGAS\_SIMU

XM 246/250, 1.6 % not found

Property	Obtained Value	Confidence Interval	Expected Value	PASS/FAIL	Related test	
					Test	p-value
Outliers	28.86% (71)	$5\sigma$	5.73e-05	-	-	-
Mean	$0.393 \pm 0.019$	$3\sigma$	-	FAIL	t test	1.77e-28
Standard Dev.	1.415	-	-	-	-	-
Median Abs. Dev.	1.197	-	-	-	-	-
Distr. Skewness	0.181	-	0	-	agostino	0.314
Normality (KS test)	p-v: 0	p-v: 0.05	-	FAIL	-	-

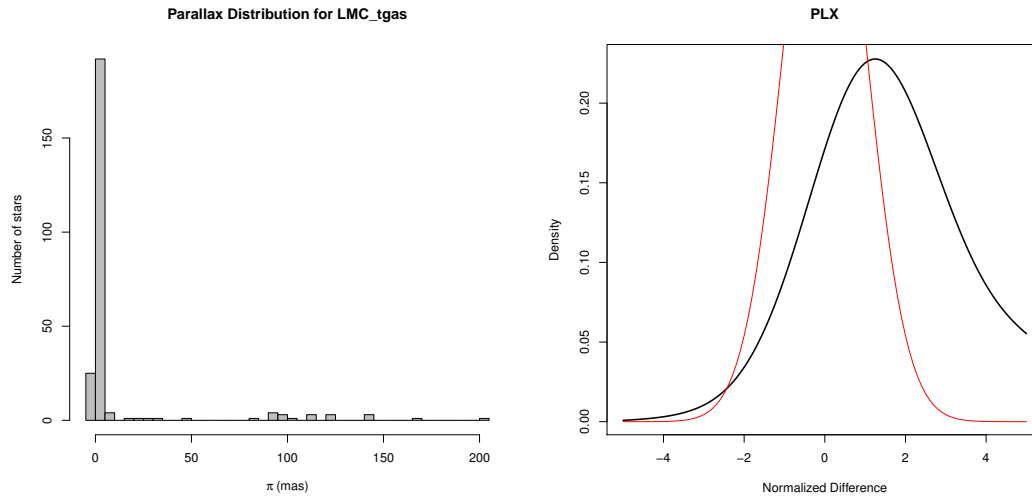


FIGURE 25: Left: TGAS\_SIMU parallax distribution for the LMC stars. Right: Normalized LMC TGAS\_SIMU parallax density distribution (in red the expected normal distribution)

### LMC ZP parallax results on HIPPARCOS

XM 39/250, 84.4 % not found

Property	Obtained Value	Confidence Interval	Expected Value	PASS/FAIL	Related test	
					Test	p-value
Outliers	5.13% (2)	$5\sigma$	$5.73e-05$	-	-	-
Mean	$0.340 \pm 0.213$	$3\sigma$	-	PASS	t test	0.277
Standard Dev.	1.257	-	-	-	-	-
Median Abs. Dev.	1.205	-	-	-	-	-
Distr. Skewness	0.279	-	0	-	agostino	0.432
Normality (KS test)	p-v: 0.236	p-v: 0.05	-	PASS	-	-



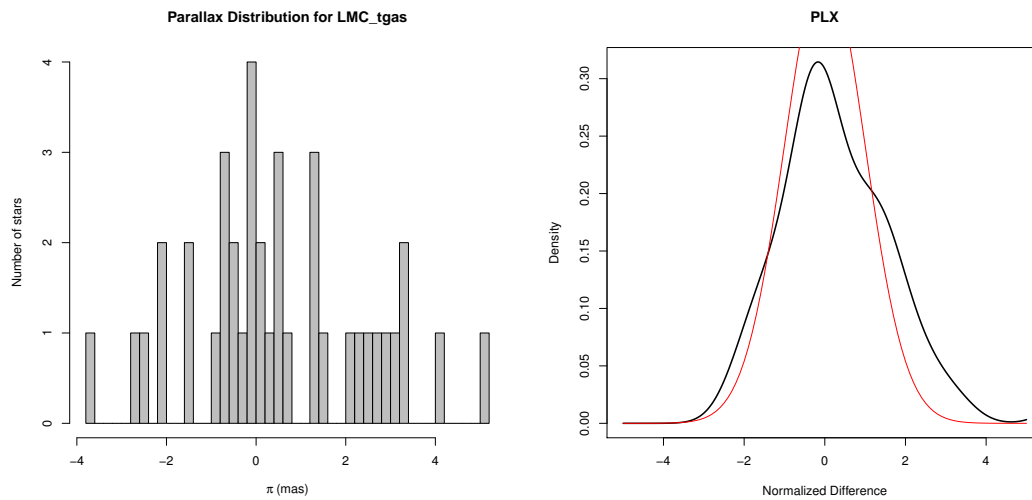


FIGURE 26: Left: Hipparcos parallax distribution for the LMC stars. Right: Normalized LMC Hipparcos parallax density distribution (in red the expected normal distribution)

### 8.5.3 Detailed results for SMC

The following tables and plots show the results obtained with TGAS\_SIMU and with Hipparcos, for SMC stars.

**SMC ZP parallax results on TGAS\_SIMU**  
XM 70/79, 11.39 % not found

Property	Obtained Value	Confidence Interval	Expected Value	PASS/FAIL	Related test	
					Test	p-value
Outliers	47.14% (33)	$5\sigma$	$5.73e-05$	-	-	-
Mean	$0.391 \pm 0.025$	$3\sigma$	-	FAIL	t test	$3.73e-13$
Standard Dev.	1.367	-	-	-	-	-
Median Abs. Dev.	1.407	-	-	-	-	-
Distr. Skewness	-0.102	-	0	-	agostino	0.771
Normality (KS test)	p-v: 0	p-v: 0.05	-	FAIL	-	-

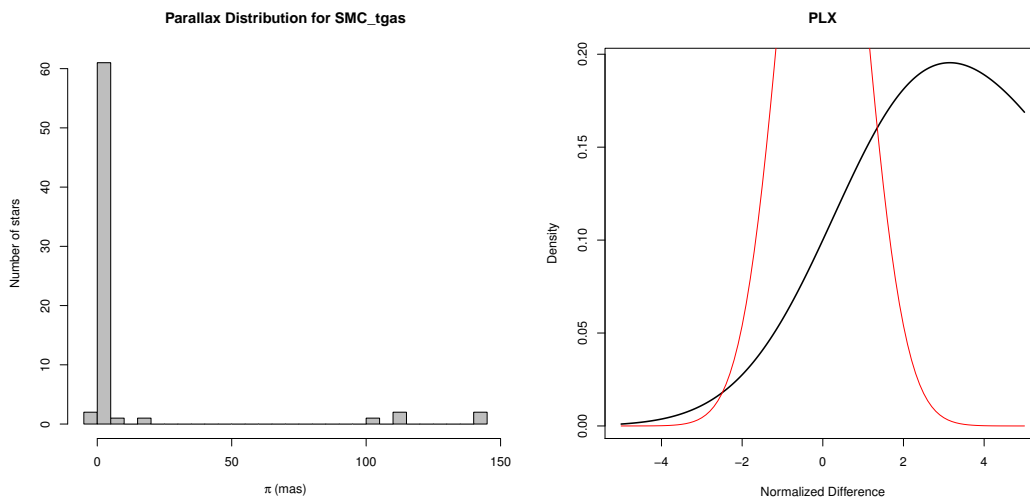


FIGURE 27: Left: TGAS\_SIMU parallax distribution for the SMC stars. Right: Normalized SMC TGAS\_SIMU parallax density distribution (in red the expected normal distribution)

**SMC ZP parallax results on HIPPARCOS**  
XM 12/79, 84.81 % not found

Property	Obtained Value	Confidence Interval	Expected Value	PASS/FAIL	Related test	
					Test	p-value
Outliers	0% (0)	$5\sigma$	5.73e-05	-	-	-
Mean	$-0.619 \pm 0.479$	$3\sigma$	-	PASS	t test	0.226
Standard Dev.	0.839	-	-	-	-	-
Median Abs. Dev.	0.780	-	-	-	-	-
Distr. Skewness	-0.652	-	0	-	agostino	0.228
Normality (KS test)	p-v: 0.519	p-v: 0.05	-	PASS	-	-

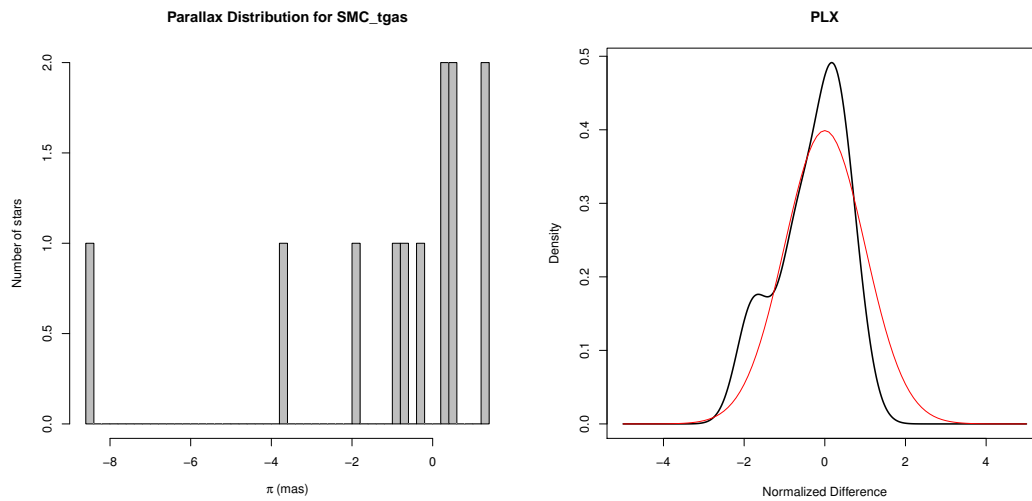


FIGURE 28: Left: Hipparcos parallax distribution for the SMC stars. Right: Normalized SMC Hipparcos parallax density distribution (in red the expected normal distribution)

#### **8.5.4 Detailed results for LMC + SMC in Hipparcos**

The following tables and plots show the results obtained with TGAS\_SIMU and with Hipparcos, for the ensemble of LMC and SMC checked to be in Hipparcos.

**LMC+SMC in Hipparcos ZP parallax results on TGAS\_SIMU**

XM 45/46, 2.17 % not found

Property	Obtained Value	Confidence Interval	Expected Value	PASS/FAIL	Related test	
					Test	p-value
Outliers	75.56% (34)	$5\sigma$	5.73e-05	-	-	-
Mean	$0.381 \pm 0.033$	$3\sigma$	-	FAIL	t test	3.75e-07
Standard Dev.	1.058	-	-	-	-	-
Median Abs. Dev.	1.019	-	-	-	-	-
Distr. Skewness	-0.487	-	0	-	agostino	0.375
Normality (KS test)	p-v: 0	p-v: 0.05	-	FAIL	-	-

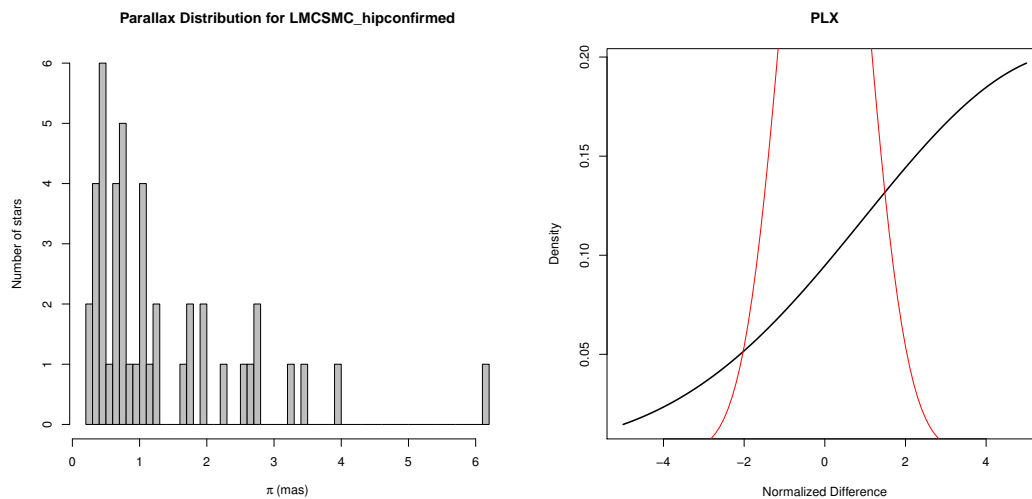


FIGURE 29: Left: TGAS\_SIMU parallax distribution for the confirmed Hipparcos LMC and SMC stars. Right: Normalized LMC+SMC TGAS\_SIMU parallax density distribution (in red the expected normal distribution)

**LMC+SMC in Hipparcos ZP parallax results on HIPPARCOS**

XM 46/46, 0% not found

Property	Obtained Value	Confidence Interval	Expected Value	PASS/FAIL	Related test	
					Test	p-value
Outliers	0% (0)	$5\sigma$	5.73e-05	-	-	-
Mean	$-0.042 \pm 0.204$	$3\sigma$	-	PASS	t test	0.897
Standard Dev.	1.065	-	-	-	-	-
Median Abs. Dev.	1.031	-	-	-	-	-
Distr. Skewness	0.036	-	0	-	agostino	0.912
Normality (KS test)	p-v: 0.986	p-v: 0.05	-	PASS	-	-

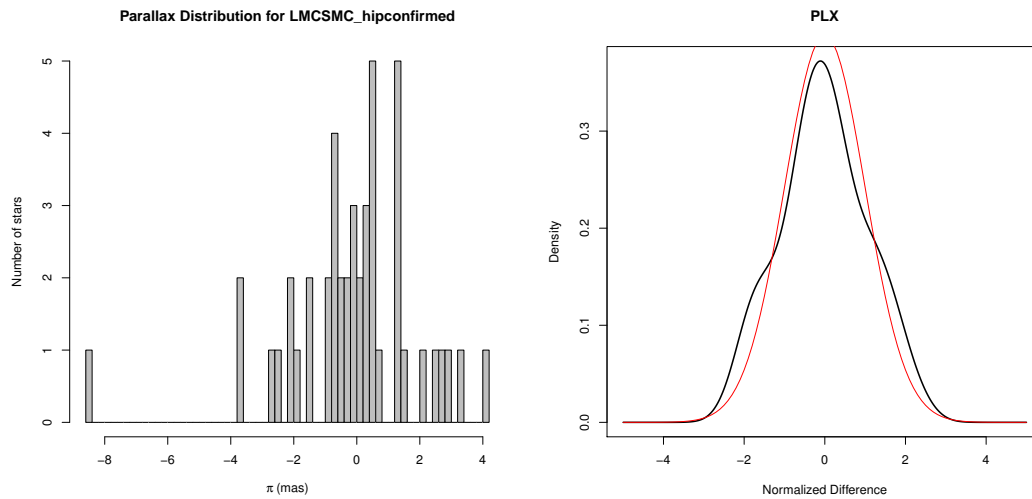


FIGURE 30: Left: Hipparcos parallax distribution for the confirmed Hipparcos LMC and SMC stars. Right: Normalized LMC+SMC Hipparcos parallax density distribution (in red the expected normal distribution)

## 9 VTS\_060\_003: Check QSO parallaxes and proper motions

### 9.1 Goals

The test intends to check the absence of parallax and proper motions for known quasars or AGN detected by Gaia from a cross-identification.

### 9.2 Input: data selection

Known QSOs or AGNs have been cross-matched with Tycho2, leading to 4 sources (including one HIP: 3C273).

### 9.3 Process

The 4 Tycho2 IDs are XM with the Gaia catalogue. A  $\chi^2$  test is done to check that their parallax and proper motion are consistent with zero.

### 9.4 Statistical Techniques

The  $\chi^2$  technique applied here is defined in Annex A.2.

### 9.5 Test Results

#### 9.5.1 Summary

TEST CASE	TGAS_SIMU	SUMMARY	HIP	SUMMARY
VTS_060_003_TGAS_AGN	PASSED	No AGN present in TGAS	PASSED	1 AGN found

#### 9.5.2 Detailed results

VTS\_060\_003\_tgas\_agn : PASSED : No AGN present in TGAS (it was intentionally removed from the simulation)

VTS\_060\_003\_tgas\_agn : PASSED: 1 AGN found in Hipparcos

## 10 Summary of the Tests Results

### 10.1 Summary Table

The following table summarizes the completion status of each test case run in local (WP944 version 18.1) on the TGAS\_SIMU and Hipparcos catalogue.

TEST CASE ID	TGAS_SIMU	HIPPARCOS
VTS_010_001_Tycho2Duplicate	PASSED	NA
VTS_010_003_hip67626	PASSED	NA
VTS_030_001_VLBI_tgas_CHI2	FAILED	FAILED
VTS_030_001_VLBI_tgas_PLX	PASSED	PASSED
VTS_030_001_VLBI_tgas_PM	FAILED	FAILED
VTS_030_003_goodHip_PLX	FAILED	NA
VTS_030_003_goodHip_PM	FAILED	NA
VTS_030_003_badchi2HipTycho2	FAILED	NA
VTS_030_004_tycho2_HV	FAILED	FAILED
VTS_030_004_simbadHV_tgas	WARNING	WARNING
VTS_030_004_TGASHPM	FAILED	FAILED
VTS_030_005_Ngeow_tgas	PASSED	PASSED
VTS_030_005_RRLyrae	FAILED	PASSED
VTS_030_005_rave_tgas	FAILED	PASSED
VTS_030_006_LMCSMC_hipconfirmed	FAILED	PASSED
VTS_030_006_SMC_tgas	FAILED	PASSED
VTS_030_006_LMC_tgas	FAILED	PASSED
VTS_060_003_TGAS_AGN	PASSED	PASSED

### 10.2 Overall Assessment

No specific problem to report.

### 10.3 Impact of Test Environment

None.

### 10.4 Recommended Improvements

Number of outliers are currently manually checked. Automatic test should be added.



## A ANNEX A: Global Statistical Techniques

This section describes the statistical analysis applied in several VTS. That is: outliers detection, means, standard and absolute deviations, distribution skewness, correlations with specific parameters,..., and some known statistical tests such as Student's t.test, D'Agostino's test, Kolmogorov-Smirnov's test or  $\chi^2$  test. The goal is to check if what is obtained corresponds, within a given confidence interval, to what is expected. Some of these statistics, as specified below, will determine whether the tests passes or fails, while the others provide extra information. The parameters to which these analysis are applied are detailed on each VTS.

### A.1 Main Statistical Analysis

The normalized difference of the observed values of a given parameter  $\theta$  coming from two different catalogues (here the External and Gaia) must be a normal distribution with zero mean and unit dispersion:

$$\Delta\theta_{norm} = \frac{\theta_{EXT} - \theta_{GAIA}}{(\sigma_{EXT}^2 + \sigma_{GAIA}^2)^{1/2}}$$

All the following statistic analysis is applied to this resulting distribution  $\Delta\theta_{norm}$ , from now on just called  $P(\theta)$  for simplicity.

- **Outliers**

The cumulative distribution function (c.d.f) computes the probability that a normally distribution random number will be less than that number:

$$F(\theta) = P(\Theta \leq \theta)$$

The outliers are checked out of  $5\sigma$ . Therefore the probability of finding values outside this limit is given by the difference of c.d.f at  $\theta = 5$  and  $\theta = -5$ , and subtracted this from 1:

$$ExpectedValue = [1 - [P(\Theta \leq 5) - P(\Theta \leq -5)]] * 100$$

Since our distributions are normal, the expected value will always be 5.733031e-05.

This value is compared to the real number of outliers found out of  $5\sigma$ . The sample is cleaned at  $5\sigma$  for the following analysis.

- **Mean**

This is calculated as the weighted mean of  $\theta_{EXT} - \theta_{GAIA}$  on the sample already cleaned of outliers. The result is accepted up to  $3\sigma$ .

$$\bar{\theta} < 3 \sigma_{\bar{\theta}} \rightarrow \text{Test of mean: PASS}$$

To complete the information of the mean a Student's t-test is applied on  $\Delta\theta_{norm}$ . This statistical test allows to determine how different the mean of the distribution of  $\theta$  differs from zero. However this test is not used to determine if the test over the mean passes or not as it is less direct than the weighted mean; however the weighted mean needs the errors to be correctly estimated, which is tested afterwards. The two tests are therefore quite complementary.

- **Standard Deviation**

It gives the amount of dispersion of the set of data values:

$$sd = \sqrt{\frac{1}{N} \sum_{i=1}^N (\theta_i - \bar{\theta})^2}$$

- **Median Absolute Deviation**

This is the median of the set comprising the absolute values of the differences between the median (M) and each data point ( $\theta$ ), in other words, a robust measure of the central tendency (not sensitive to the presence of outliers). A factor  $b$  needs to be added to consider the *mad* as a consistent estimator for the estimation of the standard deviation. For normally distributed data, as is the case here,  $b = 1.4826$ :

$$mad = b M_i(|\theta_i - M_j(\theta_j)|)$$

This parameter is calculated with the data sample cleaned of outliers.

- **Distribution skewness**

It measures the degree of asymmetry of a probability distribution. If the left tail is more pronounced than the right tail, the function is said to have negative skewness. If the reverse is true, it has positive skewness. If the two are equal, it has zero skewness. For a normal distribution, since it is a symmetric distribution, the skewness is always zero.

A D'Agostino's test for skewness is also performed (as long as the length of the sample be in the interval [8,46340], limit driven by the R implementation), in order to detect non-normality caused by asymmetry.

- **Kolmogorov-Smirnov test**

A two-sample KS test is applied in order to compare the empirical cumulative distribution function of the sample and of a normal c.d.f. The limit of the confidence interval is fixed at a p-value of 0.05.

- **Correlations**

The correlation of the given parameter  $\theta$  versus other parameters  $c_i$  can be checked using a test for Correlation Between Paired Samples. It estimates the association between paired samples and compute a test of the value being zero, i.e., no correlation.

If there are correlations to check out, the sample is first cleaned of outliers and non-valid numbers. Then, if there are more than 50 stars the 2D histogram of  $c_i$  versus  $\theta$  is done and a polynomial regression is adjusted; else a simple scatter plot of  $c_i$  versus  $\theta$  is done. For all cases, the correlation test is applied: the p-value of the test obtained for each pair of parameters is compared to a p-value limit of 0.05.

## A.2 Parallax and proper motion consistency: $\chi^2$ test

In order to validate the astrometric solution (coherence in proper motion and parallax) between the Gaia catalogue and another known external catalogue, we consider that our variables follow a Normal Density Distribution.

Each object considered for this test has three correlated variables (without taking into account the position  $[\alpha, \delta]$ ): the two components of the proper motion ( $\mu_\alpha, \mu_\delta$ ) and the parallax ( $\varpi$ ), with their respective variances. This is a 3-dimensional Multivariate Density Distribution case:

$$\phi(x_1, \dots, x_n) = \frac{1}{\sqrt{(2\pi)^n |\Sigma|}} \exp \left\{ -\frac{1}{2} (\mathbf{x} - \mathbf{x}_0)^T \Sigma^{-1} (\mathbf{x} - \mathbf{x}_0) \right\} \quad -\infty < x_i < \infty, \quad i = 1, 2, \dots, n \quad (2)$$

where  $n$  is the dimension of the distribution and  $\Sigma$  the covariance matrix.

Note that the term:

$$(\mathbf{x} - \mathbf{x}_0)^T \Sigma^{-1} (\mathbf{x} - \mathbf{x}_0) = \mathbf{X}^T \Sigma^{-1} \mathbf{X} \quad (3)$$

corresponds to the quadratic form of the random variables  $x_i$ , so that it corresponds to a chi-square variate:

$$\mathbf{X}^T \Sigma^{-1} \mathbf{X} = \chi^2 \quad (4)$$

Therefore we just need to calculate this  $\chi^2$  value to validate if our components (the astrometric solution) follow or not a normal distribution. However, what we want to validate is not whether each of these variables is coherent compared to an expected value:

$$\begin{aligned} |\mu_\alpha - \mu_{\alpha_0}| &= 0 \\ |\mu_\delta - \mu_{\delta_0}| &= 0 \\ |\varpi - \varpi_0| &= 0 \end{aligned}$$

, but whether they are coherent compared to another catalogue (G: Gaia, E: External):

$$\begin{aligned} |\mu_{\alpha_G} - \mu_{\alpha_E}| &= 0 \\ |\mu_{\delta_G} - \mu_{\delta_E}| &= 0 \\ |\varpi_G - \varpi_E| &= 0 \end{aligned}$$

That means that each catalogue has its own density distribution:  $\phi_G(\mu_{\alpha_G}, \mu_{\delta_G}, \varpi_G)$  for the Gaia catalogue, and  $\phi_E(\mu_{\alpha_E}, \mu_{\delta_E}, \varpi_E)$  for the External catalogue, but the expected value ( $x_0$ ) for each of the variables is the same in both:

$$\begin{aligned} \phi_G(\mu_{\alpha_G}, \mu_{\delta_G}, \varpi_G) &= \frac{1}{\sqrt{(2\pi)^3 |\Sigma_G|}} \exp \left\{ -\frac{1}{2} (\mathbf{x}_G - \mathbf{x}_0)^T \Sigma_G^{-1} (\mathbf{x}_G - \mathbf{x}_0) \right\} \\ \phi_E(\mu_{\alpha_E}, \mu_{\delta_E}, \varpi_E) &= \frac{1}{\sqrt{(2\pi)^3 |\Sigma_E|}} \exp \left\{ -\frac{1}{2} (\mathbf{x}_E - \mathbf{x}_0)^T \Sigma_E^{-1} (\mathbf{x}_E - \mathbf{x}_0) \right\} \\ \text{with } \mathbf{x}_G &= (\mu_{\alpha_G}, \mu_{\delta_G}, \varpi_G), \mathbf{x}_E = (\mu_{\alpha_E}, \mu_{\delta_E}, \varpi_E), \mathbf{x}_0 = (\mu_{\alpha_0}, \mu_{\delta_0}, \varpi_0) \end{aligned}$$

Once both distributions convoluted and taking into account that the expected values of both catalogues are equal ( $x_{G_0} - x_{E_0} = 0$ ), the final distribution becomes:

$$\phi_C(\mu_\alpha, \mu_\delta, \varpi) = \frac{1}{\sqrt{(2\pi)^3 |\Sigma|}} \exp \left\{ -\frac{1}{2} (\mathbf{x}_G - \mathbf{x}_E)^T \Sigma^{-1} (\mathbf{x}_G - \mathbf{x}_E) \right\} \quad (5)$$

with

$$\begin{aligned} \Sigma &= \Sigma_G + \Sigma_E = [\Sigma_{\mathbf{x}_G - \mathbf{x}_E}]_{ij} = \Sigma_{G_{ij}} + \Sigma_{E_{ij}} \\ &\text{(by the properties of covariance matrices)} \end{aligned}$$

Therefore the  $\chi^2$  matrix to be calculated for each star is:

$$\chi^2 = (\mu_{\alpha_G} \cos \delta_{\alpha_G} - \mu_{\alpha_E} \cos \delta_{\alpha_E}, \mu_{\delta_G} - \mu_{\delta_E}, \varpi_G - \varpi_E) \begin{pmatrix} (\sigma_{\mu_{\alpha}}^2)_G + (\sigma_{\mu_{\alpha}}^2)_E & (\rho_{\mu_{\alpha}\mu_{\delta}} \sigma_{\mu_{\alpha}} \sigma_{\mu_{\delta}})_G + (\rho_{\mu_{\alpha}\mu_{\delta}} \sigma_{\mu_{\alpha}} \sigma_{\mu_{\delta}})_E & (\rho_{\mu_{\alpha}\varpi} \sigma_{\mu_{\alpha}} \sigma_{\varpi})_G + (\rho_{\mu_{\alpha}\varpi} \sigma_{\mu_{\alpha}} \sigma_{\varpi})_E \\ (\rho_{\mu_{\alpha}\mu_{\delta}} \sigma_{\mu_{\alpha}} \sigma_{\mu_{\delta}})_G + (\rho_{\mu_{\alpha}\mu_{\delta}} \sigma_{\mu_{\alpha}} \sigma_{\mu_{\delta}})_E & (\sigma_{\mu_{\delta}}^2)_G + (\sigma_{\mu_{\delta}}^2)_E & (\rho_{\mu_{\delta}\varpi} \sigma_{\mu_{\delta}} \sigma_{\varpi})_G + (\rho_{\mu_{\delta}\varpi} \sigma_{\mu_{\delta}} \sigma_{\varpi})_E \\ (\rho_{\mu_{\alpha}\varpi} \sigma_{\mu_{\alpha}} \sigma_{\varpi})_G + (\rho_{\mu_{\alpha}\varpi} \sigma_{\mu_{\alpha}} \sigma_{\varpi})_E & (\rho_{\mu_{\delta}\varpi} \sigma_{\mu_{\delta}} \sigma_{\varpi})_G + (\rho_{\mu_{\delta}\varpi} \sigma_{\mu_{\delta}} \sigma_{\varpi})_E & (\sigma_{\varpi}^2)_G + (\sigma_{\varpi}^2)_E \end{pmatrix}^{-1} \begin{pmatrix} \mu_{\alpha_G} \cos \delta_{\alpha_G} - \mu_{\alpha_E} \cos \delta_{\alpha_E} \\ \mu_{\delta_G} - \mu_{\delta_E} \\ \varpi_G - \varpi_E \end{pmatrix} \quad (6)$$

The p-value of the obtained  $\chi^2$  for each star is then compared to a significance level, denoted by  $\alpha$ . That tells us if there is significant evidence to reject or not the null hypothesis (the hypothesis of no change or no difference).

In this test we validate the astrometric solution (proper motions and parallaxes) at  $\alpha = 0.01$ . Therefore, if the observed p-value is smaller than 0.01, the null hypothesis is rejected and so the proper motion and parallax of both catalogues for the given star do not match (the astrometric solution is considered not good): the test Fails. If the observed p-value is greater than  $\alpha$ , then the test Passes.

## B ANNEX B: LMC/SMC catalogue generation

### B.1 LMC Catalogue

The following contains a list of the references from which the LMC catalogue was compiled. Some notes and points of interest are given in each case. Beginning with the largest catalogues (OGLE), catalogues were joined one by one, cross matching each time to make sure there was no multiple occurrences for a single target.

#### OGLE-III. RR Lyrae Stars in the Large Magellanic Cloud

No. Sources 24,644 from a possible 24,906

Source Soszyński et al. (2009a)

Remarks The complete catalogue contains 24906 sources. Soszyński performed a cut for galactic field stars, and found 66 in front of the LMC. The catalogue contains a flag for galactic RR Lyrae, in the remarks section of catalogue these targets are referred to as uncertain. This selection of targets only includes the targets not designated as Galactic RR Lyraes and that are not classified as uncertain.

#### OGLE-III: Classical Cepheids in the Large Magellanic Cloud

No. Sources 3,373

Source Soszynski et al. (2008)

Remarks Complete sample contains 3,398 sources. This catalogue selection does not include low amplitude Cepheids, and any targets with uncertain periods are removed.

#### OGLE-III: Eclipsing Binary Stars in the Large Magellanic Cloud

No. Sources 26,120 targets

Source Graczyk et al. (2011)

Remarks Complete sample contains 26138 sources. Star '17681' was removed as within paper it is mentioned that this is a likely galactic ECB. Nowhere is there mention of further Galactic ECBs.

The paper mentions other surveys "Grison et al. (1995) found 79 candidate EBs from the EROS survey, the MACHO survey identified 611 EBs (Alcock et al. 1997).

Derekas et al. (2007) presented clean list of 3 031 EBs from MACHO database and Faccioli et al. (2007) published an extension of the preliminary catalog by Alcock et al. (1997) containing 4 634 stars.” Sample ”gives the relative completeness of  $\sim 95\%$ ” compared to the same area covered by ”the MACHO catalogs of eclipsing binaries (Derekas et al. 2007, Faccioli et al. 2007) and the OGLE-II catalogs of eclipsing binaries (Wyrzykowski et al. 2003, Graczyk and Eyer 2010)”.

### **OGLE-III: Long-Period Variables in the Large Magellanic Cloud**

No. Sources 1,663

Source Soszyński et al. (2009b)

Remarks The selection was limited to Miras which should be better understood and better characterised within the sample. This sample comprises of 56,453 stars. Extensive cleaning and catalogue comparison was done, however the authors suggest a small contribution of difference types of variable red objects: foreground objects, YSI, quasars, etc may remain in the sample.

”To test the completeness of the catalog we compared it with the MACHO catalogue of LPV in the LMC recently published by Fraser et al. (2008). 48,141 of stars in this catalogue could be found in the OGLE-III fields, but only 28 257 of these variable stars are assigned by Fraser et al. (2008) to one of the sequences in the PL space. The remaining objects are classified as one-year artifacts, are outside the boundaries of any PL sequence, or have no classification at all. These stars are not firm LPVs, so we removed them from the sample which were matched with our catalog”

”We cross-identified the OGLE and MACHO lists and found no counterparts for 1208 MACHO variables. We carefully checked all these stars which are not present in our sample. 716 of these objects are sequence E stars (ellipsoidal and eclipsing binaries) which will be published in the future. 98 of missing objects occurred to be Cepheids, RR Lyr stars, young stellar objects, quasars, nearby red dwarfs or other types of variable objects. From the remaining 394 stars (1.4% of the MACHO sample) 134 objects had extremely low amplitudes (we included some of them in our catalog, for others we did not detect any significant variability), 81 were affected by a small number of observations due to their location close to the edge of the OGLE field and 179 were saturated in our frames.”

### **Spitzer SAGE Infrared Photometry of Massive Stars in the Large Magellanic Cloud**

No. Sources 1,750 (entire catalogue)

Source Bonanos et al. (2009)

Remarks Not much mention of contamination.

“We first compiled a catalog of massive stars with known spectral types in the LMC from the literature. We then cross-matched the stars in the SAGE database, after incorporating optical and near-infrared photometry from recent surveys of the LMC. The resulting photometric catalogue was used to study the infrared properties of the stars.”

### **High- and Intermediate-Mass Young Stellar Objects in the Large Magellanic Cloud**

No. Sources 1,697

Source Gruendl & Chu (2009)

Remarks Total catalogue 2,910. Full catalogue includes background stars, so catalogue was complied with the constraint Class ! —G —D —GC, where G = Background Galaxy, D = Diffuse Source, GC = Possible young stellar object or background galaxy.

### **Spectroscopy of Red Giants in the Large Magellanic Cloud Bar**

No. Sources 373 targets

Source Cole et al. (2005)

Remarks Radial velocity measurements in the LMC.

“We are interested in stellar radial velocities in order to reject possible foreground Milky Way stars, and to search for correlations between the moments of the velocity distribution and metallicity that could help distinguish between different stellar populations.” “The mean radial velocity of our sample is  $V_{\odot} = 257 \text{ km.s}^{-1}$ , with an rms dispersion of  $25 \text{ km.s}^{-1}$  about the mean. We found no stars with velocities characteristic of the Milky Way disk ( $V_{\odot} \lesssim 100 \text{ km.s}^{-1}$ ), and the observed velocity range of  $174 \text{ km.s}^{-1} \lesssim V_{\odot} \lesssim 336 \text{ km.s}^{-1}$  is entirely consistent with the known range of LMC radial velocities (e.g., Zhao et al. 2003). Some Galactic halo giants have similar velocities, but since they are far fewer in number than disk stars, the contamination rate is negligible. ”

### **Yellow and Red Supergiants in the Large Magellanic Cloud**

No. Sources 506 candidates in all.

Source Neugent et al. (2012)



Remarks Uses proper motions and radial velocities to weed out galactic contamination.

“Owing to the large radial velocity separation between the Milky Way and the LMC, membership determination proved to be straightforward.” “We concluded that stars with radial velocities higher than  $200 \text{ km.s}^{-1}$  (317 YSG candidates and 505 RSG candidates) are probable LMC supergiants, and these were labelled category 1. Stars with radial velocities lower than  $155 \text{ km.s}^{-1}$  (1129 YSG candidates and 17 RSG candidates) are probable foreground dwarfs and were labelled category 3. The six candidates (all YSG candidates) between these two radial velocity cutoffs were then labelled category 2 or possible, but not probable, LMC supergiants.” “While foreground dwarfs can be eliminated using their radial velocities, this is not the case with halo giants as their radial velocities may overlap with that of the LMC, as much of the LMCs systemic velocity is actually the reflex motion of the Sun (Courteau & van den Bergh 1999).”

### **Metallicities, Age-Metallicity Relationships, and Kinematics of Red Giant Branch Stars in the Outer Disk of the Large Magellanic Cloud**

No. Sources 249

Source Carrera et al. (2011)

Remarks Here B or V mag. is given. V magnitude for northern field ( $-61 \leq DE \leq -59$ ), B magnitude for eastern and western fields ( $-71 \leq DE \leq -70$ ). The catalogue split between that which has B phot and that which has V phot.

“... we considered as LMC members those stars with radial velocities in the range  $170 \text{ km.s}^{-1} < V_r < 380 \text{ km.s}^{-1}$  (Zhao et al., 2003). In total, we have determined ages and metallicities for 147 stars confirmed as LMC members from their radial velocities in the six fields observed.” “Carrera et al. (2008b, hereafter Paper II) derived ages and metallicities for about 100 red giant branch (RGB) stars confirmed as LMC members from their radial velocities in each of these fields using the infrared Ca II triplet lines”. “In this paper, we extend our study of the LMC disk at larger galactocentric distances ( $R > 7 \text{ kpc}$ ).”

### **The Chemical Enrichment History of the Large Magellanic Cloud**

No. Sources 542 targets in total

Source Carrera et al. (2008)

Remarks Removed “No member” and “No member, low s/n” from catalogues.

“Stars with radial velocity in the range  $170 \text{ km.s}^{-1} < V_r < 380 \text{ km.s}^{-1}$  (Zhao et al., 2003) are considered LMC members. Only about 20 stars in each configuration of the instrument have been excluded based on their radial velocity.”

### B.1.1 LMC Catalogues not Used

The following catalogues were not included for various reasons. They have been included here for completeness.

1. *Photometric determination of the mass accretion rates of pre-main sequence stars - III. Results in the Large Magellanic Cloud*, Spezzi et al. (2012). 1,057 targets, selected via photometry. Not the most worth while catalogue to include.
2. *The MACHO Project Large Magellanic Cloud Variable-Star Inventory. XIII. Fourier Parameters for the First-Overtone RR Lyrae Variables and the LMC Distance*. Alcock et al. (2004). Covered same footprint as OGLE, only contains 54 targets with coordinates, 24 of which match targets in OGLE
3. *Chemical evolution of the Magellanic Clouds. VI. Chemical composition of nine F supergiants from different regions of the large Magellanic Cloud*. Hill et al. (1995). Lines used in the analysis for nine LMC supergiants.
4. *The Evolution Of Massive Young Stellar Objects in the Large Magellanic Cloud. I. Identification and Spectral Classification*. Seale et al. (2009). Based on spectra, but no mention of possible source contamination.
5. *Outside-In Disk Evolution in the Large Magellanic Cloud*. Gallart et al. (2008). Here just photometry was used so it is not considered.
6. *The Magellanic Clouds Photometric Survey: The Large Magellanic Cloud Stellar Catalog and Extinction Map*. Zaritsky et al. (2004). 24,107,004 stars.
7. *A UBV<sub>R</sub> CCD Survey of the Magellanic Clouds*. Massey (2002). 1,79,655 sources with V and B-V mag. 12,1366 pairs found with Zarisky. 58,289 missing from Zaritsky- within gaps is coverage.
8. *The Optical Gravitational Lensing Experiment. BVI Maps of Dense Stellar Regions. II. The Large Magellanic Cloud*. Udalski et al. (2000). 7,146,832 targets.
9. *A deep objective-prism survey for Large Magellanic Cloud members*. Sanduleak (1970). 1,275 targets, 416 pairs found with (Massey, 2002) and 1,061 pairs found with (Zaritsky et al., 2004).
10. *Emission-line stars in the LMC: the Armagh survey and a metacatalogue*. (Howarth, 2013). Containing also the Cats of Lindsay 1963, and Andrew and Lindsay 1964 with more precise positions. 825 spectroscopically identified candidate H $\alpha$  emission stars in the LMC. Full cat contains 1,417 proper motions from UCAC4, B mag, RA and DEC J2000 but no V mag.

11. *The Wolf-Rayet stars in the Large Magellanic Cloud*. Hainich et al. (2014). Based on the fourth catalog of WR stars in the LMC (Breysacher+1999), 107 stars in the LMC containing coordinates, V mag. but no colour.
12. *Carbon stars in the halo of the Magellanic Clouds: Identification and radial velocity data* Kunkel et al. (1997). Membership in this sample is not clean, “These passbands are particularly good for an AGB carbon survey because the colour at the tip of the giant branch  $B_J - R \gtrsim 2.4$ , equivalent to  $B - V \gtrsim 2$ , lies redward of Galactic foreground contamination or any other possible population.”... “Apart from photometric errors throwing a small number, <10%, of Galactic stars into the sample, virtually all the selected candidates are either late-M giants in the Clouds or cool AGB carbon stars.”
13. *Catalogues of  $H\alpha$ -EMISSION Stars and Nebulae in the Magellanic Clouds*. Henize (1956). 120 in LMC and 115 in SMC (according to Simbad comments)
14. *Studies of the LMC stellar content. I. A catalogue of 272 new O-B2 stars*. Brunet et al. (1975) 272 OB stars.
15. *The ultraviolet view of the Magellanic Clouds from GALEX: A first look at the LMC source catalog*. Simons et al. (2014). This was correlated with optical counterparts in Zaritsky et al. (2004) which is probably a better place to look for sources.
16. *The DENIS Point Source Catalogue towards the Magellanic Clouds* Cioni et al. (2000). Contains 1,300,000 sources towards the LMC and 300,000 towards the SMC but only 70% are members.
17. *The LMC geometry and outer stellar populations from early DES data*. Balbinot et al. (2015). No catalogue given.
18. *First Results from the NOAO Survey of the Outer Limits of the Magellanic Clouds*. Saha et al. (2010). No catalogue.

## B.2 SMC Catalogue

The following contains a list of the references from which the LMC catalogue was compiled. Some notes and points of interest are given in each case. Beginning with the largest catalogues (OGLE), catalogues were joined one by one, cross matching each time to make sure there was no multiple occurrences for a single target.

### OGLE-III: RR Lyr Stars in the Small Magellanic Cloud

No. Sources 2,434 targets from 2,475

Source Soszyński et al. (2010b)

Remarks This was cross correlated with previous RR Lyr star cats (the biggest of which is OGLE II). 558 stars are independently verified from 571 objects classified as RR Lyr in OGLE II. 8 of the missing objects were reclassified, and remaining 5 are RR Lyr and have been included here. Similarly for those in the GCVS Vol. V (Artyukhina+1996) - variable star catalogue, here most of the RR Lyr came from the work Graham+1975. However the majority of these targets lie outside the field of the OGLE III field. Of the 27 stars that are in the OGLE foot print - they identified nearly all targets.

### **OGLE-III: Type II Cepheids in the Small Magellanic Cloud**

No. Sources 4,618

Source Soszyński et al. (2010a)

Remarks Several targets show Cepheid like light curves, but are located outside the instability strip in the colour-magnitude diagram. These 12 objects are marked as uncertain in the catalogue remarks, which are removed from the catalogue for this case.

Final catalogue included visual inspection of all light curves. Proper motions from OGLE data bases were used to remove a few evident galaxy members from the list of candidates.

### **OGLE-III: Long-Period Variables in the Small Magellanic Cloud**

No. Sources 19,384

Source Soszyński et al. (2011)

Remarks “Large- and medium-amplitude light curves (SRVs and Miras) were visually inspected and the classification was done based on the light curve morphology, position of a star in the PL, colormagnitude and periodamplitude diagrams”...”With the aim of excluding foreground Galactic objects from our catalog, we checked the astrometric measurements of the initially selected sources, and removed stars with detectable proper motions. Only in the field SMC140, centered on the Galactic globular cluster 47 Tuc, SMC stars showed proper motions relative to the cluster members (of course, we observed the motion of 47 Tuc, but the coordinate grid was established on the cluster stars, which dominate in the field SMC140). LPVs that are members of 47 Tuc are saturated in the OGLE data, so they are not included in our catalog.”

### **A 2dF survey of the Small Magellanic Cloud**

No. Sources 4,161

Source Evans et al. (2004)

Remarks Cat. of hot luminous stars, with spectral types in the SMC. Though MOST of targets are in Zaritsky+, Massey+ or OGLE

### **Spitzer SAGE-SMC Infrared Photometry of Massive Stars in the Small Magellanic Cloud**

No. Sources 3,654

Source Bonanos et al. (2010)

Remarks Cat. compiled from literature and then cross matched with SAGE-SMC database. Criteria: accurate coordinates and spectral classifications that correspond to  $>8 M_{\odot}$ . Catalogue of Appozardi is included here with updated coordinates from Brain Skiff lists (<ftp://ftp.lowell.edu/pub/bas/starcats/>). DfF Survey (Evans et al. (2004) with 4,161) is the largest cat of SMC stars with SpT makes up backbone. Includes not only hot massive stars, but also the survey of B and Be stars in NGC 330 (Martayan et al. 2007a, 2007b, 314 stars), the VLT-FLAMES Survey of Massive Stars in NGC 330 and NGC 346 (Evans et al. 2006, 232 stars), and unpublished additional FLAMES observations in NGC 346 (PI: Evans; I.Hunter et al. 2010, in preparation, 244 OB & A-type stars). Also the three known luminous blue variables (LBVs), five supergiant B[e] stars (Zickgraf 2006), 69 Be/X-ray binaries (with spectral types from the updated online catalog 17 of Raguzova & Popov 2005), early-type eclipsing and spectroscopic binaries, and red supergiants (Massey & Olsen 2003; Levesque et al. 2006, 2007, 107 stars). Completeness levels: 4% for unresolved stars in cat (estimates total 2800), while W-R cat (Massey+2003) is thought to be close to complete. This catalogue is used in later papers such as Kouriotis to search for variability in the OGLE database.

Very large overlap between Evans and Bonanos but not complete.

### **The RR Lyrae stars in the Small Magellanic Cloud**

No. Sources 81 targets.

Source Graham (1975)

Remarks Two variable stars rediscovered in NGC104 (which is about 1/10th the distance of the SMC). There should be 92 targets in sample. Membership cut should be based on magnitudes. Here I took the sample from catalogue of Artukhina *et al* 1996 (<http://vizier.cfa.harvard.edu/viz-bin/VizieR?-source=II/205>), only contains 81 targets. Artukhina may perform another cut - original table in vizier contains 86 targets, but I removed five as they had non-member flags. All these 81 targets are covered in the newer catalogue by Kholopov *et al* 1998 (<http://vizier.cfa.harvard.edu/viz-bin/VizieR?-source=II/214>). None have reference to non-membership.

### **Spectroscopic Survey of Red Giants in the Small Magellanic Cloud. I. Kinematics**

No. Sources 2,046 objects

Source Harris & Zaritsky (2006)

Remarks "... our final kinematic sample contains 2,079 stars throughout the central 3:9 kpc; 1:8 kpc, of which 2,046 are determined to be SMC members". The targets were photometrically identified. The removal of foreground objects is done firstly by radial velocities, but some overlap is inevitable, so a second criteria is applied, where the Na doublet absorption feature (sensitive to surface gravity of giants). Foreground targets then have low velocity and large EW(Na). 14 foreground objects were found, in addition to 14 showing discrepant radial velocities.

### **Yellow Supergiants in the Small Magellanic Cloud: Putting Current Evolutionary Theory to the Test**

No. Sources 192

Source Neugent et al. (2010)

Remarks "The SMC's large radial velocity ( $\sim 160 \text{ km.s}^{-1}$ ) allows us to separate members from foreground stars. Observations of  $\sim 500$  candidates yielded 176 near-certain SMC supergiants, 16 possible SMC supergiants, along with 306 foreground stars, and provide good relative numbers of yellow supergiants down to  $12 M_{\odot}$  sun."

### **The Araucaria Project. The Distance to the Small Magellanic Cloud from Late-type Eclipsing Binaries**

No. Sources 4

Source Graczyk et al. (2014)

Remarks Four eclipsing binaries in the OGLE footprint, used to determine distance. All included.

### **A Search for Wolf-Rayet Stars in the Small Magellanic Cloud**

No. Sources 11

Source Massey & Duffy (2001)

Remarks Identified with photometry, and confirmed with spectroscopic observations.

### **Ca II Triplet Spectroscopy of Small Magellanic Cloud Red Giants. I. Abundances and Velocities for a Sample of Clusters**

No. Sources 102

Source Parisi et al. (2009)

Remarks Sample used to derive velocities of clusters within the SMC, no photometry in catalogue.

### **Forty eclipsing binaries in the Small Magellanic Cloud: fundamental parameters and Cloud distance**

No. Sources 41

Source Hilditch et al. (2005)

Remarks Derived distances for these Eclipsing binaries.

**SMC Catalogues not Used** The following catalogues were not included for various reasons. They have been included here for completeness.

1. *A finding list of proven or probable Small Magellanic Clouds members.* Sanduleak (1968). This was a primary input cat. for Hipparcos SMC, contains 169 stars (9 are flagged as SMC members in Hipparcos catalogue. Sanduleak is much brighter sample than Evans 2004. Only covers very brightes of Evans.

2. *A catalogue of late-type supergiant stars in the Small Magellanic Cloud.* Prevot et al. (1983). Survey covers large area between SMC and LMC so it cannot rule out possibility that some stars belong to LMC. 221 stars found. However 14 are galactic giants, as  $V < 11.0$  or from radial velocities. 9 stars have  $11.5 < V < 12.5$  (agreement with brightness limit of M supergiants in LMC). 7 stars in common with Dachs (1970) - classified as foreground objects and 2 doubtful members. Some disagreement with these classifications as 4 of these, their radial velocities support their membership. Similarly for Appozardi.
3. *Variability of massive stars with known spectral types in the Small Magellanic Cloud using 8 years of OGLE-III data.* Kourniotis et al. (2014). This should be contained in OGLE.
4. *Catalogue of candidate emission-line objects in the Small Magellanic Cloud.* Murphy & Bessell (2000). No mention of contamination, targets selected based on  $H\alpha$  emission.
5. *Self-consistent physical parameters for five intermediate-age SMC stellar clusters from CMD modelling.* Dias et al. (2014). They do mention cut-offs for cluster membership, could use these, but no mention of confirmed SMC membership.
6. *Review of Discrete X-Ray Sources in the Small Magellanic Cloud: Summary of the ASCA Results and Implication on the Recent Star-Forming Activity.* Yokogawa et al. (2003). No mention of contamination.
7. *A slitless spectroscopic survey for  $H\alpha$  emission-line objects in SMC clusters.* Martayan et al. (2010). No mention of contamination by field stars from the Milky way.
8. *Chemical Abundances of Metal-poor RR Lyrae Stars in the Magellanic Clouds.* Haschke et al. (2012). 6 RR Lyrae stars in the LMC and 3 in the SMC, no mention of membership.
9. *Hot Variable Stars in the SMC Cluster NGC 330.* Schmidtke et al. (2008). Based on OGLE photometry with no mention of contamination.
10. *Photometry of the young cluster NGC 330 in the Small Magellanic Cloud.* Vallenari et al. (1994). Photometry only.
11. *Ages and luminosities of young SMC/LMC star clusters and the recent star formation history of the Clouds.* Glatt et al. (2010) The most recent catalog that cross-correlates all known objects of the LMC, SMC, and the Magellanic Bridge region has been published by Bica et al. (2008). However, the cluster sample is still highly incomplete as pointed out by the authors. Leads to a catalogue of extended objects in the MC clouds.



12. *Effects of metallicity, star-formation conditions, and evolution in B and Be stars. II. Small Magellanic Cloud, field of NGC 330.* Martayan et al. (2007). No mention of field stars, contamination or membership of SMC.
13. *On Be star candidates and possible blue pre-main sequence objects in the Small Magellanic Cloud.* Mennickent et al. (2002). No mention of membership or contamination.
14. *RR Lyrae stars in the southern globular cluster NGC 362.* Székely et al. (2007). Membership flag is not given in online data.
15. *The Optical Gravitational Lensing Experiment. Catalog of High Proper Motion Stars Towards the Magellanic Clouds.* Soszynski et al. (2002). High proper motion catalogue of foreground stars to the Magellanic clouds, contains 2,162 objects.
16. *The Carnegie Hubble Program: The Distance and Structure of the SMC as Revealed by Mid-infrared Observations of Cepheids.* Scowcroft et al. (2015). 92 confirmed (cepheid) members of the SMC, no catalogue yet.





## Résumé

La mission Gaia de l'Agence Spatiale Européenne (ESA) a pour objectif de cartographier notre galaxie avec une précision astrométrique jamais atteinte auparavant. Il est donc particulièrement important que les données qui seront publiées soient rigoureusement validées afin d'assurer une qualité optimum au Catalogue. Ces validations sont faites par l'une des équipes de l'unité de coordination CU9 du Consortium Gaia DPAC (Data Processing and Analysis Consortium) chargé par l'ESA de la production du Catalogue Gaia. Dans le cadre de cette thèse, nous avons mis en place toute l'infrastructure nécessaire à la validation du catalogue Gaia par comparaison avec des catalogues externes. Celle-ci gère toutes les interactions avec l'environnement global des validations et avec la base de données Gaia. Ensuite nous avons développé un ensemble de tests statistiques pour valider les données du premier catalogue Gaia (DR1). Ces tests concernent notamment l'homogénéité des données sur le ciel, la qualité des positions et de la photométrie de l'ensemble des étoiles de DR1 (plus d'un milliard d'étoiles,  $V < 20$ ) ainsi que celle des parallaxes et mouvements propres des étoiles de Tycho-Gaia Astrometric Solution (TGAS), environ deux millions d'étoiles communes aux catalogues Gaia et Tycho-2 ( $V < 12$ ).

Ces tests statistiques sur la DR1 sont opérationnels et ont déjà été appliqués très récemment sur des données préliminaires. Cela a déjà permis d'améliorer ces données (donc la qualité du catalogue), et d'en caractériser les propriétés statistiques. Cette caractérisation est essentielle à une exploitation scientifique correcte des données. Le premier catalogue Gaia sera publié à la fin de l'été 2016.

Parmi les objets observés par Gaia, il y a une population d'étoiles particulièrement intéressantes, les étoiles du Red Clump (RC), très utilisées comme étalons de distance. Nous avons développé et testé deux méthodes pour modéliser les relations couleur-couleur (CC) et température effective-couleur dans toutes les bandes photométriques, de l'ultraviolet au proche-infrarouge. Elles permettront de caractériser le RC dans la bande G de Gaia dès la publication du catalogue : 1) en utilisant des modèles théoriques, et 2) empiriquement, en se basant sur une méthode Monte Carlo Markov Chain (MCMC). Pour cela nous avons très rigoureusement sélectionné des échantillons d'étoiles avec une bonne qualité photométrique, une bonne métallicité, déterminée par spectroscopie, une température effective homogène et une faible extinction interstellaire.

À partir de ces calibrations CC et température-couleur, nous avons ensuite développé une méthode qui permet de déterminer les magnitudes intrinsèques, les températures et les extinctions des étoiles du RC. Les couleurs et extinctions ainsi obtenues ont été testées sur des étoiles avec des températures effectives mesurées spectroscopiquement et une extinction déterminée par la mesure des Bandes Diffuses Interstellaires (DIB). Ces propriétés intrinsèques des étoiles du RC vont permettre de caractériser le Red Clump Gaia et de calibrer, dans la bande Gaia, la magnitude absolue de cet étalon de distance, premier échelon essentiel de la détermination des distances dans l'Univers.

## Mots Clés

Gaia, Validation, Étoiles du Red Clump, Calibration photométrique

## Abstract

The Gaia mission of the European Space Agency (ESA) aims to map our galaxy with an unprecedented astrometric precision. It is therefore very important that the data that will be published be rigorously validated to ensure an optimal quality in the Catalogue. These validations are done by one of the teams of the coordination unit CU9 of the Gaia DPAC Consortium (Data Processing and Analysis Consortium) commissioned by ESA of the Gaia catalogue production. As part of this thesis, we implemented all the necessary infrastructure to validate the Gaia catalogue by comparison with external catalogues. This last manages all the interactions with the global environment of validations and with the Gaia database. Then we developed a set of statistical tests to validate the data from the first Gaia catalogue (DR1). These tests relate in particular to the homogeneity of data on the sky, the quality of the positions and of photometry of all the stars of DR1 (more than a billion stars,  $V < 20$ ) as well as that of the parallaxes and proper motions for Tycho-Gaia Astrometric Solution (TGAS) stars, around two million stars in common in Gaia and Tycho-2 catalogues ( $V < 12$ ).

These DR1 statistical tests are operational and were already applied very recently on preliminary data. This has improved the data (thus the quality of the catalogue) as well as allowed to characterize the statistical properties. This characterisation is essential for a correct scientific exploitation of the data. The first Gaia catalogue will be released in late summer 2016.

Among the objects that Gaia observes, there is a population of stars particularly interesting, the Red Clump (RC) stars, widely used for distance indicators. We developed and tested two methods to model the colour-colour (CC) and effective temperature-colour relations in all photometric bands, from the ultraviolet to the near infrared. They will allow us to characterize the RC in the Gaia G band upon publication of the catalogue: 1. using theoretical models, and 2. empirically, based on a Monte Carlo Markov Chain (MCMC) method. To do so, we have very carefully selected samples of stars with a good photometric quality, a good metallicity determined by spectroscopy, an homogeneous effective temperature and a low interstellar extinction.

From these CC and temperature-colour calibrations, we then developed a method that allows to derive intrinsic magnitudes, temperatures and extinctions of the RC stars. Estimates of colours and extinctions are tested on stars with spectroscopically measured effective temperatures and an extinction determined by the measurement of Diffuse Interstellar Bands (DIB). These intrinsic properties of RC stars will allow to characterize the Gaia RC and calibrate, within the Gaia G band, the absolute magnitude of this standard candle, first essential step of determining distances in the Universe.

## Keywords

Gaia, Validation, Red Clump stars, Photometric calibration