Astronomy
&
Astrophysics

The *Gaia*-ESO Survey: radial metallicity gradients and age-metallicity relation of stars in the Milky Way disk[*]

M. Bergemann[1], G. R. Ruchti[2], A. Serenelli[3], S. Feltzing[2], A. Alves-Brito[4,23], M. Asplund[4], T. Bensby[2], P. Gruiters[5], U. Heiter[5], A. Hourihane[1], A. Korn[5], K. Lind[1], A. Marino[4], P. Jofre[1], T. Nordlander[5], N. Ryde[2], C. C. Worley[1], G. Gilmore[1], S. Randich[6], A. M. N. Ferguson[10], R. D. Jeffries[11], G. Micela[12], I. Negueruela[13], T. Prusti[14], H.-W. Rix[15], A. Vallenari[16], E. J. Alfaro[21], C. Allende Prieto[7], A. Bragaglia[16], S. E. Koposov[1,8], A. C. Lanzafame[24], E. Pancino[17,9], A. Recio-Blanco[18], R. Smiljanic[19,20], N. Walton[1], M. T. Costado[21], E. Franciosini[6], V. Hill[18], C. Lardo[17], P. de Laverny[18], L. Magrini[6], E. Maiorca[6], T. Masseron[1], L. Morbidelli[6], G. Sacco[6], G. Kordopatis[1], and G. Tautvaišienė[22]

(Affiliations can be found after the references)

ABSTRACT

We study the relationship between age, metallicity, and α-enhancement of FGK stars in the Galactic disk. The results are based upon the analysis of high-resolution UVES spectra from the *Gaia*-ESO large stellar survey. We explore the limitations of the observed dataset, i.e. the accuracy of stellar parameters and the selection effects that are caused by the photometric target preselection. We find that the colour and magnitude cuts in the survey suppress old metal-rich stars and young metal-poor stars. This suppression may be as high as 97% in some regions of the age-metallicity relationship. The dataset consists of 144 stars with a wide range of ages from 0.5 Gyr to 13.5 Gyr, Galactocentric distances from 6 kpc to 9.5 kpc, and vertical distances from the plane $0 < |Z| < 1.5$ kpc. On this basis, we find that i) the observed age-metallicity relation is nearly flat in the range of ages between 0 Gyr and 8 Gyr; ii) at ages older than 9 Gyr, we see a decrease in [Fe/H] and a clear absence of metal-rich stars; this cannot be explained by the survey selection functions; iii) there is a significant scatter of [Fe/H] at any age; and iv) [Mg/Fe] increases with age, but the dispersion of [Mg/Fe] at ages >9 Gyr is not as small as advocated by some other studies. In agreement with earlier work, we find that radial abundance gradients change as a function of vertical distance from the plane. The [Mg/Fe] gradient steepens and becomes negative. In addition, we show that the inner disk is not only more α-rich compared to the outer disk, but also older, as traced independently by the ages and Mg abundances of stars.

Key words. stars: abundances – stars: fundamental parameters – solar neighborhood – Galaxy: disk – Galaxy: formation – surveys

1. Introduction

One of the main quests in modern astrophysics is to understand the formation history of the Milky Way. This requires observational datasets that provide full phase-space information, ages, and element abundances for a large number of stars throughout the Galaxy. The distribution functions of these quantities present major constraints on the models of the formation and evolution of the Milky Way (e.g. Burkert et al. 1992; Pagel & Tautvaisiene 1995; Velazquez & White 1999; Hou et al. 2000; Chiappini et al. 2001; Brook et al. 2004, 2005; Naab & Ostriker 2006; Schönrich & Binney 2009b; Kubryk et al. 2013; Minchev et al. 2013; Rix & Bovy 2013).

In this context, the prime observables are the radial metallicity profiles and the age-metallicity relation in the Galactic disk (see e.g. the review by Feltzing & Chiba 2013). For a long time, the mere existence of any age-metallicity relation has been a matter of great debate. The study by Twarog (1980), which concluded that ages and metallicities of stars in the disk are uniquely correlated, was the first important step in putting this discussion onto a modern footing. Edvardsson et al. (1993), however, was able to show that there is no strong correlation. In addition, Feltzing et al. (2001), and later the Geneva-Copenhagen Survey

(Nordström et al. 2004; Holmberg et al. 2007, 2009; Casagrande et al. 2011a), found a large scatter in metallicity at all ages, and confirmed the tentative existence of an old metal-rich population. The most recent in-depth analysis of the age-metallicity relation in the solar neighbourhood is that by Haywood et al. (2013) who come to similar conclusions.

Combined radial and vertical element abundance profiles provide the second major observational constraint to the Galaxy formation models. Attempts have been made to quantify the metallicity distribution function and abundance gradients of the disk using data from recent large-scale spectroscopic surveys, such as SEGUE/SDSS (e.g. Lee et al. 2011; Bovy et al. 2012b,c; Schlesinger et al. 2012; Cheng et al. 2012), RAVE (e.g. Boeche et al. 2013), and APOGEE (e.g. Anders et al. 2014; Hayden et al. 2013), as well as smaller samples of stars with high-resolution observations (e.g. Bensby et al. 2003; Fuhrmann 2008; Ruchti et al. 2011; Bensby et al. 2014). However, the bulk of the data has so far come from the traditional techniques, such as OB stars, Cepheids, and H II regions (Andrievsky et al. 2002b,a; Lépine et al. 2011; Lemasle et al. 2013). These populations are young, <1 Gyr, only providing a snapshot of the present day abundance pattern. Open clusters and planetary nebulae also provide information about older populations (Friel et al. 2002; Stanghellini & Haywood 2010; Yong et al. 2012; Heiter et al. 2014): open clusters up to 8−9 Gyr and planetary nebulae up to 6 Gyr (Maciel et al. 2010). However, in order to fully study the evolution and

build-up of radial and vertical abundance distributions in the Galactic disk, we need tracers of stellar populations of all ages, from the earliest epoch of Galaxy formation to the present day.

With the data taken with the *Gaia*-ESO Survey, which form the basis of our work, we are in a good position to address these two fundamental problems and set the stage for the larger datasets coming in the future releases from the survey. The *Gaia*-ESO Survey was awarded 300 nights on the Very Large Telescope in Chile. In the high-resolution ($R = \lambda/\Delta\lambda \sim 47\,000$) mode it will acquire spectra for about 5000 field stars, probing distances up to 2 kpc from the Sun; 100 000 spectra will be acquired in medium-resolution mode ($R \sim 16\,000$), with stars probing distances up to 15 kpc. For the analysis of the spectra, several state-of-the-art spectrum analysis codes are used (Gilmore et al. 2012; Randich et al. 2013; Smiljanic et al., in prep.). In addition, sophisticated methods based on Bayesian inference are now available that combine in a systematic way observational information and stellar evolution theory (Pont & Eyer 2004; Jørgensen & Lindegren 2005; Schoenrich & Bergemann 2013; Serenelli et al. 2013).

In this study, we use the first results obtained with the *Gaia*-ESO survey to study the relationship between age and metallicity in the Galactic disk. We note that this is an exploratory study and more definitive results will be available from the next *Gaia*-ESO data releases with larger samples. Furthermore, there are other survey papers in preparation that will also explore the disk and other components in the Milky Way (Recio-Blanco et al.; Mikolaitis et al.; Rojas-Arriagada et al., in prep.). In the upcoming papers, we will also address other aspects of the Galactic disk population, including the now-controversial dichotomy of the disk (Bovy et al. 2012a), by deriving the kinematics of stars and adding individual abundances of different chemical elements. We apply the Bayesian method developed by Serenelli et al. (2013) to derive stellar ages. We explore the limitations of the observed datasets and perform a detailed analysis of the derived stellar properties (mass, ages, metallicity). Finally, we use these data that span a full range of ages from 1 Gyr to 14 Gyr, to study the relationship between elemental abundances and ages of stars, and discuss the radial abundance profiles in the Galactic disk. We also discuss the correlation between Mg abundance and ages of stars.

The paper is structured as follows. In Sect. 2 observations and data reduction are described. Section 3 presents the details of stellar evolution analysis and various tests applied to the input datasets. The sample selection bias is discussed in Sect. 4. Finally, Sect. 5 describes the results in the context of the Galactic evolution, with the focus on the age-abundance and spatial distribution of stars in the Milky Way disk, and the conclusions are drawn in Sect. 6.

2. Observations and stellar parameters

This work makes use of the high-resolution UVES observations[1] taken within the first half year of observations with the *Gaia*-ESO survey (Jan. 2012–June 2012), which comprises 576 stars. The details on the data reduction will be provided in Sacco et al. (in prep.). Of those, 348 are field stars while the others are members of open or globular clusters. The distribution of the observed field sample in the NIR colour–magnitude diagram and the selection box is shown in Fig. 1. The UVES solar neighbourhood targets were chosen according to their colours to maximise the fraction of un-evolved FG stars within 2 kpc in the
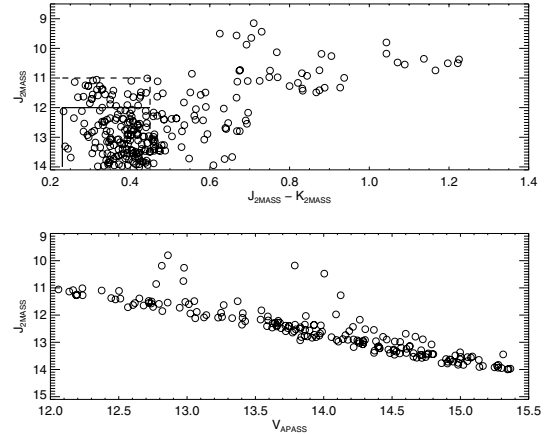
---

[1] Resolution $R \sim 47\,000$.

Fig. 1. Distribution of the observed UVES data sample in the 2MASS (*top*) colour–magnitude plane and as a function of V magnitudes from APASS (*bottom*). The photometric box is shown (see text).

solar neighbourhood. The survey selection box was defined using the 2MASS photometry: $12 < J < 14$ and $0.23 < J - K < 0.45 + 0.5E(B-V)$; if there were not enough targets, the red edge was extended[2]. With these criteria, we are predominantly selecting FG stars with magnitudes down to $V = 16.5$ (Gilmore et al., in prep.).

The stellar parameters, $T_{\rm eff}$, $\log g$, [Fe/H], and abundances for the UVES spectra are determined as follows. The observed spectra were processed by 13 research groups within the *Gaia*-ESO survey collaboration with the same model atmospheres and line lists, but different analysis methods, e.g. full spectrum chi-square minimisation, using pre-computed grids of model spectra or calculating synthetic spectra on the fly, or an analysis based on equivalent widths. The model atmospheres are 1D LTE spherically-symmetric ($\log g < 3.5$) and plane-parallel ($\log g \geq 3.5$) MARCS models (Gustafsson et al. 2008). The determination of the final set of stellar parameters based on all those analyses is described in Smiljanic et al. (in prep.). In short, the final parameter homogenisation involves a multi-stage process, in which both internal and systematic uncertainties of different datasets are carefully evaluated. Various consistency tests, including the analysis of stellar clusters, benchmark stars with interferometric and astroseismic data (Jofre et al. 2014; Blanco-Cuaresma et al. 2014; Heiter et al., in prep.), have been used to assess each group's performance. The final parameters are taken to be the median of the multiple determinations, and the uncertainties of stellar parameters are median absolute deviations, which reflect the method-to-method dispersion (see Sect. 5 for more details). This dataset is available internally to the *Gaia*-ESO collaboration, and will be referred to as *iDR1*.

2.1. Comparison with photometry

The spectroscopic $T_{\rm eff}$ scale can be tested using the infrared flux method (IRFM; Blackwell et al. 1980). We computed the effective temperature $T_{J-K_{\rm S}}$ using the colour-temperature calibrations from González Hernández & Bonifacio (2009) as described in Ruchti et al. (2011). The J, H, and $K_{\rm S}$ magnitudes

---

[2] The targets selected before April (2012) had the brightest cut on J of 11 instead of 12. If the number of objects in the field within the box was less than the number of UVES fibres, then the red-edge of the colour-box was shifted to have enough targets to fill the fibres. In addition, very red objects were selected with a colour-dependent J cut to prevent low S/N in the optical.
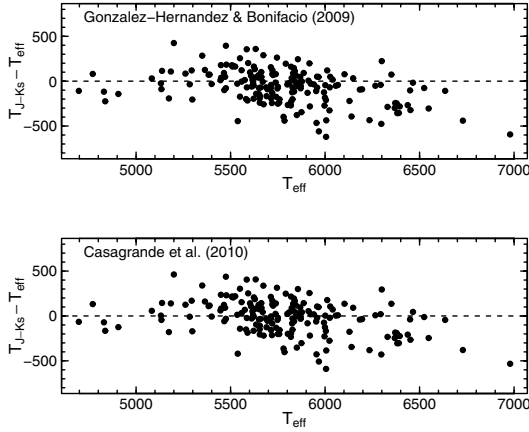
Fig. 2. Comparison of spectroscopic temperatures and photometric T_{J-K} temperatures for a subsample of stars (see Sect. 2.1).

were taken from 2MASS and reddening was estimated in a iterative procedure starting from that found with the Schlegel et al. (1998) dust maps. Those stars with $E(B-V) > 0.1$ were then corrected according to the prescription described in Bonifacio et al. (2000):

$$E(B-V)_{\text{corrected}} = 0.035 + 0.65 * E(B-V)_{\text{Schlegel}}.$$

Finally, we applied the correction for the dust layer, such that the reddening to a star at distance D and Galactic longitude b is reduced by a factor $1 - \exp(-| D \sin b |/h)$ where $h = 125$ pc. This reduction affects the stars that are relatively nearby and lie close to the Galactic plane.

We compare the results only for a subsample of the stars with $E(B-V) < 0.05$. Figure 2 shows the differences between the spectroscopic and photometric temperatures for the calibration from González Hernández & Bonifacio (2009) (top panel) and for the calibration by Casagrande et al. (2010) (bottom panel). The calibrations provide very similar results; the Casagrande et al. (2010) calibration is 50 K warmer. The both temperature scales appear to be in agreement on average, but there is a systematic drift at higher T_{eff}. The origin of this drift is not clear: it could be caused by reddening, uncertainties in theoretical photometric magnitudes, or by spectroscopic uncertainties. For this study, we have decided to keep all stars. The difference between our scale and that given by IR photometry is $\Delta T_{\text{eff}} = -20 \pm 170$ K (calibration by González Hernández & Bonifacio 2009). The intrinsic uncertainty of the $J - K$ IRFM calibration is 150 K, which may contribute to the wide apparent spread.

2.2. NLTE effects

One important problem that needs to be addressed is the influence of the LTE approximation on the determination of stellar parameters. As shown by Bergemann et al. (2012) and Lind et al. (2012), non-local thermodynamic equilibrium (NLTE) effects change surface gravities and metallicities of stars, in a way such that $\log g$ and [Fe/H] obtained from Fe lines are systematically higher. However, these studies also showed that the differences are typically small at metallicity [Fe/H] > -1.

We have utilised the Spectroscopy Made Easy (SME) spectral synthesis code (Valenti & Piskunov 1996) to study the difference between stellar parameters derived in LTE versus those derived in NLTE. SME has been recently upgraded by the module to solve for NLTE line formation using the Fe grids described in Lind et al. (2012). A complete description of the new

module will be presented elsewhere. Application of the NLTE SME module to the complete *iDR1* UVES dataset analysed in this work revealed that the NLTE effects are minor. For completeness, we summarise here our results for dwarfs and giants, even though the latter are not used in the final analysis of the age-metallicity relation and abundance distributions.

The NLTE excitation equilibrium of Fe I lines results in somewhat lower effective temperatures for giants, with stronger effects toward higher effective temperatures. The effect on surface gravity is more complex. The direct effect of higher Fe I abundances results in higher $\log g$. At the same time, the lower T_{eff} counteract the direct effect on Fe I abundances, and so $\log g$ comes out with slightly lower values. The net result is that [Fe/H] stays essentially constant. For stars with $\log g < 3.5$, the mean difference between the NLTE and LTE parameters is: -0.25 ± 17 K (T_{eff}), -0.05 ± 0.06 dex ($\log g$), -0.01 ± 0.02 dex ([Fe/H]).

For dwarfs, the effect is reversed. Surface gravities and metallicities increase, although the difference with respect to an LTE analysis is very small. For stars with $3.5 < \log g < 4.5$, the mean bias in $\log g$ and [Fe/H] is 0.03 ± 0.04 dex and 0.01 ± 0.01 dex, respectively. There is no effect on T_{eff}.

SME is not yet able to compute NLTE abundances for Mg, thus a realistic assessment of the NLTE effect cannot be done yet. We have checked other theoretical studies in the literature and found that the bias is also very small. For example, for the two most important lines of Mg I, 5711 and 5183 Å, the difference between the NLTE and LTE abundance is not larger than ± 0.05 dex for stars with $\log g > 3.5$ (e.g. Shimanskaya et al. 2000).

The differences between our LTE and NLTE results are well within the uncertainties of the stellar parameters and abundances. We thus conclude that the effect of NLTE in the studied range of stellar parameters will not significantly bias the resulting age and abundance distributions.

3. Determination of ages and distances

The ages and absolute magnitudes of stars were determined using the Bayesian pipeline BeSPP[3] developed by Serenelli et al. (2013). The grid of stellar evolutionary tracks has been computed with the GARSTEC code (Weiss & Schlattl 2008). In GARSTEC, nuclear reaction rates are those recommended by Adelberger et al. (2011), OPAL opacities by Iglesias & Rogers (1996), complemented at low temperatures by those from Ferguson et al. (2005), the FreeEOS equation of state from Cassisi et al. (2003). Convection is treated with the standard mixing length theory ($\alpha_{\text{MLT}} = 1.811$, from a solar calibration), and overshooting has been included as a diffusive mixing process (Freytag et al. 1996). More details can be found in Weiss & Schlattl (2008) and Serenelli et al. (2013).

Our grid of stellar evolution models spans the mass range $0.6 \leq M/M_\odot \leq 3.5$ with steps of $\Delta M = 0.01 \ M_\odot$ and the metallicity range $-5.0 \leq$ [Fe/H] ≤ 0.5 with steps of Δ[Fe/H] $= 0.1$ for [Fe/H] ≤ 0 and Δ[Fe/H] $= 0.05$ for [Fe/H] > 0. Models with [Fe/H] ≤ -0.6 include 0.4 dex α-enhancement. A complete description of the pipeline can be found in Serenelli et al. (2013). In that paper, we showed that below $\log g \sim 3.5$, stellar mass and age cannot be determined because of the degeneracy of stellar evolutionary tracks in the $T_{\text{eff}} - \log g$ plane (see also Allende Prieto & Lambert 1999). Therefore, in this work we consider only stars with $3.5 < \log g < 4.5$. Due to the $\log g$ cut, 54 evolved stars are removed from our sample.

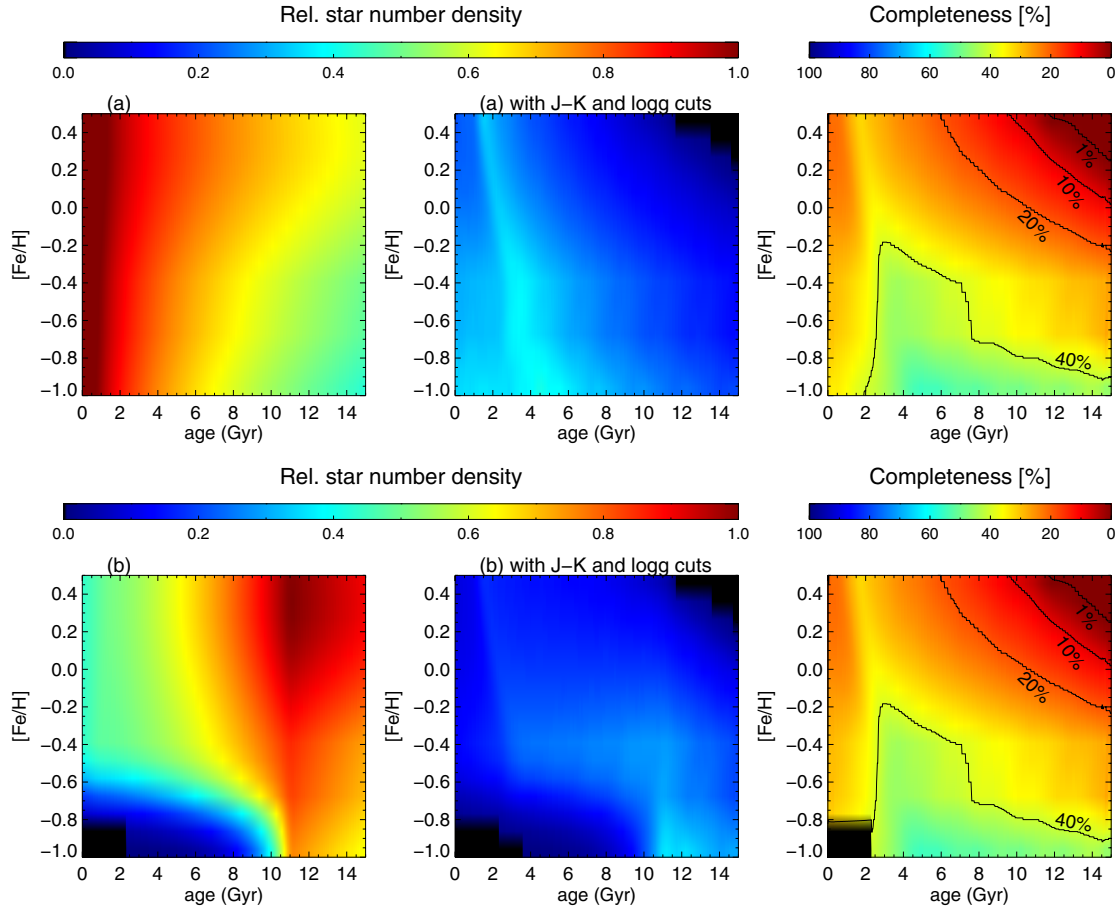---

[3] Bellaterra Stellar Parameter Pipeline.

Fig. 3. Stellar number density computed from the evolution tracks assuming only IMF (**a**), *top*) and IMF+SFR (**b**), *bottom*) without colour cuts. *Middle panels*: with the colour cut ($0.23 < J − K < 0.45$) and stellar parameter limits ($3.5 \leq \log g \leq 4.5$) to simulate the selection effect of the survey and our stellar sample. *Right*: the suppression factor, defined as the ratio of stars with cut to the number of stars without cut, i.e. the larger the number the more stars are suppressed (see text).

Age was determined by fitting a normal distribution of the probability distribution function (PDF) in the HR-diagram where the probability density is higher than 20% of the maximum. The best-estimate age is taken to be the centre of the Gaussian fit and the uncertainties are determined from the full PDF as $1−\sigma$ confidence level. We also tested other statistics using a sample of synthetic stars generated from the tracks (see Appendix A), but found that neither the weighted mean nor the median of the underlying age PDF is a satisfactory approximation of the age when the PDF is broad and asymmetric. Means or medians in truncated PDFs on the young side (or on the old side) also lead to an overestimation (underestimation) of the true age of the star.

The BeSPP pipeline has been tested in a variety of ways in Serenelli et al. (2013). However, in this work we are interested in how uncertainties on stellar parameters influence the determination of ages. This test was not done in our previous work, which focused on the relative changes of stellar ages caused stellar parameters derived under LTE or NLTE. To assess systematic biases, which could be introduced by the Bayesian method, we performed test simulations with mock samples of stars drawn randomly from the evolutionary tracks. We have also checked for the possibility of a systematic mismatch between $T_{\rm eff}$ and $\log g$ scales from stellar evolution models and from spectroscopy. From these tests (see Appendix A), we found that the age estimate is robust and un-biased when its uncertainty is either smaller than $\sim$30% or smaller than 2 Gyr. The latter

condition is relevant for younger stars that naturally tend to have larger fractional errors.

Distances were determined following Ruchti et al. (2011), using the absolute magnitudes as described above and the $K_{\rm S}$ magnitudes from 2MASS. Reddening is described in Sect. 2.1. The Galactocentric radial R distance and vertical Z distance from the plane were derived using the combination of each star's distance from the Sun and Galactic (ℓ, b) coordinates. We note that we have assumed that the Sun lies at $R = 8.3$ kpc in the plane. Uncertainties in both R and Z were propagated from the uncertainty in the distance.

4. Sample completeness

One of the most difficult but fundamental tasks is to assess the completeness of the observed dataset. Various selection functions (e.g. in colours, magnitudes, or even $T_{\rm eff}$ and $\log g$) will reduce the relative number of stars with certain properties. In principle, accounting for the sample selection is mathematically straightforward (e.g. Rix & Bovy 2013); however this has rarely been applied rigorously.

To assess the influence of survey selection cuts, we designed a simple test. First, we create a stellar population with a Salpeter initial mass function (IMF), a constant star formation rate (SFR), and a uniform metallicity distribution. The resulting density of sample stars is shown on the upper-left panel of Fig. 3. The upper-central panel of Fig. 3 shows the distribution
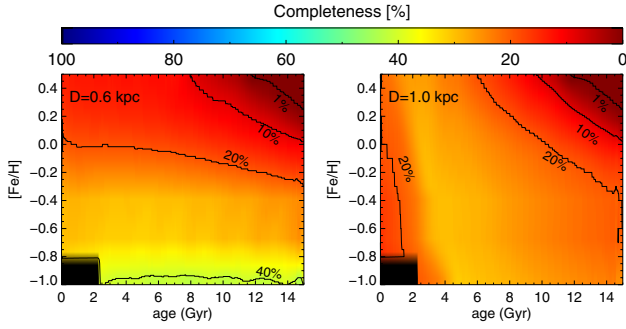
Fig. 4. Same as Fig. 3 but also including the magnitude cut of the survey: $12 < J < 14$.
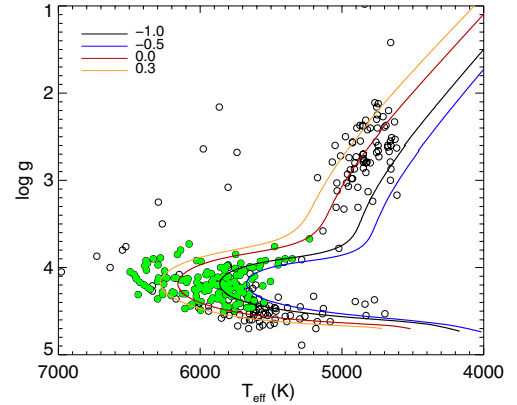

Fig. 5. Selected UVES sample: (black open circles) – all iDR1 stars; (filled green circles) – stars that satisfy our selection criteria. 8 Gyr GARSTEC isochrones for different [Fe/H] are also shown (see Sect. 3).

retained after the colour cut used for the UVES sample is applied, $0.23 < J - K < 0.45$, and with the cuts on surface gravity, $3.5 < \log g < 4.5$ (see Sect. 2.1). The density of stars in the sample is highest where $J - K = 0.23$, that approximately corresponds to $\sim$6500 K, i.e. the blue cut of the UVES sample. As the figure shows, the cuts reduce the overall number of stars, with a stronger effect on the oldest metal-rich stars. By taking the ratio of the stellar number densities computed with (N_{cut}) and without (N_{nocut}) the colour and gravity cuts, we can estimate what fraction of stars is retained in the sample. The sample completeness is thus defined as:

$$f = N_{cut}/N_{nocut}$$

and it is shown in the upper-right panel of Fig. 3, along with the contours of equal probability. The old metal-rich stars, [Fe/H] > 0.2 can be suppressed almost completely, but also the fraction of old stars with solar metallicity quickly declines.

We have also considered a more sophisticated model, assuming an SFR varying with time and with [Fe/H] (see Appendix B). The results of this simulation are shown in the lower three panels of Fig. 3. The IMF + SFR distribution of the stellar density (the lower-left panel) is markedly different from the case of a constant SFR. Because the SFR decreases with time, the stellar number density always peaks at old ages. In particular, the short e-folding time at the lowest [Fe/H] is evident by the almost complete absence of young ($<$5 Gyr) metal-poor stars ([Fe/H] < −0.7). After applying the colour and gravity cuts (low-middle panel), the notable difference with respect to the constant SFR case (top middle panel), in addition to the absence of young metal-poor stars, is the overall lower density at younger ages for any [Fe/H]. The sample completeness is shown in the lower-right panel and it is identical to the case with a flat SFR (upper-right panel). This result can be understood easily. Stellar number density is constructed by integrating for each age and [Fe/H], over the IMF and the SFR. However, the contribution of the SFR to the stellar number density is only a function of age and [Fe/H] and therefore cancels out when computing the sample completeness as a function of these two quantities. Adding a spatial distribution to the discussion also does not affect this conclusion, as long as it does not include a dependence on the stellar mass.

In the next step, we have imposed the magnitude cut according to the survey selection. This was achieved by transforming the absolute magnitudes of synthetic stars to the apparent magnitudes for a range of distances, from 100 pc to 2 kpc (typical of the observed sample). The distributions were then restricted to J magnitudes in the range from 12 to 14 mag. Figure 4 shows two representative cases, where the distance is 0.6 and 1 kpc (left

and right panel), also including the colour and gravity cuts as described above. Beyond 0.9 kpc, the density of young stars with super- and subsolar metallicity drops, also at distances smaller than 0.6 kpc metal-rich stars are suppressed. In either case, the most vulnerable region is the upper right corner of the age-metallicity plot, that corresponds to old metal-rich stars.

To summarise, the combined effect of the colour and magnitude limits in the *Gaia*-ESO survey catalogue is to under-sample stars that are either young *and* metal-poor, or old *and* metal-rich. However, the extent to which these stars are under-sampled, depends quantitatively on their distance and actual age distribution, which is not known a priori. Furthermore, we have not accounted for the anisotropy of the inter-stellar reddening, which could have a differential effect on the typical characteristics of stars observed within a given photometric box as a function of spatial direction. Therefore, we see these tests only as a numerical illustration to help us understand the picture qualitatively, and the results cannot be used to correct the observed sample.

5. Results

5.1. Age-[α/Fe] and age-[Fe/H] relations

The main goal of this work is to study the relation between ages and abundances of stars in the Milky Way disk. As discussed in Sect. 3, we apply cuts on $\log g$ and on age errors. We also remove stars with uncertainties in Mg abundances larger than 0.15 dex, which leaves us with 144 stars in the final sample (Fig. 5). The mean uncertainties are 80 K in T_{eff}, 0.15 dex in $\log g$, 0.06 in [Fe/H], and 0.06 in [Mg/Fe]. The results for all selected stars in the age-metallicity and age-[Mg/Fe] planes are shown in Fig. 6. The contours in the top panel of Fig. 6 indicate the relative sample completeness; the percentages were normalised to its peak value, which is $\sim$35% (Fig. 3).

First, we turn to the discussion of the age-metallicity relation (Fig. 6, top panel). The main features demonstrated in earlier observational studies of the disk by Feltzing et al. (2001); Haywood et al. (2013); Bensby et al. (2014) are clearly visible. Between 0 Gyr and 8 Gyr the trend is predominantly flat, and the scatter in [Fe/H] is large at any given age. This is the most populated locus on the diagram, which is unbiased according to our tests. After 8 Gyr, there is a clear decline in [Fe/H], such that the older stars are progressively more metal-poor. Our result does not support the analysis by Casagrande et al. (2011b), which is based on the photometric metallicities and ages of stars in the
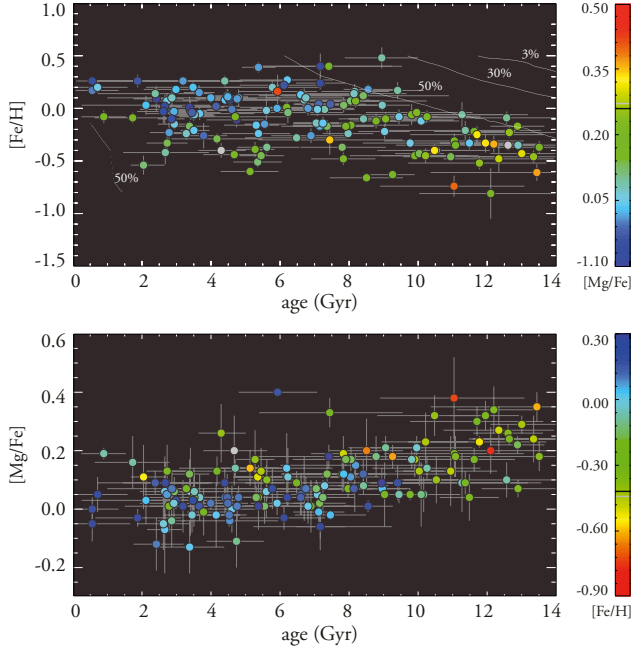
Fig. 6. *Top*: age-metallicity plot for the Milky Way disk. The contours indicate the relative sample completeness, i.e. the percentage of stars that would remain in the sample due to the *Gaia*-ESO survey selection functions, i.e. IR magnitude and colour cuts, and restrictions imposed on stellar parameters. Here the magnitude cuts refer to the distance of 1 kpc. For clarity, this fraction was normalised to its peak value. *Bottom*: the distribution of stars in the [Mg/Fe] – age plane.

Geneva-Copenhagen Survey. The authors find no age-metallicity relation; the stars are homogeneously distributed in metallicity in any age bin up to 12 Gyr (Casagrande et al. 2011b, their Fig. 16). While qualitatively, the mean metallicity of the sample for old ages could be affected by our sampling bias against old and metal-rich stars, Fig. 6 shows that the suppression relative to the most populated part of the plot is not larger than 50−70%. The fact that no metal-rich star is observed with age >10 Gyr may indicate that such stars are rare, if they exist at all in the solar neighbourhood.

Figure 6 (bottom panel) shows [Mg/Fe] ratios as a function of age, colour-coded with [Fe/H]. The oldest stars with ages >12 Gyr show [Mg/Fe], from 0 to 0.4 dex, and a broad range of metallicity, from solar to [Fe/H] ∼ −1. There is little evidence that the relation tightens at ages greater than 9 Gyr in our sample, as advocated e.g. by Haywood et al. (2013) who used a subsample of 363 stars from the Adibekyan et al. (2012) sample of 1111 FGK stars. Bensby et al. (2014) also found a knee at 9 Gyr, with a clear increase in [Mg/Fe] with age (their Fig. 21), albeit with a notably larger scatter at ages >11 Gyr than in Haywood et al. (2013). It is possible that the larger scatter in our sample at old ages is due to the fact that we include stars with relatively large age uncertainties. However, from our analysis of the selection effects, it is to be expected that some fraction of α-poor old stars could be artificially suppressed for the same reasons as discussed above, akin to the [Fe/H] suppression shown in the age-metallicity plot. Regardless of these effects, the trends at old age, as seen by Haywood et al. (2013) and Bensby et al. (2014) fit within our [Mg/Fe]-age relation.

Finally, one interesting feature of the age-metallicity relation deserves a comment. Figure 7 (top panel) shows the observed stars in the age-metallicity plane colour-coded with their $T_{\rm eff}$. The obvious correlation with effective temperature is striking yet
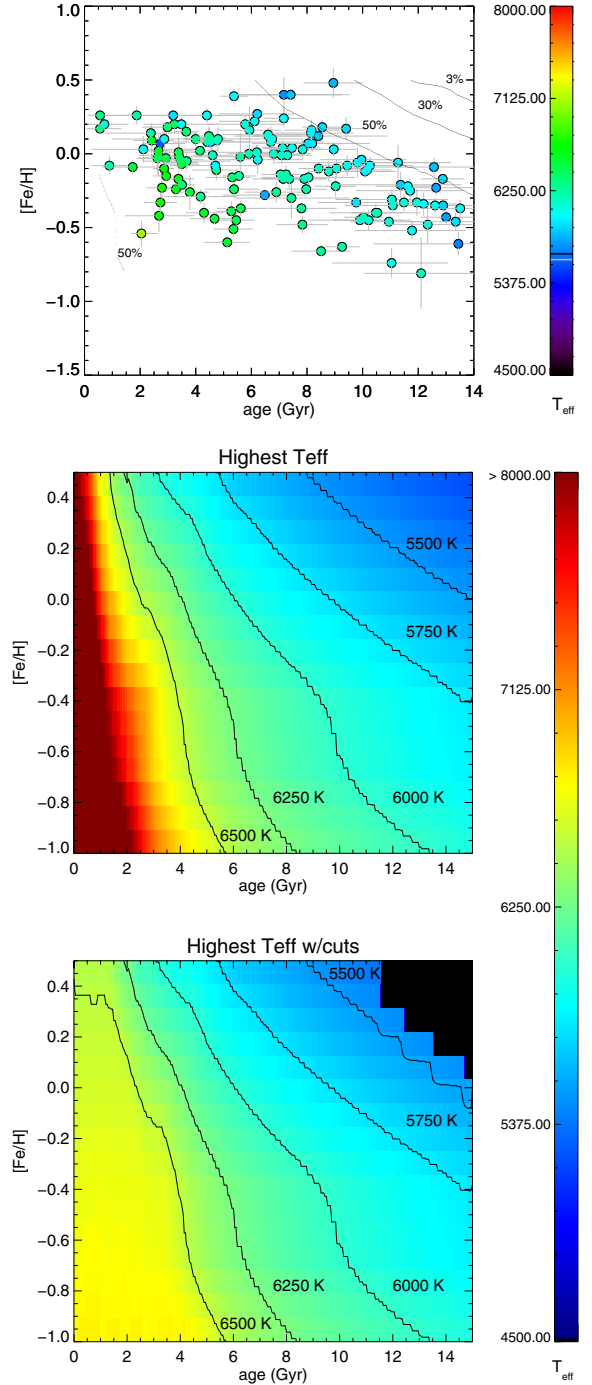
Fig. 7. *Top panel*: observed stars in the age-[Fe/H] plane, colour-coded with their $T_{\rm eff}$. *Bottom two panels*: highest possible $T_{\rm eff}$ for a given age and [Fe/H] obtained from the stellar evolution tracks without (middle) and with (bottom) photometry and $\log g$ cuts. Selected curves of constant $T_{\rm eff}$ are given for reference. For $T_{\rm eff} \leq 6500$ K, the highest "observable" values of $T_{\rm eff}$ do not depend on the cuts imposed on the sample and are simply the result of stellar evolution effects.

it can be easily explained based on the similar considerations as in Sect. 4. In the middle and bottom panel of Fig. 7, we also show the maximum $T_{\rm eff}$ to be expected in the [Fe/H] – age plane based on stellar evolution models, without (middle) and with (bottom) cuts on the photometry and stellar parameters. The colour and $\log g$ cuts affect the bottom left corner of the age-metallicity plot. Young hot stars are suppressed because of the blue colour cut in the *Gaia*-ESO UVES sample selection. However, from the
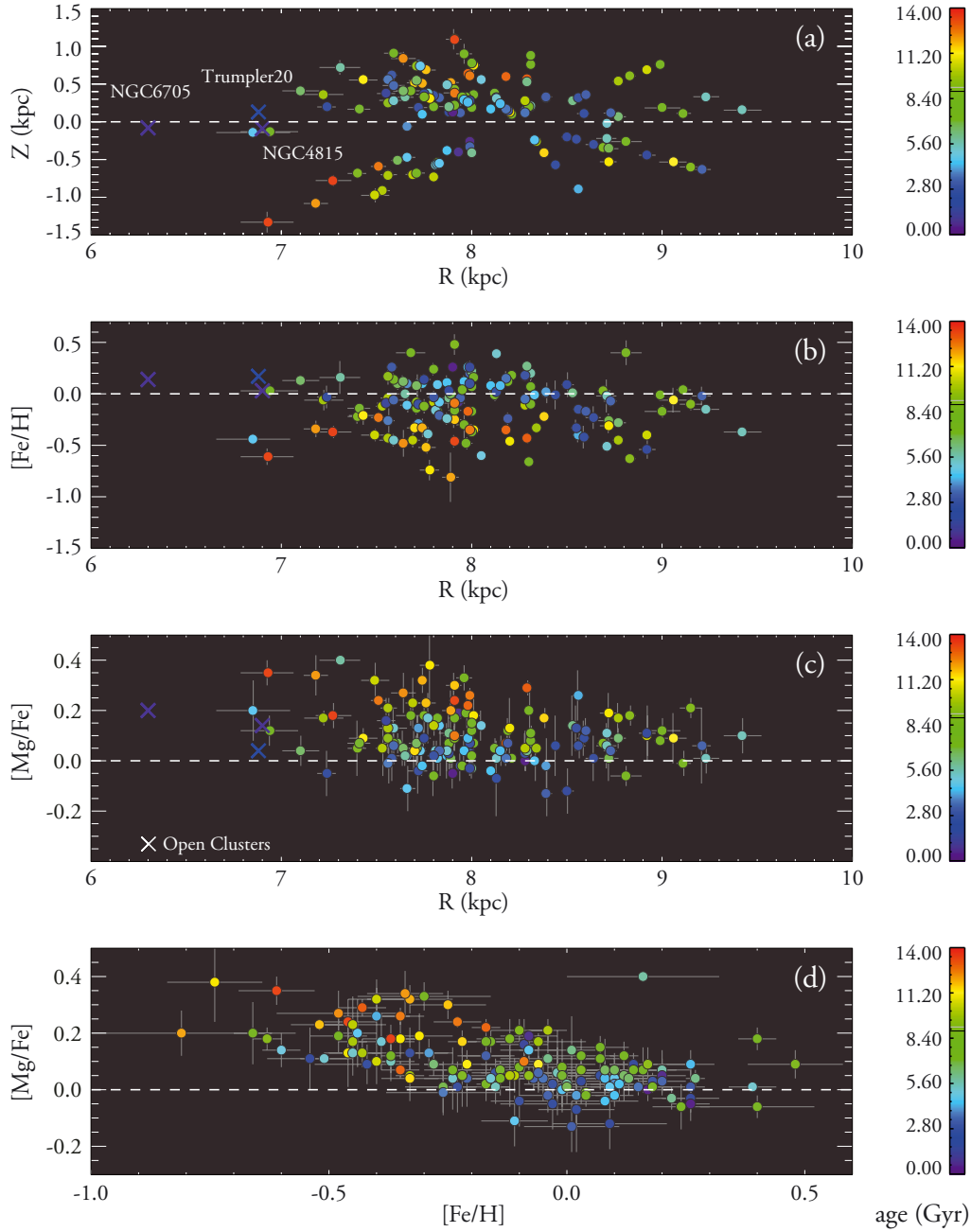
Fig. 8. Final stellar sample. *Top to bottom*: **a)** the distribution of stars in the *Z* vs. *R* plane; **b)** and **c)** radial gradients in the MW disk for [Fe/H] and [Mg/Fe]; **d)** the [Fe/H] − [Mg/Fe] relation. Colour represents age of a star. The three young open clusters, which are available in UVES iDR1, are shown with large crosses.

comparison of these two plots we see that the trend for $T_{\rm eff} \leq$ 6500 K does not depend on selection effects. The segregation of the sample according to the $T_{\rm eff}$ is thus simply the consequence of stellar evolution: the highest "observable" temperature of stars decreases with age and metallicity, and naturally leads to the $T_{\rm eff}$ distribution shown in Fig. 7. Such a trend must be present in any survey, as long as the probed $T_{\rm eff}$ range is not too narrow.

5.2. Abundance gradients as a function of radial and vertical distance

Figure 8 shows how stars are distributed with the vertical distance from the plane |*Z*| and Galactocentric distance *R*, colour-coded according to their ages. The middle panels show the

gradients of [Fe/H] and [Mg/Fe] as a function of *R*; the bottom panel is the [Mg/Fe]-[Fe/H] relation. The three open clusters that are available in the UVES iDR1, NGC 6705, Trumpler 20, and NGC 4815 (Magrini et al. 2014) are also plotted.

There are obvious differences in the age, metallicity, and Mg gradients with *R*. The younger stars with ages ≤7 Gyr show an outward decrease in the mean [Fe/H], slightly stronger compared to the previous studies of young disk populations, i.e. Cepheids, OB stars, and H II regions (Andrievsky et al. 2002a,b; Lemasle et al. 2013). Using only stars close to the plane, within |*Z*| of 300 pc, we derive Δ[Fe/H]$/\Delta R = -0.068 \pm$ 0.014 dex/kpc or Δ[Fe/H]$/\Delta R = -0.076 \pm 0.010$ dex/kpc (stars with ages ≤7 Gyr). Analysing RAVE spectra of disk dwarfs, Boeche et al. (2013) found that Δ[Fe/H]$/\Delta R = -0.065$ dex/kpc

for $Z_{max} < 400$ pc. For $300 < Z_{max} < 800$ pc, we obtain $\Delta[Fe/H]/\Delta R = -0.114 \pm 0.009$ dex/kpc. The steepening of the gradient is surprising given the other studies found flatter or even positive radial metallicity gradients at larger vertical distances from the plane (Allende Prieto et al. 2006; Hayden et al. 2013; Anders et al. 2014; Katz et al. 2011), but this could reflect our small dataset or the fact that the studies probe much larger $|Z|$.

The old stars in our sample tend to have lower [Fe/H] even in the inner disk (Fig. 8b), and our [Fe/H] gradient for old stars with ages above 12 Gyr is $\Delta[Fe/H]/\Delta R = 0.069 \pm 0.044$, positive but with a larger uncertainty and potentially consistent with being flat. However, we also note that this apparent age dependence of the radial metallicity gradient is also related to the fact that we observe older stars at larger distances from the plane. As seen in the top (a) panel of Fig. 8, the majority of old stars are also located at $|Z| > 800$ pc and our sample does not have directional isotropy: we probe larger distances in the inner disk. Therefore, we cannot yet make more definitive statements about the gradients for old stellar population, or at high $|Z|$. Nordström et al. (2004) noted a change of slope of the radial [Fe/H] gradient in the disk. They find a progressively flattening $\delta[Fe/H]/\delta R_{mean}$ with increasing age, which even becomes positive for ages >10 Gyr[4]. It is likely that their results also reflect the effect of observing older stars at larger $|Z|$ from the plane.

The radial [Mg/Fe] gradient is shown in the panel c of Fig. 8 colour-coded with age. For the stars with $|Z| \leq 300$ pc, the radial gradient in the sample mean of the [Mg/Fe] abundance is close to zero, $\Delta[Mg/Fe]/\Delta R = 0.021 \pm 0.014$. However, the gradient becomes negative and steepens with increasing distance from the plane. For the stars with $0.3 < |Z| < 0.8$ kpc in our sample, the most populated bin in our sample, we infer $\Delta[Mg/Fe]/\Delta R = -0.045 \pm 0.011$. It appears that the gradient steepens further at even larger vertical distance, $|Z| > 0.8$ kpc. Moreover, we find that the gradient of [Mg/Fe] becomes negative and very steep for older stars, i.e. $\Delta[Mg/Fe]/\Delta R = 0.015 \pm 0.014$ (ages ≤ 7 Gyr) and $\Delta[Mg/Fe]/\Delta R = -0.071 \pm 0.029$ (ages ≥ 12 Gyr). In words, the outer disk has *less high-Mg and old stars* than the inner disk. These results are very interesting, and they support and extend the results of the earlier studies, including Bensby et al. (2010, 2011) and Bovy et al. (2012c). The former focussed on the inner ($4 < R < 7$ kpc) and outer ($9 < R < 13$ kpc) disk giants. Their main conclusion is that while the inner disk giants chemically behave as the solar neighbourhood stars in terms of the "two" disk components (low [α/Fe] thin-disk stars vs. high [α/Fe] thick-disk stars), the outer disk is only formed of "thin-disk"-like stars. Bovy et al. (2012c) found that the incidence of α-enhanced stars falls off strongly towards the outer disk using G-dwarfs from the SEGUE survey. In this work, we are also able to add, for the first time, ages to the inner-outer disk picture. The scatter of [Fe/H], [Mg/Fe], and ages of stars towards the inner disk is *larger than in the outer disk*. This means that the chemical evolution models for the Galactic disk which predict smaller abundance scatter in the inner disk (Hou et al. 2000) are disfavoured by our results.

The bottom d panel of Fig. 8 shows the [Mg/Fe]-[Fe/H] relation. The results are not un-expected: the solar-metallicity stars are predominantly young. At lower metallicity, we see younger stars with low Mg/Fe abundance ratios, but also very old Mg-enhanced stars. Whether these are two different components, each with its own history, or a continuous distribution of stars cannot be firmly established yet.

---

[4] Nordstrom et al. (2004) studied gradients as a function of the mean radius R_{mean}.

In general, our results may have several interpretations (see Feltzing & Chiba 2013, for a review of disk formation scenarios). Some observational studies (e.g. Jurić et al. 2008) decompose the disk into the thin and thick components, which formed in different episodes of Galaxy formation. Models which assume different evolutionary history of the two disks (e.g. Chiappini et al. 1997, 2001) are broadly consistent with observations. Stars in the thin disk are young <8 Gyr and α-poor, while the thick disk is older, metal-poor and α-enhanced (Bensby et al. 2003; Kordopatis et al. 2011, 2013). In this interpretation, at the smallest distances from the plane, our age-metallicity relation is dominated by the thin disk. In the intermediate vertical distance bin, our data would indicate a presence of both the thin and thick disk.

While the discrete nature of the disk is one possibility, our data can be also explained by other Galaxy formation models, which do not explicitly divide the disk into subcomponents (Hou et al. 2000; Roškar et al. 2008; Rahimi et al. 2013; Kubryk et al. 2013; Minchev et al. 2013, 2014). For example, the semi-analytical model by (Schönrich & Binney 2009b, see their Fig. 6), which assumes radial migration driven by transient spiral arms, is consistent with our data both what concerns the age-metallicity relation shape and the [Fe/H] scatter at a given age. In accord with our results, this model also predicts the apparent bimodality in the [Fe/H]-[α/Fe] plane, just as we observe (Fig. 8, bottom panel). This effect is not a consequence of a star formation hiatus (as e.g. in Chiappini et al. 2001), but the 1 Gyr timescale of SN Ia coupled with secular heating and churning (Schönrich & Binney 2009a, a simple analytical explanation is given in their Appendix B). There exist other types of models that also show consistencies with our data. For example, hydrodynamical N-body simulations (Brook et al. 2012; Minchev et al. 2013) form a thick-disk like component through early gas-rich mergers and radial migration driven by external and internal perturbations. The most recent simulations by Minchev et al. (2014, their Fig. 9) predict that mono-age populations have radial abundance gradients that do not change with vertical distance from the plane. The change of radial gradients with $|Z|$ is thus caused by a different mix of stellar ages at different altitudes. We cannot yet test this hypothesis because of the sparse sampling of stars. However, we confirm that the positive gradient in [Mg/Fe] at low $|Z|$ changes to negative at larger vertical distance from the plane, we also observe a similar structure in the mono-age bins. To put more quantitative constraints on their prediction, more data is needed. Therefore we leave this analysis for the next paper, which will include a much larger number of stars from the *Gaia*-ESO DR2 dataset.

6. Summary

The main goal of our work is to study the relationship between age and metallicity and spatial distribution of stars in the Galactic disk from observational perspective.

The stellar sample includes several hundreds of stars observed at high-resolution (with the UVES instrument at VLT) within the first six months of the *Gaia*-ESO survey. Stellar parameters and abundances were determined with 1D local thermodynamic equilibrium (LTE) MARCS model atmospheres, by combining the results obtained by several different spectroscopic techniques. We then derived the mass and age of each star, as well as its distance from the Sun, using the Bayesian method developed in Serenelli et al. (2013). The results were validated against astero-seismology and the IRFM method. We took special care to quantify the survey selection effects in the

interpretation of the final parameter distributions. For this, we performed a series of tests using a mock stellar sample generated from theoretical stellar models, which was subject to various cuts in the space of observables. The outcome of these tests allowed us to define the optimal dataset to address the science problems. In the final sample, we have 144 stars with a wide range of ages (0.5−13.5 Gyr), Galactocentric distances from 6 kpc to 9 kpc, and vertical distances from the plane $0 < |Z| < 1.5$ kpc. The sample is not large, however it should be noted that the study is exploratory, i.e. laying out the method and the properties of the observed field sample of stars in the *Gaia*-ESO Survey. At the end of the survey we expect to have a one order of magnitude increase in the stellar sample.

Our findings can be summarised as follows.

First, the observed age-metallicity relation is fairly flat in range of ages between 0 and 8 Gyr (or equivalently in the young disk, for stars close to the plane). We also confirm the previous results that there is a significant scatter of metallicity at any age. A steep decline in [Fe/H] is seen for stars with ages above 9 Gyr. The colour and magnitude cuts on the survey suppress very old metal-rich stars and young metal-poor stars. However, the suppression relative to the most populated locus on the age-metallicity relation is not larger than 50−70%. In our sample, no solar-metallicity star is observed with age >10 Gyr. This may indicate that such stars are rare, if they exist at all in the solar neighbourhood.

[Mg/Fe] ratios correlate with age, such that α-rich stars are older, but the dispersion of [Mg/Fe] abundances is not small at any age. In particular, the old stars with ages above 9 Gyr have a range of α-enhancement, from 0 to 0.4 dex, and metallicity from solar to [Fe/H] ∼ −1. This contrasts with the very tight correlation of [Mg/Fe] and ages for old stars as suggested by Haywood et al. (2013). However, the trends found by Haywood et al. (2013) and Bensby et al. (2014) overlap with our [Mg/Fe]-age relation.

We also show, in agreement with earlier observational and theoretical work, that the radial abundance gradients (Fe, Mg) change as a function of vertical distance from the plane, or the mean age of a population. At $|Z| \leq 300$ pc, we find $\Delta[\text{Fe/H}]/\Delta R = -0.068 \pm 0.014$ and $\Delta[\text{Mg/Fe}]/\Delta R = 0.021 \pm 0.014$. For the most populated $|Z|$ bin in our sample, $300 \leq |Z| \leq 800$ pc we infer $\Delta[\text{Mg/Fe}]/\Delta R = -0.045 \pm 0.011$, and $\Delta[\text{Fe/H}]/\Delta R = -0.114 \pm 0.009$. The picture is not too different when separating stars according to their age: the gradient of [Mg/Fe] is close to zero for younger stars, but becomes negative and very steep for the older stellar population, i.e. $\Delta[\text{Mg/Fe}]/\Delta R = 0.015 \pm 0.014$ (ages ≤7 Gyr) and $\Delta[\text{Mg/Fe}]/\Delta R = -0.071 \pm 0.029$ (ages ≥12 Gyr). The anisotropic distribution of stars in our sample, i.e. larger fraction of stars observed in the inner disk, should be kept in mind when interpreting these gradients.

To summarise, perhaps the most important result of this analysis is how the properties of the dominant stellar populations in the disk change as we probe stars with different ages, distances, Mg abundances, and metallicities. We find more older, α-rich, and metal-poor stars in the inner disk. In the outer disk, stars are on average younger and α-poor. Although our current sample is small, our results lend support to current pictures of the formation of the Galactic disk, such as the inside-out formation. This initial dataset illustrates the potential of the *Gaia*-ESO Survey. With upcoming data releases, we will have a large database of spectra that have been reduced and analysed in a systematic and homogeneous manner. We will thus be able to distinguish among the many theories of the formation of the disk of the Milky Way.

Appendix A: Tests

To check for systematic biases caused by the method itself, we performed Monte-Carlo simulations by generating a random sample of synthetic stars assuming a Salpeter IMF, a constant SFR, and a uniform [Fe/H] distribution over time. The sample is restricted to the low-mass range $0.7 \leq M/M_\odot \leq 1.5$ and $-2.2 \leq [\text{Fe/H}] \leq 0.4$, representative of the *Gaia*-ESO survey UVES sample. Mass and [Fe/H] are discretised to match the values present in the grid. Along each track, T_{eff} and $\log g$ are interpolated to the age of the synthetic star. The synthetic stars were then assigned individual uncertainty values, comprising two sources. The random uncertainty component was drawn from the empirical probability distributions constructed from the uncertainty distribution in the actual UVES sample. In addition, we accounted for the possibility of a systematic error. The central values of the simulated data were perturbed by introducing noise drawn from uniform distributions in the range $(-100, 100)$ K, $(-0.1, 0.1)$ dex and $(-0.05, 0.05)$ dex for T_{eff}, $\log g$, and [Fe/H] respectively. The tests showed that the age estimate is robust if the uncertainty defined as described above is smaller than ∼30% or than 2 Gyr. The latter condition is relevant for younger stars that naturally have larger fractional errors.

Figure A.1 (top panels) shows the results for the stars in the synthetic samples that fulfil the age uncertainty criteria. The top and bottom panels correspond to $4.0 \leq \log g \leq 4.6$ and $3.5 \leq \log g < 4.0$, respectively. The quantity u_i is our measure of bias and dispersion, and it is defined as $u_i = (X_{\text{B},i} - X_{\text{in},i})/\sigma_{\text{X},i}$, where X_i denotes the mass or age of a synthetic star i and $\sigma_{\text{X},i}$ the uncertainty returned by BeSPP. The Gaussian fit to the distribution is overplotted in red. The average difference and dispersion are also quoted. We find that for subgiant stars, $3.5 \leq \log g < 4.0$, systematics are virtually absent and the dispersion is well below unity, a result that probably derives from the faster evolution of stars in this phase. In the range $4.0 \leq \log g \leq 4.6$ the method tends to overestimate masses and to underestimate ages by 44% and 36% of the mean uncertainty, respectively. The dispersion is very close to unity, consistent with the expected distribution of errors. The systematic uncertainties introduced by BeSPP in the determination of the age of dwarf stars will be typically 1.3 Gyr for a 12 Gyr old star, down to ∼0.5 Gyr for a 5 Gyr star.

One may also question whether there are systematic differences in the parameter scales between theory and observations. For example, the treatment of convection and surface boundary effects in stellar evolution models may lead to a difference in the T_{eff} scales compared to spectroscopic predictions. Likewise, systematic uncertainties in stellar parameters may cause serious offsets in surface vs. stellar evolution scales. To test this, we compared the stellar parameters (T_{eff}, $\log g$, [Fe/H]) recovered

with independent values derived from astero-seismic analysis (available for 12 stars). This comparison revealed a mean off-set between BeSPP and astero-seismology of $0.07 \pm 0.08\ M_\odot$. The agreement is satisfactory, and it confirms the accuracy of the algorithm that we use for the age and mass calculations.

Appendix B: Star formation rate

We follow the recipe of Schoenrich & Bergemann (2013), where the SFR is assumed proportional to

$$P(\tau, [\text{Fe/H}]) = \begin{cases} 0 & \text{if} & \tau > 15\ \text{Gyr} \\ 1 & \text{if} & 11\ \text{Gyr} \le \tau \le 15\ \text{Gyr} \\ \exp\left(\frac{\tau - 11\ \text{Gyr}}{\sigma_\tau}\right) & \text{if} & \tau < 11\ \text{Gyr} \end{cases} \quad \text{(B.1)}$$

where

$$\sigma_\tau = \begin{cases} 1.5\ \text{Gyr} & \text{if} & [\text{Fe/H}] < -0.9 \\ 1.5 + 7.5\frac{[\text{Fe/H}]+0.9}{0.4}\ \text{Gyr} & \text{if} & -0.9 \le [\text{Fe/H}] \le -0.5 \\ 9\ \text{Gyr} & \text{if} & -0.5 < [\text{Fe/H}]. \end{cases}$$

References

Adelberger, E. G., García, A., Robertson, R. G. H., et al. 2011, Rev. Mod. Phys., 83, 195
Adibekyan, V. Z., Sousa, S. G., Santos, N. C., et al. 2012, A&A, 545, A32
Allende Prieto, C., & Lambert, D. L. 1999, A&A, 352, 555
Allende Prieto, C., Beers, T. C., Wilhelm, R., et al. 2006, ApJ, 636, 804
Anders, F., Chiappini, C., Santiago, B. X., et al. 2014, A&A, 564, A115
Andrievsky, S. M., Bersier, D., Kovtyukh, V. V., et al. 2002a, A&A, 384, 140
Andrievsky, S. M., Kovtyukh, V. V., Luck, R. E., et al. 2002b, A&A, 381, 32
Bensby, T., Feltzing, S., & Lundström, I. 2003, A&A, 410, 527
Bensby, T., Alves-Brito, A., Oey, M. S., Yong, D., & Meléndez, J. 2010, A&A, 516, L13
Bensby, T., Alves-Brito, A., Oey, M. S., Yong, D., & Meléndez, J. 2011, ApJ, 735, L46
Bensby, T., Feltzing, S., & Oey, M. S. 2014, A&A, 562, A71
Bergemann, M., Lind, K., Collet, R., Magic, Z., & Asplund, M. 2012, MNRAS, 427, 27
Blackwell, D. E., Petford, A. D., & Shallis, M. J. 1980, A&A, 82, 249
Blanco-Cuaresma, S., Soubiran, C., Jofré, P., & Heiter, U. 2014, A&A, in press, DOI: 10.1051/0004-6361/201323153
Boeche, C., Siebert, A., Piffl, T., et al. 2013, A&A, 559, A59
Bovy, J., Rix, H.-W., & Hogg, D. W. 2012a, ApJ, 751, 131
Bovy, J., Rix, H.-W., Hogg, D. W., et al. 2012b, ApJ, 755, 115
Bovy, J., Rix, H.-W., Liu, C., et al. 2012c, ApJ, 753, 148
Brook, C. B., Kawata, D., Gibson, B. K., & Freeman, K. C. 2004, ApJ, 612, 894
Brook, C. B., Gibson, B. K., Martel, H., & Kawata, D. 2005, ApJ, 630, 298
Brook, C. B., Stinson, G. S., Gibson, B. K., et al. 2012, MNRAS, 426, 690
Burkert, A., Truran, J. W., & Hensler, G. 1992, ApJ, 391, 651
Casagrande, L., Ramirez, I., Melendez, J., Bessell, M., & Asplund, M. 2010, A&A, 512, A54
Casagrande, L., Schoenrich, R., Asplund, M., et al. 2011a, VizieR On-line Data Catalog: J/A+A/530/A138
Casagrande, L., Schönrich, R., Asplund, M., et al. 2011b, A&A, 530, A138
Cassisi, S., Salaris, M., & Irwin, A. W. 2003, ApJ, 588, 862
Cheng, J. Y., Rockosi, C. M., Morrison, H. L., et al. 2012, ApJ, 752, 51
Chiappini, C., Matteucci, F., & Gratton, R. 1997, ApJ, 477, 765
Chiappini, C., Matteucci, F., & Romano, D. 2001, ApJ, 554, 1044
Edvardsson, B., Andersen, J., Gustafsson, B., et al. 1993, A&A, 275, 101
Feltzing, S., & Chiba, M. 2013, New Astron. Rev., 57, 80
Feltzing, S., Holmberg, J., & Hurley, J. R. 2001, A&A, 377, 911
Freytag, B., Ludwig, H.-G., & Steffen, M. 1996, A&A, 313, 497
Friel, E. D., Janes, K. A., Tavarez, M., et al. 2002, AJ, 124, 2693
Fuhrmann, K. 2008, MNRAS, 384, 173
Gilmore, G., Randich, S., Asplund, M., et al. 2012, The Messenger, 147, 25
González Hernández, J. I., & Bonifacio, P. 2009, A&A, 497, 497
Gustafsson, B., Edvardsson, B., Eriksson, K., et al. 2008, A&A, 486, 951
Hayden, M. R., Holtzman, J. A., Bovy, J., et al. 2013, AJ, submitted [arXiv:1311.4569]
Haywood, M., Di Matteo, P., Lehnert, M., Katz, D., & Gomez, A. 2013, A&A, 560, A109
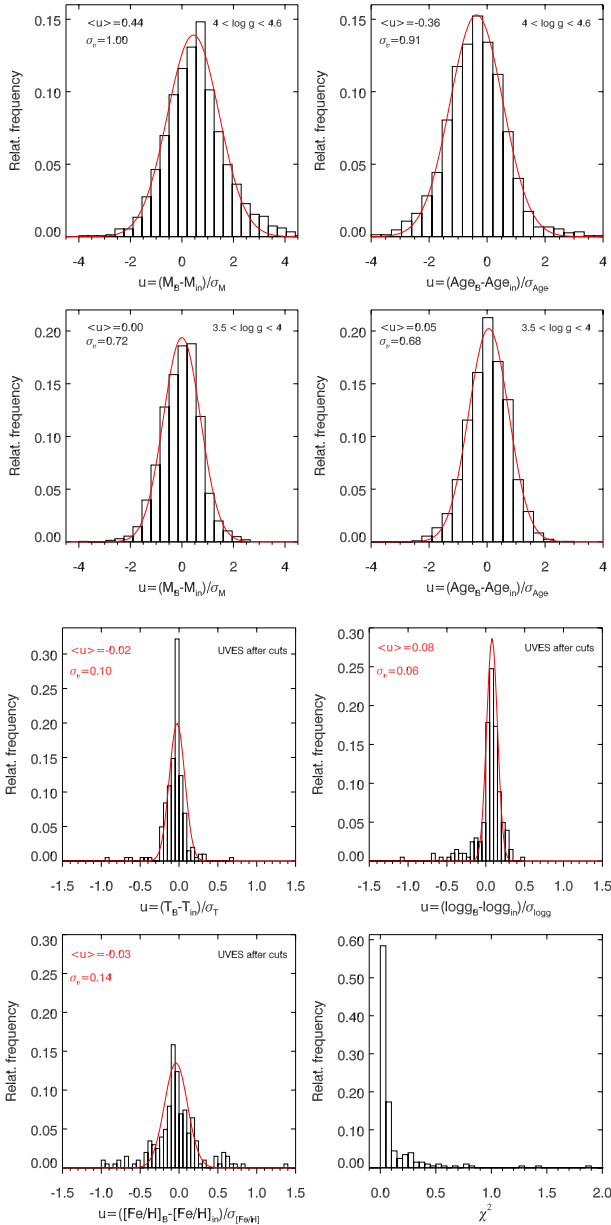
Fig. A.1. *Top four panels*: stellar parameters recovered by BeSPP for the ideal synthetic stellar sample drawn from the tracks (see text). *Bottom four panels*: stellar parameters (T_{eff}, $\log g$, [Fe/H]) recovered by BeSPP compared with the input spectroscopic values for the UVES stars. The total χ^2 was computed summing over the three spectroscopic parameters and using only the observational errors.

by BeSPP with the input spectroscopic values (Fig. A.1, bottom panel). The left-bottom panel shows the total χ^2 computed summing over the three spectroscopic parameters and using only the observational errors. The agreement is excellent, with dispersion values σ of the order of 0.1, ruling out the possibility of a systematic mismatch between the observed data and stellar models.

Another method to validate the accuracy of the Bayesian results is to compare them with that obtained by independent methods, e.g. from astero-seismology. The number of stars observed in iDR1 is too small to draw rigorous conclusions. While there is no independent information for the majority of the UVES stars, we can use the library of FGK Benchmark stars, which was built for the *Gaia*-ESO survey (Jofre et al. 2014; Blanco-Cuaresma et al. 2014; Heiter et al., in prep.). We have compared the masses from BeSPP for the reference parameters

Heiter, U., Soubiran, C., Netopil, M., & Paunzen, E. 2014, A&A, 561, A93

Holmberg, J., Nordström, B., & Andersen, J. 2007, A&A, 475, 519

Holmberg, J., Nordström, B., & Andersen, J. 2009, A&A, 501, 941

Hou, J. L., Prantzos, N., & Boissier, S. 2000, A&A, 362, 921

Iglesias, C. A., & Rogers, F. J. 1996, ApJ, 464, 943

Jofre, P., Heiter, U., Soubiran, C., et al. 2014, A&A, 564, A133

Jørgensen, B. R., & Lindegren, L. 2005, A&A, 436, 127

Jurić, M., Ivezić, Ž., Brooks, A., et al. 2008, ApJ, 673, 864

Katz, D., Soubiran, C., Cayrel, R., et al. 2011, A&A, 525, A90

Kordopatis, G., Recio-Blanco, A., de Laverny, P., et al. 2011, A&A, 535, A107

Kordopatis, G., Gilmore, G., Wyse, R. F. G., et al. 2013, MNRAS, 436, 3231

Kubryk, M., Prantzos, N., & Athanassoula, E. 2013, MNRAS, 436, 1479

Lee, Y. S., Beers, T. C., An, D., et al. 2011, ApJ, 738, 187

Lemasle, B., François, P., Genovali, K., et al. 2013, A&A, 558, A31

Lépine, J. R. D., Cruz, P., Scarano, S. J., et al. 2011, MNRAS, 417, 698

Lind, K., Bergemann, M., & Asplund, M. 2012, MNRAS, 427, 50

Maciel, W. J., Costa, R. D. D., & Idiart, T. E. P. 2010, A&A, 512, A19

Magrini, L., Randich, S., Romano, D., et al. 2014, A&A, 563, A44

Minchev, I., Chiappini, C., & Martig, M. 2013, A&A, 558, A9

Minchev, I., Chiappini, C., & Martig, M. 2014, ArXiv e-prints

Naab, T., & Ostriker, J. P. 2006, MNRAS, 366, 899

Nordström, B., Mayor, M., Andersen, J., et al. 2004, A&A, 418, 989

Pagel, B. E. J., & Tautvaisiene, G. 1995, MNRAS, 276, 505

Pont, F., & Eyer, L. 2004, MNRAS, 351, 487

Rahimi, A., Carrell, K., & Kawata, D. 2013 [arXiv:1308.2061]

Randich, S., Gilmore, G., & Gaia-ESO Consortium 2013, The Messenger, 154, 47

Rix, H.-W., & Bovy, J. 2013, A&ARv, 21, 61

Roškar, R., Debattista, V. P., Quinn, T. R., Stinson, G. S., & Wadsley, J. 2008, ApJ, 684, L79

Ruchti, G. R., Fulbright, J. P., Wyse, R. F. G., et al. 2011, ApJ, 737, 9

Schlegel, D. J., Finkbeiner, D. P., & Davis, M. 1998, ApJ, 500, 525

Schlesinger, K. J., Johnson, J. A., Rockosi, C. M., et al. 2012, ApJ, 761, 160

Schoenrich, R., & Bergemann, M. 2013, MNRAS, submitted [arXiv:1311.5558]

Schönrich, R., & Binney, J. 2009a, MNRAS, 396, 203

Schönrich, R., & Binney, J. 2009b, MNRAS, 399, 1145

Serenelli, A. M., Bergemann, M., Ruchti, G., & Casagrande, L. 2013, MNRAS, 429, 3645

Shimanskaya, N. N., Mashonkina, L. I., & Sakhibullin, N. A. 2000, Astron. Rep., 44, 530

Stanghellini, L., & Haywood, M. 2010, ApJ, 714, 1096

Valenti, J. A., & Piskunov, N. 1996, A&AS, 118, 595

Velazquez, H., & White, S. D. M. 1999, MNRAS, 304, 254

Weiss, A., & Schlattl, H. 2008, Ap&SS, 316, 99

Yong, D., Carney, B. W., & Friel, E. D. 2012, AJ, 144, 95

[1] Institute of Astronomy, University of Cambridge, Madingley Road, CB3 0HA, Cambridge, UK
e-mail: mbergema@ast.cam.ac.uk

[2] Lund Observatory, Department of Astronomy and Theoretical Physics, Box 43, 221 00 Lund, Sweden

[3] Institute of Space Sciences (IEEC-CSIC), Campus UAB, Fac. Ciències, Torre C5 parell 2, 08193 Bellaterra, Spain

[4] Research School of Astronomy & Astrophysics, Mount Stromlo Observatory, The Australian National University, ACT 2611, Australia

[5] Department of Physics and Astronomy, Division of Astronomy and Space Physics, Ångström laboratory, Uppsala University, Box 516, 75120 Uppsala, Sweden

[6] INAF – Osservatorio Astrofisico di Arcetri, Largo E. Fermi 5, 50125 Florence, Italy

[7] Instituto de Astrofísica de Canarias, 38205 La Laguna, Tenerife, Spain

[8] Moscow MV Lomonosov State University, Sternberg Astronomical Institute, 119992 Moscow, Russia

[9] ASI Science Data Center, via del Politecnico SNC, 00133 Roma, Italy

[10] Institute of Astronomy, University of Edinburgh, Blackford Hill, Edinburgh EH9 3HJ, UK

[11] Astrophysics Group, Research Institute for the Environment, Physical Sciences and Applied Mathematics, Keele University, Keele, Staffordshire ST5 5BG, UK

[12] INAF – Osservatorio Astronomico di Palermo, Piazza del Parlamento 1, 90134, Palermo, Italy

[13] Departamento de Física, Ingeniería de Sistemas y Teoría de la Señal, Universidad de Alicante, Apdo. 99, 03080, Alicante, Spain

[14] ESA, ESTEC, Keplerlaan 1, PO Box 299, 2200 AG Noordwijk, The Netherlands

[15] Max-Planck Institut für Astronomie, Königstuhl 17, 69117 Heidelberg, Germany

[16] INAF – Padova Observatory, Vicolo dell'Osservatorio 5, 35122 Padova, Italy

[17] INAF – Osservatorio Astronomico di Bologna, via Ranzani 1, 40127 Bologna, Italy

[18] Laboratoire Lagrange (UMR7293), Université de Nice Sophia Antipolis, CNRS, Observatoire de la Côte d'Azur, CS 34229, 06304 Nice Cedex 4, France

[19] Department for Astrophysics, Nicolaus Copernicus Astronomical Center, ul. Rabiańska 8, 87-100 Toruń, Poland

[20] European Southern Observatory, Karl-Schwarzschild-Str. 2, 85748 Garching bei München, Germany

[21] Instituto de Astrofísica de Andalucía-CSIC, Apdo. 3004, 18080 Granada, Spain

[22] Institute of Theoretical Physics and Astronomy, Vilnius University, Goštauto 12, 01108 Vilnius, Lithuania

[23] Instituto de Fisica, Universidade Federal do Rio Grande do Sul, Av. Bento Goncalves 9500, Porto Alegre, RS, Brazil

[24] Dipartimento di Fisica e Astronomia, Sezione Astrofisica, Universitá di Catania, via S. Sofia 78, 95123 Catania, Italy