

# Effects of the selection function on metallicity trends in spectroscopic surveys of the Milky Way

G. Nandakumar<sup>1</sup>, M. Schultheis<sup>1</sup>, M. Hayden<sup>1</sup>, A. Rojas-Arriagada<sup>1,2,3</sup>, G. Kordopatis<sup>1</sup>, and M. Haywood<sup>4</sup>

<sup>1</sup> Laboratoire Lagrange, Université Côte d'Azur, Observatoire de la Côte d'Azur, CNRS, Blvd de l'Observatoire, 06304 Nice, France  
e-mail: govind.nandakumar@oca.eu

<sup>2</sup> Instituto de Astrofísica, Facultad de Física, Pontificia Universidad Católica de Chile, Av. Vicuña Mackenna 4860, Casilla 306, Santiago, Chile

<sup>3</sup> Millennium Institute of Astrophysics, Av. Vicuña Mackenna 4860, 782-0436 Macul, Santiago, Chile

<sup>4</sup> GEPI, Observatoire de Paris, PSL Research University, CNRS, Univ. Paris Diderot, Sorbonne Paris Cité, Place Jules Janssen, 92195 Meudon, France

Received 4 May 2017 / Accepted 18 July 2017

## ABSTRACT

**Context.** Large spectroscopic Galactic surveys imply a selection function in the way they performed their target selection.

**Aims.** We investigate here the effect of the selection function on the metallicity distribution function (MDF) and on the vertical metallicity gradient by studying similar lines of sight using four different spectroscopic surveys (APOGEE, LAMOST, RAVE, and *Gaia*-ESO), which have different targeting strategies and therefore different selection functions.

**Methods.** We use common fields between the spectroscopic surveys of APOGEE, LAMOST, RAVE (ALR) and APOGEE, RAVE, *Gaia*-ESO (AGR) and use two stellar population synthesis models, GALAXIA and TRILEGAL, to create mock fields for each survey. We apply the selection function in the form of colour and magnitude cuts of the respective survey to the mock fields to replicate the observed source sample. We make a basic comparison between the models to check which best reproduces the observed sample distribution. We carry out a quantitative comparison between the synthetic MDF from the mock catalogues using both models to understand the effect of the selection function on the MDF and on the vertical metallicity gradient.

**Results.** Using both models, we find a negligible effect of the selection function on the MDF for APOGEE, LAMOST, and RAVE. We find a negligible selection function effect on the vertical metallicity gradients as well, though GALAXIA and TRILEGAL have steeper and shallower slopes, respectively, than the observed gradient. After applying correction terms on the metallicities of RAVE and LAMOST with respect to our reference APOGEE sample, our observed vertical metallicity gradients between the four surveys are consistent within  $1\sigma$ . We also find consistent gradient for the combined sample of all surveys in ALR and AGR. We estimated a mean vertical metallicity gradient of  $-0.241 \pm 0.028$  dex  $\text{kpc}^{-1}$ . There is a significant scatter in the estimated gradients in the literature, but our estimates are within their ranges.

**Conclusions.** We have shown that there is a negligible selection function effect on the MDF and the vertical metallicity gradients for APOGEE, RAVE, and LAMOST using two stellar population synthesis models. Therefore, it is indeed possible to combine common fields of different surveys in studies using MDF and metallicity gradients provided their metallicities are brought to the same scale.

**Key words.** solar neighborhood – Galaxy: evolution – Galaxy: disk – Galaxy: abundances – Galaxy: stellar content – Galaxy: general

## 1. Introduction

The Milky Way is the primary laboratory where we can obtain the detailed chemical, kinematic, and spatial distribution of individual stars that make up the different components (thin and thick disc, bulge, halo) of a typical spiral galaxy. Stellar atmospheres retain the composition of chemical elements present in the interstellar medium at the time and place of their formation. Thus tracing chemical abundances of individual stars combined with their kinematic properties and current phase-space location helps us to model and test various formation and evolution scenarios of the Milky Way components to which they belong.

Metallicity is a crucial ingredient used to decipher the Milky Way's chemical history (e.g. Freeman & Bland-Hawthorn 2002; Ivezić et al. 2012). The mean metallicity of stellar populations is found to vary with galactocentric radius and height from the Galactic mid-plane (Hayden et al. 2015; Schlesinger et al. 2012). The radial and vertical metallicity gradients observed in the Milky Way (Hartkopf & Yoss 1982; Kordopatis et al. 2011b; Cheng et al. 2012; Hayden et al. 2014; Mikolaitis et al. 2014; Cunha et al. 2016; Jacobson et al. 2016) are strong signatures

of formation and evolution of the different substructures of the Milky Way. Different disc formation scenarios explaining the observed abundance distribution are proposed: a radial gradient may result from the inside-out disc formation and be partially erased by radial mixing, while vertical gradients can be generated via disc heating by spiral arm interaction and mergers (see Quinn et al. 1993; Chiappini et al. 2001; Schönrich & Binney 2009a,b; Rix & Bovy 2013; Mikolaitis et al. 2014, etc.).

Tracers like open clusters, HII regions, Cepheid variables, FGK dwarfs, planetary nebulae, and red giant field stars have been used to determine the radial/vertical gradients (Chen et al. 2003; Costa et al. 2004; Yong et al. 2006; Maciel & Costa 2006; Maciel & Costa 2010; Stanghellini & Haywood 2010; Balser et al. 2011; Kordopatis et al. 2011b; Carrell et al. 2012; Frinchaboy et al. 2013; Gazzano et al. 2013; Bergemann et al. 2014; Hayden et al. 2014, etc.). Generally, these studies yield negative slopes for both radial and vertical gradients. However, there is a significant scatter among the estimated gradients. The radial gradient is found to vary from  $-0.028$  dex  $\text{kpc}^{-1}$  (Mikolaitis et al. 2014, for thin disc stars) to  $-0.17$  dex  $\text{kpc}^{-1}$  (Sestito et al. 2008, using open clusters) in the inner disc

(Galactocentric radius,  $R < 11$  kpc) close to the Galactic mid-plane. The radial gradient is found to get shallower and become positive as we move away from the plane (Boeche et al. 2013, 2014; Hayden et al. 2014). Similarly, a large dispersion in the estimated vertical metallicity gradient is found over the years, ranging from  $-0.112 \text{ dex kpc}^{-1}$  (Boeche et al. 2014) to  $-0.31 \text{ dex kpc}^{-1}$  (Soubiran et al. 2008; Hayden et al. 2014) for stars in the solar neighbourhood ( $7 < R < 9$  kpc and  $|Z| < 2$  kpc). This large uncertainty of the observed metallicity gradient makes it difficult to constrain chemo-dynamical evolution models of the Milky Way (e.g. Hou et al. 2000; Chiappini et al. 2001; Snaith et al. 2015). Clearly there is a pressing need on the observational side to reduce the uncertainty of this fundamental parameter for these models.

During the last decade, the number of low, medium, and high resolution spectroscopic surveys of stellar populations in our Galaxy have increased drastically (Wyse 2016). There are several multi-object spectroscopic surveys that have been completed or are underway, such as the RAdial Velocity Experiment (RAVE; Steinmetz et al. 2006), the LAMOST Experiment for Galactic Understanding and Exploration (LEGUE; Deng et al. 2012), the Apache Point Observatory Galactic Evolution Experiment (APOGEE; Majewski et al. 2017), the *Gaia*-ESO survey (GES; Gilmore et al. 2012), and the Sloan Extension for Galactic Understanding and Exploration (SEGUE; Yanny et al. 2009). They differ in spectral resolution, wavelength coverage, and in their selected targets (giant stars, dwarf stars, clusters, etc.) based on their science goals. These unique target selection schemes can lead to biases in which stellar populations are observed, and affect measurements of the observed properties of the Milky Way; these targeting biases are known as the selection function. The selection function is defined as the fraction of objects in a certain colour and magnitude range successfully observed spectroscopically compared to the underlying stellar populations, and determines how representative the observed sample is compared to the full existing stellar population of the Milky Way.

The target selection schemes limit the coverage of parameter space of  $T_{\text{eff}}$ ,  $\log g$ , and  $[\text{Fe}/\text{H}]$  that could potentially lead to biases while carrying out analyses that measure the gradients and metallicity distributions of certain stellar populations. Cheng et al. (2012) and Schlesinger et al. (2012) used different weighting schemes to correct for the metallicity bias introduced by the target selection in their sample of SEGUE main-sequence turn-off stars, and G and K dwarf stars, respectively. Meanwhile, Bovy et al. (2012) determined a plate-dependent selection function for a G-dwarf sample along  $\sim 150$  lines of sight in SEGUE using the dereddened colour-magnitude boxes. The selection effects in the APOGEE red clump (RC) sample is discussed in Bovy et al. (2014) and Nidever et al. (2014) using the much simpler and well-defined target selection algorithm of APOGEE (Zasowski et al. 2013). Sharma et al. (2014) constrained the kinematic parameters of the Milky Way disc using stars from RAVE and the Geneva-Copenhagen Survey (GCS; Nordström et al. 2004) using kinematic analytic models. The RAVE selection function was taken into account while modelling using GALAXIA (Sharma et al. 2011). On comparing the temperature and colour distributions of RAVE stars with that predicted using GALAXIA, they found a reasonably good match except for  $J - K$  colours for stars in the low latitude fields. An extinction correction was performed to correct this. Wojno et al. (2017) described the RAVE selection function in detail and studied the selection function effect on the RAVE metallicity and velocity parameters. For this, they created a mock-RAVE catalogue using the GALAXIA stellar population synthesis model.

They found that RAVE stars do not show any selection effects in terms of kinematics and metallicities using the selection cuts in magnitude and colour of RAVE. Anders et al. (2016) created a mock sample of more than 600 solar-like oscillating red giant stars observed by both CoRoT and APOGEE based on their selection functions. They found some small systematic biases of  $\pm 0.02$  dex in the radial gradient, most notably in the age bin 2–4 Gyr. Recently there have been many more attempts to provide a detailed description of the selection function for other major spectroscopic surveys (e.g. GES: Stonkutė et al. 2016; LAMOST: Yuan et al. 2015, Carlin et al. 2012).

In this paper, we study the effect of the selection function on the metallicity gradient and the metallicity distribution function (MDF) for four different spectroscopic surveys of the Milky Way tracing different stellar populations: APOGEE, LAMOST, RAVE, and GES. We chose similar lines of sight considering common fields for three surveys at a time: APOGEE-LAMOST-RAVE (ALR) and APOGEE-GES-RAVE (AGR). We tried to emulate the selection function using the colour and magnitude cuts as they are defined for the respective surveys. We used two stellar population synthesis models to compare the distribution of the synthesized model sources with the original input catalogue in the respective colour-magnitude diagrams of each survey. The effects of the selection function are studied in detail by applying the selection function to the two models and comparing the MDF of the selected sources with that of the underlying sample.

The paper is structured as follows. In Sect. 2 we describe the four spectroscopic surveys and their respective target selection schemes. Section 3 presents the comparison of stellar parameters between the surveys. The determination of common fields between the surveys, sample selection, and their spectro-photometric distance calculation are presented in Sect. 4. Section 5 describes in detail the stellar population synthesis models that we use to create mock fields and understand the selection function effect in MDF. Section 6 describes the determination of the vertical metallicity gradient, makes a comparison with the literature values, and discusses the influence of the selection function. The final conclusions of our study are given in Sect. 7.

## 2. Spectroscopic surveys

We use the latest available data release of four large-scale spectroscopic surveys: APOGEE (DR13), RAVE (DR5), GES (DR4), and LAMOST (DR2). In this section, we describe the details of each survey. Each survey has an observing strategy and a specific target selection method designed on the basis of the respective science goals. Since most of the target selection schemes make use of simple colour and magnitude cuts, the chosen photometric input catalogue(s) of the sources play an important role in selecting the stars to be observed for each survey. In this section, we describe the chosen input catalogues and the target selection schemes (colour and magnitude cuts) for each survey in detail.

We show the colour magnitude diagrams (CMDs) with the selection box (based on the colour and magnitude cuts) that defines the selection fraction ( $N_{\text{observed}}/N_{\text{photometric sample}}$ ) for the respective fields of each survey. We further bin the selection box into smaller boxes, where the observed sources are located, of 0.3 mag in magnitude and 0.05 mag in colour, which we call “masks”. These masks are used to create the mock fields from the stellar population synthesis models as described in detail in Sect. 5.2.

## 2.1. APOGEE

APOGEE (Majewski et al. 2017) is one of the four programs in the Sloan Digital Sky Survey III (SDSS-III; Eisenstein et al. 2011), which performed a three-year survey of our Galaxy using the Sloan 2.5 m Telescope at the Apache Point Observatory (APO). APOGEE observes in the near-infrared  $H$ -band ( $1.5\text{--}1.7\ \mu\text{m}$ ) at high spectral resolution ( $R \sim 22\,500$ ) and high signal-to-noise ratios,  $S/N$  ( $>100$ ). Each plate has a field of view (FOV) ranging from  $1\text{--}3^\circ$ ; the number of visits per field varies from 1 to  $\sim 24$  depending on the type and location of the field. The APOGEE Stellar Parameters and Chemical Abundances Pipeline (ASPCAP; García Pérez et al. 2016) is used to determine the stellar parameters and chemical abundances of up to 15 elements based on a  $\chi^2$  minimization between observed and synthetic model spectra.

We use the DR13 catalogue which has 164 558 sources (SDSS Collaboration et al. 2016). We select only main survey targets, removing calibration, telluric, and ancillary targets, for a total sample of 109 376 stars. APOGEE provides also “calibrated” stellar parameters which were calibrated using a sample of well-studied field and cluster stars, including a large number of stars with asteroseismic stellar parameters from NASA’s *Kepler* mission. Using calibrated parameters implies a limit in  $\log g < 3.8$ . Nearly 16% of our sample lacks calibrated surface gravities and nearly 4% of the sources have no calibrated effective temperature and metallicity values. For this reason, we chose the uncalibrated ASPCAP values of fundamental stellar parameters for our study.

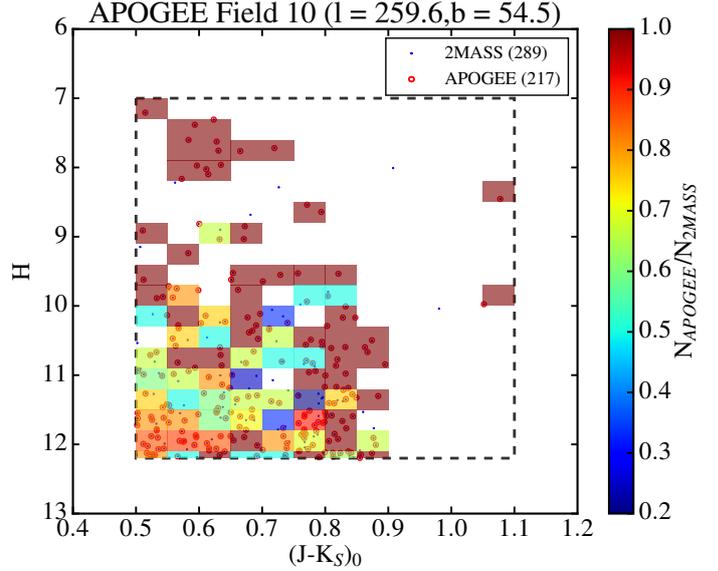
APOGEE has a well-defined input catalogue and colour selection scheme, as described in Zasowski et al. (2013). The Two Micron All Sky Survey (2MASS) Point Source Catalog (Skrutskie et al. 2006) is used as the base catalogue, and the targets are chosen based on their  $H$ -band magnitude and a colour limit to the dereddened  $(J - K_S)_0$  colour (Zasowski et al. 2013). The extinction corrections are derived using the Rayleigh Jeans Colour Excess (RJCE) method (Majewski et al. 2011), which calculates the reddening values combining the 2MASS photometry with mid-IR data (*Spitzer*-IRAC Galactic Legacy Infrared Mid-Plane Survey, Benjamin et al. 2003; Churchwell et al. 2009; and Wide-field Infrared Survey Explorer, Wright et al. 2010), as

$$A(K_S) = 0.918 \times (H - [4.5\mu] - (H - [4.5\mu]_0)) \quad (1)$$

$$E(J - K_S) = 1.5 \times A(K_S) \quad (2)$$

A colour cut at  $(J - K_S)_0 \geq 0.5$  mag was used to include stars cool enough that the stellar parameters and abundances can be reliably derived by ASPCAP (García Pérez et al. 2016), and to lower the fraction of nearby dwarf star “contaminants” in the sample. For the halo fields ( $|b| > 16^\circ$ ), the limit is extended to a bluer colour cut of 0.3 mag. In addition, for some dwarf-dominated halo fields, Washington+DDO51 photometry was used to choose more giants stars than dwarfs. The bit 7 of the APOGEE\_TARGET1 flag is set for sources that fulfil the Washington+DDO51 photometric giant star criteria (see Zasowski et al. 2013). The general  $H$ -magnitude limit is taken to be  $7 \leq H \leq 13.8$ , though the upper limit varies depending on the field and the plate design.

Using the input 2MASS sample and their respective  $A(K_S)$  values for our fields of interest, we estimated the fraction of the observed stars with respect to 2MASS stars in small rectangular bins in the CMD. Figure 1 shows a typical example where we see that the selection function shows variations along the CMD. For the full field the fraction of observed stars to the 2MASS sample



**Fig. 1.**  $(J - K_S)_0$  vs.  $H$  showing the selection function for one of the fields located towards  $l \sim 259.6^\circ$ ,  $b \sim 54.5^\circ$ . The bins are colour-scaled based on the  $N_{\text{APOGEE}}/N_{\text{2MASS}}$  with sizes of 0.05 mag in  $(J - K_S)_0$  colour and 0.3 mag in  $H$ . The dashed box shows the overall colour and magnitude cuts used for APOGEE.

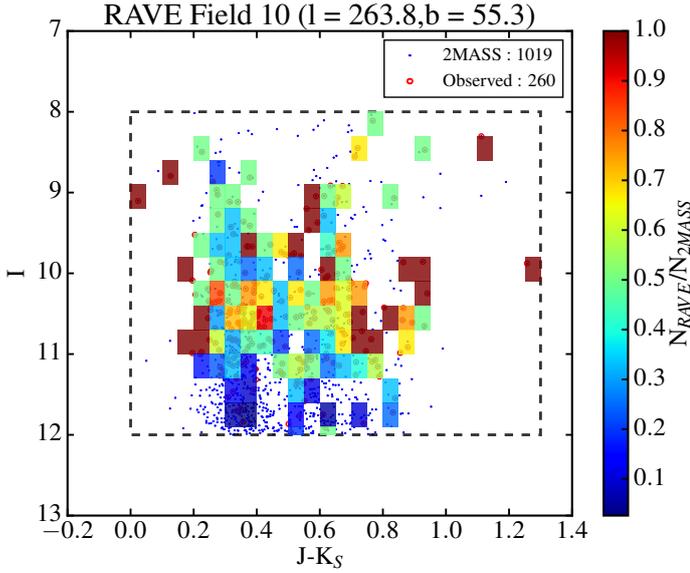
is  $\sim 0.75$ . However, with the rectangular binning in the CMD, we see a lower fraction for fainter stars ( $H > 11$ ) and for bluer stars ( $(J - K_S)_0 < 0.7$ ).

## 2.2. RAVE

The RAdial Velocity Experiment (RAVE; Steinmetz et al. 2006) is a multi-fibre spectroscopic survey that covers the entire southern celestial hemisphere except at low  $|b|$  and  $|l|$ . The observations were carried out at the Anglo-Australian Observatory (AAO) in Siding Spring, Australia, using the 1.2 m UK Schmidt telescope. A 6 degree field multi-object spectrograph was used to obtain the spectra in the infrared CaII-triplet region ( $8410 \text{ \AA} < \lambda < 8795 \text{ \AA}$ ) with a spectral resolution of  $R \sim 7500$ . The stellar atmospheric parameters were estimated using the pipeline designed for the RAVE spectra (Kordopatis et al. 2011a, 2013) making use of the MATrix Inversion for Spectral SynthEsis algorithm (MATISSE, Recio-Blanco et al. 2006) and the DEcision tree alGorithm for ASTrophysics (DEGAS).

We used the DR5 catalogue of RAVE, which has more than  $\sim 520\,000$  sources (Kunder et al. 2017). There are repeated observations ( $\sim 15\%$ ) with the same RAVEID, DENIS\_ID, and 2MASS\_ID, but which differ in the stellar parameters. In these cases, we chose the sources with the highest  $S/N$ . For our study, we used the calibrated fundamental stellar parameters which are labelled with the suffix “\_N\_K” (e.g. TEFF\_N\_K, LOGG\_N\_K, etc.; Kunder et al. 2017).

For RAVE, the selection function is defined based on the  $I$  magnitude and  $(J - K_S)$  colour cut. The initial target selection is based on the apparent  $I$ -band magnitude, for  $9 < I < 12$ , but the input sample is not obtained from a single catalogue. For the regions towards the Galactic disc and bulge (Galactic latitude  $|b| < 25^\circ$ ), a colour criterion  $J - K_S \geq 0.5$  is imposed to select cool giant stars (Kordopatis et al. 2013). The original input catalogue for earlier data releases of RAVE was constructed by deriving  $I$  magnitudes from the Tycho-2 catalogue (Høg et al. 2000), photographic  $I$  magnitudes from the SuperCOSMOS Sky



**Fig. 2.**  $(J - K_S)$  vs.  $I$  (CMD) showing the selection function for one of the fields located towards  $l \sim 263.8^\circ$ ,  $b \sim 55.3^\circ$ . The bins are colour-scaled based on the  $N_{\text{RAVE}}/N_{\text{2MASS}}$  with sizes of 0.05 mag in  $(J - K_S)$  colour and 0.3 mag in  $I$ . The dashed box shows the overall colour and magnitude cuts for RAVE.

Survey (Hambly et al. 2001), and later using the Gunn  $I$ -band photometry from the DENIS catalogue (Epchtein et al. 1997). The latest DR4 release includes observations drawn from a new input catalogue based on DENIS DR3 (DENIS Consortium 2005) cross-matched with the 2MASS point source catalogue. We adopted the same strategy as Wojno et al. (2017) by choosing 2MASS as the input catalogue, and calculated an approximate  $I_{\text{2MASS}}$  magnitude via the following formula:

$$I_{\text{2MASS}} - J = (J - K_S) + 0.2 \times e^{\frac{(J - K_S) - 1.2}{0.2}} + 0.12. \quad (3)$$

For our fields of interest, we searched for 2MASS sources using the approximate field centres (Tables 2 and 3) with a radius of  $2.85^\circ$ , and used the above-mentioned formula to calculate the  $I_{\text{2MASS}}$  magnitude. We used the same approach as in the case of APOGEE to determine the selection function for each field, i.e. by defining selection bins in the CMD,  $(J - K_S)$  vs.  $I_{\text{2MASS}}$ . We used sources with  $I_{\text{2MASS}}$  between 8 and 12 mag as the  $I_{\text{2MASS}}$  distribution for DR5 is shown to be broader than  $I_{\text{DENIS}}$  (Wojno et al. 2017). Figure 2 shows the CMD for one such field located towards high galactic latitude. Here again, we see a clear drop in the selection fraction to about 0.1–0.2 for the fainter magnitudes ( $I > 11$ ).

### 2.3. GES

The *Gaia*-ESO survey (GES) is a public spectroscopic survey aimed at targeting  $\sim 10^5$  stars covering the major components of the Milky way (Gilmore et al. 2012). The observations are carried out using the Fibre Large Array Multi Element Spectrograph (FLAMES; Pasquini et al. 2002) on the Very Large Telescope array (VLT) in Cerro Paranal, Chile. This fibre facility has a FOV of 25 arcmin diameter for two different spectrographs, GIRAFFE and UVES. The stellar parameters were derived by different nodes (using the MATISSE, SME: Valenti & Piskunov 1996 and FERRE: Allende Prieto et al. 2006, codes for GIRAFFE spectra, and about a dozen different methods for UVES spectra). The final recommended GES parameters come from

careful homogenization and calibration of the different results for a given star.

For our study, we chose the sources observed using GIRAFFE with two set-ups, HR10 ( $\lambda = 5339\text{--}5619 \text{ \AA}$ ,  $R \sim 19\,800$ ) and HR21 ( $\lambda = 8484\text{--}9001 \text{ \AA}$ ,  $R \sim 16\,200$ ). We used the homogenized set of parameters from the three nodes in DR4, which are available on the public ESO webpage<sup>1</sup>, and were left with 29 591 sources.

The GES selection function is defined in Stonkutė et al. (2016) based on the VISTA Hemisphere Survey (VHS, McMahon et al. 2013) magnitudes. We obtained the VHS catalogue for our fields of interest from the ESO archive by searching using the field centre and a search radius of  $0.2^\circ$ . We used the APERMAG4 magnitude in the VHS catalogue for the corresponding  $J$ ,  $H$ , and  $K$  magnitudes as it gives the closest magnitudes to those provided in the GES catalogue.

The target selection scheme of GES is built on stellar magnitudes and colours by defining two boxes, one blue and one red. The blue box is used for the selection of the turn-off and main-sequence targets, while the red box is for the red clump targets (Stonkutė et al. 2016). The colour and magnitude cuts for the blue and red boxes are given below:

- Blue box:  $0.0 \leq (J - K_S) \leq 0.45$  for  $14.0 \leq J \leq 17.5$ .
- Red box:  $0.4 \leq (J - K_S) \leq 0.70$  for  $12.5 \leq J \leq 15.0$ .

That said, the actual target selection scheme also takes into account the extinction by shifting the boxes by  $0.5 \times E(B - V)$ , where  $E(B - V)$  is taken as the median reddening in the field measured from the Schlegel et al. (1998) maps. Furthermore, additional targets were assigned by relaxing the red edge of the colour-cut if enough targets were not available within the colour cuts (e.g. high latitude Milky Way fields). Thus, the target selection scheme becomes

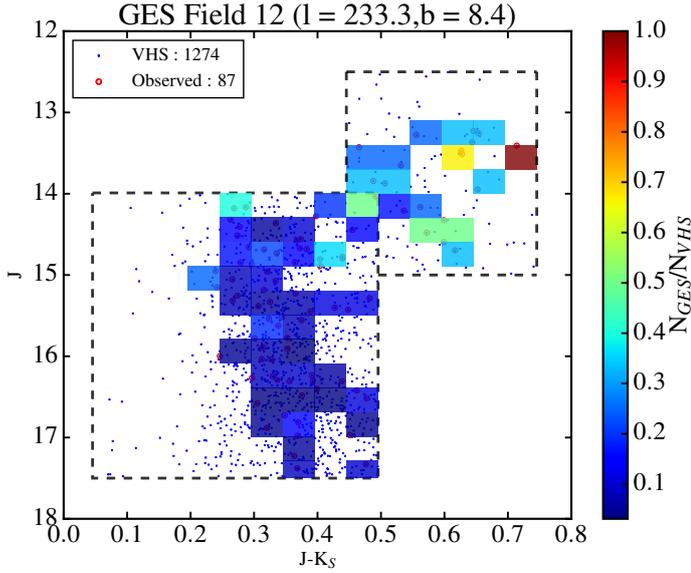
- Blue box:  $0.5E(B - V) + [0.0 \leq (J - K_S) \leq 0.45]$  for  $14.0 \leq J \leq 17.5$ .
- Red box:  $0.5E(B - V) + [0.4 \leq (J - K_S) \leq 0.70]$  for  $12.5 \leq J \leq 15.0$ .
- Extra box:  $0.5E(B - V) + [0.0 \leq (J - K_S) \leq 0.45 + \Delta_G]$  for  $J \geq 14.0$  and  $J + 3 \times ((J - K_S) - 0.35) \leq 17.5$ .

Where  $\Delta_G$  is the right-edge extension of the extra box, and the values of  $\Delta_G$  and  $E(B - V)$  are provided in Table 1 of Stonkutė et al. (2016) for the required fields. Figure 3 shows the selection scheme for field 12, which is a low latitude field with a higher stellar density, that leads to a lower selection fraction. In this field, we found no stellar parameters for about 15% of the GES sources. This fraction can increase to about 40% in other fields.

### 2.4. LAMOST

The Large Sky Area Multi-Object Fiber Spectroscopic Telescope (LAMOST) is another extensive ground-based spectroscopic survey of the Galaxy being carried out with the Guo Shou Jing reflecting Schmidt Telescope. It is equipped with 16 low resolution spectrographs capable of recording the spectra of up to 4000 objects simultaneously in a FOV of  $5^\circ$ , covering all optical wavelengths with a spectral resolution of  $\sim 1800$  (Cui et al. 2012; Zhao et al. 2012). The survey contains the LAMOST ExtraGalactic Survey (LEGAS) and the

<sup>1</sup> <http://www.eso.org/gi/>



**Fig. 3.**  $(J - K_S)$  vs.  $J$  (CM diagram) showing the selection function for field 12 located approximately towards  $l \sim 233.3^\circ$ ,  $b \sim 8.4^\circ$ . The bins are colour-scaled based on the  $N_{\text{GES}}/N_{\text{VHS}}$  with sizes of 0.05 mag in  $(J - K_S)$  colour and 0.3 mag in  $J$ . The dashed box shows the overall colour and magnitude cuts for red and blue boxes used for GES.

LAMOST Experiment for Galactic Understanding and Exploration (LEGUE; Deng et al. 2012; Smith et al. 2012), which itself is composed of three separate surveys with different input catalogues and selection functions.

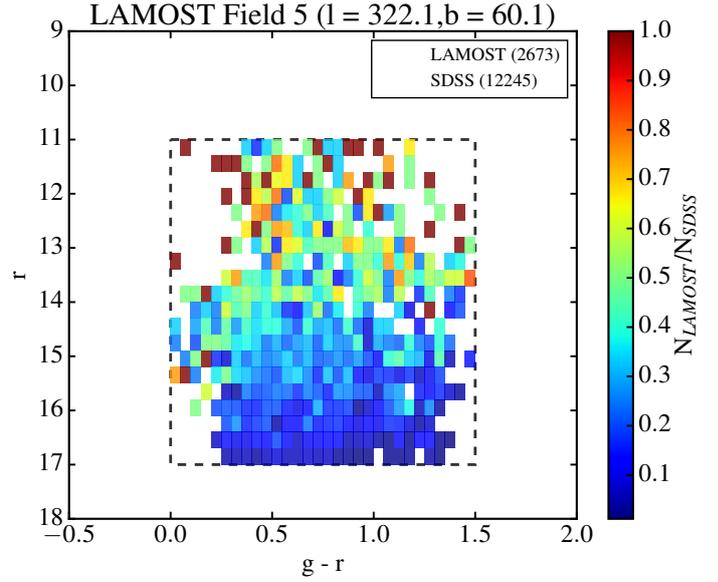
We use the DR2 catalogue, which has the calibrated stellar parameters for 2 207 803 sources estimated using the LAMOST Stellar Parameter Pipeline (LASP, Wu et al. 2011).

Like RAVE, LAMOST does not make use of a single input catalogue and unlike the other three surveys, LAMOST uses *ugriz* photometry for their target selection. Catalogues such as UCAC4 (Zacharias et al. 2013) and Pan-STARRS 1 (Tonry et al. 2012) have been used to select targets for the observing plates in the main survey regions. To make sure that we use a homogeneous photometric input sample, we use only the SDSS photometry<sup>2</sup> in order to define the selection function. For the respective fields, we searched for SDSS sources within a radius of  $2.5^\circ$  around the field centres. We use only those fields where we have the full SDSS footprint covered.

The targeting algorithm for LEGUE designed by Carlin et al. (2012) was not applied, due to sparse stellar sampling. This results from the limited dynamic range of magnitudes observed on a single LAMOST plate and from the brighter  $r$  magnitude limit at the faint end compared to the designed goal (Liu et al. 2017). Finally, the target selection was carried out on a plate-by-plate basis with different plates covering  $9 < r < 14$ ,  $14 < r < 16.8$ ,  $r < 17.8$ , and  $r < 18.5$ .

Based on the distribution of LAMOST sources of each field in the  $g - r$  vs.  $r$  CMD, we used the following colour and magnitude cuts:  $0.0 < g - r < 1.5$  and  $11 < r < 17.8$ . We neglected the small fraction of very red ( $g - r > 1.5$ ) and blue ( $g - r < 0.0$ ) sources. Figure 4 shows a typical example for a field towards  $l = 322.1^\circ$ ,  $b = 60.1^\circ$ . Statistically, LAMOST is more prominent than the other three surveys, and this is seen in the dramatic rise in the number of masks within the colour and

<sup>2</sup> 2MASS was more complete in each field, but the LAMOST sources are fainter than the reliable magnitude limits ( $\sim 14.3$ ) of 2MASS pass band magnitude.



**Fig. 4.**  $(g - r)$  vs.  $r$  (CM diagram) showing the selection function for one of the fields located towards approximate  $l \sim 322.1^\circ$ ,  $b \sim 60.1^\circ$ . The bins are colour-scaled based on the  $N_{\text{LAMOST}}/N_{\text{SDSS}}$  with sizes of 0.05 mag in  $(g - r)$  colour and 0.3 mag in  $r$ . The dashed box shows the overall colour and magnitude cuts used for LAMOST.

magnitude range. The gradual decrease in the selection fraction towards fainter magnitudes is seen here as well.

### 3. Comparison of stellar parameters between the surveys

Despite the large amount of spectroscopic data available, few studies exist (e.g. Schultheis et al. 2017; Kunder et al. 2017; Lee et al. 2015; Chen et al. 2015; Kordopatis et al. 2013) comparing stellar parameters and chemical abundances of the same stars. Chen et al. (2015) studied common stars between APOGEE DR12 (Alam et al. 2015) and the LAMOST DR2 catalogue. They found that the LAMOST photometrically calibrated  $T_{\text{eff}}$  were consistent with the spectroscopic  $T_{\text{eff}}$  from APOGEE, but found systematic biases in  $\log g$  and  $[\text{Fe}/\text{H}]$  for common stars in the APOGEE DR12 and LAMOST DR2 catalogues. Lee et al. (2015) applied the SEGUE Stellar Parameter Pipeline (SSPP; Allende Prieto et al. 2008; Lee et al. 2008a,b, 2011; Smolinski et al. 2011) to the spectra from LAMOST and compared the stellar parameters with the common stars in APOGEE (DR12), RAVE (DR4), and SEGUE. For the RAVE DR5 release (Kunder et al. 2017) a detailed comparison of the derived stellar parameters in RAVE with that of APOGEE, GES, and LAMOST for the common stars has been obtained as a part of the external RAVE DR5 verification.

Recently, a data-driven approach known as *The Cannon* (Ness et al. 2015) has been introduced. *The Cannon* uses stellar spectra along with the derived stellar parameters from well-characterized stars (estimated with pipelines using synthetic model spectra) in higher resolution surveys as a training set to derive stellar parameters. This method was used to derive the stellar parameters for around 450 000 giant stars in LAMOST (low spectral resolution survey) by bringing them to the scale of APOGEE (high spectral resolution), showing that two very different spectroscopic surveys can be combined together (Ho et al. 2017). But still there are limitations in this data-driven approach, as the accuracy of *the Cannon* depends on the chosen training

**Table 1.** Median offset and dispersion estimated for the comparison of fundamental stellar parameters in APOGEE – RAVE and APOGEE – LAMOST.

APOGEE – RAVE (907)		
	Median offset	Dispersion
$T_{\text{eff}}$ (K)	–109.4	229.2
Log $g$ (dex)	0.02	0.60
[Fe/H] (dex)	0.10	0.16
APOGEE – LAMOST (11 203)		
	Median offset	Dispersion
$T_{\text{eff}}$ (K)	–42.3	170.4
Log $g$ (dex)	–0.04	0.34
[Fe/H] (dex)	0.05	0.09

set. In the cases of APOGEE and LAMOST, which target different populations – red giants stars for APOGEE vs. dwarf stars for LAMOST – only a limited training set is available. *The Cannon* was also used to re-analyse the RAVE spectra and a new catalogue (RAVE-on) of stellar parameters and abundances was produced (Casey et al. 2017). The training set for red giant stars was made using common stars in APOGEE DR13 (SDSS Collaboration et al. 2016), while the stars in common in K2/EPIC catalogue (Huber et al. 2016) made up the main-sequence and subgiant branch training set.

For our study, we need to make sure that systematic offsets between surveys are corrected. This is accomplished by comparing the derived stellar parameters for the common sources between the surveys. We have arbitrarily chosen the APOGEE data set as a reference frame due to its high spectral resolution and S/N. We used a cross-identification radius of  $2''$  to identify the common sources of the three surveys with respect to APOGEE.

In this section, we investigate the systematic offsets seen in  $T_{\text{eff}}$ , log  $g$ , and [Fe/H]<sup>3</sup> between APOGEE and the other data sets, RAVE, GES, and LAMOST, respectively. We found only six sources in common between GES and APOGEE, making any comparison difficult. Therefore, we carried out the comparison only between APOGEE and RAVE and between APOGEE and LAMOST. Table 1 lists the median offsets and the standard deviation estimated for APOGEE – RAVE and APOGEE – LAMOST.

### 3.1. APOGEE-RAVE

There are 907 sources in common between APOGEE and RAVE. Figure 5 (top panel) shows the comparison of the stellar parameters. We find that RAVE has systematically higher temperatures than APOGEE with a median difference of about 110 K. The log  $g$  values show a peculiar shape (Fig. 5, top middle panel), which has been already noted by Kunder et al. (2017). They consider this behaviour to be the consequence of degeneracies in the Ca IR triplet region that affects the determination of log  $g$  (Kordopatis et al. 2011a). Overall we see a large dispersion (0.6 dex) for log  $g$  between APOGEE and RAVE. In terms of the metallicity comparison, we note that APOGEE gives higher metallicities for metal-poor stars ([Fe/H] < –0.2 dex), while much lower metallicities for metal-rich stars ([Fe/H] > 0.2 dex)

<sup>3</sup> APOGEE, RAVE, and GES use the [M/H] notation for the overall content of metallic elements, rather than [Fe/H]. However, throughout this paper, we use [Fe/H] as the global metallicity, i.e. [Fe/H] = [M/H]. For LAMOST, only [Fe/H] is provided in the catalogue.

in comparison with RAVE. We have calculated the median offsets in bins of 0.25 dex in RAVE metallicities (indicated by the red points) and did a linear fit to them, as indicated by the red line.

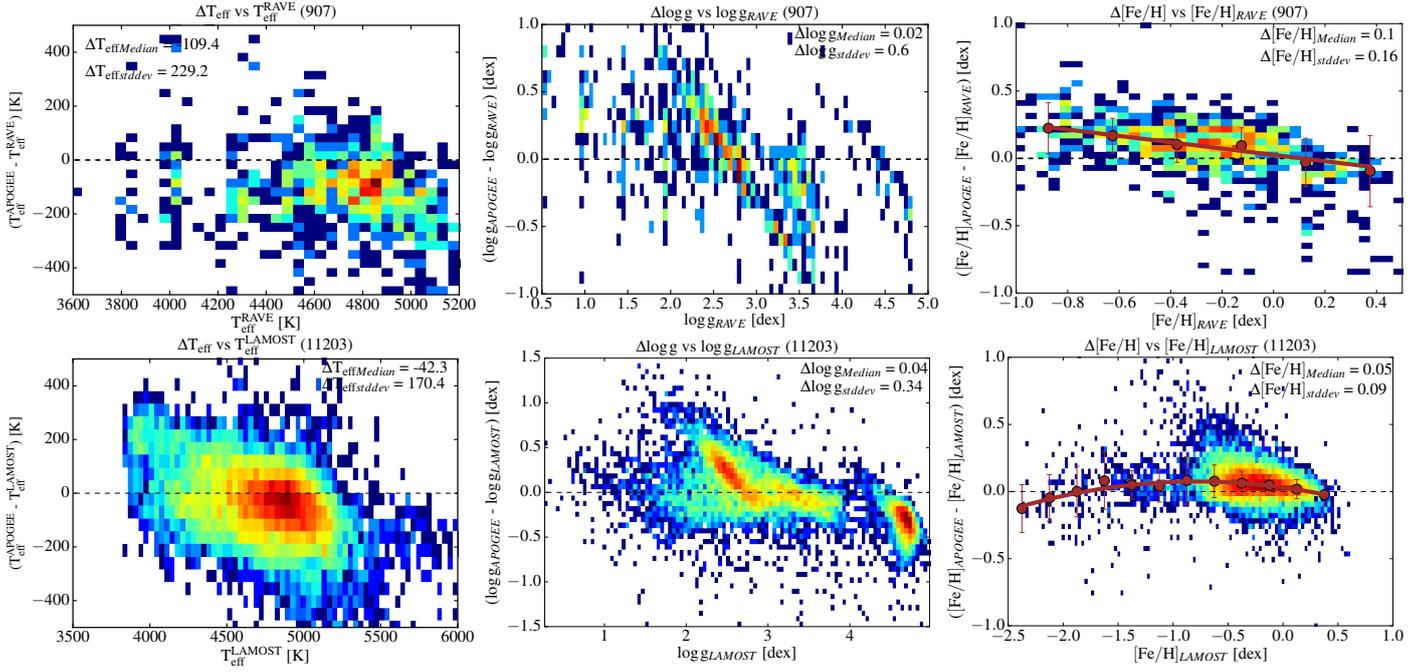
### 3.2. APOGEE-LAMOST

LAMOST has 11 203 sources in common with APOGEE, which is statistically the highest number. The plots used for comparison are shown in the bottom panel of Fig. 5. The median offset of  $T_{\text{eff}}$  between LAMOST and APOGEE is about 42 K with a dispersion of 170 K. A slight trend with  $T_{\text{eff}}$  is visible in the sense that for stars hotter than 5000 K LAMOST predicts higher temperatures. However, the APOGEE pipeline is more adapted to getting stellar parameters for cooler stars below 5000 K (Holtzman et al. 2015). The log  $g$  correlation shows distinct behaviour for the APOGEE values above and below  $\sim 4.0$  dex. Below 4.0 dex we see in general good agreement except around log  $g \sim 2.5$ , the area where the red clump is dominant. The log  $g$  values above 4.0 dex are underestimated in APOGEE because of the lack of reliable calibrators for stars with high surface gravity values (Holtzman et al. 2015). This can be seen in the behaviour of difference in log  $g$  with log  $g_{\text{LAMOST}}$  for log  $g > 4.0$  dex. Hence, we estimate the median offset and dispersion for log  $g_{\text{APOGEE}} < 4.0$  dex. Overall, the median offset is negligible ( $\sim -0.04$  dex), but with a large dispersion ( $\sim 0.34$  dex). In the case of metallicity, the median offset between APOGEE and LAMOST for sources within  $-1 < [\text{Fe}/\text{H}] < 0.5$  is 0.05 dex with a dispersion of 0.14 dex. Here again, we added correction terms for the metallicities by calculating the median offset in bins of 0.25 dex in metallicity and fitting a second-degree polynomial to them (red points and line).

## 4. Common fields and distance determination

We have chosen to select common fields distributed along similar lines of sight to study the effect of the selection function on the observed MDF and the vertical gradient. In total, there are only three common fields between the four surveys, which provide a relatively small sample size, and for this reason we have chosen to restrict ourselves to the common fields between three surveys at a time: APOGEE, LAMOST, and RAVE (hereafter ALR) and APOGEE, GES, and RAVE (AGR). We found eight common fields in the ALR case and ten in AGR. Tables 2 and 3 list the common fields in each survey of ALR and AGR, respectively, with the approximate mean field centres and number of sources in each field.

Distances for the sources in each survey are estimated by isochrone fitting as described in Rojas-Arriagada et al. (2017), which is similar to other methods in the literature (e.g. Zwitter et al. 2010) using the derived stellar parameters  $T_{\text{eff}}$ , log  $g$ , [M/H] together with the  $J, H, K_S$  (for APOGEE, RAVE, and GES) or SDSS  $u, g, r, i$  photometry (for LAMOST). A set of Padova and TRIESTE Stellar Evolution Code (PARSEC) isochrones with ages ranging from 1 to 13 Gyr in 1 Gyr step and metallicities from –2.2 dex to +0.5 dex in 0.1 dex step are chosen for this. PARSEC is the stellar evolutionary code used to compute sets of stellar evolutionary tracks for stars of different initial masses, evolutionary phases, and metallicities. Isochrones in several photometric systems are derived from these tracks (Bressan et al. 2012). The distance of the observed star to the set of all model stars from the whole set of isochrones is calculated in the  $T_{\text{eff}}\text{-log } g\text{-[Fe/H]}$  parameter space. This distance is weighted to account for the evolutionary speed and



**Fig. 5.** Comparison of common sources in APOGEE and RAVE (*top*), and APOGEE and LAMOST (*bottom*) for  $T_{\text{eff}}$  (*left*),  $\log g$  (*middle*), and  $[\text{Fe}/\text{H}]$  (*right*). For the metallicities, the median of the difference in parameters and its dispersion is also shown as red circles with error bars in the plots.  $[\text{Fe}/\text{H}]$  here denote the global metallicity for APOGEE and RAVE (see footnote 2).

**Table 2.** Details of the ALR fields.

Field	Mean ( $l^\circ$ , $b^\circ$ )	$N(\text{APOGEE})$	$N(\text{RAVE})$	$N(\text{LAMOST})$
5	(320.7, 59.4)	222 (222)	497 (495)	2844 (2677)
10_1	(262.6, 55.2)	221 (221)	277 (277)	503 (475)
10_2	(259.1, 55.4)	–	–	1306 (1175)
13_1	(58.7, $-45.5$ )	228 (228)	566 (566)	719 (709)
13_2	(61.7, $-46.5$ )	–	–	1185 (1160)
14	(312.4, 59.1)	225 (225)	250 (248)	1208 (777)
15_1	(292.1, 59.6)	221 (221)	362 (362)	994 (898)
15_2	(293.2, 60.7)	–	–	990 (846)
20_1	(331.6, 58.6)	227 (227)	306 (306)	2822 (2488)
20_2	(330.7, 60.7)	–	–	3021 (2870)
25_1	(273, 58.3)	252 (252)	282 (282)	2668 (2500)
25_2	(269.9, 57.9)	263 (262)	–	385 (360)
25_3	(270.7, 58.8)	260 (260)	–	1542 (1460)
32	(254, 51.4)	225 (225)	300 (300)	1793 (1698)

**Notes.** The field numbers assigned by us, mean of field centres of each field for three surveys, along with the number of observed sources in each field are listed. The number of observed sources having their distances calculated (based on availability of derived  $T_{\text{eff}}$ ,  $\log g$ ,  $[\text{Fe}/\text{H}]$ , and photometric magnitudes) are indicated in parentheses alongside the observed source numbers. There are overlapping fields for certain surveys, indicated by \_1, \_2, or \_3.

non-uniformity of model stars along the isochrone tracks. Using these weights, the most likely values of absolute magnitudes of the star in each band is calculated as the weighted mean or median of the model stars' absolute magnitudes. We also compute the line-of-sight reddening from the observed and theoretical colours. Finally, we compute the distance modulus and the

line-of-sight distance in each passband from the absolute magnitudes and the estimated reddening (Rojas-Arriagada et al. 2017). We use the average value of the distances from different passbands as the final line-of-sight distance of each source. Using the same approach to calculate the distances for each of the four surveys makes sure that no biases are introduced. The typical

**Table 3.** Details of the AGR fields. The columns are the same as in Table 2.

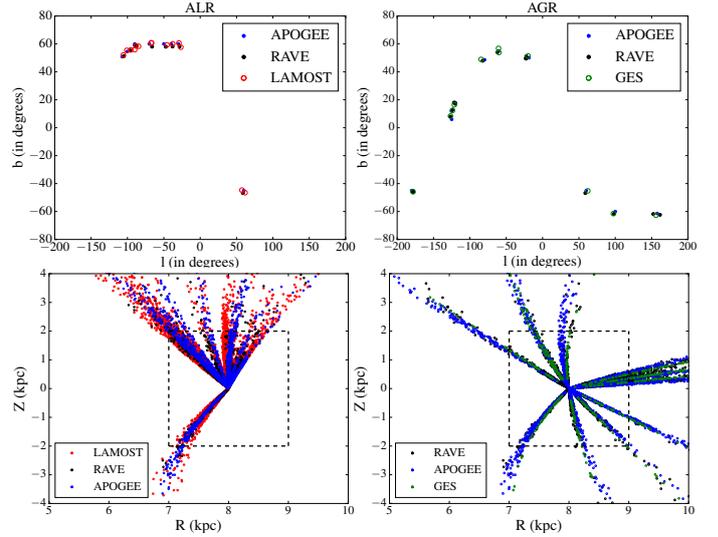
Field	Mean ( $l^\circ, b^\circ$ )	$N(\text{APOGEE})$	$N(\text{RAVE})$	$N(\text{GES})$
1_1	(339.6, 50.8)	228 (228)	485 (485)	103 (54)
1_2	(337.1, 49.7)	–	521 (521)	–
1_3	(337.2, 49.5)	–	534 (534)	–
3	(182, –45.6)	222 (222)	155 (155)	91 (47)
4_1	(158.4, –62.1)	259 (259)	389 (389)	82 (45)
4_2	(152.2, –61.8)	–	233 (233)	–
5_1	(239.6, 17.2)	230 (230)	614 (614)	104 (81)
5_2	(238.2, 18)	–	479 (479)	–
8	(60.3, –45.7)	228 (228)	566 (566)	103 (59)
10	(277.7, 48.5)	220 (220)	582 (582)	102 (51)
12	(233.8, 7.4)	229 (229)	801 (801)	104 (87)
14	(235.9, 12.4)	229 (229)	451 (451)	108 (88)
23	(98.4, –61.1)	279 (279)	320 (320)	86 (47)
31_1	(299.1, 54.3)	231 (231)	348 (348)	108 (60)
31_2	(299.4, 56.9)	–	–	96 (54)

uncertainty of the distances is in the order of  $\sim 20\%$ . The Galactocentric distance  $R$  (kpc) and the vertical height  $Z$  (kpc) from the Galactic mid-plane for the sources are calculated by assuming the Sun to be located at  $R \sim 8.0$  kpc. The distribution of the fields in the Galactic plane and that of the sources in the  $R - Z$  plane are shown in Fig. 6. We see from the  $R - Z$  distribution in the bottom panel of Fig. 6 that there is a wide range of  $Z$  allowing us to determine the vertical gradient (see Sect. 6), while the range in the radial gradient is limited. Even though the range in  $R$  is broader in AGR than that in ALR, to be consistent between the two cases and to minimize the impact of any radial gradient on the vertical gradient, we restrict the samples in  $R$  from 7 to 9 kpc and  $|Z|$  from 0 to 2 kpc for our study (see the dashed box in the bottom panel of Fig. 6).

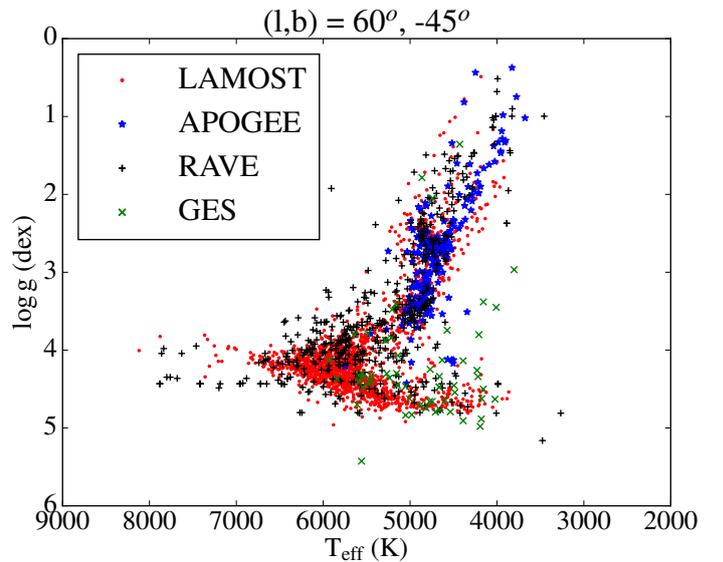
In Fig. 7, we show the Hertzsprung-Russell diagram for the sources in the field common to all four surveys located towards  $(l, b) \sim 60^\circ, -45^\circ$ . The location of LAMOST and RAVE sources are mainly concentrated along the main-sequence and the turn-off stars, while APOGEE traces mainly giant stars and red clump stars. The lowest density comes from GES tracing mainly main-sequence stars.

## 5. MOCK fields using stellar population synthesis models

Stellar population synthesis models make use of Galaxy formation and evolution scenarios along with some physical assumptions to generate a picture of the Milky Way in different photometric systems. Their prime objective is to compare and interpret different observational data currently available and also to test the theories on which the models are based. These models use the fundamental equation of stellar statistics (Bahcall 1986) to compute the star counts for the distinct components (thin



**Fig. 6.** Distribution of the common fields in ALR and AGR shown in the Galactic plane (top); the  $R - Z$  distribution of the sources in those fields with the dashed box indicating the  $7 \leq R \leq 9$  kpc and  $|Z| \leq 2$  kpc range we chose to select the sources for our study (bottom).



**Fig. 7.**  $T_{\text{eff}}$  vs.  $\log g$  diagram of sources for the four surveys in the field located towards  $(l, b) \sim 60^\circ, -45^\circ$ .

and thick discs, halo, bulge) that make up the Milky Way. Fundamental stellar parameters are derived assuming each component's respective density distribution laws, initial mass function and luminosity function, star formation rate, age-metallicity relation, etc., using libraries of stellar evolutionary tracks and synthetic spectra. GALAXIA (Sharma et al. 2011) and TRILEGAL (Girardi et al. 2005) are the two most commonly used population synthesis models differing in their assumed component parameters, density laws, star formation history, stellar libraries, and other dynamical constraints. We chose them to create the respective mock fields for each survey with the aim of understanding the selection function effect in the MDFs and to attempt a basic comparison of the models. We chose two stellar population synthesis models with different input physics to test the robustness of our analysis.

GALAXIA uses a modified version of the BESANÇON model (Robin et al. 2003), which is dynamically self consistent,

constraining the scale height of populations (assumed to be isothermal and relaxed) by its velocity dispersion and Galactic potential. Among the four main populations, the thin disc is divided into seven age components (0 to 10 Gyr) and has a two-slope initial mass function. The thick disc has a mean metallicity of  $-0.78$  dex simulated as a single burst of age 11 Gyr, while the halo (mean metallicity  $\sim -1.78$  dex) and bulge (mean metallicity  $\sim 0.0$  dex) are simulated as single bursts of ages of 14 and 10 Gyr, respectively. Instead of the Schlegel extinction maps (Schlegel et al. 1998) used by GALAXIA to calculate the extinction along a given line of sight, we chose the more sophisticated model of the 3D dust distribution provided by Drimmel et al. (2003).

TRILEGAL does not include the dynamical consistency to constrain the scale height, but like GALAXIA it is able to deal with a full set of different photometric systems. TRILEGAL also has four galactic components with certain input parameters that can be modified, such as thin and thick discs, halo, and bulge. We assumed a thin disc with a total mass surface density of  $55.4 M_{\odot} \text{pc}^{-2}$ , a scale length  $h_R = 2.15$  kpc, and an age dependent scale height  $h_Z(t_{\text{Gyr}}) = 245(1+t/5.5)^{1.66}$  pc. The abundances of Rocha-Pinto et al. (2000) are adopted for the thin disc and a two-step SFH with a 1.5 times enhancement in the SFR between the ages of 1 and 4 Gyr. The thick disc is assumed to have a local mass volume density of  $0.008 M_{\odot} \text{pc}^{-3}$ ,  $h_R = 3.2$  kpc, and  $h_Z = 0.74$  kpc. The SFH is constant over an age range of 11–12 Gyr with mean metallicity  $\sim -0.67 \pm 0.1$  dex and an  $\alpha$ -enhancement of  $\sim 0.3$  dex. The halo is modelled using an axisymmetric power law with a power-law index of 2.75 (de Jong et al. 2010) and local mass volume density of  $4 \times 10^{-4} M_{\odot} \text{pc}^{-3}$ . The SFH is constant over an age range of 12–13 Gyr with mean metallicity  $\sim -1.6 \pm 1.0$  dex and a corresponding  $\alpha$ -enhancement of  $\sim 0.3$  dex. We use the Drimmel et al. (2003) dust distribution to calculate the extinction in TRILEGAL as well. The model scheme and other details are described in Girardi et al. (2005).

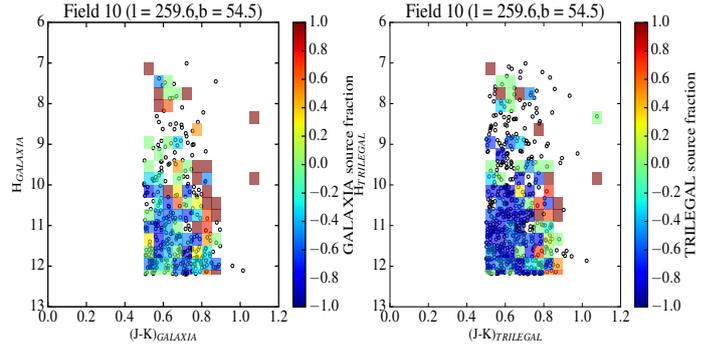
We generate the mock catalogues along each line of sight using the field centre and the field radius of the respective surveys for each of the model. The 2MASS+SDSS photometric system was used for both the models.

### 5.1. Applying uncertainties and related checks on the models

Both GALAXIA and TRILEGAL predict the stellar parameters and photometric magnitudes for each star at a given line of sight. Each of the four surveys has intrinsic errors in the measured stellar parameters and observed photometric magnitudes, which should be simulated accordingly in the model in order to make it more realistic. Since we use only the metallicity values from the models to compare the MDFs and vertical gradients, we do this only for the metallicity among the stellar parameters in the model. In order to simulate the metallicity errors, we have fitted a fourth-degree polynomial in the  $\sigma_{[\text{Fe}/\text{H}]}$  vs.  $[\text{Fe}/\text{H}]$  plane for APOGEE, LAMOST, and GES. For RAVE, we used the same metallicity error description as described in Kordopatis et al. (2013). We apply a Gaussian filter to the metallicities of the mock catalogues.

Similarly we need to apply uncertainties to the photometric 2MASS or SDSS magnitudes provided by each model. We used an exponential function for 2MASS and a fourth-degree polynomial for SDSS to define the relation between the mag vs.  $\sigma_{\text{mag}}$ , which we model as a Gaussian for each model source, as mentioned earlier.

In addition, we have simulated errors in the distance distribution in the model to verify the percentage of sources lost



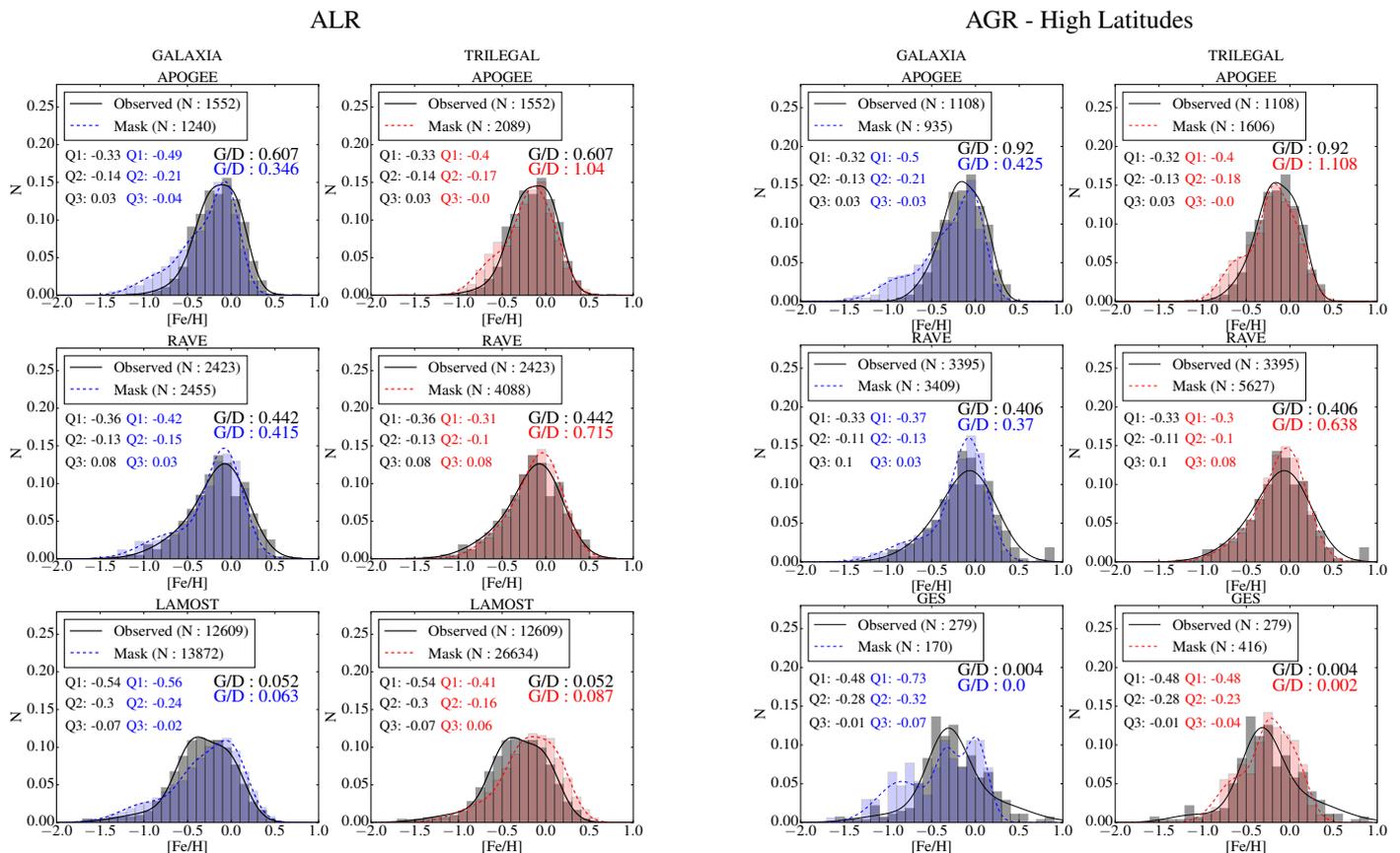
**Fig. 8.** CMD diagrams for the APOGEE field located towards  $l, b \sim 59.6^\circ, 54.5^\circ$  with GALAXIA and TRILEGAL source distribution shown in the left and right panels, respectively. The open circles in each panel represent the respective model sources. The rectangular boxes are the masks where the real observed sources are, each colour-coded with the fraction  $(N_{2\text{MASS}} - N_{\text{model}})/N_{2\text{MASS}}$  if  $N_{2\text{MASS}} > N_{\text{model}}$  or  $(N_{2\text{MASS}} - N_{\text{model}})/N_{\text{model}}$  if  $N_{2\text{MASS}} < N_{\text{model}}$ . The redder colours indicate that the 2MASS sources are in equal number or greater than the number of model sources, while bluer colours denote more model sources. From the colours of the bins, there are more TRILEGAL sources than GALAXIA sources towards the faint magnitudes.

or gained, due to our limiting  $R - Z$  cut. The typical errors in the spectro-photometric distances are in the order of 20% (see Hayden et al. 2015; Schultheis et al. 2015). We carried out ten trials, each time introducing a 20% error in distance calculated by the models to check whether this drastically affects the number of sources at the boundaries of the  $R - Z$  range that are thrown out or that come in. We have found that a 20% error in the distances would affect less than 5% of stars in the selected  $R - Z$  ranges. This is a relatively small change in the sample size, but to make the mock sample realistic, we kept the 20% distance uncertainty in the models.

### 5.2. Comparison between GALAXIA and TRILEGAL

In Sect. 2, we show the CMDs for each survey with their respective colour and magnitude cuts that were used to select the target sample. We carry out the same exercise for the mock fields generated using the models. The masks (in the form of small rectangular boxes within the selection box) are used to denote the bins in colour and magnitude where the sources are observed. Figure 8 shows an example of our method for the APOGEE field located towards  $l \sim 259.6^\circ$  and  $b \sim 54.5^\circ$  for both GALAXIA and TRILEGAL. The masks are colour-coded with the fractional percentage of model sources compared to the input photometric sample (2MASS or SDSS). We find that TRILEGAL predicts more sources at the faint end than does GALAXIA. A very similar trend is seen for the simulated RAVE, LAMOST, and GES fields.

Using the masks in the CMD we force each model to have the same fraction ( $N_{\text{observed}}/N_{\text{photometric sample}}$ ) as the real targeted and observed sample. To understand which model best replicates the observed MDF, we compare the mask and the observed sample MDFs combining all common fields for each survey in ALR and AGR, restricted in the  $R - Z$  range of  $7 \leq R \leq 9$  kpc and  $|Z| \leq 2$  kpc, as shown in Fig. 9. Each model uses different stellar libraries and stellar evolutionary tracks, which could lead to systematic offsets in the abundance scale between the models and observations. However, as discussed in Sects. 5.3 and 6.1, we investigate the selection effect within each model.



**Fig. 9.** Mask and observed normalized MDF for ALR (*left*) and AGR (*right*) in the  $R - Z$  range of  $7 \leq R \leq 9$  kpc and  $0 \leq |Z| \leq 2$  kpc. The survey histograms are in black, while GALAXIA and TRILEGAL histograms are in blue and red, respectively. Histograms are normalized by dividing the counts in each 0.1 dex bin by the total number of sources. The distributions in black line represent the observed MDF, while those in blue and red lines represent the mask MDF for GALAXIA and TRILEGAL, respectively. APOGEE and RAVE distributions are shown in the *top and middle* rows, respectively, and the LAMOST (*left*) and GES (*right*) in the bottom row. Quartile values for both distributions are given in each panel colour-coded according to the distribution. Indicated is also the giant-to-dwarf ratio for mask and observed samples of each survey for both models.

Except for APOGEE and GES, GALAXIA predicts a slightly higher number of sources, while TRILEGAL overestimates the star density for all four surveys. This difference in the number of predicted sources with respect to the observed sample could be explained by the different assumption in the density normalization of the two models. The fractional percentage of mask sources compared to the observed sources is listed in Table 4, where the differences between the two models can be clearly seen.

The MDFs are binned in 0.1 dex metallicity bins and then normalized by the total number of sources. The fit of the distributions are carried out using a Gaussian mixture model (GMM; as in Rojas-Arriagada et al. 2017). A mixture model (M) is a weighted sum of a number of probability distribution functions, while the mixture in a GMM in 2D is defined by a sum of bivariate normal distributions. For a given data structure composed of certain underlying substructures/features, the parameters ( $\mu$ ,  $\sigma$ , etc.) that define the best mixture model with a given number of modes is determined using the expectation-maximization algorithm that maximizes the likelihood function of the mixture. Since we do not know a priori the exact number of components in the data, we use the Bayesian Information Criterion (BIC) as a cost function to assess the relative fitting quality between different proposed mixtures and determine the best solution (the one with the lowest BIC value) to the number of Gaussian components that constitutes the distribution.

**Table 4.** Fractional percentage of mask sources compared to the observed sources for GALAXIA and TRILEGAL in ALR and AGR.

ALR	APOGEE (%)	RAVE (%)	LAMOST (%)
GALAXIA	-20.16	+1.32	+10.06
TRILEGAL	+34.41	+68.72	+111.23
AGR	APOGEE (%)	RAVE (%)	GES (%)
GALAXIA	-15.62	+0.41	-39.07
TRILEGAL	+45.04	+65.74	+49.10

In order to quantify the differences between the observed and the mask MDFs, we estimate and compare the quartile values for each distribution as in Wojno et al. (2017). The quartile values designated as Q1, Q2, Q3 represent the 25th, 50th and 75th percentiles of the distribution, respectively, as indicated in each panel of Fig. 9. We choose 0.1 dex (considered to be the general metallicity uncertainty) as the threshold for the difference between the respective quartiles of samples below which the distributions are considered to agree. We note the following results by comparing the mask and observed MDFs for each survey (see Fig. 9):

- Both models show a significant metal-poor tail in the mask MDF compared to the APOGEE MDF, which is more prominent in the case of GALAXIA but absent in the observations.

This can be seen in the difference in Q1 quartile between GALAXIA and APOGEE distributions, exceeding the adopted threshold of 0.1 dex.

- For RAVE, TRILEGAL matches the observed distribution very well for both AGR and ALR. In GALAXIA the mask MDF is distributed as broadly as the RAVE MDF, though there are subtle differences in the source fraction throughout.
- For LAMOST, the observed MDF has a broad peak that is skewed towards subsolar metallicities, which both the models are unable to replicate. The difference in Q1 and Q2 quartiles exceed the 0.1 dex threshold between TRILEGAL and LAMOST distributions.
- The small sample size in the case of GES makes it hard to decipher the exact shape of the MDFs, especially for GALAXIA. The majority of the 0.1 dex bins in the GALAXIA MDF have fewer than 15 sources, except for the bins closer to the peak with more than 25 sources. Thus the normalized GALAXIA mask MDF is dominated by noise in the form of multiple peaks, which we are not able to properly fit using GMM. The TRILEGAL mask MDF has better statistics, resulting in significantly less noise in the distribution. This is also evident in the quartile analysis for GALAXIA, with the Q1 quartile difference between GALAXIA and GES distributions exceeding 0.1 dex, which is not the case with TRILEGAL.

We also estimated the giant-to-dwarf ratio in the model and observed samples in all cases. The giants and dwarfs are separated based on their  $\log g$  value, i.e. dwarfs:  $\log g > 3.5$  and giants:  $\log g < 3.5$ . These ratios show how well each model is able to replicate the observed stellar population. Among the four surveys, the giant fraction is highest in APOGEE followed by RAVE, LAMOST, and GES. We find that both models are unable to represent the ratio we find in the observed sample for APOGEE as GALAXIA gives a lower value, while TRILEGAL overpredicts the ratio by a factor of  $\sim 2$  in ALR. The prediction by TRILEGAL is close to the observed case for APOGEE in AGR, although still slightly overpredicted. For RAVE, we find the ratio in GALAXIA to be very close to that of the observed case, while TRILEGAL again gives comparatively higher values for both ALR and AGR. The giant-to-dwarf ratios predicted by both GALAXIA and TRILEGAL are quite similar for LAMOST and for GES, though the sample size is quite limited for the GES mask in GALAXIA.

Overall, neither model is able to reproduce both the MDF and the giant-to-dwarf ratio of the APOGEE sample. Both conditions are found to be very consistently satisfied by GALAXIA in the case of RAVE. Even though the shape of the MDFs are slightly different from the observed MDF, we find consistency in the MDF and giant-to-dwarf ratio between GALAXIA and TRILEGAL for LAMOST. In the case of GES, TRILEGAL reproduces the observed MDF better than GALAXIA, likely because of the lack of targets in the latter model.

### 5.3. Effects of the selection function in MDF

With the models described above, we are able to study the effect of the selection function on the MDF for the sample from the common fields of each survey in ALR and AGR. We categorize the sources in the mock fields by the limiting magnitude of the respective surveys and restricted in  $R - Z$  range as the parent population. This represents the underlying sample from which the selection function in the form of colour and magnitude cuts are applied to create a subset of mask sources. These

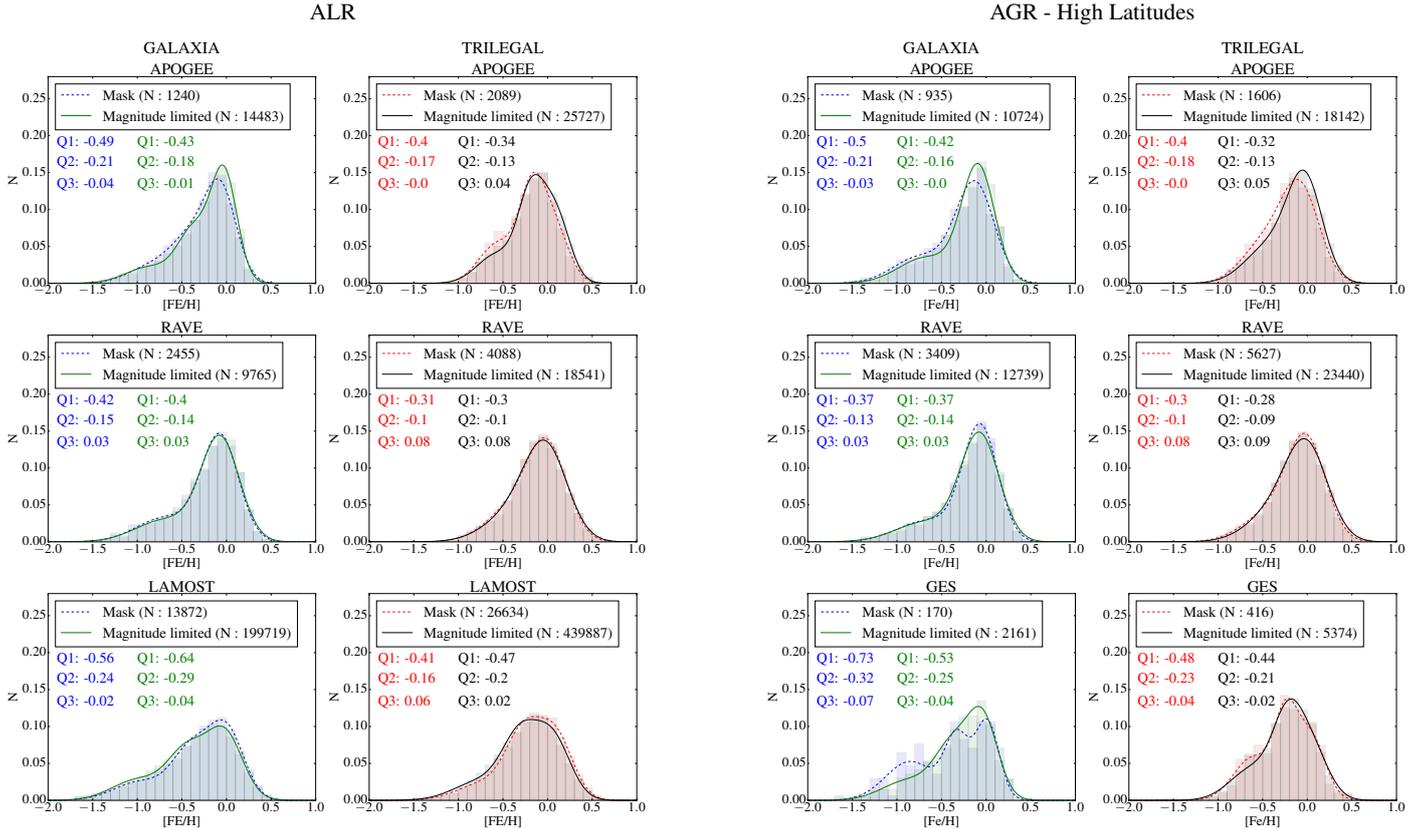
mask sources in turn represent the observed sources. Thus by comparing the MDF of the underlying sample, hereafter called the magnitude limited sample, with that of the mask sample, we can assess the effect of the selection function, if any, on the underlying MDF for each survey. For ALR and AGR, we restrict both the magnitude limited and mask sample in the  $R - Z$  range of  $7 \leq R \leq 9$  kpc and  $|Z| \leq 2$  kpc and all fields are combined together. The  $|Z|$  values for sources in the three low latitude fields in AGR do not exceed 1 kpc for the selected  $R$  range. As these low latitude fields have different selection cuts and low numbers of stars, we restrict our analysis only towards high latitude fields.

We compare the magnitude limited MDFs and the effect of the selection function on the MDF for ALR and AGR in Fig. 10. Here again we use the GMM method to fit the multiple number of Gaussians to the MDF. In addition, we use the quartile values to carry out a quantitative comparison of two distributions. Wojno et al. (2017) carried out a similar comparison of distributions using the quartile values for RAVE. As mentioned in Sect. 5.2, we choose 0.1 dex as the threshold for the difference between the respective quartiles of samples below which the distributions are considered to agree thus implying that the selection function has a minimal effect on the MDF. The quartile values for mask and magnitude limited samples are listed in the respective panels in Fig. 10. We find that all the quartile values estimated for the mask and the magnitude limited sample in GALAXIA are more metal-poor than those in TRILEGAL.

As APOGEE and RAVE are the common surveys in ALR and AGR, we check the consistency of the quartile values of their mask and magnitude limited distributions. Since we are using only high latitude fields, the colour and magnitude cuts are consistent for both surveys in ALR and AGR. As a first step, we compare the quartile values of mask and magnitude limited distributions of APOGEE in ALR and AGR for each model separately. The same model-wise quartile comparison is carried out for RAVE in ALR and AGR. We find that model-wise, the individual quartiles differ only by a maximum of  $\sim 0.03$  dex for both surveys between ALR and AGR. This ensures that APOGEE and RAVE distributions in ALR and AGR are consistent.

To quantitatively compare the mask and magnitude limited distributions of each model in ALR and AGR, we check for differences in their individual quartiles:

- For APOGEE, we find the differences in  $Q1$  to be  $\sim 0.05$ – $0.08$  dex, for  $Q2 \sim 0.03$ – $0.05$  dex, and for  $Q3 \sim 0.03$ – $0.05$  dex. Thus, although we find the quartile differences to vary widely, they are all within the 0.1 dex difference threshold, implying no large selection function effect. As mentioned in Sect. 2.1, there are certain halo fields in APOGEE where the giant targets are preselected using the Washington+DDO51 photometry. These photometric bands are not available in the two the models we use, so we attempted an approximate simulation of this preselection for such fields in our mask sample by using the APOGEE\_TARGET1 flag (see Sect. 2.1). We estimated the observed giant-to-dwarf fraction in each small rectangular bins, which we tried to replicate in the models by choosing approximately the same giant fraction. We carried out the comparison of mask and magnitude limited samples with this approximate giant preselection using quartiles. We find there is an overselection of metal-poor stars in the  $-0.5$  to  $-1.0$  dex range in  $[\text{Fe}/\text{H}]$  in GALAXIA, which is not evident in TRILEGAL. The Q1 quartiles for mask and magnitude limited samples show differences higher than 0.1 dex in such fields for GALAXIA, but in Fig. 9, we find that GALAXIA



**Fig. 10.** MDFs of magnitude limited and mask sample for the high latitude fields of each survey in ALR (*left*) and AGR (*right*). The GALAXIA and TRILEGAL MDFs are shown respectively in the *left and right columns* of each panel. The histograms are normalized by dividing the counts in each 0.1 dex bin by the total number of sources (mentioned in each panel). The blue and red lines represent the mask distribution, while the green and black lines for the magnitude limited distribution fitted using GMM for GALAXIA and TRILEGAL respectively. APOGEE and RAVE distributions are shown in the *top and middle rows*, respectively, and the LAMOST (*left*) and GES (*right*) in the bottom row. Quartile values for the two distributions are given in each panel, colour-coded according to the distribution.

already (i.e. without any giant preselection) overpredicts the metal-poor stars in the same  $[\text{Fe}/\text{H}]$  regime compared to the observed APOGEE MDF. It is likely that this selection effect seen in such giant dominated fields is the result of model parameterization, as it already significantly overpredicts contributions from the metal-poor populations.

- For RAVE, the individual quartile differences are minimal:  $\sim 0.0\text{--}0.02$  dex. Thus we find the RAVE MDF to be least affected by the selection effect. [Wojno et al. \(2017\)](#) have found very similar result for the RAVE DR5 sample, for separate giant, main-sequence, turn-off samples of stars and for a mixed population sample and at different distance bins from the Galactic mid-plane.
- Like APOGEE, the individual quartile differences in LAMOST show some variations ( $Q1$ :  $\sim 0.06\text{--}0.08$  dex,  $Q2$ :  $\sim 0.04\text{--}0.05$  dex, and  $Q3$ :  $\sim 0.03$  dex). However, as per our criteria, the selection function effect is not prominent.
- In the case of GES, we find inconsistency in the quartiles between GALAXIA and TRILEGAL. The quartile differences of the mask and magnitude limited sample in TRILEGAL agrees within the 0.1 dex threshold ( $Q1$ :  $\sim 0.04$  dex,  $Q2$ :  $\sim 0.02$  dex, and  $Q3$ :  $\sim 0.02$  dex), but the  $Q1$  values in GALAXIA differ by  $\sim 0.2$  dex. We find that GALAXIA masks do not have enough sources in the metal-poor regime (170 stars in total), making this difference highly susceptible to Poisson noise.

## 6. Vertical metallicity gradients in ALR and AGR

We measure the vertical metallicity gradient of our source sample in the solar neighbourhood for each survey in ALR and AGR. We investigate the possible selection effect on the metallicity gradient using both stellar population synthesis models (GALAXIA and TRILEGAL). We further determine the vertical metallicity gradient for each survey independently after accounting for metallicity offsets between them.

### 6.1. Effects of selection function in vertical metallicity gradients

We use stellar population synthesis models as described in Sect. 5.3 to simulate any influence of the selection function on the vertical metallicity gradient.

Here we estimate and compare the vertical metallicity gradients for the mask and magnitude limited sample of each survey. In both models, the gradient in metallicity in the vertical direction is not incorporated as an input parameter. Instead, the mean metallicity of different Galaxy components like thin and thick discs, combined with their different scale heights, leads to the vertical metallicity gradient. In TRILEGAL, this is found to be shallow ( $\sim -0.1$  dex  $\text{kpc}^{-1}$ ), while that in GALAXIA is steeper ( $\sim -0.4$  dex  $\text{kpc}^{-1}$ ). This can be attributed to the wide range of the age–metallicity relation for thin disc in GALAXIA

**Table 5.** Vertical metallicity gradients measured for mask and magnitude limited sample for GALAXIA and TRILEGAL in ALR.

Model	Survey	Mask (dex kpc <sup>-1</sup> )	Mag limited (dex kpc <sup>-1</sup> )
GALAXIA	APOGEE	-0.359 ± 0.033	-0.382 ± 0.025
	LAMOST	-0.396 ± 0.019	-0.378 ± 0.017
	RAVE	-0.438 ± 0.047	-0.424 ± 0.042
TRILEGAL	APOGEE	-0.085 ± 0.018	-0.054 ± 0.01
	LAMOST	-0.037 ± 0.005	-0.044 ± 0.005
	RAVE	-0.121 ± 0.019	-0.103 ± 0.016
	GES	-0.075 ± 0.039	-0.072 ± 0.011

( $\sim -0.01$  to  $-0.37$  dex [Sharma et al. 2011](#)). Thus the two models differ largely in their vertical metallicity gradients. We use only the ALR sample to carry out our simulations to ensure sufficient statistics.

As mentioned in Sect. 5.2, the mask sample is made by randomly choosing the model sources within each 0.05 mag by 0.3 mag bins in the CMD. The slopes of the gradients are measured by finding the median metallicity in 0.2 kpc bins in  $|Z|$  and then using a linear least-squares regression fit to the median values. Unlike in the case of MDF, we find that the source distribution in the  $|Z|$ - $[\text{Fe}/\text{H}]$  plane is sensitive to the random distribution of stars calculated by the models for the mask sample. We find that the mask gradient varies each time the model is run since the location of mask sample stars at the high  $Z$  boundaries keep varying. In order to account for this, we performed ten different mask samples for each survey in ALR. We use the median value of the vertical metallicity gradient and its error estimated from all ten trials as the final gradient for the mask sample. If there is a major influence of the selection function, we expect to find different metallicity gradients between the mask and magnitude limited samples. Table 5 gives the fitted values of the metallicity gradient for ALR using GALAXIA and TRILEGAL. We find that the variation in the gradient estimated between the mask and magnitude limited samples are consistent within the  $1\sigma$  limit for all three surveys (see Table 5). This indicates that the selection function does not have a strong impact on the vertical metallicity gradient. We investigate this further in the following section using the vertical metallicity gradients estimated for the observed sample from each survey.

## 6.2. Vertical metallicity gradients for the observed sample

To estimate the vertical metallicity gradient and compare them between the different surveys, we have to ensure that the metallicities of the different surveys are on the same scale. We applied a small offset to the RAVE and LAMOST metallicities with respect to our reference sample APOGEE using the linear and second-degree polynomial functions that we fitted in Sects. 3.1 and 3.2, respectively. While estimating the functional relation of metallicity offsets of RAVE and LAMOST with respect to APOGEE, we make sure that the relation holds true separately for both high and low S/N samples of RAVE and LAMOST. So we can proceed without any major quality cuts for each survey, ensuring a statistically significant sample for our study.

Here again we restrict our study to the high latitude fields. Figure 11 shows the vertical metallicity gradients plotted separately for ALR and AGR. We follow the same fitting routine mentioned in the previous section to estimate the slopes. Table 6 lists the slopes of the gradients calculated for each survey in ALR and AGR, along with the slope of the combined sample.

**Table 6.** Vertical metallicity gradients measured for ALR and AGR high latitude fields.

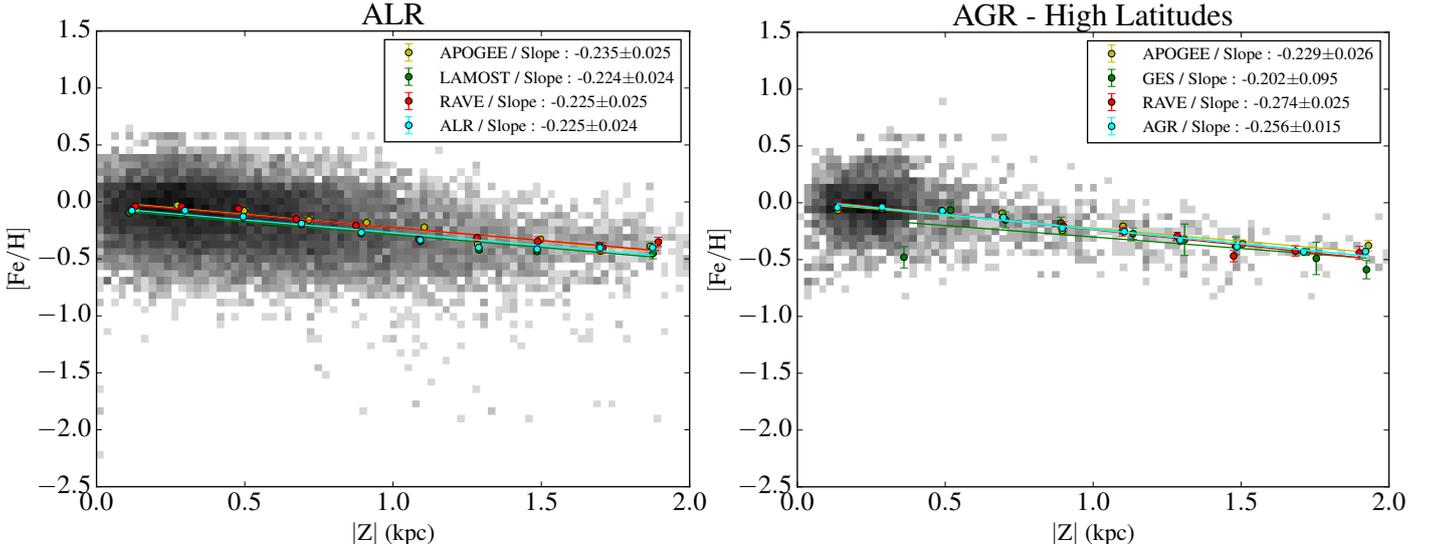
Survey	$\frac{d[\text{Fe}/\text{H}]}{dz}$ (dex kpc <sup>-1</sup> )
APOGEE	-0.235 ± 0.025
LAMOST	-0.224 ± 0.024
RAVE	-0.225 ± 0.025
<b>ALR</b>	<b>-0.225 ± 0.024</b>
APOGEE	-0.229 ± 0.026
GES	-0.202 ± 0.095
RAVE	-0.274 ± 0.025
<b>AGR</b>	<b>-0.256 ± 0.015</b>
Mean $\frac{d[\text{Fe}/\text{H}]}{dz}$	<b>-0.241 ± 0.028</b>

We also list the mean vertical metallicity gradient estimated from combined samples of ALR and AGR in the last row of the table.

We measured the vertical gradients of  $-0.235 \pm 0.025$  dex kpc<sup>-1</sup> and  $-0.229 \pm 0.026$  dex kpc<sup>-1</sup> for APOGEE in ALR and AGR, respectively. [Hayden et al. \(2014\)](#) measured a slightly steeper slope of  $-0.31 \pm 0.01$  dex kpc<sup>-1</sup> for their APOGEE DR10 sample located within the solar neighbourhood,  $7 < R < 9$  kpc and  $|Z| < 2$  kpc. Their sample in the solar neighbourhood is more complete and homogeneous than the volume limited sample we are dealing with. We used the same criterion in [Hayden et al. \(2014\)](#) to distinguish the  $\alpha$ -poor and  $\alpha$ -rich sources in our sample, and measured  $\frac{d[\text{M}/\text{H}]}{dz} = -0.175 \pm 0.045$  dex kpc<sup>-1</sup> and  $-0.164 \pm 0.035$  dex kpc<sup>-1</sup> for the low- $\alpha$  samples in ALR and AGR, respectively. [Hayden et al. \(2014\)](#) measured a slightly steeper gradient of  $-0.21 \pm 0.02$  dex kpc<sup>-1</sup> for their set of low- $\alpha$  samples. Meanwhile, the low number statistics of the high- $\alpha$  sources in our sample prevented us from calculating the same. The vertical metallicity gradient for the DR13 APOGEE sources in the same  $R - Z$  range was recalculated and found to be consistent with our slope ([Hayden et al., in prep.](#)).

The RAVE vertical metallicity gradient calculated for ALR and AGR are also similar, with slopes of  $-0.225 \pm 0.025$  dex kpc<sup>-1</sup> and  $-0.274 \pm 0.025$  dex kpc<sup>-1</sup>, respectively. The steeper slope in AGR could be due to the comparatively low number statistics in the bins at high  $|Z|$  with respect to that in ALR, which makes the slope steeper. [Boeche et al. \(2014\)](#) used giants stars in the RAVE DR4 sample and measured a shallower slope of  $\frac{d[\text{Fe}/\text{H}]}{dz} = -0.112 \pm 0.007$  dex kpc<sup>-1</sup> for  $\sim 10\,511$  stars (RC sample) in the region extending from 7.5 to 8.5 kpc in  $R$  and  $|Z| \leq 2$  kpc. They have also carried out a study of the gradients seen in the  $\alpha$ -poor and  $\alpha$ -rich sample, but we were not able to identify any clear separation in the  $[\alpha/\text{Fe}]$  vs.  $[\text{Fe}/\text{H}]$  plane to define a  $\alpha$ -poor and  $\alpha$ -rich sample.

The low number statistics of GES is a significant issue when analysing metallicity trends. Since we use only the high latitude fields, there are no sources populated in the first 0.2 kpc bins of  $|Z|$ . We also find a comparatively metal-poor median metallicity value in the 0.2–0.4 kpc  $|Z|$  bin, which leads to a very high uncertainty of  $\sim 0.1$  dex kpc<sup>-1</sup> in the estimated gradient. We calculated  $\frac{d[\text{Fe}/\text{H}]}{dz}$  of  $-0.202 \pm 0.095$  dex kpc<sup>-1</sup>, which is still inside  $1\sigma$  with respect to APOGEE and RAVE. [Mikolaitis et al. \(2014\)](#) measured vertical metallicity gradients of  $-0.079 \pm 0.013$  and  $-0.046 \pm 0.010$  dex kpc<sup>-1</sup> for thin and thick-disc FGK stars separately in the solar circle ( $7 \leq R \leq 9$  kpc and  $|Z| \leq 3.5$  kpc), which are much shallower than our measurements. They use the first internal data release of the GES (GES iDR1) and chemically



**Fig. 11.** Vertical metallicity gradients calculated for all sources in each survey belonging to ALR (*left*) and AGR (*right*). The slope estimated for each survey is also shown in the plots. The gradient for the combined sample of surveys is shown for ALR and AGR.

differentiate the thin and thick discs based on their  $\alpha$  abundances, whereas our sample is made using the DR4 release and we do not make any separation based on the  $\alpha$ -abundances.

The vertical metallicity gradient measured for LAMOST is steep,  $-0.224 \pm 0.024$  dex  $\text{kpc}^{-1}$ , but within the  $1\sigma$  limit with respect to APOGEE and RAVE. Using nearly 70 000 red clump stars covering  $7 \leq R \leq 14$  kpc and  $|Z| \leq 3$  kpc from the LAMOST Spectroscopic Survey of the Galactic Anti-Centre (LSS-GAC) survey, Huang et al. (2015) measured the radial and vertical metallicity gradients. Our estimate is steeper than their slopes of  $-0.146 \pm 0.012$  dex  $\text{kpc}^{-1}$  and  $-0.149 \pm 0.012$  dex  $\text{kpc}^{-1}$  measured for the sample in  $7 \leq R \leq 8$  kpc and  $8 \leq R \leq 9$  kpc, respectively. Xiang et al. (2015) measured a vertical metallicity gradient that is in the range of  $\sim -0.2$  to  $-0.3$  dex  $\text{kpc}^{-1}$  in the  $R$  bin of 8 to 9 kpc and  $|Z| < 2$  kpc, for a sample of main-sequence turn-off stars from LSS-GAC, which is consistent with our value.

Our mean vertical metallicity gradient from the combined samples of ALR and AGR is  $-0.241 \pm 0.028$  dex  $\text{kpc}^{-1}$ . Among recent studies, Schlesinger et al. (2014) carried out a detailed study of the vertical metallicity gradient using over 40 000 G-dwarf stars from the SEGUE DR9 catalogue, volume complete in the range of 6.7 to 9.5 kpc in  $R$  and 0.27 to 1.62 kpc in  $|Z|$ . Their range in  $R - Z$  is very close to the coverage of our samples. They estimated the gradient to be  $-0.243^{+0.039}_{-0.053}$  dex  $\text{kpc}^{-1}$ , which is in good agreement with our derived mean value.

In addition to the results from the surveys used in our analysis, there are other studies calculating the vertical metallicity gradients near the solar circle, Chen et al. (2003):  $-0.295 \pm 0.005$  dex  $\text{kpc}^{-1}$ , Borkova & Marsakov (2003):  $-0.29 \pm 0.06$  dex  $\text{kpc}^{-1}$ , Soubiran et al. (2008):  $-0.31 \pm 0.03$  dex  $\text{kpc}^{-1}$  for thin disc clump giants within  $|Z| \leq 1$  kpc, Ak et al. (2007):  $-0.22 \pm 0.03$  dex  $\text{kpc}^{-1}$  for G-type stars in  $|Z| \leq 3$  kpc. All these gradients are very close to the gradients we estimated for each survey. Vertical gradients have been measured for thick discs alone; Chen et al. (2011) have measured a gradient of  $-0.22 \pm 0.07$  dex  $\text{kpc}^{-1}$  for RHB stars in  $1 \leq |Z| \leq 3$  kpc, and Carrell et al. (2012)  $-0.113 \pm 0.010$  dex  $\text{kpc}^{-1}$  for SEGUE FGK dwarf stars in  $1 \leq |Z| \leq 3$  kpc. Kordopatis et al. (2011b) have estimated a vertical metallicity gradient of  $-0.14 \pm 0.05$  dex  $\text{kpc}^{-1}$

using roughly 700 stars at  $1 \leq |Z| \leq 4$  kpc from the solar neighbourhood towards the Galactic coordinates ( $l \sim 277^\circ$ ,  $b \sim 47^\circ$ ).

While there is a large variation in the vertical metallicity gradient in the literature, we find consistent measurements of the vertical gradient in the four spectroscopic surveys analysed in this paper. This implies that the effect of the selection function on the vertical metallicity gradient is very small, if any at all.

As seen in Sect. 6.1 the metallicity gradients for the mask sample in both models (Table 5) do not match the observed metallicity gradient (Table 6). However, as we compare the magnitude limited sample and the mask sample for each model separately to study the influence of the selection effect, we neglect the discrepancy between the observed gradient and both models. Nevertheless, it shows that both models need to be improved in order to reproduce the observed quantities (e.g. metallicity gradient).

Thus, using the models and the observed sample, we find negligible selection function effects on the vertical metallicity gradient. This in turn means that it is indeed possible to merge different spectroscopic surveys to obtain a broader  $Z$  range that traces the vertical metallicity gradient to higher precision, provided they are on the same metallicity scale.

## 7. Conclusions

We investigated the effect of the selection function on the MDF and on the vertical metallicity gradient using common fields between APOGEE-LAMOST-RAVE (ALR) and APOGEE-GES-RAVE (AGR) around the solar neighbourhood. Our results can be summarized as follows.

We compared and discussed stellar parameters of the common sources between APOGEE, RAVE, and LAMOST. In order to bring the surveys to the same metallicity scale, we applied corrections for  $[\text{Fe}/\text{H}]$  with respect to APOGEE, which we used as the reference sample. Distances were calculated for all four surveys and we restricted our sample in  $R$  from 7 to 9 kpc and  $|Z|$  from 0 to 2 kpc allowing the determination of the MDF and vertical metallicity gradient.

We generated mock fields using two commonly used population synthesis models, GALAXIA and TRILEGAL, to investigate the selection effect in MDFs. We divided the corresponding

CMDs into small bins (called masks) which we applied to each model trying to replicate the observed MDF. Based on the comparison of the shape of mask MDFs and giant-to-dwarf ratio we find that

- for APOGEE, both models have a dominant metal-poor tail absent in the observed MDF. In addition, GALAXIA underestimates the giant-to-dwarf ratio while TRILEGAL overestimates it;
- GALAXIA traces both the shape of the MDF and the giant-to-dwarf ratio of the RAVE sample better than TRILEGAL;
- for LAMOST there is good consistency between the giant-to-dwarf ratio of mask samples and observed samples, even though the shape of the mask MDF relative to the observed is found to be different in the two models.

To understand the selection function effect in MDF, we compared the mask MDF with the magnitude limited MDF for the survey-replicas of both models using quartiles. We found APOGEE, RAVE, and LAMOST to have negligible selection effects using both models, while GES suffers from number statistics that are too low to be conclusive.

We simulated the vertical metallicity gradient with the two models, and they both failed to reproduce our observed metallicity gradient; clearly some improvements in both models are necessary. However, when we compare the vertical metallicity gradients between mask and magnitude limited samples for APOGEE, RAVE, and LAMOST, which are sensitive to the selection function, we did not find any significant difference. In addition, the agreement found in our observed values of the metallicity gradient between the four different surveys again strengthens the argument that the selection effect plays a negligible role when determining the metallicity gradient.

We scaled the metallicity values in RAVE and LAMOST to that in APOGEE and compared the vertical metallicity gradients for combined fields for each survey in ALR and AGR. The estimated vertical metallicity gradient for each survey is consistent within  $1\sigma$  indicating the negligible effect of selection function. Finally, we estimated a mean vertical metallicity gradient of  $-0.241 \pm 0.028$  dex  $\text{kpc}^{-1}$ .

We conclude that in the era of rising large spectroscopic surveys, in principle common fields of the surveys could be combined once they are put on the same metallicity scale. This will increase significantly the statistics without imposing any selection effect when studying the MDF and the metallicity gradient. With the forthcoming *Gaia* data releases, we plan to extend this study to a much larger volume.

*Acknowledgements.* We wish to thank the anonymous referee for the extremely useful comments that greatly improved the quality of this paper. G.N. and M.S. acknowledge the Programme National de Cosmologie et Galaxies (PNCG) of CNRS/INSU, France, for financial support. Funding for SDSS-III has been provided by the Alfred P. Sloan Foundation, the Participating Institutions, the National Science Foundation, and the US Department of Energy Office of Science. The SDSS-III web site is <http://www.sdss3.org/>. SDSS-III is managed by the Astrophysical Research Consortium for the Participating Institutions of the SDSS-III Collaboration including the University of Arizona, the Brazilian Participation Group, Brookhaven National Laboratory, Carnegie Mellon University, University of Florida, the French Participation Group, the German Participation Group, Harvard University, the Instituto de Astrofísica de Canarias, the Michigan State/Notre Dame/JINA Participation Group, Johns Hopkins University, Lawrence Berkeley National Laboratory, Max Planck Institute for Astrophysics, Max Planck Institute for Extraterrestrial Physics, New Mexico State University, New York University, the Ohio State University, Pennsylvania State University, University of Portsmouth, Princeton University, the Spanish Participation Group, University of Tokyo, University of Utah, Vanderbilt University, University of Virginia, University of Washington, and Yale University. Funding for RAVE has been provided by the Australian Astronomical Observatory; the

Leibniz-Institut fuer Astrophysik Potsdam (AIP); the Australian National University; the Australian Research Council; the French National Research Agency; the German Research Foundation (SPP 1177 and SFB 881); the European Research Council (ERC-StG 240271 Galactica); the Istituto Nazionale di Astrofisica at Padova; Johns Hopkins University; the National Science Foundation of the USA (AST-0908326); the W. M. Keck foundation; the Macquarie University; The Netherlands Research School for Astronomy; the Natural Sciences and Engineering Research Council of Canada; the Slovenian Research Agency; the Swiss National Science Foundation; the Science & Technology Facilities Council of the UK; Opticon; Strasbourg Observatory; and the Universities of Groningen, Heidelberg, and Sydney. Based on data products from observations made with ESO Telescopes at the La Silla Paranal Observatory under programme ID 188.B-3002. These data products have been processed by the Cambridge Astronomy Survey Unit (CASU) at the Institute of Astronomy, University of Cambridge, and by the FLAMES/UVES reduction team at INAF/Osservatorio Astrofisico di Arcetri. These data have been obtained from the *Gaia*-ESO Survey Data Archive, prepared and hosted by the Wide Field Astronomy Unit, Institute for Astronomy, University of Edinburgh, which is funded by the UK Science and Technology Facilities Council. This work was partly supported by the European Union FP7 programme through ERC grant number 320360 and by the Leverhulme Trust through grant RPG-2012-541. We acknowledge the support from INAF and Ministero dell' Istruzione, dell' Università e della Ricerca (MIUR) in the form of the grant "Premiale VLT 2012". The results presented here benefit from discussions held during the *Gaia*-ESO workshops and conferences supported by the ESF (European Science Foundation) through the GREAT Research Network Programme. The Guo Shou Jing Telescope (the Large Sky Area Multi-Object Fiber Spectroscopic Telescope, LAMOST) is a National Major Scientific Project built by the Chinese Academy of Sciences. Funding for the project has been provided by the National Development and Reform Commission. LAMOST is operated and managed by National Astronomical Observatories, Chinese Academy of Sciences.

## References

- Ak, S., Bilir, S., Karaali, S., Buser, R., & Cabrera-Lavers, A. 2007, *New Astron.*, **12**, 605
- Alam, S., Albareti, F. D., Allende Prieto, C., et al. 2015, *ApJS*, **219**, 12
- Allende Prieto, C., Beers, T. C., Wilhelm, R., et al. 2006, *ApJ*, **636**, 804
- Allende Prieto, C., Sivarani, T., Beers, T. C., et al. 2008, *AJ*, **136**, 2070
- Anders, F., Chiappini, C., Rodrigues, T. S., et al. 2016, *Astron. Nachr.*, **337**, 926
- Bahcall, J. N. 1986, *ARA&A*, **24**, 577
- Balser, D. S., Rood, R. T., Bania, T. M., & Anderson, L. D. 2011, *ApJ*, **738**, 27
- Benjamin, R. A., Churchwell, E., Babler, B. L., et al. 2003, *PASP*, **115**, 953
- Bergemann, M., Ruchti, G. R., Serenelli, A., et al. 2014, *A&A*, **565**, A89
- Boeche, C., Siebert, A., Piffl, T., et al. 2013, *A&A*, **559**, A59
- Boeche, C., Siebert, A., Piffl, T., et al. 2014, *A&A*, **568**, A71
- Borkova, T. V., & Marsakov, V. A. 2003, *A&A*, **398**, 133
- Bovy, J., Rix, H.-W., Liu, C., et al. 2012, *ApJ*, **753**, 148
- Bovy, J., Nidever, D. L., Rix, H.-W., et al. 2014, *ApJ*, **790**, 127
- Bressan, A., Marigo, P., Girardi, L., et al. 2012, *MNRAS*, **427**, 127
- Carlin, J. L., Lépine, S., Newberg, H. J., et al. 2012, *Res. Astron. Astrophys.*, **12**, 755
- Carrell, K., Chen, Y., & Zhao, G. 2012, *AJ*, **144**, 185
- Casey, A. R., Hawkins, K., Hogg, D. W., et al. 2017, *ApJ*, **840**, 59
- Chen, L., Hou, J. L., & Wang, J. J. 2003, *AJ*, **125**, 1397
- Chen, Y. Q., Zhao, G., Carrell, K., & Zhao, J. K. 2011, *AJ*, **142**, 184
- Chen, Y.-Q., Zhao, G., Liu, C., et al. 2015, *Res. Astron. Astrophys.*, **15**, 1125
- Cheng, J. Y., Rockosi, C. M., Morrison, H. L., et al. 2012, *ApJ*, **746**, 149
- Chiappini, C., Matteucci, F., & Romano, D. 2001, *ApJ*, **554**, 1044
- Churchwell, E., Babler, B. L., Meade, M. R., et al. 2009, *PASP*, **121**, 213
- Costa, R. D. D., Uchida, M. M. M., & Maciel, W. J. 2004, *A&A*, **423**, 199
- Cui, X.-Q., Zhao, Y.-H., Chu, Y.-Q., et al. 2012, *Res. Astron. Astrophys.*, **12**, 1197
- Cunha, K., Frinchaboy, P. M., Souto, D., et al. 2016, *Astron. Nachr.*, **337**, 922
- de Jong, J. T. A., Yanny, B., Rix, H.-W., et al. 2010, *ApJ*, **714**, 663
- Deng, L.-C., Newberg, H. J., Liu, C., et al. 2012, *Res. Astron. Astrophys.*, **12**, 735
- DENIS Consortium 2005, *VizieR Online Data Catalog: II/263*
- Drimmel, R., Cabrera-Lavers, A., & López-Corredoira, M. 2003, *A&A*, **409**, 205
- Eisenstein, D. J., Weinberg, D. H., Agol, E., et al. 2011, *AJ*, **142**, 72
- Epchein, N., de Batz, B., Capolani, L., et al. 1997, *The Messenger*, **87**, 27
- Freeman, K., & Bland-Hawthorn, J. 2002, *ARA&A*, **40**, 487
- Frinchaboy, P. M., Thompson, B., Jackson, K. M., et al. 2013, *ApJ*, **777**, L1
- García Pérez, A. E., Allende Prieto, C., Holtzman, J. A., et al. 2016, *AJ*, **151**, 144
- Gazzano, J.-C., Kordopatis, G., Deleuil, M., et al. 2013, *A&A*, **550**, A125

- Gilmore, G., Randich, S., Asplund, M., et al. 2012, *The Messenger*, 147, 25
- Girardi, L., Groenewegen, M. A. T., Hatziminaoglou, E., & da Costa, L. 2005, *A&A*, 436, 895
- Hambly, N. C., Davenhall, A. C., Irwin, M. J., & MacGillivray, H. T. 2001, *MNRAS*, 326, 1315
- Hartkopf, W. I., & Yoss, K. M. 1982, *AJ*, 87, 1679
- Hayden, M. R., Holtzman, J. A., Bovy, J., et al. 2014, *AJ*, 147, 116
- Hayden, M. R., Bovy, J., Holtzman, J. A., et al. 2015, *ApJ*, 808, 132
- Ho, A. Y. Q., Ness, M. K., Hogg, D. W., et al. 2017, *ApJ*, 836, 5
- Høg, E., Fabricius, C., Makarov, V. V., et al. 2000, *A&A*, 355, L27
- Holtzman, J. A., Shetrone, M., Johnson, J. A., et al. 2015, *AJ*, 150, 148
- Hou, J. L., Prantzos, N., & Boissier, S. 2000, *A&A*, 362, 921
- Huang, Y., Liu, X.-W., Zhang, H.-W., et al. 2015, *Res. Astron. Astrophys.*, 15, 1240
- Huber, D., Bryson, S. T., Haas, M. R., et al. 2016, VizieR Online Data Catalog: J/ApJS/224/2
- Ivezić, Ž., Beers, T. C., & Jurić, M. 2012, *ARA&A*, 50, 251
- Jacobson, H. R., Friel, E. D., Jiřková, L., et al. 2016, *A&A*, 591, A37
- Kordopatis, G., Recio-Blanco, A., de Laverny, P., et al. 2011a, *A&A*, 535, A106
- Kordopatis, G., Recio-Blanco, A., de Laverny, P., et al. 2011b, *A&A*, 535, A107
- Kordopatis, G., Gilmore, G., Steinmetz, M., et al. 2013, *AJ*, 146, 134
- Kunder, A., Kordopatis, G., Steinmetz, M., et al. 2017, *AJ*, 153, 75
- Lee, Y. S., Beers, T. C., Sivarani, T., et al. 2008a, *AJ*, 136, 2022
- Lee, Y. S., Beers, T. C., Sivarani, T., et al. 2008b, *AJ*, 136, 2050
- Lee, Y. S., Beers, T. C., Allende Prieto, C., et al. 2011, *AJ*, 141, 90
- Lee, Y. S., Beers, T. C., Carlin, J. L., et al. 2015, *AJ*, 150, 187
- Liu, C., Xu, Y., Wan, J.-C., et al. 2017, *Res. Astron. Astrophys.*, 17, 096
- Luck, R. E., Kovtyukh, V. V., & Andrievsky, S. M. 2006, *AJ*, 132, 902
- Maciel, W. J., & Costa, R. D. D. 2010, in *Chemical Abundances in the Universe: Connecting First Stars to Planets*, eds. K. Cunha, M. Spite, & B. Barbuy, *IAU Symp.*, 265, 317
- Majewski, S. R., Zasowski, G., & Nidever, D. L. 2011, *ApJ*, 739, 25
- Majewski, S. R., Schiavon, R. P., Frinchaboy, P. M., et al. 2017, *AJ*, 154, 96
- McMahon, R., Banerji, M., Gonzalez, E., et al. 2013, *The Messenger*, 154, 35
- Mikolaitis, Š., Hill, V., Recio-Blanco, A., et al. 2014, *A&A*, 572, A33
- Ness, M., Hogg, D. W., Rix, H.-W., Ho, A. Y. Q., & Zasowski, G. 2015, *ApJ*, 808, 16
- Nidever, D. L., Bovy, J., Bird, J. C., et al. 2014, *ApJ*, 796, 38
- Nordström, B., Mayor, M., Andersen, J., et al. 2004, *A&A*, 418, 989
- Pasquini, L., Avila, G., Blecha, A., et al. 2002, *The Messenger*, 110, 1
- Quinn, P. J., Hernquist, L., & Fullagar, D. P. 1993, *ApJ*, 403, 74
- Recio-Blanco, A., Bijaoui, A., & de Laverny, P. 2006, *MNRAS*, 370, 141
- Rix, H.-W., & Bovy, J. 2013, *A&ARv*, 21, 61
- Robin, A. C., Reylé, C., Derrière, S., & Picaud, S. 2003, *A&A*, 409, 523
- Rocha-Pinto, H. J., Maciel, W. J., Scalo, J., & Flynn, C. 2000, *A&A*, 358, 850
- Rojas-Arriagada, A., Recio-Blanco, A., de Laverny, P., et al. 2017, *A&A*, 601, A140
- 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
- Schlesinger, K. J., Johnson, J. A., Rockosi, C. M., et al. 2014, *ApJ*, 791, 112
- Schönrich, R., & Binney, J. 2009a, *MNRAS*, 396, 203
- Schönrich, R., & Binney, J. 2009b, *MNRAS*, 399, 1145
- Schultheis, M., Kordopatis, G., Recio-Blanco, A., et al. 2015, *A&A*, 577, A77
- Schultheis, M., Rojas-Arriagada, A., García Pérez, A. E., et al. 2017, *A&A*, 600, A14
- SDSS Collaboration 2016, *ApJS*, submitted [arXiv:1608.02013]
- Sestito, P., Bragaglia, A., Randich, S., et al. 2008, *A&A*, 488, 943
- Sharma, S., Bland-Hawthorn, J., Johnston, K. V., & Binney, J. 2011, *ApJ*, 730, 3
- Sharma, S., Bland-Hawthorn, J., Binney, J., et al. 2014, *ApJ*, 793, 51
- Skrutskie, M. F., Cutri, R. M., Stiening, R., et al. 2006, *AJ*, 131, 1163
- Smith, M. C., Okamoto, S., Yuan, H.-B., & Liu, X.-W. 2012, *Res. Astron. Astrophys.*, 12, 1021
- Smolinski, J. P., Lee, Y. S., Beers, T. C., et al. 2011, *AJ*, 141, 89
- Snaith, O., Haywood, M., Di Matteo, P., et al. 2015, *A&A*, 578, A87
- Soubiran, C., Bienaymé, O., Mishenina, T. V., & Kovtyukh, V. V. 2008, *A&A*, 480, 91
- Stanghellini, L., & Haywood, M. 2010, *ApJ*, 714, 1096
- Steinmetz, M., Zwitter, T., Siebert, A., et al. 2006, *AJ*, 132, 1645
- Stonkutė, E., Koposov, S. E., Howes, L. M., et al. 2016, *MNRAS*, 460, 1131
- Tonry, J. L., Stubbs, C. W., Lykke, K. R., et al. 2012, *ApJ*, 750, 99
- Valenti, J. A., & Piskunov, N. 1996, *A&AS*, 118, 595
- Wojno, J., Kordopatis, G., Piffl, T., et al. 2017, *MNRAS*, 468, 3368
- Wright, E. L., Eisenhardt, P. R. M., Mainzer, A. K., et al. 2010, *AJ*, 140, 1868
- Wu, Y., Luo, A.-L., Li, H.-N., et al. 2011, *Res. Astron. Astrophys.*, 11, 924
- Wyse, R. F. G. 2016, in *Multi-Object Spectroscopy in the Next Decade: Big Questions, Large Surveys, and Wide Fields*, eds. I. Skillen, M. Barcells, & S. Trager, *ASP Conf. Ser.*, 507, 13
- Xiang, M.-S., Liu, X.-W., Yuan, H.-B., et al. 2015, *Res. Astron. Astrophys.*, 15, 1209
- Yanny, B., Rockosi, C., Newberg, H. J., et al. 2009, *AJ*, 137, 4377
- Yong, D., Carney, B. W., Teixeira de Almeida, M. L., & Pohl, B. L. 2006, *AJ*, 131, 2256
- Yuan, H.-B., Liu, X.-W., Huo, Z.-Y., et al. 2015, *MNRAS*, 448, 855
- Zacharias, N., Finch, C. T., Girard, T. M., et al. 2013, *AJ*, 145, 44
- Zasowski, G., Johnson, J. A., Frinchaboy, P. M., et al. 2013, *AJ*, 146, 81
- Zhao, G., Zhao, Y.-H., Chu, Y.-Q., Jing, Y.-P., & Deng, L.-C. 2012, *Res. Astron. Astrophys.*, 12, 723
- Zwitter, T., Matijević, G., Breddels, M. A., et al. 2010, *A&A*, 522, A54