The importance of selection functions - biases and sample selection effects in the Gaia-ESO Milky Way fields

Edita Stonkutė

Vilnius University
Institute of Theoretical Physics and Astronomy

In collaboration with S. Feltzing, L. Howes, G. Ruchti, T. Bensby, S. Koposov & C. Worley
Galactic disc and Halo

The multi-object spectroscopic surveys have been gathering data with continually improving spectral resolution and/or for larger number of stars.


- the Sloan Extension for Galactic Understanding and Exploration – SEGUE-1: 230,000 unique stars & SEGUE-2: 119,000 stars, focusing on the in situ stellar halo of the Galaxy (Yanny et al. 2009)

- the RAdial Velocity Experiment – RAVE, DR5: 457,588 unique stars (Kunder et al. 2017)

- the APO Galactic Evolution Experiment – APOGEE: 150,000 stars & APOGEE-2: expected 300,000 stars (Majewski et al. 2015; 2016)

- the Gaia-ESO Survey – GES: targeting ~ $10^5$ stars (Gilmore et al. 2012; Randich et al. 2013)

- the Large sky Area Multi-Object fiber Spectroscopic Telescope experiment LAMOST DR1: 1 944 329 stellar spectra (Lou et al. 2015)
Large-scale spectroscopic surveys

• Each survey has an **observing strategy** and a specific **target selection method** designed for survey to achieve their science goals;

• Surveys differ:
  - spectral resolution;
  - wavelength coverage;
  - selected main targets
    (giant stars, dwarf stars, clusters, streams etc.)

• These unique target selection schemes can lead to **biases** and **affect measurements** of the observed properties of the Milky Way.

(see reviews by Bland-Hawthorn & Gerhard 2016; Wyse 2016)
The selection functions

• The **selection function** is defined as:

  — the fraction of objects in a certain colour and magnitude range successfully observed spectroscopically ($N_{\text{obs}}$) compared to the underlying stellar populations ($N_{\text{phot}}$).

\[ f(S) = \frac{N_{\text{obs}}}{N_{\text{phot}}} \]

• The selection function **determines**:
  • how **representative** the observed sample is **compared** to the full existing stellar population of the Milky Way.
Quantification of the selection function of a survey has been demonstrated for:

- **GCS** by Schönrich & Binney, 2009; Sharma et al., 2014.

- **SEGUE** by Bovy et al., 2012; Cheng et al., 2012; Schlesinger et al., 2012.

- **APOGEE** by Bovy et al., 2014; Nidever et al., 2014; Anders et al., 2016.

- **LAMOST** by Carlin et al., 2012; Yuan et al., 2015.

- **RAVE** by Sharma et al., 2011, 2014; Francis, 2013; Wojno et al., 2017.

- **GES** by Stonkutė et al., 2016.

- **TGAS** by Bovy, 2017.
Selection function is important

• Most of the target selection schemes make use of simple color 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.
• We have to guarantee a reproducible and accurate selection function during the survey planning, execution, and data reduction.

The population synthesis models:
- Besançon (Robin et al., 2003), TRILEGAL (Girardi et al., 2005), GALAXIA (Sharma et al., 2011).
Applying the selection function
The SEGUE 24,270 G- and 16,847 K-dwarfs sample

- **Photometric errors:**
  - ✓ the uncertainty in each bin of the MDF is \(~6\%\) and should not create any [Fe/H] biases in MDF.

- **Undetected binarity:**
  - ✓ Does not preferentially affect stars of a particular [Fe/H], with an uncertainty of \(4\%\) for each MDF bin.

- **Subgiant contamination:**
  - ✓ For [Fe/H] \(\geq -1.0\), the change in each MDF bin is \(~3\%)\. Below [Fe/H] of -1, the change is \(~10\%)\.

- **Extinction:**
  - ✓ The uncertainty is typically around \(2\%\).

- **Volume completeness:**
  - ✓ For [Fe/H] \(\geq -2.0\), variation in the systematic and random distance will change each MDF bin by \(~3\%-7\%)\. At the metal-poor end: \(~35\%-40\%)\.

- **Bootstrap analysis:**
  - ✓ Below [Fe/H] of -1, the change is \(~40\%). For [Fe/H] \(\geq -1.0\) is typically around \(~8\%).
The target-selection of SEGUE does bias the sample in metallicity space, favoring metal-poor stars.
The selection function of the RAVE survey:
Cylindrical Galactocentric velocity components

Wojno et al. 2017

Giants ($\log(g)<3.5$)
mock- RAVE
parent Galaxia
parent Galaxia I < 12

\[ S \propto S_{\text{select}}(\alpha, \delta, I, J - K_s), \]

\[ S_{\text{pipeline}} = S_{\text{pipeline}}(T_{\text{eff}}, \log g, [\text{Fe/H}]) \]

\[ S = S_{\text{pipeline}} \times S_{\text{select}}. \]
The selection function of the RAVE survey: Cylindrical Galactocentric velocity components

- The Selection function (SF) in the magnitude range $9 < I < 12$ does not intrinsically induce biases in the kinematics with respect to expectations from the Besançon model available in Galaxia.

  $4000 \text{K} < T_{\text{eff}} < 8000 \text{K}$,
  $0.5 < \log g < 5$.

- Some small biases when a parent sample extending to $I = 13$.
The selection function of the RAVE: Metallicity distribution

- The SF in the magnitude range $9 < I < 12$ does not intrinsically induce biases in the chemistry of stars with respect to expectations from Galaxia.

$$4000 \text{K} < T_{\text{eff}} < 8000 \text{K}, \quad 0.5 < \log g < 5.$$  

- RAVE stars provide unbiased samples in terms of kinematics and metallicities that are well suited for kinematic modeling without taking into account the detailed selection function via volume corrections.
Combining the data from different surveys

- The effect of SF on the MDF and on the vertical metallicity gradient by studying **similar lines of sight:**
  - APOGEE (Majewski et al. 2015)
  - LAMOST (Deng et al. 2012)
  - RAVE (Steinmetz et al. 2006)
  - GES (Gilmore et al. 2012)

- Stellar population synthesis models:
  - GALAXIA (Sharma et al. 2011)
  - TRILEGAL (Girardi et al. 2005)
Unlike in the case of MDF, we find that the source distribution of vertical metallicity gradient. In TRILEGAL, this is found to be different scale heights, leading to the discs, combined with their different Galaxy components like thin and thick.

As mentioned in Section 5.2, the mask sample is made by randomly choosing the model sources within each 0.05 mag bin by sets between them.

Here we estimate and compare the vertical metallicity gradient for each survey independently after accounting for metallicity obtained largely from Poisson noise.

Vertical metallicity gradients measured for mask and magnitude sample for GALAXIA and TRILEGAL in ALR. We further determine the vertical metallicity gradient using both stellar population synthesis models.

**MDF:**

\[7 < R < 9, \, [\text{kpc}]\]

\[0 < |Z| < 2, \, [\text{kpc}]\]

_1
giant-to-dwarf ratio for mask and observed samples of each survey for both models._

_Nandakumar et al., 2017_
Vertical metallicity gradients

Nandakumar et al., 2017

GES 2017: Gaia-ESO Survey Fourth Science Meeting
Combining the data from different surveys: Results

• There is a negligible selection function effect on the MDF and the vertical metallicity gradients for APOGEE, RAVE, and LAMOST using GALAXIA & TRILEGAL stellar population synthesis models.

• GES suffers from number statistics that are too low to be conclusive.

• Therefore, it is possible to combine common fields of different surveys in studies using MDF and metallicity gradients provided their metallicities are brought to the same scale.

Nandakumar et al., 2017
Tycho-Gaia Astrometric Solution (TGAS) selection function

48 % of the sky

\[ S(J, J-K_s, \text{Ra}, \text{Dec}) = \frac{\text{# of stars in TGAS}}{\text{# of stars in 2MASS}} \]

Gaia-ESO Survey

Target selection of the Milky Way fields

More & more details!
GES MW field target selection schemes

• VISTA VHS for GIRAFFE:
  - Blue box:
    \[0.00 \leq (J - K_S) \leq 0.45;\]
    \[14.0 \leq J \leq 17.5.\]
  - Red box:
    \[0.40 \leq (J - K_S) \leq 0.70;\]
    \[12.5 \leq J \leq 15.0.\]

• 2MASS PSC for UVES:
  - Blue box:
    \[0.23 \leq (J - K_S) \leq 0.45;\]
    \[12.0 \leq J \leq 14.0.\]

The line-of-sight interstellar extinction was taken into account by shifting the colour-boxes of GIRAFFE targets by 0.5*E(B − V) using Schlegel et al. (1998) maps.
Case 1 – too few stars to fill all fibres using our boxes:

**GIRAFFE:**

Extra box:

\[ 0.00 \leq (J - K_S) \leq 0.45 + \Delta_G; \]

\[ J \geq 14.0; \]

\[ J + 3.0 \times ((J - K_S) - 0.35) \leq 17.50. \]

**UVES:**

Extra box:

\[ 0.23 \leq (J - K_S) \leq 0.45 + \Delta_U; \]

\[ J \geq 12.0; \]

\[ J + 3.0 \times ((J - K_S) - 0.35) \leq 14.00. \]

Here \( \Delta_G \) and \( \Delta_U \) are the red-edge extensions of the blue box for GIRAFFE and UVES.
The frequency distribution of extensions $\Delta G$ and $\Delta U$

**The dashed line** shows the right-edge extensions $\Delta G$ for GIRAFFE.

**The Solid line** shows the right-edge extensions $\Delta U$ for UVES.
GES_MW_142000-050000 centred at $l = 339.9^\circ$ & $b = 51.4^\circ$, FoV = 35 arcmin in diameter.
Case 2 – too many stars to allocate all fibres using boxes:

**GIRAFFE:**

Blue box:
- \(0.00 \leq (J - K_S) \leq 0.45\);
- \(14.0 \leq J \leq 17.5\).

Red box:
- \(0.40 \leq (J - K_S) \leq 0.70\);
- \(12.5 \leq J \leq 15.0\).

**UVES:**

Blue box:
- \(0.23 \leq (J - K_S) \leq 0.45\);
- \(12.0 \leq J \leq 14.0\).

The blue box is divided in \(4 \sim\) equal-size magnitude bins, with \(J_{1,2,3,4} = (J_{\text{max}} - J_{\text{min}})/4\).
GES_MW_201959-470000 centred at $l = 352^\circ.7$ & $b = -34^\circ.2$, FoV = 35 arcmin in diameter.

GIRAFFE Case 2

UVES Case 2
Way fields in iDR4. For these fields targets were selected as outlined in Sections 3.1 and 3.2.

The adopted data quality flags to select Milky Way field targets from the VHS catalogue are as presented in Section 3.2.4 Naming conventions.

VISTA VHS has a sufficient sky coverage to meet the full science goals of the survey. The VHS survey data consists of three survey components: VHS Galactic Plane Survey (VHS-GPS); VHS-ATLAS and VHS-ESO Survey up to 2015 June and the number of Milky Way fields observed with UVES. Green circles – fields with the selection based on VISTA VHS photometry; blue circles – fields with the selection based on 2MASS photometry; and red triangles – fields with the selection based on SkyMapper photometry in order to study metal-poor stars and K giants. Some details on the selection of these targets are presented in Sections 3.1 and 3.2.

The target selection magnitude and colour limits for GIRAFFE are listed in Table 1131–1146 (2016). The distribution of observed Milky Way fields in the Galactic Plane is shown in Figure 1131–1146 (2016) with the airmass. Fig.

Case 1

(a) Milky Way fields, GIRAFFE Case 1.

(c) Milky Way fields, UVES Case 1.

Case 2

(b) Milky Way fields, GIRAFFE Case 2.

(d) Milky Way fields, UVES Case 2.
Allocation of the GES MW targets

**Case 1:** GE_MW_142000_050000 centered at $l=339.9^\circ$ and $b=51.40$

**Case 2:** GE_MW_201959_540000 centered at $l=352.7^\circ$ and $b=-34.20$

- Grey dots are targets in a 1-square-degree FoV.
- Potential targets with 35' FoV in diameter.
- Allocated FLAMES targets, with 25' FoV in diameter.
The CMD Weights: targeted and allocated weights

• Primary target selection:

\[ W_{T,F} = \frac{N_T}{N_F}, \]

- \( N_T \) is the number of targeted objects in the field within 35 arcmin FoV in diameter.
- \( N_F \) is the number of objects in the field within a 1 deg FoV in diameter.
- \( W_{T,F} \) is the weight of targeted objects versus objects in the 1 deg FoV in diameter field.

• Final target selection:

\[ W_{A,T} = \frac{N_A}{N_T}, \]

- \( N_A \) is the number of allocated objects in the field within the FLAMES 25 arcmin FoV in diameter.
- \( W_{A,T} \) is the weight of allocated objects versus targeted for a given Milky Way field.
The CMD weights

Figure 11. Weights of each field for targeted stars versus stars in the field within a 1 deg FoV in diameter compared with weights for allocated versus targeted stars in (a) blue, (b) red and (c) extra boxes for GIRAFFE. The colour coding indicates Galactic latitude in degrees.

Figure 12. Weights of each field for targeted stars versus stars in the field within a 1 deg FoV in diameter compared with weights for allocated versus targeted stars in (a) blue and (b) extra boxes for UVES. The colour coding indicates Galactic latitude in degrees.

Figure 13. Weights of each field for targeted stars versus stars in the field within a 1 deg FoV in diameter compared with weights for allocated versus targeted stars in blue box for GIRAFFE. (a)–(d) show weights in $J_{1–4}$ magnitude bins in fields within given FoV. The colour coding indicates Galactic latitude in degrees.

In order to use Milky Way field stars for a specific science question, we must understand how the spectroscopic sample is drawn from the underlying population. As can be seen from Fig. 15, the completeness of the Gaia-ESO Survey varies substantially between fields. For each field, we must therefore assess how representative the spectroscopic sample is of the underlying population. To correct for these types of biases the presented weights, $W_{T, F}$, should be used.

7.2 Using iDR4: stellar weights for the colour–magnitude diagram

The selection function presented in this paper corrects for the discrepancy between the number of stars allocated to be observed and the number of stars originally available from the photometry for each field. This enables any comparison of fields to account for the varying population densities associated with different lines of sight. In order to ensure a completely fair comparison of the data, however, a second correction is needed. Within each field, the density of stars available for observation varies considerably with respect to both colour and magnitude. Furthermore, not all observed stars end up with reasonable parameters; a significant proportion of observations fail to produce high enough quality spectra to enable robust stellar parameter determination. Naturally, the fainter targets...
The CMD grids

GIRAFFE Blue box bins of size: 0.05 in \((J - K_s)\) & 0.4375 in \(J\). GIRAFFE Red & UVES blue boxes: 0.05 in \((J - K_s)\) & 0.5 in \(J\).

- Weights of successfully observed targets:

\[
W_{O_{\text{bin}},F_{\text{bin}}} = \frac{N_{O_{\text{bin}}}}{N_{F_{\text{bin}}}} ,
\]

- \(N_{O_{\text{bin}}}\) is the number of successfully observed targets in that bin.
- \(N_{F_{\text{bin}}}\) is the number of objects in the VISTA VHS or 2MASS photometry for that bin.
We can therefore, when trying to accurately sample the Milky Way disc, characterize the importance of an observed star in iDR4 as 1/

We have discussed the details of the selection function for the Milky Way field stars observed in the Gaia-ESO Survey iDR4 and (black dashed line) when all Milky Way fields are analysed together. For example, looking at the radial metallicity distribution the weights can be used to correct for any incompleteness effects on the survey. Furthermore, correct for any incompleteness effects on the survey. For each bin as a running average. However, instead of spectroscopic fibre, i.e. FLAMES/GIRAFFE fibres, in order to allocate on FLAMES/GIRAFFE and FLAMES/UVES.

The total population. In this case, we do not want stars with very small magnitude, we can perform a weighted average, and compute the mean of the data in Galactocentric radial distance potentially selected targets but also on allocated targets. It is crucial to limit the data to only those stars that most represent the underlying population. A simple way for allocation on FLAMES/GIRAFFE and FLAMES/UVES.

Total, normalize between the different lines of site; while each field in a given FoV. In this example we chose to remove stars more than in the previous example. For example, looking at the radial box of Milky Way fields were selected to target rare but astrophysically important stellar populations (e.g. metal-poor stars, K giants), where the density of stars is not enough to fill the FLAMES fibres). Case 2 select second priority targets (i.e. for those Milky Way fields where the VHS (McMahon et al. 2006) and/or vertical metallicity distribution the weights can be used to correct versus uncorrected (see Figs. 19).

The presented field CMD weights, the total weight, and normalise between the different lines of sight.

Photometric sample – black contours;
Spectroscopic sample – green contours;
Spectroscopic sample w/the SNR > 20 – yellow contours.

GES 2017: Gaia-ESO Survey Fourth Science Meeting
The GES Milky Way fields MDF

(W_{Total} = W_{A,T} \times W_{T,F} \times W_{O_{bin},F_{bin}},

The blue & red lines show the observed MDF; the black dashed lines - weighted MDF.)
Summary

• The chosen photometric input catalogue(s) of the sources play an important role in selecting the stars to be observed for each survey.

• We have to guarantee a reproducible and accurate selection function during the survey planning, execution, and data reduction.

• The Gaia-ESO survey target selection is based on stellar magnitudes and colours, using photometry from the VISTA Hemisphere Survey (McMahon et al. 2013) and 2MASS photometry (Skrutskie et al. 2006).

• The actual target selection scheme is divided into two cases:
  • In Case 1 target selection algorithm, in addition to the two main selection CMD boxes (i.e. blue, red) extends the colour limits of blue box to select second priority targets (i.e. for those Milky Way fields where the density of stars are not enough to fill the FLAMES fibres).
  • Case 2 is used to select targets near the Galactic plane. In this case the target selection algorithm is configured to select the same number of targets per magnitude bin (i.e. not to have bias towards very faint stars).

• We presented the CMD weights witch could be used to understand Gaia-ESO Survey results and correct it for selection biases.
Thank you!