

The *Gaia*-ESO Survey: matching chemodynamical simulations to observations of the Milky Way

B. B. Thompson,^{1,2,3,4★} C. G. Few,^{2,4,5} M. Bergemann,⁶ B. K. Gibson,^{2,4★}
 B. A. MacFarlane,¹ A. Serenelli,⁷ G. Gilmore,⁸ S. Randich,⁹ A. Vallenari,¹⁰
 E. J. Alfaro,¹¹ T. Bensby,¹² P. Francois,¹³ A. J. Korn,¹⁴ A. Bayo,¹⁵ G. Carraro,¹⁶
 A. R. Casey,¹⁷ M. T. Costado,⁹ P. Donati,¹⁸ E. Franciosini,¹⁶ A. Frasca,¹⁹
 A. Hourihane,⁸ P. Jofré,⁸ V. Hill,²⁰ U. Heiter,¹⁴ S. E. Koposov,⁸ A. Lanzafame,^{19,21}
 C. Lardo,²² P. de Laverny,²³ J. Lewis,⁸ L. Magrini,⁹ G. Marconi,¹⁶ T. Masseron,⁸
 L. Monaco,²⁴ L. Morbidelli,⁹ E. Pancino,⁹ L. Prisinzano,²⁵ A. Recio-Blanco,²³
 G. Sacco,⁹ S. G. Sousa,²⁶ G. Tautvaišienė,²⁷ C. C. Worley⁸ and S. Zaggia¹⁰

Affiliations are listed at the end of the paper

Accepted 2017 September 4. Received 2017 September 4; in original form 2016 December 10

ABSTRACT

The typical methodology for comparing simulated galaxies with observational surveys is usually to apply a spatial selection to the simulation to mimic the region of interest covered by a comparable observational survey sample. In this work, we compare this approach with a more sophisticated post-processing in which the observational uncertainties and selection effects (photometric, surface gravity and effective temperature) are taken into account. We compare a ‘solar neighbourhood analogue’ region in a model Milky Way-like galaxy simulated with RAMSES-CH with fourth release *Gaia*-ESO survey data. We find that a simple spatial cut alone is insufficient and that the observational uncertainties must be accounted for in the comparison. This is particularly true when the scale of uncertainty is large compared to the dynamic range of the data, e.g. in our comparison, the [Mg/Fe] distribution is affected much more than the more accurately determined [Fe/H] distribution. Despite clear differences in the underlying distributions of elemental abundances between simulation and observation, incorporating scatter to our simulation results to mimic observational uncertainty produces reasonable agreement. The quite complete nature of the *Gaia*-ESO survey means that the selection function has minimal impact on the distribution of observed age and metal abundances but this would become increasingly more important for surveys with narrower selection functions.

Key words: methods: numerical – Galaxy: abundances – galaxies: evolution – galaxies: formation.

1 INTRODUCTION

The characteristic abundance ratios found in different stellar populations provide us with an opportunity to uncover the history of galaxy formation. Using what is known as galactic archaeology to link the chemistry, ages and dynamics of stars allows us to trace the origins of the components of the Milky Way (Eggen, Lynden-

Bell & Sandage 1962; Freeman & Bland-Hawthorn 2002). We have learned a great deal about the processes associated with galaxy formation using the essential tools of chemical evolution models and simulations of galaxy formation (e.g. Scannapieco et al. 2005; Sommer-Larsen & Fynbo 2008; Sánchez-Blázquez et al. 2009; Kobayashi & Nakasato 2011; Wiersma, Schaye & Theuns 2011; Calura et al. 2012; Pilkington et al. 2012; Few et al. 2012b; Gibson et al. 2013; Miranda et al. 2016) and semi-analytic tools (Calura & Menci 2009; Yates et al. 2013). More recently, we have gone beyond tracing dynamics and global metallicity within simulations to include chemical evolution in such a way that individual elements and

* E-mail: ben@benjaminbthompson.com (BBT); brad.gibson@hull.ac.uk (BKG)

isotopes can be traced in combination with self-consistent galaxy formation scenarios (e.g. Steinmetz & Muller 1995; Lia, Portinari & Carraro 2002; Kawata & Gibson 2003; Tornatore et al. 2004; Oppenheimer & Davé 2008; Few et al. 2014).

Comparison of these chemodynamical models with observed trends is fundamental to establishing the validity of the models and understanding the observations. A wealth of high precision observational data is required for such comparisons to be made, and thus detailed testing of chemical evolution models is only achievable through comparison to Milky Way stars with external galaxies providing only mean trends. Yet despite improvements to the abundance of observational data sets and in simulation resolution, the way in which these comparisons are conducted has remained unaltered for decades. It is straightforward, and indeed common to simply take the results of a simulation of a Milky Way-like galaxy and compare it like-for-like with observations of the Milky Way itself. Typically, a spatial region in a simulation that is similar to the one covered by observational data of interest is sampled and the stellar properties are directly compared (e.g. Martínez-Serrano et al. 2008; Kobayashi & Nakasato 2011; Calura et al. 2012; Few et al. 2014).

One strong argument against this simple comparison method is that it ignores observational biases. First, the observed data sets have inherent uncertainties, either systematic (because of the stellar model atmospheres) or random (instrumental effects). Secondly, observational surveys usually observe stars within some range of stellar parameters or distances, which is usually dictated by the intention to study specific types of stars (low- or high-mass, low- or high-metallicity) in certain Galactic populations. This selection function (Stonkutė et al. 2016) creates biases in the distribution functions of the observed data set. Most commonly, selection based on colour and apparent magnitude of stars is reflected in the shape of the metallicity and age distribution functions (Bergemann et al. 2014).

In galaxy formation simulations, stellar properties are typically represented by ‘star particles’, which describe the combined properties of a coeval group of stars (a simple stellar population), its total stellar mass and metallicity.¹ Thus, one is limited primarily to the integrated luminosities and averaged chemical composition on the scale of open clusters within simulations, i.e. one star particle represents the mean properties of an open cluster.

As models improve, the detailed distribution of stellar ages and metallicities – in addition to their mean – become increasingly important. It is thus crucial that the approach to derive ‘observables’ from the simulated data for comparison with real observations is as close as possible to the methodology employed by observers.

In this work, we discuss a ‘solar neighbourhood analogue’ region in a model Milky Way-like galaxy simulated with the RAMSES-CH code (Teyssier 2002; Few et al. 2012a, 2014), which is post-processed using the SYNCMD toolkit (Pasetto, Chiosi & Kawata 2012) to mimic observational selection functions. The simulated data are compared with the *Gaia*-ESO spectroscopic stellar survey (Gilmore et al. 2012; Randich et al. 2013). The *Gaia*-ESO survey is the largest ongoing high-resolution spectroscopic survey of stars in the Milky Way. In the high-resolution ($R \sim 47\,000$) mode, the goal is to acquire spectra for about 5 000 field stars, probing distances ~ 2 kpc from the Sun. Here, we use the results from the fourth data release of the survey (hereafter, *GES-IDR4*), which includes all

stellar spectra for the first 18 months of the survey. Our simulated solar neighbourhood analogue encapsulates a 2 kpc spherical region of space in our simulated galaxy.

We apply different degrees of post-processing on the simulated data to mimic observational effects. Within our simulation data, we sample a spatial region analogous to the solar neighbourhood region covered by the *Gaia*-ESO survey and discuss three different methods of transforming the simulated data into the ‘observer plane’. Our first method is simply the traditional spatial selection of the solar neighbourhood analogue. The second method applies a scatter to the age, metallicity and Mg abundance based on the uncertainty in the observed data sets. The final method applies SYNCMD in addition to the observationally motivated scatter of the data to include the survey selection functions.

The paper is organized as follows. We describe the methodology employed in this work in Section 2 and the chemodynamical simulation code used in Section 2.1, the properties of the simulated galaxy in Section 2.2 and the SYNCMD toolkit in Section 2.3. We describe the *Gaia*-ESO survey data used in this work in Section 2.4 and the data reduction and the observational survey selection function in Section 2.5. We introduce the methods of sampling the simulated solar neighbourhood analogue in Section 2.6 and expand on that for the unaltered simulated galaxy in Section 2.6.1 convolving that with *GES-IDR4* errors in Section 2.6.2 and the application of SYNCMD in Section 2.6.3. Finally we describe our results in Section 3 and conclude with Section 4.

2 METHODOLOGY

In this work, we use SYNCMD to ‘observe’ a cosmologically simulated galaxy to demonstrate the significance of different aspects of observational effects when comparing models with empirical distributions. In this section, we describe the simulations, initial conditions and the observations as well as the post-processing applied to mimic observational effects.

2.1 RAMSES – CH

Our cosmological simulations are performed with the grid code RAMSES (Teyssier 2002). To trace the chemical evolution of the simulated galaxy we employ a chemodynamical patch called RAMSES-CH (Few et al. 2012a, 2014). RAMSES-CH is able to perform N -body and hydrodynamical simulations including stars, dark matter and gas. The adaptive mesh refinement method used in RAMSES allows for refinement of the grid on a cell-by-cell basis increasing the resolution in dense regions of the volume. This refinement allows for a reduction in computing time while maintaining a high-resolution spatial grid around the galaxy and also capturing large-scale cosmological phenomena. RAMSES-CH includes treatments of self-gravity, hydrodynamics, star formation, supernova feedback, gas cooling and chemical enrichment.

Dense gas cells form star particles if the density surpasses a number density threshold of $n_0 = 0.1 \text{ cm}^{-3}$. The rate at which the stellar population particles are produced is $\dot{\rho}_* = \epsilon_* \rho_g / t_{\text{ff}}$, where $t_{\text{ff}} = (3\pi/32G\rho_g)^{1/2}$ is the local gas free-fall time, ρ_g is the gas mass density and star formation efficiency, $\epsilon_* = 0.01$. Our choices of both n_0 and ϵ_* are the same as in Few et al. (2014), where ϵ_* is chosen to reproduce the Schmidt–Kennicutt relation (Schmidt 1959; Kennicutt 1998). The particles used to trace the stellar mass phase are commonly and colloquially referred to as ‘star particles’ by the community, however, they do not represent single stars but coeval stellar populations. In the simulation presented here, they

¹ In this work, metallicity is defined as the iron abundance, [Fe/H].

have a birth mass of $3.3 \times 10^4 M_{\odot}$. To avoid confusion, we define these coeval stellar populations in the simulations as ‘stellar population particles’. In contrast, the synthetic particles representing individual stars (see Section 2.3) are defined as ‘synthetic star particles’.

Radiative gas cooling in the simulation is metallicity- and density-dependent. Cooling rates are calculated assuming photoionization equilibrium with a redshift-dependent uniform UV background (Haardt & Madau 1996). Cooling rates due to the presence of metals are calculated from the total metallicity of the gas which is interpolated from the CLOUDY (Ferland et al. 1998) cooling rates at zero and solar metallicity at temperatures exceeding 10^4 K; colder gas takes its metal cooling rates from Rosen & Bregman (1995). We also employ the delayed cooling feedback mechanism from Teyssier et al. (2013) to account for the unresolved multiphase nature of the gas and avoid the spurious loss of thermal energy following SN feedback. In addition to these prescriptions for gas heating/cooling, we also impose a polytropic equation of state as a temperature floor. This temperature floor prevents the gas from reaching the low temperatures at which the Jeans length of the gas is unresolved and unphysical fragmentation may occur. The gas is therefore prevented from falling below $T_{\min} = T_{\text{th}}(n_g/n_0)^{\gamma-1}$ where $\gamma = 2$, $T_{\text{th}} = 188$ K and $n_0 = 0.1 \text{ cm}^{-3}$.

RAMSES-CH allows us to track the elements H, He, C, N, O, Ne, Mg, Si and Fe from their dominant production sites (SNII, SNIa and AGB stellar winds) into the ISM where they are advected with the gas flow and become imprinted on the stellar population particles. The details of RAMSES-CH are described fully in Few et al. (2014) but we briefly summarize the main components here. Energetic feedback from both type Ia and type II supernovae (SNIa and SNII, respectively) is included with each SN injecting 10^{51} erg as thermal energy into the local grid cell, AGB stars eject their mass passively into the enclosing grid cell. We use the model B SNII yields of Woosley & Weaver (1995) with a correction applied to the yields after Timmes, Woosley & Weaver (1995) which halves the quantity of Fe produced by massive stars and AGB yields are taken from van den Hoek & Groenewegen (1997). We consider stars in the mass range $0.1\text{--}8 M_{\odot}$ to evolve along the AGB while stars with masses $8\text{--}100 M_{\odot}$ eject mass and energy as SNII.

The mass distribution of stars in each stellar population particle is determined by the initial mass function (IMF). In this work, we use the Salpeter (1955) IMF, where we treat the IMF as a single power law of slope -1.35 with lower and upper mass limits of 0.1 and $100 M_{\odot}$, respectively. The number of SNIa per unit initial stellar mass is also determined by the IMF via the number of stars with masses $3\text{--}8 M_{\odot}$ in binary systems with either a red giant or main sequence star. The lifetime of these systems is taken as the main sequence lifetime of the secondary star (Kodama & Arimoto 1997). This combination of chemical evolution model parameters is described in Few et al. (2014) as model *S55-uM100-IaK* where the impact of the choice of IMF and SNIa model on simulations similar to that presented here is also discussed. In brief, the particular IMF that is chosen in the model offsets the [O/Fe] ratio by up to 0.25 dex as a result of the greater number of SNII progenitors created by a top-heavy IMF. The Salpeter IMF chosen for this work results in an [O/Fe] and [Fe/H] distribution that lies between the two rather extreme IMFs from Kroupa, Tout & Gilmore (1993) and Kroupa (2001). The Chabrier (2003) IMF is now occasionally favoured for chemical evolution and galaxy modelling. The similarity of the Chabrier (2003) IMF to that of Kroupa (2001), which is used in Few et al. (2014) means that we can estimate that the mean [Mg/Fe] is lessened by only

around 0.05 dex through our choice of IMF than if we opted for Chabrier (2003).

There are now numerous hydrodynamical simulation codes to choose from: grid- and particle-based codes and the recently emerging moving-mesh and meshless approaches (Springel 2010; Hopkins 2015). The strengths and weaknesses of these codes are explored in idealized test cases (Agertz et al. 2007; Price 2008; Tasker et al. 2008), but also in cosmological galaxy formation (Scannapieco et al. 2012; Kim et al. 2014) and in isolated galaxy discs (Hopkins 2015; Few et al. 2016). Specifically, a key property of our code, RAMSES-CH, is its ability to capture metal mixing. This is extremely pertinent to this work as it directly affects the dispersion in the abundance ratios of the gas which becomes imprinted on the stars. In general, grid-based codes handle metal diffusion better than particle-based codes (Pilkington et al. 2012; Revaz et al. 2016) which makes RAMSES-CH a reasonable choice for the study of the chemical evolution of galaxies.

2.2 Galaxy initial conditions: *Selene-CH*

We employ a cosmological ‘zoom-in’ simulation technique using RAMSES-CH to simulate the galaxy: ‘*Selene-CH*’. This galaxy exists in a box $20 h^{-1}$ Mpc in size created with cosmological parameters $(H_0, \Omega_m, \Omega_{\Lambda}, \Omega_b, \sigma_8) = (70 \text{ km s}^{-1}, 0.28, 0.72, 0.045, 0.8)$ based on the WMAP 7-year results (Komatsu et al. 2011) and the simulation is run to $z = 0$. The adaptive grid can refine up to 17 levels corresponding to a maximum resolution of 218 pc with a dark matter particle mass resolution of $5.64 \times 10^6 M_{\odot}$ and a stellar population particle birth mass of $3.3 \times 10^4 M_{\odot}$.

The galaxy presented here is a chemodynamical re-simulation of the *Selene* initial conditions first presented in Few et al. (2012b). The feedback scheme used in simulating *Selene-CH* is different to the original version and so, while the galaxy has roughly the same environment and assembly history, some differences are to be expected. The evolution of the stellar distribution in this resimulation of *Selene* is analysed in Dobbs et al. (2017).

The galaxy inhabits a dark matter halo with a mass of $5.245 \times 10^{11} M_{\odot}$ and has a stellar mass of $5.603 \times 10^{10} M_{\odot}$. The dark matter halo mass may not be a precise match to the halo mass of the Milky Way but in a previous study using the initial conditions for this galaxy (Few et al. 2012b) it is found that assembly history is far more significant than small changes in halo mass. *Selene* as a galaxy is selected because of its environment (more than 3 Mpc distant from any other haloes more massive than $3 \times 10^{11} M_{\odot}$) and quiescent assembly history, (no major mergers after redshift $z = 1.0$) that means it lends itself to comparison with the Milky Way. The halo and its properties are identified using the AMIGA halo finder (Gill, Knebe & Gibson 2004; Knollmann & Knebe 2009). The assembly history of the original version of *Selene* is described in Few et al. (2012b) and more extensively with relation to the effect of its assembly on the metallicity and age distribution in Ruiz-Lara et al. (2016).

A gas surface density projection of *Selene-CH* is shown in Fig. 1 demonstrating the presence and shape of the spiral arms. The cross at $x = 4.0$ kpc and $y = 6.93$ kpc is the region where we place our simulated observer as described in Section 2.6.1, 8 kpc from the galactic centre in a spiral arm. We use stars from a spherical region 2 kpc in radius around this point which is treated as our simulated solar neighbourhood analogue. The size of this region is discussed in Section 2.6.1.

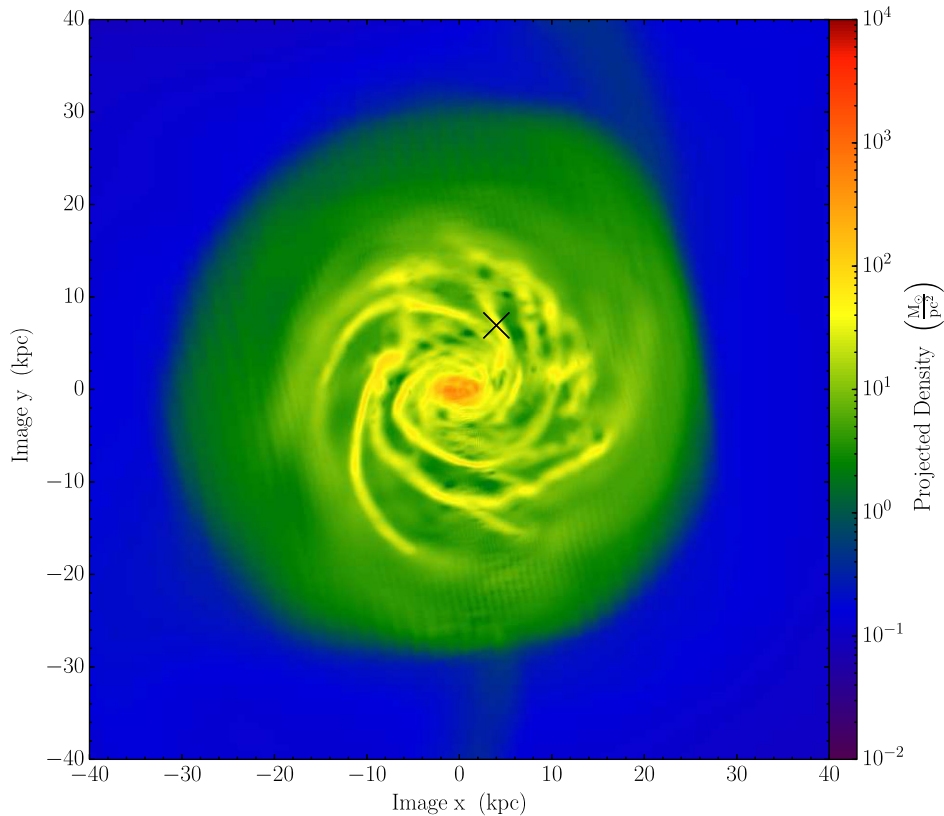


Figure 1. Face on gas surface density projection of *Selene-CH*. The galaxy was visualized using the yt visualization toolkit (Turk et al. 2011) with a projection depth of 20 kpc. The black cross at $x = 4.0$ kpc and $y = 6.93$ kpc is the position of our simulated observer as described in Section 2.6.1. The region selected in this work is a 2 kpc sphere around the indicated position, similar to the coverage of the *Gaia*-ESO survey.

2.3 SYNCMD

The SYNCMD synthetic stellar populations generation tool (Pasetto et al. 2012) is a toolkit designed to examine simulation data in a similar manner to how an observational survey would sample real life stellar populations. The toolkit is used to apply observationally motivated selection functions to simulated stellar population particles. As discussed in Section 2.1, each such particle represents a coeval monoabundance stellar population, its mass is simply the total stellar mass. The details of the SYNCMD code are given in Pasetto et al. (2012), and we summarize the process here. A preliminary application of SYNCMD has been undertaken using a RAVE-like selection function (Miranda, Macfarlane & Gibson 2014).

SYNCMD allows us to split a stellar population particle into individual stars by stochastically populating a colour–magnitude diagram (CMD). Due to resolution limits in simulations, stellar population particles represent ‘averaged’ stellar populations rather than individual stars. Stellar population particles typically have a mass of $\sim 10^4$ – $10^6 M_{\odot}$ and therefore a stochastic approach is valid. We consider each stellar population particle to consist of 10^5 synthetic star particles, which sample the IMF from a mass of $0.15 M_{\odot}$ to the main sequence turnoff mass at the current age of the stellar population particle, up to a maximum of $20 M_{\odot}$. All stellar population particles consist of this number of synthetic star particles and we weight their contribution to distribution functions by the initial mass of the stellar population particle and by the mass remaining in that particle as a fraction of the initial mass, calculated by integrating over the Salpeter (1955) IMF.

The physical properties of the synthetic stellar particles are based on theoretical stellar models. These properties are T_{eff} , $\log(g)$, magnitudes in different colour bands, ages, metal abundances and masses. Here, we use Bertelli et al. (2008, 2009) isochrones, which cover a wide grid of helium and metal abundances (Y and Z , respectively), an enrichment ratio $\Delta Y/\Delta Z$, and include mass loss by stellar wind and the thermally pulsing AGB phase (Marigo & Girardi 2007). The isochrones are used to calculate a data base of simple stellar populations using a modified version of YZVAR, which has been used in many studies (for instance Bertelli et al. 2003, and references therein). We place the observer in a region of the simulated galaxy analogous to the location of the Sun in the Milky Way and generate synthetic stars that trace a synthetic CMD by linearly interpolating in age and metallicity between isochrones of simple stellar populations. The interpolation is described in Pasetto et al. (2012).

This methodology enables each individual stellar population particle to be mapped to 10^5 synthetic star particles. The mean stellar properties of age, metallicity and metal abundance are that of the parent stellar population particle. The synthetic star particles are allocated masses at random from the chosen IMF. The masses of these particles are then used to populate an isochrone using the metal abundance and age of the original stellar population particle from the simulation data. Properties of the synthetic star particles are calculated from the data base and from the stellar population particle’s age, metallicity and distance from the observer. The synthetic star particle’s properties that are calculated are the age, luminosity, T_{eff} , $\log(g)$, metal abundance, H abundance, He abundance

and magnitudes in the *UBVRIJHK* bands. Photometric colours and magnitude values for each synthetic star will be adjusted according to the distance of the star to the simulated observer. The stars retain the age and chemical abundances of the simulation particle they are created from. We eliminate synthetic star particles from the sample that do not fall within the selection criteria. For comparison with an observational data set, one can apply a selection function to the synthetic star particles of the observational survey that the user wishes to emulate. The remaining stars are used to analyse whatever physical property one wishes, having essentially removed the fraction of each stellar population particle that would not lie within the selection functions.

2.4 The *Gaia*-ESO Survey

In this work, we focus on the high-resolution *GES-IDR4* UVES data of the field stars, for which accurate effective temperatures T_{eff} , surface gravities $\log(g)$, $[\text{Fe}/\text{H}]$ and Mg abundances are available. These targets were chosen according to their colours to maximize the fraction of un-evolved foreground (FG) stars within 2 kpc in the solar neighbourhood (see Stonkutė et al. 2016, for more details on target selection). The selection box was defined using the 2MASS photometry (Skrutskie et al. 2006; Huchra et al. 2012): $12 < J < 14$ and $0.23 < J-K < 0.45 + 0.5E(B-V)$. Application of the colour excess $E(B-V)$ and target selection are described in Stonkutė et al. (2016). According to these selection criteria, the majority of stars are FG stars with magnitudes down to $V = 16.5$.

For the analysis of the spectra, several state-of-the-art spectrum analysis codes are used (Smiljanic et al. 2014). The observed spectra were processed by 13 research groups within the *Gaia*-ESO survey collaboration with the same model atmospheres and line lists (Heiter et al. 2015), but different analysis methods: full spectrum template matching, line formation on-the-fly and the equivalent width method. The model atmospheres are 1D LTE spherically symmetric ($\log(g) \leq 3.5$ dex) and plane-parallel ($\log(g) \geq 3.5$ dex) MARCS (Gustafsson et al. 2008). The final parameter homogenization involves a multistage process, in which both internal and systematic errors of different data sets are carefully investigated. Various consistency tests, including the analysis of stellar clusters, benchmark stars with interferometric and asteroseismic data, have been used to assess each group’s performance. A comprehensive description of the pipelines and tests of the UVES results can be found in Smiljanic et al. (2014), specifically sections 4–7 of that paper.

The final stellar parameters are median of the multiple determinations, and the uncertainties of stellar parameters are median absolute deviations, which reflect the method-to-method dispersion. For most stars, the uncertainties are within 100 K in T_{eff} , 0.15 dex in $\log(g)$, and 0.1 dex in $[\text{Fe}/\text{H}]$ and Mg abundances. This accuracy could be achieved because of very careful selection of diagnostics features, very broad wavelength coverage and good signal-to-noise ratio of the observed spectra, and validation of the results on the accurate stellar parameters and NLTE estimates of chemical abundances of the *Gaia* Benchmark stars (Jofré et al. 2015). In this work, we focus on the high-resolution UVES data of the field stars. UVES is the ultraviolet and visual cross-dispersed echelle spectrograph installed at the second unit telescope of the VLT (Dekker et al. 2000). The stars were observed using the UVES U-580 setting, which covers the wavelength range from 480 to 680 nm, with a small beam-splitter gap at 590 nm. Most spectra have signal-to-noise ratio between 30 and 100 per pixel. For these stars, accurate effective temperatures T_{eff} , surface gravities $\log(g)$, $[\text{Fe}/\text{H}]$, and Mg abundances are available. The distribution of the *GES-IDR4* sample

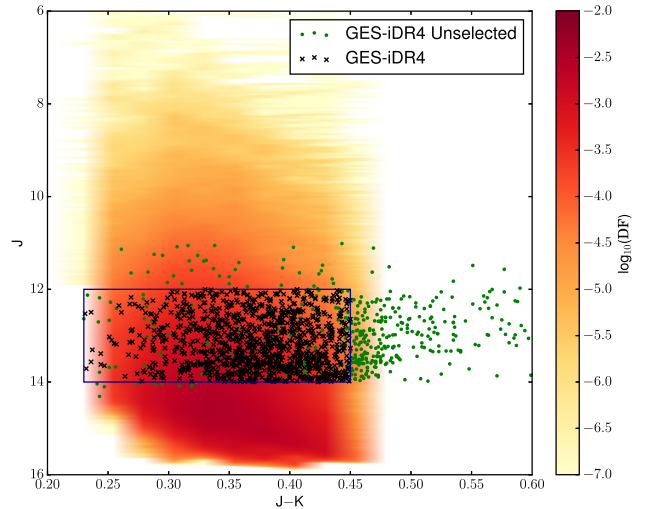


Figure 2. Synthetic CMD of J versus $J-K$ in apparent magnitude space of the simulated solar neighbourhood analogue and of stars from *GES-IDR4*. The red heatmap represents the synthetic stellar populations from *Selene-SYN* which is derived from the simulated galaxy *Selene-CH* as shown in Fig. 1. The black crosses represent stars selected from the *GES-IDR4* data set whereas the stars labelled with green circles representing those removed from *GES-IDR4* by the selection function as described in Section 2.5. The blue rectangle highlights the J and $J-K$ region selection function boundary conditions of $12 < J < 14$ and $0.23 < J-K < 0.45$. Both data sets include the application of surface gravity and effective temperature filters of; $3.5 \leq \log(g) \leq 4.5$ dex and $5400 \leq T_{\text{eff}} \leq 6400$ K. A reddening correction is applied when selecting the observed targets in *GES-IDR4* [$J-K + 0.5 E(B-V)$], which we describe in Section 2.5.

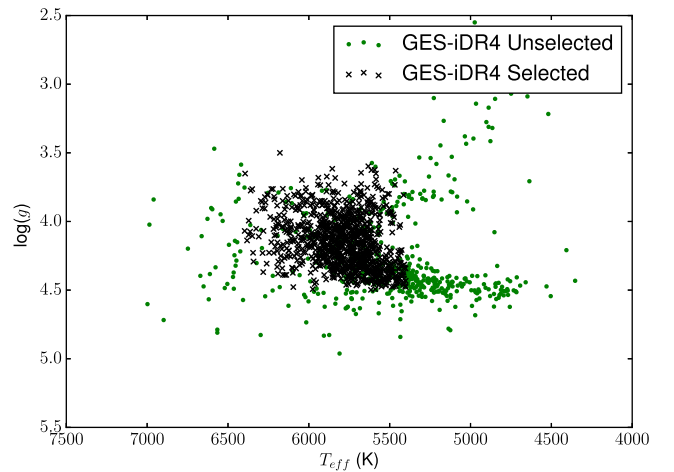


Figure 3. T_{eff} versus $\log(g)$ for the *GES-IDR4* observed sample plotted as black crosses. For reference, we also include the stars which are removed *GES-IDR4* data set due to falling outside the selection criteria as green circles. The selection function is described in section Section 2.5.

in the colour-magnitude plane is shown in Fig. 2. The $T_{\text{eff}}-\log(g)$ diagram of the *GES-IDR4* sample is shown in Fig. 3.

The ages and masses were determined using the Bayesian code BeSPP (Serenelli et al. 2013). We use the grid of input stellar evolution models computed using the GARSTEC code (Weiss & Schlattl 2008); it covers a wide range of masses, $0.6 \leq M_{\odot} \leq 1.4 M_{\odot}$ in steps of $0.01 M_{\odot}$, and metallicities, $-5 \leq [\text{Fe}/\text{H}] \leq +0.5$ dex. The models more metal-poor than -0.6 dex assume α -enhancement of 0.4 dex. Distances were computed using the

2MASS photometry. The ages of the stars are computed from the mode of the posterior PDF. The uncertainties in age were determined as ± 34 per cent around the median value.

2.5 Survey selection function

Before applying the observational selection function to the galaxy simulation, we further post-process the *GES-iDR4* data for the analyses presented here. We require that the observed stars must satisfy the following selection criteria for them to be included in the data set:

- (i) The star is not a member of a star cluster and belongs to the Milky Way field population (using the tag ‘GES_MW’ in the *GES-iDR4* catalogue);
- (ii) The star is not tagged as a member of a special field such as the CoRoT asteroseismic targets or deep fields in the Galactic Centre and anti-centre directions;
- (iii) *J*-band magnitude $12.0 \leq J \leq 14.0$;
- (iv) *J*–*K* colour of $0.23 \leq J-K \leq 0.45 + 0.5 E(B-V)$;
- (v) Heliocentric radial distance of $r \leq 2.0$ kpc;
- (vi) Surface gravity of $3.5 \leq \log(g) \leq 4.5$ dex;
- (vii) Effective temperature of $5400 \leq T_{\text{eff}} \leq 6400$ K.

The colour-magnitude cuts were used to create the input catalogues for the *Gaia*-ESO survey to maximize the number of unevolved FG stars within 2 kpc in the solar neighbourhood.

The T_{eff} and the $\log(g)$ fields of the selection function are chosen because ages of stars with $T_{\text{eff}} \approx 5400$ may not be accurate. Likewise, stellar ages are not well determined for hotter stars with $T_{\text{eff}} \approx 6400$ K or $\log(g) \approx 4.5$ dex, and for more evolved stars on the red giant branch, $\log(g) \approx 3.5$ dex. Our selection would thus include subgiants and main-sequence dwarfs. After applying the selection function as described, we have 1024 stars and use this as our definitive *GES-iDR4* data set.

2.6 Analysing the simulated solar neighbourhood

The aim of this work is to best examine how different ways of processing the same simulation data give variations in the results obtained for the distribution of chemical abundances in the solar neighbourhood analogue. A first-order approach that is commonly used is to simply take a spatial region within a simulation that matches the region of interest in a galaxy and compare that with observational data (e.g. Chiappini, Matteucci & Romano 2001; Martínez-Serrano et al. 2008; Few et al. 2012b, 2014). This approach samples the entire stellar population and requires volume completeness for each type of stars, something that no observational survey does. The fact that simulations are not subject to observational errors is also usually ignored. As a result, the observed and simulated distributions are not directly comparable.

In this work, we compare the *GES-iDR4* results, after the survey selection function is applied, with the following variants of the simulated galaxy *Selene-CH* in order to demonstrate the influence of each component of the process used to mimic observational limits;

- (i) *Selene-CH* is the unaltered and unmodified galaxy. We select all of the stellar population particles that reside within a 2 kpc sphere around the simulated observer, 15 562 in total. These particles are compared directly with the selected *GES-iDR4* results. This kind of direct comparison demonstrates the methodology employed in the ‘traditional sense’, i.e. with spatial cuts alone.

- (ii) *Selene-GES* is a modified version of *Selene-CH*. In this case, we apply stochastic scattering to the ages and abundance ratios of the stellar population particles to emulate observational uncertainties. The magnitude of the scattering is based on the mean errors taken from the *Gaia*-ESO data set.

- (iii) *Selene-SYN* is the result of applying the SYN^{CMD} toolkit, as described in Section 2.3, to the scattered stellar population particles ages and metallicities in *Selene-GES* (i.e. the statistically scattered results of *Selene-CH*). This data set includes the application of selection functions for $\log(g)$, T_{eff} , *J*-band magnitude and *J*–*K* colour and is a more rigorous attempt to mimic the *GES-iDR4* data.

In short, *Selene-CH* represents a first-order analysis of the simulations similar to that found in the majority of the literature, *Selene-GES* shows the effect of applying observational scatter, and *Selene-SYN* demonstrates the influence of selection effects. We now describe the post-processing used to create each of these data sets in detail.

2.6.1 *Selene-CH*: the ‘standard’ simulation approach

We identify a solar neighbourhood analogue within the simulated galaxy *Selene-CH*, a position 8 kpc from the galactic centre on a spiral arm. This position is shown with a cross in Fig. 1 at $(x, y) = (4.0, 6.93)$ kpc; the solar neighbourhood analogue is centred on this point relative to the galactic centre. We have repeated the analysis that follows with stars from different positions on a circle with a galactocentric radius of 8 kpc and find that our results are robust to changes in the position of the simulated observer. This is due to azimuthal homogeneity in the age and chemical abundances; the mean azimuthal variations at 8 kpc from the galactic centre for $[\text{Fe}/\text{H}]$ and $[\text{Mg}/\text{Fe}]$ are 0.02 dex and 0.005 dex, respectively, and the mean age variation is only 0.5 Gyr. This local robustness to changes in position means that attempts to imitate the right ascension and declination distribution of the *GES-iDR4* data would only reduce the number of stellar population particles selected and increase noise in the simulated sample.

Previous works have applied different variations of a *posteriori* re-normalizations (e.g. Pagel & Tautvaisiene 1995; François et al. 2004; Henry et al. 2010) and/or employed GCE models to infer revised sets of stellar yields (François et al. 2004). We have followed this method with our element abundances normalized to the Asplund et al. (2009) solar values and then further shifted by $\Delta[\text{Fe}/\text{H}] = -0.066$ dex and $\Delta[\text{Mg}/\text{Fe}] = 0.078$ dex. This shift in the abundance ratios brings the average abundance of the stars aged between 4.0 and 5.0 Gyr to the solar abundance. This may seem arbitrary, however the amount by which we normalize is small compared to the width of the distribution and we are primarily concerned with the dispersion of the element ratios. Furthermore, it is demonstrated in Few et al. (2014) that variations in abundance ratios (particularly those of α -elements to Fe) are effectively shifted in the same way depending on the IMF. The need to apply such a shift implies that the sub-grid chemical evolution model is not quite correct which is hardly surprising given the uncertainties in the underlying yields and chemical evolution model. Therefore, while the re-normalization of abundance ratios introduces a slight inconsistency to the model, it by no means negates our results.

2.6.2 *Selene-GES*

This data set extends the methodology described above to generate *Selene-CH* by applying a stochastic scattering based on the *GES-iDR4* error bars for age, metallicity and $[\text{Mg}/\text{Fe}]$ abundance ratio, to

mimic the effect of the unavoidable uncertainties found in observations on the precisely known (but not necessary accurate) simulated values.

We degrade the precision of our simulated metallicity and [Mg/Fe] data on a particle-by-particle basis using a Gaussian distribution, centred on the original simulated value with a standard deviation equal to the mean error found in the *GES-iDR4* data set: $\sigma_{[\text{Fe}/\text{H}]} = 0.101$ dex and $\sigma_{[\text{Mg}/\text{Fe}]} = 0.120$ dex. New abundance ratios for each stellar population particle are chosen randomly from this distribution.

The age value for each stellar population particle is also scattered this way except that the distribution from which the new value is chosen at random is not symmetric. The ages of the observed stars in the *GES-iDR4* data set have a mean lower age error of $\sigma_{\text{age}, l} = 3.20$ Gyr and a mean upper age error of $\sigma_{\text{age}, u} = 2.37$ Gyr. We construct a piecewise function from two half-Gaussians with these standard deviations, respectively, to scatter the simulated ages. This process not only broadens the distributions but also makes stellar population particles slightly older.

2.6.3 *Selene-SYN*

Our final version of *Selene* takes the scattered stellar population particles from *Selene-GES* and inputs those particles to *SYN*CMD creating a third data set referred to here as *Selene-SYN*. The mechanics of *SYN*CMD are described in Section 2.3 but the key here is to split the stellar population particles into individual synthetic star particles with a realistic distribution of star properties so that we can apply photometric, $\log(g)$ and T_{eff} cuts to exactly mimic the observed *GES-iDR4* data set. The selection criteria are stated in Section 2.5, however we do not apply the dust extinction correction to the $J-K$ upper limit because our synthetic star particles are unaffected by dust. The synthetic star particles that remain after this are used as our sample of stars analogous to the *GES-iDR4* data set so that we can compare the simulations in a more like-for-like manner. The CMD for *Selene-SYN* is shown in Fig. 2.

As stellar population particles represent different masses of stars, the contribution of each one in terms of synthetic star particles is weighted by the initial mass of the stellar population particle to correctly account for the mass. The initial mass is used because any stars that have evolved and no longer form part of the stellar population are removed from the 10^5 synthetic star particles. The distribution functions shown in this work are described as ‘mass-weighted’, this means that we have weighted the *Selene-CH* and *Selene-GES* particles by their mass to be consistent with the *Selene-SYN* distribution function.

3 RESULTS AND DISCUSSION

We now present and discuss the impact that ‘observing’ our simulations has on the distribution of selected stars in age, [Fe/H] and [Mg/Fe] in comparison with *GES-iDR4* data.

3.1 Ages

We begin the discussion with the analysis of the age distribution in the observed and simulated data sets, however, we remind the reader that age determinations for the observed stars are notoriously difficult, because they rely on the knowledge of surface stellar parameters, metallicities and α -element abundances (Section 2.4). Typically, ages of stars in *GES-iDR4* have an uncertainty of ~ 30 per cent,

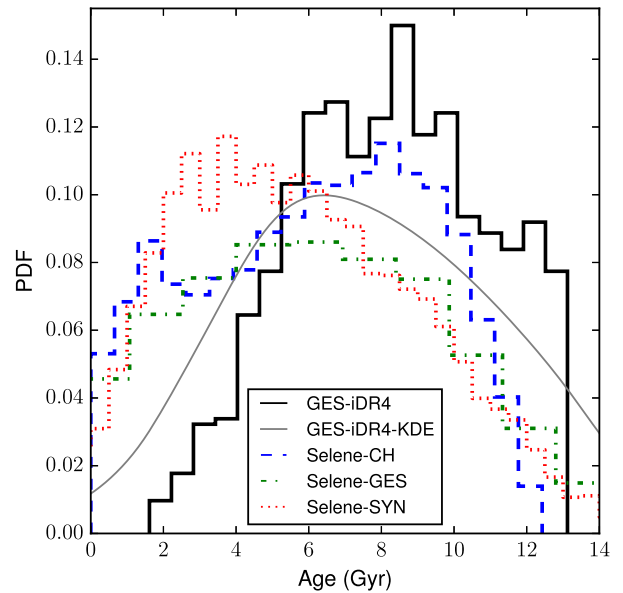


Figure 4. A plot of the present day ($z=0$) age distribution functions for the data sets discussed. The *Gaia*-ESO DR4 data (*GES-iDR4*) is shown as a black line, the grey line is a kernel density estimate of the *GES-iDR4* data set to represent the uncertainties on each point; this has the net effect of shifting the distribution towards younger stellar ages. The simulation data sets *Selene-CH*, *Selene-GES* and *Selene-SYN* are shown as blue dashed, green dot-dashed and red triple-dot-dashed lines respectively.

which is a statistical error and does not include any systematic component. Systematic errors cannot be easily quantified, because of the complex interdependence of different parameters and correlated errors (for example, the error in [Fe/H] is correlated with the error in T_{eff} and in $\log(g)$). Therefore, some mismatch between the observed and model data sets is expected and should not be taken as the evidence of the failure of the galaxy simulations.

The age distributions of our three versions of *Selene* and the *GES-iDR4* stars are shown in Fig. 4. Clearly, there is a systematic difference between the *GES-iDR4* and the simulation data, with an obvious offset to younger stars seen in the simulated data. The application of the stellar age scattering to the simulated data (*Selene-GES*) has the effect of flattening the somewhat truncated older part of the age distribution, removing the peak at 8–10 Gyr and reducing the number of young stars which is entirely expected from the sharp edge of the underlying distribution. Finally, when sampling the CMD of the stellar population particles and applying the *GES-iDR4* selection criteria to the scattered data (*Selene-SYN*) we find that the old end of the age distribution is unaffected, but that the GES photometric filters have the effect of producing a peak between 2 and 5 Gyr and removing many of the stars with ages below 1 Gyr from the distribution. The latter effect brings the young end of the distribution function closer to the observed distribution but there is still a significant discrepancy in both the shape and mean age. To show how the uncertainties associated with the *GES-iDR4* observations alter the shape of the distribution, we have used kernel density estimation to calculate the uncertainty weighted probability distribution function, shown as *GES-iDR4-KDE* in Fig. 4. This distribution is not the correct one with which to compare the processed simulations and we provide it simply for accuracy in representing the observed data.

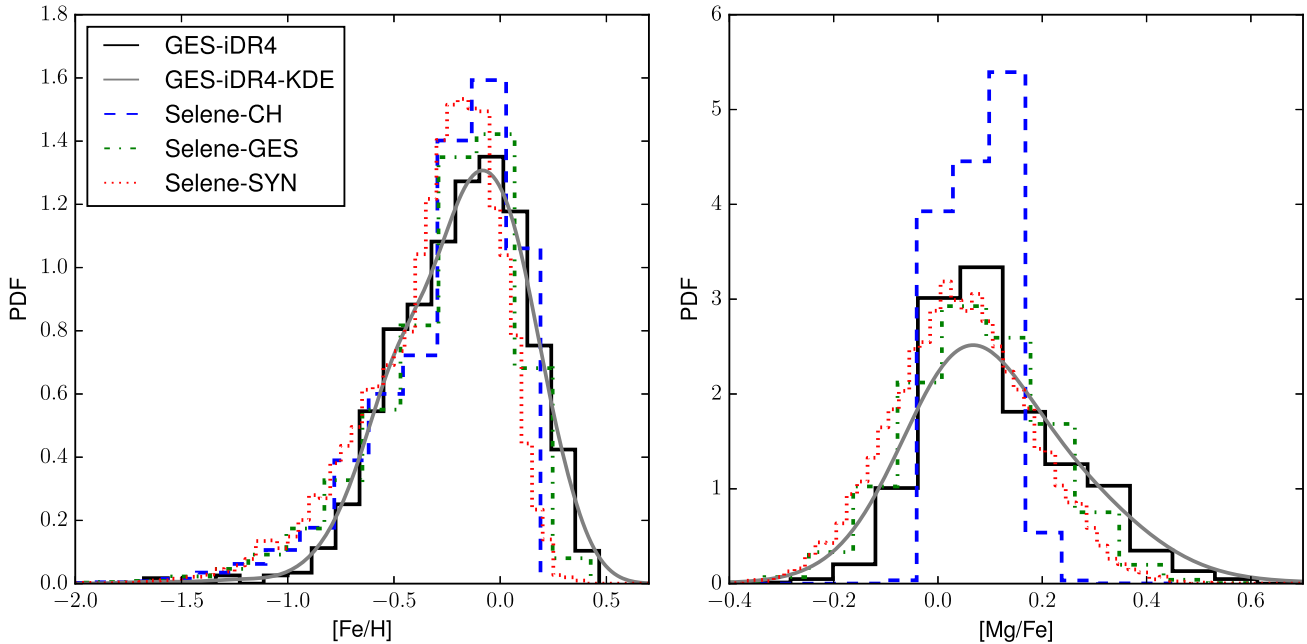


Figure 5. [Fe/H] and [Mg/Fe] distribution functions in the solar neighbourhood analogues of the mass-weighted *Selene-CH* and *Selene-GES* data sets, the synthetic observation data set *Selene-SYN* and the observation data set *GES-iDR4*. *Selene-CH*, *Selene-GES* and *Selene-SYN* represented as blue dashed, green dot-dashed and red triple-dot-dashed lines, respectively. The distribution functions of the *GES-iDR4* stars are shown in black, the grey line is a kernel density estimate of the *GES-iDR4* data set to represent the uncertainties on each point; this has little effect on the [Fe/H] distribution but does diminish the peak in [Mg/Fe]. The bin widths for *Selene-CH*, *Selene-GES* and *Selene-SYN* in [Fe/H] is 0.05 dex, and for *GES-iDR4* this is 0.1 dex. The bin widths for *Selene-CH* and *GES-iDR4* in [Mg/Fe] is 0.05 dex, for *Selene-GES* and *Selene-SYN* is 0.02 dex.

The differences in the age distributions could be caused by several effects. First, this could be due to the differences in the underlying distributions of stellar parameters in the observed and simulated samples. In particular, the combination of the SFH from *Selene-CH* and the SSPs data base used in *SYN*CMD produce a temperature distribution with a ~ 400 K hotter mean T_{eff} value than our chosen *GES-iDR4* sample. This is a very significant difference and is most likely central to understanding of the discrepancy, but currently, we have no suitable framework to explore the effect. If the spectroscopic determinations of T_{eff} are biased, this could explain the deficiency in age for young stars in the *GES-iDR4* sample. Bergemann et al. (2014; Fig. 2) showed that T_{eff} measurements, especially for stars with $T_{\text{eff}} > 6000$ K, appear to be overestimated when compared to the more accurate methodology (infra-red flux method). If this also holds true for *GES-iDR4* data sets then by imposing a T_{eff} cut of 6400 K we actually remove stars, which may have even lower T_{eff} and this pushes the observed distribution towards colder (and older) stars.

One should keep in mind that typically, the mean age of stars is a function of galactocentric radius and *Selene-CH* does form inside-out (see Pilkington et al. 2012). This means that the age distribution of stars would shift to older values with decreasing radii, and thus a more appropriate solar neighbourhood analogue may exist for this galaxy, however given our uncertainty regarding the true distribution we have opted to select our region of interest based on the distance from the galactic centre. Finally, the discrepancy in the age distributions could be due to the differences in the star formation history. Currently, we have no robust constraints on the star formation history of the Milky Way disc over the past 10 Gyr and a detailed analysis of this very complex problem is beyond the scope of this work.

3.2 Distribution functions of [Fe/H] and [Mg/Fe]

Fig. 5 shows the distribution functions of [Fe/H] (left-hand panel) and [Mg/Fe] (right-hand panel) for the simulated data sets compared with *GES-iDR4*. As we have for Fig. 4, we also provide kernel density estimates of the [Fe/H] and [Mg/Fe] *GES-iDR4* distributions in Fig. 5 to show the effect of the uncertainties on these measurements. The uncertainties have a much smaller impact on these abundance ratios than on the age distribution but the peak in [Mg/Fe] is noticeably broadened. Again, we stress that the processed simulation data should be compared with *GES-iDR4* and not *GES-iDR4-KDE* which is simply shown to illustrate the effect of broadening in accordance with uncertainties.

The unaltered simulated stellar population particles (*Selene-CH*, blue line) have a more peaked [Fe/H] distribution compared to the observations (black line), and an extremely narrow distribution in [Mg/Fe]. Furthermore, the mass-weighted *Selene-CH* distribution functions are truncated, at high [Fe/H] and at both low- and high [Mg/Fe]. Table 1 shows the interquartile range (IQR), skewness (σ_3) and kurtosis² (σ_4), which allow us to perform a quantitative analysis of the effect of the observationally motivated changes to the simulated data (see below).

The differences between the width of the *Selene-CH* distributions and the observations do not indicate a failure of the simulation. The simulated [Mg/Fe] distribution is smoothed and broadened significantly when the observationally motivated scattering with errors is applied (*Selene-GES*, green line). This is seen as the increase in the IQR for the [Mg/Fe], and to a lesser extent, in the [Fe/H]

² We use the definition of kurtosis whereby a normal distribution has a kurtosis = 0 (the excess kurtosis).

Table 1. [Fe/H] and [Mg/Fe] distribution function characteristics for the three simulation data sets (*Selene-CH*, *Selene-GES* and *Selene-SYN*), and the observational data set *GES-IDR4* in dex. The interquartile range (IQR), skewness (σ_3) and kurtosis (σ_4) values for the distribution functions of [Fe/H] and [Mg/Fe] as shown in Fig. 5. Columns 1 gives the name of the data set, columns 2, 3 and 4 are the IQR, σ_3 , and σ_4 of the [Fe/H] distribution, respectively. Columns 5, 6 and 7 are IQR, σ_3 and σ_4 , respectively, of the [Mg/Fe] distribution.

Name	IQR	[Fe/H]		IQR	[Mg/Fe]	
		σ_3	σ_4		σ_3	σ_4
<i>Selene-CH</i>	0.398	-1.47	3.52	0.113	-0.282	4.26
<i>Selene-GES</i>	0.409	-1.28	2.98	0.180	-0.0165	0.358
<i>Selene-SYN</i>	0.350	-0.98	1.15	0.160	0.0416	0.099
<i>GES-IDR4</i>	0.414	-0.63	0.95	0.175	1.04	3.64

distributions. The effect is most pronounced for [Mg/Fe], where wings are created in the data on both sides of the distribution. For [Fe/H], the change is only noticeable for higher [Fe/H] values, with the low-metallicity tail being largely unaffected. The key result of Fig. 5 is that the observational uncertainties in age, metallicity and [Mg/Fe] have a much greater effect on the resulting distribution functions than do the photometric selection filters. This means that the observational uncertainties place a fundamental limit on detection of any substructure and on our ability to quantify the slope of any astrophysical relevant relationship in the data.

As expected, scattering the data leads to an increase in the IQR for both the [Fe/H] and [Mg/Fe] distributions providing a good agreement with the observed IQR compared to the original values. When followed up by imposing selection functions we find that the distributions are slightly narrowed but not so much that the reasonable agreement in the spread of the distributions are lost. The simulated [Fe/H] distribution is improved by both stages of our post-processing with the distribution becoming less skewed and reducing in kurtosis to approach the observed values largely due to the enhanced positive tail of the distribution. The [Mg/Fe] distribution is slightly more complicated in that the post-processing does not give a particularly good qualitative fit to the observations in terms of skewness and kurtosis despite the success of reproducing the IQR. As with the [Fe/H] distribution, the scattering and selection effects make the initially negatively skewed distribution more positive but does not go far enough to be in line with the positively skewed, *GES-IDR4*, [Mg/Fe] data. The observed [Mg/Fe] kurtosis indicates a higher likelihood of outliers than found with a normal distribution, the even higher value of the *Selene-CH* distribution is due to the extremely narrow distribution (kurtosis is not a measure of peakedness). The post-processing greatly reduces the kurtosis to be much closer to zero which is entirely expected as the scattering in particular pushes the distribution to be almost normal. The conformity of the *Selene-SYN* distribution to a normal curve is because of the initially narrow distribution and the large scale of the scattering from the [Mg/Fe] uncertainty.

While the width of the observed distribution functions can be reproduced by application of observationally motivated scattering, our simulations do not recover the detailed shape of the [Fe/H] or [Mg/Fe] distribution functions. The shape of the simulated [Fe/H] distribution is promisingly close but still defies similarity with an excess of stars between -0.4 and -0.2 and a deficit between -0.6 and -0.4 ; however the mean values of the [Fe/H] and [Mg/Fe] distributions do both match the observed data. The post-processing

does not give a particularly good qualitative fit to the observations in terms of skewness and kurtosis. As with the [Fe/H] distribution, the scattering and selection effects make the initially negatively skewed distribution more positive but does not go far enough to be in line with the positively skewed, *GES-IDR4*, [Mg/Fe] data.

The mismatch between observed and simulated data for [Mg/Fe] could also hint at the problem with the observations or with stellar yields in our chemical evolution model. In fact, our results confirm the earlier studies (Timmes et al. 1995; François et al. 2004; Andrews et al. 2017) that show that chemical evolution models of the solar neighbourhood systematically underpredict [Mg/Fe] at any metallicity, also the solar values are too low compared to the observed [Mg/Fe] in the solar photosphere. This could be either due to poorly understood stellar yields of SNIa or SNI (see François et al. 2004), or because of the systematic errors in the observed data. It is known that Mg lines in cool stars are affected by NLTE (Bergemann et al. 2015; Merle et al. 2011). In fact, Bergemann et al. (2016, submitted) show that the NLTE [Mg/Fe] trend is lower than LTE trend, that would help to improve the agreement with the simulations.

3.3 Relationships between [Fe/H], [Mg/Fe] and stellar ages in the observed and simulated solar neighbourhood

While the quantitative analysis of 1D distribution functions is precise, it does not aid our understanding of which stars are responsible for the differences between the models and simulations. Greater insight can be provided by examining the evolution of [Fe/H] and [Mg/Fe] with age which are shown in Figs 6 and 7, respectively.

The distribution of the raw simulation data with only a geographical cut (*Selene-CH*) in Fig. 6 is significantly narrower than the observed distribution at all ages. The distribution of stellar population particles is particularly narrow for the oldest stars which are responsible for the low-metallicity tail in Fig. 5. The narrow distributions of the *Selene-CH* stellar population particles in both [Fe/H] (Fig. 6) and [Mg/Fe] (Fig. 7) are a result of the overall trend with age which contrasts with the observed distribution where observational uncertainties dominate. As discussed in the previous section, when the scatter is applied in *Selene-GES*, the initially narrow underlying distribution is no longer discernible in the simulated data and the distribution is instead far more consistent with the observed data albeit with a slight offset to younger ages at a given [Fe/H].

The original, unscattered distribution of the stellar population particles underestimates [Fe/H] compared to the observed distribution for particles older than 8 Gyr (left-hand panel of Fig. 6). This underestimation in the old stars is not so prominent in the *Selene-GES* results (middle panel of Fig. 6), because scattering with observational age errors brings some of the young metal-rich stars to greater apparent ages. Applying the *SYN*CMD tool to produce *Selene-SYN* does not lead to any significant improvement over *Selene-GES* in terms of fitting the observations. While some of the youngest stars are removed (see Fig. 4), it is not sufficient to match the dearth of young stars in *GES-IDR4*.

The trend of [Mg/Fe] with age for the unaltered stellar population particles (left-hand panel of Fig. 7) is a roughly linear increase with a very small up-turn seen in the oldest stars. Again the scattering broadens the distribution significantly (as shown in the middle panel of Fig. 7), but the application of the selection function makes only small changes to the distribution of the simulated stars (right-hand panel of Fig. 7). The main difference between simulation and

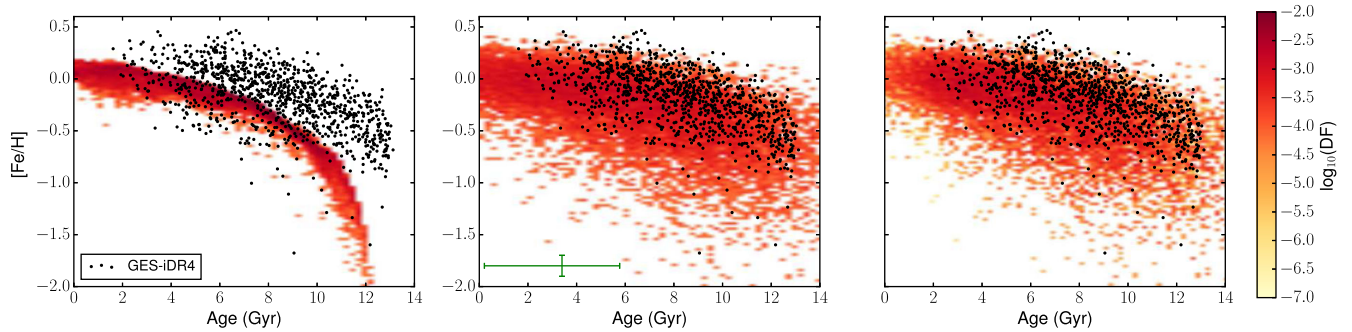


Figure 6. Age-metallicity relation for our various simulated data sets compared with the *GES-iDR4* distribution. The *GES-iDR4* stars are plotted as black points in each panel. *Selene-CH*, *Selene-GES* and *Selene-SYN* are represented as normalized heat maps with increasingly red colours indicating an increase in the abundance of composite or synthetic star particles in bins of 0.025 dex in $[\text{Fe}/\text{H}]$ and 0.2 Gyr in age. we include a representative error bar in green which represents the size of the scatter in $[\text{Fe}/\text{H}]$ and age between *Selene-CH* and *Selene-GES* ($\sigma_{[\text{Fe}/\text{H}]} = 0.101$ dex, $\sigma_{\text{age}, l} = 3.20$ Gyr and $\sigma_{\text{age}, u} = 2.37$ Gyr, values which are computed from the mean of the errors of the *GES-iDR4* data.).

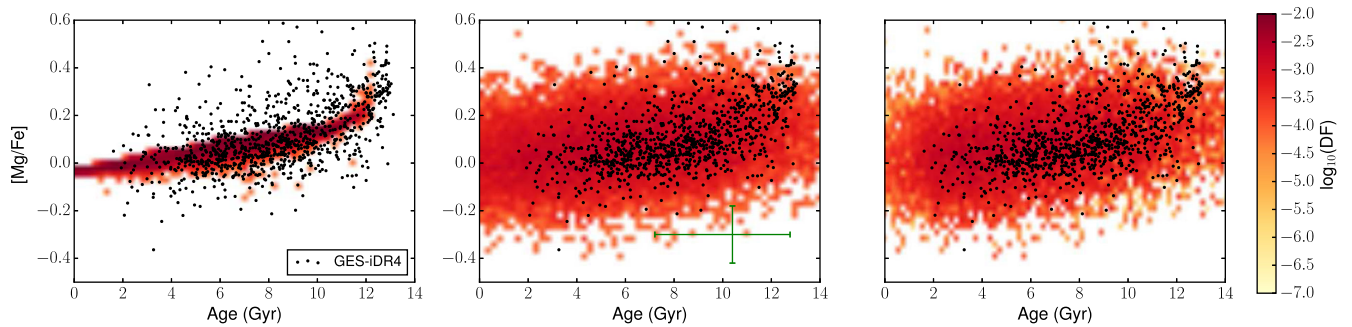


Figure 7. $[\text{Mg}/\text{Fe}]$ versus stellar age. The *GES-iDR4* stars are plotted as black points in each panel. *Selene-CH*, *Selene-GES* and *Selene-SYN* are represented as normalized heat maps with increasingly red colours indicating an increase in the abundance of composite or synthetic star particles in bins of 0.025 dex in $[\text{Fe}/\text{H}]$ and 0.2 Gyr in age. we include a representative error bar in green which represents the size of the scatter in $[\text{Mg}/\text{Fe}]$ and age between *Selene-CH* and *Selene-GES* ($\sigma_{[\text{Mg}/\text{Fe}]} = 0.120$ dex, $\sigma_{\text{age}, l} = 3.20$ Gyr and $\sigma_{\text{age}, u} = 2.37$ Gyr, values which are computed from the mean of the errors of the *GES-iDR4* data.).

observation is in the stars older than 10 Gyr; the up-turn in the underlying *Selene-CH* is not as strong as for the observed old and high- $[\text{Mg}/\text{Fe}]$ stars. As discussed earlier, there are two possible explanations: (a) the neglected NLTE effects in our observed $[\text{Mg}/\text{Fe}]$ distributions and (b) erroneous stellar yields in the chemical evolution model.

Finally, we should note that the GES high-resolution data set does not show any evidence of the bimodal $[\text{Mg}/\text{Fe}]$ distribution with $[\text{Fe}/\text{H}]$, which have been proposed as a chemical separator of the thin and thick discs. This is also consistent with our simulations, which do not have any discontinuity in the SFH at ~ 1 Gyr.

4 CONCLUSIONS

We compare the results of a Milky Way-like galaxy simulation created using *RAMSES-CH* (Few et al. 2014) with the fourth data release of the *Gaia*-ESO survey considering the 1D distribution functions of age, $[\text{Fe}/\text{H}]$ and $[\text{Mg}/\text{Fe}]$ as well as the age evolution of the latter two properties. The comparison is conducted in three stages:

- (i) The simulated stellar population particles are compared directly with the observed distributions.
- (ii) Typical observational uncertainty (from *GES-iDR4*) as the standard deviation of a Gaussian function used to stochastically scatter the simulated data to mimic observational uncertainty.
- (iii) The simulated stellar population particles are stochastically scattered as above and are then split into individual stars based on stellar population models and only those accepted by the *GES-iDR4* selection functions are retained for comparison.

Each of these stages mimics the effects found in observations as a way of placing the simulated data in the ‘observer frame’. The application of stochastic scattering based on the errors of an observational survey has the effect of smoothing out the age-distribution function. The further application of the *GES-iDR4* selection function has the effect of removing young stars (ages < 1 Gyr) and shifting the peak age to between 2 and 5 Gyr. Despite both these effects bringing the simulated age distribution closer to the observed one there is still a significant offset between the distributions. One possible explanation to the remaining discrepancies is that our simulated galaxy has a different assembly history. *Selene* (original simulation from Few et al. 2012b) is not constrained to be identical to the Milky Way and thus determining the location in *Selene-CH* which is the best analogue to the solar neighbourhood is somewhat open to interpretation.

Our key finding is that there is no need for the chemical evolution models to reproduce the large dispersion in metallicity and age seen in observational studies (e.g Bergemann et al. 2014). We show here that this dispersion is apparent and can be fully accounted for by the observational uncertainty. The typical observational uncertainties of our data set are of the order 30 per cent for age and 0.1 dex for $[\text{Fe}/\text{H}]$ and $[\text{Mg}/\text{Fe}]$. When we apply errors to our simulated stars,

we get the age–[Fe/H]–[Mg/Fe] distribution functions that are more consistent with the observed data. In particular, observationally motivated scattering spreads the [Fe/H] distribution towards larger values, which creates an apparent high-metallicity tail. This leads to a larger fraction of metal-rich stars at all ages, improving the agreement with the observed metallicity distribution for the oldest stars.

The scattering and observational selection function improve not only the fit in the age-abundance and age-metallicity plane, but also the fit of the simulated [Fe/H] and [Mg/Fe] distribution functions to the *GES-iDR4* data. Scattering according to observational uncertainties broadens the distribution functions (increased IQR) and swamps their excessive negative skewness, driving them towards a normal curve. The scattering also produces wings on both sides of the initially narrow simulated [Mg/Fe] distribution, making it more consistent with observations. The application of an observational selection function to the simulated data acts to slightly reduce the IQR by culling a number of outliers from the distribution tails. This effect, in particular, removes most metal-rich and metal-poor stars, which are most likely to be the oldest and youngest stars. As a consequence, the kurtosis of both [Fe/H] and [Mg/Fe] distribution functions is reduced. Also our normal scattering function drives the distributions to conform more closely to a normal curve.

One peculiar feature of the observed [Mg/Fe] distributions is that they have a positive skew, which is very difficult to induce in the simulations. It is possible that our assumption of Gaussian errors is wrong. However, at present we do not have a full error matrix for *Gaia*-ESO data set and cannot account for more complex correlations and systematic effects in the observational data.

We also stress that neither in the data, nor in the simulations do we detect the bimodality in the [Mg/Fe] space with metallicity, which has been claimed by some studies (e.g. Fuhrmann 1998) to trace the Galactic thin and thick discs. The assembly history of the simulated galaxy is relatively quiescent with no major mergers after redshift $z = 1.0$, and there are no prominent sub-structures in the simulated [Fe/H]–[Mg/Fe] plane, which could result from an episodic star-formation history (see Few et al. 2014).

Our analysis also reveals that there is more fundamental mismatch between the models and observations: the high-[Mg/Fe] stars generally lack in the original and scattered simulated data sets. This problem most likely has a different cause. First, it has been demonstrated that the NLTE corrections to Mg abundances for the late F- and GK-type stars at lower metallicity, [Fe/H] ~ -0.5 to -1.5 dex, are usually negative (Bergemann et al. 2016; Osorio & Barklem 2016), that is the LTE abundances of Mg are overestimated. Therefore, the high-[Mg/Fe] plateau could be a consequence of the systematics in the observed LTE data set. Secondly, the stellar yields of Mg and/or Fe are not well understood. Standard chemical evolution models severely underproduce the solar Mg abundance (Andrews et al. 2017), but also the shape of the [Mg/Fe] with metallicity does not conform to observations. While this problem is beyond the scope of our paper we would like to encourage future work to explore the causes of this phenomena.

To conclude, it is fundamentally important to reduce the uncertainty of the observed data sets in order to constrain the models of Galaxy formation. The typical observational uncertainty of 0.1 dex in chemical abundances and ~ 30 per cent in ages are too large to provide meaningful information on the substructure in the age-chemical abundance space, which is relevant to the interpretation of the evolution of the Galactic disc. Survey selection functions, like the colour–magnitude selection in the *Gaia*-ESO survey, may

or may not have a sizeable effect on the results, however for the *Gaia*-ESO, this effect is extremely small compared to the effect of observational uncertainties.

ACKNOWLEDGEMENTS

This is based on data products from observations made with ESO Telescopes at the La Silla Paranal Observatory under programme ID 188.B-3002 (the *Gaia*-ESO Public Spectroscopic Survey). We acknowledge the insightful comments and support provided by our colleagues Stefano Pasetto, Daisuke Kawata, Rob Thacker and Dimitris Stamatellos. We would thank the anonymous referee for a very constructive report of the work presented here. BBT acknowledges the support of STFC through its PhD Studentship Programme (ST/F007701/1). We also acknowledge the generous allocation of resources from the Partnership for Advanced Computing in Europe (PRACE) via the DEISA Extreme Computing Initiative (PRACE-3IP Project RI-312763 and PRACE-4IP Project 653838) and STFC’s DiRAC Facility (COSMOS: Galactic Archaeology). CGF acknowledges funding from the European Research Council for the FP7 ERC starting grant project LOCALSTAR and the DiRAC Complexity system, operated by the University of Leicester IT Services, which forms part of the STFC DiRAC HPC Facility (www.dirac.ac.uk). This equipment is funded by BIS National E-Infrastructure capital grant ST/K000373/1 and STFC DiRAC Operations grant ST/K0003259/1. DiRAC is part of the National E-Infrastructure. Continued access to the University of Hull’s High Performance Computing Facility (‘viper’), the HPC facility at the University of Central Lancashire and the computational facilities at Saint Mary’s University are likewise gratefully acknowledged. TB was funded by the project grant ‘The New Milky Way’ from the Knut and Alice Wallenberg Foundation. SGS acknowledges the support by Fundação para a Ciência e Tecnologia (FCT) (ref: UID/FIS/04434/2013 & PTDC/FIS-AST/7073/2014 & Investigador FCT contract of reference IF/00028/2014) through national funds and by FEDER through COMPETE2020 (ref: POCI-01-0145-FEDER-007672 & POCI-01-0145-FEDER-016880). UH acknowledges support from the Swedish National Space Board (SNSB/Rymdstyrelsen). The *Gaia*-ESO Survey 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. MTC acknowledge the financial support from the Spanish Ministerio de Economía y Competitividad, through grant AYA2013-40611-P. UH acknowledges support from the Swedish National Space Board (SNSB/Rymdstyrelsen). This work was supported by Sonderforschungsbereich SFB 881 ‘The Milky Way System’ (subprojects A5, C9) of the German Research Foundation (DFG). This work benefited from discussions at GNASH workshop, Victoria supported by the National Science Foundation under

Grant No. PHY-1430152 (JINA Center for the Evolution of the Elements). ARC is supported by Australian Research Council Grant DP160100637.

REFERENCES

- Agertz O. et al., 2007, *MNRAS*, 380, 963
- Andrews B. H., Weinberg D. H., Schönrich R., Johnson J. A., 2017, *ApJ*, 835, 224
- Asplund M., Grevesse N., Sauval A. J., Scott P., 2009, *ARA&A*, 47, 481
- Bergemann M. et al., 2014, *A&A*, 565, A89
- Bergemann M., Kudritzki R.-P., Gazak Z., Davies B., Plez B., 2015, *ApJ*, 804, 113
- Bergemann M., Collet R., Schoenrich R., Andrae R., Kovalev M., Ruchti G., Hansen C. J., Magic Z., 2016, *ApJ*, preprint ([arXiv:1612.07363](https://arxiv.org/abs/1612.07363))
- Bertelli G., Nasi E., Girardi L., Chiosi C., Zoccali M., Gallart C., 2003, *AJ*, 125, 770
- Bertelli G., Girardi L., Marigo P., Nasi E., 2008, *A&A*, 484, 815
- Bertelli G., Nasi E., Girardi L., Marigo P., 2009, *A&A*, 508, 355
- Calura F., Menci N., 2009, *MNRAS*, 400, 1347
- Calura F. et al., 2012, *MNRAS*, 427, 1401
- Chabrier G., 2003, *ApJ*, 586, L133
- Chiappini C., Matteucci F., Romano D., 2001, *ApJ*, 554, 1044
- Dekker H., D’Odorico S., Kaufer A., Delabre B., Kotzlowski H., 2000, in Iye M., Moorwood A. F., eds, *Proc. SPIE Conf. Ser. Vol. 4008, Optical and IR Telescope Instrumentation and Detectors*. SPIE, Bellingham, p. 534
- Dobbs C. L. et al., 2017, *MNRAS*, 464, 3580
- Eggen O. J., Lynden-Bell D., Sandage A. R., 1962, *ApJ*, 136, 748
- Ferland G. J., Korista K. T., Verner D. A., Ferguson J. W., Kingdon J. B., Verner E. M., 1998, *PASP*, 110, 761
- Few C. G., Courty S., Gibson B. K., Kawata D., Calura F., Teysier R., 2012a, *MNRAS*, 424, L11
- Few C. G., Gibson B. K., Courty S., Michel-Dansac L., Brook C. B., Stinson G. S., 2012b, *A&A*, 547, A63
- Few C. G., Courty S., Gibson B. K., Michel-Dansac L., Calura F., 2014, *MNRAS*, 444, 3845
- Few C. G., Dobbs C., Pettitt A., Konstandin L., 2016, *MNRAS*, 460, 4382
- François P., Matteucci F., Cayrel R., Spite M., Spite F., Chiappini C., 2004, *A&A*, 421, 613
- Freeman K., Bland-Hawthorn J., 2002, *ARA&A*, 40, 487
- Fuhrmann K., 1998, *A&A*, 338, 161
- Gibson B. K., Pilkington K., Brook C. B., Stinson G. S., Bailin J., 2013, *A&A*, 554, A47
- Gill S. P. D., Knebe A., Gibson B. K., 2004, *MNRAS*, 351, 399
- Gilmore G. et al., 2012, *Messenger*, 147, 25
- Gustafsson B., Edvardsson B., Eriksson K., Jørgensen U. G., Nordlund Å., Plez B., 2008, *A&A*, 486, 951
- Haardt F., Madau P., 1996, *ApJ*, 461, 20
- Heiter U. et al., 2015, *Phys. Scr*, 90, 054010
- Henry R. B. C., Kwitter K. B., Jaskot A. E., Balick B., Morrison M. A., Milingo J. B., 2010, *ApJ*, 724, 748
- Hopkins P. F., 2015, *MNRAS*, 450, 53
- Huchra J. P. et al., 2012, *ApJS*, 199, 26
- Joffré P. et al., 2015, *A&A*, 582, A81
- Kawata D., Gibson B. K., 2003, *MNRAS*, 340, 908
- Kennicutt R. C., Jr, 1998, *ApJ*, 498, 541
- Kim J.-h. et al., 2014, *ApJS*, 210, 14
- Knollmann S. R., Knebe A., 2009, *ApJS*, 182, 608
- Kobayashi C., Nakasato N., 2011, *ApJ*, 729, 16
- Kodama T., Arimoto N., 1997, *A&A*, 320, 41
- Komatsu E. et al., 2011, *ApJS*, 192, 18
- Kroupa P., 2001, *MNRAS*, 322, 231
- Kroupa P., Tout C. A., Gilmore G., 1993, *MNRAS*, 262, 545
- Lia C., Portinari L., Carraro G., 2002, *MNRAS*, 330, 821
- Marigo P., Girardi L., 2007, *A&A*, 469, 239
- Martínez-Serrano F. J., Serna A., Domínguez-Tenreiro R., Mollá M., 2008, *MNRAS*, 388, 39
- Merle T., Thévenin F., Pichon B., Bigot L., 2011, *MNRAS*, 418, 863
- Miranda M. S., Macfarlane B. A., Gibson B. K., 2014, *PoS(NIC XIII)*, p. 149
- Miranda M. S. et al., 2016, *A&A*, 587, A10
- Oppenheimer B. D., Davé R., 2008, *MNRAS*, 387, 577
- Osorio Y., Barklem P. S., 2016, *A&A*, 586, A120
- Pagel B. E. J., Tautvaisiene G., 1995, *MNRAS*, 276, 505
- Pasetto S., Chiosi C., Kawata D., 2012, *A&A*, 545, A14
- Pilkington K. et al., 2012, *A&A*, 540, A56
- Price D. J., 2008, *J. Comput. Phys.*, 227, 10040
- Randich S., Gilmore G., Gaia-ESO Consortium, 2013, *Messenger*, 154, 47
- Revaz Y., Arnaudon A., Nichols M., Bonvin V., Jablonka P., 2016, *A&A*, 588, A21
- Rosen A., Bregman J. N., 1995, *ApJ*, 440, 634
- Ruiz-Lara T., Few C. G., Gibson B. K., Pérez I., Florido E., Minchev I., Sánchez-Blázquez P., 2016, *A&A*, 586, A112
- Salpeter E. E., 1955, *ApJ*, 121, 161
- Sánchez-Blázquez P., Courty S., Gibson B. K., Brook C. B., 2009, *MNRAS*, 398, 591
- Scannapieco C., Tissera P. B., White S. D. M., Springel V., 2005, *MNRAS*, 364, 552
- Scannapieco C. et al., 2012, *MNRAS*, 423, 1726
- Schmidt M., 1959, *ApJ*, 129, 243
- Serenelli A. M., Bergemann M., Ruchti G., Casagrande L., 2013, *MNRAS*, 429, 3645
- Skrutskie M. F. et al., 2006, *AJ*, 131, 1163
- Smiljanic R. et al., 2014, *A&A*, 570, A122
- Sommer-Larsen J., Fynbo J. P. U., 2008, *MNRAS*, 385, 3
- Springel V., 2010, *MNRAS*, 401, 791
- Steinmetz M., Müller E., 1995, *MNRAS*, 276, 549
- Stonkutė E. et al., 2016, *MNRAS*, 460, 1131
- Tasker E. J., Brunino R., Mitchell N. L., Michielsen D., Hopton S., Pearce F. R., Bryan G. L., Theuns T., 2008, *MNRAS*, 390, 1267
- Teyssier R., 2002, *A&A*, 385, 337
- Teyssier R., Pontzen A., Dubois Y., Read J. I., 2013, *MNRAS*, 429, 3068
- Timmes F. X., Woosley S. E., Weaver T. A., 1995, *ApJS*, 98, 617
- Tornatore L., Borgani S., Matteucci F., Recchi S., Tozzi P., 2004, *MNRAS*, 349, L19
- Turk M. J., Smith B. D., Oishi J. S., Skory S., Skillman S. W., Abel T., Norman M. L., 2011, *ApJS*, 192, 9
- van den Hoek L. B., Groenewegen M. A. T., 1997, *A&AS*, 123
- Weiss A., Schlattl H., 2008, *Ap&SS*, 316, 99
- Wiersma R. P. C., Schaye J., Theuns T., 2011, *MNRAS*, 415, 353
- Woosley S. E., Weaver T. A., 1995, *ApJS*, 101, 181
- Yates R. M., Henriques B., Thomas P. A., Kauffmann G., Johansson J., White S. D. M., 2013, *MNRAS*, 435, 3500

¹Jeremiah Horrocks Institute, University of Central Lancashire, Preston, PR1 2HE, UK

²E.A. Milne Centre for Astrophysics, University of Hull, Hull, HU6 7RX, UK

³Institute for Computational Astrophysics, Dept of Astronomy & Physics, Saint Mary’s University, Halifax, BH3 3C3 Nova Scotia, Canada

⁴Joint Institute for Nuclear Astrophysics - Center for the Evolution of the Elements (JINA-CEE)

⁵Department of Physics & Astronomy, University of Exeter, Exeter, EX4 4QL, UK

⁶Max Planck Institute for Astronomy, Königstuhl 17, D-69117 Heidelberg, Germany

⁷Institute of Space Sciences, Carrer de Can Magrans, Barcelona, E-08193, Spain

⁸Institute of Astronomy, University of Cambridge, Madingley Road, Cambridge CB3 0HA, UK

⁹INAF - Osservatorio Astrofisico di Arcetri, Largo E. Fermi 5, 50125, Florence, Italy

¹⁰INAF - Padova Observatory, Vicolo dell'Osservatorio 5, I-35122 Padova, Italy

¹¹Instituto de Astrofísica de Andalucía-CSIC, Apdo. 3004, E-18080, Granada, Spain

¹²Lund Observatory, Department of Astronomy and Theoretical Physics, Box 43, SE-221 00 Lund, Sweden

¹³GEPI, Observatoire de Paris, CNRS, Université Paris Diderot, 5 Place Jules Janssen, F-92190 Meudon, France

¹⁴Department of Physics and Astronomy, Uppsala University, Box 516, SE-751 20 Uppsala, Sweden

¹⁵Instituto de Física y Astronomía, Universidad de Valparaíso, Chile

¹⁶European Southern Observatory, Alonso de Cordova 3107 Vitacura, Santiago de Chile, Chile

¹⁷School of Physics & Astronomy, Monash University, Clayton 3800, Australia

¹⁸INAF - Osservatorio Astronomico di Bologna, via Ranzani 1, I-40127, Bologna, Italy

¹⁹INAF - Osservatorio Astrofisico di Catania, via S. Sofia 78, I-95123, Catania, Italy

²⁰Université Côte d'Azur, OCA, CNRS, Laboratoire Lagrange, CS 34229, F-06304 Nice cedex 4, France

²¹Università di Catania, Dipartimento di Fisica e Astronomia, Sezione Astrofisica, via S. Sofia 78, I-95123 Catania, Italy

²²Astrophysics Research Institute, Liverpool John Moores University, 146 Brownlow Hill, Liverpool L3 5RF, UK

²³Université Côte d'Azur, Observatoire de la Côte d'Azur, CNRS, Laboratoire Lagrange, Bd de l'Observatoire, CS 34229, F-06304 Nice Cedex 4, France

²⁴Departamento de Ciencias Físicas, Universidad Andres Bello, Republica 220, Santiago, Chile

²⁵INAF - Osservatorio Astronomico di Palermo, Piazza del Parlamento 1, I-90134, Palermo, Italy

²⁶Instituto de Astrofísica e Ciências do Espaço, Universidade do Porto, CAUP, Rua das Estrelas, P-4150-762 Porto, Portugal

²⁷Institute of Theoretical Physics and Astronomy, Vilnius University, Sauletekio av. 3, 10222 Vilnius, Lithuania

This paper has been typeset from a \TeX/L\AA\TeX file prepared by the author.