StarHorse: A Bayesian tool for determining stellar masses, ages, distances, and extinctions for field stars

A. B. A. Queiroz,1,2⋆ F. Anders,3,2 B. X. Santiago,1,2 C. Chiappini,3,2 M. Steinmetz,3 M. Dal Ponte,1,2 K. G. Stassun,4 L. N. da Costa,2,5 M. A. G. Maia,2,5 T. C. Beers,6 J. Crestani,1,2 J. G. Fernández-Trincado,7,8 D. A. García-Hernández,9 A. Roman-Lopes,10 O. Zamora,9

1 Instituto de Física, Universidade Federal do Rio Grande do Sul, Caixa Postal 15051, Porto Alegre, RS - 91501-970, Brazil
2 Laboratório Interinstitucional de e-Astronomia, - LIneA, Rua Gal. José Cristino 77, Rio de Janeiro, RJ - 20921-400, Brazil
3 Leibniz-Institut für Astrophysik Potsdam (AIP), An der Sternwarte 16, 14482 Potsdam, Germany
4 Vanderbilt University, Department of Physics & Astronomy, VU Station B 1807, Nashville, TN 37235
5 Observatório Nacional, Rua Gal. José Cristino 77, Rio de Janeiro, RJ - 20921-400, Brazil
6 Department of Physics and JINA Center for the Evolution of the Elements, University of Notre Dame, Notre Dame, IN 46556, USA
7 Departamento de Astronomia, Universidade de Concepcion, Casilla 160-C, Concepcion, Chile
8 Institut Utinam, CNRS UMR6213, Univ. Bourgogne Franche-Comté, OSU THETA, Observatoire de Besançon, BP 1615, 25010 Besançon Cedex, France
9 Instituto de Astrofísica de Canarias, 38205 La Laguna, Tenerife, Spain
10 Departamento de Física, Facultad de Ciencias, Universidad de La Serena, Cisternas 1200, La Serena, Chile

Accepted XXX. Received YYY; in original form ZZZ

ABSTRACT

Understanding the formation and evolution of our Galaxy requires accurate distances, ages and chemistry for large populations of field stars. Here we present several updates to our spectro-photometric distance code, that can now also be used to estimate ages, masses, and extinctions for individual stars. Given a set of measured spectro-photometric parameters, we calculate the posterior probability distribution over a given grid of stellar evolutionary models, using flexible Galactic stellar-population priors. The code (called StarHorse) can accommodate different observational datasets, prior options, partially missing data, and the inclusion of parallax information into the estimated probabilities. We validate the code using a variety of simulated stars as well as real stars with parameters determined from asteroseismology, eclipsing binaries, and isochrone fits to star clusters. Our main goal in this validation process is to test the applicability of the code to field stars with known Gaia-like parallaxes. The typical internal precision (obtained from realistic simulations of an APOGEE+Gaia-like sample) are ≈ 8% in distance, ≈ 20% in age, ≈ 6% in mass, and ≈ 0.18 mag in A_V. The median external precision (derived from comparisons with earlier work for real stars) varies with the sample used, but lies in the range of ≈ [0.2,1]% for distances, ≈ [12,31]% for ages, ≈ [4,12]% for masses, and ≈ 0.07 mag for A_V. We provide StarHorse distances and extinctions for the APOGEE DR14, RAVE DR5, GES DR3 and GALAH DR1 catalogues.

Key words: Stars: distances – fundamental parameters – statistics; Galaxy: stellar content

1 INTRODUCTION

Improving the accuracy and precision of stellar distances and ages, as well as individual interstellar extinction measurements, is one of the major tasks of stellar astrophysics in
the Gaia era. Although the parallaxes from the first data release of the Gaia mission (Gaia Collaboration et al. 2016b)\(^1\) provide a major improvement for stars in the solar vicinity (d \(\lesssim\) 200 pc), they do not yet reach the precision of spectro-photometric methods for the much larger distances probed by spectroscopic stellar surveys. Even after the final Gaia data release, foreseen for 2022, spectro-photometry will provide more precise distances for stars beyond 10 kpc.

A large amount of spectroscopic data for individual stars has become available in recent years from dedicated surveys such as the Sloan Extension for Galactic Understanding and Exploration (SEGUE, Yanny et al. 2009), the Apache Point Observatory Galactic Evolution Experiment (APOGEE, Majewski et al. 2017), the RADial Velocity Experiment (RAVE; Steinmetz et al. 2006), the Galactic Archaeology with HERMES survey (GALAH; Martell et al. 2017), the LAMOST Experiment for Galactic Understanding and Exploration (LEGUE, Deng et al. 2012), and the Gaia ESO Survey (GES; Gilmore 2012). The combination of such datasets with broad-band photometric data and the astrometric solutions from Gaia allow for a much more detailed modelling of the chemo-dynamical history of the Milky Way. On the one hand, Gaia’s proper motions and parallaxes, complemented with radial-velocity measurements, enable us to measure stellar phase-space distribution and parallaxes, and/or age estimates (e.g., Brogaard et al. 2016; Yong et al. 2016; Fuhrmann & Chini 2017; Lagarde et al. 2017). The second assumption is that the observational uncertainties of the measured parameters follow a normal distribution. The third assumption is that generalises the method presented in Santiago et al. (2016).

The calculations rely on three important assumptions: Most importantly, we assume that the stellar models are correct, which might not be true for metal-poor stars as well as other limitations in the current stellar models, i.e., that the object of interest follows a canonical single-star evolutionary track. We caution that this assumption, even if a star is apparently single, can be violated to various degrees in practice, leading first and foremost to incorrect stellar mass and/or age estimates (e.g., Brogaard et al. 2016; Yong et al. 2016; Fuhrmann & Chini 2017; Lagarde et al. 2017). We can then calculate the probability that a set of independent measured parameters \(\tilde{x} = (x_1, ..., x_n)\) with associated Gaussian uncertainties \(\sigma_\tilde{x}\) is drawn from a set of theoretical values \(\tilde{x}_0\), predicted by some model \(M\), by writing:

\[
P(\tilde{x}, \sigma_\tilde{x}|\tilde{x}_0) = \prod_i \frac{1}{\sqrt{2\pi} \sigma_{x_i}} \exp \left( -\frac{(x_i - x_{0i})^2}{2\sigma_{x_i}^2} \right).
\]

The above expression is called the likelihood of measuring the set \(\{\tilde{x}, \sigma_\tilde{x}\}\) given a model \(M(\tilde{x}_0|\tilde{\theta})\). Using Bayes’ theorem, we now compute the posterior probability distribution (the probability of the model, given the data) as:

\[
P(\tilde{x}_0|\tilde{x}, \sigma_\tilde{x}) = \frac{P(\tilde{x}, \sigma_\tilde{x}|\tilde{x}_0) P(\tilde{x}_0)}{P(\tilde{x}, \sigma_\tilde{x})}.
\]

The numerator contains the likelihood and the model priors \(P(\tilde{x}_0)\), and the denominator depends only on the measured parameters and their uncertainties. Since we are only interested in inferring the best model parameters for a given set of theoretical models, this denominator merely becomes a constant that can be normalised out.

In our case, the model family \(M(\tilde{\theta})\) consists of a grid of stellar models computed for different ages, metallicities, \(\text{metallicity} (M/H)\), photometric magnitudes, \(m_\text{1}\), and parallax, \(\pi\), to estimate the mass, \(m_\text{s}\), age, \(\tau\), distance, \(d\), and extinction (in V band, \(A_V\)) for individual stars.

1 \url{http://sci.esa.int/gaia/}
and initial masses, convolved with a grid of distances and extinctions (modifying the apparent magnitudes of each stellar model). To evaluate the probability of some specific model quantity, \( \theta \) (usually one that cannot be measured directly), we now compute the marginal posterior probability distribution function (PDF) for this quantity, by integrating over all variables of Equation (2), except \( \theta \):

\[
p(\theta) = \int \cdots \int d\omega_0 P(\omega_\theta | \omega_{\theta}) d\omega_\theta.
\]

As mentioned earlier, a typical set of measured parameters includes \( \omega = \{ [M/H], T_\text{eff}, \log g, m_1, \pi \} \), or any subset of these. The model parameters \( \theta \) we compute are mass, \( m_* \), age, \( \tau \), distance, \( d \), and V-band extinction, \( A_V \). Our code delivers various statistics for the desired quantities. As in Rodrigues et al. (2014) and Santiago et al. (2016), for each quantity we compute the median of the marginalised posterior probability distribution, \( p(\theta) \) (Eq. 3), along with its 5%, 16%, 84% and 95% percentiles.

3 CODE UPDATES

In this section, we explain the technical details of our code in more detail (for an overview see the flow diagram in Fig. 1). Via a parameter file, the user can choose the set of stellar models to be used, the available photometric and spectroscopic data, the treatment of extinction (whether to correct photometry for reddening or whether to include extinction as a parameter to be estimated), and the set of priors, among other options. Once the evolutionary models and the data are read in, the code operates according to the options chosen in the parameter file. These options, along with the other updates since Santiago et al. (2016), are detailed in the next subsections. Readers interested only in the overall performance of the code may skip these.

3.1 Including Parallax as a measured parameter

To adapt our model to the new era of astrometric surveys like Gaia, JASMINE (Gouda 2012), VLBA (Melis et al. 2014), and SKA (Imai et al. 2016), we introduced parallax as an optional measured input parameter for our code. As explained in the previous section, the likelihood can be extended for a generic group of measured parameters, so the method presents no difficulties to introduce the parallax in the likelihood, and it allows for much more precise estimates of stellar masses, ages, and extinction. When the user decides to use parallaxes as the primary input, we fix the range of distances for all models to be consistent with that measurement within 3σ (see §3.2.1 below). If this is not specified, the possible range of distances to be probed for each stellar model is derived by matching an observed apparent magnitude, \( m_\lambda \) (within ±3σ\( m_\lambda \)), in some filter to the corresponding model absolute magnitude.

3.2 Stellar parameters posterior

Currently our code can determine distances, ages, masses, and extinctions, given a set of measured parameters by marginalising the joint posterior PDF. Below we explain in more detail how we build the values of distance, extinction, ages, and masses covered by the PDF.

3.2.1 Distance

If no reliable parallaxes are available, or if they are only available for a subset of stars, the range of distances to be probed comes from the available measured apparent magnitudes. We choose a master filter, \( m_{\lambda,0} \), and create an array of length \( N_d \) that ranges from \( m_{\lambda,0} \pm 3\sigma_{m_{\lambda,0}} \). For each value of this array we compute the distance modulus, \( (m_{\lambda,0} - M_\lambda) \), for the absolute magnitudes in the model grid; these values are then finally transformed into an array of possible distances, \( d \). Evidently, this process assumes that the apparent magnitudes have previously been corrected for extinction. As explained in §3.1, if the user decides to use parallax measurements as the primary input, we build the distance array by inverting the array of allowed parallaxes. We then transform \( d \) to intrinsic distance moduli that do not depend on colour or extinction. The \( d \) and \( (m-M) \) values are then used in the priors and likelihood to build the posterior PDF.

3.2.2 Extinction

When multi-band photometry over a sufficient wavelength range is available, one can use the measured colours of a star to estimate interstellar dust extinction. When the intrinsic magnitudes are constrained by spectroscopic measurements, this extinction measurement can become very precise (e.g., Rodrigues et al. 2014). Our code can now also be used to determine extinction towards stars, by adding another free dimension to the model space.

When parallaxes are not available, and we build a distance moduli that comes from the apparent magnitude \( (m_\lambda - M_\lambda) \), as explained in §3.2.1, the distance moduli are corrected by an a priori unknown extinction: \( (m_{\lambda,0} - M_\lambda) = (m_\lambda - M_\lambda) - A_\lambda \). For each stellar model and each possible distance modulus, \( (m_\lambda - M_\lambda) \), we thus create \( N_{A_\lambda} \) random \( A_\lambda \) values from a previously defined range of possible \( A_\lambda \). If there is no initial guess of the \( A_\lambda \) for the given star, this range of \( A_\lambda \) values is kept fixed as \([-0.1, 3.0]\). If some expectation for \( A_\lambda \) is available (an \( A_\lambda \) prior, \( A_\lambda,\text{prior} \)), we probe extinction in the range \([A_{\lambda,\text{prior}}/3, 3 \cdot A_{\lambda,\text{prior}}]\). We then transform the \( A_\lambda \) to \( A_\lambda \) values using a chosen extinction law (Cardelli et al. 1989 by default), and subtract it from \((m_\lambda - M_\lambda)\). Since the model space is large, we usually use \( N_{A_\lambda} = 3 \) to lower the computational cost. As long as the spectroscopic measurements do not confine the solution to a very small volume in model space, the marginalised PDFs over extinction and distance remain well-sampled.

When parallax information is available, the dereddened distance modulus array, \((m-M)_{0}\), is determined directly from the parallax. To determine the extinction we then use the reddened distance modulus arrays built from the apparent magnitudes, \((m_{\lambda} - M_{\lambda})\), and the difference between those two naturally delivers \( A_\lambda \).

3.2.3 Masses and Ages

Because masses and ages are quantities provided by the grid of evolutionary models, they are simply repeated over the
additional dimensions of distance and extinction. Therefore, once we have a PDF from equation (3), we can directly estimate these parameters by marginalising over the distance and extinction dimensions.

### 3.3 New spatial, chemical, and age priors

Our code uses several priors that summarise our prior knowledge about the initial mass function and the stellar population structure of the Milky Way, including the thin and the thick disk, and the stellar halo. For each Galactic component, these priors include the spatial density distribution, a metallicity distribution function (MDF) and an age distribution. In Santiago et al. (2016), we adopted the same priors as Binney et al. (2014). In the current version, we updated our default structural parameters (scale lengths, scale heights, normalisations) from Binney et al. (2014) to Bland-Hawthorn & Gerhard (2016).

Since APOGEE and other surveys are now probing also the inner regions of the Milky Way, we have also added simple spatial priors for the bulge/bar. The simplest choice of bulge spatial prior is a spherical exponential model with an exponential e-folding length of $0.5 - 1.0$ kpc. We also added the oblate model described by Dehnen & Binney (1998). Finally, as our default model we included a bar-like bulge model from Robin et al. (2012). Their models assume the bulge to be a triaxial ellipsoid, either boxy or disky (or yet a combination of both depending on the plane of projection), and with density laws that can be a sech$^2$, an exponential, or a Gaussian. Our code has been tested with one- and two-component model priors. Our default model is the S ellipsoid (bar component) taken from the “S+E” case listed in Table 2 of Robin et al. (2012). This “S+E” model was the minimum likelihood one among the models presented by the authors. The E component was removed based on the revision of the thick-disk structure and its extrapolation towards the inner Galactic regions made by Robin et al. (2014), which effectively rendered the classical ellipsoid bulge unnecessary. We refer the interested reader to Robin et al. (2012) and Robin et al. (2014) for more details.

Fig. 2 shows the contributions of each Galactic component spatial distribution as a function of distance, for four representative directions. In the upper left panel we show a direction towards the inner Galaxy $(l, b) = (20, 8)$; in this case the bar/bulge component is the dominant population at heliocentric distances of ~ 6 kpc, its density decreases rapidly toward greater distances. The upper right panel shows a direction toward the same Galactic longitude, but at higher latitude. In this case the we miss the bulge, and notice that each of the other components dominates at a certain distance range. The lower panels shows directions away from the Galactic centre. In the lower left panel $(l, b) = (90, 30)$, we see that the contribution for the disks dominates out to 3.5 kpc, in the lower right panel $(l, b) = (150, 60)$ the halo dominates already for $d > 2$ kpc.

Our new age and metallicity priors for the four Galactic components are all assumed to be Gaussians. The corresponding mean and standard deviation values for each case are provided in Table 1. The motivation for this change is twofold: 1 - simplicity: they are simple functions, easily computed, which makes them ideal for the computationally intensive parameter estimate process used here; 2 - they are made broad enough to accommodate most or all of the recent age and metallicity distributions found in the literature, which are not only diverse, but also often conflicting. By taking this approach, we avoid making our priors too specific, but also avoid completely overlooking the knowledge accumulated about the different Galactic components.

We note that the previous age priors from Santiago et al. (2016) assigned zero probability to disk stars older than 10 Gyr and to thick disk or halo stars younger than this value. Recent results found in the literature pose a challenge to such simple age step functions. One example is the discovery of young α-enhanced stars, likely thick-disk members (Chippin et al. 2015). The previous metallicity priors were also narrower, specially for the thin-disk, essentially ruling out any thin disk star more metal-poor than $[Fe/H] = -0.6$. The changes made to these components also make them more in sync with our current understanding of the bulge populations, for which there is also recent evidence for a larger fraction of stars younger than $\sim 5$ Gyr (Bensby et al. 2013; Valle et al. 2015).

### 4 TESTS WITH SIMULATED STARS

To validate our code, we carried out several tests using simulated stars. Since one of our main goals is to apply StarHorse to field stars with reliable parallaxes to infer masses and ages, we include parallax in the set of observed quantities of our simulated stars, and use them to constrain our distance range.

The first sample of simulated stars we used is identical with the set used in Santiago et al. (2016), except for the distance range and the assumed spectroscopic uncertainties. It consists of 5000 randomly drawn PARSEC models (Bressan et al. 2012), convolved with Gaussian errors in the photometric parameters and parallaxes (see details in Table 3). The error values for the “high-res” version of this simulation were inspired by the spectroscopic uncertainties from the APOGEE DR14 results, while the low-res case was based on typical uncertainties from the SEGUE and RAVE surveys. In both cases the samples contain stars with distances between 0.05 and 1 kpc (typical $Gaia$-TGAS distances), random galactic positons, and extinctions from Schlegel et al. (1998).

The PARSEC simulated stars are useful to map the internal accuracy and precision of our estimated distances, ages, masses, and extinctions over a wide input parameter range. The caveat is that the PARSEC sample sets are not representative of a real magnitude-limited sample of stars in our Galaxy. By randomly picking PARSEC models, we tend to oversample young stars, regardless of metallicity, relative to most survey data in the nearby Galaxy. This also means

\begin{table}
\centering
\begin{tabular}{|c|c|c|c|c|}
\hline
Component & Mean age & $\sigma$ age & Mean [M/H] & $\sigma$[M/H] \\
\hline
Thin disk & 5.0 Gyr & 4.0 Gyr & -0.2 & 0.3 dex \\
Thick disk & 10.5 Gyr & 2.0 Gyr & -0.6 & 0.5 dex \\
Halo & 12.5 Gyr & 1.0 Gyr & -1.6 & 0.5 dex \\
Bulge & 10.0 Gyr & 3.0 Gyr & 0.0 & 0.5 dex \\
\hline
\end{tabular}
\caption{Adopted parameters of the Gaussian age and metallicity priors for the Galactic components.}
\end{table}
that our prior knowledge about Galactic stellar populations does not apply to these simulations, so that in this case we set all priors to unity.

To test the code in more realistic scenario, we also use a TRILEGAL (Girardi et al. 2012) population-synthesis simulation of an APOGEE-TGAS-like sample of giant stars. The details of this simulation are given in Table 2; we describe the main features briefly here. The underlying stellar models of TRILEGAL are from Marigo et al. (2008), which are similar but not identical to our default PARSEC 1.2S models.

We used a Chabrier (2003) log-normal initial mass function (IMF) for all Galactic components (thin disk, thick disk, bulge, and stellar halo), and the default spatial distribution, density normalization, star-formation rate (SFR), and age-metallicity relation (AMR) for all components (see Table 2). Extinction was assumed to result from an exponential dust disk with calibration at infinity of $A_V = 0.0378$ mag for the Galactic poles, and the photometry is in the $UBVRIJHK$ system (Maíz Apellániz 2006; Bessell 1990). The Solar position and Solar height above the disk were assumed to be...
Table 2. Adopted geometry, local or central calibration and SFR+AMR for the Galactic components simulated with TRILEGAL

| Component   | Spatial distribution | Local/Central Calibration | SFR+AMR                      |
|-------------|----------------------|---------------------------|------------------------------|
| Thin disk   | Squared hyperbolic secant | Local 55.40 $M_{\odot}$pc$^{-2}$ | 2-step age + Fuhrmann (1998) + α-enh Girardi et al. (2005) |
| Thick disk  | Squared hyperbolic secant | Local 0.001 $M_{\odot}$pc$^{-3}$ | 11-12 Gyr const. + $Z = 0.008$ with $\delta(M/H) = 0.1$ dex |
| Halo        | Power law            | Local 0.0001 $M_{\odot}$pc$^{-3}$ | 12-13 Gyr + Ryan et al. (1996) $[M/H]$ distribution |
| Bulge       | Triaxial bulge       | Central 406. $M_{\odot}$pc$^{-3}$ | 10 Gyr + Zoccali et al. (2003) $[M/H] + 0.3$ dex |

4.1 Internal accuracy and precision

Figure 3 shows the results of our PARSEC simulated-star tests (high-res case). The first two rows show the relative errors in distance, $(d_{SH} - d_{true})/d_{true}$, where $d_{SH}$ are the distances estimated by our code (SH standing for StarHorse). Each panel shows these same errors as a function of a different parameter. The last panels on the right show the relative distance errors (top row) and uncertainties (2nd row) mapped onto the log $g$ vs. $T_{eff}$ diagram. The first line of Table 4 shows the relative distance error values that correspond to the 5% -ile, 16% -ile, 50% -ile, 84% -ile, and 95% -ile positions of the relative error distribution, when their signs are omitted. We also list the median value of the full error distribution, to quantify the presence of systematic trends. For example, we see that 50% of the simulated stars have distance errors of less than 6%, and 84% have distance errors below 16%. There is no strong systematic trend with any of the parameters, apart from an increase in the errors for larger distances and for lower log $g$ values (giants). We also note that the discreteness of the PARSEC model grid used is visible in most of the panels.

The remaining rows of Fig. 3 show the same type of plots, but now with the relative errors in age, mass, and $A_V$. As before, the percentiles of the relative error distributions are listed in Table 4. As in the case of distances, the mass estimates (rows 5 and 6 in Figure 3) do not suffer from any clear systematics with the main parameters, apart from the trend of increasing errors with distance. There is a subset of mostly subgiant and dwarf stars ($log_g > 3.5$, and $m_{ast} < 0.8m_0$) with very-well determined masses. From Table 4, we see that 50% (84%) of the estimated masses agree with the true values within 7 (22)%., and that the outliers are predominantly young (massive) evolved stars, which are rare in the Milky Way field.

As for ages, shown in the 3rd and 4th rows of Figure 3, catastrophic errors, of 100% or more, occur for about 15% of the stars. These stars are not restricted to a small subset of parameter space. But age is the main parameter leading to these catastrophic errors, which are more frequent for $t < 1$ Gyr. There is also some dependence on distance (i.e., low parallaxes, for which the parallax error is relatively large) and mass (or log $g$). Since dwarfs change their position in the spectroscopic Hertzsprung-Russell diagram only slightly on long timescales, ages for dwarfs are more uncertain; the age PDF will tend to be very flat. The $A_V$ errors shown in
the last two rows are within 0.05 (0.1) mag for 50% (84%) of the stars, with no clear systematic effects.

To test the influence of our prior knowledge, the TRILEGAL results shown in Fig. 4 (all priors included) can be compared to Fig. A2 in Appendix A. That figure shows the corresponding results for the case when no priors were adopted. In general, the results in the all priors and in the no priors cases are quite similar, indicating that the accuracy of our code is robust to prior assumptions. We also tested the case where only the spatial priors were used, but not the age and MDF ones. As expected for this intermediate case, the panels are again very similar to those shown in Figures 4 and A2.

This is further corroborated by Figure 5, where we show the relative error distributions of age, distance, mass, and $A_V$ for the three combinations of priors, and separated for subgiants/hot dwarfs and giants. The histograms confirm the trends seen in the scatter plots, and they also show that the parameter estimates are not strongly dependent on the priors adopted, at least out to our maximum distance of 1 kpc (TGAS volume). We note that the spatial density profiles, SFH and MDF used by TRILEGAL are not the same as those in StarHorse, which shows the importance of adding basic and non-restrictive priors in the parameter inference.

### 4.2 Effect of systematic stellar parameter errors

As in Santiago et al. (2016), we also tested the effect of systematic offsets in the observed quantities on our results, using the PARSEC samples. The results are shown in Figure 6. Again we split the simulated samples into giants and dwarfs. Each column corresponds to a given parameter for which shifts were applied, keeping the other parameters at their observed (but systematics-free) value. Each panel shows the mean, the median, and the dispersion around the mean values of the relative error (over all stars), as a function of the shift parameter.

In almost all panels, the difference between the mean and median relative errors is very small, attesting to the existence of relatively few outliers. The dispersion around the mean is rarely larger than 20% (or 0.2 mag in the case of extinction) in most cases studied. The exception is the relative age error, for which the mean is much farther from zero than the median, and the dispersion is of order ±10% in the case of giants, and even larger for dwarfs (see also discussion above). On average, however, the effect of systematic errors on our estimated parameters is typically less than ±10%. In the following we discuss some more conspicuous effects.

Systematic errors in $T_{\text{eff}}$ affect almost all inferred pa-

---

**Table 3. Summary of the reference data: parameter ranges, uncertainties, and provenance.**

| Sample               | $\sigma(\mu)$ | $d$ range [kpc] | $\sigma(T_{\text{eff}})$ [K] | $\sigma$ log g range | $\sigma$ [M/H] range | $\sigma$ mag range [V mag] |
|----------------------|---------------|-----------------|-------------------------------|----------------------|------------------------|-----------------------------|
| PARSEC high-res      | 0.3           | 0.05 – 1        | 70.0                          | 3000 – 7000          | 0.08                   | 1 – 5                      | 9.05                         |
| PARSEC low-res       | 0.3           | 0.05 – 1        | 95.0                          | 3000 – 7000          | 0.24                   | 1 – 5                      | 9.05                         |
| TRILEGAL             | 0.3           | 0.05 – 1        | 70.0                          | 3000 – 7000          | 0.08                   | 1 – 4.1                    | 9.17                         |
| Detached Eclipsing Binaries | 0.04       | 0.01 – 0.65     | 80.0                          | 4320 – 5730          | 0.02                   | 1 – 3.6                    | 9.17                         |
| Other Eclipsing Binaries | 0.36       | 0.03 – 1        | 2560.0                        | 3880 – 30300         | 0.02                   | 2.9 – 4.5                  | 9.17                         |
| CoRoGEE stars        | 0.007         | 0.8 – 10        | 90.0                          | 4000 – 5500          | 0.05                   | 1.4 – 3.0                  | 9.30                         |
| OCCASO clusters stars| 0.06          | 0.05 – 6        | 60.0                          | 4300 – 5300          | 0.1                    | 1.7 – 3.2                  | 9.00                         |
Figure 3. StarHorse relative errors for the PARSEC high-res simulations. For each parameter, the six panels on the left show the relative errors, $d$ (top row), $\tau$ (third and fourth row), $m$ (fifth and sixth row), $A_{V}$ (last rows) as a function of the true parameters. The solid blue line is the identity line, and the dashed red lines correspond to ±20% errors (except for ages: 40%). The panels on the right show the relative errors (top panel) and uncertainties (bottom panel) in the $\log g$ vs. $T_{\text{eff}}$ plane.
Figure 4. Relative distance, age, mass, and V-band extinction errors for the TRILEGAL simulations, in the same style as Fig. 3.
rameters for both dwarfs and giants. The effect of (under)overestimating temperatures on dwarfs is perhaps simpler to interpret, as it leads to best matching models of (lower)higher masses, therefore (less)more luminous. The apparent distance modulus is correspondingly biased too (low)high, yielding either (smaller)larger inferred distances or $A_V$, or a mixture of both. For the giants, evolutionary timescales are shorter, making age a central parameter. In their case, an (under)overestimated $T_{\text{eff}}$ requires a (older)younger (and therefore (less)more massive and luminous) progenitor, and leads to concordance models of higher apparent distance moduli.

The systematic effects of metallicity biases on the inferred parameters are of lower amplitude as compared to $T_{\text{eff}}$. This is consistent with $T_{\text{eff}}$ having a tighter correlation with the photometric parameters, and hence more strongly affecting the likelihood functions. Estimated masses are affected when the metallicity is biased. An (under)overestimated metallicity leads to a better match of a given star to models of (higher)lower metallicity, which will be of (higher)lower mass for a fixed luminosity (i.e., fixed apparent magnitudes and distance). For giants, the higher(lower) masses will again require younger(older) progenitors.

The case of $\log g$ is such that it affects distances more strongly. This has been investigated before by Santiago et al. (2016), with similar results. An (under)overestimated $\log g$ leads to an (over)underestimated in the distances, since the data for a star become more consistent with models of stars (more)less luminous than it actually is. Parallaxes (last column of Figure 6) also predominantly affect distances, in the expected sense. A slight effect on masses and ages of giants can also be seen, since an (under)overestimated parallax will require (more)less luminous giants, therefore shifting the models towards (younger)older ages with (higher)lower mass progenitors.

### 5 EXTERNAL VALIDATION

Up to this point, we have shown StarHorse results for simulated stars, for which we previously know all their stellar parameters. Although TRILEGAL delivers realistic star counts simulations for the Galaxy, real data can present different behaviour from the assumptions we made on the simulations. In this section we test StarHorse results for observed data. We choose samples of stars from eclipsing binaries, asteroseismology, and open clusters. Another important difference with respect to the validation carried out in the previous section is that here we will compare our estimated distances, ages, masses, and extinction values out to distances larger than 1 kpc.

#### 5.1 Eclipsing Binaries

In this section we show StarHorse validations with samples of Eclipsing Binaries (EBs). EBs can give precise stellar masses and radii, which in turn yield precise surface gravities. If temperature is also available, these stars can provide a good benchmark for estimating distances, ages, masses, and extinctions. Here we show tests on two samples: one is made up of detached binary systems with individual values of stellar parameters, the other has EBs that are photometrically unresolved.

##### 5.1.1 Detached Eclipsing Binaries

We use the sample of EBs from Ghezzi & Johnson (2015) as a first comparison sample to our StarHorse estimates. Those authors carried out a literature search for detached binary systems with at least one evolved star to be used as a benchmark for the determination of masses. This sample contains a total of 26 binaries with photometry, atmospheric parameters, parallaxes, ages, masses, and extinctions. The majority of the sample is composed of stars from the Large and Small Magellanic Clouds (LMC, SMC), but some stars are near the Solar vicinity ($d < 1$ kpc). As we see in Table 3, this sample contains mainly giant and subgiant stars ($\log g < 3.6$) with low metallicity. To run StarHorse we used as input the photometry, $T_{\text{eff}}$, [Fe/H], and $\pi$ given in Table 1 of Ghezzi & Johnson (2015), and $\log g$ given in their Table 2.

StarHorse recovered distances, ages, masses and extinctions for 24 out of 26 stars. Fig. 7 compares these results to the more fundamental determinations of Ghezzi & Johnson (2015). For the distances, shown in the upper left panel of Fig. 7, the reference distance is the inverse of the Hipparcos input parallaxes used by those authors. Because the sample includes stars either in the Solar Neighbourhood, ($d < 1$ kpc), as well as stars at LMC/SMC distances, we show these latter as a separate inset in the figure. The agreement is excellent, with only a small degradation for the larger distances to LMC and SMC stars, whose parallax uncertainties are larger (see the panel inset). The mean uncertainty on the parallax of this sample is lower than the simulated samples; see Table 3. This is probably the reason for such a good agreement in distances. In the same figure, we show the comparison between StarHorse and PARAM da Silva et al. (2006) ages (which incorporate results from asteroseismology using a method similar to StarHorse; upper right panel), and the comparison between StarHorse masses and those taken from asteroseismology scaling relations (lower left panel; for the references on the masses, see Table 2 Ghezzi & Johnson (2015)). We also show the comparison between StarHorse $A_V$ and $3.1E(B-V)$ (lower right panel), where the references for $E(B-V)$ are in Table 1 of Ghezzi & Johnson (2015).

From Figure 7, we see that StarHorse yields ages that are systematically larger than those from PARAM. The median age offset is 22%. Still, most of the StarHorse ages appear to agree with PARAM ages to within 50%. The two age determinations are also consistent with each other for most stars, considering both error bars. Table 4 shows that 50% (84%) of the estimated ages have errors below 25% (76%). As for the masses, the agreement between our estimates and those from asteroseismology are somewhat better. StarHorse masses tend to be smaller by 12%. The error bars and discrepancies relative to PARAM values are relatively smaller as well, with most of stars having errors of 17% or smaller. The extinction estimates also agree well with the ones from the literature, despite the large error bars. Most of the stars have an $A_V$ error below 0.18 mag, with a moderate systematic effect ($-0.07$ mag).
5.1.2 Other Eclipsing Binaries

As a second comparison with EBs, we use the sample from Stassun & Torres (2016), which contains 156 systems. Their sample is composed of stars with precise stellar radii and effective temperatures. Most of the stars have also available masses, Gaia parallaxes, metallicities, and magnitudes in at least one of the following filter systems: Tycho (Høg et al. 2000), APASS (Henden & Munari 2014), Strömgren (Casagrande et al. 2014) and 2MASS (Cutri et al. 2003). Their distances and extinctions were estimated by performing fits to the broad-band photometric spectral energy distributions of the binary systems. The range of the parameters is described in Table 3. Although most of the parameters are individual for each star, the magnitudes are systemic; these magnitudes describe the binary system, not each star individually. This can include a bias in our likelihood, and as most of the sample is made by systems with similar masses, this will affect both primary and secondary stars, though the effect would be grater on the secondary stars. It is very important to proceed with this test, since approximately 50% of Solar-type field stars are binary or multiple systems (Raghavan et al. 2010), (Moe & Di Stefano 2017), and most of them are photometrically unresolved.

To run StarHorse we only select stars with available masses and limit the distance to 1 kpc. Figure 8 compares the StarHorse results for these stars with the ones of Stassun & Torres (2016). The StarHorse distances are in general underestimated in relation to the ones estimated by Stassun & Torres (2016). Most of our distance estimates are smaller by 20% (see Table 4), the effect being larger for the secondary members. This is probably a direct result of the systemic magnitudes we are using, since each star is assigned a brighter magnitude than it actually is, the amplitude of the effect being larger for secondaries. Therefore, our code will tend to match it with models at nearer distances. The masses estimated by StarHorse are in reasonable agreement with the reference sample, as shown in Table 4: 50% of the stars present relative errors smaller then 10%. We should keep in mind that the surface gravities for this EBs sample are based on the quoted radii and masses. The estimated extinction values do not present systematic deviation with the extinction given by Stassun & Torres (2016), but as the extinction strongly depends on the distance models, they also should be affected by the systemic magnitudes.

5.2 Asteroseismology: CoRoGEE

We also use the CoRoT-APOGEE sample (CoRoGEE; Anders et al. 2017) to evaluate the accuracy of our StarHorse results. CoRoGEE contains seismic measurements (from CoRoT) combined with high resolution spectra (from APOGEE) for more then 600 stars, this sample has also estimates of distance, age, mass, and Av from the PARAM code. They cover a wide range of Galactocentric distances, metallicities, and ages (see Table 3), but are all red giants stars. To run StarHorse, we use as input the atmo-
Figure 6. Relative errors per parameter (First two rows: \(d\), third and fourth rows: \(\tau\), fifth and sixth rows: \(m_a\), seventh and eighth rows: \(A_V\)) obtained when different systematic offsets are introduced to the stellar parameter at a time. For this test we use the simulated PARSEC sample with high-resolution errors, but only 500 stars. The panels show the results for giant stars with \(\log g < 4\), and dwarf stars with \(\log g > 4\) separately. From the left, the first column shows the results for offsets to \(T_{\text{eff}}\), the second column shows the offsets to \([M/H]\), the third column shows the offsets to \(\log g\), and the last column shows the offsets to parallax \(\pi\). In each panel, the black (green) dots represent the mean (median) of the distribution of relative errors in each parameter, whereas the standard deviation from the mean is shown as the error bars.
spheric parameters and total metallicity given by APOGEE, and the parallaxes as the inverse of the distance given by PARAM. We then compare our estimates with the ones from PARAM. The comparison is shown in Figure 9. The estimated distances are in excellent agreement (upper left panel), since the well-constrained input parallaxes are used by StarHorse when building the marginalized distance PDF for each star.

The upper right panel of Figure 9 shows the comparison between StarHorse and PARAM ages. The discreteness of our age grid is visible in the plot. The scatter is large, but most of the stars have ages that agree within ±50% of each other. The median and 84%-ile positions in the distribution of age discrepancies are 12% and 65%, respectively (Table 4). As in the case of the EBs, StarHorse tends to yield larger ages than those based on asteroseismology, but this time with a smaller median systematic (16%). In general, the results are quite similar to those from the previous section. This is also true for the extinction estimates, shown in the bottom right panel of Fig. 9. The mass estimates tend to show a better agreement in the case of CoRoGEE, with no systematic trend and smaller errors (median value of 4%) when compared to the EBs.

5.3 The OCCASO Clusters

As a third comparison sample, we use the data from the Open Clusters Chemical Abundances from Spanish Observatories (OCCASO) sample (Carrera et al. 2016; Casamiquela et al. 2016). This sample contains a total of 128 stars from 18 clusters, covering Galactocentric distances out to 6 kpc, and a small range in metallicity. OCCASO contains only red clump stars, for a better spectroscopic resolution in spectroscopy. The age and distance estimates for these clusters are based on isochrone fitting, with a typical uncertainty of 0.2 mag in distance modulus and a mean age uncertainty of 0.2 Gyr. The input parameters used for StarHorse were parallaxes, converted from the isochrone distances, metallicities, and atmospheric parameters for each star from high-resolution spectroscopy from OCCASO survey; the mean uncertainty on these parameters can be seen in Table 3.

Figure 10 shows the comparison to StarHorse distances and ages. Each point is a star, for which the cluster’s distance and age are attributed. The distances agree with no systematics and median (84%-ile) discrepancies of 2% (13%). Only a few of stars from the most distant cluster resulted with StarHorse distances significantly above the one from isochrone fitting. The ages exhibit larger relative discrepancies (31% and 63%, respectively for the 50%-ile and 84%-ile positions). We notice from the figure that the stars with a high discrepancy on distance also results in a high discrepancy in age (dark blue points), these points belong to the NGC 6791 cluster – one of oldest open clusters in the Milky Way. Still, for most of the stars the age estimates are consistent within the relatively large error bars.

6 RELEASED DATA PRODUCTS

We applied StarHorse to a few spectroscopic surveys together with astrometric and photometric measurements to deliver public distances and extinction catalogues. For the moment we are not releasing ages and masses because their estimates are still subject to considerable improvement. Among the several spectroscopic surveys available to the community, we chose among those that have been more widely used and which have high to medium resolution. All catalogues are available to the community via the LineA web page.

The following subsections explain the procedure used to determine distances and extinctions for each spectroscopic survey. We applied StarHorse in these samples with all priors described in §3, and in all cases the following corrections were made to each survey catalogue:

(i) Polish photometry and astrometry

In some few cases in which a 2MASS (Cutri et al. 2003) magnitude of a star exists, but the associated uncertainty is >9999.99, we substitute this value by 0.2 mag. Similarly, if the quoted APASS (Henden & Munari (2014)) magnitude uncertainty is 0, we set it to be 0.15 mag, and introduce an error floor of 0.02 mag.

For the cases where parallaxes from Gaia-DR1 are available, we followed the recommendation of Arenou et al. (2017) to correct for a global mean shift of parallax by setting \( \pi = \pi_{\text{TGAS}} + 0.036 \text{ mas} \).

2 http://www.linea.gov.br/020-data-center/acesso-a-dados-3/spectrophotometric-distances-starhorse-code/
Table 4. Results of the external validation tests. We report the percentiles of the modulus of the relative deviation distribution and the median of the relative deviation distribution for each parameter estimated by StarHorse. Data are shown for all reference samples used.

| Sample               | Parameter | $P_5$ | $P_{16}$ | $P_{50}$ | $P_{84}$ | $P_{95}$ | median |
|----------------------|-----------|-------|----------|----------|----------|----------|--------|
| PARSEC high-res      | $d$       | 0.005 | 0.015    | 0.063    | 0.163    | 0.258    | 0.0    |
|                      | $\tau$   | 0.002 | 0.109    | 0.350    | 0.972    | 4.203    | 0.0    |
|                      | $m_*$     | 0.001 | 0.012    | 0.074    | 0.222    | 0.391    | 0.0    |
|                      | $A_V$     | 0.005 | 0.014    | 0.048    | 0.107    | 0.162    | 0.0    |
| PARSEC low-res       | $d$       | 0.005 | 0.016    | 0.060    | 0.172    | 0.289    | -0.001 |
|                      | $\tau$   | 0.014 | 0.110    | 0.398    | 1.120    | 4.550    | 0.002  |
|                      | $m_*$     | 0.003 | 0.017    | 0.083    | 0.247    | 0.429    | -0.001 |
|                      | $A_V$     | 0.016 | 0.049    | 0.169    | 0.354    | 0.499    | 0.004  |
| TRILEGAL all priors  | $d$       | 0.007 | 0.023    | 0.079    | 0.171    | 0.240    | -0.04  |
|                      | $\tau$   | 0.017 | 0.054    | 0.188    | 0.516    | 0.937    | 0.11   |
|                      | $m_*$     | 0.006 | 0.019    | 0.064    | 0.132    | 0.206    | -0.04  |
|                      | $A_V$     | 0.017 | 0.054    | 0.184    | 0.354    | 0.529    | -0.11  |
| Detached Eclipsing Binaries | $d$ | 0.001 | 0.002    | 0.017    | 0.041    | 0.047    | -0.015 |
|                      | $\tau$   | 0.009 | 0.022    | 0.246    | 0.764    | 1.94     | 0.22   |
|                      | $m_*$     | 0.008 | 0.033    | 0.124    | 0.165    | 0.194    | -0.12  |
|                      | $A_V$     | 0.017 | 0.030    | 0.186    | 0.354    | 0.529    | -0.07  |
| Eclipsing Binaries   | $d$       | 0.032 | 0.071    | 0.178    | 0.298    | 0.419    | -0.178 |
|                      | $\tau$   | 0.005 | 0.021    | 0.099    | 0.168    | 0.23     | 0.048  |
|                      | $m_*$     | 0.021 | 0.044    | 0.130    | 0.269    | 0.338    | 0.075  |
| CoRoGEE              | $d$       | 0.000 | 0.001    | 0.003    | 0.008    | 0.014    | 0.0    |
|                      | $\tau$   | 0.013 | 0.047    | 0.120    | 0.650    | 1.253    | 0.16   |
|                      | $m_*$     | 0.003 | 0.011    | 0.043    | 0.097    | 0.164    | 0.0    |
|                      | $A_V$     | 0.009 | 0.024    | 0.070    | 0.158    | 0.270    | -0.04  |
| OCCASO cluster members | $d$   | 0.000 | 0.001    | 0.0244   | 0.130    | 0.310    | 0.0    |
|                      | $\tau$   | 0.028 | 0.090    | 0.310    | 0.626    | 0.968    | -0.07  |

...
Table 5. General description of the release distances and extinctions

| Column      | Description                  | units  |
|-------------|------------------------------|--------|
| OBJECT ID   | Survey’s object ID name      |        |
| glon        | Galactic longitude           | degrees|
| glat        | Galactic latitude            | degrees|
| dist05      | 5th percentile of the star’s distance PDF | kpc    |
| dist16      | 16th percentile of the star’s distance PDF | kpc    |
| dist50      | 50th percentile of the star’s distance PDF | kpc    |
| dist84      | 84th percentile of the star’s distance PDF | kpc    |
| dist95      | 95th percentile of the star’s distance PDF | kpc    |
| meandist    | Mean of the star’s distance PDF | kpc    |
| diststd     | Standard deviation of the star’s distance PDF | kpc    |
| AV05        | 5th percentile of the star’s extinction PDF | mag   |
| AV16        | 16th percentile of the star’s extinction PDF | mag   |
| AV50        | 50th percentile of the star’s extinction PDF | mag   |
| AV84        | 84th percentile of the star’s extinction PDF | mag   |
| AV95        | 95th percentile of the star’s extinction PDF | mag   |
| meanAV      | Mean of the star’s extinction PDF | mag   |
| stdAV       | Standard deviation of the star’s extinction PDF | mag   |
| SH_INPUTFLAGS | StarHorse flags regarding the input data |        |
| SH_OUTPUTFLAGS | StarHorse flags regarding the output data |        |

Figure 8. Comparison between our distance (upper panel), mass (left lower panel) and extinction (right lower panel), results with those from the EB sample of Stassun & Torres (2016). The dashed lines correspond to the same deviates as in Figure 7. The legend shows the orange dots representing the primary star, the purple dots the secondary star, and the blue dots represents detected triple systems.

Figure 9. Same panels as in Figure 7, but now showing the comparison to the CoRoGEE sample.

these cases amount to 150 K in \( T_{\text{eff}} \), 0.3 dex in \( \log g \), 0.15 dex in [M/H], and 0.1 dex in \([\alpha/M]\).

(iv) \( A_V \) prior As priors for the V-band extinction we used the APOGEE targeting extinction values (Zasowski et al. 2013), derived by the RJCE method (Majewski et al. 2011), by setting \( A_{V,\text{prior}} = A(K_s)_{\text{Targ}}/0.12 \). When RJCE estimates were not available, we used \( E(B-V) \) estimates from (Schlegel et al. 1998) to estimate the prior \( A_V \). As explained in Sec. 3.2.2, the posterior \( A_V \) values are allowed to lie in a very broad interval around the prior values.

We then applied StarHorse to the resulting catalogue. Our results are available through an SDSS-IV value-added catalogue (VAC) of APOGEE stellar distances, as part of SDSS DR14. The output format of this catalogue is described

\[ \text{http://www.sdss.org/vac/apogee-dr14-based-distance-estimations/} \]
Therefore, we converted the GES abundances in the total $\alpha$ The PARSEC models do not list individual elemental abundances. From the 7870 field stars, 6316 of them meet these criteria. than 0.2 dex, with no cuts in the abundances. From the quality criteria: relative errors in Teff less than 5%; errors in logg lower than 0.4 dex; and errors in metallicity lower leaves us with 7870 stars. We then adopted the following StarHorse From this catalogue we selected only Milky Way field stars, 2MASS $JHK$ columns, from which we selected the necessary information to run StarHorse. This catalogue contains 100 clusters, and calibration stars. The catalogue contains 100 columns, from which we selected the necessary information to run StarHorse: atmospheric parameters, elemental abundances, and 2MASS $JHK_s$ pass-bands (Cutri et al. 2003). From this catalogue we proceeded with the following steps, before applying StarHorse.

(i) Select only field stars and reliable sources: From this catalogue we selected only Milky Way field stars, which are the main targets intended for StarHorse. This leaves us with 7870 stars. We then adopted the following quality criteria: relative errors in Teff less than 5%; errors in logg lower than 0.4 dex; and errors in metallicity lower than 0.2 dex, with no cuts in the abundances. From the 7870 field stars, 6316 of them meet these criteria.

(ii) Calculate an overall $[\alpha/Fe]$ abundance: The PARSEC models do not list individual elemental abundances, only the total metallicity value, [M/H]. The models do not include non-Solar abundances in $\alpha$ elements either. Therefore, we converted the GES abundances in the total metallicity [M/H] before running StarHorse. For that purpose, we calculated the overall $[\alpha/Fe]$ abundance as follows:

$$[\alpha/Fe] = \frac{1}{n} \sum_{i}^{n} [X_i/Fe] - [Fe/Fe],$$

where $X_i$ refers to the elements O, S, Ti, Ca and Mg, and the $[X_i/Fe]$ abundances were calculated using the solar values from Asplund et al. (2009). The error in $[\alpha/Fe]$ was propagated in quadrature from the error on each elemental abundance and the error in $[Fe/Fe]$.

(iii) Cross-match with photometry: To obtain more precise and reliable extinction estimates we decided to include APASS (Henden & Munari 2014) magnitudes for this sample. For this we carried out a positional crossmatch with APASS DR9 (Henden & Munari 2014) using TOPCAT (Taylor 2005), with a maximum separation of 1 arcsec. Of the 6316 stars, 5719 stars have APASS magnitudes.

(iv) $A_V$ Prior: No $A_V$ prior was applied to this sample, therefore the posterior probability function for $A_V$ always ranges from 0 to 3 (mag).

After carrying out these steps, we used this final catalog with 6316 stars as input to StarHorse. The code delivered an output catalog with 6011 stars with available distances and extinctions. The columns of the released distances and extinction catalogue are shown in Table 5, and the uncertainties distribution are shown in Figure 12.

6 http://www.eso.org/rm/api/v1/public/releaseDescriptions/91

6.2 The Gaia-ESO sample

The Gaia-ESO survey (GES Gilmore et al. 2012) is a large public spectroscopic survey with high resolution that covers all Milky Way components and open star clusters of all ages and masses. The final GES release is expected to include about 10$^5$ stars.

We downloaded the Gaia-ESO data release 3 (DR3) from the ESO catalogue facility. This catalogue contains a total of 25533 stars, including the Milky-Way field, open clusters, and calibration stars. The catalogue contains 100 columns, from which we selected the necessary information to run StarHorse: atmospheric parameters, elemental abundances, and 2MASS $JHK_s$ pass-bands (Cutri et al. 2003). From this catalogue we proceeded with the following steps, before applying StarHorse.

(i) Select only field stars and reliable sources: From this catalogue we selected only Milky Way field stars, which are the main targets intended for StarHorse. This leaves us with 7870 stars. We then adopted the following quality criteria: relative errors in Teff less than 5%; errors in logg lower than 0.4 dex; and errors in metallicity lower than 0.2 dex, with no cuts in the abundances. From the 7870 field stars, 6316 of them meet these criteria.

(ii) Calculate an overall $[\alpha/Fe]$ abundance: The PARSEC models do not list individual elemental abundances, only the total metallicity value, [M/H]. The models do not include non-Solar abundances in $\alpha$ elements either. Therefore, we converted the GES abundances in the total metallicity [M/H] before running StarHorse. For that purpose, we calculated the overall $[\alpha/Fe]$ abundance as follows:

$$[\alpha/Fe] = \frac{1}{n} \sum_{i}^{n} [X_i/Fe] - [Fe/Fe],$$

where $X_i$ refers to the elements O, S, Ti, Ca and Mg, and the $[X_i/Fe]$ abundances were calculated using the solar values from Asplund et al. (2009). The error in $[\alpha/Fe]$ was propagated in quadrature from the error on each elemental abundance and the error in $[Fe/Fe]$.

(iii) Cross-match with photometry: To obtain more precise and reliable extinction estimates we decided to include APASS (Henden & Munari 2014) magnitudes for this sample. For this we carried out a positional crossmatch with APASS DR9 (Henden & Munari 2014) using TOPCAT (Taylor 2005), with a maximum separation of 1 arcsec. Of the 6316 stars, 5719 stars have APASS magnitudes.

(iv) $A_V$ Prior: No $A_V$ prior was applied to this sample, therefore the posterior probability function for $A_V$ always ranges from 0 to 3 (mag).

After carrying out these steps, we used this final catalog with 6316 stars as input to StarHorse. The code delivered an output catalog with 6011 stars with available distances and extinctions. The columns of the released distances and extinction catalogue are shown in Table 5, and the uncertainties distribution are shown in Figure 12.

6.3 GALAH sample

The Galactic Archaeology with HERMES (GALAH; Martell et al. 2017) is a spectroscopic survey that will target about 1 million stars with the high-resolution ($R \approx 28,000$) instrument High Efficiency and Resolution Multi-Element Spectrograph (HERMES; Raskin et al. 2011), at the Anglo-Australian Telescope (AAT). The main goal of the project is to provide a detailed star-formation history for the thick and thin disks. Therefore, the survey covers mainly the disk, but it also has some fields towards the bulge and halo. A first public data release of GALAH is already available. We then applied StarHorse to estimate distances and extinction for GALAH DR1 stars. The catalogue contains 10680 stars with [Fe/H], [α/Fe], log g, and $T_{\text{eff}}$ measurements derived by the Cannon method (Ness et al. 2015). We proceeded with the following steps to have an input catalogue ready for StarHorse.

(i) Stellar parameter uncertainties: The GALAH DR1 catalogue does not provide individual uncertainties for the stars. We therefore used the values recommended by Martell et al. (2017): $\sigma([Fe/H]) = 0.056$ dex, $\sigma(\log g) = 0.17$ dex and $\sigma(T_{\text{eff}}) = 51$ K. As there is no mention of the uncertainty in $[\alpha/Fe]$, so we assumed the
error to be same as for $\sigma_{[\text{Fe/H}]}$.

(ii) **Cross-match with photometry and astrometry:**
Since the GALAH DR1 catalog has stars in common with the Gaia DR1, we carried out a positional crossmatch with Gaia using TOPCAT (Taylor 2005), with a maximum separation of 1 arcsec. From the 10680 DR1 stars, 7919 stars have parallax available. We obtained photometry for the sample by crossmatching with the 2MASS (Cutri et al. 2003) (10680 matches) and APASS (Henden & Munari 2014) (9263 matches) catalogues.

Figure 11. Illustration of the APOGEE DR14 distance and extinction results from StarHorse. Top panel: Aitoff projection of median APOGEE distances per HealPix cell in Galactic coordinates. Middle panel: Resulting median $A_V$ per HealPix cell in the inner Galaxy. Bottom left panel: Spectroscopic Hertzsprung-Russell diagram, colour-coded by median distance in each pixel. Bottom right panel: 2MASS colour-magnitude diagram, colour-coded by median extinction in each pixel.
(iii) \textit{Av Prior}

We use the reddening given by the GALAH DR1 catalogue, which is derived by comparison of absolute magnitudes with the apparent magnitude \( V \) from APASS (Henden & Munari 2014) and \( J \) magnitude from 2MASS (Cutri et al. 2013). We assumed then that \( A_V = 3.1(E(B-V)) \).

After these steps were completed, we used this final file as input to \textit{StarHorse} and derived distances and extinctions. The code returned 10,623 distances and extinctions. The columns for the released catalogue are described in Table 5 and the uncertainty distribution is shown in Figure 12.

6.4 The RAVE catalogues

The RAdial Velocity Experiment (RAVE, Steinmetz (2003)) is one of the largest spectroscopic surveys of the Milky Way. RAVE has already delivered spectra for almost 500K stars, that were randomly targeted in an area of 20K square degrees of the Galactic Southern Hemisphere. In addition, RAVE is currently the survey that contains the largest number of stars in common with TGAS (\( \sim 200K \) stars). The survey works with a multi-object spectrograph deployed on 1.2-m UK Schmidt Telescope of the Australian Astronomical Observatory (AAO). The spectra have a medium resolution of (\( R \sim 7,500 \)) and cover the Ca-triplet region (8410-8795 \( \text{Å} \)).

Given that RAVE has a medium resolution, is pioneering among the large spectroscopic surveys, covers a large area, and has the largest overlap with the Gaia sample, we decided to apply \textit{StarHorse} to the entire survey, and to make the estimated distances and extinctions available to the community. The following subsections explain how we proceed with the RAVE input catalogues to execute \textit{StarHorse} and the description of the released distances-extinction catalogues. All RAVE catalogues were downloaded from the RAVE website\(^8\).

6.4.1 The RAVE DR5 catalogue

The data release 5 (DR5 Kunder et al. 2017) is the lastest RAVE data release. It contains spectra for 483,330 stars. We downloaded the publicly released catalogue called \textit{RAVE}_{DR5}, which contains spectral parameters and radial velocities derived by the SPARV pipeline (Zwitter et al. 2008; Siebert et al. 2011). The catalogue also contains astrometry from \textit{Gaia-DR1} (215,590), and photometry from 2MASS and APASS. We note that very recently, McMillan et al. (2017) updated the DR5 catalogue parameters and derived distances using feedback from the \textit{Gaia} DR1 parallaxes, but our VAC presented here is based on the public DR5 data. From this catalogue, we proceeded with the following steps before running \textit{StarHorse}:

(i) \textit{Spectral parameters}

We use the calibrated atmospheric parameters, which are named in the catalogue as \( \text{Teff}_K \), \( \log g_N K \), \( \text{M}_K \). For the uncertainties, if the error spectral analysis is available, we use the maximum between the two values: \( \sigma \text{Teff}_K \) and StdDev\( \text{Teff}_K \), otherwise we use the maximum between: 70K

\( ^8 \) \url{https://www.rave-survey.org/project/}
and $\sigma_{\text{Teff}}$. We worked analogously with the other parameters.

(ii) $\Delta V$ prior
As explained in section 3.2.2, we can use a prior value of extinction to build the $\Delta V$ posterior probability function. We use as extinction prior in $V$-band the maps of $E(B - V)$ from Schlegel et al. (1998) for this catalogue.

6.4.2 RAVE-SC catalogue
The RAVE-SC catalogue has stars from DR5 with gravity from seismic calibrations (Valentini et al. 2017). We downloaded the catalogue named as RAVE\_Gravity\_SC. Step 1 from the previous subsection 6.4.1 was also applied to this sample. We use the overall $[\alpha/\text{Fe}]$ abundance and the $[\text{Fe}/\text{H}]$, given by the catalogue to calculate a total metallicity as defined by (Salaris et al. 1993), with a fixed uncertainty of 0.2 dex. The atmospheric parameters were used as they were given by the catalogue; we use the following temperature and surface gravity columns: $\text{Teff}_{\text{IR}}$ and $\log g_{\text{SC}}$. For the $A_V$ prior we use the Schlegel et al. (1998) $E(B - V)$ maps.

6.4.3 RAVE-on catalogue
The RAVE-on catalogue (Casey et al. 2016a) has stars from DR5 with parameters derived by the Cannon method (Casey et al. 2016b). We downloaded the catalogue named as RAVE-ON. The atmospheric parameters and $[\text{Fe}/\text{H}]$ were used directly from this catalogue. The following steps were applied before applying StarHorse:

(i) Calculate an overall $[\alpha/\text{Fe}]$ abundance
The Cannon provides the individual abundances for the stars. We then calculated $[\alpha/\text{Fe}]$ as the simple average between the individual abundances when they are available, exactly as described in section 6.2, with $X_i$ as O, Mg, Ca, and Si.

(ii) $\Delta V$ Prior:
No $\Delta V$ prior was applied to this sample, therefore the posterior probability function for $A_V$ always ranges from 0 to 3 (mag).

6.5 StarHorse FLAGS
All released data products catalogues have two columns that describe the StarHorse input data, SH\_INPUTFLAGS, and the StarHorse output data, SH\_OUTPUTFLAGS, as shown in Table 5. The input flags specify which parameters were used in the likelihood calculation to estimate the distances and extinctions given. For example, if the temperature was available for that star a $\text{Teff}$ flag will appear. The other parameters are specified as follows in the input flag: LOGG (surface gravity), PARALLAX (parallax), MH (metallicity), JHKs (2Mass filters) and BVgriz (APASS filters). If the input flag contains ALPHAM it means that the alpha elements were available in the calculation of the total metallicity of the star. The input flags also indicate you if we use a $A_V$ prior as the AVprior flag. The output flags tell us if the number of models which have converged in the likelihood calculation is too small. If less than 10 models are consistent with the star a NUMMODELS\_BAD flag will appear, while if the number of models is between 10 and 30 a NUMMODELS\_WARM flag will appear. The output flags also indicate if the estimated extinction is negative (NEGATIVE\_EXTINCTION\_WARN), if it is too high (HIGH\_EXTINCTION\_WARN), or if the estimated extinction has a bright 2Mass source (EXTINCTION\_BAD\_BRIGHT2MASS).

7 SUMMARY AND FUTURE PERSPECTIVES
We have presented a code that computes distances, ages, masses, and extinctions for field stars with photometric, spectroscopic, and astrometric data. It is based on Bayesian inference, computing the marginal posterior distributions for the data given a set of stellar models. The code represents a significant improvement over the one presented by Santiago et al. (2016) in several aspects. The most important one is the ability to estimate ages, masses, and extinction, in addition to the spectrophotometric distances presented by those authors. The updated code, which we call StarHorse, is also capable of incorporating the parallax as an additional observational quantity in the statistical analysis (Figure 1). Updated spatial, metallicity, and age priors for the Galactic components, now including the bulge as well, are presented (Table 1). In addition, StarHorse is now more flexible in terms of the input data and the choice of observational quantities to be used within them.

The new code was validated using simulated and real stars. These latter are samples with reliable parallax (or distance) data, including field giants with asteroseismic data, EBs used as benchmarks for stellar evolutionary codes, or cluster stars with well-known distances and ages, usually from isochrone fitting, often in combination with spectroscopic data. For EBs that are not detached, the distances present an offset in relation to the reference ones, our distances being usually smaller by 20% in this specific case. The discrepancy is larger for secondary stars than for the primary ones, which is what one expects from using systemic photometric measurements. In all cases, age is the single most difficult parameter to infer, yielding median errors that range from 12% to 35% for quality spectrophotometric data, depending on the sample (Table 4). Errors larger than 100% in age may result for ≃ 15% of the stellar models, most of them younger than $\tau \simeq 1$ Gyr. In a realistic flux-limited sample, as simulated by the TRILEGAL code or for real stars, the fraction of such catastrophic age errors is reduced to ≃ 5% of the stars. Our results for stellar ages, either based on simulated or real stars, also indicate a systematic trend of StarHorse overestimating ages by 10 − 20%.

As for spectrophotometric masses, we obtain consistent results over all validation samples used, in the sense that errors < 20% are observed for most (84%) of the stars in any sample. The median error varies depending on the quality of the parallax used as a constraint. For typical Gaia-TGAS errors of 0.3 mas, the typical distance errors are around 15%. For real stars used as reference, median StarHorse $A_V$ errors are of 0.07 mag, with the 84% − iles error at 0.15 − 0.20 mag. For TRILEGAL synthetic stars, median and 84% − iles $A_V$ errors are larger, 18% and 38%, respectively. One reason
may be the larger parallax (and hence distance) errors attributed to the TRILEGAL stars compared to the real ones, since the estimated extinction is coupled to distance through the comparison of model absolute magnitudes to observed apparent magnitudes.

We note that the error estimates based on comparison to real samples may be overestimated, considering that the some of the discrepancy may be attributed to the methods used to obtain the reference quantities for comparison. In fact, Rodrigues et al. (2017) report that ages and masses from asteroseismology are typically obtained with a precision of 19% and 5%, respectively, which are comparable to the errors we quote in this analysis.

StarHorse has already been used to infer distances and extinction values for stars from APOGEE DR14. These parameters, in turn, may be used in connection to APOGEE abundances and radial velocities, to study the properties of the main Galactic populations, and their spatial variations, as was previously done by Anders et al. (2014), and Fernández-Alvar et al. (2016) using distances from Santiago et al. (2016). For more local samples, such as Gaia-TGAS and RAVE, reliable parallax information can be included in the Bayesian method to yield masses and ages, as validated in this paper, allowing for a more detailed modelling of the chemo-dynamical history of our Galaxy (Anders et al. 2017).

Finally, we have run StarHorse on different public catalogs from the RAVE collaboration, as well as on GES, and GALAH public data releases. These are available for download at the LIneA web site.9

ACKNOWLEDGEMENTS

We thank Laia Casamiquela for providing data from the OCCASO survey.

Funding for the SDSS Brazilian Participation Group has been provided by the Ministério de Ciência e Tecnologia (MCT), Fundação Carlos Chagas Filho de Amparo à Pesquisa do Estado do Rio de Janeiro (FAPERJ), Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), and Financiadora de Estudos e Projetos (FINEP).

Funding for the Sloan Digital Sky Survey IV has been provided by the Alfred P. Sloan Foundation, the U.S. Department of Energy Office of Science, and the Participating Institutions. SDSS-IV acknowledges support and resources from the Center for High-Performance Computing at the University of Utah. The SDSS web site is www.sdss.org.

SDSS-IV is managed by the Astrophysical Research Consortium for the Participating Institutions of the SDSS Collaboration including the Brazilian Participation Group, the Carnegie Institution for Science, Carnegie Mellon University, the Chilean Participation Group, the French Participation Group, Harvard-Smithsonian Center for Astrophysics, Instituto de Astrofísica de Canarias, The Johns Hopkins University, Kavli Institute for the Physics and Mathematics of the Universe (IPMU) / University of Tokyo, Lawrence Berkeley National Laboratory, Leibniz-Institut für Astrophysik Potsdam (AIP), Max-Planck-Institut für Astronomie (MPIA Heidelberg), Max-Planck-Institut für Astrophysik (MPA Garching), Max-Planck-Institut für Extraterrestrische Physik (MPE), National Astronomical Observatory of China, New Mexico State University, New York University, University of Notre Dame, Observatório Nacional / MCTI, The Ohio State University, Pennsylvania State University, Shanghai Astronomical Observatory, United Kingdom Participation Group, Universidad Nacional Autónoma de México, University of Arizona, University of Colorado Boulder, University of Oxford, University of Portsmouth, University of Utah, University of Virginia, University of Washington, University of Wisconsin, Vanderbilt University, and Yale University.

CC acknowledges support from DFG Grant CH1188/2-1 and from the ChETEC COST Action (CA16117), supported by COST (European Cooperation in Science and Technology).

ABAQ and FA acknowledge support from the Leibniz Graduate School for Quantitative Spectroscopy at AIP, in particular the Outgoing and Incoming Mobility Programme.

TCB acknowledges partial support from grant PHY 14-30152: Physics Frontier Center/JINA Center for the Evolution of the Elements (JINA-CEE), awarded by the US National Science Foundation.

This work has made use of data from the European Space Agency (ESA) mission Gaia (http://www.cosmos.esa.int/gaia), processed by the Gaia Data Processing and Analysis Consortium (DPAC, http://www.cosmos.esa.int/web/gaia/dpac/consortium). Funding for the DPAC has been provided by national institutions, in particular the institutions participating in the Gaia Multilateral Agreement. This work has also made use of data from Gaia-ESO based on data products from observations made with ESO Telescopes at the La Silla Paranal Observatory under programme ID 188.B-3002.

REFERENCES

Abolfathi B., et al., 2017, preprint, (arXiv:1707.09322)
Anders F., et al., 2014, A&A, 564, A115
Anders F., et al., 2017, A&A, 597, A30
Arenou F., et al., 2017, preprint, (arXiv:1701.00292)
Asplund M., Grevesse N., Sauval A. J., Scott P., 2009, ARA&A, 47, 481
Bensby T., et al., 2013, A&A, 549, A147
Bessell M. S., 1990, PASP, 102, 1181
Binney J., et al., 2014, MNRAS, 437, 351
Bland-Hawthorn J., Gerhard O., 2016, ARA&A, 54, 529
Blanton M. R., et al., 2017, AJ, 154, 28
Boeche C., et al., 2013, A&A, 553, A19
Boeche C., et al., 2014, A&A, 568, A71
Bovy J., Rix H.-W., Schlafly E. F., Nidever D. L., Holtzman J. A., Shetrone M., Beers T. C., 2016, ApJ, 823, 30
Bressan A., Marigo P., Girardi L., Salasnich B., Dal Cero C., Rubele S., Nanni A., 2012, MNRAS, 427, 127
Brogdata K., et al., 2016, Astronomische Nachrichten, 337, 793
Burnett B., Binney J., 2010, MNRAS, 407, 339
Burnett B., et al., 2011, A&A, 532, A113
Cardelli J. A., Clayton G. C., Mathis J. S., 1989, ApJ, 345, 245

9 http://www.linea.gov.br/020-data-center/acesso-a-dados-3/spectrophotometric-distances-starhorse-code/

MNRAS 000, 1–21 (2017)
APPENDIX A: ADDITIONAL SIMULATION RESULTS

Here we show the results of PARSEC simulations when parallaxes are not used to constrain distances or the likelihood. Figure A1 can then be compared to the case shown in the main body of the paper, Figure 3. All estimated parameters are subject to larger errors, especially $\alpha V$. The systematic distance and mass error dependences on true stellar age and mass become very pronounced when parallaxes are not used.

We also show the results of TRILEGAL simulations for the case where no priors are adopted, to be compared to those shown in Figure 4, for the All priors case. Figure A2 show the results for TRILEGAL simulations with no spatial, MDF or ADF see $\S 3.3$.

This paper has been typeset from a TeX/LaTeX file prepared by the author.
Figure A1. Same panels as in Figure 3, but now showing the results from StarHorse when the constraint provided by the parallax is not used.
Figure A2. Same panels as in Figure 4, but now showing the results from StarHorse when no priors in metallicity, age and spatial distribution are adopted.