

Fundamental stellar parameters and metallicities from Bayesian spectroscopy: application to low- and high-resolution spectra

Ralph Schönrich^{1,2*}, and Maria Bergemann^{3,4}

¹ *Hubble Fellow, Department of Astronomy, The Ohio State University, 140 West 8th Avenue, 43210 Columbus, Ohio, USA*

² *Rudolf-Peierls Centre for Theoretical Physics, University of Oxford, 1 Keble Road, OX1 3NP, Oxford, United Kingdom*

³ *Max-Planck-Institut für Astrophysik, Karl-Schwarzschild-Str. 1, 85741 Garching, Germany*

⁴ *Institute of Astronomy, University of Cambridge, Madingley Road, CB3 0HA, Cambridge, United Kingdom*

Draft, 25 November 2013

ABSTRACT

We present a novel method based on Bayesian inference to derive physical parameters of stars from any observed data. Here we focus on $\log(g)$, T_{eff} , $[\text{Fe}/\text{H}]$, age, mass, and distance of a star. As an input, we use spectra, photometric magnitudes, parallaxes, and stellar evolution models. The method simultaneously computes full probability distributions for all desired parameters and delivers the first comprehensive and objective error estimates. The classical spectroscopic analysis, in the form of global spectrum synthesis, is directly integrated into the Bayesian framework, which also allows to include chemical abundances in the scheme. We lay out the mathematical framework and apply it to high-resolution spectra (UVES, HARPS, NARVAL instruments), as well as low-resolution spectra from SDSS/SEGUE survey. The method is flexible and can be applied to the analysis of single stars, large stellar datasets, or unresolved stellar populations. By its flexibility and the simultaneous analysis of multiple independent measurements for a star, it will be ideal to analyse and cross-calibrate the large ongoing and forthcoming surveys, like Gaia-ESO, SDSS, Gaia and LSST.

Key words: stars: fundamental parameters – stars: distances – techniques: photometric – techniques: spectroscopic – methods: statistical – methods: data analysis

1 INTRODUCTION

Observations are a central source of knowledge on almost any entity in astrophysics. Over several centuries of intense research, several principal observational techniques have been developed that are now routinely used to study stars and stellar populations in the Milky Way and other galaxies. We have information from astrometry, photometry, spectroscopy, but also interferometry, and astroseismology, that give complementary information on the physical parameters of stars (detailed chemical composition, gravities, temperatures, masses and ages) and their kinematics (radial velocities, distances, and orbital characteristics). However, in contrast to e.g. cosmology, where sophisticated Bayesian schemes have been established (e.g. Drell et al. 2000; Kitaura & Enßlin 2008), stellar parameter determinations are still widely based on best-fit estimates and simple averages between different methods.

The advent of large stellar spectroscopic and photometric surveys like SEGUE/SDSS (Yanny et al. 2009), RAVE (Steinmetz et al. 2006), APOGEE (Majewski et al. 2007), GCS (Nordström et al. 2004), and the Gaia-ESO survey (Gilmore et al. 2012), as well as astroseismic surveys like Kepler (Chaplin et al. 2011), makes it necessary to develop fully automated methods

for data analysis and determination of stellar parameters. Standard spectroscopic inversion methods are commonly assumed to be accurate, however, they usually involve subjective and hardly reproducible elements, like line fitting and normalisation, or decisions on spectral diagnostic features. Manual analysis of stars is limited to sample sizes of ~ 1000 stars, unsuitable for large surveys. Existing automated methods usually suffer from weakly constrained systematics as well as idealised error estimates. So far, all attempts to overcome these problems have concentrated on simple weighted averaging between different methods (e.g. Lee et al. 2008a,b).

The large stellar surveys change stellar astronomy into a precision science, where we cannot limit ourselves to pointing out structures in diagrams, but where knowledge of the error distributions is key to make meaningful model comparisons, e.g. of Galactic evolution and stellar structure. The approach we need must be flexible, objective, applicable to very large datasets, and provide an optimal combination of the different bodies of observational data. The only mathematical apparatus known to permit a systematic combination of different quantities are Bayesian frameworks. The first steps in this direction were made by Pont & Eyer (2004), Jørgensen & Lindegren (2005), Shkedy et al. (2007), Burnett & Binney (2010), Casagrande et al. (2011), Binney et al. (2013), Serenelli et al. (2013), and Gruberbauer & Guenther (2013). However, the studies are very limited in scope and applica-

* E-mail: ralph.schoenrich@physics.ox.ac.uk

bility: they either addressed the problem of fitting a spectrum only (Shkedy et al. 2007), partly focussed on the problem of finding the maximum likelihood solution, or leave have to rely on treating the data with overly simplified parametrisations (e.g. Burnett & Binney 2010; Casagrande et al. 2011). In contrast, a Bayesian scheme can only fulfill its claim of unbiased information, if the full dimensionality of the constraints in parameter space is preserved.

In the following we will present a new method for the determination of stellar parameters that provides an optimal exploitation of different observational information. The method offers a homogeneous full-scale quantitative recovery of the full probability distributions in parameter space, which are given by the available observations, i.e. photometry, astrometry, spectroscopy, and well-established knowledge from stellar evolution theory and Galaxy structure. The method is objective, computationally efficient, can be readily applied to data from all existing surveys and is robust to missing bits of data, e.g. damaged pixels in a spectrum or low-quality photometry. By embedding spectroscopic analysis directly in scheme, the Bayesian method allows for consideration of all pieces of relevant information at once, thus avoiding unnecessary information loss.

In this first paper of the series, our main goal is to determine effective temperature, surface gravity, metallicity, mass, age and distances of individual stars. Thus, we limit the input data to spectroscopy, photometry, stellar evolution models and facultatively parallax measurements. However, the method can be readily generalised to any number of parameters, such as kinematics or stellar rotation, and include other input information, e.g. asteroseismology and interferometric angular diameters. Furthermore, it is straightforward to analyse star formation history of a whole stellar population, e.g. a young cluster or an old galaxy, using its integrated colours and spectra. Thus the Bayesian method has a very broad scope to applications both in the context of Galactic and extra-galactic research.

The paper is structured as follows. In Sec. 2 and 3, we present the details of the algorithm and its implementation. Sec. 4 presents the results for a sample of stars with very high-resolution observations and for a sub-sample of calibration stars from the SDSS/SEGUE catalogue. We close with discussion and conclusions are drawn in Sec. 5.

2 METHOD OUTLINE

2.1 Bayesian scheme

So far, the majority of observational studies of stars, be it photometric or spectroscopic, have focussed on providing best-fit estimates of stellar parameters. However, accurate comparisons to theoretical models of e.g. galaxy evolution, require the full probability distribution of the derived parameters given the available observations.

This demands a Bayesian formalism. In this context we need to express the probability of a set of parameters $\mathbf{X} = X_1, \dots, X_n$ given a set of observations $\mathbf{O} = O_1, \dots, O_m$ by the probability that this observation could take place given the set of parameters. By definition the conditional probability $P(\mathbf{X}|\mathbf{O})$, that \mathbf{X} given \mathbf{O} , derives from the combined probability $P(\mathbf{X}, \mathbf{O}) = P(\mathbf{X}|\mathbf{O})P(\mathbf{O})$. We can hence write down:

$$P(\mathbf{X}|\mathbf{O}) = \frac{P(\mathbf{X})}{P(\mathbf{O})} P(\mathbf{O}|\mathbf{X}), \quad (1)$$

where the *posterior* probability $P(\mathbf{X}|\mathbf{O})$ is the conditional probability of the parameter set \mathbf{P} given \mathbf{O} . $P(\mathbf{O}|\mathbf{X})$, which we call observed

likelihood, is the probability of making the set of observations \mathbf{O} given the set of parameters \mathbf{X} and $P(\mathbf{X})$ is the *prior probability* we ascribe to that set of parameters. $P(\mathbf{O})$ is the probability that the set of observations was made, which we set to 1 (Pont & Eyer 2004). This simplifies our problem to

$$P(\mathbf{X}|\mathbf{O}) = P(\mathbf{X})P(O_1, \dots, O_m|\mathbf{X}), \quad (2)$$

where $P(\mathbf{X}|\mathbf{O})$ is the posterior probability distribution function (PDF) on the chosen parameter space. In our work, observations are conditionally independent given the parameters, i.e. if all parameters are perfectly known, the observations do not provide additional information about each other. Hence we can disentangle the observations by:

$$P'(\mathbf{X}) = P(O_1, \dots, O_m|\mathbf{X}) = \prod_{(j=1)}^m P(O_j|\mathbf{X}). \quad (3)$$

2.2 Core parameter space

The parameter set \mathbf{X} contains all parameters relevant to the problem under investigation and important to the description of a star. This may include surface and interior structure parameters (total emitted flux, surface gravity, mean density, etc). Since we are dealing with a single object, of course, all these parameters are related in some way. However, we can break their dependencies into main groups, using the fact that each type of observations constraints only a subset of these parameters, whereas it bears no information on others.

In this work, we define the 'core' parameter space $R_c \equiv ([\text{Fe}/\text{H}], T_{\text{eff}}, \log(g))$ of metallicity (expressed by iron abundance), effective temperature and surface gravity. The parameters in R_c impact all our observations and models.

Other parameters are constrained by only a subset of observations: e.g. detailed abundances are of importance for spectroscopic observations, while stellar magnitudes in different colour bands C span the space of the photometric parameters. Age τ , initial mass M_{init} and present mass M fall into the domain of stellar models. Distance s and parallax ω are determined either from direct astrometric observations or via the distance modulus when comparing stellar models with photometry.

Thus the full parameter space can be disentangled into individual contributions:

$$R \equiv R_c + R_{sp} + R_{ph} + R_{mod} + R_{others}, \quad (4)$$

where R_c is the core parameter space the other R_j are the parameters of importance to different types of observations or prior expectations (see Sec. 3.2 to 3.6).

2.3 Observations

In contrast to parameters, which span the n -dimensional space of the posterior probability distribution, the nature of observations is irrelevant. An observation can be anything, from the numbers of electrons on a CCD to a needle on a scale, while the corresponding observed likelihood is a function on parameter space.

A clear distinction between observations and "observables" is key to understand Bayesian schemes. Commonly (e.g. Burnett & Binney 2010) Bayesian schemes are introduced using the highly problematic concept of "observables", which denote best-fit values for some parameters (like T_{eff}), which happen to be relatively well-constrained by (single) observations. These "observables" approximate a true observation by a couple of quantities,

which are insufficient to describe its full information content: The observed likelihood is almost never a delta-function, so that one has to include separate variables for the variance, for higher orders, for cross-correlations between different parameters, etc. This results in a rather clumsy bulk of variables, which do not fully describe the true observed likelihood. Further, other than their name suggests, observable parameters are not fixed - the Bayesian formalism will in general give an estimate different from the best-fit value.

While selection functions are in most cases essential for understanding observations with theoretical models, this does not apply to the discussed Bayesian schemes. Yet, some studies introduce a selection function in their equations (see a longer discussion in Sec. 8.1 of the Appendix). We refrain from using such a selection function, because only selection criteria based on the parameter space would affect the Bayesian scheme, while a survey selection must be based on random choice or previous observations.

2.4 Summary of notation

To facilitate reading the equations we quickly summarize the main notations: We denote the set of observations by \mathbf{O} , the set of parameters by \mathbf{X} , the parameter space by R and all probability distributions by P . To cope with the different sources of information, we introduce indices: "ph" for photometry, "sp" for spectroscopy, "astr" for astrometry (parallaxes), in addition we use "mod" for knowledge from stellar models and "pr" for priors. Hence the observational likelihood in full parameter space from spectroscopic observations reads $P_{\text{sp}}(O_{\text{sp}}|\mathbf{X})$. To facilitate the reading we contract the notation for the conditional probabilities by decorating P with a prime: e.g. $P'_{\text{sp}} \equiv P'_{\text{sp}}(\mathbf{X}) \equiv P_{\text{sp}}(O_{\text{sp}}|\mathbf{X})$.

Commonly used variables are age τ , stellar mass M , solar mass M_{\odot} , initial mass M_{init} , logarithmic iron abundance $[\text{Fe}/\text{H}]$, general metallicity $[\text{Me}/\text{H}]$, parallax ω , distance s , and distance modulus μ .

3 DETAILED ALGORITHM

In this pilot study, we restrict ourselves to the most important basic case: the calculation of stellar parameters, when we have spectroscopic, and/or photometric observations. We will show how to expand this to include parallax measurements. After validating the method on the high-resolution spectroscopic data of nearby stars, we apply it to a sample of low-resolution spectra from SEGUE/SDSS (Yanny et al. 2009; Allende Prieto et al. 2008) calibration sample.

3.1 Contributors to the posterior PDF

With the conditional independence (equation 3) we simplify the calculation of our posterior:

$$P(\mathbf{X}|O_{\text{sp}}, O_{\text{ph}}) \sim P(O_{\text{sp}}|\mathbf{X}) \cdot P(O_{\text{ph}}|\mathbf{X})P(O_{\text{astr}}|\mathbf{X}) \cdot P_{\text{mod}}(\mathbf{X}) \cdot P_{\text{pr}}(\mathbf{X}) \quad (5)$$

$$\sim P'_{\text{sp}} \cdot P'_{\text{ph}} \cdot P'_{\text{astr}} \cdot P_{\text{mod}} \cdot P_{\text{pr}},$$

where $O_{\text{sp}}, O_{\text{ph}}, O_{\text{astr}}$ denote the photometric and spectroscopic and astrometric observations, P_{mod} the probability derived from stellar models and P_{pr} the prior probability distribution function. In the second line we abbreviate our observational likelihoods, representing their conditional nature based on the observations in short by a prime. Note that any combination of observational constraints can

be dropped from these equations, as well as new observations (e.g. interferometry) can be added by multiplication.

The PDFs from Equ. 5 have two interesting qualities:

- some PDFs describe sharp structures in the n -dimensional parameter space, thus lowering the dimensionality of the probability distribution and reducing computational costs (by the multiplication, the combined PDF cannot have higher dimensionality than its components). In other words, the space volume where their PDF is non-negligible has a lower dimensionality than the overall space. For example, stellar models together with model atmospheres map directly from the fundamental stellar parameters ($M_{\text{init}}, \tau, [\text{Me}/\text{H}]$) to their observed space ($T_{\text{eff}}, \log(g), C$).

- some PDFs constrain only a subset of $j < n$ parameters, i.e. they are flat the other dimensions. Though they can be generalised to the n -dimensional space of the aggregate PDF, most of these dimensions will be redundant, i.e. we have $P(X_1, \dots, X_n|\mathbf{O}) \sim P(X_1, \dots, X_j|\mathbf{O})$. It can be efficient to merge them in an early step with another PDF that carries more dimensions. An especially valuable case are parameters that are nearly conditionally independent from the other parameters. E.g. the detailed abundances for the majority of chemical elements hardly affect temperature and gravity estimates.

The meaning and structure of the single contributors to the posterior P will be examined below.

3.2 Priors P_{pr}

The priors encode our previous knowledge on the distribution of the examined stellar population in parameter space. The model knowledge P_{mod} will be treated here as a separate prior, though it could be in fact understood as an observation. Appropriate priors are essential to avoid biases in weakly constrained data (see Fig. 10 for an example). Further, to have set "no prior" means to have adopted a flat prior, which is not fixed under parameter transformations. However, priors must be handled with great caution to obey Cromwell's rule (Lindley 1982) and to avoid overconfidence biases and reproducibility problems.

There are strong dependencies between most parameters. As the posterior PDF has lower effective dimensionality than the parameter space R , we can not set constraints on every single dimension in R without risk of over-constraining the priors. We circumvent this problem by limiting our effective priors to age τ , initial mass M_{init} , metallicity $[\text{Me}/\text{H}]$, and distance s ; all other dimensions are indirectly constrained by these priors and we adopt no additional constraints on them.

Throughout this work we will use the following priors:

$$P_{\text{pr}}(\mathbf{X}) := p(\tau, M, [\text{Fe}/\text{H}], s) \sim \quad (6)$$

$$\sim P(\tau|[\text{Fe}/\text{H}]) \cdot P(M_{\text{init}}) \cdot P(s, l, b) \cdot P([\text{Me}/\text{H}], [\text{Fe}/\text{H}]).$$

$P([\text{Me}/\text{H}], [\text{Fe}/\text{H}])$ is a fixed relation between metallicity (required for the isochrones) and the iron abundance, which we have to introduce, since we do not measure detailed abundances in this paper. The adopted relation is given in Sec. 8.2 of the Appendix. For $P(M_i)$ we employ a Salpeter IMF (Salpeter 1955), with exponent -2.35 and independent from metallicity and age. For simplicity, we separate the distance prior $P(s, l, b)$ as a purely geometric term defined by Galactic structure. This avoids the danger of merely recovering the distance-metallicity dependencies invoked by that prior. However, we have to account for the metallicity-dependent age distribution by adopting a shorter timescale in the star formation history of metal-poor stars. Details are given in the Appendix.

3.3 Stellar models P_{mod}

Stellar models describe an effectively three-dimensional constraint in the full parameter space. The corresponding PDF can be represented in the core parameter space $R_c = (T_{\text{eff}}, \log(g), [\text{Fe}/\text{H}])$. In this work, we neglect other dimensions, adopting a simple relationship for alpha enhancement and neglecting initial/present stellar rotation.

The calculation of the PDF is performed by summing the weights of available stellar model points falling into the cells of a dense grid in the target space R_c , folded by a Gaussian kernel with widths $\sigma_c = (30 \text{ K}, 0.04 \text{ dex}, 0.02 \text{ dex})$ in $(T_{\text{eff}}, \log(g), [\text{Fe}/\text{H}])$. The error represents the internal uncertainty of the stellar models and fills gaps caused by the discrete data representation. In addition, the effective width is augmented by the grid spacing on which the PDF is calculated.

We use a dense grid of stellar isochrones from the BASTI database (Pietrinferni et al. 2004, 2006, 2009), kindly provided to us by S. Cassisi for the stellar parameter determinations in Casagrande et al. (2011). For details on the colour calibrations for SDSS colours, see Marconi et al. (2006). We interpolate the models in the initial mass M_{init} to ensure a narrow mass spacing, but do not attempt an interpolation in age τ or metallicity $[\text{Fe}/\text{H}]$. When summing over the isochrones, we assign to each point i a weight W_i proportional to the parameter space volume it represents:

$$W_i = N_W \Delta[\text{Fe}/\text{H}] \cdot \Delta\tau \cdot \Delta M_{\text{init}}, \quad (7)$$

where N_W is the normalisation, $\Delta M_{\text{init}} = 0.5(M_{\text{init},i+1} - M_{\text{init},i-1})$ is the average distance to its neighbours in initial mass, $\Delta[\text{Fe}/\text{H}]$ is the average distance in metallicity between the isochrone and its nearest neighbours, and $\Delta\tau$ is the average distance in age τ . On the boundary of the grid we take the distance to the neighbouring point. Note that the approach is identical for stellar tracks instead of isochrones. The model probability at each point \mathbf{X} in our parameter space can then be represented as a weighted sum over all relevant models points i :

$$P_{\text{mod}}(\mathbf{X}) = \sum_i W_i g((\mathbf{X} - \mathbf{X}_i), \sigma), \quad (8)$$

where \mathbf{X}_i is the vector in parameter space given by the model grid. Here we represent the uncertainty of the models by an n -dimensional Gaussian g with a dispersion vector σ . Specifically, we assume σ_c on our core parameter space as above, and no additional uncertainties in the other dimensions.

3.4 Photometric data P'_{ph}

Stellar models couple photometric colours with other parameters. Thus we best calculate the photometric PDF simultaneously with the stellar models. Denoting the stellar model magnitudes at model point i and colour band k by $C_{i,k}$ and the photometric observation in band k with O_k , we have:

$$P_{i,\text{ph}}(O_{\text{ph}}|\mathbf{C}, s, r) = \prod_k P(O_k|C_{i,k}, s, r), \quad (9)$$

with distance s and reddening r . Lacking sufficient data on the true PDFs, we represent the observational likelihoods of photometric colours, the reddening values and the model uncertainties by a Gaussian $g(x - \mu, \sigma) = \exp(-(x - \mu)^2 / (2\sigma^2))$, which enables us to combine them into:

$$P(O_k|C_{i,k}, \mu(s), r) = g(C_{i,k} + \mu(s) + r \cdot \hat{e}_k - O_k, \sigma_k), \quad (10)$$

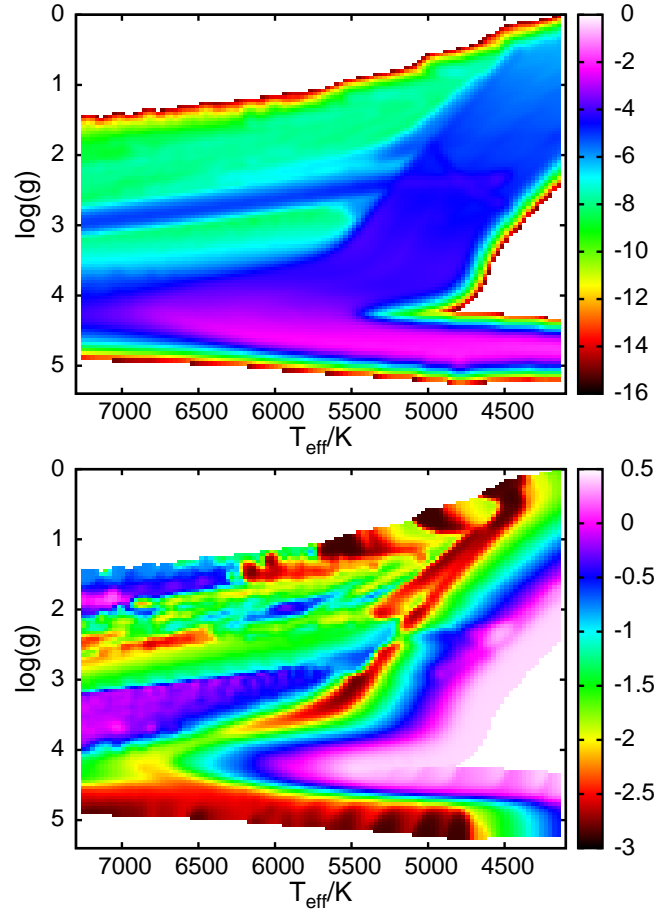


Figure 2. Prior probability densities from stellar models at any metallicity in the $(T_{\text{eff}}, \log(g))$ -plane (upper panel). We use the Salpeter IMF prior and the metallicity-dependent age prior. The lower panel shows the metallicity expectation value $\langle [\text{Fe}/\text{H}] \rangle (T_{\text{eff}}, \log(g))$ at each point in the $(T_{\text{eff}}, \log(g))$ -plane. The structures are dominated by the metallicity-dependent T_{eff} -shift and evolutionary differences on the giant branch, e.g. the absence of a horizontal branch for metal-rich stars in exchange for the enhanced red clump.

with the distance modulus $\mu(s) = 5 \log(s/10 \text{ pc})$, the reddening strength r multiplied with the reddening vector in each colour \hat{e}_k , and with $\sigma_k^2 = \sigma_{\text{mod}}^2 + \sigma_{\text{obs}}^2 + \sigma_{r \cdot e_k}^2$ as the combined variance/uncertainty of models, observations and reddening.

The other assumption is the universal reddening vector. This may have to be relaxed when dealing with very different ISM environments. The dust peak and the slope of the reddening spectrum can be shifted, or stars may be individually reddened, e.g. by a circumstellar envelope.

We note that this method can be used to create reddening maps. Since that is beyond the scope of this work, we restrict the sample to stars with relatively low reddening, use reddening values from other sources assuming a fractional reddening error of 10%.

3.5 Spectroscopic data P_{sp}

The observational likelihood P_{sp} in Eq. (5) in principle incorporates all available spectral information. This comprises spectral type, element abundances, rotation, stellar activity (chromospheric emission in cores of strong lines, magnetic effects), inter- and circum-stellar reddening, convection characteristics, etc. However, information is limited by imperfect spectra and even more by the input physics

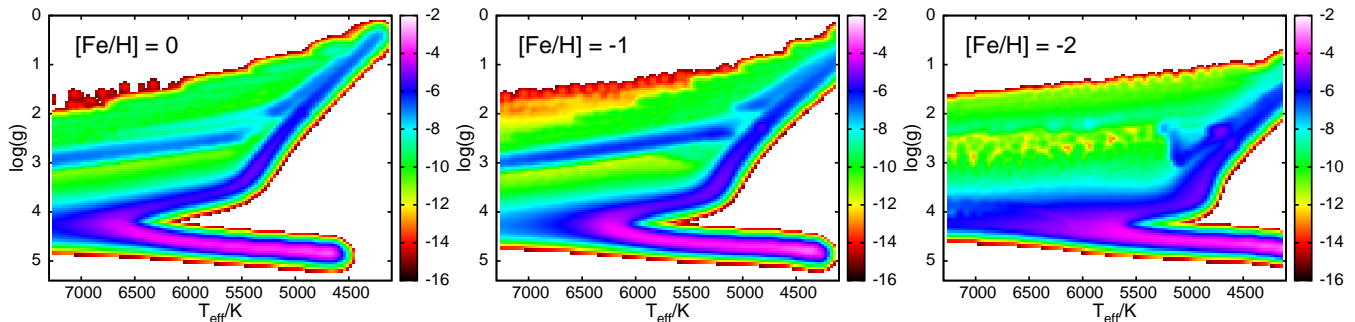


Figure 1. A priori densities from stellar models (P_{model}) at metallicities $[Fe/H] = -2, -1, 0$ from left to right and accounting for a Salpeter IMF prior, and the metallicity-dependent age prior.

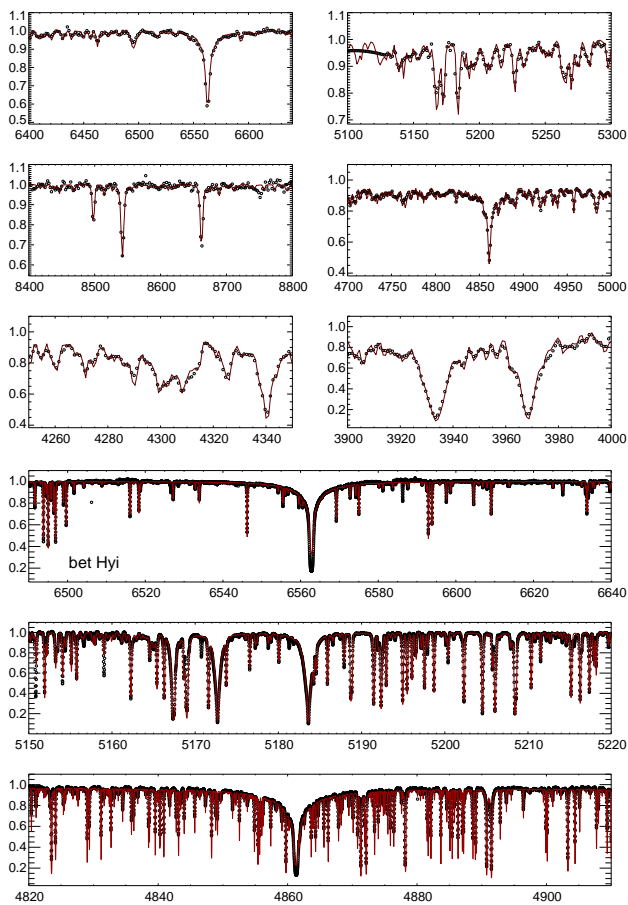


Figure 3. Top six graphs: A low-resolution spectrum from the SDSS/SEGUE survey (black, plate no. 2038, fiber 564) compared to the best-fitting theoretical model (red). Bottom three graphs: same for a high-resolution spectrum from UVES of β Hyi.

underlying the theoretical spectra. At present, calculations of large grids are only possible with 1D hydrostatic codes assuming local thermodynamical equilibrium (LTE), while full hydrodynamic 3D non-LTE calculations are slowly becoming feasible. Further, the high dimensionality of the problem forbids computing separate grids for all possible detailed abundances.

Here we use a grid designed for late-type (spectral type FGKM) stars, with $4400 < T_{eff} < 6800$ K, $1.4 < \log(g) < 4.6$,

$-3.9 < [Fe/H] < 0.9$ as given by the original MAFAGS-ODF models:¹ the grid spacing is kept at 200 K, 0.4 dex and 0.3 dex in T_{eff} , $\log(g)$, and $[Fe/H]$ to make interpolation between the points reasonable, requiring a total of 6912 theoretical spectra. The grids also cover 4 values of micro-turbulence from 1 to 2.5 km s^{-1} . For the sake of simplicity, we adopt a micro-turbulence of 2 km/s for giants and 1 for dwarfs ($\log(g) \geq 3.5$) (Bergemann et al. 2012), and an α -enhancement of 0.4 dex for $[Fe/H] < -0.6$ (cf. Gehren et al. 2004). Note that any hard cuts on parameters will lead to anomalies in the spectroscopic PDF. The latter two assumptions are owed to the size limitations of our preliminary theoretical grid and will be obsolete in future papers, where we will fit the α -enhancement directly.

The synthetic spectral grids were computed by the updated version of the SIU code (Reetz 1999; Bergemann et al. 2012). The unique parameters are $[Me/H]$, T_{eff} , $\log(g)$. The model atmospheres are MAFAGS-ODF (Grupp 2004a,b). The 3-dimensional spectral grids cover the full HRD, as shown in Fig. 5, and sample the wavelength windows around the spectral features important for diagnostic of FGKM stars: $3850 - 4050 \text{ \AA}$ (Ca I lines), $4350 - 4450 \text{ \AA}$ (G-band, CN sensitive), $4600 - 4900 \text{ \AA}$ (H_β), $5100 - 5300 \text{ \AA}$ (Mg I triplet, main gravity diagnostics), $6400 - 6640 \text{ \AA}$ (H_α), $8400 - 8800 \text{ \AA}$ (Ca II triplet, also used in Gaia and in RAVE stellar survey).

To obtain the spectroscopic observational likelihood $P(O_{sp}|X_i)$ at each point in parameter space, we resample the synthetic spectrum to the wavelength scale and resolution of the observations and evaluate the goodness-of-fit-statistics χ^2 at each pixel i of the observed spectrum:

$$\chi^2 = \sum_{i=1}^n (O_i - S_i)^2 / \sigma^2, \quad (11)$$

where S the template comparison spectrum, O the observed spectrum, σ the weighted observational uncertainty. Noisy and uninformative regions are given less weight using special masks. The final PDF is gained by summing over all pixels within a given segment, and over all segments.

The original resolution of the synthetic grid is 500 000. Thus, the method can be potentially applied to any observed dataset, e.g., low-resolution and high-resolution spectra. For the analysis of the

¹ We note, however, that any other model grid can be easily implemented, with no requirement of symmetry or shape, since our code includes a robust interpolation scheme. Alternatively, one could perform calculations of line formation on the fly on a grid of model atmospheres. This cleaner approach is, however, still computationally too costly.

SEGUE spectra, we post-convolved the spectral grids with instrumental resolution, $R = 2000$. A typical fit to a SEGUE spectrum is shown in Fig. 3. In the high-resolution mode, we use the resolution of the UVES-instrument ($R = 50\,000$). Before evaluating the test statistics, the spectra are continuum-normalised and radial-velocity corrected by cross-correlating with the template theoretical spectra.

To limit the costly spectral comparisons, we first calculated the combined PDF from model, photometric, astrometric and prior information $P_{\text{astr,mod,ph,pr}}$. Where this likelihood is larger than a threshold value, we define a coarse grid of points, on which we calculate the spectroscopic observational likelihood $P_{\text{sp}}(O_{\text{sp}}|R)$. With this we approximate the joint posterior PDF (Equ. 5) on our parameter space and iterate two steps with increasingly fine spacing in the allowed parameter range as shown in Fig. 6.

From the spectral data points we interpolate to the finer grid used in the calculations by summing up the weighted contributions from all neighbouring points using a kernel that drops linearly out to one spectral grid spacing and as a squared exponential beyond this point.

3.6 Parallaxes and other additional data

The Gaia mission will derive parallax measurements for nearly all stars with spectroscopic information. Parallax measurements only affect the distance s (and distance modulus μ), so that it is straightforward to combine the observational likelihood from parallax measurements P'_{astr} with the photometric and model information.

In the following, we assume a Gaussian parallax error. Cromwell's rule does not apply to mathematical truths, so negative parallaxes are excluded by setting the prior to 0. This yields:

$$P'_{\text{astr}} = P_{\text{astr}}(O_{\text{astr}}|\omega) = N\Theta(\omega)g(\omega - \omega_0, \sigma_\omega), \quad (12)$$

where N is a normalisation, Θ is the Heaviside-function (1 for $\omega \geq 0$ and 0 for $\omega < 0$), $g(\omega - \omega_0, \sigma_\omega)$ is again a Gaussian distribution around the measured parallax ω_0 (which can be negative) with standard deviation σ .

It is important not to clip negative values of ω_0 : a small negative value of ω_0 has still a different information content than a large negative value. In the case of a Gaussian error distribution, the probability ratio between a smaller parallax and a larger parallax rises, the further the measurement is away from zero. Or to use an example: the likelihood ratio between having failed by 3σ and by 4σ is larger than the likelihood ratio between having failed by 3σ and by 2σ .² Fig. 4 demonstrates how the parallax distributions get more concentrated towards zero, the more negative the measured value is.

To combine P'_{astr} it with the photometric and model PDF, we integrate over the possible distance moduli μ , at each stellar model point i :

$$P_{i,\text{astr,ph}} = \int P_{i,\text{ph}}(O_{\text{ph}}|\mathbf{C}, \mu(s), r)P_{\text{astr}}(\mu)J_{\mu/\omega}d\mu, \quad (13)$$

with the Jacobian $J_{\mu/\omega} = 20 \ln(10)10^{-0.2\mu}$.

In this work, P_{astr} is considered for the stars with high-resolution spectra only (see Sec. 4.3.1).

² This would only not be true if the error distribution gives constant likelihood ratios for identical distances from the measurement value, i.e. for a declining single exponential.

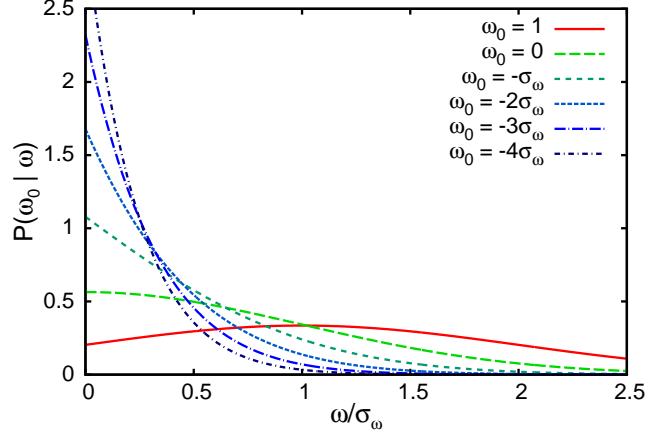


Figure 4. Conditional likelihood of the observation of $\omega_0 = 0, -\sigma_\omega, \dots$ under the real parallax ω , assuming a Gaussian error distribution with standard deviation σ_ω . The distributions become significantly narrower around the 0 for more negative parallax values, reducing the parallax expectation value and raising the expected distance.

3.7 Combining the PDFs

Equipped with these results, we can now assemble the combined PDF in equation 5. In simple words, the strategy is separate all PDFs into PDFs on the core parameter space $R_c = (T_{\text{eff}}, \log(g), [\text{Me}/\text{H}])$ and the conditional PDFs on the remaining parameter space given that point in R_c . Depending on our needs we can then represent those remaining parameter estimates either as simple moments (expectation value, variance, etc.) at each point in R_c , or as full distributions.

Formalising this is a bit tedious, since it involves a conditional probability derived from a conditional probability. To simplify the notation, we use the previous abbreviation of observational dependence with a prime. The combined calculation of photometric and model part yields:

$$P'_{\text{astr,mod,ph,pr}} = P'_{\text{ph}} \cdot P'_{\text{mod}} \cdot P'_{\text{pr}} = P'_{\text{astr,mod,ph,pr}}(\mathbf{X}_c)P'_{\text{astr,mod,ph,pr}}(\mathbf{a}_{\text{ph}}|\mathbf{X}_c), \quad (14)$$

where \mathbf{X}_c is the vector of parameters in our core parameter space R_c and \mathbf{a}_{ph} is the vector of remaining parameters constrained by the photometric and astrometric observations, models and priors, i.e. $\mathbf{a}_{\text{ph}} = (M_i, \tau, \mathbf{C}, r, s, \dots)$. Similarly, we separate the spectroscopic information:

$$P'_{\text{sp}} = P'_{\text{sp}}(\mathbf{X}_c)P'_{\text{sp}}(\mathbf{a}_{\text{sp}}|\mathbf{X}_c), \quad (15)$$

where \mathbf{a}_{sp} denotes the all other parameters constrained by spectroscopic observations, like detailed abundances, or stellar rotation. In this work we do not use this supplementary information, so that we can drop the term $P'_{\text{sp}}(\mathbf{a}_{\text{sp}}|\mathbf{X}_c)$. Most of the parameters in \mathbf{a}_{sp} will not coincide with the parameters in \mathbf{a}_{ph} , but if they correspond, they must be written into the core parameter space. For example, rotation and stellar activity available from high-quality spectra constrain stellar ages. We will discuss this in a future work.

We can now calculate the final probability distribution function:

$$P(\mathbf{X}|\mathbf{O}) = P'_{\text{astr,mod,ph,pr,sp}}(\mathbf{X}_c)P'_{\text{sp}}(\mathbf{a}_{\text{sp}}|\mathbf{X}_c)P'_{\text{astr,mod,ph,pr}}(\mathbf{a}_{\text{ph}}|\mathbf{X}_c), \quad (16)$$

where

$$P'_{\text{astr,mod,ph,pr,sp}}(\mathbf{X}_c) = P'_{\text{sp}}(\mathbf{X}_c)P'_{\text{astr,mod,ph,pr}}(\mathbf{X}_c). \quad (17)$$

3.8 Calculating projections, central values, and uncertainties

We can gain the conditional probability distribution in a lower number of parameters by marginalising, i.e. by integrating out the other dimensions in the joint conditional probability distribution function. E.g., to exclude the parameter x_{j+1} , we write:

$$P(X_1, \dots, X_j, X_{j+2}, \dots, X_n | \mathbf{O}) = \int P(X_1, \dots, X_n | \mathbf{O}) dx_{j+1} \quad (18)$$

$$P(X_j | \mathbf{O}) = \int \int P(X_1, \dots, X_n | \mathbf{O}) dx_1 \dots dx_{j-1} dx_{j+1} \dots dx_n \quad (19)$$

From this we can obtain the moments of the probability distribution in each variable or group of variables:

$$\langle X_j \rangle = \int x_j P(X_j | \mathbf{O}) dx_j \quad (20)$$

$$\langle X_j^2 \rangle = \int x_j^2 P(X_j | \mathbf{O}) dx_j \quad (21)$$

$$(22)$$

where $\langle X_j \rangle$ denotes the expectation value of the parameter X_j , and the standard deviation $\sigma_{X_j} = \sqrt{\langle X_j^2 \rangle - \langle X_j \rangle^2}$.

3.9 Short recipe of the algorithm

In short the steps are as follows:

- 1) Combine photometric and astrometric information together with the priors and sum over all stellar model points to obtain a preliminary PDF P_{prel} in core parameter space, calculate moments or full PDFs for the remaining dimensions.
- 2) In regions of parameter space, where the probability is larger than a threshold value³, calculate a coarse grid of spectroscopic probabilities and approximate the PDF P_{sp} by interpolation.
- 3) Multiply P_{prel} with P_{sp} to obtain an approximate posterior PDF P . Determine a refined grid in parameter space to better sample the spectroscopic PDF and iterate steps 2) + 3) (Fig. 6).

The threshold value on a binned PDF was chosen as $t \ll 1/(N \cdot M)$ where N is the number of stars and M is the effective number of bins, because in a large sample we have to expect the presence of rare objects, which will have low preliminary probabilities. As a different condition one can formulate that the integral of the PDF over parameter space must be

$$N \cdot \int_{P(\mathbf{X}) < t} P(\mathbf{X}) d\mathbf{X} \ll 1. \quad (23)$$

3.10 Selected examples

To illustrate the algorithm, we describe here the results for two stars. One is β Hyi from our high resolution data sample, for which we have basic Johnson photometry, high resolution spectroscopy and a Hipparcos parallax. The other star, randomly selected from the SEGUE data sample (plate number 2038 and fiber number 564), is a turn-off subgiant. In this case we have SDSS photometry and a low-resolution spectrum from SEGUE.

The resulting probability distribution functions in the $(T_{\text{eff}}, \log g)$ -plane are shown in Fig. 7 and in Fig. 8. The top left

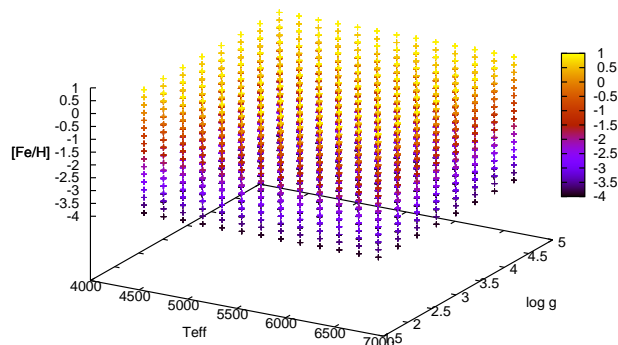


Figure 5. The grid of synthetic spectra.

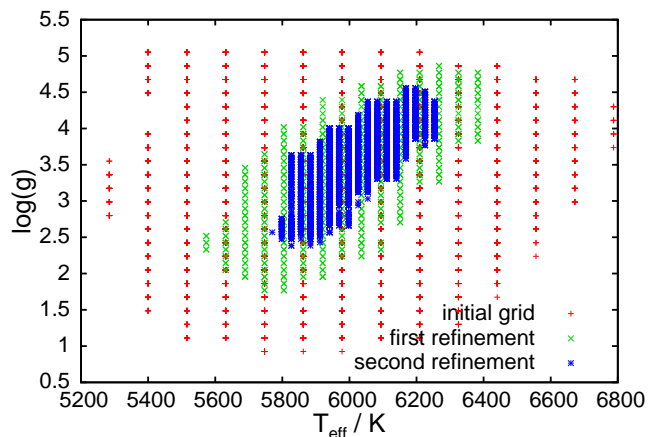


Figure 6. The process of iterative grid refinement in our pipeline for our example SEGUE star (plate 2038, fiber 564). Most points in the $(T_{\text{eff}}, \log(g))$ -plane have several metallicity values.

panel shows the combined PDF from photometry, prior and stellar evolution; the top right panel shows the spectroscopic PDF in photometrically allowed space. These two estimates combine to the final posterior PDF in the bottom left panel. The corresponding metallicities are shown by colour coding in the bottom right panel. The individual probability densities from photometry-stellar evolution and spectroscopy are clearly different in shape and in location.

The Hipparcos parallax combined with photometry and stellar models puts tight constraints on the surface gravity of β Hyi in Fig. 7. This leads also to a tight correlation between metallicity and gravity as evident from the coloured dots in the bottom right panel. A moderate step in the spectroscopic PDF at $\log(g) \sim 3.5$ is produced by a step in micro-turbulence in our current grid of theoretical spectra, which will disappear with the improved grids in preparation. The calculation does not cover the full allowed region of the spectroscopic PDF (see the coarse behaviour at smaller gravities in the top right panel), saving computation time, since the joint PDF (lower left panel) is fully represented. The final expectation values and uncertainties are $T_{\text{eff}} = (5837 \pm 72)$ K, $\log(g) = (3.981 \pm 0.068)$ dex, and $[\text{Fe}/\text{H}] = (-0.196 \pm 0.074)$ dex versus $T_{\text{eff}} = (5873 \pm 38)$ K, $\log(g) = (3.98 \pm 0.02)$ dex, and

³ The threshold should be sufficiently small to ensure coverage of the final PDF. Here we use a generous 10^{-30} per bin.

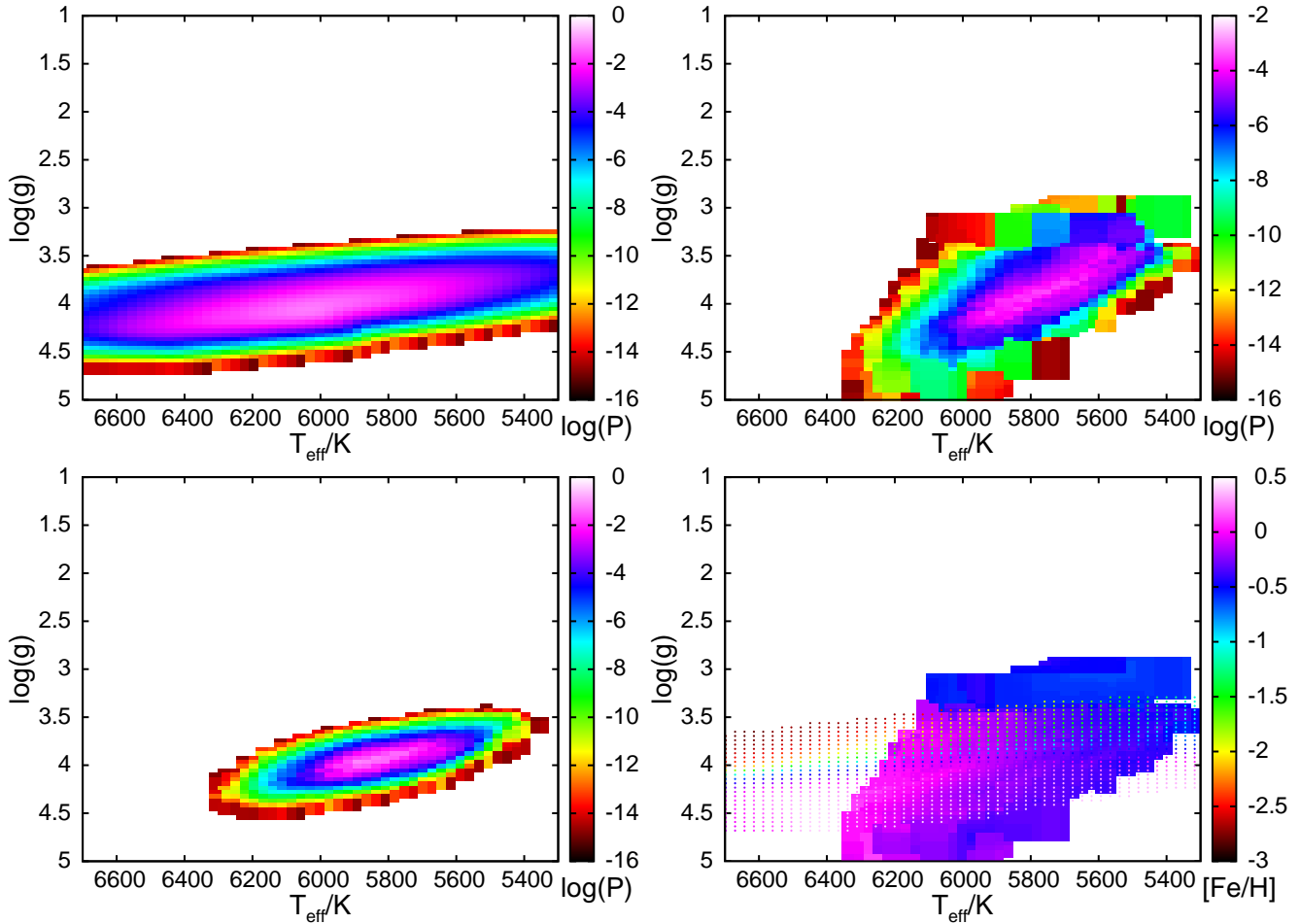


Figure 7. The PDF's for β Hvi from the high-res sample. Top left: the combined PDF from photometry-prior-stellar evolution projected onto $(T_{\text{eff}}, \log g)$ -space; top right: the spectroscopic PDF in photometrically allowed space. These two estimates combine to the overall PDF in the bottom left panel. The bottom right: metallicity expectation values from spectroscopy (coloured area) and the photometric-model part (colored dots).

$[\text{Fe}/\text{H}] = (-0.08 \pm 0.02)$ dex in the reference sample (described in the next Section).

While neither the photometric part nor the spectroscopic constraints are very tight for the SEGUE star in Fig. 8, the combined PDF is very well defined. This shows the benefits of solving the problem in the full parameter space R_c . While points in the $\log(g) - T_{\text{eff}}$ plane may be allowed by both derivations, the corresponding limits on the third dimension $[\text{Fe}/\text{H}]$ are in disagreement, ruling them out. These are the regions in the bottom right panel of Fig. 8, where the colours are mis-matched. To stress this point we show the one-dimensional probability distributions in T_{eff} in Fig. 9. While our parameters are nicely between the values of the SEGUE follow-up study Allende Prieto et al. (2008) and SEGUE DR9 (see Sec. 4.1), the behaviour of our PDF is more interesting: The combined PDF is not even remotely a simple combination of its two contributors. Most interestingly, the expectation value of the combined estimate is not situated between the estimates from each spectroscopy (6027 K) and photometry (6027 K), but significantly higher (6072 K). This complex behaviour can only be accounted for within a full Bayesian approach.

Our final expectation values and uncertainties for this SEGUE test star are $T_{\text{eff}} = 6072 \text{ K} \pm 44 \text{ K}$, $\log(g) = 3.86 \pm 0.14$, $[\text{Fe}/\text{H}] = -0.47 \pm 0.07$, for comparison SEGUE DR9 provides $T_{\text{eff}} = 6181 \text{ K} \pm 19 \text{ K}$, $\log(g) = 3.90 \pm 0.03$, $[\text{Fe}/\text{H}] = -0.459 \pm 0.006$. Note that we

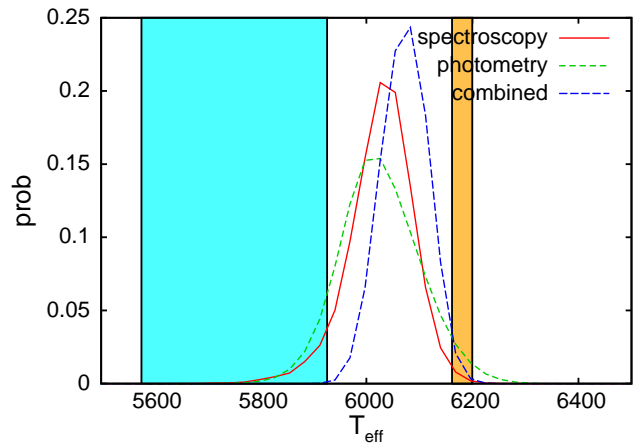


Figure 9. Projected 1D T_{eff} distributions for the discussed SEGUE star versus the 1σ intervals from Allende Prieto et al. (2008, light blue) and from SEGUE DR9 (orange). Note how the combined estimate differs from a naive expectation when looking at photometric and spectroscopic information separately.

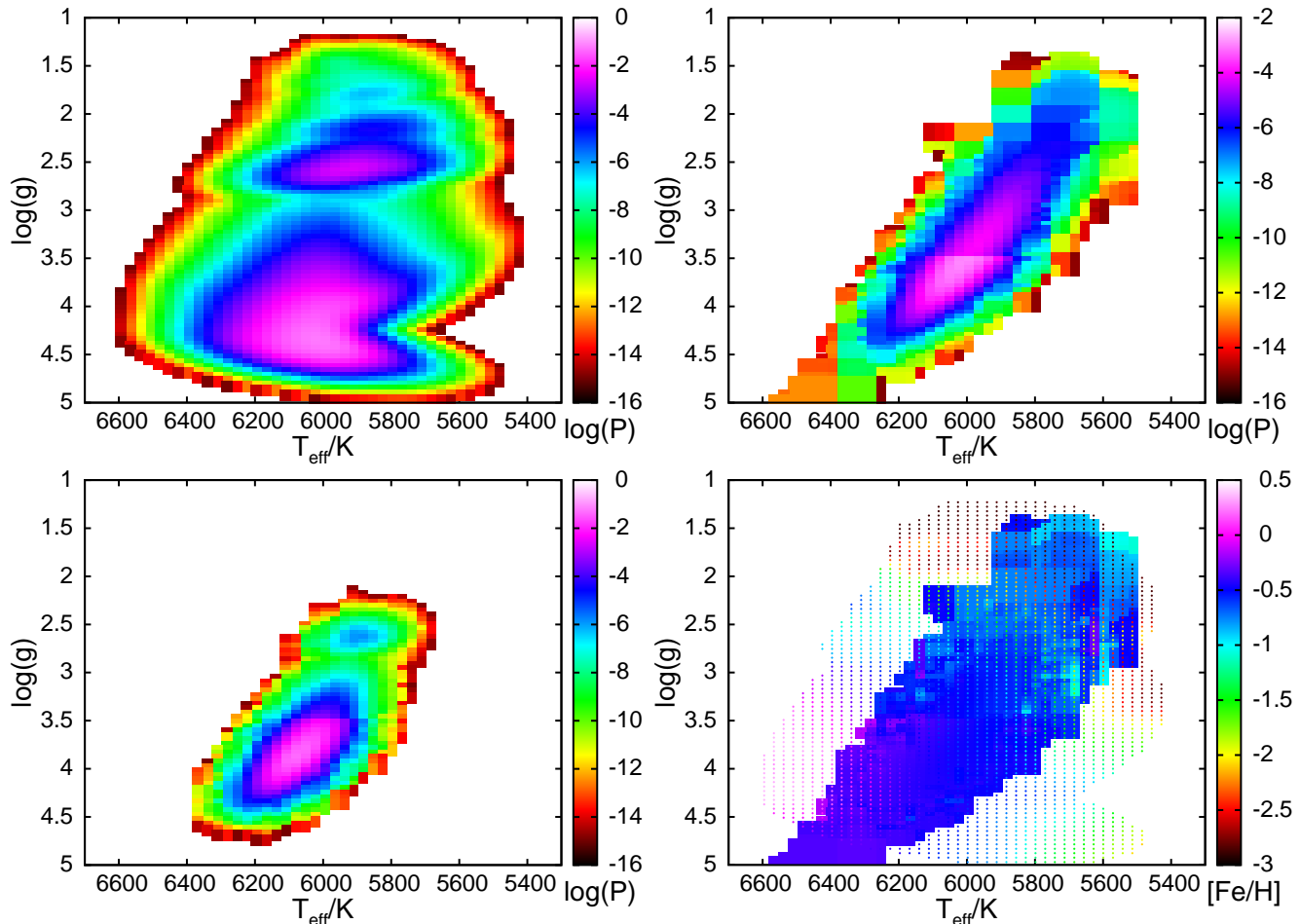


Figure 8. The PDF’s for one of the sample stars from the SEGUE sub-sample (see Sec. 3.8). Top left: the combined PDF from photometry-prior-stellar evolution projected onto $(T_{\text{eff}}, \log g)$ -space; top right: the spectroscopic PDF in photometrically allowed space. These two estimates combine to the overall PDF in the bottom left panel. The bottom right: metallicity expectation values from spectroscopy (coloured area) and the photometric/model part (coloured dots).

add the reported uncertainties just for the sake of completeness. The formally reported errors from the SEGUE pipeline cannot be considered realistic. They are severely under-estimated (by about an order of magnitude) as shown by the comparisons in Lee et al. (2008a,b) as well the discussion later in this work. The spectral fits in our six standard bands for the best spectroscopic solution are shown in Fig. 3.

This discussion also shows that even a relatively uncertain information can give an improvement to more precise values that is beyond a simple one-dimensional combination. More importantly, mismatches between different sources of information help to flag pathologies in a sample by unexpectedly small overlap of the contributing PDFs.

4 APPLICATION TO OBSERVATIONS

4.1 Datasets

Our approach is most needed and also most powerful, when different observations are available for a star and the information content is complementary but limited. With this in mind and to test the stability of our method, we choose both a sample featuring high-resolution spectra ($R \geq 40000$), as well as one with low-resolution

spectra ($R \sim 2000$) from SEGUE. For the high-resolution sample we have a combined reference dataset from Heiter et al. (in prep.), for the low resolution spectra we have two sets of comparison values: The SEGUE parameter pipeline (SSPP) values and a low/intermediate resolution follow-up by Allende Prieto et al. (2008).

The high-resolution sample comprises 87 high-resolution spectra of 40 nearby stars including the Sun, taken with the HARPS and UVES instruments at VLT, and with NARVAL at the Pic du Midi observatory. This dataset was kindly provided by P. Jofre. In addition, there are Hipparcos parallaxes (van Leeuwen 2007), making the sample closely resemble future data from Gaia astrometry combined with Gaia-ESO spectroscopy. The sample is particularly valuable because the spectra were taken on different instruments and there are independent parameter determinations, including interferometric angular diameters and astroseismic surface gravities.⁴ The stars cover a very wide range in metallicities, gravities and temperatures in parameter space (see a complete description in Blanco-Cuaresma et al. in prep). Photometry in the

⁴ The parameters were kindly provided by U. Heiter and will be published in Heiter et al.(in prep.).

U, B, V, I, J, H, K bands, was compiled from the Hipparcos catalogue (Perryman et al. 1997), from 2MASS (Skrutskie et al. 2006), and from Johnson et al. (1966). U -band photometry for HD22879 stems from Koen et al., (2010), improved JHK -photometry for ξ Hya from Laney, Joner & Pietrzyński (2012). Solar photometry was adopted from Binney & Merrifield (1998), updated with the values of Ramírez et al. (2012). We increased the errors in light of the general uncertainty of the Sun’s photometry to 0.03 mag.

Our low-resolution sample was selected from SEGUE by Allende Prieto et al. (2008), who did an intermediate-resolution follow-up.⁵ It consists of 78 stars within the parameter range $-2.5 < [\text{Fe}/\text{H}] < 0.3$, $4000 < T_{\text{eff}} < 7000$ K, and $1.5 < \log(g) < 4.5$. For these stars, we have low-resolution SEGUE spectra, ($R \sim 2000$), photometry in the SDSS $ugriz$ bands, Schlegel, Finkbeiner & Davis (1998) reddening estimates and positional data from SDSS DR9 (Ahn et al. 2012). One star was removed from the sample, as it was flagged for strongly disagreeing observational information (very low quality measure $Q < -20$, cf. equation 24).

4.2 Using photometry only

Before we turn to testing the performance of the full algorithm against our reference samples, we test it for the simpler case, where spectroscopic information is not available. For the vast majority of stars in the Galaxy, we will also after the big spectroscopic surveys just have photometric and astrometric information. However, the Bayesian method is capable of deriving stellar parameters also in this case.

In Fig. 16, we compare temperatures, gravities and metallicities derived from photometry and parallaxes only with our high resolution reference sample (top row). The bottom row shows results from photometry alone for the SEGUE sample compared to values from SEGUE DR9. While in the high resolution sample the temperatures (top row) are mildly larger than in the reference, the random scatter is as low as ~ 110 K. This is particularly remarkable, since the photometric information for most of these stars is not even remotely competitive with typical modern photometric surveys. We only have “good” values for the B, V, I Johnson bands (accurate to of order 0.03 mag), U -band information for a minority of stars. The infrared bands from 2MASS are highly uncertain with errors of about 0.3 mag, since our stars are beyond the bright limit for this survey. The excellent outcome for T_{eff} and $\log(g)$ is reasonable, however, as both gravity and temperature are indirectly constrained by the stellar models in combination with the Hipparcos parallaxes. Note that even uncertain parallaxes would in most cases be sufficient to fix $\log(g)$ and distance s , since it is sufficient to have a determination of the stellar branch. In contrast, the metallicities are only very weakly determined. Johnson colours are per se not a good diagnostic for stellar metallicities and for most stars we have no U -band magnitude. The resulting wide error bars lead to a concentration of the expectation values towards the middle of our grid. This underlines the need for photometric surveys that cover

⁵ We do not use this study as a reference, because it is very difficult to assess their accuracy and homogeneity: different parts of the sample were analysed with different methods (equivalent width method for the higher-resolution stars vs spectrum synthesis for the lower-resolution stars). For most of these stars, the spectra were degraded to $R \sim 7000$ from the original $R \sim 15000$ with unclear consequences.

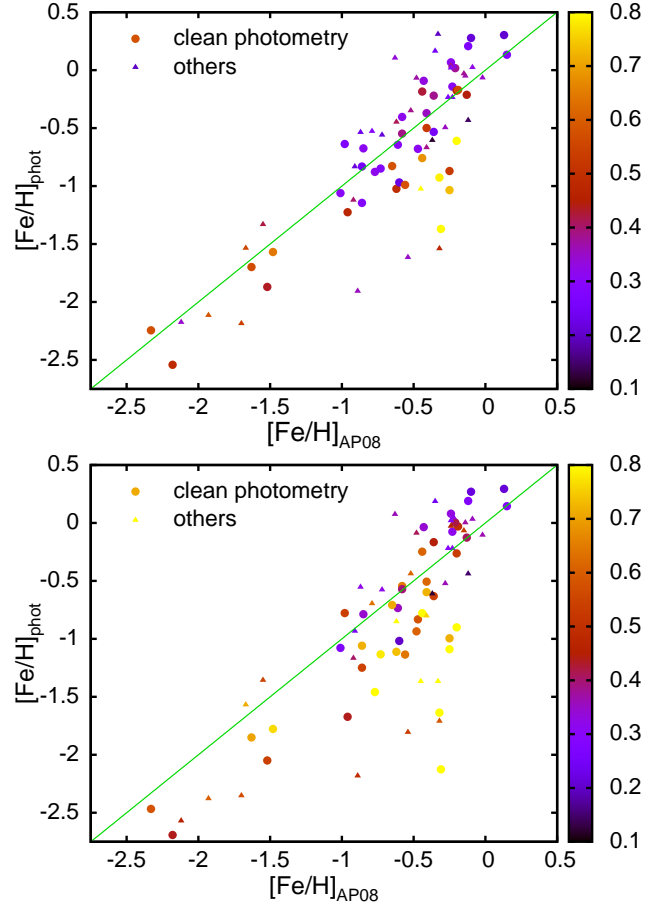


Figure 10. Photometric metallicities (expectation values of the posterior distributions) in the SEGUE/SDSS sample versus the determinations from Allende Prieto et al. (2008). Top panel includes the age prior from eq.27, while the bottom panel does not, resulting in a bias towards younger ages. Colours code the standard deviation of the posterior distributions, capped at 0.8 dex. Stars with clean photometry according to the SDSS database are depicted with discs, while stars with problematic photometry are shown as smaller triangles.

intermediate or narrow band filters capable of constraining stellar parameters.

The precise SDSS photometry and the location of SDSS colour bands allow a better handle on metallicities. In the following, we have to look more closely at the importance of our assumptions.

4.2.1 The importance of priors

In Fig. 10 we compare our photometric metallicities (y-axis) to the metallicities from Allende Prieto et al. (2008) for our SEGUE sample (y-axis), plotting stars with clean photometry with larger discs and stars with bad photometry with smaller triangles. Colours code the error estimate from the Bayesian method. Evidently, there is enough information to constrain metallicities at least in the higher metallicity range to an accuracy of about 0.2 dex. Contrary to common derivations like Ivezić et al. (2008), which fail at metallicities ≥ -0.5 (cf. Árnadóttir, Feltzing & Lundström 2010), our approach is valid throughout the entire metallicity range. However, it is important to realize how important the age prior becomes in

parameter	Comparisons to reference sample			
	$\Delta\mu$	σ	$\Delta\mu_{\text{astros.}}$	$\sigma_{\text{astros.}}$
T_{eff}/K	(65 ± 19)	141	(69 ± 13)	66
$\log(g)/\text{dex}$	(-0.024 ± 0.017)	0.13	(0.031 ± 0.009)	0.046
$[\text{Fe}/\text{H}]/\text{dex}$	(-0.099 ± 0.026)	0.19	(-0.049 ± 0.014)	0.07

Table 1. Differences in the mean expectation values of our sample minus the reference sample $\Delta\mu$ and rms scatter of σ for the entire reference sample (left two columns) and the subsample with astroseismic determinations. While the astroseismic subset has competitive accuracy and precision, the remainder of the reference sample strongly scatters against our values.

this case. In the lower panel we show the same data with a fully flat age prior instead of using eq. 27. This flat age prior implies a far larger uncertainty in the gravity of a star, which severely affects objects that cannot be clearly identified as subgiants, or main sequence stars. Via the degeneracy of u -band information, their potentially lower gravities allow for a wider range of (mostly lower) metallicities, which lowers the expectation values and boosts the error estimates. Despite this problem, the situation is far better than in the traditional approach: the classical metallicity calibrations like Ivezić et al. (2008) or An et al. (2013) rely on stars falling not only on a fixed age bin, but also onto a single evolutionary sequence. This undue neglect of surface gravity uncertainties leads to a metallicity bias and overconfidence concerning the uncertainties. In contrast, the full Bayesian approach makes optimal use of all available colour information, while it allows to explore the uncertainty in the assumptions.

4.3 Tests of the complete algorithm

4.3.1 High-resolution spectra

Currently, our spectral grids do not cover stars with $T_{\text{eff}} < 4400\text{ K}$, $\log(g) < 1.4$ and assume a slow rotation of 1 km s^{-1} , which is typical for most G and K stars (Fuhrmann 2004). Hence, for the spectroscopic comparison sample we have excluded stars with $T_{\text{eff}} > 4500\text{ K}$ and drop the fast rotating stars η Bootis and μ Leonis, which have $v \sin i \sim 15\text{ km s}^{-1}$ and $\sim 5\text{ km s}^{-1}$. We also remove ξ Hya due to contradictory results from different astroseismic derivations (Stello et al. 2006). This leaves 20 stars with 53 spectra.

As a comparison or "reference" sample we use stellar parameters from Heiter et al. (in prep.). Their effective temperatures were derived from the interferometric angular diameters or calibration relations. The gravities stem from astroseismology or Hipparcos parallaxes, and their metallicities are based on the analysis of Fe II lines, which are not affected by non-LTE effects (Bergemann et al. 2012).

Parameters from the comparison sample (x-axis) are compared to expectation values from our full Bayesian analysis (y-axis) in the top row of Fig. 11, while we give statistics for the comparison in Table 1. Colour codes the quality measure

$$Q = \log_{10} \frac{\int P_{\text{ph,mod,pr,astr}} \cdot P_{\text{sp}} d\mathbf{X}}{2 \int P_{\text{ph,mod,pr,astr}}^2 d\mathbf{X} + \int P_{\text{sp}}^2 d\mathbf{X}}, \quad (24)$$

which gives a simplified indication on how well the spectroscopic PDF agrees with the remaining information.

The two metallicity outliers at high metallicity in Fig. 11 are ϵ Vir and β Gem. Both have very high macro-turbulence values ($\sim 5\text{ km s}^{-1}$ Hekker & Meléndez 2007), which contradict the current assumptions of our spectral pipeline.

From Table 1, it is apparent that if we confine the sample to the subset with astroseismic determinations, the random mean scatter is cut by more than half. This implies that only the astroseismic subset can match or exceed our precision, while the Bayesian method is clearly superior to the traditional analysis on comparable data. Some of the reference gravities were derived from Hipparcos parallaxes, which should make them similar to our results. In this case, our fully Bayesian determinations in gravity appear more reliable than the less sophisticated reference because they also take into account physical information from colours and spectra. Interferometric T_{eff} , although they are usually taken to be mildly model-dependent, still require an estimate of limb darkening and bolometric fluxes. The former are determined with 1D LTE model atmospheres, while Chiavassa et al. (2010) showed that 3D hydrodynamical models predict different center-to-limb variation, which may cause systematic biases in angular diameter estimates. Bolometric fluxes are estimated by interpolating between observed photometric magnitudes with the help of theoretical spectra, giving rise to another systematic uncertainty.

It is very instructive to compare the full method results to the spectroscopic results. In the bottom row of Fig. 11 we show expectation values and parameter uncertainty from purely spectroscopic information (green error bars) and when using spectroscopy plus the model prior (colored points with blue error bars). Spectroscopic values alone are generally too low by (0.17 ± 0.06) dex with a residual scatter of about 0.4 dex compared to the full solution (see Ruchti 2013, for discussion of similar spectroscopic underestimates). Using the Bayesian framework (but excluding the parallax and photometric information) this greatly improves to ~ 0.3 dex residual scatter with no systematic offset at all. Hence, while spectroscopic information alone cannot compete with astrometric information, it gives sufficient information on surface gravity to allow for decent values derived by the Bayesian framework.

In Table 2 we provide a synopsis of the best available stellar parameters and ages from full the Bayesian method. When more than one spectrum is present for a star, we provide the weighted average of the expectation values and errors (we have to assume that the errors between the single determinations are highly dependent) for single spectra. Where no spectral information is available, we fill in the results from the combination of photometry and parallax measurements.

4.3.2 Comparison with SEGUE DR9

Fig. 12 shows the comparison of our parameter expectation values with the SEGUE DR9 data release.

Our temperatures are systematically colder than SEGUE DR9 by about 130 K. This is a consequence of our spectral and photometric T_{eff} scales being 160 K and respectively 60 K colder, suggesting that SEGUE DR9 overestimates stellar temperatures. The results from (Allende Prieto et al. 2008, , hereafter AP08), are on average $\sim 200\text{ K}$ below the SEGUE DR9 derivations and about 70 K below ours. The strength of our approach becomes apparent in the residual scatter of the temperature values after correcting for the systematic offset: while spectroscopic and purely photometric temperatures give a residual rms of $\sim 138\text{ K}$ and $\sim 156\text{ K}$ relative to SEGUE DR9, the full approach excels with 87 K.

The Bayesian gravities are systematically higher for suspected main sequence stars ($\log(g) > 4$), reflecting the systematic gravity underestimates of SEGUE DR9 in this range (also confirmed by SEGUE not matching expectations for the main sequence). The purely spectroscopic gravities of our method are significantly lower

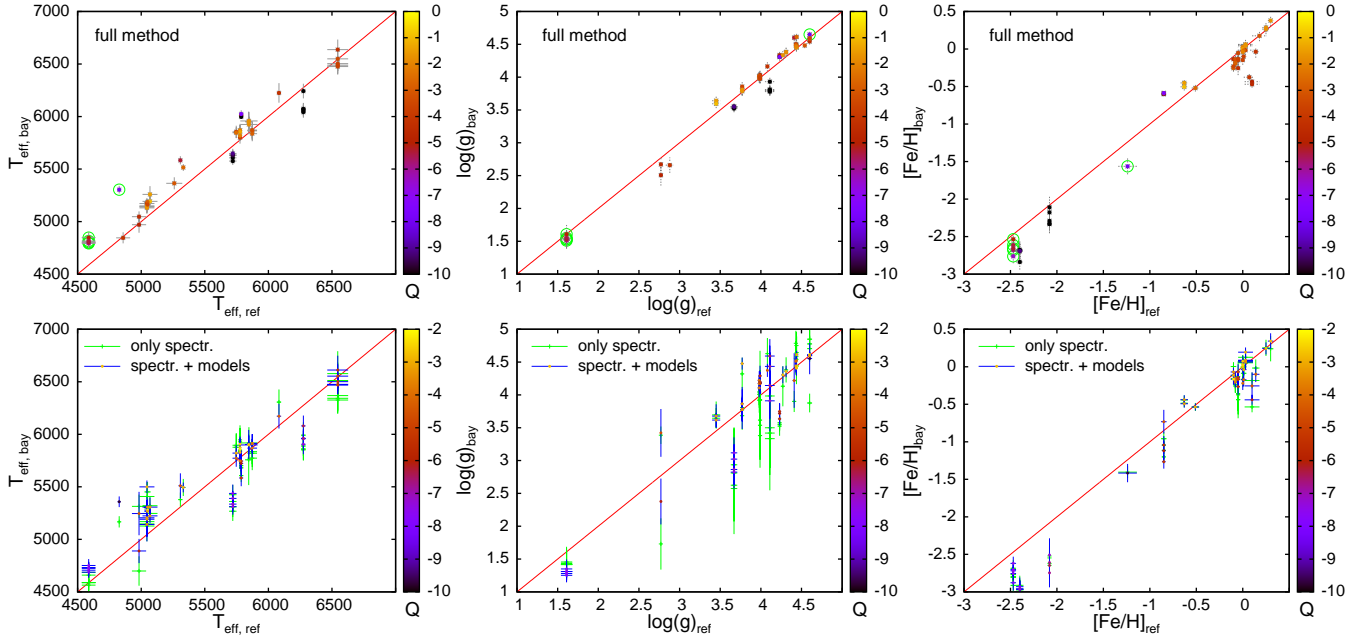


Figure 11. Expectation values from our Bayesian parameter determinations versus the reference values for stars with high-resolution spectra. In the top row we show the full Bayesian determinations using all available information, i.e. parallaxes, photometry, spectroscopy and stellar model compared to reference values from Heiter et al. (in prep.), which are derived from interferometry, asteroseismology and parallaxes. The bottom row depicts the parameter expectation values from spectroscopy alone (green error bars) and from spectroscopy + the model prior (but no photometry or astrometry).

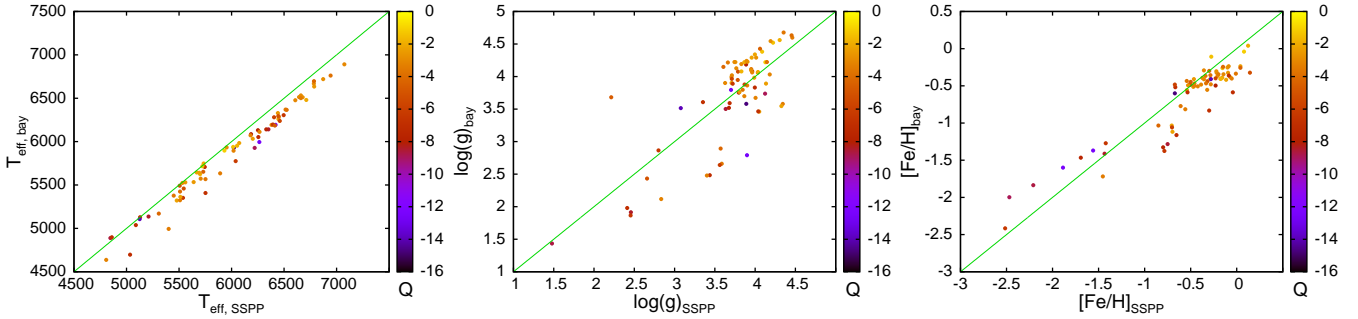


Figure 12. Bayesian expectation values for the SEGUE sub-sample compared to the results from the SEGUE parameter pipeline. The colour codes the quality measure Q , as in Fig. 11.

than DR9 and AP08 by ~ 0.5 dex in the intermediate and lower gravity range (about 0.3 dex for all stars). This is clearly identified as a bias, since the Bayesian approach reports too young ages, especially for several metal-poor stars. Though the Bayesian approach cannot completely eradicate a systematic bias in one of its inputs, it strongly reduces this problem by systematically increasing the surface gravities by an average of 0.25 dex compared to the purely spectroscopic value.

The metallicity determinations for $[\text{Fe}/\text{H}]$ between -2 and ~ -0.6 dex are robust. However, metal-rich stars have a recognizable metallicity difference between our photometric and our spectroscopic determinations, with the latter being systematically lower. For the open cluster *M67* ($[\text{Fe}/\text{H}] \sim 0.0$ or ~ 0.05 Magic et al. 2010; Gratton 2000) our spectroscopy alone gives $[\text{Fe}/\text{H}] - 0.17$ versus a photometric estimate of ~ 0.05 . As in the case of the gravities, the Bayesian method partly mitigates this problem: photometric metallicities in this range push the combined estimates towards higher values; however, due to the intrinsic uncertainty of 0.2 dex,

the corrections are minor. This also shows the importance of fair error assessment: in this case a too low error estimate from spectroscopy prevents a stronger correction of the value by the photometric information, which has intrinsic uncertainties of ~ 0.25 dex in this range. The metallicity underestimates result also in a slope against the SEGUE DR9 values at high metallicities as seen in Fig. 12. Tests show stability of our results down to a signal to noise ratio of ~ 30 and checks on the continuum setting yielded no conclusive evidence. It is very likely that a finer resolution of the synthetic spectral grid and inclusion of the α -enhancement dimension will solve the problem, for remaining differences we can use the Bayesian framework to detect and rectify remaining biases. This work is in progress and will be presented in a future paper dealing specifically with the analysis of SEGUE spectra.

The most important result is, that even at significant systematic biases, the Bayesian method itself remains robust, i.e. other parameters are not strongly affected, and the solutions are pushed towards a significantly less biased result.

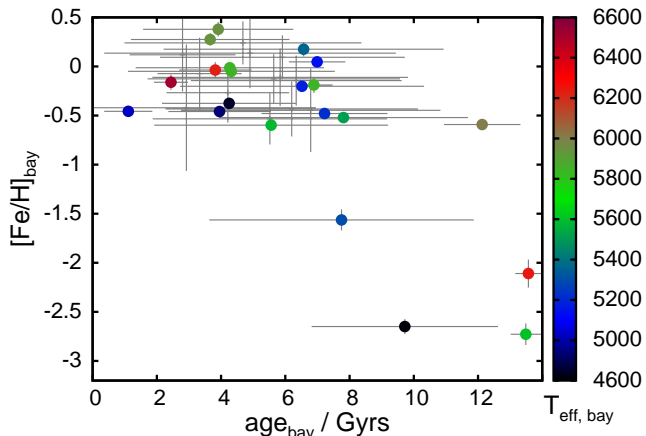


Figure 13. Ages and metallicities for the high-resolution reference sample with Bayesian stellar parameters. Colours encode the temperature estimate. For better visibility we merged the different values of each star according to Table 2.

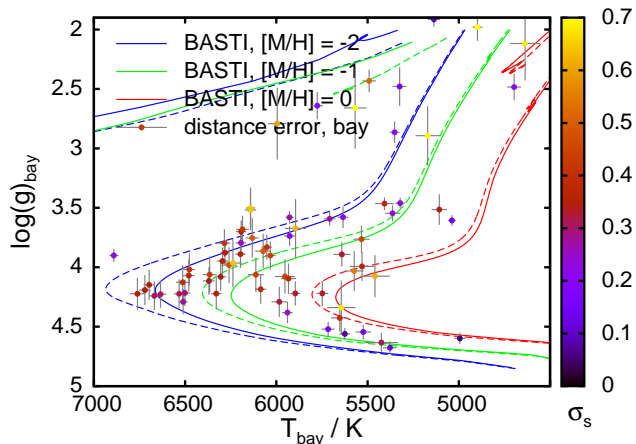


Figure 14. Distance error estimates for the Bayesian method in the temperature-gravity plane. Evidently, distance errors and classification uncertainties are not a clear function of magnitude. While clearly identified main sequence stars offer the best accuracy, even stars with high gravity estimates can have low confidence in their actual classification and distance estimates.

4.3.3 More advanced results

Inspection of Fig. 15 gives us a clue about the reliability of each parameter determination. This figure shows the HR diagrams in the $(T_{\text{eff}}, \log(g))$ -plane with expectation values from the Bayesian method (top row), versus the reference samples (bottom). On the left hand side, we show the values for SEGUE spectra, on the right hand side our analysis of the high-resolution reference sample compared to the results from Heiter et al. (in prep.). To facilitate the interpretation, we plot isochrones at 10 and 13 Gyr at three different metallicities $(-2, -1, 0)$, matching the colour scale of the stars.

The key differences between our results and that determined by conventional methods are obvious. Despite the mildly biased spectroscopic gravity estimates, our results show a clearly superior performance in this plot. The Bayesian results cover the main sequence, while SEGUE DR9 does not attain main sequence values. Even more striking is the appearance of unphysical stars: Both SEGUE DR9 and the reference sample from Heiter et al. have stars

in highly unphysical positions, with the error estimates not even close to the offset from the nearest evolutionary sequence. E.g. both SEGUE DR9 and the high-resolution reference sample place three stars around $[\text{Fe}/\text{H}] \sim -1$ far right of the turn-off or respectively right of the main-sequence. The plot suggests that the gravity offsets between the high-resolution reference values and the Bayesian method track back to a neglected metallicity effect in the reference sample. In principle the Bayesian method could yield stars in between the sequences, since we here give expectation values. A hint of this tendency can be seen, but by construction our errors will correspond to the offset, because the actual likelihood at the unphysical points is near zero.

The resulting age-metallicity relation for the high-resolution sample is displayed in Fig. 13. To make the plot easier to read, we merged the entries for different spectra as in Table 2. The picture very much resembles the results of Casagrande et al. (2011). The younger expectation values for one of the very metal-poor stars corresponds to a larger error estimate, forcing the expectation value away from the hard boundary given by the age of the universe. Further there is no striking trend in metallicity at younger ages. The importance of a reliable assessment of all stellar parameters in one single approach is demonstrated in Fig. 14. Here we plot the same stars from SEGUE as in the top left panel of Fig. 15, but now colour coded with the estimated fractional distance error. It is apparent that even some very high gravity estimates are no guarantee for a good main sequence classification, vice versa stars with lower gravity can have high distance confidence. As expected, these stars are usually cleanly identified subgiants, giants, or even better, red-clump stars. While the distance and its uncertainty are in principle enough to support estimates of mean motion and velocity dispersions in a population, we point out that an investigation of velocity distributions themselves requires accurate estimates of the exact shape of the probability distribution in distance space, which the Bayesian method can deliver.

5 DISCUSSION AND FUTURE DEVELOPMENTS

The method presented in this work is essential for accurate determination of astrophysical parameters of stars. Though the demonstrated scheme appears to be the most efficient way to extract information from the current and upcoming Galactic surveys, several shortcomings need and will be addressed.

- We are working on extending the grids of stellar spectrum models, i.e. wider wavelength coverage (UV to IR) and finer grid resolution, inclusion of α -enhancement and rotation as extra dimensions in the grids.
- Especially on the low-resolution side, the continuum finding algorithms need to be improved.
- Parameters, like micro-turbulence, which in fact parametrise the deficits of the current 1D-models in physical realism, must be better constrained or best be made obsolete by the use of more physical models. In the short and intermediate range we will find smoother corrections on a denser grid that allow for more precise evaluation. In the far future, this problem should be solved by better physics, i.e. 3D-NLTE calculations for stars, which are at present still too costly.
- It is also interesting to include age- and mass- sensitive diagnostics (such as, Ca UV lines), that would in principle allow us to choose spectroscopic models which are more appropriate in a given domain of the HRD. At present, the analysis of OBA stars relies on

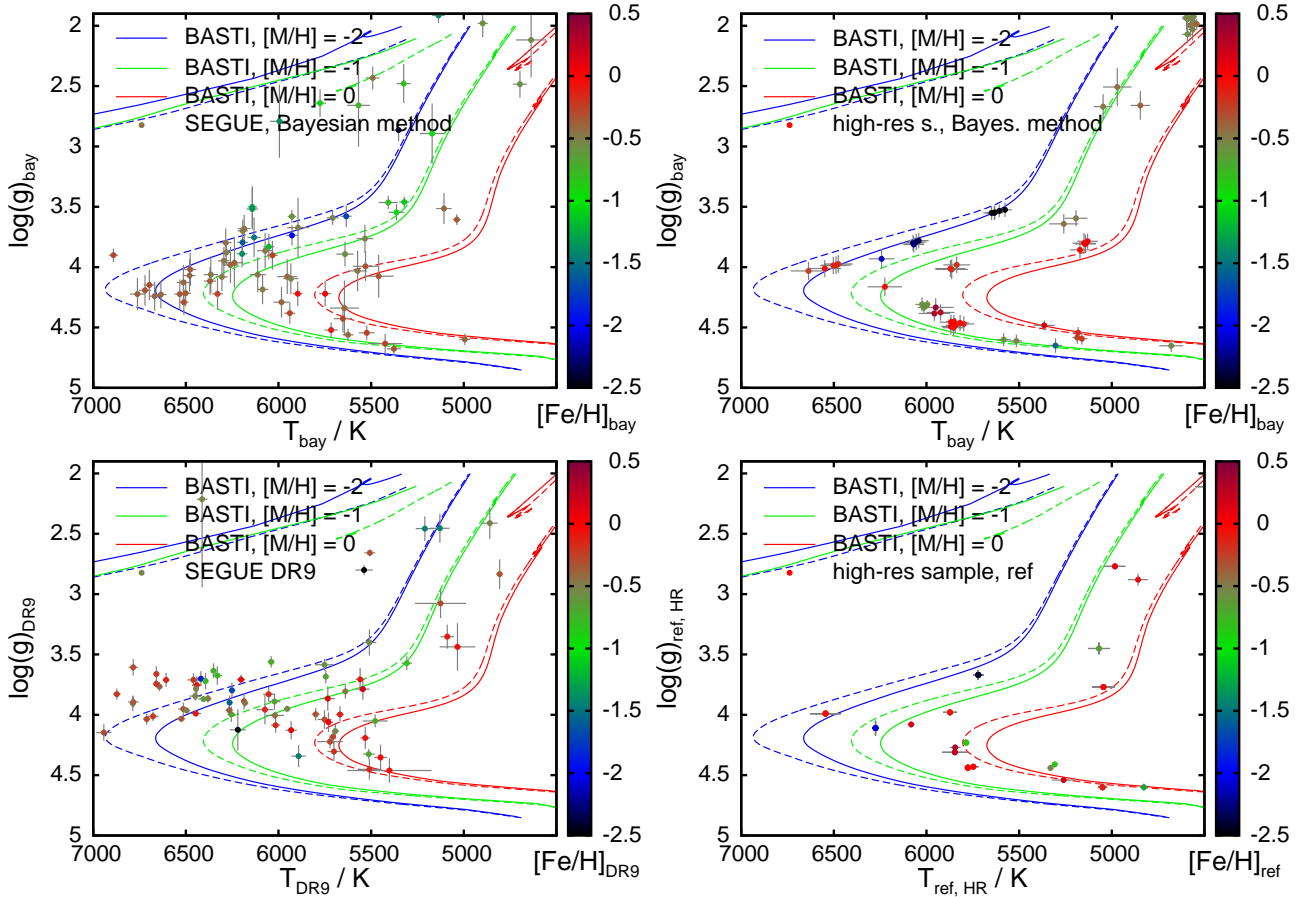


Figure 15. Expectation values from our Bayesian pipeline (top), versus SEGUE DR9 (bottom left) and the high-resolution reference sample from Heiter et al. (in prep., bottom right). Metallicities from each derivation are coded in colours, which are also used for the 10 Gyr (dashed) and 13 Gyr (solid) isochrones for metallicity $[M/H] = -2, -1, 0$. Note the disappearance of stars in unphysical positions (right of the turn-off) in the Bayesian method, which are present both in the low and the high resolution reference samples.

NLTE model atmospheres, whereas LTE models are standard for FGK stars.

- The stellar evolution models still apply rather simplistic (frequently gray) atmosphere models. Consequent systematics can be explored via residuals from this Bayesian method e.g. in magnitude space, as well as aberrations of physical parameters. A long-term goal would be to gear the stellar evolution codes with the same atmosphere models used for the spectroscopic modelling to avoid biases by partly contradictory models.

- The photometric information in our scheme is affected by reddening. Colour distortions and mismatches between the photometric and other information can be directly used to determine reddening, in addition spectral information can be extracted at high resolution e.g. from interstellar Na D lines. Since we simultaneously derive probability distribution functions for stellar distances, the method can be adapted for reddening reconstructions (like the ones by Schlafly, Green & Finkbeiner 2013).

6 CONCLUSIONS

In this paper we present the first generalised Bayesian approach for stellar parameter determination.

The essence of the Bayesian method is a combination of several probability distribution functions in the multi-dimensional pa-

rameter space, which can be expanded arbitrarily depending on a) the available observational information for a star, and b) the desired physical quantities. The presented framework simultaneously evaluates the spectroscopic informations (gained from comparisons to theoretical spectra) and all other sources of information. This allows to calculate the full probability distributions in parameter space and helps to cut computational costs by pre-constraining the parameter space that has to be searched with the spectroscopic method.

In this work we showed how to combine low or high-resolution spectroscopy, photometry, parallax measurements and reddening estimates to estimate central physical parameters $T_{\text{eff}}, \log(g), [M/H]$ of a star, as well as its mass, age, distance, or detailed chemical composition. The exploitation of theoretical constraints like stellar models, as well as strong mutual dependence or independence of different parameters reduce the complexity and effective dimensionality of the problem and make the computation possible. The scheme can be easily expanded to other sources of information, in particular to astroseismic e.g. from Corot or Kepler.

The presented method has unique advantages compared to other available approaches:

- It makes an optimal and unbiased use of all observational data and theoretical information for a star, thus providing the parameter estimates that satisfy all observational constraints;

- The method is robust with respect to missing data, such as low quality or missing spectral or photometric information.
- The method is vital to gain a grip on derived quantities. E.g. to determine the distance of a star, it is not sufficient to know its best-fit values for surface gravity, temperature, metallicity and their errors; a fair assessment is only possible if we know the full combined PDF in all parameters. We showed that indeed the Bayesian estimates in particular for uncertainties differ from simple expectations.
- Data from different surveys can be analysed with exactly the same scheme: stellar models are available in most photometric systems and the synthetic spectra grids can be folded with any instrument response function. This avoids systematic offsets caused by applying different analysis methods to different surveys and the Bayesian method can serve as a benchmark for cross-calibration between surveys.

The method requires unbiased assessments from all its sources of information. However, we know that systematic biases (e.g. theoretical atmosphere flaws, stellar evolution uncertainties like convection, nuclear reaction rates, etc.) currently affect these sources. This vulnerability can bias the entire derived parameter set. To test the performance of our method we compared both to reference samples for low-resolution and for high-resolution spectra. In all cases where we encounter problems, e.g. lower spectroscopic gravities, the Bayesian method remains robust and pushes all values towards the benchmark. Comparisons with astroseismically and traditionally derived parameters shows that the Bayesian method provides excellent results on the astroseismic sample and clearly superior performance compared to our traditionally derived reference. We provide parameter estimations for these stars in Table 2.

Similarly the photometric information is affected by reddening. However, this impact can be directly used to determine reddening especially in a larger sample. By the simultaneous determination of distance distributions, the method offers an excellent basis for reddening measurements similar to Schlafly, Green & Finkbeiner (2013).

Up to the last decade, sample sizes of Galactic surveys determined the scope of model comparisons: at sample sizes of ~ 1000 stars, Poisson noise was usually of the same importance as systematic uncertainties and knowledge of the detailed error distributions. In the future we can advance from a more qualitative understanding of best-fit parameters for our Galaxies to full quantitative analysis. This implies, however, that progress in evaluating the upcoming and present large stellar surveys for the Milky Way critically depends on our ability to cope both with the systematic biases and more importantly derive precise and accurate error distributions, and hence on the development and success of methods like the presented.

7 ACKNOWLEDGEMENTS

It is a pleasure to thank David Weinberg and Sergey Koposov for fruitful discussions and advice and James Binney for helpful comments to the text. We thank U. Heiter, P. Jofre, S. Cuaresma, T. Gehren, and F. Grupp for providing the observed high-resolution data and stellar atmosphere models used in this work. R.S. acknowledges financial support by NASA through Hubble Fellowship grant HF-51291.01 awarded by the Space Telescope Science Institute, which is operated by the Association of Universities for Research in Astronomy, Inc., for NASA, under contract NAS 5-26555. This work was partly supported by the European Union FP7 programme through ERC grant number 320360. We thank for the

great hospitality of the Aspen physics center, where parts of this paper were written.

REFERENCES

- Ahn C.P. et al., 2012, *ApJS*, 203, 21
 Allende Prieto, C., Sivarani, T., Beers, T. C., et al. 2008, *AJ*, 136, 2070
 An D. et al., 2013, *ApJ*, 763, 65
 Árnadóttir A.S., Feltzing S., Lundström I., 2010, *A&A*, 521, 40
 Aumer M., Binney J., 2009, *MNRAS*, 397, 1286
 Bergemann M., Lind K., Collet R., Magic Z., Asplund M., 2012, *MNRAS*, 427, 27
 Bevington P.R., Robertson D.K., 1992, *Data reduction and error analysis for the physical sciences*, McGraw-Hill, New York
 Binney J., Merrifield M., 1998, *Galactic Astronomy*, PUP, Princeton, NJ
 Binney J. et al., 2013, *MNRAS*, tmp.2584B, arXiv: 1309.4270
 Burnett B., Binney J., 2010, *MNRAS*, 407, 339
 Casagrande L., Schönrich R., Asplund M., Cassisi S., Ramírez I., Meléndez J., Bensby T., Feltzing S., 2011, *A&A*, 530, 138
 Chaplin W.J. et al., 2011, *ApJ*, 713, 169
 Chiavassa A, Collet R., Casagrande L., Asplund M., 2010, *A&A*, 524, 93
 Chieffi A., Straniero O., Salaris M., in *The Formation and Evolution of Star Clusters*, ed. K. Janes, ASPCS, 13, 219
 Drell P.S., Loredó T.J., Wasserman I., 2000, *ApJ*, 530, 593
 Fuhrmann, K., 2004, *AN*, 325, 3
 Gehren T., Liang Y.C., Shi J.R., Zhang H.W. Zhao G., 2004, *A&A*, 413, 1045
 Gilmore G. et al., 2012, *Msngr*, 147, 25
 Gratton R., 2000, *ASPC*, 198, 225
 Gruberbauer M., Guenther D.B., 2013, *MNRAS*, 432, 417
 Grupp F., 2004, *A&A*, 420, 289
 Grupp F., 2004, *A&A*, 426, 309
 Hekker S., Meléndez J., 2007, *A&A*, 475, 1003
 Ivezić Ž et al., 2008, *ApJ*, 684, 287
 Jofre P. et al., 2013, arXiv:1309.1099
 Johnson H.L., Mitchell R.I., Iriarte B., Wisniewski W.Z., 1966, *CoLPL*, 4, 99
 Jørgensen B., Lindegren L., 2005, *A&A*, 436, 127
 Just A., Jahrreiss H., 2007, arXiv:0706.3850
 Kitaura F.S., Enßlin T.A., 2008, *MNRAS*, 389, 497
 Koen C., Kilkenny D., van Wyk F., Marang F., 2010, *MNRAS*, 403, 1949
 Laney C.D., Joner M.D., Pietrzyński G., 2012, *MNRAS*, 419, 1637
 Lee Y.S. et al., 2008, *AJ*, 136, 2022
 Lee Y.S. et al., 2008, *AJ*, 136, 2050
 Lindley D.V., 1982, Academic Press, London, *The Bayesian approach to statistics, in: Some Recent Advances in Statistics*, Eds. J. Tiago de Oliveira and B. Epstein
 Madau P., Pozzetti L., Dickinson M., 1998, *ApJ*, 498, 106
 Magic Z., Serenelli A., Weiss A., Chaboyer B., 2010, *ApJ*, 718, 1378
 Majewski, S. R., Skrutskie, M. F., Schiavon, R. P., et al. 2007, *Bulletin of the American Astronomical Society*, 39, #132.08
 Marconi M., Cignoni M., Di Criscienzo M., Ripepi V., Castelli F., Musella I., Ruoppo A., 2006, *MNRAS*, 371, 1503
 McMillan P., 2011, *MNRAS*
 Nordström B. et al., 2004, *A&A*, 418, 989

- Perryman M., 1997, *A&A*, 323, 49
 Pietrinferni A., Cassisi S., Salaris M., Castelli F., 2004, *ApJ*, 612, 168
 Pietrinferni A., Cassisi S., Salaris M., Castelli F., 2006, *ApJ*, 642, 797
 Pietrinferni A., Cassisi S., Salaris M., Percival S., Ferguson J.W., 2009, *ApJ*, 697, 275
 Pont F., Eyer L., 2004, *MNRAS*, 351, 487
 Ramírez I. et al., 2012, *ApJ*, 752, 5
 Reetz J., 1999, *Ap&SS*, 265, 171
 Ruchti G., Bergemann M., Serenelli A., Casagrande L., Lind K., 2013, *MNRAS*, 429, 126
 Salaris M., Weiss A., 1998, *A&A*, 335, 943
 Salpeter E., 1955, *ApJ*, 121, 161
 Schlafly E., Green G., Finkbeiner D.P., 2013, *AAS*, 22114506
 Schlegel D.J., Finkbeiner D.P., Davis M., 1998, *ApJ*, 500, 525
 Schönrich R., 2012, *MNRAS*, 427, 274
 Schönrich R., Binney J., 2009, *MNRAS*, 396, 203
 Serenelli A., Bergemann M., Ruchti G., Casagrande L., 2013, *MNRAS*, 429, 3645
 Shkedy Z., Decin L., Molenberghs G., Aerts C., 2007, *MNRAS*, 377, 120
 Skrutskie et al., 2006, *AJ*, 131, 1163
 Steinmetz M. et al., 2006, *AJ*, 132, 1645
 Stello D., Kjeldsen H., Bedding T.R., Buzasi D., 2006, *A&A*, 448, 709
 van Leeuwen F., 2007, *A&A*, 474, 653
 Yanny B. et al., 2009, *AJ*, 137, 4377

8 APPENDIX

8.1 Selection function

Previous approaches (e.g., Burnett & Binney (2010)) introduced a selection function. With our choice of symbols, this would read:

$$P(\mathbf{X}|S, \mathbf{O}) = P(\mathbf{O}|\mathbf{X})P(\mathbf{X}) \frac{P(S|\mathbf{O}, \mathbf{X})}{P(S, \mathbf{O})} \quad (25)$$

where S denotes the selection function. Burnett & Binney (2010) then split the selection function into two parts: the one that depends on the parameters \mathbf{X} and the other one, that does not and is thus of no importance. However, there appears to be no reason to introduce the other term: selections of a sample are nearly almost made on observations and not on stellar parameters that are not known a priori. The one example of such a selection function acting on parameter space we could find in the literature, is actually based on a misunderstanding by Burnett & Binney (2010): Knowing the available parallax measurement and its error for a star, they try to mimic a typical kinematic quality cut in a sample by nulling all probability that produces too low parallaxes in proportion to the measured parallax error. However, it is not clear why one should not use the full parallax information here: applying the selection function implies that one has the knowledge necessary to compute the full likelihood, the selection function instead gives an undesirable one-sided constraint against far-away stars, and when pretending not to have the parallax information for testing purposes, the selection function will arbitrarily cut away the tail of effective distance overestimates, leading to wrong confidence and biased error estimates.

8.2 details on priors

For the metallicity-iron abundance prior we assume a fixed alpha enhancement. It is known that also alpha enhanced stellar models are very well approximated by scaled solar abundance models (cf. Chieffi et al. 1991; Salaris & Weiss 1998). We use this fact by setting the relation:

$$[\text{Me}/\text{H}] = \begin{cases} [\text{Fe}/\text{H}] + 0.1 & \text{if } [\text{Fe}/\text{H}] < -1.0 \\ [\text{Fe}/\text{H}] - 0.2([\text{Fe}/\text{H}] + 0.5) & \text{if } -1.0 \leq [\text{Fe}/\text{H}] < -0.5 \\ [\text{Fe}/\text{H}] & \text{if } [\text{Fe}/\text{H}] \geq -0.5 \end{cases} \quad (26)$$

The combined prior probability density of age and metallicity is used as:

$$P(\tau, [\text{Fe}/\text{H}]) = N \cdot P([\text{Fe}/\text{H}]) \cdot \begin{cases} 0 & \text{if } \tau > 14 \text{ Gyr} \\ 1 & \text{if } 11 \text{ Gyr} \leq \tau \leq 14 \text{ Gyr} \\ \exp\left(\frac{\tau - 11 \text{ Gyr}}{\sigma_\tau}\right) & \text{if } \tau < 11 \text{ Gyr} \end{cases} \quad (27)$$

where

$$\sigma_\tau = \begin{cases} 1.5 \text{ Gyr} & \text{if } [\text{Me}/\text{H}] < -0.9 \\ \left(1.5 + 9 \cdot \frac{0.9 + [\text{Fe}/\text{H}]}{0.4}\right) \text{ Gyr} & \text{if } -0.9 \leq [\text{Fe}/\text{H}] \leq -0.5 \\ 9 \text{ Gyr} & \text{else.} \end{cases} \quad (28)$$

For the sake of simplicity we give each population the same upper limit of 14 Gyr and allow for a constant density in age down to 11 Gyr. Cosmological studies as well as observations in the Milky Way disc (Madau, Pozzetti & Dickinson 1998; Aumer & Binney 2009; Schönrich & Binney 2009) measure a significant decline of star formation rates with time even for Galactic disc stars. Observations and these theoretical models also derive a significantly older age for more metal-poor populations, which motivates the decreasing time constant towards lower metallicities. The high altitude of the SDSS/SEGUE sample additionally favours older ages (cf. Just & Jahreiss 2007), but in order not to conflict with Cromwell's rule on the other hand, we lean towards a relatively moderate decline with time. In addition we use a prior on the spatial distribution of stellar densities. It cannot be justified to apply a flat distance prior to a general sample. This is founded in the specific sampling in space implied by our observations. In general every sample will cover some fixed angle on the sky (be it so-called pencil beams like in SEGUE or a complete sky coverage), so the actual volume is a cone that covers an area $A(d) = k \cdot d^2$, where d is the distance and k given by the sky coverage and selection probabilities. Well within the magnitude limits stars are hence preferentially found far away. However, the likelihood to end up in the sample is also proportional to the density of the population in the observed region, i.e.:

$$P_s(s) = k d^2 \int \int \rho(s, \omega) s^2 \omega \quad (29)$$

where we integrate over the sky position ω . SEGUE measures mostly stars in the high disc, so we describe the spatial distribution for our stars by a primitive thick disc plus halo model, i.e.:

$$\rho(R, z) = e^{-z/z_0} e^{-(R-R_0)/R_d} + 0.03 \cdot \left(\frac{r}{R_0}\right)^{-2.5} \quad (30)$$

where R is the cylindrical galactocentric radial coordinate, r the galactocentric distance, z the altitude above the plane, $z_0 = 0.9$ kpc the assumed scale height of the Galactic disc, $R_d = 2.5$ kpc the scale length of the Galactic disc, $R_0 = 8.27$ kpc the assumed galactocentric distance of the Sun from McMillan (2011); Schönrich (2012).

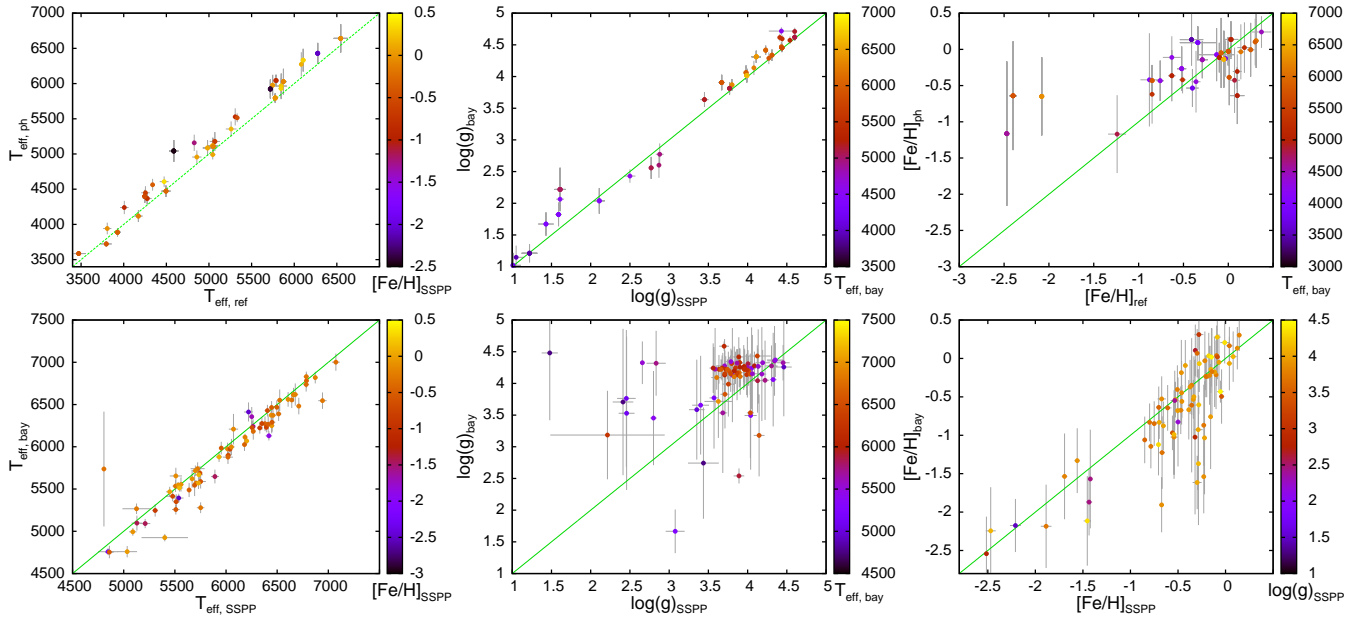


Figure 16. Parameter estimates without using spectroscopic information. The top row shows our high resolution comparison sample, the bottom row shows our comparison with the SEGUE stars. Temperatures are fine in both occasions. In the absence of parallaxes, gravity information is marginal, leading to large uncertainties. Similarly, our high resolution sample has virtually no U -band photometry, such that metallicities are weakly determined. This pushes the expectation values strongly towards the middle.

name	HIP	spectra	[Fe/H]	$\sigma_{\text{[Fe/H]}}$	T_{eff}	$\sigma_{T_{\text{eff}}}$	log(g)	$\sigma_{\text{log(g)}}$	τ	σ_{τ}	remark
HD 107238	60172	0	-0.14	0.26	4473	81	2.04	0.20	6.3	3.3	phot.
HD 122563	68594	4	-2.650	0.076	4809	47	1.54	0.13	9.7	2.9	comb., bad photometric T
HD 140283	76976	4	-2.73	0.11	5608	40	3.539	0.056	13.49	0.47	comb.
HD 173819	92202	0	-0.42	0.64	4240	95	1.05	0.25	2.9	4.0	phot
HD 190056	98842	0	-0.11	0.29	4449	94	2.07	0.17	5.8	4.0	phot
HD 220009	115227	0	-0.43	0.28	4369	85	1.67	0.19	6.2	3.9	phot
HD 22879	17147	3	-0.592	0.024*	6006	19*	4.316	0.044*	12.1	1.2*	comb. ¹
HD 84937	48152	1	-2.11	0.14	6242	70	3.931	0.082	13.57	0.41	comb. ²
ksi Hya	56343	1	-0.458	0.032*	4933	35	2.476	0.080	3.9	1.6	comb., metallicity fit questionable
Procyon	37279	4	-0.161	0.078	6515	79	3.993	0.073	2.44	0.53	comb.
alpha Cen A	71683	2	0.275	0.063	5939	79	4.380	0.066	3.7	2.5	comb.
alpha Cen B	71681	1	0.175	0.072	5364	58	4.482	0.041	6.6	4.4	comb.
Psi Phe	8837	0	0.14	0.36	3586	31*	0.65	0.22	4.9	4.5	phot. ³
Sun	0	4	-0.013	0.046	5842	49	4.464	0.063	4.3	2.9	comb.
18 Sco	79672	2	-0.050	0.059	5849	54	4.492	0.064	4.3	3.2	comb.
61 Cyg A	104214	0	-0.45	0.43	4563	83	4.717	0.060	6.8	4.0	phot.
alpha Tau	21421	0	0.09	0.22	3889	57	1.21	0.14	5.9	3.8	phot.
Arcturus	69673	0	-0.27	0.31	4399	91	1.82	0.19	4.2	1.9	phot.
alpha Cet	14135	0	-0.53	0.26	3723	41	0.50	0.16	5.5	3.7	phot.
tau Cet	8102	1	-0.520	0.047	5515	32	4.612	0.053	7.8	3.9	comb.
beta Ara	85258	0	-0.07	0.36	4118	83	1.02	0.20	3.3	1.3	phot.
mu Ara	86796	1	0.379	0.073	5950	97	4.334	0.077	3.9	2.3	comb.
Pollux	37826	1	-0.376	0.043	4846	53	2.66	0.12	4.3	2.1	comb., high macroturbulence
eps For	14086	2	-0.479	0.049	5218	68	3.614	0.072	7.2	2.0	comb.
eps Vir	63608	2	-0.457	0.050	5024	58	2.62	0.12	1.11	0.75	comb., high macroturbulence
beta Vir	57757	1	-0.037	0.080	6225	93	4.163	0.077	3.8	1.1	comb.
eta Boo	67927	0	0.12	0.37	6332	162	3.868	0.093	2.8	1.7	phot., fast rotator
delta Eri	17378	2	0.047	0.050	5139	59	3.791	0.071	6.99	0.88	comb. ⁴
eps Eri	16537	3	-0.202	0.050	5184	27	4.562	0.049	6.5	3.8	comb.
gam Sge	98337	0	-0.13	0.25	3942	83	1.15	0.19	5.6	3.9	phot.
gmb 1830	57939	1	-1.56	0.11	5304	36	4.649	0.060	7.8	4.1	comb.
mu Cas	5336	1	-0.598	0.010*	5584	39	4.601	0.053	5.6	3.6	comb. ¹
mu Leo	48455	0	0.24	0.22	4607	74	2.43	0.11	4.7	3.7	phot., low T_{eff} and rotating
beta Hyi	2021	3	-0.189	0.081	5848	79	3.997	0.074	6.89	0.58	comb.

Table 2. Parameter expectation values and errors for metallicity [Fe/H] in dex, temperature T_{eff} in K, surface gravity log(g) in dex, and age τ in Gyr, all values rounded to two significant digits in the formal error. The second column provides the Hipparcos catalogue number for each star, the third column the number of spectra involved. Stars outside the spectral grid or with bad spectra have 0 used spectra and are denoted with *phot.* in the last column, as their parameters stem from photometry, stellar models and parallax measurements, while "comb." in the last column denotes a full Bayesian approach. Detailed remarks on single stars: ¹internal rim solution by $[\alpha/\text{Fe}]$ step at -0.6 dex, metallicity and errors biased. ²UVES and HARPS spectra dropped. ³Outside model grid (rim solution). ⁴NARVAL bad spectral fit. disregarded, though Bayesian values in line with other estimates.