

# Ensemble age inversions for large spectroscopic surveys

Alexey Mints<sup>1,2,3</sup>, Saskia Hekker<sup>2,3</sup>, and Ivan Minchev<sup>1</sup>

<sup>1</sup> Leibniz-Institut für Astrophysik Potsdam (AIP), An der Sternwarte 16, 14482 Potsdam, Germany  
e-mail: [amints@aip.de](mailto:amints@aip.de), [minzastro@gmail.com](mailto:minzastro@gmail.com)

<sup>2</sup> Max Planck Institute for Solar System Research, Justus-von-Liebig-Weg 3, 37077 Göttingen, Germany

<sup>3</sup> Stellar Astrophysics Centre, Department of Physics and Astronomy, Aarhus University, Ny Munkegade 120, 8000 Aarhus C, Denmark

Received 10 May 2019 / Accepted 10 August 2019

## ABSTRACT

**Context.** Galactic astrophysics is now in the process of building a multi-dimensional map of the Galaxy. For such a map, stellar ages are an essential ingredient. Ages are measured only indirectly however, by comparing observational data with models. It is often difficult to provide a single age value for a given star, as several non-overlapping solutions are possible.

**Aims.** We aim at recovering the underlying  $\log(\text{age})$  distribution from the measured  $\log(\text{age})$  probability density function for an arbitrary set of stars.

**Methods.** We build an age inversion method, namely we represent the measured  $\log(\text{age})$  probability density function as a weighted sum of probability density functions of mono-age populations. Weights in that sum give the underlying  $\log(\text{age})$  distribution. Mono-age populations are simulated so that the distribution of stars on the  $\log g$ -[Fe/H] plane is close to that of the observed sample.

**Results.** We tested the age inversion method on simulated data, demonstrating that it is capable of properly recovering the true  $\log(\text{age})$  distribution for a large ( $N > 10^3$ ) sample of stars. The method was further applied to large public spectroscopic surveys. For RAVE-on, LAMOST and APOGEE we also applied age inversion to mono-metallicity samples, successfully recovering age-metallicity trends present in higher-precision APOGEE data and chemical evolution models.

**Conclusions.** We conclude that applying an age inversion method as presented in this work is necessary to recover the underlying age distribution of a large ( $N > 10^3$ ) set of stars. These age distributions can be used to explore age-metallicity relations, for instance.

**Key words.** methods: statistical – stars: fundamental parameters – Galaxy: stellar content

## 1. Introduction

Stellar ages and distances are important ingredients to compose a reliable picture of the Galaxy. In combination with chemical abundances and kinematics, they help us recover the history of star formation and satellite accretion for the Galaxy from its formation to the present day. Large astrometric surveys, of which *Gaia* (Perryman et al. 2001) with its more-than  $10^9$  objects is now the dominating one, provide us with kinematic information such as positions, proper motions, and parallaxes. Spectroscopic data can be used to obtain radial velocities and the physical properties of stars such as temperatures, surface gravities, and chemical compositions. Modern spectroscopic surveys provide such data for millions of stars.

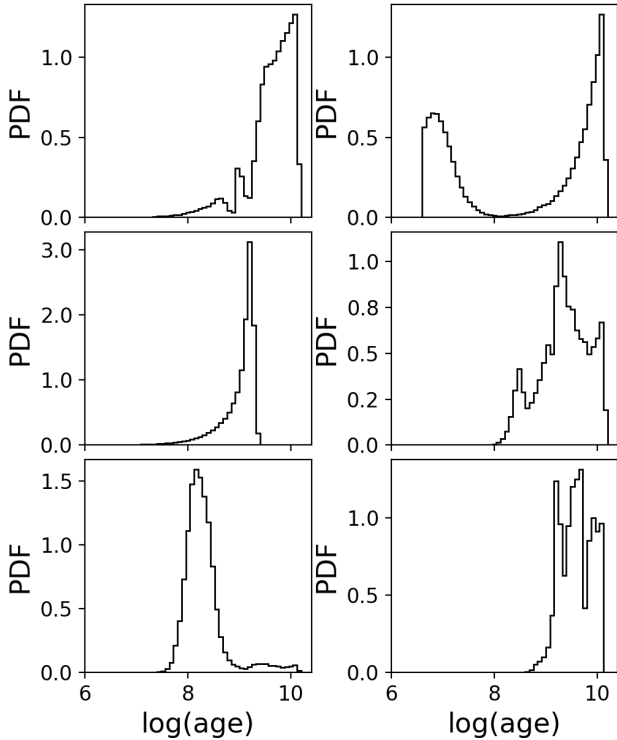
While distances, chemical abundances, and kinematics are typically contained in survey results, ages are not. This is because they are not related directly to any observable parameter and therefore estimating ages is more complicated and in most cases involves models of stellar evolution. Soderblom (2010) gives a large overview of different methods used to derive stellar ages. One of the methods listed by these latter authors, the so-called isochrone matching, can be used to derive both ages and distances, and is widely applied to spectroscopic data. The isochrone matching method is based on comparison of quantities measured spectroscopically (like the effective temperature  $T_{\text{eff}}$ , surface gravity  $\log g$ , and metallicity [Fe/H]) to a set of

models. A subset of models with parameters close to the observed ones gives an estimate of the age and luminosity of a star. The luminosity combined with visible magnitudes from photometric surveys and extinction values gives an estimate of the distance, independent of the astrometric parallax.

Recent work by Minchev et al. (2019) argues that the lack of age information in Galactic archaeology can lead to severe misinterpretation of the Milky Way formation and evolution. The authors showed that a number of chemo-kinematical relations used to study the Milky Way are plagued by a phenomenon known as Yule-Simpson's paradox, which has the effect of erasing or completely reversing the trends seen in mono-age populations when marginalizing over age (or birth radius).

In Mints & Hekker (2017; hereafter Paper 1) we introduced the implementation of an approach named Unified tool for Distance, Age, and Mass estimations (UniDAM<sup>1</sup>). This tool uses PARSEC models (Bressan et al. 2012) and infrared photometry from 2MASS (Skrutskie et al. 2006) and AllWISE (Cutri et al. 2014) surveys to produce probability density functions (PDFs) in distance,  $\log(\text{age})$ , and mass for a given star. A further extension of UniDAM that includes the use of *Gaia* parallaxes was presented in Mints & Hekker (2018) and Mints (2018). The output of UniDAM contains for a given star one or several solutions, with each solution having a unimodal PDF in  $\log(\text{age})$ , mass, and distance (labelled as unimodal sub-PDF, or USPDF). We report for each solution, along with mean, median, and mode val-

<sup>1</sup> <http://www2.mps.mpg.de/homes/mints/unidam.html>



**Fig. 1.** Examples of  $\log(\text{age})$  PDF for various stars. We note that PDFs often show non-Gaussian features: heavy tails, truncation, multiple modes.

ues, the standard deviation and confidence intervals, as well as a label indicating the type of best-fitting unimodal function used to fit the USPDF and parameters of that function. This is done in order to overcome problems that are inherent to the isochrone-matching method: non-Gaussianity and multi-modality of the produced PDFs. There are cases where age PDFs are close to being Gaussian, and thus ages can be inferred with high precision. This occurs when high-quality data are used and a specific subset of stars is analysed, as was done, for example, by [Tucci Maia et al. \(2016\)](#) with high-resolution spectra of solar twins and by [Wu et al. \(2017\)](#) with spectroscopic and asteroseismic data for main sequence turn-off stars. In a general case, as in a large spectroscopic survey, derivation of an age value from an age PDF is often more problematic. This is illustrated in Fig. 1, where we show examples of  $\log(\text{age})$  PDFs for several stars. Some  $\log(\text{age})$  PDFs are close to exponential functions, which correspond to flat PDFs in linear ages. For such stars we can put little or no constraints on their age. An important consequence of these problems is the fact that it is difficult or even impossible to provide for a given star a single age estimate that will be both accurate and precise. In the literature mean values (see, e.g. [Feuillet et al. 2016](#); [Queiroz et al. 2018](#)) or modes ([Xiang et al. 2017](#)) of PDFs were used as age proxies. The implicit assumption is that these proxies are unbiased – at least in a statistical sense. In this work, we argue that this assumption is not always valid, and we provide a way of reconstructing the age distribution of a stellar population.

## 2. Quality of probability density functions

For our task we use UniDAM results presented in [Mints & Hekker \(2017\)](#). Fits to  $\log(\text{age})$  PDFs produced by UniDAM allow us to quickly reconstruct these PDFs for any subset of the

survey. Here, we first want to show that these fits give a reliable representation of  $\log(\text{age})$  PDFs. We do that by comparing for each survey the stacked PDF for all stars with the sum of PDF fits produced by UniDAM. In this section, we analyse  $\log(\text{age})$  distributions to show that using USPDF fits from our catalogue confer an advantage over using mean values alone or mean values and uncertainties. This is illustrated in Fig. 2, which shows several representations of  $\log(\text{age})$  distribution, for every survey processed by UniDAM:

– Histogram of mean (blue line in Fig. 2), median (purple line), or mode (brown line) values of USPDFs for all stars:

$$N(\tau_b) = \sum_{i: \tau_i \in (\tau_b - \Delta, \tau_b + \Delta)} w_i, \quad (1)$$

where  $\tau_b$  are the centres of bins in  $\log(\text{age})$ ,  $\Delta$  is the bin half-width, and  $w_i$  is the weight of the USPDF  $i$ <sup>2</sup>. These representations thus retain only one parameter per USPDF, and therefore become very noisy on small stellar samples.

– Histograms of mean (orange line), median, or mode values of USPDFs that are smoothed with  $\log(\text{age})$  uncertainties as

$$N_{\text{unc}}(\tau) = \sum_i w_i \mathcal{N}(\tau | \tau_i, \sigma_{\tau,i}), \quad (2)$$

where  $\sigma_{\tau,i}$  is the uncertainty in  $\log(\text{age})$  for USPDF  $i$ ,  $\mathcal{N}(\tau | m, v)$  is the PDF of the normal distribution with mean  $m$  and standard deviation  $v$ , and the summation is done over all USPDFs for all stars in the survey. This representation is much smoother than histograms of mean, median, or mode values, although it retains only two parameters ( $\tau_i$  and  $\sigma_{\tau,i}$ ) for each USPDF. The smoothed distribution of median and mode values is not shown in Fig. 2 to avoid overloading the plot.

– Stacked fits to USPDFs (green dashed line in Fig. 2):

$$N_{\text{fit}}(\tau) = \sum_i w_i F_i(\tau, p_i), \quad (3)$$

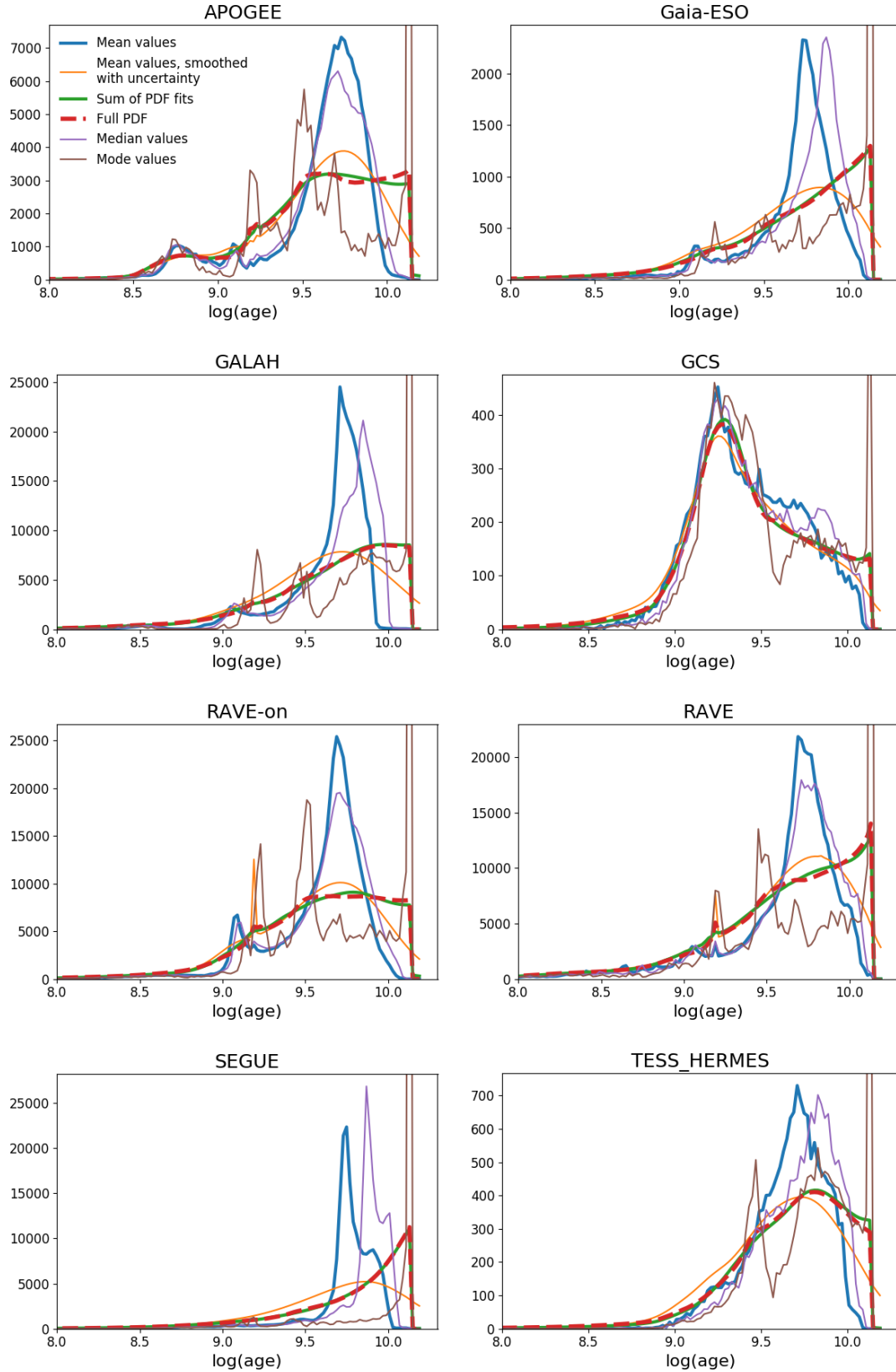
where  $F_i$  is the best-fitting function for the PDF of a solution in  $\log(\text{age})$  and  $p_i$  are the parameters of the fit. Again, the summation is done over all solutions (USPDFs) for all stars in the survey.

– “Full PDF” – the sum of all PDFs for all stars in the survey (red line in Fig. 2). This sum is produced by UniDAM by directly adding up all PDFs for each solution in the fitting process. For each solution the detailed PDFs are however not stored, as this would require much more space than already used and would be difficult to work with.

These last two representations save more detailed information about  $\log(\text{age})$  PDFs than histograms and smoothed histograms of single parameters, and hence are more helpful in recovering real underlying age distributions.

Figure 2 shows that stacked fits to USPDFs  $N_{\text{fit}}(\tau)$  (see Eq. (3)) are close to the full PDF, which confirms that fits are a reliable proxy for the full PDF. This means that for a subset of survey stars one can use stacked fits to USPDFs for stars in this subset in place of the full PDF for the same subset. The advantage of doing so is that UniDAM output allows us to reconstruct the  $\log(\text{age})$  PDF for an arbitrary subset of stars without re-running the fit and without storing full PDF data for each star, which would take about 30 times more space. For that reason, in our subsequent analysis we use  $N_{\text{fit}}(\tau)$ .

<sup>2</sup> In UniDAM, USPDF weight is the fraction of the total PDF contained in the USPDF.



**Fig. 2.** Different representations of  $\log(\text{age})$  distribution for public spectroscopic surveys. See Sect. 2 for a discussion.

From Fig. 2, it is also clear that a different representation of the age distribution has a substantially different shape for most surveys. Therefore, the choice of the representation might affect further conclusions based on the age distribution of

stars. It is however unclear which distribution is closest to the real age distribution. As we show below with the use of simulated data for which the age distribution is known, none of them are in fact a good representation of the real age distribution, so

none of them can be reliably used as an age tracer, at least not directly.

To further illustrate the problem, we build an artificial survey using stellar isochrone models with a constant age  $\tau_0$  and feed it into UniDAM. The real distribution in  $\log(\text{age})$  will be a delta function  $F(\tau) = \delta(\tau - \tau_0)$ . However, the  $\log(\text{age})$  PDF produced by UniDAM will not be a delta function. A few examples of these  $\log(\text{age})$  PDFs as produced by UniDAM are shown in Fig. 3. There we show  $\log(\text{age})$  PDFs for mono-age populations simulated in such a way as to mimic APOGEE (Majewski et al. 2017) and RAVE-on (Casey et al. 2017) surveys (see Sect. 4 for details on how the simulations were done). The strong difference in PDF shapes between APOGEE- and RAVE-on-based populations are due to the different fractions of main sequence stars in APOGEE and RAVE-on surveys: main sequence stars contribute more to the high-age part of the  $\log(\text{age})$  PDF. In most cases,  $\log(\text{age})$  PDFs have a peak near the population age, however the distributions are clearly non-Gaussian; they are very broad and in some cases show more than one peak. This is not a mistake, but an inherent property of the Bayesian isochrone fitting method used in UniDAM. In fact, almost any other isochrone fitting method will have the same property: isochrones overlap on the Hertzsprung–Russell diagram, resulting in the degenerate relation between observed spectroscopic parameters and physical properties of stars.

We can try to reconstruct the real  $\log(\text{age})$  distribution from the observed  $\log(\text{age})$  PDF for a given stellar population. To do that, we need to simulate a set of mono-age mock catalogues (MAMCs) and produce  $\log(\text{age})$  PDFs for each catalogue in this set. If we have an observed mono-age population (e.g. an open cluster), we can compare its  $\log(\text{age})$  PDFs to the  $\log(\text{age})$  PDFs for MAMCs. The age for the MAMC with the closest PDF will be the estimate of the age of the observed population. If the population consists of stars with a set of ages  $\tau_1, \tau_2, \dots, \tau_n$ , its  $\log(\text{age})$  PDF will be the sum of  $\log(\text{age})$  PDFs for MAMCs with the same ages  $\tau_1, \tau_2, \dots, \tau_n$ . We can formulate an inverse problem and try for an arbitrary observed population to represent its  $\log(\text{age})$  PDF as a linear combination of the MAMC  $\log(\text{age})$  PDFs, with coefficients as a function of  $\log(\text{age})$  giving an estimate of the true age distribution in the observed population. The details of this approach are explained below.

### 3. Inversion method

#### 3.1. Main equation

The aim of the method proposed here is to try to represent the stacked  $\log(\text{age})$  PDF or analogous function for a given set of stars with the linear combination of  $\log(\text{age})$  PDFs for mono-age mock catalogues (MAMC). The method of MAMC construction is described below in Sect. 4. The important point here is that MAMC should be constructed in such a way that the distribution of its stars in  $\log g$  and metallicity [Fe/H] is close to the distribution in these parameters for the considered set of stars. In our approach, MAMC with  $\log(\text{age})$  of  $\tilde{\tau}$  contains stars with their spectrophotometric parameters taken from PARSEC model isochrones with the same  $\log(\text{age})$   $\tilde{\tau}$ . Uncertainties were assigned to be close to observational ones (see more on how the uncertainties were chosen in Sect. 4). Because of the uncertainties in parameters, UniDAM will give for each star a  $\log(\text{age})$  PDF which is different from a delta function  $\delta(\tau - \tilde{\tau})$ . In some cases, especially for stars on the main sequence, the PDF in  $\log(\text{age})$  will not have its peak around  $\tilde{\tau}$ , but will rather be close to an exponential function with its peak at the largest value of

$\log(\text{age})$  covered by models (see Fig. 1). The exponential distribution in  $\log(\text{age})$  is equivalent to a flat distribution in linear age, which means that age is poorly constrained or unconstrained in these cases. A stacked PDF of a MAMC, similarly, does not necessarily have its maximum at  $\tilde{\tau}$ . In Fig. 3 we show PDFs for several MAMC constructed to have their  $\log g$  and [Fe/H] distributions similar to those for APOGEE (Majewski et al. 2017) and RAVE-on (Casey et al. 2017) surveys.

We designate the  $\log(\text{age})$  PDF, as produced by UniDAM for a given survey, as  $C(\tau)$ . We aim at estimating the underlying  $\log(\text{age})$  distribution of stars  $N(\tilde{\tau})$  for the same survey. For the MAMC with  $\log(\text{age})$   $\tilde{\tau}$ , we designate the  $\log(\text{age})$  PDF as  $P(\tau, \tilde{\tau})$ . If the sizes of the survey and all MAMCs are infinite, the following equation should hold:

$$C(\tau) = \int_0^{\infty} P(\tau, \tilde{\tau})N(\tilde{\tau})d\tilde{\tau}. \quad (4)$$

Thus, to obtain the unknown age distribution  $N$  from the observed PDF  $C$  we need to solve the integral Eq. (4). We note however that in reality Eq. (4) holds only approximately, because the number of stars in the survey and in each MAMC is finite, and therefore both  $C(\tau)$  and  $P(\tau, \tilde{\tau})$  are subject to stochastic deviations. These deviations limit our ability to find the solution for Eq. (4).

Let us consider as an example a case where the survey contains several mono-age populations (e.g. a set of open clusters). This implies that  $N(\tilde{\tau}) = \sum_i N_i \delta(\tilde{\tau} - \tau_i)$ , with  $N_i$  being the fractional weight of the  $i$ th population, which has  $\log(\text{age})$   $\tau_i$ . Following Eq. (4), we obtain that  $C(\tau) = \sum_i N_i P(\tau, \tau_i)$ , in other words,  $\log(\text{age})$  PDF for such a survey is a weighted sum of PDFs of mono-age populations. In fact, the PDF from UniDAM  $C(\tau)$  is defined over the grid of  $\log(\text{ages})$  ( $\tau_i, i = 1, \dots, n$ ). Similarly, MAMC PDFs  $P(\tau, \tilde{\tau})$  are defined only for  $\tau \in \tau_i$ . Thus we can substitute the function  $P(\tau, \tilde{\tau})$  with a quadratic matrix of the size  $n \times n$  and functions  $N(\tilde{\tau})$  and  $C(\tau)$  with vectors  $N(\tilde{\tau})$  and  $C(\tau)$  of the size  $n$ , and Eq. (4) can be rewritten as a system of linear equations:

$$C_j = \sum_i P_{ji}N_i. \quad (5)$$

There is an obvious non-negativity constraint for this system:  $N_i \geq 0$  for all values of  $i$ . Therefore, similarly to the example provided above, we try to represent the survey population as a superposition of  $i$  mono-age populations with ages  $\tau_i$ .

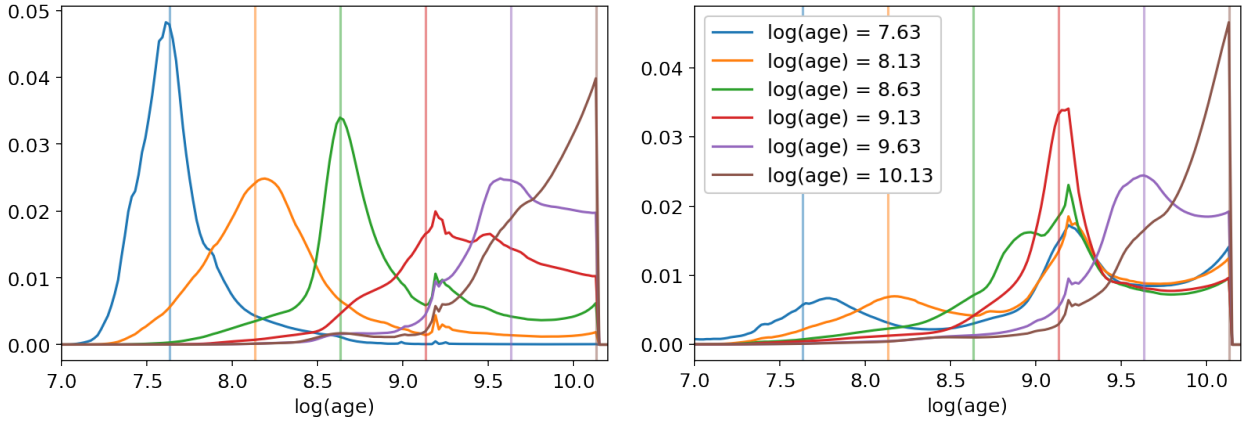
#### 3.2. Solving the main equation

In the system of linear equations (Eq. (5)),  $C_j$  are taken from the survey  $\log(\text{age})$  PDF and  $P_{ji}$  comes from MAMC  $\log(\text{age})$  PDFs. We want to solve this system of linear equations for  $N_i$ , bearing in mind the non-negativity constraint. It is possible to find a solution by means of the non-negative least squares method (NNLS; see Lawson & Hanson 1995), which maximises the following function:

$$L_0 = - \sum_{j=1}^n \left( C_j - \sum_{i=1}^n P_{ji}N_i \right)^2. \quad (6)$$

This will typically lead to a result in which only a few of the components of  $N$  will be non-zero, which is not physical – we expect a rather smooth  $\log(\text{age})$  distribution.

In order to get a smooth result we use regularized likelihood maximization. We chose a Tikhonov regularization that favours



**Fig. 3.** Examples of  $\log(\text{age})$  PDFs derived with UniDAM (same as red lines in Fig. 2) for simulated mono-age populations with  $\log g$  and  $[\text{Fe}/\text{H}]$  distributions taken from APOGEE (*left*) and RAVE-on (*right*). Vertical lines indicate the true age for each population.

smoother solutions by adding a sum of squares of the first derivatives of the solution<sup>3</sup>. We thus maximise the following function:

$$L = - \sum_{j=1}^n \left( C_j - \sum_{i=1}^n P_{ji} N_i \right)^2 - \lambda \sum_{i=1}^n \left( \frac{dN_i}{d\tau} \right)^2, \quad (7)$$

where  $\lambda$  is a regularization parameter. Solutions with narrow spikes will produce large absolute values of  $\frac{dN_i}{d\tau}$  and will have smaller values of likelihood function  $L$ , as compared to smoother solutions, even if the latter produce larger differences between  $C_j$  and  $\sum_{i=1}^n P_{ji} N_i$ .

We can use finite difference formulas to calculate  $\frac{dN_i}{d\tau}$  from  $N_i$ , and rewrite:

$$\frac{dN_i}{d\tau} = \mathcal{T} N, \quad (8)$$

where,  $\mathcal{T}$  is a Toeplitz matrix representation of the first derivative:

$$\mathcal{T} = \begin{pmatrix} -1.5 & 2. & -0.5 & 0 & \dots & \dots & \dots & 0 \\ 0 & -1.5 & 2. & -0.5 & 0 & \dots & \dots & 0 \\ 0 & 0 & -1.5 & 2. & -0.5 & 0 & \dots & 0 \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots \\ 0 & \dots & 0 & 0 & -1.5 & 2. & -0.5 & 0 \\ 0 & \dots & \dots & 0 & 0 & -1.5 & 2. & -0.5 \\ 0 & \dots & \dots & 0 & 0.5 & -2 & 1.5 & 0 \\ 0 & \dots & \dots & \dots & 0 & 0.5 & -2 & 1.5 \end{pmatrix} \quad (9)$$

In this matrix, the  $i$ th row represents coefficients for the second-order forward formula for the first derivative:

$$f'(x_i) = -\frac{3}{2}f(x_i) + 2f(x_{i+1}) - \frac{1}{2}f(x_{i+2}), \quad (10)$$

with the exception of the last two rows, which represent similar coefficients for the backward formula:

$$f'(x_i) = \frac{1}{2}f(x_{i-2}) - 2f(x_{i-1}) + \frac{3}{2}f(x_i). \quad (11)$$

Thus, Eq. (5) can be rewritten as

$$\mathbf{C}' = \mathbf{P}' \mathbf{N}, \quad (12)$$

<sup>3</sup> We performed tests with a version of Tikhonov regularization where the sum of squares of the solution values is used rather than derivatives, which resulted in very similar but slightly noisier results.

where

$$\mathbf{P}' = \begin{pmatrix} \mathbf{P} \\ \lambda \mathcal{T} \end{pmatrix} \quad (13)$$

$$\mathbf{C}' = (\mathbf{C} | 0 \dots 0)^T. \quad (14)$$

In this notation, maximization of  $L$  from Eq. (7) is equivalent to minimization of  $\|\mathbf{C}' - \mathbf{P}' \mathbf{N}\|_2$ , with a constraint that all components of  $\mathbf{N}$  are non-negative. Again, we can use NNLS to obtain the optimal vector  $\mathbf{N}$ . The solution can be characterised by the sum of residuals  $\mathbf{R} = \|\mathbf{C}' - \mathbf{P}' \mathbf{N}\|_2$ , which indicates how close the  $\log(\text{age})$  PDF for the derived age distribution is to  $\mathbf{C}(\tau)$ .

### 3.3. Searching for optimal $\lambda$ value

The problem of the regularization approach is that the parameter  $\lambda$  has to be properly chosen in order to obtain the correct solution of the problem. Two extreme cases are  $\lambda = 0$  and  $\lambda = \infty$ . In the first case no regularization is active, and Eq. (12) is reduced to Eq. (5). In that case the solution is the most precise, however it is not smooth. If  $\lambda = \infty$ , the regularization dominates the solution, such that  $N(\tau)$  becomes a constant, equal to the mean value of  $\mathbf{C}(\tau)$  – this is the smoothest possible solution, which is very imprecise. In this section we describe an empirical method for obtaining the value of  $\lambda$  that gives the result we consider to be optimal.

First of all, we note that both  $\mathbf{P}$  and  $\mathbf{C}$  are subject to statistical variations due to the limited number of sources in both the observed survey and in the simulated MAMC. Because each star can contribute to the PDF in a wide range of  $\log(\text{age})$  values, variations of the PDF at different values of  $\log(\text{age})$   $\tau$  are not independent. In order to account for this effect properly, we chose to produce five realizations of MAMC (differing only by random number algorithm seed) and to split the observed survey into five parts. This gives us five  $\mathbf{P}$  matrices  $\mathbf{P}_p, p = 1, 2, \dots, 5$  and five  $\mathbf{C}$  vectors  $\mathbf{C}_q, q = 1, 2, \dots, 5$ . Using this information, we can make use of the cross-validation technique, requiring that the solution obtained with one combination of  $p$  and  $q$  values should be good for all other combinations.

For a given value of the regularization parameter  $\lambda$  we can thus build a set of 25 equations similar to Eq. (12), with all possible combinations of  $\mathbf{P}_p$  and  $\mathbf{C}_q$ . Let us now focus on the combination with  $p = p_0$  and  $q = q_0$ . For this combination, we write Eq. (12) as:

$$\mathbf{C}'_{q_0} = \mathbf{P}'_{p_0} \mathbf{N}, \quad (15)$$

and designate a solution of this equation as  $N_{p_0, q_0}$ . We can further build a matrix of residuals with respect to all combinations of  $C$  and  $P$ :

$$R_{m,n,p_0,q_0} = \|C_n - P_m N_{p_0,q_0}\|_2. \quad (16)$$

We take the following expression as our quality parameter:

$$Q_{p_0,q_0} = \frac{1}{24} \left( \sum_p \sum_q R_{m,n,p_0,q_0} - R_{p_0,q_0,p_0,q_0} \right). \quad (17)$$

All  $R_{m,n,p_0,q_0}$  and  $Q_{p_0,q_0}$  are functions of the regularization parameter  $\lambda$ . We show an example of  $R_{m,n,p_0,q_0}(\lambda)$  and  $Q_{p_0,q_0}(\lambda)$  functions in Fig. 4.  $R_{p_0,q_0,p_0,q_0}$  decreases as  $\lambda$  decreases – with less regularization it is possible to obtain a more accurate solution of Eq. (12). For values of  $R_{m,n,p_0,q_0}$  for  $m \neq p_0$  or  $n \neq q_0$  the decreasing trend with decreasing  $\lambda$  stops or even reverses at some point. This happens because solutions  $N_{p_0,q_0}$  of Eq. (12) that are accurate for the combination  $P_{p_0}$  and  $C_{q_0}$  become too “specialised”, and are not as accurate for other combinations of  $P_p$  and  $C_q$ . As an optimal value of  $\lambda$  we take the point where the rapid decrease of  $Q_{p_0,q_0}$  (hereafter labelled simply as  $Q$ ) with decreasing  $\lambda$  stops or slows down. To find this point, we first fit the following piecewise linear function to the  $Q$ -function in log-log space:

$$Q_{\text{fit}}(\log \lambda) = \begin{cases} a + c(\log \lambda - \log b_1), & \text{if } \lambda \leq b_1 \\ a + d(\log \lambda - \log b_1), & \text{if } b_1 < \lambda \leq b_2 \\ a + d(\log b_2 - \log b_1), & \text{if } \lambda > b_2 \end{cases}. \quad (18)$$

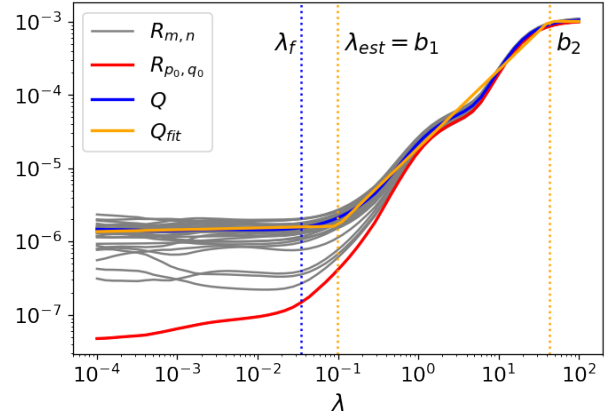
Function  $Q_{\text{fit}}$  has three linear segments, with transitions at values of  $\lambda = b_1$  and  $\lambda = b_2$ . Slopes are  $c$  in the first segment (close to zero in Fig. 4),  $d$  in the second segment and zero in the third segment. Parameter  $a$  represents the value of  $Q_{\text{fit}}(b_1)$ . This choice of the representation for the piecewise linear function is motivated by the ease of the initial guess for the fitted parameters. All five parameters ( $a, b_1, b_2, c, d$ ) are fitted simultaneously.

As the first estimate for the optimal regularization parameter value we take the first turning point of the fitted function  $\lambda_{\text{est}} = b_1$ . In many cases the turn of the  $Q$  function is not as sharp as that of the fitted function, and  $\lambda_{\text{est}}$  is an overestimation of the optimal value of  $\lambda$ . To correct for this, we take the first point to the left of  $\lambda_{\text{est}}$ , where  $Q$  and  $Q_{\text{fit}}$  intersect, as our final estimate,  $\lambda_f$ , as shown in Fig. 4. This estimate in some cases tends to produce a noisier “over-fitted” result, while the  $\lambda_{\text{est}}$  tends to produce smoother “under-fitted” one. On average, however,  $\lambda_f$  produces better results than  $\lambda_{\text{est}}$ . Our choice of  $\lambda_f$  is further validated with tests described below in Sect. 5.

We can repeat the above procedure for every possible combination of  $p_0$  and  $q_0$ , thus obtaining 25 different values of  $\lambda_f$  and corresponding solutions  $N_{p_0,q_0}$  of Eq. (15). As the final solution we can take the mean of  $N_{p_0,q_0}$ , and as a measure of the uncertainty – their standard deviation. For a more detailed uncertainty analysis see Sect. 5.3.

#### 4. Mono-age mocks construction

In order to properly recover the age distribution for a given survey, which we refer to as the base survey, we need to construct MAMCs in such a way that the distribution of stars in each MAMC in physical parameters and their uncertainties is close to that of the base survey, and at the same time retain age information. To achieve this, we follow the procedure below.

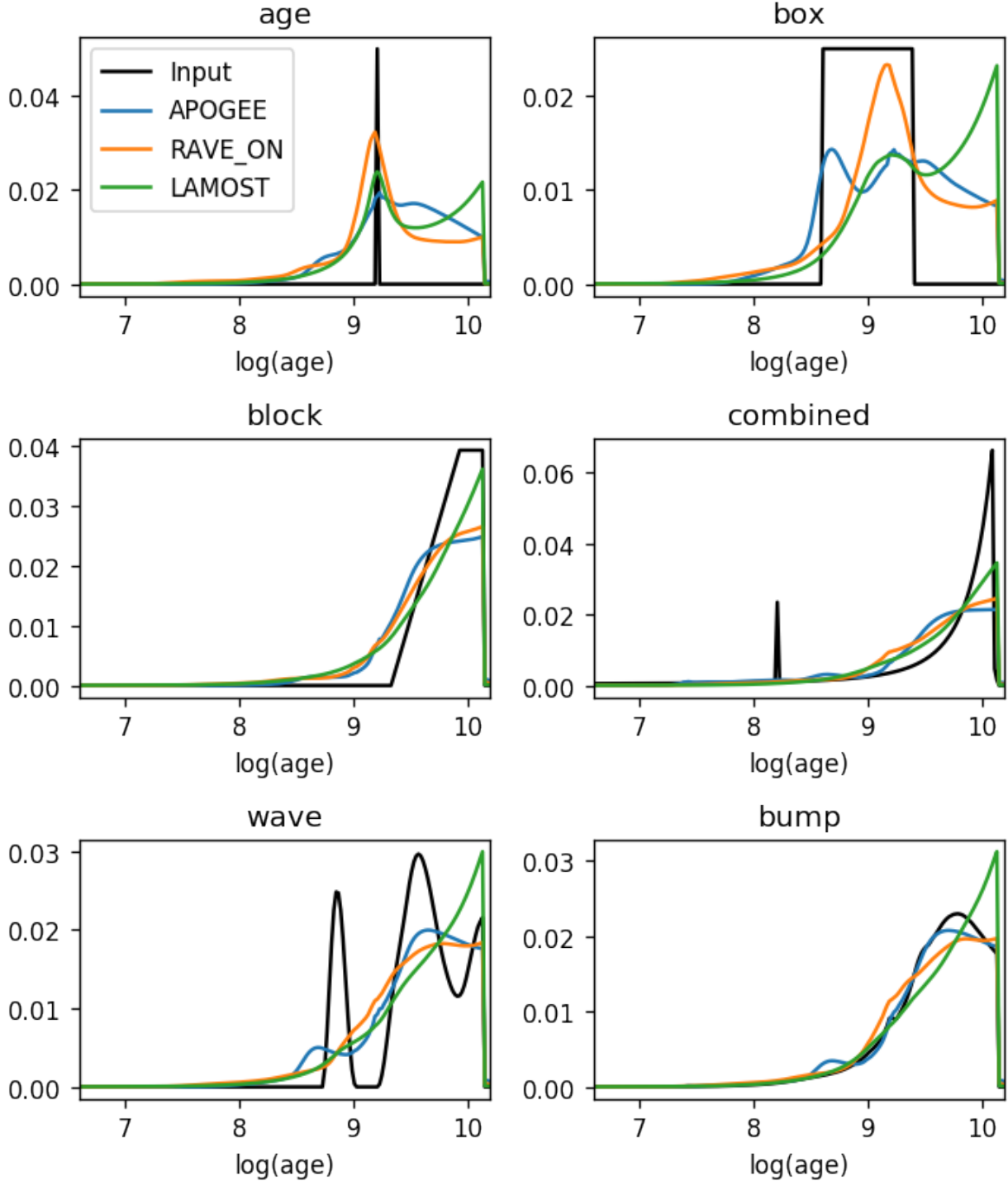


**Fig. 4.** Illustration for the choice of optimal  $\lambda$ . The plot shows all residual functions  $R_{m,n}$  (grey lines, see Eq. (16)) and highlights the  $R_{p_0,q_0}$  in red and  $Q$  (Eq. (17)) in blue. The fitted piecewise linear function  $Q_{\text{fit}}$  is shown in orange (see Eq. (18)). Vertical lines indicate the turning points of the fitted piecewise linear function  $Q_{\text{fit}}$  (orange dashed lines) and the finally adopted  $\lambda$  value at the intersection point of  $Q$  and  $Q_{\text{fit}}$  (blue dashed line).

We build MAMCs using PARSEC models (Bressan et al. 2012) to simulate stars. In order to mimic the random observational scatter, instead of taking each model, we take a sample of 25 models with random normally distributed perturbations added in  $T_{\text{eff}}$ ,  $\log g$ , and  $[\text{Fe}/\text{H}]$ . To assign a proper amplitude of perturbations, we randomly select uncertainties  $\sigma_P = \{\sigma_T, \sigma_{\log g}, \sigma_{[\text{Fe}/\text{H}]}\}$  from the base survey data for stars that have  $P = \{T_{\text{eff}}, \log g \text{ and } [\text{Fe}/\text{H}]\}$  close to the model, such that the difference between the physical parameters for the model and those for the base survey star is smaller than the mean uncertainty in the respective parameter for the base survey  $P_{\text{model}} - P_{\text{star}} < \text{mean}(\sigma_P)$ . This provides a way to reconstruct the scatter and the systematic variations of  $\sigma_P$  across the parameter space. Selected uncertainties are not only used as perturbation amplitudes, they are also assigned as “observational” uncertainties for models. The photometric magnitudes of all models are perturbed with a Gaussian noise with a scale of 0<sup>m</sup>025, to reproduce typical photometric uncertainties in 2MASS and AllWISE.

The set of models is then truncated in the  $T_{\text{eff}}, \log g, [\text{Fe}/\text{H}]$  space to the footprint of the base survey in that space, and models outside of that footprint are excluded. In order to build MAMC from that set of models, we sample a pre-defined number of models  $n_{\text{mock}}$  as follows. We bin stars from the base survey in  $\log g - [\text{Fe}/\text{H}]$  space. The fraction of models sampled for a given MAMC from each bin is equal to the fraction of base survey stars in that bin. Within each bin, models are selected randomly with probabilities proportional to the fraction of the initial mass function represented by that model.

We cannot use all three physical parameters ( $T_{\text{eff}}, \log g, [\text{Fe}/\text{H}]$ ) for binning, as in this case the resulting MAMC will be too close to the base survey itself, retaining almost no information about the MAMC age value. On the other hand, using just one parameter will produce an MAMC that will not resemble the base survey, making inversion impossible. We decided to use  $[\text{Fe}/\text{H}]$  as a first parameter for the binning, as it is the only parameter out of the three that is independent of stellar mass, and as such is expected to be the same for all stars of the same origin. The second parameter is  $\log g$ , as it helps to distinguish between main sequence stars and giants, which have very different contributions to the  $\log(\text{age})$  PDF. Such a distinction is impossible to make with  $T_{\text{eff}}$ .



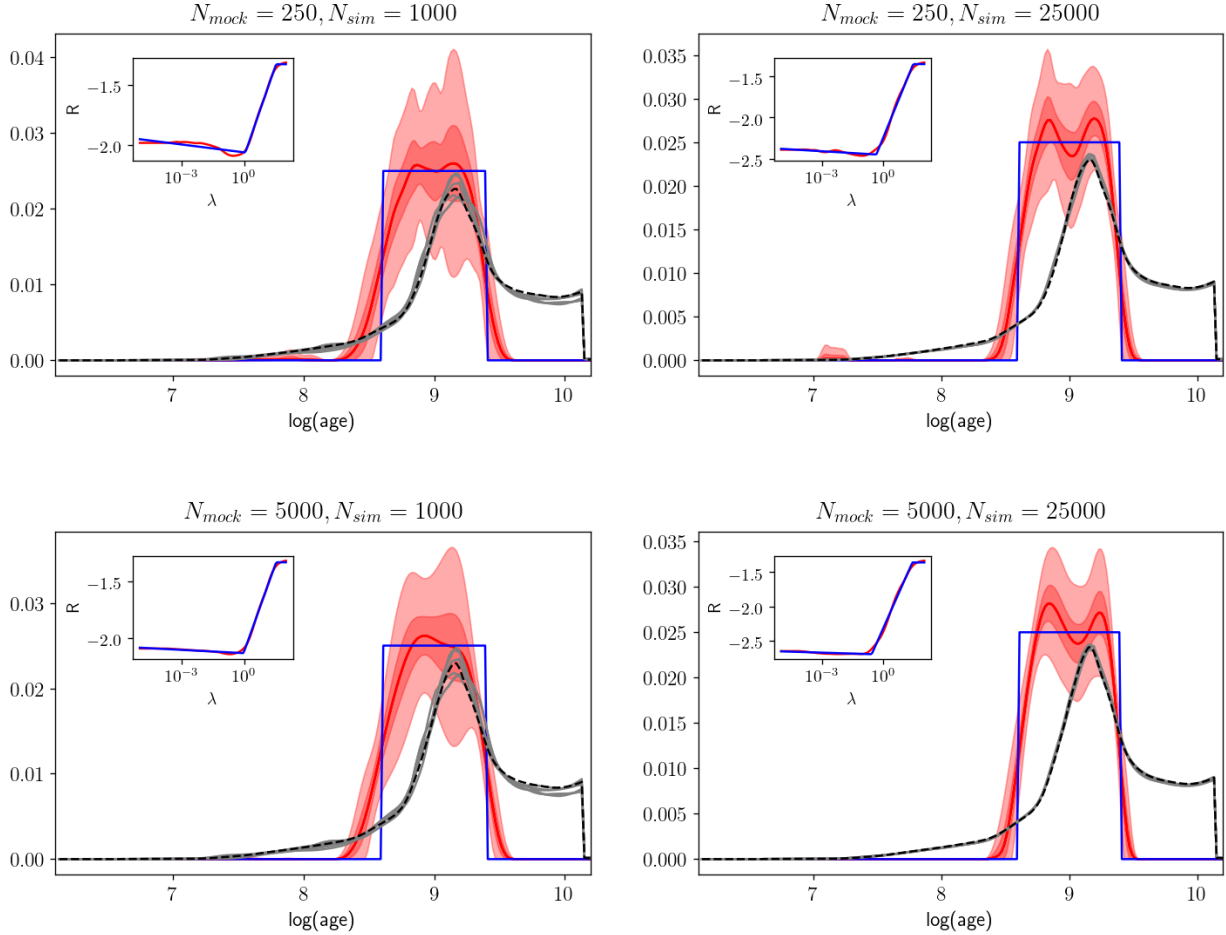
**Fig. 5.** Input log(age) distributions and log(age) PDFs produced with UniDAM for different base surveys. Plots are for six cases considered in this work. For the “age” input the true PDF was re-scaled to make other lines visible.

The above procedure is repeated for each value of log(age)  $\tau$  over the considered grid:  $6.61 \leq \tau \leq 10.13$  dex with a step of 0.02 dex or age between approximately  $4 \times 10^6$  and  $13.5 \times 10^9$  years. As a result we get a set of MAMCs that is fed into UniDAM to obtain a set of log(age) PDFs  $P_{j,i}$ .

## 5. Tests with mock data

In this section, we use simulated data to validate the choice of the optimum regularization parameter  $\lambda$  and to determine the

influence of the survey size  $n_{\text{sim}}$  and MAMC size  $n_{\text{mock}}$  on the precision and the accuracy of the inversion result. We simulate a survey with pre-defined age distribution  $N_{\text{input}}$  as a concatenation of MAMC simulated in a way described in Sect. 4, with the size of  $i$ th MAMC defined by  $N_{\text{input},i}$ . We then obtain with UniDAM the log(age) PDF for the simulated survey and apply the inversion as described in Sect. 3.2. We can then compare the result of the inversion with the input age distribution to get an estimate of the accuracy and precision of the method. Below we describe the process in detail.



**Fig. 6.** Results of the age inversion for box test data. *Upper panels* are for the smaller MAMC ( $n_{\text{mock}} = 250$ ), and *lower panels* are for the larger MAMC ( $n_{\text{mock}} = 5000$ ). *Left panels* are for the smaller simulated survey ( $n_{\text{sim}} = 1000$ ), and *right panels* are for the larger simulated survey ( $n_{\text{sim}} = 25000$ ). The solid blue line shows the input  $\log(\text{age})$  distribution, and the solid red line is the result of the inversion with 68- (dark shading) and 95 (light shading) percent confidence intervals. Input  $\log(\text{age})$  PDFs for five realizations of the simulated catalogue are plotted with grey lines and the black dashed line shows the  $\log(\text{age})$  PDF inferred from the inversion result. The inset in each panel shows the cross-validation curves (red) and piecewise-linear fit (blue) for one realization of the simulated catalogue (see Sect. 3.3 for the description of the cross-validation curve).

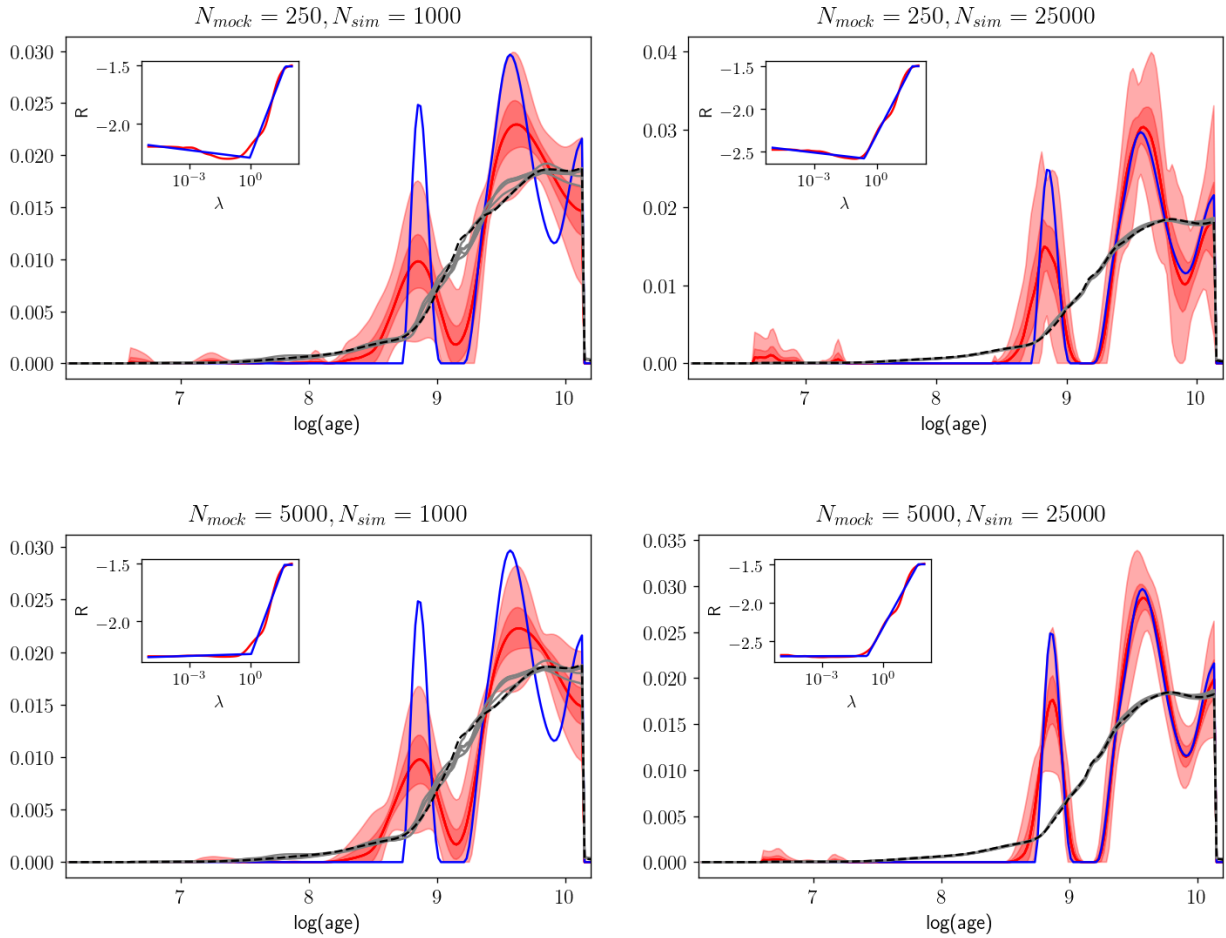
### 5.1. Input distributions

For this work we chose six input age distributions, aiming to emulate critical as well as more common cases. Every distribution was generated using one of the three base surveys: APOGEE DR14 (Majewski et al. 2017), LAMOST DR3 (Luo et al. 2015), and RAVE-on (Casey et al. 2017). In Fig. 5 we show the input  $\log(\text{age})$  distributions and PDFs produced by UniDAM for the three base surveys. “Age” and “box” inputs implement a spike and step function. Such functions are hard to reproduce with our method, as the uncertainties in age determination tend to smooth out the distribution. “Block” and “bump” inputs show the opposite case of smooth, slowly changing PDFs, and therefore we expect that the inversion will work best for them. “Combined” and “wave” inputs represent intermediate cases. Interestingly,  $\log(\text{age})$  PDFs for bump and wave simulations have very similar shapes for all three base surveys, despite the very different input  $\log(\text{age})$  distributions. We note that the choice of base survey affects the  $\log(\text{age})$  PDFs generated by UniDAM. Most importantly, the  $\log(\text{age})$  PDF is typically not close to the true age distribution, with the only exception being the APOGEE result for the bump distribution.

### 5.2. Test results

In Figs. 6 and 7 we show examples of test results, with different input distributions and varying  $n_{\text{sim}}$  and  $n_{\text{mock}}$ . There we compare input distributions (blue lines) with age inversion results (red lines), and “observed”  $\log(\text{age})$  PDFs (dashed black lines) with  $\log(\text{age})$  PDFs as predicted by age inversion (grey lines, one for each combination of simulated survey and MAMC realizations).

Figures 6 and 7 illustrate that the result of the inversion is almost insensitive to  $n_{\text{mock}}$  for small values of  $n_{\text{sim}}$ , as expected (see discussion in Sect. 5.3). On the other hand, for  $n_{\text{sim}} \geq 25000$ , MAMC size  $n_{\text{mock}}$  plays a larger role. It is also clear that an increase in  $n_{\text{sim}}$  and  $n_{\text{mock}}$  increases the sensitivity of the inversion to rapid changes in  $\log(\text{age})$  distribution. For example, in the case of the wave test,  $n_{\text{mock}} = 250$  and  $n_{\text{sim}} = 25000$  is enough to properly recover the second peak of  $\log(\text{age})$  distribution at  $\tau \approx 9.6$  dex, though we need to increase  $n_{\text{mock}}$  to 5000 in order to properly recover the first, much narrower peak at  $\tau \approx 8.9$  dex. For all simulations, observed (black lines) and predicted (grey lines)  $\log(\text{age})$  PDFs become almost indistinguishable for  $n_{\text{mock}} = 5000$  and  $n_{\text{sim}} = 25000$ .



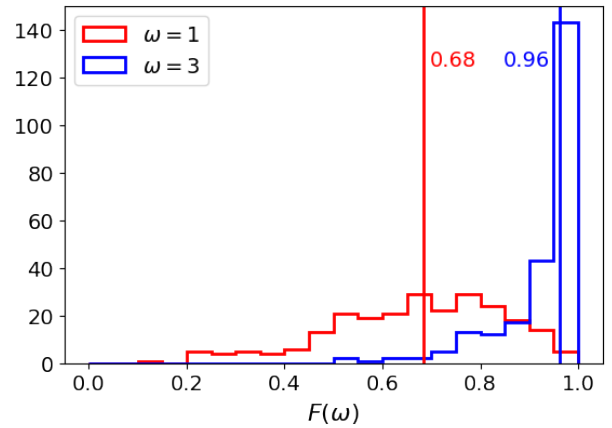
**Fig. 7.** Same as Fig. 6, but for wave input. Line styles and colours are as in Fig. 6.

### 5.3. Uncertainty analysis

An important part of every scientific result is a proper uncertainty. We can use our tests to provide a way to estimate uncertainties for age inversion results.

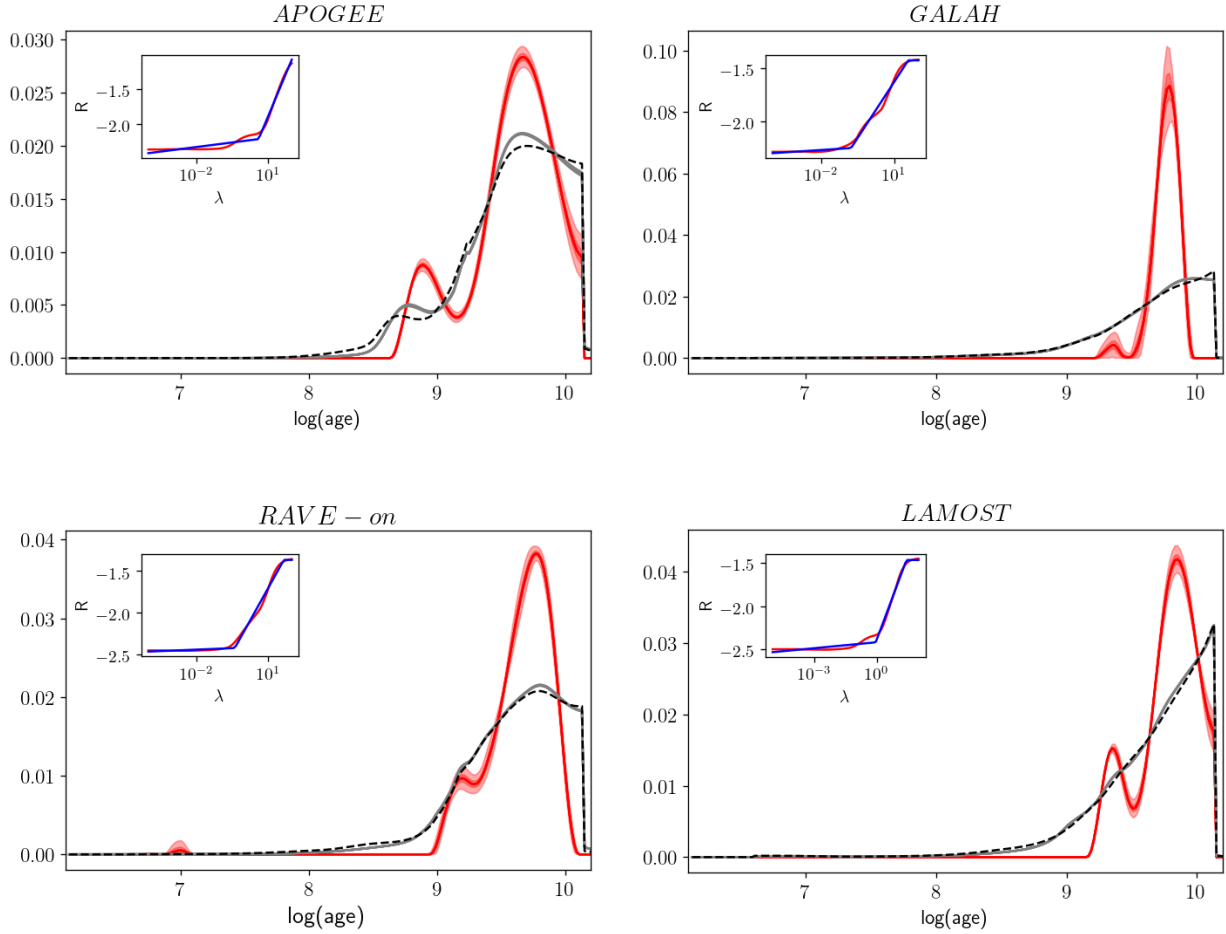
In Sect. 3.3 we give a method for obtaining the optimal value of the regularization parameter  $\lambda$ . This parameter and a corresponding solution of Eq. (12) can be obtained for every combination of the five survey  $\log(\text{age})$  PDFs  $C_q$  and five MAMC  $\log(\text{age})$  PDF sets  $P_p$ . Hence we can have 25 different solutions for Eq. (15), with the differences between them being due to statistical variations in  $P$  and  $C$ . Variations between these solutions can be used to estimate the uncertainty of the solutions. Because of the smoothness of the solution, which is imposed by the regularization, and because of the similarity between  $\log(\text{age})$  PDFs for MAMCs with similar  $\log(\text{age})$  values, differences between solutions at different values of  $\log(\text{age})$   $\tau$  are highly correlated. For these reasons, the distribution of differences between the inversion result and the true underlying age distribution will not be a Gaussian but a heavy-tailed distribution.

In the case of the tests that we are performing, we also use the knowledge of the true age distribution to find the relation between the variation between solutions  $\sigma(\tau)$  and the true error  $\delta(\tau)$  (the true error is defined as the absolute difference between input and output  $\log(\text{age})$  distributions). The fraction  $F(\omega)$  of values of  $\tau$  with  $\delta(\tau) < \omega\sigma(\tau)$  can be calculated for each result. In Fig. 8 we show distributions of  $F(\omega)$  for  $\omega = 1$  and  $\omega = 3$ . If  $\sigma(\tau)$  is a correctly determined Gaussian uncertainty, then  $F(1) = 0.68$  and  $F(3) = 0.997$ , as expected for



**Fig. 8.** Distributions of the fraction  $F(\omega)$  of values of  $\tau$  with true error  $\delta(\tau) < \omega\sigma(\tau)$ , where  $\sigma(\tau)$  is the estimated uncertainty. Shown are distributions for  $\omega = 1$  and  $\omega = 3$  for simulated data. Vertical lines and numbers represent median values of each distribution.

1- and 3- $\sigma$  confidence intervals. Measured  $F(1)$  values for all our tests show a broad distribution with a mean of 0.67 and median of 0.685 – very close to expectations. Measured  $F(3)$  values also have a broad distribution, with a mean of 0.93 and median of 0.96 – lower than the expected value, making  $3\sigma(\tau)$  effectively a 2-sigma rather than 3-sigma confidence interval estimate. This is because uncertainties have a heavy-tail non-Gaussian distribution. There are no clear trends visible for  $F(\omega)$  values with  $n_{\text{sim}}$



**Fig. 9.** Results of age inversion for several surveys. The solid red line is the computed underlying age distribution with 68- (dark shading) and 95 (light shading) percent confidence intervals. Input  $\log(\text{age})$  PDFs for five parts of each survey are plotted with grey lines and the black dashed line shows the  $\log(\text{age})$  PDF inferred from the inversion result. The inset in each panel shows the cross-validation curves (red) and piecewise-linear fit (blue) for one part of the survey (see Sect. 3.3 for the description of the cross-validation curve).

and  $n_{\text{mock}}$ , at least within the considered ranges of these parameters.

We note here that the scatter of  $F(\omega)$  around the median values is high, which means that values of  $\sigma(\tau)$  can be seen only as an approximation of the uncertainty.

One would expect the fractional uncertainty values to decrease as  $(n_{\text{sim}}n_{\text{mock}})^{-1/2}$ . In reality, their relation is not as strong due to complex correlations between the real age distribution and the corresponding  $\log(\text{age})$  PDF. Our estimate shows that the fractional uncertainty scales as  $n_{\text{sim}}^{-0.4}n_{\text{mock}}^{-0.15}$ . This means that one has to increase  $n_{\text{mock}}$  by an order of magnitude to improve the result by about 30%. Importantly, for a given  $n_{\text{mock}}$  there seems to exist a maximum value of  $n_{\text{sim}}$ , above which the fractional uncertainty does not decrease with  $n_{\text{sim}}$  at all. This is caused by the fact that in that regime, fractional uncertainty becomes dominated by variations between realizations of MAMC, and not by the survey. The opposite is also true – for a given  $n_{\text{sim}}$  there is a maximum value of  $n_{\text{mock}}$ , above which the fractional uncertainty does not decrease as it is dominated by survey PDF variations. The exact value of the maximum of  $n_{\text{mock}}$  for a given  $n_{\text{sim}}$  depends on the base survey and on the shape of the underlying age distribution and can vary by over an order of magnitude. The maximum considered value  $n_{\text{mock}} = 5000$  is sufficient for surveys as large as  $n_{\text{sim}} = 50000$ . This is still below typical sizes of spectroscopic surveys. However, if we want to slice the survey into parts, to trace age–metallicity relations, for

example, as is done below in Sect. 6.2, the size of each part will be on the order of  $n_{\text{sim}} = 50000$  or even smaller. Even more importantly, all these measurements are done for artificial data, where systematic uncertainties are zero by definition, as simulated, and mock catalogues are created from the same set of models. For real data, as we show below, systematic errors will likely be the dominating source of uncertainty.

## 6. Real data applications

### 6.1. Inversion of full surveys

We apply the inversion method described above to several large spectroscopic surveys, namely APOGEE (Majewski et al. 2017), GALAH (Martell et al. 2017), RAVE-on (Casey et al. 2017), and LAMOST (Luo et al. 2015). The results are presented in Fig. 9. These plots indicate that all surveys have a bimodal age distribution, although the location, width, and relative amplitudes for the two modes are different. There are several possible explanations for this fact. The first possibility is that the trend is real and is caused by the target selection in the survey, with stars of different metallicity being observed in different parts of the Galaxy. The second possibility is that one of the implicit assumptions of the method does not hold, namely, that in a given sample age does not depend on  $\log g$  or  $[\text{Fe}/\text{H}]$ . Within the metallicity bin for a magnitude-limited survey, observed stars with lower

values of  $\log g$ , and thus with higher luminosities, are on the average located at larger distances than those with higher  $\log g$ , and thus with lower luminosities. Therefore, age distribution might vary over the observed  $\log g$  range. This may cause the inversion method to give incorrect results. It remains unclear however, which of the effects dominates.

Last but not least, PARSEC models and spectroscopic measurements can have a systematic offset between them, which can cause a complex systematic age bias, which cannot be accounted for in modelling.

Possible systematic effects also manifest themselves in the difference between the inversion result and  $\log(\text{age})$  PDFs of the survey, indicated in Fig. 9 with black dashed and grey solid lines. This difference is considerably larger than the one we obtain for the simulated data, where the systematic offset is zero, even though the survey size  $n_{\text{sim}}$  is larger than those considered in the simulation.

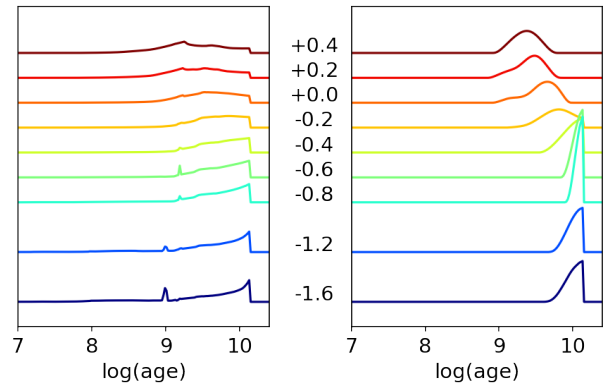
## 6.2. Inversion of mono-metallicity populations

The limitation of the inversion of the full survey is that the age distribution is assumed to be the same for all distances and all metallicities covered by the survey, which in general does not hold, as more metal poor stars tend to be older. In order to mitigate this problem we made a separate study of RAVE-on, LAMOST (Luo et al. 2015), DR4, and APOGEE stars, splitting them into several metallicity bins and applying age inversion to each bin separately.

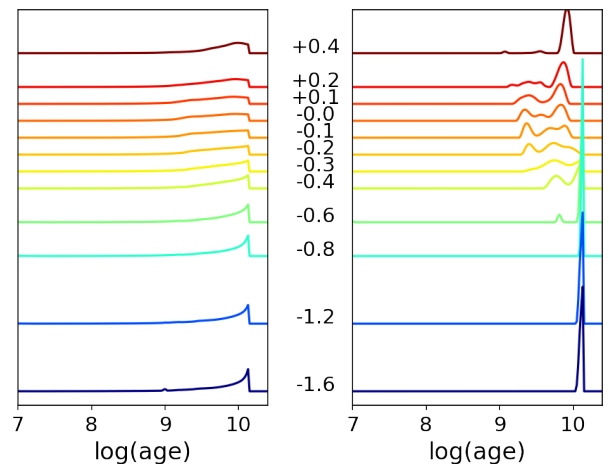
In Fig. 10 we present the inputs ( $\log(\text{age})$  PDFs) and outputs (underlying age distributions) of the age inversion procedure for RAVE-on stars. Inputs ( $\log(\text{age})$  PDFs) are typically broad and for metal-poor populations show an extra narrow peak at  $\log(\text{age}) \tau \approx 9$ , which is associated with red-clump stars. The underlying age distributions computed through the inversion are a lot smoother and narrower, with no secondary peaks. The peak of the underlying age distribution goes from  $\log(\text{age}) \tau = 9.37$  for the highest metallicity bin  $[\text{Fe}/\text{H}] = 0.4$  dex to the maximum possible  $\log(\text{age})$  of  $\tau = 10.13$  for  $[\text{Fe}/\text{H}] < -0.4$  dex. At very low metallicities,  $[\text{Fe}/\text{H}] \leq -1.2$ , a gradual broadening of the underlying age distribution is observed, which is caused by the presence of a lower number of stars in metal-poor bins and thus a reduced age resolution (see test results with low  $n_{\text{sim}}$  values in Figs. 6 and 7).

For LAMOST, we use a constraint of  $T_{\text{eff}} < 7000$  K in this work, as there are visible pipeline artefacts in the Hertzsprung–Russell diagram at higher temperatures, which we are not able to simulate. Even with that cut, LAMOST contains about an order of magnitude more stars than RAVE-on or APOGEE. The results of the age inversions for the LAMOST sample are shown in Fig. 11 and for the APOGEE sample in Fig. 12. Results for LAMOST show narrower underlying age distributions than RAVE-on or APOGEE results, likely due to larger statistics and hence higher age resolution.

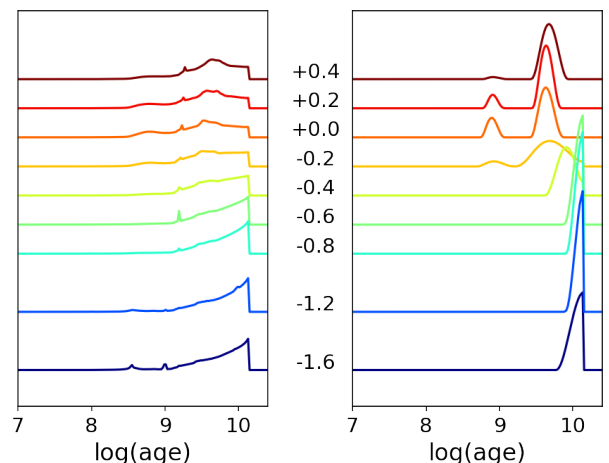
For LAMOST, two peaks are visible in the underlying age distributions for metallicities  $[\text{Fe}/\text{H}] \geq -0.2$  dex: one at around  $\tau \approx 9.35$  and one at  $\tau \approx 9.83$ . The latter peak becomes dominant and shifts to higher  $\tau$  values, as  $[\text{Fe}/\text{H}]$  increases. Similarly, for APOGEE a secondary peak is visible in the underlying age distribution for metallicities  $[\text{Fe}/\text{H}] \geq -0.2$  dex at around  $\tau \approx 8.9$ . In both cases, the result is that the mean age for bins with  $[\text{Fe}/\text{H}] \geq +0.2$  dex is higher than that for solar-like metallicities. This does not seem to be physical – we expect more metal-rich stars to be systematically younger. This might be attributed to the limitations of the method listed above in the Sect. 6.1. An



**Fig. 10.** Results of age inversion for RAVE-on metallicity slices:  $\log(\text{age})$  PDFs as produced by UniDAM (left) and underlying age distribution computed through the inversion (right). Plots for various metallicities are offset in vertical direction for visual purposes, with numbers between plots indicating  $[\text{Fe}/\text{H}]$  values.



**Fig. 11.** Same as Fig. 10, but for the LAMOST survey.



**Fig. 12.** Same as Fig. 10, but for the APOGEE survey.

alternative explanation is that the effect is physical, and is caused by the fact that metal rich ( $[\text{Fe}/\text{H}] \geq +0.2$  dex) stars originate in the inner part of the Galaxy, and it takes a certain amount of time for the migration process to bring them to the solar vicinity, where they can be observed (see Minchev et al. 2018). Thus a small deficit of young metal-rich stars can be expected.

It is very likely that all three effects are acting simultaneously. For better results, selection effects (like those presented in [Mints & Hekker 2019](#)) have to be taken into account. Binning the data in distance or galactic latitude is also desirable. Both tasks are beyond the scope of the current work.

Overall, the underlying age distributions obtained through the inversion allows us to trace age–metallicity trends that are much less obvious if  $\log(\text{age})$  PDFs are considered. It also allows structures like low-age spikes to be removed, which do not seem to be real and are likely due to observational scatter in physical parameters of older stars.

### 6.3. Metallicity–age and age–metallicity relations

Results from Sect. 6.2 allow us to study metallicity–age and age–metallicity relations for RAVE-on LAMOST and APOGEE. Both relations can be defined through a two-dimensional distribution function  $C(\tau, [\text{Fe}/\text{H}]) = C_{[\text{Fe}/\text{H}]}(\tau)N([\text{Fe}/\text{H}])$ , where  $C_{[\text{Fe}/\text{H}]}(\tau)$  is the  $\log(\text{age})$  PDF in the bin with metallicity  $[\text{Fe}/\text{H}]$ , and  $N([\text{Fe}/\text{H}])$  is the number of stars in the survey as a function of metallicity. From that we can define  $\log(\text{age})$  as a function of metallicity as:

$$\tau([\text{Fe}/\text{H}]) = \frac{\int_{\tau_{\min}}^{\tau_{\max}} C(\tau, [\text{Fe}/\text{H}]) d\tau}{\tau_{\max} - \tau_{\min}}, \quad (19)$$

and similarly metallicity as a function of  $\log(\text{age})$ :

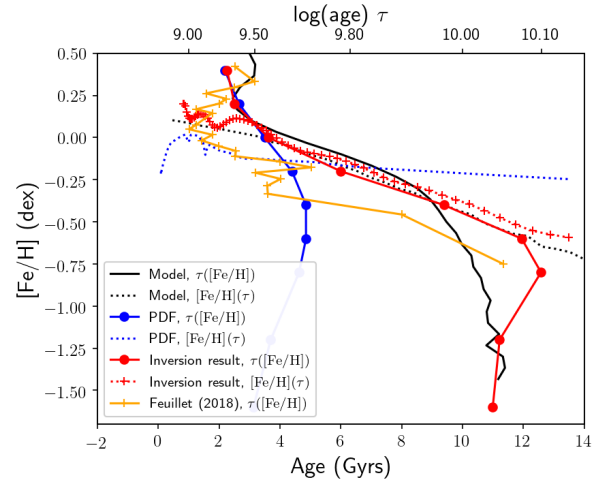
$$[\text{Fe}/\text{H}](\tau) = \frac{\int_{[\text{Fe}/\text{H}]_{\min}}^{[\text{Fe}/\text{H}]_{\max}} C(\tau, [\text{Fe}/\text{H}]) d[\text{Fe}/\text{H}]}{[\text{Fe}/\text{H}]_{\max} - [\text{Fe}/\text{H}]_{\min}}. \quad (20)$$

Here,  $(\tau_{\min}, \tau_{\max})$  is the considered range in  $\log(\text{age})$  and  $([\text{Fe}/\text{H}]_{\min}, [\text{Fe}/\text{H}]_{\max})$  is the range in metallicity.

In Figs. 13–15 we show  $\tau([\text{Fe}/\text{H}])$  with blue solid and  $[\text{Fe}/\text{H}](\tau)$  with blue dotted lines. In these plots, we continue to use solid lines for  $\tau([\text{Fe}/\text{H}])$  relations and dotted lines for  $[\text{Fe}/\text{H}](\tau)$  relations. These two lines are nearly orthogonal, which is the consequence of the  $C(\tau, [\text{Fe}/\text{H}])$  distribution being very broad.

We can now replace  $\log(\text{age})$  PDFs  $C(\tau, [\text{Fe}/\text{H}])$  with the underlying age distribution from the inversion  $N(\tau, [\text{Fe}/\text{H}])$  in Eqs. (19) and (20). The results are shown in Figs. 13–15 with red solid and dotted lines. These two red lines are a lot closer to each other, which is a sign of a much tighter  $C(\tau, [\text{Fe}/\text{H}])$  function.

These data were compared to the results of the recent work by [Feuillet et al. \(2018\)](#), who presented the analysis of a sample of 721 nearby red giant stars selected from APOGEE ([Majewski et al. 2017](#)). All these stars are closer than 400 pc and have reliable TGAS parallaxes, which allows the determination of  $\log(\text{age})$  to 0.07 dex precision. This sample was used, among other applications, to derive mean ages as a function of metallicity, shown in Figs. 13–15 with orange solid lines. We note that the data presented in [Feuillet et al. \(2018\)](#), as well as our results for  $\tau([\text{Fe}/\text{H}])$ , give age distribution and mean age in each metallicity bin, thus measuring age as a function of metallicity  $\tau([\text{Fe}/\text{H}])$ . At the same time, models of chemical evolution typically focus on the metallicity as a function of age  $[\text{Fe}/\text{H}](\tau)$  (see for example Fig. 4 in [Minchev et al. 2013](#)), which is strictly speaking a different function. Functions  $\tau([\text{Fe}/\text{H}])$  and  $[\text{Fe}/\text{H}](\tau)$  can be close to each other only if the two-dimensional distribution  $C(\tau, [\text{Fe}/\text{H}])$  is tight. Hence in a general case care must be taken in comparing  $\tau([\text{Fe}/\text{H}])$  and  $[\text{Fe}/\text{H}](\tau)$ , as this comparison might lead to inaccurate conclusions.



**Fig. 13.** Age–metallicity ( $\tau([\text{Fe}/\text{H}])$ ) relations for nearby APOGEE stars (orange line, [Feuillet et al. 2018](#)) and for RAVE-on stars (solid lines, this work). Blue lines are values derived from  $\log(\text{age})$  PDFs, red lines are values derived from the underlying age distributions obtained from inversion, and black lines are values derived from [Minchev et al. \(2013\)](#) chemical evolution models. Dotted lines show mean metallicity as a function of  $\log(\text{age})$  for the same data.

If we compare the age–metallicity distributions from  $\log(\text{age})$  PDFs  $C(\tau, [\text{Fe}/\text{H}])$  with those obtained from the underlying age distributions computed through the inversion,  $N(\tau, [\text{Fe}/\text{H}])$ , we see that the latter are much closer to [Feuillet et al. \(2018\)](#) than the former. The small systematic offset can be attributed to the fact that, as opposed to the solar neighbourhood APOGEE subsample used by [Feuillet et al. \(2018\)](#), surveys used here contain more thick-disk and halo stars that are systematically older.

We can compare our age–metallicity and metallicity–age trends with those predicted by chemical evolution models. For this comparison, we use the model described in [Minchev et al. \(2013\)](#). We added an additional smoothing to the two-dimensional age–metallicity distribution predicted by this model. The smoothing scale was chosen to be close to the typical uncertainty in age (2 Gyr) and metallicity (0.1 dex). We then calculated  $\tau([\text{Fe}/\text{H}])$  and  $[\text{Fe}/\text{H}](\tau)$  for this distribution and show it in Figs. 13 and 14 with black solid lines and Fig. 15 with black dotted lines. We note that mean age  $\tau$  starts to increase as a function of  $[\text{Fe}/\text{H}]$  for  $[\text{Fe}/\text{H}] > 0$  for the model data as well as LAMOST and APOGEE inversion results. This is very likely caused by the absence of young metal-rich stars in the solar vicinity, which are formed in the inner Galaxy and need time to migrate outwards. Mean ages for metal-poor stars are also systematically larger for inversion results than those predicted by models. This might be caused by the fact that UniDAM accommodates stellar ages up to the age of the Universe ( $\approx 13.5$  Gyr), while the maximum stellar age in the [Minchev et al. \(2013\)](#) model is 11.175 Gyr (before the smoothing was applied). In general, on all three plots  $\tau([\text{Fe}/\text{H}])$  and  $[\text{Fe}/\text{H}](\tau)$  from the model and from age inversion results are close to each other.

## 7. Summary and outlook

In this work we present age inversions – a method for revealing the underlying distribution of a large ( $N > 10^3$ ) ensemble of stars in  $\log(\text{age})$  from their cumulative  $\log(\text{age})$  PDFs produced by UniDAM. This allows us to remove biases inherent to isochrone

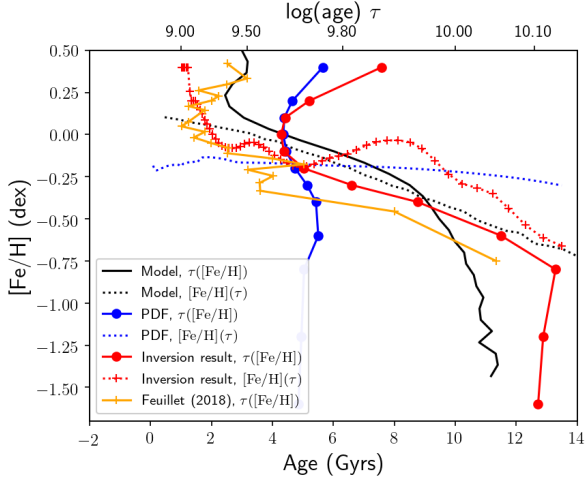


Fig. 14. Same as Fig. 13, now for LAMOST stars.

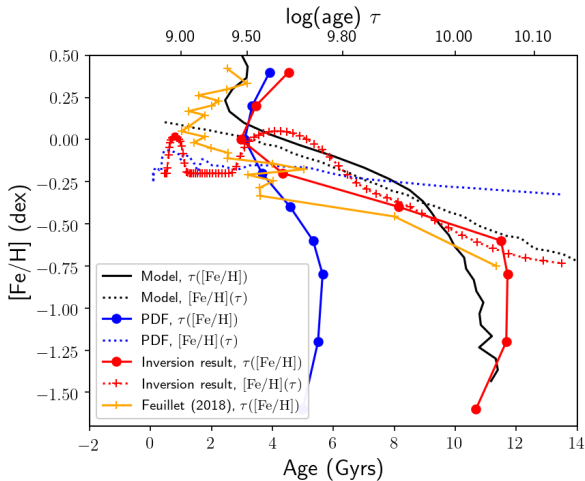


Fig. 15. Same as Fig. 13, now for APOGEE stars.

fitting. The method was tested on simulated data showing that it allows us to reconstruct the underlying age distribution of stars in the simulated sample. The inversion results are a lot closer to the real underlying age distribution than the stacked  $\log(\text{age})$  PDF or the distribution of mean PDF values.

The method was further applied to data produced by UniDAM for real surveys, deriving different age distributions for different surveys. This is expected, as surveys use different observational strategies and focus on different samples of stars. Systematic offsets between survey data and PARSEC models used in this work and survey pipeline artefacts are the main limitations of the method. Due to this we are limited in our ability to generate mono-age populations that will be distributed closely to the survey stars, which is needed for a reliable age inversion.

We also apply the age inversion method to mono-metallicity slices of RAVE-on, LAMOST, and APOGEE surveys. This allows us to trace how the age distribution changes as a function of metallicity, and to reconstruct both metallicity–age and age–metallicity relations for RAVE-on, LAMOST, and APOGEE samples, successfully removing artefacts of the isochrone fitting. We obtain results similar to those published in Feuillet et al. (2018) for a much smaller, high-precision APOGEE sample and to those predicted by chemical evolution models. The number of stars in both RAVE-on and LAMOST surveys, at least for

metallicities close to solar, is high enough to make further divisions in, for example, galactic latitude, distance, and  $\alpha$ -abundance bins. This is considered beyond the scope of the current work.

In the future, systematic offsets between data and models are expected to decrease as the uncertainties (both random and systematic) of observations decrease and as models improve. Furthermore, the future spectroscopic surveys such as 4MOST (de Jong et al. 2016) and WEAVE (Dalton et al. 2014) will provide data for millions of stars. The age inversion method presented here will be suited to providing the underlying age distribution of such large survey data sets as a whole and as functions of, for example, metallicity, location in the Galaxy, and  $\alpha$ -elements abundances.

One could be tempted to use the vast amount of data contained in *Gaia* to derive age distributions from photometry alone, as was done, for example, in Dolphin (2013). However, this might be difficult given the lack of metallicity information in photometric data, which is essential for unambiguous age determination. In any case, such a study would differ substantially from the present one, and the new method to solve Eq. (4) might be needed.

It is possible to make use of *Gaia* parallax data to improve the precision of ages derived from spectrophotometric data, as was done in Mints & Hekker (2018) and Mints (2018). However, this would require a more complex simulation strategy that takes into account proper parallax uncertainty distribution. In addition, *Gaia* parallax zero-point offset (see, e.g. Leung & Bovy 2019) would have to be treated properly, as it can inflict a systematic age bias.

*Acknowledgements.* Authors thank the SAGE group at Max Planck Institute for Solar System Research for many fruitful discussions. Authors thank the anonymous referee for a detailed report with many useful suggestions, which helped us to improve the manuscript substantially. The research leading to the presented results has received funding from the European Research Council under the European Community’s Seventh Framework Programme (FP7/2007–2013)/ERC grant agreement (No 338251, StellarAges). This research made use of Astropy, a community-developed core Python package for Astronomy (Astropy Collaboration 2013). This research made use of matplotlib, a Python library for publication quality graphics (Hunter 2007). This research made use of SciPy (SciPy Team 2001). This research made use of TOPCAT, an interactive graphical viewer and editor for tabular data (Taylor 2005). Funding for RAVE has been provided by: the Australian Astronomical Observatory; the Leibniz-Institut fuer Astrophysik Potsdam (AIP); the Australian National University; the Australian Research Council; the French National Research Agency; the German Research Foundation (SPP 1177 and SFB 881); the European Research Council (ERC-StG 240271 Galactica); the Istituto Nazionale di Astrofisica at Padova; The Johns Hopkins University; the National Science Foundation of the USA (AST-0908326); the W. M. Keck foundation; the Macquarie University; the Netherlands Research School for Astronomy; the Natural Sciences and Engineering Research Council of Canada; the Slovenian Research Agency; the Swiss National Science Foundation; the Science and Technology Facilities Council of the UK; Opticon; Strasbourg Observatory; and the Universities of Groningen, Heidelberg and Sydney. The RAVE web site is at <https://www.rave-survey.org>. 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](http://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 Observatories 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. Guoshoujing Telescope (the Large Sky Area Multi-Object Fiber Spectroscopic Telescope LAMOST) is a National Major Scientific Project built by the Chinese Academy of Sciences. Funding for the project has been provided by the National Development and Reform Commission. LAMOST is operated and managed by the National Astronomical Observatories, Chinese Academy of Sciences. The GALAH survey is based on observations made at the Australian Astronomical Observatory, under programmes A/2013B/13, A/2014A/25, A/2015A/19, A/2017A/18. We acknowledge the traditional owners of the land on which the AAT stands, the Gamilaraay people, and pay our respects to elders past and present.

## References

- Astropy Collaboration (Robitaille, T. P., et al.) 2013, *A&A*, 558, A33
- Bressan, A., Marigo, P., Girardi, L., et al. 2012, *MNRAS*, 427, 127
- Casey, A. R., Hawkins, K., Hogg, D. W., et al. 2017, *ApJ*, 840, 59
- Cutri, R. M., Wright, E. L., Conrow, T., et al. 2014, *VizieR Online Data Catalog: II/328*
- Dalton, G., Trager, S., Abrams, D. C., et al. 2014, in *Ground-based and Airborne Instrumentation for Astronomy V*, SPIE Conf. Ser., 9147, 91470L
- de Jong, R. S., Barden, S. C., Bellido-Tirado, O., et al. 2016, in *Ground-based and Airborne Instrumentation for Astronomy VI*, SPIE Conf. Ser., 9908, 99081O
- Dolphin, A. E. 2013, *ApJ*, 775, 76
- Feillet, D. K., Bovy, J., Holtzman, J., et al. 2016, *ApJ*, 817, 40
- Feillet, D. K., Bovy, J., Holtzman, J., et al. 2018, *MNRAS*, 477, 2326
- Hunter, J. D. 2007, *Comput. Sci. Eng.*, 9, 90
- Lawson, C. L., & Hanson, R. J. 1995, *Solving Least Squares Problems* (Philadelphia: SIAM)
- Leung, H. W., & Bovy, J. 2019, *MNRAS*, 489, 2079
- Luo, A.-L., Zhao, Y.-H., Zhao, G., et al. 2015, *Res. Astron. Astrophys.*, 15, 1095
- Majewski, S. R., Schiavon, R. P., Frinchaboy, P. M., et al. 2017, *AJ*, 154, 94
- Martell, S., Sharma, S., Buder, S., et al. 2017, *MNRAS*, 465, 3203
- Minchev, I., Chiappini, C., & Martig, M. 2013, *A&A*, 558, A9
- Minchev, I., Anders, F., Recio-Blanco, A., et al. 2018, *MNRAS*, 481, 1645
- Minchev, I., Matijevic, G., Hogg, D. W., et al. 2019, *MNRAS*, 487, 3946
- Mints, A. 2018, ArXiv e-prints [arXiv:1805.01640]
- Mints, A., & Hekker, S. 2017, *A&A*, 604, A108
- Mints, A., & Hekker, S. 2018, *A&A*, 618, A54
- Mints, A., & Hekker, S. 2019, *A&A*, 621, A17
- Perryman, M. A. C., de Boer, K. S., Gilmore, G., et al. 2001, *A&A*, 369, 339
- Queiroz, A. B. A., Anders, F., Santiago, B. X., et al. 2018, *MNRAS*, 76, 2556
- Scipy Team 2001, *SciPy: Open Source Scientific Tools for Python*, <http://www.scipy.org/>
- Skrutskie, M. F., Cutri, R. M., Stiening, R., et al. 2006, *AJ*, 131, 1163
- Soderblom, D. R. 2010, *ARA&A*, 48, 581
- Taylor, M. B. 2005, in *Astronomical Data Analysis Software and Systems XIV*, eds. P. Shopbell, M. Britton, & R. Ebert, *ASP Conf. Ser.*, 347
- Tucci Maia, M., Ramírez, I., Meléndez, J., et al. 2016, *A&A*, 590, A32
- Wu, Y.-Q., Xiang, M.-S., Zhang, X.-F., et al. 2017, *Res. Astron. Astrophys.*, 17, 5
- Xiang, M., Liu, X., Shi, J., et al. 2017, *ApJS*, 232, 2