Bayesian inference of stellar parameters and interstellar extinction using parallaxes and multiband photometry

C. A. L. Bailer-Jones[1,2]★

[1]*Max-Planck-Institut für Astronomie, Königstuhl 17, 69117 Heidelberg, Germany*
[2]*Laboratory for Particle Physics and Cosmology, Department of Physics, Harvard University, 38 Oxford Street, Cambridge, MA 02138, USA*

ABSTRACT

Astrometric surveys provide the opportunity to measure the absolute magnitudes of large numbers of stars, but only if the individual line-of-sight extinctions are known. Unfortunately, extinction is highly degenerate with stellar effective temperature when estimated from broadband optical/infrared photometry. To address this problem, I introduce a Bayesian method for estimating the intrinsic parameters of a star and its line-of-sight extinction. It uses both photometry and parallaxes in a self-consistent manner in order to provide a non-parametric posterior probability distribution over the parameters. The method makes explicit use of domain knowledge by employing the Hertzsprung–Russell Diagram (HRD) to constrain solutions and to ensure that they respect stellar physics. I first demonstrate this method by using it to estimate effective temperature and extinction from *BVJHK* data for a set of artificially reddened *Hipparcos* stars, for which accurate effective temperatures have been estimated from high-resolution spectroscopy. Using just the four colours, we see the expected strong degeneracy (positive correlation) between the temperature and extinction. Introducing the parallax, apparent magnitude and the HRD reduces this degeneracy and improves both the precision (reduces the error bars) and the accuracy of the parameter estimates, the latter by about 35 per cent. The resulting accuracy is about 200 K in temperature and 0.2 mag in extinction. I then apply the method to estimate these parameters and absolute magnitudes for some 47 000 F, G, K *Hipparcos* stars which have been cross-matched with Two-Micron All-Sky Survey (2MASS). The method can easily be extended to incorporate the estimation of other parameters, in particular metallicity and surface gravity, making it particularly suitable for the analysis of the 10^9 stars from *Gaia*.

Key words: methods: data analysis – methods: statistical – surveys – stars: fundamental parameters – Hertzsprung–Russell and colour–magnitude diagrams – dust, extinction.

1 INTRODUCTION

The upcoming astrometric survey *Gaia* will provide accurate parallaxes (better than 10 per cent) for about one hundred million stars out to 10 kpc, representing an enormous increase in the number of stars for which accurate distances can be derived (e.g. Mignard & Drimmel 2007; Lindegren et al. 2008; Bailer-Jones 2009). When combined with a measurement of the apparent magnitude, m, this allows us to estimate the absolute magnitude, M, of a star via the simple geometric relationship

$$m + 5 \log \varpi = M + A - 5, \qquad (1)$$

where ϖ is the parallax (in arcsec) and A is the interstellar extinction in magnitudes. Fundamental stellar parameters are usually inferred using just the spectral energy distribution (SED), yet clearly the parallaxes provide an important additional constraint on M. Yet the above equation can only be applied in a simple manner to deduce M if A is known. In principle, it too may be estimated from the SED, but in practice there is a strong degeneracy between extinction and effective temperature (T) in optical/near-infrared multiband photometry which limits the accuracy with which either can be estimated. Such a degeneracy exists with the very low resolution *Gaia* spectroscopy (Bailer-Jones 2010a). Moreover, M is a strong function of T, so we cannot estimate A and T from the spectrum and then expect equation (1) to give a consistent solution for M.

A common way to avoid this dilemma is to assume a value for A – often zero or a value from an extinction map – but this is rarely

★E-mail: calj@mpia.de

valid and certainly not admissible for a deep, all sky survey such as *Gaia*. The basic issue is that A and T are not independent of parallax and apparent magnitude. The solution is to solve for A and T simultaneously using both the SED and the parallax/apparent magnitude. This must be done probabilistically in order to properly characterize the intrinsic degeneracy between A and T.

This paper introduces a general way to do this which uses Bayes' theorem to ensure that all of the information is taken into account self-consistently. The basic idea is to estimate $P(A, T|\boldsymbol{p}, q)$, the posterior probability density function (PDF) over the two parameters given two pieces of information, the normalized SED, $\boldsymbol{p}$, and the quantity $q = m + 5\log\varpi$. This normalized SED describes just the shape of the SED, ignoring the overall flux level. It is used in a conventional, multivariate, forward modelling approach to compare the data with a set of labelled templates in order to obtain $P(\boldsymbol{p}|A, T)$, the likelihood over A, T. The quantity q constrains the sum $M + A$ (equation 1). Used alone it can do little other than place plausible, but not very useful, limits on extreme values of A and M. I therefore explicitly incorporate the knowledge embodied in the Hertzsprung–Russell Diagram (HRD), the distribution of stars in the (M, T) plane. The physics of stellar structure forbids stars from occupying large areas of this plane, and the nature of stars' structure and changing rates of evolution mean that the remaining parts are far from being uniformly populated. This well-established information should not be ignored when inferring astrophysical parameters (APs). The method makes uses of this information in a consistent and quantitative probabilistic framework.

In Section 2, I describe the method in detail and derive the basic equations. I demonstrate it in Section 3 by using it to estimate parameters for a set of 5280 stars covering a range of A and T using *BVJHK* photometry and parallax. These data are based on 880 *Hipparcos* stars (ESA 1997) for which effective temperatures were estimated by Valenti & Fischer (2005) (hereafter VF05) from echelle spectra. I artificially redden the data in order to introduce extinction variance. As the 'true' parameters of these data are known, it can be shown that the method improves the parameter estimation accuracy compared to using just the four colours. I then apply this method in Section 4 to 'blindly' estimate A and T for 85 000 *Hipparcos* stars.

The motivation for this work is to make best use of the parallax in order to improve the estimation of stellar APs. In principle one could add q as another input alongside $\boldsymbol{p}$ to a pattern recognition algorithm such as a neural network or a support vector machine. But such tools fail to recognize the astrophysical significance of this extra input, and unpublished tests by the author show that this approach indeed does not work.

The present paper is not the first to combine astrometric and SED data for stellar parameter estimation in a probabilistic manner. But it is, to the author's knowledge, the first to introduce extinction as a free parameter and to include the HRD in the estimation process. Many authors first derive T and then use the parallax to derive M assuming zero extinction. Alternatively T is derived assuming a value for extinction, a prerequisite for many methods. For example, Takeda et al. (2007) use the inferred stellar parameters $(T, \log g,$ and [Fe/H]) from VF05 to predict the parallax, and then use this in a likelihood model together with evolutionary tracks to infer luminosity, mass and age. This approach does not use the HRD prior nor does it solve for extinction (although assuming zero extinction is probably a valid assumption for these very nearby stars). Pont & Eyer (2004) and Jørgensen & Lindegren (2005) develop Bayesian methods for estimating stellar ages, but they both assume that T is already known and that A is either known or zero. However, this

overlooks the fact that in most large surveys T must be estimated from multiband photometry or low-resolution spectroscopy, and that it is degenerate with A (which is rarely known independently). Estimating both T and A is non-trivial so they should not be considered as 'input data' for an inference. Rather they should be part of that inference in order that their uncertainties and degeneracies be correctly propagated.

Using the Bayesian framework, we can also turn the problem around in order to estimate, for example, stellar distances given some measured properties of the stars. Burnett & Binney (2010) recently outlined a method for obtaining 'spectroscopic parallaxes' in this way.

2 THEORY

2.1 Problem statement

We would like to determine the PDF over the stellar parameters, $\boldsymbol{\phi}$, given measurements of the SED, apparent magnitude and parallax. I will restrict the problem to $\boldsymbol{\phi} = (A_0, T)$, i.e. to determining $P(A_0, T|\boldsymbol{p}, q)$, although a generalization is straightforward (see Section 5). Before deriving an expression for this in Section 2.7, I must first introduce and explain a few concepts. The method involves calculating likelihoods based on forward modelling of the SED (Sections 2.3 and 2.4) for which we need a template grid which shows variance in effective temperature and extinction. The parallax and apparent magnitude are then introduced using the q constraint and the HRD prior (Sections 2.5 and 2.6). Table 1 summarizes the main notation I use.

2.2 Interstellar extinction

In order to construct a grid showing a range of interstellar extinction, we need to adopt an extinction law. I adopt the widely used form from Cardelli, Clayton & Mathis (1989). This gives the monochromatic extinction in a narrow band at wavelength λ in terms of two extinction parameters A_0 and R_0 as

$$A_\lambda = A_0[a_\lambda + b_\lambda/R_0], \tag{2}$$

where a_λ and b_λ are fixed polynomials. A_0 is frequently written as A_V in this equation, but this is confusing because A_0 *is not the extinction in the V band*. The extinction in the *V* filter (or indeed any filter) with pass band function h_λ is a consequence of integration over the stellar SED, F_λ, i.e.

$$A_V = -2.5\log\left(\frac{\int F_\lambda h_\lambda 10^{-0.4A_\lambda}d\lambda}{\int F_\lambda h_\lambda d\lambda}\right). \tag{3}$$

Table 1. Notation.

V	Apparent magnitude in the V band (mag)
M_V	Absolute magnitude in the V band (mag)
A_V	Extinction in the V band (mag)
A_0	Extinction parameter (mag)
R_0	Selective extinction parameter
T	Stellar effective temperature (K)
Z	Stellar metallicity (fraction)
ϖ	Parallax (arcsec)
q	$\equiv V + 5\log\varpi$ (mag)
$\boldsymbol{p}$	Normalized SED with elements $\{p_i\}$
$\boldsymbol{\phi}$	Set of stellar APs
P	Probability density
log	Base 10 logarithm

Thus A_V depends on the SED of the specific star observed, and hence on its intrinsic parameters (in particular effective temperature). Two stars with different T will generally have different A_V for the same A_0. A_0, in contrast, is a property of the interstellar medium only and so is a better parameter with which to characterize the interstellar extinction.

As the q-constraint depends fundamentally on A_V rather than A_0, we need to express the former in terms of the latter. This will be done in Section 3.5. It turns out that for F, G, K stars with extinctions up to 3.5 mag, the difference between A_0 and A_V is less than 0.2 mag.

The artificial reddening will be done using the specific extinction curves from Fitzpatrick (1999) with $R_0 = 3.1$.

2.3 Forward model

The forward model predicts the observed stellar SED, $\boldsymbol{p}$, given the stellar APs, $\boldsymbol{\phi}$. How many APs we need to consider for an accurate prediction depends in particular on the type of stars we want to model and on the resolution of $\boldsymbol{p}$. Note that $\boldsymbol{p}$ is a *normalized* SED, i.e. it contains no apparent magnitude information. Here, the SED is a set of colours derived from broad-band photometry, so I limit the parameters to A_0 and T, $\hat{\boldsymbol{p}} = \boldsymbol{f}(A_0, T)$. All other APs are assumed either to be fixed (R_0) or to have negligible impact on the normalized SED ([Fe/H] and $\log g$). Although [Fe/H] has a significant and usable effect on broad U-band photometry (e.g. Ivezić et al. 2008), its impact on the redder bands considered here is minimal and is neglected. $\log g$ is an even weaker parameter (Bailer-Jones 2010a) so its variance is neglected too. The method can none the less be generalized to incorporate these extra parameters as appropriate.

The forward model is calculated by a smooth fit to a set of templates using the method developed for the ILIUM algorithm (Bailer-Jones 2010a). It involves fitting a two-dimensional smoothing spline (a thin-plate spline) as a function of A_V and T for each element of $\boldsymbol{p}$ separately.

2.4 The likelihood model

The likelihood of the spectral data given the APs is $P(\boldsymbol{p}|\boldsymbol{\phi}) = P(\boldsymbol{p}|A_0, T)$. Assuming Gaussian errors on a measurement of $\boldsymbol{p} = (p_1, \ldots, p_i, \ldots, p_I)$ with covariance matrix $\mathbf{C}_p$, the likelihood model is an I-dimensional Gaussian

$$P(\boldsymbol{p}|\boldsymbol{\phi}) \propto e^{-D^2/2} = \exp\left(-\frac{1}{2}[\boldsymbol{p} - \boldsymbol{f}(\boldsymbol{\phi})]^T \mathbf{C}_p^{-1}[\boldsymbol{p} - \boldsymbol{f}(\boldsymbol{\phi})]\right).$$

(4)

If the elements of $\boldsymbol{p}$ were uncorrelated then $\mathbf{C}_p = \text{diag}(\sigma_{p_i}^2)$, where σ_{p_i} is the uncertainty in p_i, so the exponent could be simplified to

$$D^2 = \sum_{i=1}^{i=I}\left[\frac{p_i - f_i(\boldsymbol{\phi})}{\sigma_{p_i}}\right]^2.$$

(5)

2.5 Parallax/magnitude (q) constraint

As outlined in the Introduction, simple geometry and the definition of absolute magnitude and extinction place the following constraint on noise-free quantities

$$V + 5\log\varpi = M_V + A_V - 5$$

(6)

(I assume we measure the apparent magnitude in the V band, although any other band would do). The goal is to use this equation to

constrain M_V and A_V from noisy measurements of parallax and magnitude. To do this we need a noise model. For brevity define

$$q \equiv V + 5\log\varpi.$$

(7)

Since equation (6) only holds in the absence of noise, consider the random variable

$$x = q - (M_V + A_V - 5).$$

(8)

The noise model for x is $P(x|M_V, A_V)$, which has expectation value zero and variance σ_q^2, the variance in q (M_V and A_V are not measured so contribute no noise). For simplicity I choose to model this as a one-dimensional Gaussian in x, $\mathcal{N}_x(0, \sigma_q)$. For a given star (fixed M_V and A_V), $P(x|M_V, A_V)$ has its maximum when the measurement q equals $M_V + A_V - 5$ (i.e. $x = 0$). The further a measurement of q is away from this value the less probable it is. As q is the only measured term in equation (8) it follows that $P(x|M_V, A_V) = P(q|M_V, A_V)$.

Now consider $P(q|M_V, A_V)$ as a function of M_V and A_V for a given measurement q, as shown in Fig. 1. We can think of proposing trial solutions for M_V and A_V: the further they lie from the solid line, the lower $P(q|M_V, A_V)$ (inset in Fig. 1). How quickly the probability drops off depends on σ_q. With the Gaussian approximation of the noise model for q we have

$$P(q|M_V, A_V) = \mathcal{N}_x[0, \sigma_q(V, \varpi)]$$
$$= \mathcal{N}_x[q - (M_V + A_V - 5), \sigma_q(V, \varpi)].$$

(9)

This gives a probabilistic constraint on M_V and A_V from a measurement of q, quantified by the known statistics of the noise in the photometry and parallaxes. As noted in Section 2.2, we can write A_V as a function of A_0 and T, so this q constraint can be written as $P(q|M_V, A_0, T)$.
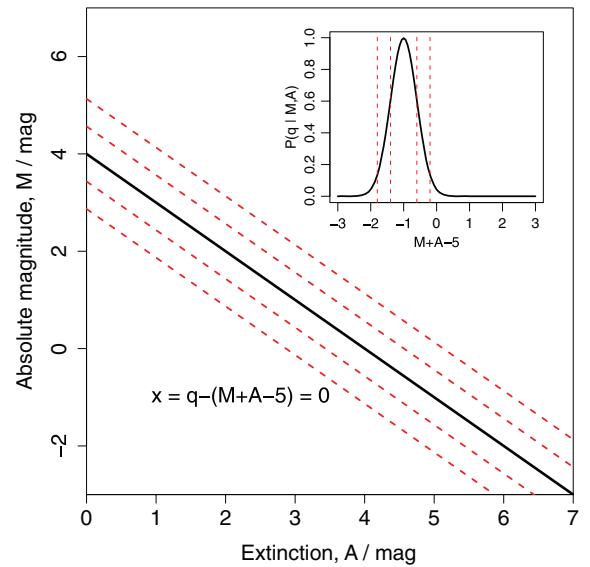

Figure 1. Illustration of using the parallax and apparent magnitude ($q = V + 5\log\varpi$) to constrain extinction and absolute magnitude. Here we measure $q = -1$ (which corresponds to a $V = 14$ star at 1 kpc, for example, or to a $V = 19$ star at 10 kpc, etc.). If this were a noise-free measurement, it would constrain the solution (M_V, A_V) to lie on the solid black line. But as q is a noisy measurement – here a Gaussian with $\sigma_q = 0.4$ (inset) – all solutions have a finite probability, decreasing with distance from the line. Specifically, any slice perpendicular to the line has the Gaussian profile shown in the inset panel, the red dotted lines in both plots showing the 1 and 2 sigma levels for this value of σ_q.

Note that this does not constrain M_V or A_V to have astrophysically 'sensible' values (e.g. the line continues to negative A_V in Fig. 1). This may be done by the HRD prior and/or a prior on extinction.

2.6 Hertzsprung–Russell Diagram prior

The HRD prior, $P(M_V, T)$, gives the relative probabilities of finding stars in different parts of the HRD. The fact that this (M_V, T) plane is far from being uniformly populated is potentially useful in constraining stellar APs: if M_V were known to lie in some range with some probability, for example, T would correspondingly be constrained. This is pertinent information independent of the specific photometric or parallax measurement.

The form we adopt for the HRD depends on the assumed stellar population and can be constructed in a number of different ways. We could, for example, take an observed sample and normalize the relative density of stars to give $P(M_V, T)$. Alternatively we could set the probability at each point to be inversely proportional to the speed of evolution of all types of stars through that point. In this article I will construct the HRD prior using a simulated population of stars evolved with a specified star formation rate, initial mass function (IMF) and metallicity distribution (Section 3.4 and Fig. 9).

2.7 Probabilistic combination

We are now in a position to derive an expression for $P(A_0, T | \boldsymbol{p}, q)$ in terms of quantities we have just introduced. From Bayes' theorem

$$P(A_0, T | \boldsymbol{p}, q) = \frac{P(\boldsymbol{p}, q | A_0, T) P(A_0, T)}{P(\boldsymbol{p}, q)} \quad (10)$$

and from the rule of joint probabilities

$$P(\boldsymbol{p}, q | A_0, T) = P(\boldsymbol{p} | q, A_0, T) P(q | A_0, T). \quad (11)$$

As $\boldsymbol{p}$ and q are independent measurements[1] we can write $P(\boldsymbol{p} | q, A_0, T) = P(\boldsymbol{p} | A_0, T)$. This and equation (11) allow us to write equation (10) as

$$P(A_0, T | \boldsymbol{p}, q) = \frac{P(\boldsymbol{p} | A_0, T) \, P(q | A_0, T) \, P(A_0) \, P(T)}{P(\boldsymbol{p}, q)} \quad (12)$$

where I have also assumed that A_0 and T are unconditionally independent. The terms $P(A_0), P(T), P(\boldsymbol{p}, q)$ are the inevitable priors over these APs or measurements. The first term in the numerator is the likelihood (Section 2.4). The second term we need to further decompose, plus we want to introduce some dependence on M_V so that we can incorporate the HRD and the q constraint. A general rule of probability allows us to write this term as a marginalization over M_V

$$P(q | A_0, T) = \int_{M_V} P(q | M_V, A_0, T) \, P(M_V | A_0, T) \, dM_V. \quad (13)$$

The first term in the integral is the q constraint (Section 2.5). As A_0 is independent of M_V and T, we can rewrite the second term in the integral as

$$P(M_V | A_0, T) = P(M_V | T) = \frac{P(M_V, T)}{P(T)}. \quad (14)$$

(Another way of thinking about this is to note that given T, A_0 tells us nothing additional about M_V.) $P(M_V, T)$ is the HRD prior.

[1] Here I only assume that $\boldsymbol{p}$ and q are independent when conditioned on A_0 and T, although normally we would further assume them to be unconditionally independent. This is the case when $\boldsymbol{p}$ is a normalized SED, as then it bears no distance or apparent magnitude information.

Substituting equation (14) into equation (13) and that into equation (12) gives the final result

$$P(A_0, T | \boldsymbol{p}, q)$$

$$= \underbrace{P(\boldsymbol{p} | A_0, T)}_{\text{likelihood}} \underbrace{\frac{P(A_0)}{P(\boldsymbol{p}, q)}}_{\text{priors}} \underbrace{\int_{M_V} \underbrace{P(q | M_V, A_0, T)}_{q \text{ constraint}} \underbrace{P(M_V, T)}_{\text{HRD prior}} dM_V}_{\text{HRD/q factor}}$$

$$(15)$$

where we see that $P(T)$ has cancelled. This equation can be seen as a product of three terms. The first term is the likelihood function. The second term comprises priors over the extinction and the data. Of these, $P(\boldsymbol{p}, q)$ is not relevant (for AP estimation) because the data are already given. The third term is an integral over two factors: the combined astrometric/photometric noise model (q constraint) and the HRD prior. The integral marginalizes over the unknown M_V leaving a term which is a function of A_0 and T.

Given measurements of $\boldsymbol{p}$ and q we can sample the terms in equation (15) on a grid of A_0 and T in order to map the full PDF. We can also separately marginalize over A_0 and T in order to get one-dimensional PDFs for each AP, i.e.

$$P(T | \boldsymbol{p}, q) = \int_{A_0} P(A_0, T | \boldsymbol{p}, q) \, dA_0 \quad (16)$$

and likewise for A_0. If appropriate we may then summarize this with the mean and a confidence interval.

If we lack information (or do not want to use it) then some terms in equation (15) simplify. For example, if we have no measurement of q then we can set the q constraint to a constant. In that case the integral over M_V makes the HRD prior into a prior on just T. If we do not want to use an informative prior on the extinction we can set $P(A_0)$ to be constant. Likewise, if we do not want to use the HRD prior, then this is equivalent to setting $P(M_V, T)$ to a flat distribution (!). In practice the q constraint is only effective if we use it together with the HRD prior and/or the extinction prior.

Throughout the rest of this paper I will use a uniform extinction prior. As it is separable in equation (15), we can easily imagine the effect of introducing this prior subsequently. I will show two sets of results for $P(A_0, T | \boldsymbol{p}, q)$ based on two different sets of assumptions (priors). The first is a uniform HRD prior and constant q constraint, in which case the posterior PDF is just equal to the likelihood function (renormalized), i.e. the APs are inferred using only the spectrum, $\boldsymbol{p}$. I will therefore refer to this as the *p-model*. This is the baseline against which I will analyse the effect of using the HRD/q factor, using specific models for the q constraint and HRD prior described in the next section. I will refer to this as the *pq-model*.

It may be useful to recognize that when $\boldsymbol{p}$ and q are unconditionally independent (the normal case), we can interpret equation (15) as the combination of two separate estimates of the PDF over (A_0, T) given each of $\boldsymbol{p}$ and q. We can see this when we use Bayes' theorem to rewrite the right-hand side of equation (12) as

$$P(A_0, T | \boldsymbol{p}, q) = \frac{P(A_0, T | \boldsymbol{p}) P(A_0, T | q)}{P(A_0, T)}. \quad (17)$$

The p-model is simply $P(A_0, T | \boldsymbol{p})$.

3 MODEL FITTING AND VALIDATION

Let us now use the probabilistic model to estimate the full posterior PDF over effective temperature and extinction for stars with measured five-band photometry and parallaxes.
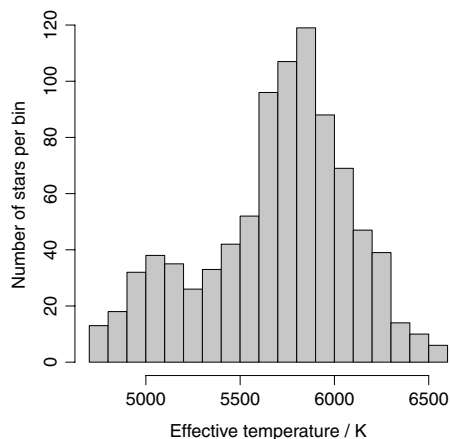
Figure 2. Distribution of effective temperature of the 880 stars from VF05 used to fit and test the model.

All calculations are actually done using $\log T$ rather than T – as then uniform samplings are more appropriate – but many results will be shown in terms of T. (The above theory holds for any monotonic transformation of the variables.) Recall that a small error in $\log T$ of $\delta(\log T)$ corresponds to a fractional error $(\delta T / T)$ of about $2.3\delta(\log T)$.

3.1 Construction of the labelled data set ('extended catalogue')

In order to validate the performance of the model and to assess whether the HRD/q factor improves the accuracy of AP estimates, we need a data set with independent estimates of the APs from spectroscopy. I use a set of 880 bright, nearby F, G, K dwarf stars observed with echelle spectroscopy for the Keck, Lick and Anglo–Australian Telescope (AAT) planet search programmes. VF05 estimated effective temperature, surface gravity, metallicity ([Fe/H]) plus various individual element abundances for these spectra. These estimates are based on Kurucz model atmospheres combined with various line lists with empirical corrections. The stars all have near-solar metallicity (90 per cent with [Fe/H] between -0.4 and $+0.45$ dex), are bright (90 per cent between $V = 5.1$ and 8.6 mag) and have parallaxes from *Hipparcos* (90 per cent have distances ranging from 13 to 67 pc). The full temperature range is 4707– 6594 K ($\log T = 3.673$–3.819 dex) with distribution shown in Fig. 2. I only use these effective temperature estimates (derived only from the spectrum) in what follows.[2] Individual internal errors are about 50 K (1σ). I obtained the corresponding broad-band *BVJHK* photometry via the *Hipparcos* identifiers from SIMBAD [*JHK* comes from Two-Micron All-Sky Survey (2MASS)].[3]

VF05 do not provide estimates of the line-of-sight extinction, but given their proximity it is reasonable to assume it is negligible for most stars. Yet in order to train and validate the model, we need a sample with variance in A_0, which I introduce via simulations. Ideally one would take a sufficiently high resolution SED for each star and calculate the extinction in each band for a range of A_0 (at fixed R_0) using equation (3). This, however, would be difficult to do reliably with the original VF05 data because of the problems of

flux calibrating echelle spectra. I instead represent the SED with a synthetic spectrum at the known effective temperature of each star. In fact, because these are broad-band filters, the extinction in a band varies smoothly with T, so we do not need a synthetic SED for every unique T. It is instead sufficient to calculate the integral on a discrete grid of six values of T (I use SEDs from MARCS models) for a range of A_0 and to then make a series of one-dimensional quadratic fits, $A_b = y_b(T; A_0)$. These functions give the extinction in band b, A_b, as a smooth function of T for each A_0. Extinction is then applied to the original *BVJHK* photometry by adding the corresponding A_b. I do this for six values of A_0 (0.0 to 2.5 mag in steps of 0.5 mag) and apply it to the 880 stars (T values) in the catalogue. This produces an extended catalogue of 5280 stars showing variance in A_0 and T with consistent APs, photometry and astrometry.

The parallaxes are not changed by this procedure. This grid of stars is therefore not characteristic of the solar neighbourhood, because I have introduced numerous fainter, extincted stars with the same T and distance as each unextincted star. This is appropriate, because the goal is only to achieve variance in T and A_0 for fitting the forward model.

Note that the MARCS models have only been used to simulate the *changes* in the colours due to extinction. (The temperature variation is still from the original data.) This makes the data set relatively insensitive to any inaccuracies in the MARCS model's predictions of the colours at the six temperatures used in the fits. Compared to the calibrations of Worthey & Lee (2006), Vallenari (private communication) finds a systematic offset in the predicted MARCS colours of -0.05 mag in $B - V$ at 4770 K, which corresponds to a systematic offset in T of $+50$ K or 1 per cent. (The systematic offsets in the redder colours are smaller.) Although these systematics only affect the non-zero extinction data, they will slightly bias the whole forward model to give slightly lower T estimates relative to the Worthey & Lee calibration. Overall, the effective temperature scale predicted by the forward model is an amalgam of the VF05 and MARCS systems.

In the method I reduce the five-band photometry to the four colours $\boldsymbol{p} = (B - V, V - J, J - H, H - K)$. While the five band measurements are independent, the four colours are not, so we must include the covariant (off-diagonal) terms into the covariance matrix in equation (4) (see Appendix). As these are bright stars their uncertainties are not set by photon statistics but rather by object-independent sources. I therefore adopt an estimate of the photometric uncertainty of 0.021 mag in each band, corresponding to $\sigma_{\mathrm{p}} = 0.03$ mag in each colour.

3.2 Forward model fitting and likelihood calculation

The forward model defined in Section 2.3 was fit to each of the four colours as a function of A_0 and $\log T$ using a quarter of the 5280 stars in the extended catalogue selected at random. A two-dimensional smoothing spline with 10 degrees of freedom gave a good fit in both directions.[4] With either AP fixed, the four colours vary almost linearly with the other AP (see Fig. 3). The sensitivities of the four colours with respect to temperature at zero extinction (the gradients $|\partial p_i/\partial \log T|_{A_0=0}$) are approximately $(-4.4, -7.3, -2.4, -0.4)$ mag dex^{-1} for $(B - V, V - J, J - H, H - K)$, respectively. (Multiply these numbers by $\log 1.1{=}0.041$ to get the approximate colour change due to increasing T by 10 per cent, for example.) The

---

[2] VF05 also estimate luminosity, radius and mass using in addition parallax and V-band photometry, but these are not used here.

[3] The full VF05 sample comprises 1040 stars, but only 885 had full five-band photometry in SIMBAD, and five had highly deviant photometry, leaving 880.

[4] The extended catalogue was initially calculated on a denser grid of A_0, but this did not significantly improve the forward modelling accuracy.
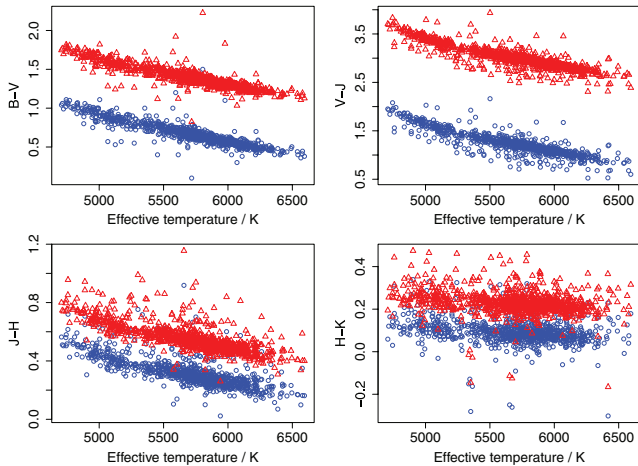
Figure 3. Colour–temperature relation for the stars used to fit and test the model, for $A_0 = 0$ mag (blue circles) and $A_0 = 2.5$ mag (red triangles).

sensitivities of these colours with respect to A_0 at $T = 5500$ K are (0.29, 0.71, 0.10, 0.05) mag/mag. The scatter in the colours at fixed A_0 and T in 100 K temperature wide bins is typically 0.05–0.1 mag for $B - V$, $J - H$ and $H - K$, but larger in $V - J$ (0.1–0.15 mag) and at around $T = 5800$ K. This scatter is partly due to photometric errors but is mostly a result of cosmic variance. This obviously limits the accuracy of AP estimation with these data. For example, if we knew the extinction, then estimating T using just the $V - J$ colour with an error of 0.1 mag would translate into an error in $\log T$ of 0.014 dex or 3 per cent. Although the use of four colours permits better estimates, neither AP is known a priori and, as we shall see, there is an intrinsic degeneracy between these APs with these colours.

Once fit, the forward model is used to calculate the likelihood $P(\boldsymbol{p}|A_0, T)$ (Section 2.4) for the remaining 3/4 of the stars in the extended catalogue. A common way of sampling probability distributions is with Monte Carlo methods, but this is inefficient for a two-dimensional parameter space. I instead sample it on a dense, regular grid with A_0 varying from 0.0 to 3.5 mag in steps of 0.05 mag and $\log T$ varying from $\log 4300$ to $\log 7000$ in steps of 0.005 dex, yielding a grid of $71 \times 43 = 3053$ points. I will refer to this as the *d-grid*. The limits on T at which we calculate this grid are set by the temperature range in the catalogue and avoiding wanting to extrapolate the forward model more than about 400 K in T.

3.3 p-model results

The p-model is the posterior PDF (equation 15) over the two APs assuming uniform priors A_0 and T, i.e. not using the HRD/q factor. This posterior, $P(A_0, T|\boldsymbol{p})$, can be summarized in a two-dimensional contour plot and is shown in Fig. 4 for some example stars, whereby the contours contain 90 per cent, 99 per cent and 99.9 per cent of the total posterior probability (found by integrating the PDF).

The most obvious feature common to all these plots (and for the overwhelming majority of stars in the extended catalogue) is the strong A_0–T degeneracy revealed by the long, narrow contours. *This degeneracy is intrinsic to the data and not a feature of the method.* This is because for given colours corresponding to some nominal A_0 and T, we will get the same colours by increasing both APs or by decreasing both APs by a certain amount. We would even have the degeneracy with negligible photometric errors, although in that case the degenerate region would be smaller (narrower region).

The contours show a slight curvature in these plots against T. In the original $\log T$ space they are straighter with a slope $dA_0/d\log T \simeq 10$ mag dex^{-1}.

Not surprisingly, the APs cannot be estimated very accurately. In most of these plots the 90 per cent contour extends over approximately 1000 K in effective temperature and over 1 mag in extinction. Using just these data we can neither estimate the parameters more precisely nor remove the degeneracy. That can only be done using additional (prior) information, e.g. if we had other data which indicated that the extinction was small.

From a Bayesian perspective the full PDF is the final answer to the inference of the APs. But for surveys of many objects we want to summarize this with a best estimate plus a confidence interval. As the PDFs are predominantly unimodal and symmetric, I summarize using the mean and 90 per cent confidence interval for each AP (plus the correlation coefficient, the slope of the major axis of the contours). These may be obtained by marginalizing the PDF over each AP (equation 16). The result of the marginalization over A_0 is shown in Fig. 5. The 90 per cent confidence interval is the range which contains the central 90 per cent of the probability. It is not necessarily symmetric about the mean. For a Gaussian distribution this interval is the $\pm 1.64\sigma$ interval, so the equivalent '1σ error' is 3.3 times smaller.

The posterior PDF is calculated on and normalized over a discrete, albeit fine, grid (the d-grid), with the 90 per cent confidence interval calculated via discrete integration with linear interpolation. This finite-sized grid truncates the PDF when it extends to the edge of the grid. Some adjustments are therefore made in order to report sensible confidence intervals when the PDF peaks very strongly at the edge of the grid. The constraint at $A_0 = 0$ is a natural one because extinction cannot be negative, but the other three constraints are artificial. We see in the right-hand column of Fig. 5 an example of how this edge peaking translates into constraints on the inferred mean and confidence interval. This actually only occurs for a small fraction of the stars in the extended catalogue. It could be avoided entirely if we chose to extend the A_0 range or to extrapolate the forward model further over T.

In the fifth column of Fig. 4 I show cases where the true APs lie outside of the 90 per cent confidence interval of the p-model estimation. This could be a consequence of (a) cosmic scatter (other parameters) affecting the data, (b) large errors in the data which are not reflected by the estimated uncertainties, $\{\sigma_{p_i}\}$, in the likelihood model, (c) an error in the VF05 T assignment, or a combination of these. There also exist cases where the contours lie at higher A_V and T than the 'true' APs. This could also be an indication that the original star had non-zero extinction.

The overall *accuracy* of the p-model results is shown in Fig. 6, which plots the AP residuals, $\delta\phi = \phi_{\text{inferred}} - \phi_{\text{true}}$, against $\log T$ and each other. There is no particular trend in the residuals, other than evidence of the d-grid limits in the form of diagonal trends at the edges of the upper two plots. The third panel shows the not surprising result that the residuals are correlated. I summarize the accuracy using the mean absolute residual, $\overline{|\delta\phi|}$ (abbreviated as ϵ_{mar}), and the mean residual, $\overline{\delta\phi}$ (abbreviated as ϵ_{sys}), which are summaries of the random errors and systematic errors, respectively. ϵ_{mar} is 0.024 dex (5 per cent) for $\log T$ (300 K for a solar-type star of 5800 K) and 0.29 mag for A_0. (Multiply these by 1.25 to get the 1σ for a Gaussian distribution.) The systematic errors, apparent from the plots, are 0.008 dex (2 per cent) and 0.08 mag, respectively.

The two main sources of the random errors are (1) the photometric errors, plus the inevitable non-Gaussianity of the error distribution, and (2) the intrinsic degeneracy, or the fact that these four colours
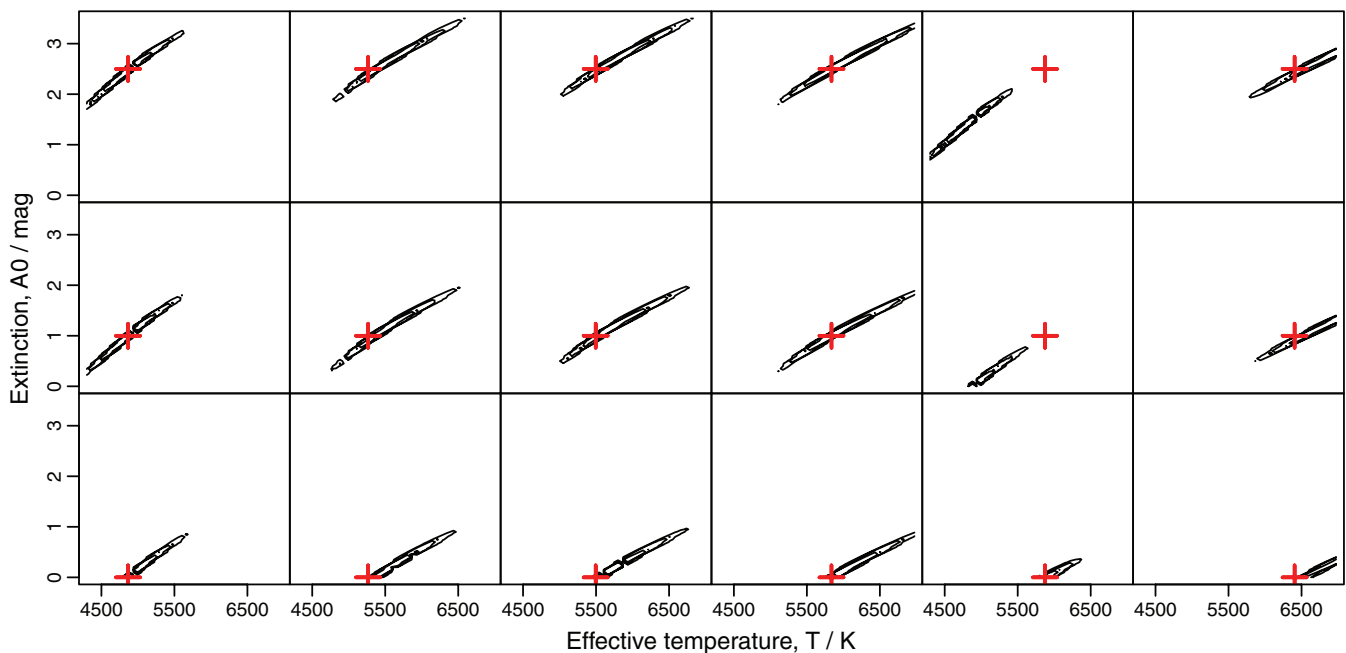
Figure 4. Posterior probability density function (PDF) from the p-model over 18 stars with six different true temperatures (columns) and three different true extinctions (rows) applied to the extended catalogue. Three contours are shown for each star, each enclosing 90 per cent, 99 per cent and 99.9 per cent of the total posterior probability. For comparison, the true AP values are shown with the red cross.
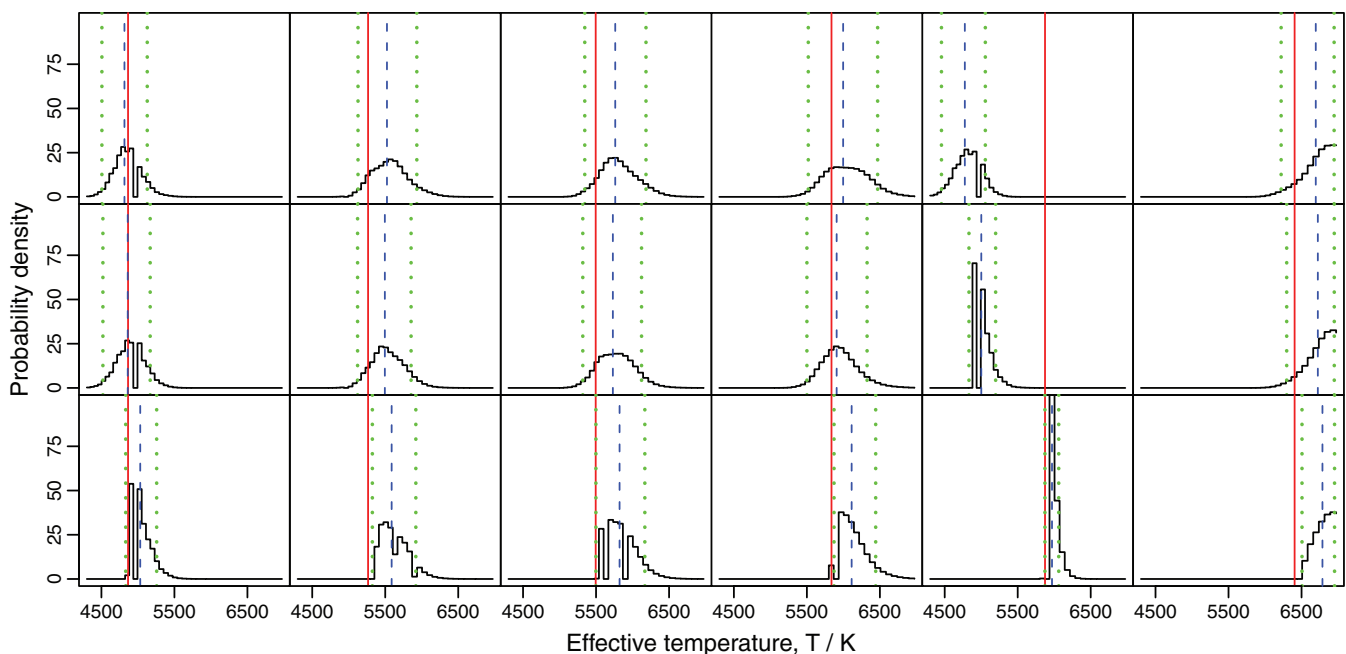

Figure 5. One-dimensional posterior PDFs (p-model) over temperature for the stars shown in Fig. 4 obtained by marginalizing those two-dimensional PDFs over A_0. The probability density is properly normalized (integral is unity). The dashed blue line is the mean of the one-dimensional PDF, and the green dotted lines denote the 90 per cent confidence interval. The 'true' temperature (from VF05) is shown with a solid red line.

alone are insufficient even in principle to estimate both APs exactly. The source of the small systematic errors is less clear. Both train and test data are drawn from the same sample, so it is not an issue of a simple data mismatch. The systematic may be a consequence of the slight biasing introduced by the use of the MARCS models (section 3), which could have introduced an inconsistency in the modelled variation of the colours over the APs.

The *precision* of the AP estimates (how good we think they are in the absence of the truth) is measured by the 90 per cent confidence intervals. These are typically 0.07 dex (16 per cent) in $\log T$ and 0.7 dex in A_0, with relatively little dependence on the APs except at the edge of the d-grid (see Fig. 7).

Let us now introduce an HRD prior and a model for the q constraint and examine how these alter the inference.

© 2010 The Author, MNRAS **411,** 435–452
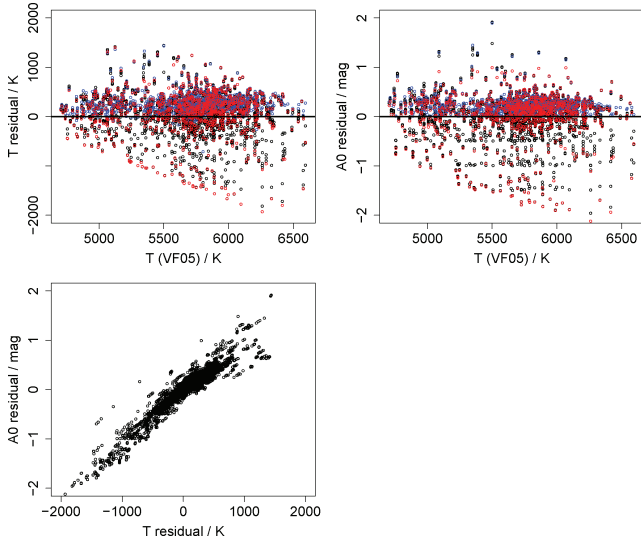Monthly Notices of the Royal Astronomical Society © 2010 RAS

Figure 6. Parameter residuals for the p-model applied to the extended catalogue. The residual is defined as p-model prediction minus true value. Points with true A_0 equal to 0.0 and 2.5 mag are coloured blue and red, respectively.
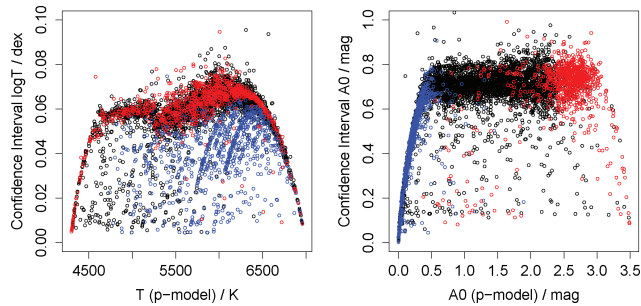

Figure 7. The 90 per cent confidence intervals of the APs inferred from the extended catalogue with the p-model, plotted against the estimated mean AP. To get the equivalent Gaussian 1σ error multiply by 0.3. Points with true A_0 equal to 0.0 and 2.5 mag are coloured blue and red, respectively.

3.4 Model for the HRD prior

The HRD prior, $P(M_V, T)$, is the prior probability over M_V and T based on our knowledge of stellar evolution and stellar populations. We have some choice to exercise in what evolution/population we represent. For the sake of this article I use the results of a complete evolution of a simulated stellar population generated by the code of Vallenari, Chiosi & Sordo (2010) based on the evolutionary models of Bertelli et al. (2008). The population comprises 200 000 stars drawn from a Salpeter IMF with initial masses ranging from 0.2 to 107 $M_\odot$ (although 99 per cent of stars have masses below 1.3 $M_\odot$). A constant star formation rate over a period of 13.7 Gyr is applied and the stars are evolved independently. Solar metallicity, $Z = 0.0190$–0.0193, is used for all stars. The final M_V and T of these stars are calculated and used to place them in the HRD. A subset is shown as the red sequence in Fig. 8. This represents the HRD prior as a set of delta functions. This could be used as is, but in order to accommodate some cosmic variation and overcome the sparseness of the simulation in some parts of the HRD, I smooth it using two-dimensional kernel density estimation. I use a Gaussian kernel with kernel widths of 0.0025 dex in $\log T$ and 0.0625 mag in M_V (and zero covariance) and calculate the density on a 200×200 grid. Normalizing so that the total integral is unity
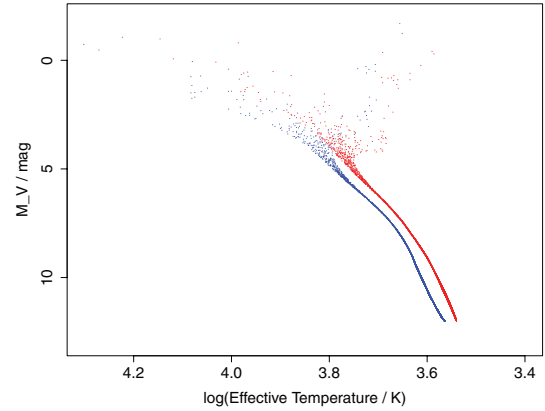

Figure 8. Stellar evolution simulations from Vallenari et al. (2010) used to build the HRD prior. The stellar locus shown in red points (main sequence to the right) is for $Z = 0.019$ and the one in blue points (to the left) is for $Z = 0.0019$. Only every tenth star and only the lower mass stars in the simulations are shown.
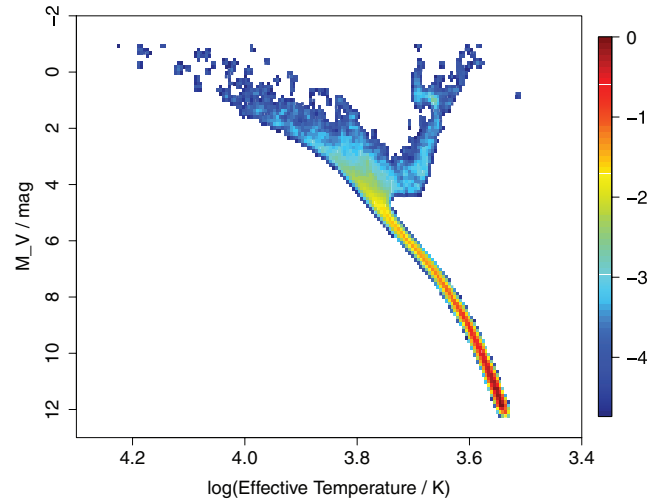

Figure 9. HRD prior. The colour scale shows $\log P(M_V, T)$ normalized for the purpose of illustration to have zero at its maximum. Unoccupied areas are shown in white.

gives the prior $P(M_V, T)$. This is shown using a log density scale in Fig. 9. Note the very large range in densities (almost five orders of magnitude).

This HRD prior is certainly not perfect. Its main defect is that it considers only a single metallicity. While that is not an issue for the solar metallicity VF05 sample, it is a limitation for the later application to the *Hipparcos*/2MASS sample in Section 4. Metallicity significantly affects the position of the stellar loci in the HRD, as we can see in Fig. 8. Comparison of the main sequences in this figure with that in the smoothed HRD prior shows that quite a liberal smoothing has been applied, making this a rather weak prior which could be interpreted as a small prior range on metallicity of the order of ± 0.5 dex. I shall later investigate the impact of the HRD metallicity empirically.

3.5 Model for the q constraint

The q constraint, $P(q|M_V, A_0, T)$, introduced in Section 2.5, is a one-dimensional PDF over $q - (M_V + A_V - 5)$ with zero mean, where A_V is a function of A_0 and T. Although, from equation (7), q will
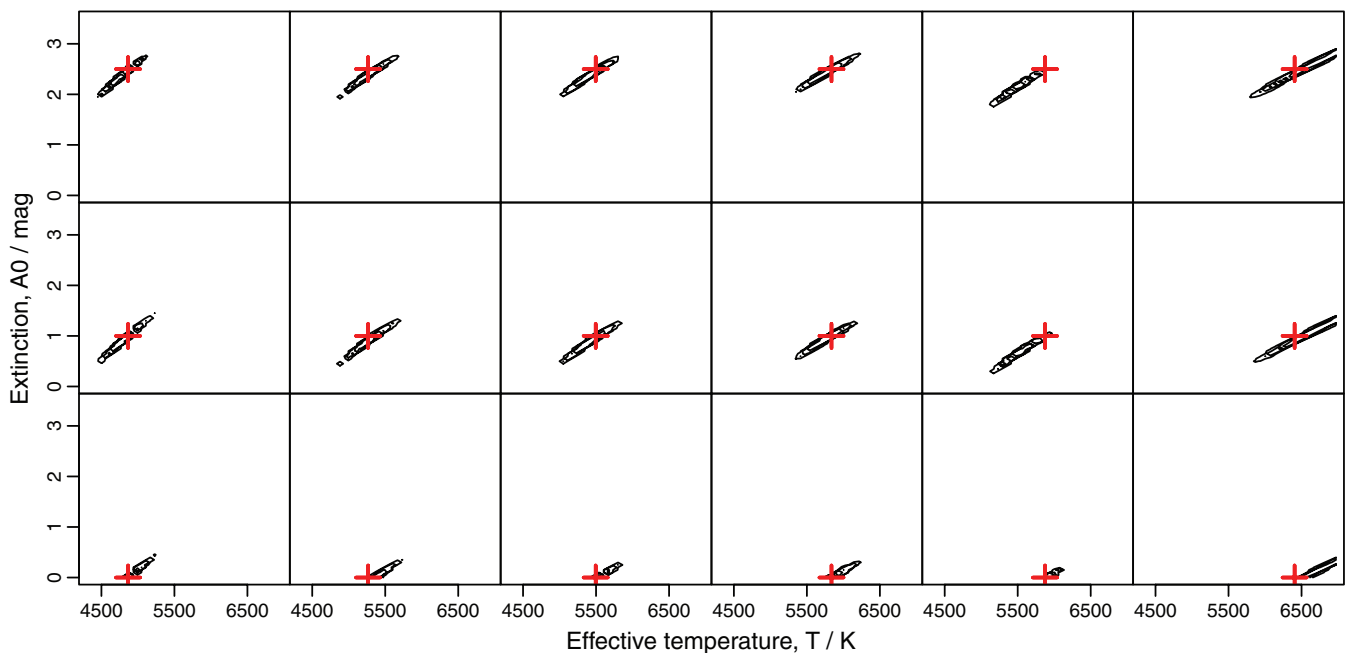
Figure 10. Posterior probability distribution for the pq-model (i.e. including the HRD/q factor) for the same stars as shown in Fig. 4.

not strictly have a Gaussian distribution when V and ϖ have Gaussian distributions, we can still approximate it as such provided we calculate its variance, σ_q^2, correctly. This must take into account the observational correlation between magnitude and parallax. Writing the variances of the photometric noise and the parallax noise as σ_V^2 and σ_ϖ^2, respectively, it follows from the definition of q and a general result of covariances (equation A1) that

$$\sigma_q(V, \varpi)^2 = \left(\frac{5}{\ln 10} \frac{\sigma_\varpi}{\varpi} \right)^2 + \sigma_V^2$$
$$+ 2 \left(\frac{5}{\ln 10} \frac{\sigma_\varpi}{\varpi} \right) \sigma_V \rho(V, \log \varpi) \qquad (18)$$

where ρ is the correlation coefficient.

Calculated over the original catalogue of 880 objects $\rho(V, \log \varpi) = -0.62$, the value I adopt for all calculations. A typical star in the original catalogue has $V = 7$ mag, $\sigma_V = 0.02$ mag, $\varpi = 30$ mas and $\sigma_\varpi = 1$ mas which gives $\sigma_q = \sqrt{0.0052 + 0.0004 - 0.0018} = 0.06$ mag (note that it is dominated by the parallax error). Even the largest value of σ_q (for a star with a small parallax and large parallax error) is just 0.62 mag (90 per cent have $\sigma_q < 0.13$).

How does the q constraint work? Having measured q, it constrains $M_V + A_0$ to within a range of order $\pm \sigma_q$ about this measurement. For a given value of A_0 this limits the range of M_V and hence, via the HRD prior, the range of T. In other words, there is a finite-width probability distribution over T for a given A_0. The ensemble of all such distributions is the two-dimensional PDF over (A_0, T) for given q, which is just $P(A_0, T|q)$ in equation (17).

One more component is required to use this in practice. The q constraint is defined in terms of the extinction in a band, A_V, whereas our goal is to get a posterior PDF over the extinction parameter A_0 (see section 2.2). The two are closely related, however, and it is reasonable to assume that to a good approximation

$$A_V = A_0 + y(A_0, T), \qquad (19)$$

where y is a smooth, two-dimensional function. I fit this using simulated V-band photometry calculated from the same MARCS synthetic spectra used to create the extended catalogue (Section 3.1). I use a regular grid in T (six values) and A_0 (21 values). The variation is quite smooth and is fitted accurately with a thin-plate spline. Because A_0 is defined at a monochromatic wavelength close to the centre of the V band, $y(A_0, T)$ is a small (negative) correction to A_0. Its largest value is -0.18 mag, occurring for the highest extinction ($A_0 = 3.5$ mag) and lowest temperature (4000 K) considered.

3.6 pq-model results

The pq-model was applied to the full extended catalogue using this HRD prior and q-model. The resulting PDFs over the same stars as shown for the p-model in Fig. 4 are shown in Fig. 10. A comparison shows that the introduction of the HRD/q factor has improved the precision (reduced the size of the contours). This can also be seen in the one-dimensional marginalized PDFs (Fig. 11) as a reduction in width and increase in height of the distributions. The HRD/q factor has not removed the degeneracy, but it has reduced it.

The critical question is whether the HRD/q factor has improved the *accuracy*. I again take the mean of the marginalized PDF distribution as the estimate of the AP. The residuals for the pq-model are shown in Fig. 12 and may be compared with the p-model residuals from before. We immediately see that the pq-model has the better performance in terms of both random and systematic errors. The third column of Table 2 summarizes the random errors with the mean absolute residuals. There is a marked improvement with the pq-model compared to the p-model: by 39 per cent in T and 33 per cent in A_0 (measured as $[\epsilon_{mar}^p - \epsilon_{mar}^{pq}]/\epsilon_{mar}^p$). The overall performance of 3.5 per cent (ϵ_{mar}) in T, which is 200 K at 5800 K, is quite good considering the limited data on which this is based, in particular when we recall that the intrinsic scatter in the photometry limits the photometric-only estimates to a similar value (as discussed in Section 3.2). At 5800 K this error corresponds to a Gaussian 1σ of

© 2010 The Author, MNRAS **411**, 435–452
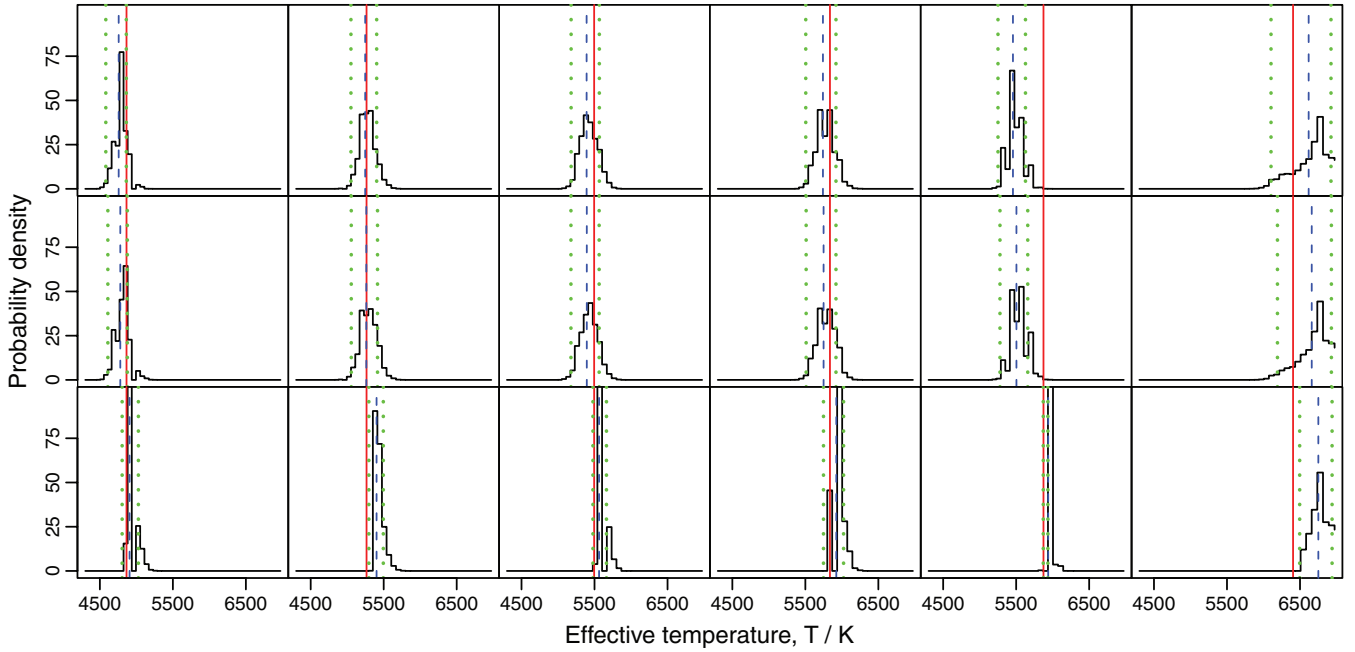Monthly Notices of the Royal Astronomical Society © 2010 RAS

Figure 11. One-dimensional posterior PDF (pq-model) over temperature for the stars shown in Fig. 10 obtained by marginalizing those two-dimensional PDFs over A_0. The dashed blue line is the mean of the one-dimensional PDF, and the green dotted lines denote the 90 per cent confidence interval. The 'true' temperature (from VF05) is shown with a solid red line.
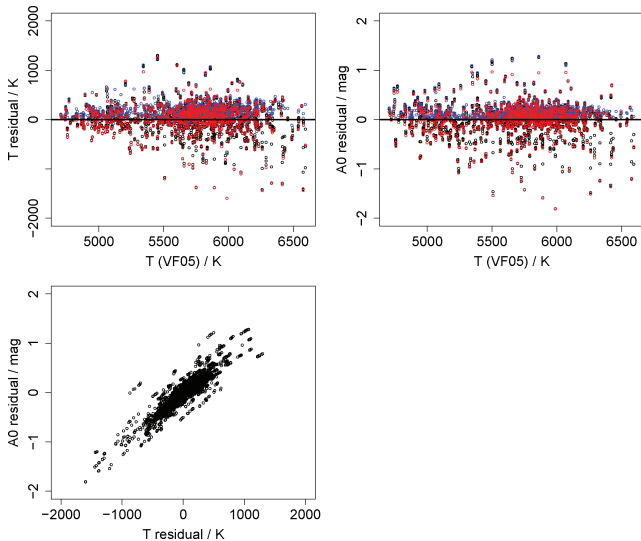

Figure 12. AP residuals for the pq-model applied to the VF05 data. The residual is defined as pq-model prediction minus VF05 ('true') value. Points with true A_0 equal to 0.0 and 2.5 mag are coloured blue and red, respectively.

250 K which may be compared to the internal errors on the training data of 50 K.

Table 2 also quantifies the improvement in the precision (smaller |CI|). Interestingly, the pq-model is more conservative than the p-model. This is measured by the *confidence*, the ratio of accuracy to precision, defined here as $\mathrm{Con} = \langle |\delta\phi_i|/\mathrm{CI}_i \rangle$. Precision has not improved as much as the accuracy, so the pq-model has slightly lower confidence.

The HRD/q factor improves AP estimation significantly. In 75 per cent of cases both A_0 and T are determined more accurately by the pq-model (there is no dependence on T or A_0). Fig. 13 shows the shift in the estimated AP from the p-model to the pq-model.

Table 2. Performance of four parameter estimation methods on the extended catalogue. ϵ_{mar} (accuracy) is the mean absolute residual; CI (precision) is the average 90 per cent confidence interval (multiply by 0.3 to get the equivalent Gaussian 1σ error); Con (confidence) is the mean of the absolute ratio of residual to precision; ϵ_{sys} (systematic) is the mean residual; SE(ϵ_{sys}) (standard error in systematic) is $\epsilon_{\mathrm{mar}} \times 1.25/\sqrt{N}$, where N is the number of objects in the test set.

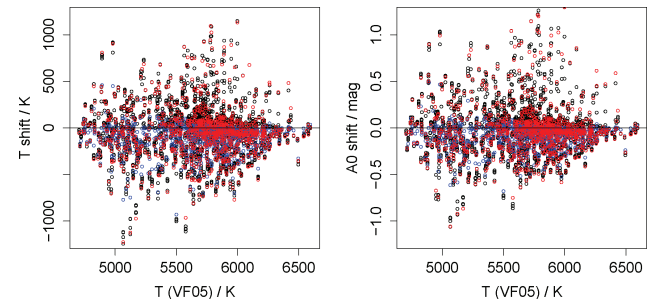| Parameter	Method	ϵ_{mar}	|CI|	Con	ϵ_{sys}	SE(ϵ_{sys})
$\log T$	p-model	0.024	0.055	0.71	0.0075	0.0004
$\log T$	pq-model	0.015	0.038	0.54	0.0037	0.0003
$\log T$	ph-model	0.020	0.051	0.67	0.0026	0.0004
$\log T$	ILIUM	0.026	0.071	0.36	0.0032	0.0005
A_0	p-model	0.29	0.63	0.88	0.075	0.005
A_0	pq-model	0.19	0.43	0.63	0.037	0.003
A_0	ph-model	0.25	0.59	0.85	0.020	0.004
A_0	ILIUM	0.38	0.92	0.41	0.079	0.008

Figure 13. Shift in mean estimated AP (pq-model minus p-model), plotted against the true (VF05) temperature. Points with true A_0 equal to 0.0 and 2.5 mag are coloured blue and red, respectively.

Shifts are both positive and negative. However, there is a slight tendency, especially at lower T, for the pq-model to give lower estimates of both A_0 and T (and in general these are also more accurate). So at least in this application, the HRD/q factor tends

to lower the AP estimates compared to a pure photometric-based estimation.

Although the pq-model yields a systematic error (ϵ_{sys}) half as large as the one from the p-model, it is still statistically significant at around 10σ (compare the values in the last two columns of the table). As already mentioned in the context of the p-model results, this may be a consequence of small inaccuracies in the MARCS models. It may also (or instead) be a result of the choice to use the mean of the one-dimensional marginalized PDFs as the estimated AP values. One could make a constant, global correction for this systematic, which is slightly less than 1 per cent in T and 0.04 mag in A_0 for the pq-model.

As a baseline comparison, Table 2 also shows the summary performance of the ILIUM algorithm, introduced by Bailer-Jones (2010a). ILIUM performs an iterative, local, non-linear interpolation of the d-grid via the forward model to estimate the parameters. It essentially tries to locate the peak of a likelihood function. Compared to the p-model, we see that ILIUM has a similar performance on T but is considerably worse in A_0. [The ILIUM confidence intervals have been estimated using the covariance transformation – equation (11) in Bailer-Jones 2010a.] As explained in the ILIUM paper, the algorithm does not explicitly take into account the degeneracy between the parameters, so this might be a reason for its inferior performance.

3.7 Interpretation of the HRD/q factor

We can understand what the HRD/q factor is doing by considering trial solutions in the HRD in Fig. 9. Consider a star with $q = 0$ and $\sigma_q = 0.1$. If extinction were zero, this would constrain M_V to be 5 ± 0.1 and the HRD prior then constrains T to have a distribution given by the corresponding horizontal slice through the HRD. The likelihood model, for its part, constrains A_0 and T from the SED. At zero extinction this also gives some probability distribution over T. The product of this with the first T distribution gives the combined probability distribution over T. It is quite possible that this is zero (the two T distributions do not overlap) in which case there is no permitted solution at zero extinction. If we consider a solution at a higher extinction, then M_V is constrained to a different range by the HRD prior and we get a different distribution over T, also from the likelihood model. These two one-dimensional distributions over T are just the 'HRD/q factor' and 'likelihood' terms in equation (15) with fixed A_0. The entire two-dimensional PDF is constructed by considering all A_0.

The posterior PDF in equation (15) can also be calculated when we have no parallaxes yet still want to use the HRD to constrain solutions. In that case, we simply set the q constraint to a constant and recalculate the integral. For want of a name let us call this the *ph-model*. The results of applying this to the extended catalogue are also shown in Table 2. Not surprisingly, its performance lies between that of the p-model and the pq-model. Yet, it is encouraging that the HRD prior alone still brings improvement over use of just the colours, by about 13 per cent in both parameters. This is relevant for classification projects which have no distance information.

3.8 Impact of metallicity

The position of the stellar locus in the HRD is quite sensitive to the stellar metallicity, a lower metallicity shifting the main sequence down/left in the HRD (to higher effective temperatures) as can be seen in Fig. 8. For the experiments discussed so far, adopting solar metallicity for the HRD prior is reasonable, because this matches

the atmospheric metallicities inferred by VF05. It is none the less instructive to investigate the impact of adopting a different metallicity. I reran the pq-model using a prior built with stars of 10-times lower metallicity ($Z = 0.0019$, or [Fe/H] = -1). This gives systematically higher estimates of the parameters compared to the original pq-model. These shifts are systematic errors with respect to the true parameter values. Quantitatively, the shifts are $\epsilon_{sys} = +0.03$ dex in $\log T$ (or 400 K in T) and $\epsilon_{sys} = +0.35$ mag in A_0. (The precisions are now also worse – larger |CI| – so at least the model recognizes there is some kind of problem.) These systematic errors degrade the overall performance to be worse than the p-model (which does not use the HRD at all). It may be obvious, but introducing additional information into an inference will degrade performance if that information is erroneous.

If we really had no prior estimate of the metallicity (and cannot estimate it from the photometry), then we should play conservative and build an HRD prior covering a range of metallicities. I have done this by combining the same stellar evolution data at the four metallicities $Z = (0.0001, 0.0019, 0.019, 0.3)$ with relative weights of $(0.1, 1, 1, 0.1)$, used in order to down weight the highest and lowest metallicities. To accommodate the large gaps between the loci (Fig. 8) I simply adopted an eight-times larger kernel bandwidth for the smoothing than was used to create the original HRD prior in Fig. 9 (there are better ways to do this in a real application). The resulting HRD prior is very broad. It is therefore not surprising that the results of the pq-model based on this are very similar to the p-model results. Accuracy and precision are only improved by a few per cent over the p-model. This is just the behaviour we expect from a prior: when the extra information it adds is weak compared to the data, it hardly alters the inference.

To use this HRD prior we should ideally have some estimate of the metallicity which is significantly better than 1 dex. This is virtually impossible with *BVJHK* photometry when we cannot restrict T and A_0 a priori. In contrast, when A_0 is known to be less than 1 mag, it is possible to estimate [Fe/H] from *griz* photometry to an (mean absolute) accuracy of 0.4 dex for stars with T between 4000 and 10 000 K, although at lower metallicities there are even larger systematic errors due to the lack of metallicity signature (C. Liu, private communication). For *Gaia* it should be possible to estimate metallicity to sufficient accuracy from the low-resolution spectroscopy for the brighter hundred million stars: Bailer-Jones (2010a) has shown that even for a broad prior extinction range (0–10 mag), [Fe/H] can be estimated to an accuracy of 0.5 dex at $G = 15$ mag. In situations where individual metallicity estimates are not possible, we should make use of any prior information about the population.

4 APPLICATION TO 85 000 *HIPPARCOS/* *2MASS STARS*

In the previous section I constructed a forward model based on real data extended to show variance in A_0. I will now use the same model, as well as the same HRD prior and q constraint model, to calculate posterior PDFs for *Hipparcos* stars which have parallaxes and *BVJHK* measurements but unknown A and T values.

In order to avoid extrapolating the forward model too far I only calculate the posterior PDF over the parameter range of the d-grid (4300–7000 K, 0.0–3.5 mag). As the *Hipparcos* sample is not a priori limited to this range – we would in particular expect it to include hotter stars – we cannot expect the model to give good results on stars which lie beyond it. Some such stars will have photometry outside the range covered by the d-grid and so will yield zero values

of the likelihood over the whole d-grid. I refer to these as *invalid stars*. Of the remaining stars, some may give non-zero values of the likelihood, typically at the edge of the d-grid, but could still have true parameters beyond the bounds of the d-grid. I shall refer to these as *exterior stars*. We could attempt to identify and remove these stars based on the input data, for example with a colour cut. However, this would not take into account the q data, may not remove all exterior stars (or may remove valid stars) and could anyway not be used efficiently with higher dimensional SED data because of the more complex mapping between data and parameters. I therefore implement a simple way of identifying exterior stars by identifying when their marginalized PDF is strongly truncated at the edge of the grid (explained below).

4.1 Catalogue cross-matching and data selection ('h2m catalogue')

I take parallaxes from the new *Hipparcos* catalogue of van Leeuwen (2007a,b), which has 117 955 entries. Using these positions and proper motions, I cross-matched these stars using the Gator tool with the 2MASS catalogue (with a 2 arcsec search radius) to obtain the *JHK* photometry. Finally I used the *Hipparcos* identification numbers to extract from SIMBAD the *BV* data and (for inspection purposes only) the spectral types (where available). I retain only those stars which have: complete photometric and astrometric data; best photometric quality in 2MASS (especially signal-to-noise ratio > 10, ph_qual = 'AAA', cc_flg = '000'); $\varpi \geq 0.1$ mas. This results in a catalogue of 85 608 stars which I shall refer to as the *h2m catalogue* (*Hipparcos*/2MASS catalogue). The 90 per cent quantile ranges are 6.8–10.2 mag in *V*, 1–21 mas in ϖ and −6.4 to 0.4 mag in *q*. The uncertainties in *JHK* are taken from the 2MASS catalogue (most lie between 0.02 and 0.05 mag). I assume 0.02 mag for the uncertainties in the *B* and *V* bands. The correlations in the colours formed from these are again taken into account in the covariance matrix in the likelihood (see Appendix).

4.2 Results of the p-model and pq-model

I first applied the p-model to the full h2m catalogue. The posterior PDFs exhibit the same A_0, T degeneracy we saw with the previous data set. Applying the pq-model, we again see that the HRD/q factor reduces the size of the posterior PDF regions, i.e. increases the precision of the estimates, although by a much smaller degree than with the previous data: the mean 90 per cent confidence interval, |CI|, reduces from 0.032 to 0.026 dex for $\log T$, and from 0.36 to 0.29 mag for A_0. Of the 85 608 stars in the h2m catalogue, 7406 (9 per cent) yielded no solution in the either model because they gave zero likelihood at every point in the d-grid (the *invalid stars*).

As before I adopt as the single best parameter estimates for A_0 and T the means of the one-dimensional marginalized PDFs. Fig. 14 shows the distribution of these estimates for the two models. The vast majority of stars are assigned low extinctions by both models. We also see a paucity of stars at intermediate temperatures. As expected, there are many solutions lying at the edge of the plot, which are the d-grid boundaries. These stars generally have marginalized PDFs which peak at the edge of the grid and so are effectively truncated by it (similar to the situation in the right-hand column of Fig. 5). These probably have true APs which lie outside the grid and therefore are beyond the range of applicability of these models. I identify such a star as an *exterior star* if (for either parameter) (a) the peak of the PDF is at the edge of the d-grid, or if (b) the mean of the PDF lies within half of the 90 per cent
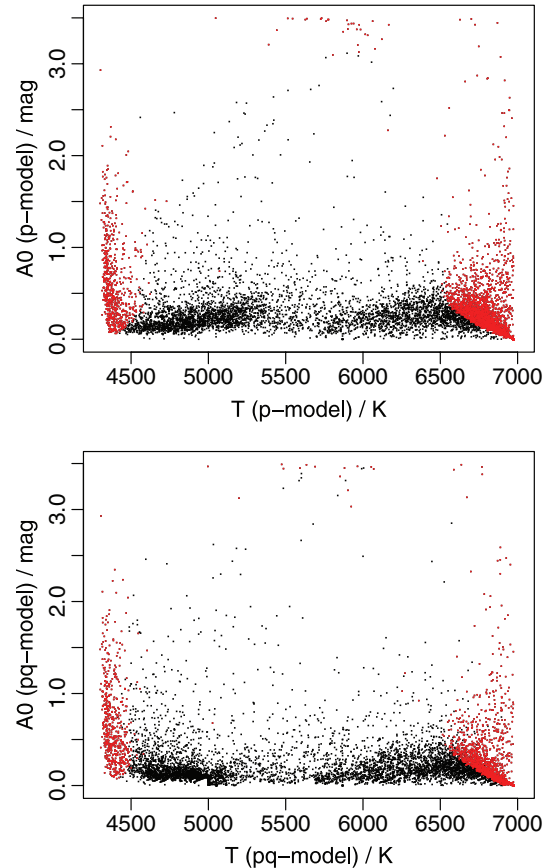

Figure 14. Estimated mean parameters for the stars in the h2m catalogue inferred from the p-model (top) and pq-model (bottom). To avoid crowding, these are results based on a randomly selected 12 per cent of the h2m catalogue, but the general features are present in the full sample. The stars plotted in red are the exterior stars (those with likely true parameters out-of-bounds of the model).

confidence interval of the edge of the d-grid. As A_0 must be strictly positive, this condition is not applied to the $A_0 = 0$ mag edge of the grid. In this way 39 678 and 31 542 exterior stars are identified in the p-model and pq-model results, respectively, which is 51 per cent and 40 per cent, respectively, of the valid stars. These are overplotted as red points in Fig. 14. These stars are removed from the subsequent analysis and from the final catalogue provided, leaving 38 524 stars from the p-model and 46 660 from the pq-model.

As we do not have 'true' parameter estimates for the h2m catalogue I cannot report accuracies for the estimates, so we must assess the results in a different way. Given the mean parameter estimates I use equation (19) to calculate A_V and then calculate the absolute magnitude for each star (from equation 6).[5] The resulting HRD for the h2m stars is shown in Fig. 15 for parameters from the p-model (upper panel) and pq-model (lower panel). This plot does not reflect the relative densities in the HRD very well, so it is redrawn (using the full sample) as a density scale plot in Fig. 16 (note that a smaller M_V scale is used). Both HRDs show a clear separation between a main sequence, dominated by stars between 5700 and 6700 K, and a giant branch at around 4500–5000 K. (We likewise see two distinct populations with different colours in colour–colour diagrams.) The

---

[5] A passionate Bayesian would propagate PDFs to the end, marginalize over A_0 and T and then calculate the mean of $P(M_V | \boldsymbol{p}, q)$.
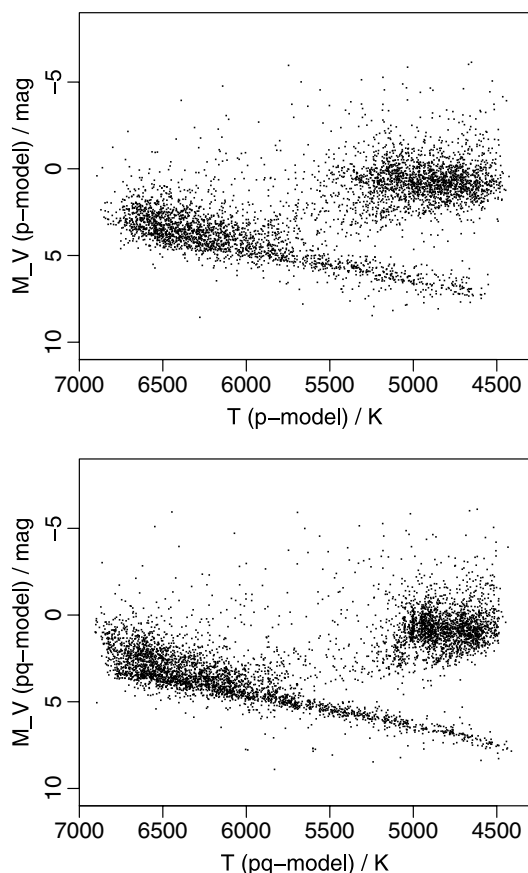
Figure 15. HRD for the h2m catalogue derived from the p-model (top) and pq-model (bottom), excluding the exterior stars. To avoid crowding, these are results based on the randomly selected 12 per cent of the h2m catalogue.

Figure 16. HRD for the h2m catalogue (excluding exterior stars) derived from the p-model (top) and pq-model (bottom) shown as a density plot (achieved via smoothing with a Gaussian kernel). The number of stars per unit area is normalized to a value of 1.0 at the maximum density (separate normalization in each plot).

HRDs are quite similar overall, but there are important differences. The main sequence in the pq-model has a narrower M_V spread for stars cooler than solar, but a broader spread for hotter stars. The giant branch in the pq-model has a narrower temperature spread and also displays two density maxima (discussed in the next section). An order of magnitude check on the correctness of these HRDs can be made by considering the position of typical stars. The Sun, for example, has $T = 5800$ K and $M_V = 4.8$ mag, which would correctly place it on the main sequence in both models.

If we now apply the ph-model to these data (i.e. we use the HRD prior but not the q factor), then the resulting HRD density map is actually quite similar to that arising from the p-model. This proves that most of the difference in the HRD morphology we see in Fig. 16 comes from the information provided by the parallax in combination with the HRD, rather than from the HRD alone.

Fig. 17 shows the HRD using a colour scale to represent the mean extinction at each point in the diagram. We see only a few regions with high extinctions, predominantly stars below the main sequence. Fig. 18 shows the same but increasing the dynamic range for the low extinction stars (0–0.5 mag), which are 90 per cent of stars in both models.

4.3 Discussion

SIMBAD provides spectral types for the majority of the h2m stars. These inevitably come from heterogeneous sources, yet provide a rough check on the inferred temperatures. The spectral types are
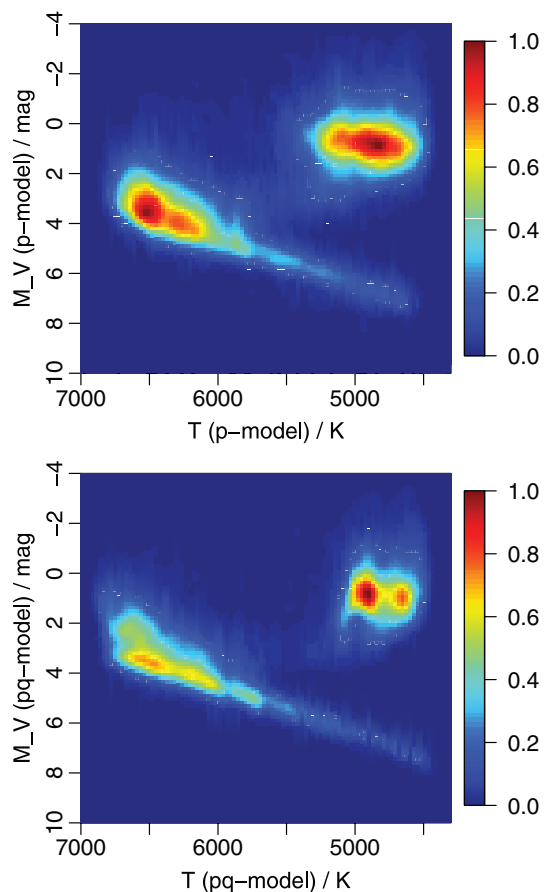
provided as character strings, sometimes with uncertain or intermediate types such as F6/F7V, in which case I just take the first spectral type. The distribution of spectral types in the whole h2m catalogue is shown in the upper panel of Fig. 19. Removing the invalid and exterior stars for the pq-model leaves those plotted in the lower panel. As expected this has removed predominantly early-type (hot) stars. Fig. 20 shows the relationship between these spectral types and the inferred T estimates for this model. The correlation is reasonable, especially when we consider that there is anyway no tight relationship between T and spectral type a priori. As comparison, this figure also shows the average T–spectral type calibration for dwarfs and giants from Bailer-Jones et al. (1997) for the Galactic thin disc at [Fe/H] $= -0.2$ dex (there is a scatter of around ± 300 K at each spectral type in the calibration). The agreement is reasonable, supporting the conclusion that the T assignments from the pq-model are appropriate.

SIMBAD also supplies a luminosity class for about half of the h2m stars. We can use these to make a check on the apparent clear separation between a main sequence (dwarf stars) and giant branch (giant stars) in Fig. 16. For this purpose I make a simple classification of the stars as giants if both $T < 5250$ K and $M_V < 4$ mag, and as dwarfs otherwise. For the pq-model this classifies 39 per cent of the stars as giants. Of these, 96 per cent of those which have a luminosity class are class III, II/III or III/IV, i.e. a giant class.
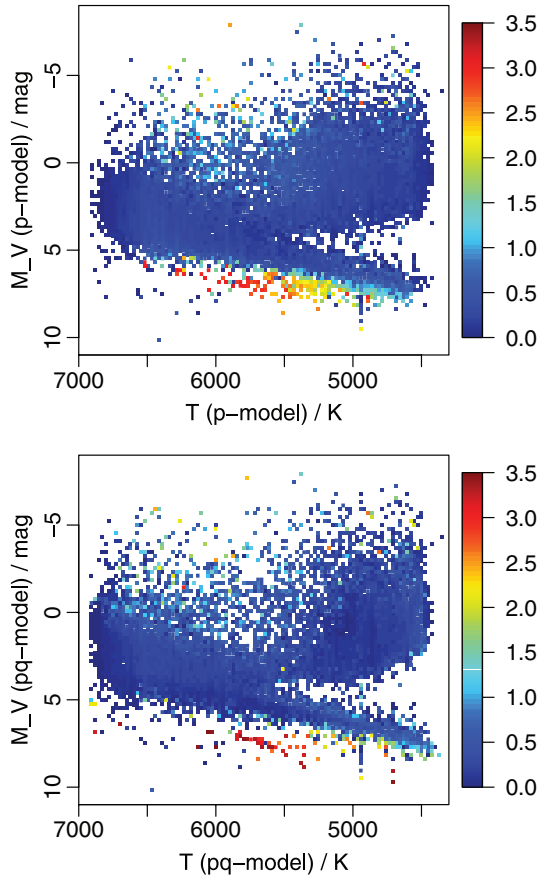
Figure 17. HRD for the complete h2m catalogue (excluding exterior stars) derived from the p-model (top) and pq-model (bottom), showing the mean extinction (A_0 in magnitudes) on a colour scale. Unoccupied areas are shown in white.

Figure 18. As Fig. 17 but showing a narrower range of low extinctions (0–0.5 mag). Cells of higher mean extinction are not plotted (about 10 per cent of stars in both models).

Of the dwarfs, 93 per cent of those which have a luminosity class are class V or IV/V, i.e. a dwarf class. Assuming we have roughly correct temperature assignments, this suggests that the extinction estimates cannot be so wrong as to have mixed up dwarfs and giants in the cooler part of the HRD.

The inferred extinction distribution of the stars shown in Fig. 14 is consistent with the three-dimensional extinction model of Vergely et al. (1997). This model, which is based on *Hipparcos* parallaxes with extinctions estimated from Strömgren photometry, gives an average extinction (A_V) in the Galactic plane in the solar neighbourhood of 1.5 mag kpc^{-1} (with a vertical scaleheight of 70 pc). This is compatible with the present result: most of the h2m stars have extinctions below 0.5 mag, and most lie between a distance of 50 pc and 1 kpc.

In both panels of Fig. 17, we see some highly extinct stars below the main sequence. In the pq-model there are 41 such stars with $M_V > 5$ mag and $A_0 > 3.0$ mag. They are all very nearby (4–18 pc) and faint ($V = 9.6$–12.0 mag) and have very red $V - J$ colours: 3.4–4.2 mag for 37 of them, whereas only 0.2 per cent of the valid, non-exterior stars in the pq-model have such red colours. This is at the limit of the colours used to fit the forward model (Fig. 3). These stars also have very late spectral types of M0–M4.5 from SIMBAD. Assuming these spectral types to be correct, this would correspond to effective temperatures between about 3800 and 3000 K (Leggett et al. 2000). This is well below the lower limit of the extended catalogue (4707 K), so the inference would depend on a significant
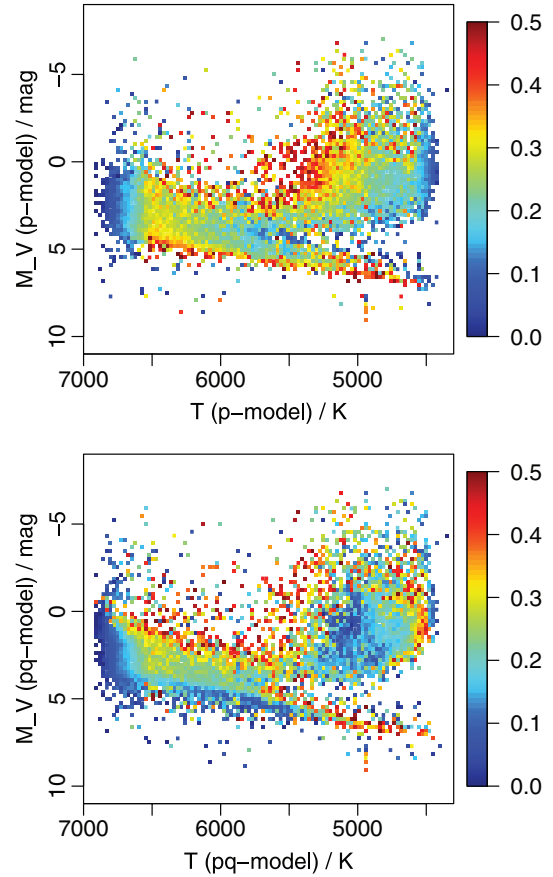
extrapolation of the forward model. As these colours and APs are not represented in the model training we cannot expect a reliable inference, so I conclude that the pq-model and p-model AP assignments are incorrect. Ideally, these stars would have been identified as 'exterior' stars, yet it is unsurprising that some slipped through the somewhat ad hoc criteria adopted.

We saw in Section 3.8 that adopting an HRD prior with a lower metallicity than the true metallicity resulted in a systematic increase (error) in both A_0 and T. A priori we expect some of the *Hipparcos* stars to have subsolar metallicity, in which case the use of a solar metallicity HRD prior for these stars is expected to yield the converse, that is, erroneously low A_0 and T estimates. A lower extinction corresponds to a larger intrinsic magnitude on account of the q constraint, although for the 10 times lower metallicity used in Section 3.8 the change in A_0 and therefore in M_V was only 0.3 mag on average.

I investigated the impact of metallicity empirically by rerunning the pq-model using the 10 times lower metallicity prior. The resulting HRD is shown in Fig. 21 and should be compared with the lower panel of Fig. 15. We do indeed see that the main sequence has been shifted to a slightly larger M_V. The main sequence is also not as tight as before and the upper part of the main sequence has a different distribution. There is also a significant change in the morphology of the giant branch. In particular it now extends to higher temperatures.

We saw in Fig. 16 that the pq-model HRD exhibits a bimodality in the giant branch. This might reflect two genuine populations with
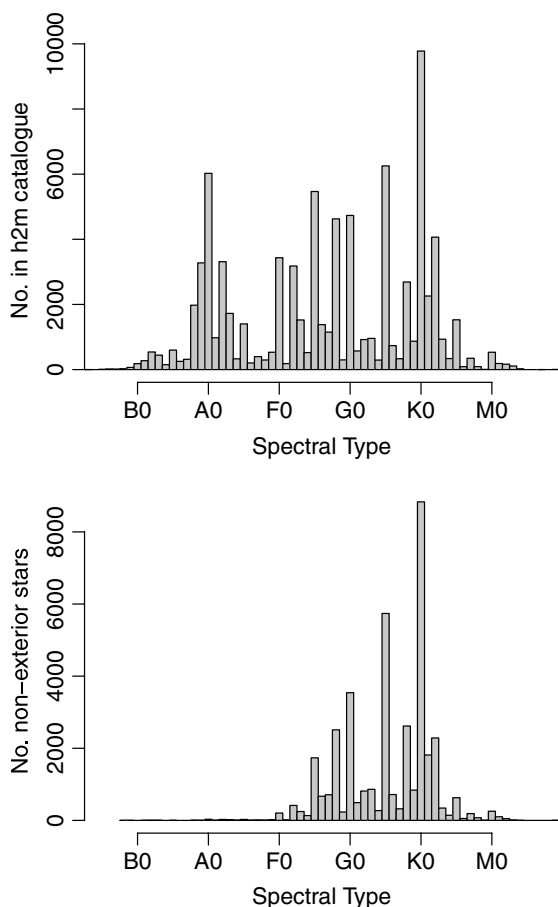
Figure 19. Distribution of SIMBAD spectral types for all stars in the h2m catalogue (top) and for the valid, non-exterior stars as identified by the pq-model (bottom).

different metallicities, as was inferred for the *Hipparcos* sample by Girardi et al. (1998).[6] Assuming one population has near-solar metallicity and the other sub-solar metallicity, the latter would be the bluer (higher T) of the two. However, as just discussed, the use of the solar metallicity HRD prior results in such stars being assigned erroneously low A_0 and T values. (So if this higher T clump really is low metallicity, in reality it should have higher T, higher A_0 and lower M_V.) We see from Fig. 18 that the higher T clump is indeed assigned a lower A_0. Assuming that, in reality, there is no systematic change in the extinction across the giant branch, this apparently lower extinction may in fact be a result of an extinction–metallicity confusion (degeneracy) introduced by using an erroneous metallicity in the HRD prior in combination with the A_0–T degeneracy. This again underlines the importance of estimating metallicity from the SED.

4.4 Catalogue

A catalogue of the AP estimates for 46 900 stars is available online with the electronic version of this article (see Supporting Information) and from CDS Strasbourg. 38 524 stars have APs from the

---

[6] The bimodality is not seen in the p-model HRD, perhaps because the APs are less precise (and presumably less accurate) in that model, or perhaps because the HRD/q factor is providing a better discrimination between the two populations.
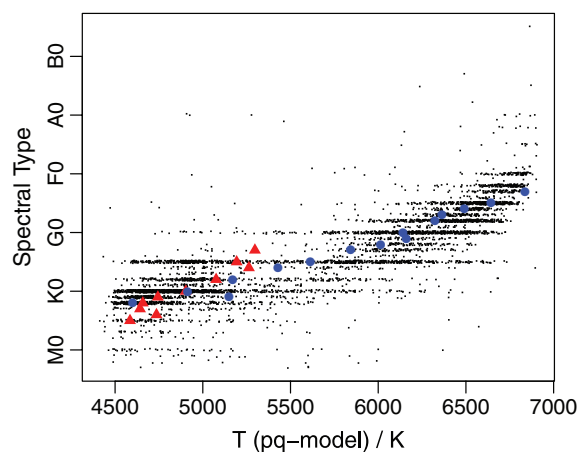
Figure 20. Relationship between SIMBAD spectral types and effective temperature inferred by the pq-model for the h2m catalogue. To avoid crowding, results for only the randomly selected 12 per cent of the catalogue are shown, plus points have been jittered by $\pm\frac{1}{4}$ spectral types. Blue circles and red triangles show the calibration for dwarfs and giants (respectively) from Bailer-Jones et al. (1997).
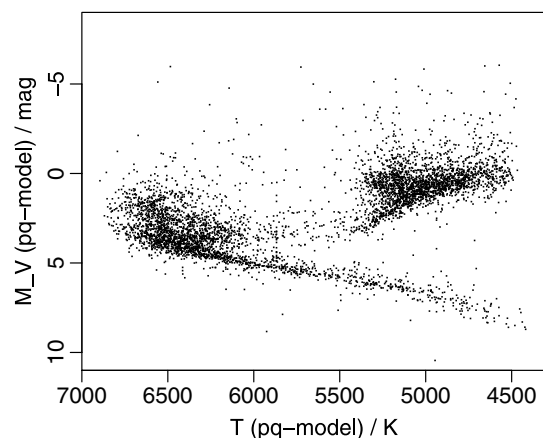

Figure 21. HRD for the h2m catalogue derived from the pq-model using the HRD prior with $Z = 0.0019$, i.e. 10 times below solar metallicity. Exterior stars are excluded.

p-model and 46 660 have APs from the pq-model (most have APs from both). T spans 4350–6900 K and A_0 0–3.45 mag. A sample of the catalogue is show in Table 3 with the columns defined in Table 4. Columns 2–13 are the estimates from the present work. NA is used to indicate where either the p-model or pq-model does not provide parameter estimates (i.e. it is an *exterior star*, as defined in Section 4; *invalid stars* are excluded from the catalogue). Columns 1, 22 and 23 are from the catalogue of van Leeuwen (2007a), Columns 14 and 15 are from SIMBAD and Columns 16–21 are taken from the 2MASS catalogue.

To convert the extinction parameter, A_0, to the extinction in the V band, A_V, one may use the following quadratic approximation to the function $y(A_0, T)$ in equation (19)

$$y(A_0, T) \simeq -5.376 + 2.884(\log T) - 0.4217 A_0$$
$$- 0.3865(\log T)^2 - 0.00374 A_0^2 + 0.1072(\log T) A_0. \quad (20)$$

The rms error of this fit is 0.0025 mag over the fitted parameter range 4000–7000 K and 0–5 mag in A_0.

Table 3. A sample of the output catalogue (rows 5000–5010). The columns are defined in Table 4. NA stands for 'not available', i.e. it is an exterior star in that model. The full version is available online (see Supporting Information).

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23
10882	3.7291	3.6982	3.7607	0.39	0.00	0.77	3.6998	3.6938	3.7040	0.02	0.00	0.10	9.71	8.86	7.178	0.021	6.738	0.029	6.588	0.017	2.67	0.80
10883	NA	NA	NA	NA	NA	NA	3.7865	3.7591	3.8107	0.34	0.03	0.61	10.69	10.14	8.866	0.023	8.519	0.042	8.460	0.021	9.56	1.28
10886	3.7540	3.7313	3.7801	0.23	0.00	0.52	3.7588	3.7334	3.7835	0.28	0.00	0.57	8.84	8.10	6.736	0.025	6.381	0.027	6.261	0.016	12.61	0.89
10887	3.6790	3.6606	3.7008	0.19	0.00	0.48	3.6739	3.6597	3.6943	0.12	0.00	0.41	9.61	8.51	6.601	0.021	6.079	0.026	5.933	0.022	2.47	1.14
10888	3.8210	3.7980	3.8404	0.24	0.00	0.46	3.8181	3.7963	3.8380	0.20	0.00	0.43	9.48	9.02	8.056	0.043	7.871	0.031	7.815	0.020	6.31	1.19
10890	NA	NA	NA	NA	NA	NA	3.8227	3.7988	3.8385	0.27	0.01	0.45	7.09	6.65	5.665	0.018	5.452	0.026	5.440	0.021	15.93	0.56
10891	3.7094	3.6794	3.7360	0.40	0.01	0.74	3.6929	3.6759	3.7036	0.18	0.00	0.35	10.43	9.48	7.630	0.020	7.119	0.026	7.021	0.021	1.88	0.98
10894	3.7397	3.7051	3.7721	0.43	0.01	0.82	3.7456	3.7045	3.7780	0.50	0.01	0.88	10.40	9.56	7.939	0.027	7.511	0.055	7.370	0.027	2.63	1.48
10895	3.6650	3.6395	3.6853	0.35	0.01	0.62	3.6661	3.6452	3.6841	0.36	0.10	0.61	9.23	8.00	5.863	0.017	5.290	0.016	5.159	0.018	3.81	0.96
10899	3.6746	3.6596	3.6881	0.16	0.00	0.35	3.6706	3.6592	3.6852	0.10	0.00	0.31	8.84	7.72	5.804	0.018	5.275	0.018	5.118	0.017	4.12	0.48
10900	3.6690	3.6556	3.6823	0.11	0.00	0.29	3.6672	3.6562	3.6770	0.09	0.00	0.23	9.12	7.99	6.079	0.020	5.516	0.024	5.369	0.020	3.79	0.89

Table 4. Definition of the columns in the output catalogue (CI stands for 'confidence interval').

1	*Hipparcos* identifier
2	$\log T/K$ from the p-model (mean estimate)
3	Lower bound of the 90 per cent CI to $\log T/K$ from the p-model
4	Upper bound of the 90 per cent CI to $\log T/K$ from the p-model
5	A_0/mag from the p-model (mean estimate)
6	Lower bound of the 90 per cent CI to A_0/mag from the p-model
7	Upper bound of the 90 per cent CI to A_0/mag from the p-model
8	$\log T/K$ from the pq-model (mean estimate)
9	Lower bound of the 90 per cent CI to $\log T/K$ from the pq-model
10	Upper bound of the 90 per cent CI to $\log T/K$ from the pq-model
11	A_0/mag from the pq-model (mean estimate)
12	Lower bound of the 90 per cent CI to A_0/mag from the pq-model
13	Upper bound of the 90 per cent CI to A_0/mag from the pq-model
14	*B* magnitude
15	*V* magnitude
16	*J* magnitude
17	1σ uncertainty in the *J* magnitude
18	*H* magnitude
19	1σ uncertainty in the *H* magnitude
20	*K* magnitude
21	1σ uncertainty in the *K* magnitude
22	Parallax/mas
23	1σ uncertainty in the parallax/mas

5 ASSUMPTIONS AND POSSIBLE EXTENSIONS

The method as presented could be improved and extended in a number of ways.

The range over which we can estimate APs is set by the range of validity of the forward model, which in turn depends on the labelled templates used to fit it. I focused on F, G, K stars. This could be extended using additional data at higher temperatures, although we must be careful to ensure that they are parametrized on a consistent temperature scale. Preliminary predictions of the method's performance on simulated *Gaia* spectrophotometry and astrometry for stars over a wider A_0 and T range are reported in Bailer-Jones (2010b).

The development of the method here has assumed for simplicity that stars are described by just three APs: T, A_0 and M_V. The SED ($\boldsymbol{p}$) is assumed to depend on only the first two of these, with dependence on M_V introduced by the measured apparent magnitude and parallax (q). I only derived PDFs over T and A_0, choosing then to simply derive M_V using equation (6), but we could also derive a PDF over M_V.

Given higher resolution data, we can usually estimate additional parameters from the SED, in particular the metallicity, [Fe/H], and the surface gravity, $\log g$. In the demonstrations above, the p-model essentially assumes all stars to be solar metallicity dwarfs, because this is what was used to fit the forward model. Although this is not correct for the h2m data, it is none the less acceptable for the broad-band colours considered here because they show little sensitivity to either metallicity or surface gravity. When it comes to the pq-model, however, we have seen that the assumed metallicity, Z, of the HRD does have a significant impact on the inferred T and A_0, so we should introduce some dependence on Z in the method. If we replace T with (T, Z) in equation (10) and follow the derivation through, we arrive again at equation (15) but with an additional Z in the conditioned terms (to the right of the '|') for the likelihood and the HRD prior. It does not explicitly appear in the q constraint because once conditioned on M_V, Z adds no further information. For

the likelihood term we would estimate atmospheric [Fe/H] and then convert to Z assuming some fixed abundance pattern (although in principle we could introduce the Helium abundance, Y, as another parameter in the HRD). Estimating Z to around 0.5 dex or better is equivalent to using a 'narrower' HRD prior, which in turn would lead to much more precise estimates of T, A_0 and M_V.

We can likewise extend the method to estimate the surface gravity, $\log g$, which is often estimated from higher resolution spectroscopy and would help to constrain M_V further. $\log g$ would appear in the same way as Z in equation (15), but unlike Z it is not an independent parameter in the HRD, but rather is determined by M_V and T.

For a large, deep survey like *Gaia* it may be of particular importance to permit the selective extinction parameter, R_0 to vary, as it is known to vary across the Galaxy (e.g. Patriarchi, Morbidelli & Perinotto 2003). This could easily be introduced via simulation in the same way that A_0 variance was introduced in Section 3.1. An R_0 dependence is introduced into equation (15) by replacing A_0 with (A_0, R_0). To be effective we need to be able to estimate R_0 from the SED. First investigations suggest that this will be possible with the *Gaia* low-resolution spectroscopy (C. Liu, private communication), although the degeneracies with A_0 and T are yet to be fully characterized.

Finally, the method as presented is predicated on all stars being single. A physical binary comprising two identical stars has M_V which is 0.75 mag lower than a single star of the same T. Such binaries could be accommodated using an adjusted HRD and introducing an extra AP, s, the probability that the star is single. To solve for the other APs (the common T and A_0) we would marginalize over s, adopting the single-star HRD when $s = 1$ and an HRD shifted by 0.75 mag when $s = 0$. We can likewise infer $P(s|\boldsymbol{p}, q)$. The situation is more complex when we allow for more general types of binary, but can be helped if we can make some estimates of the effective temperatures of the two components from their composite spectrum. We have had some limited success in this with simulated *Gaia* spectra (P. Tsalmantza, private communication).

In all of this work, there is choice in the nature of the HRD prior (just as there is a choice in what SED we observe). The HRD prior was adopted primarily to eliminate the 'forbidden' regions of the HRD (the white regions in Fig. 9), rather than to introduce specific dependence on the IMF or star formation rate used to construct it. All pq-model results were for a power-law (Salpeter) IMF, but we actually get very similar results with a flat IMF (which is not that surprising, given the relatively narrow T and hence mass range).

It is worth noticing that the HRD prior is a prior on the absolute magnitude and not on the observed magnitude. It therefore does not take into account the magnitude limit of the survey from which the data have been obtained. This could be potentially useful information. For example, the limiting magnitude may be such that we do not expect to find very many intrinsically faint white dwarfs or brown dwarfs.

The whole approach adopted has been to estimate APs for single stars by considering their line-of-sight extinction as an intrinsic parameter. This ignores any correlation in the extinction between neighbouring stars. Yet in many parts of the Galaxy it may be reasonable to assume that the density of interstellar material, and hence the line-of-sight extinction to that point, varies smoothly with three-dimensional position. We could introduce this smoothness constraint through a hierarchical inference procedure in which we solve for the APs of a set of N stars simultaneously. The idea is to parametrize the extinction as a function of (known) spatial

position, $A_0(l, b, \varpi; \boldsymbol{\beta})$, where l, b are the Galactic coordinates and $\boldsymbol{\beta}$ are the free parameters of this function. Rather than solving for N two-dimensional PDFs over (T, A_0) for each star independently, we solve for the much higher dimensional PDF over $(T_1 \ldots T_N, \boldsymbol{\beta})$. This is computationally far more complex, as it involves N non-independent integrals like that in equation (15). But it constrains the extinction to vary in a physically plausible way, plus it would help to estimate extinction in those cases where it may otherwise be very difficult. Such a large-scale inference is not unthinkable for a survey such as *Gaia*.

Let us finally consider the method's philosophy of separating the observed data into independent terms $\boldsymbol{p}$ and q. This leads to an astrophysically meaningful and useful separation of the terms in equation (15), in which we can simply use the HRD/q factor to 'update' the AP estimates obtained from the pure SED (equation 17). The price we pay for this is the need to normalize the SED (achieved here by forming colours). This introduces correlations between the elements of the SED, although this was easy to accommodate in the likelihood function in the present case. If we were willing to sacrifice the separability, then we could adopt a more purist approach and operate one step closer to the raw data directly on the fluxes and the parallax. This would be of little benefit for *Gaia*, however, because the parallax and the apparent (G-band) magnitude are derived from the same measurements so are anyway correlated. Furthermore, the planned processing of the *Gaia* SED (the 'BP/RP spectrum') will probably also result in correlated fluxes.

6 CONCLUSIONS

The main conclusions of this work and features of the method are as follows.

(i) There exists a significant degeneracy between A_0 and T when estimated from broad-band optical/infrared data.

(ii) The method developed here allows one to make self-consistent and quantitative use of the HRD prior and parallaxes in order to improve the accuracy of the AP estimation beyond using just the colours (by 35 per cent overall with the extended catalogue).

(iii) Even if parallaxes are not available, use of the HRD prior improves AP accuracy compared to using just the colours (by 13 per cent overall with the extended catalogue).

(iv) The method gives a multidimensional posterior density distribution over the APs and can easily be extended to include other APs and applied to other (spectro)photometric data.

(v) The method ensures that all derived parameters are self-consistent. It takes into account uncertainties and covariances in the data.

(vi) Accurate estimation of metallicity (to about 0.5 dex) is necessary if the use of the HRD/q factor is to improve the accuracy of the T and A_0 estimation.

The method developed in this paper has significant advantages over standard pattern recognition methods, such as neural networks or support vectors machines. These try to learn an inverse mapping from the input data to the APs, yet this would be cumbersome to achieve with these heterogeneous photometric and parallax inputs. First, we would have to ensure that the stars in the training set have consistent magnitude, parallax, SED and extinction. Secondly, we would have to train separate algorithms depending on what inputs we use (as these methods are not robust to simply omitting input variables). The method developed in this paper, in contrast, has no need for any training data on, or modelling of, the parallax or apparent magnitude (only their uncertainties are modelled). Pattern

recognition methods normally also give just a single solution and are incapable of naturally providing probability distributions (and thus expected uncertainties) over parameters. They likewise cannot easily incorporate prior information. Most significantly, however, these methods cannot explicitly take into account the constraints from physical background information we have, namely the relationship between temperature, extinction, parallax and magnitude, which we know from stellar physics and geometry. Approaches which do not take these constraints properly into account could produce unphysical solutions. Physical constraints and other prior information could only be incorporated indirectly (and probably incompletely) via some clumsy tuning of the training data set or transformation of the variables. The only genuine advantage that bare pattern recognition methods offer is speed, but the saving this offers in terms of computer power is hardly significant compared to the cost of gathering astrometric data.

REFERENCES

Bailer-Jones C. A. L., 2009, in Andersen J., Bland-Hawthorn J., Nordström B., eds, Proc. IAU Symp. Vol. 254, The Galaxy Disk in Cosmological Context. Kluwer, Dordrecht, p. 475
Bailer-Jones C. A. L., 2010a, MNRAS, 403, 96
Bailer-Jones C. A. L., 2010b, Gaia DPAC Technical note, GAIA-C8-TN-MPIA-CBJ-049
Bailer-Jones C. A. L., Irwin M., Gilmore G., von Hippel T., 1997, MNRAS, 292, 157
Bertelli G., Girardi L., Marigo P., Nasi E., 2008, A&A, 484, 815
Burnett B., Binney J., 2010, MNRAS, 407, 339
Cardelli J. A., Clayton G. C., Mathis J. S., 1989, ApJ, 329, L33
ESA, 1997, ESA SP-1200, The Hipparcos and Tycho Catalogues. ESA, Noordwijk
Fitzpatrick E. L., 1999, PASP 111, 63
Girardi L., Groenewegen M. A. T., Weiss A., Salaris M., 1998, MNRAS, 301, 149
Ivezić Z. et al., 2008, ApJ, 684, 287
Jørgensen B. R., Lindegren L., 2005, A&A, 436, 127
Leggett S. K., Allard F., Dahn C., Hauschildt P. H., Kerr T. H., Rayner J., 2000, ApJ, 535, 965
Lindegren L. et al., 2008, in Jin W., Platais I., Perryman M. A. C., eds, Proc. IAU Symp. Vol. 248, A Giant Step From Milli-to Micro-Arcsecond Astrometry, Kluwer, Dordrecht, p. 217
Mignard F., Drimmel R. (eds), 2007, DPAC Proposal for the Gaia Data Processing, Gaia DPAC Technical note, GAIA-CD-SP-DPAC-FM-030
Patriarchi P., Morbidelli L., Perinotto M., 2003, A&A, 410, 905
Pont F., Eyer L., 2004, MNRAS, 351, 487
Takeda G., Ford E. B., SIlls A., Rasio F. A., Fischer D. A., Valenti J. A., 2007, ApJS, 168, 297
Valenti J. A., Fischer D. A., 2005, ApJS, 159, 141 (VF05)
Vallenari A., Chiosi E., Sordo R., 2010, A&A, 511, 79
van Leeuwen F., 2007a, Astrophys. Space Sci. Libr. Vol. 350, Hipparcos, the New Reduction of the Raw Data. Springer, Heidelberg
van Leeuwen F., 2007b, A&A, 474, 653
Vergely J.-L., Egret D., Freire Ferrero R., Valette B., Koeppen J., 1997, in Perryman M. A. C., Bernacca P. L., Battrick B., eds, Proc. ESA Symp. 'Hipparcos – Venice '97', ESA SP-402. ESA, Noordwijk, p. 603
Worthey G., Lee H.-C., 2006, preprint (astro-ph/0604590)

APPENDIX A: COVARIANCE

Let Var() denote the variance and Cov() the covariance operators. For two random variables X and Y with arbitrary distributions, a general result is

$$\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y) + 2\text{Cov}(X, Y) \,. \tag{A1}$$

Another general result for constants a, b, c, d and random variables W, X, Y, Z is

$$\begin{aligned}
&\text{Cov}(aX + bY, cW + dZ) \\
&= ac\,\text{Cov}(X, W) + ad\,\text{Cov}(X, Z) \\
&\quad + bc\,\text{Cov}(Y, W) + bd\,\text{Cov}(Y, Z).
\end{aligned} \tag{A2}$$

From this, it follows that the covariance of two colours involving a common band, e.g. $m_1 - m_2$ and $m_2 - m_3$, assuming that each band is measured independently, is

$$\text{Cov}(m_1 - m_2, m_2 - m_3) = -\text{Var}(m_2) \,. \tag{A3}$$

The colour vector formed from the five bands *BVJHK*,

$$\boldsymbol{p} = (B - V, V - J, J - H, H - K) \,, \tag{A4}$$

therefore has covariance matrix

$$\mathbf{C}_p = \begin{pmatrix} \sigma_B^2 + \sigma_V^2 & -\sigma_V^2 & 0 & 0 \\ -\sigma_V^2 & \sigma_V^2 + \sigma_J^2 & -\sigma_J^2 & 0 \\ 0 & -\sigma_J^2 & \sigma_J^2 + \sigma_H^2 & -\sigma_H^2 \\ 0 & 0 & -\sigma_H^2 & \sigma_H^2 + \sigma_K^2 \end{pmatrix} \tag{A5}$$

where σ_i^2 is the variance in band i.

SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article:

Table 3. The output catalogue.

Please note: Wiley-Blackwell are not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.

This paper has been typeset from a TeX/LaTeX file prepared by the author.